Power-Expected-Posterior Priors for Variable Selection in Gaussian Linear Models

Ioannis Ntzoufras,

Department of Statistics, Athens University of Economics and Business, Athens, Greece; e-mail: ntzoufras@aueb.gr.

Joint work with:

Dimitris Fouskakis & David Draper

Department of Mathematics
National Technical University of Athens, Greece

Dep. of Applied Mathematics and Statistics
University of California, Santa Cruz, USA

Synopsis

1 Model Selection and the paradox

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

- response data $\boldsymbol{y}$ generated by a model $M_\ell \in \mathcal{M}$.
- Each model specifies the distribution of $\boldsymbol{y}$.
- $\boldsymbol{\beta}_\ell$ is the parameter vector for model M_ℓ.
- $f(M_\ell)$ is the prior probability of model M_ℓ.

Then posterior inference is based on posterior model probabilities

$$f(M_\ell|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|M_\ell)f(M_\ell)}{\sum_{m_k \in \mathcal{M}} f(\boldsymbol{y}|M_k)f(M_k)},$$

where $f(\boldsymbol{y}|M_\ell)$ is the marginal likelihood under model m and $f(M_\ell)$ is the prior probability of model M_ℓ.

Posterior odds and Bayes factors

Pairwise comparisons of any two models, m_k and m_ℓ , are based on the **Posterior Odds** (PO)

$$PO_{k,\ell} \equiv \frac{f(M_k|\boldsymbol{y})}{f(M_\ell|\boldsymbol{y})} = \frac{f(\boldsymbol{y}|M_k)}{f(\boldsymbol{y}|M_\ell)} \times \frac{f(M_k)}{f(M_\ell)} = B_{k,\ell} \times O_{k,\ell}$$

which is a function of the **Bayes Factor** $B_{k,\ell}$ and the **Prior Odds** $O_{k,\ell}$.

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).

⇒ Posterior density takes the shape of the likelihood and is insensitive to the exact value of the prior density function.

For multiple models inference ⇒ BFs (and POs) are quite sensitive to the choice of the prior variance of model parameters.

⇒ For nested models, we support the simplest model with the evidence increasing as the variance of the parameters increase ending up to support of more parsimonious model no matter what data we have.

⇒ Under this approach, the procedure is quite informative since the data do not contribute to the inference.

⇒ Improper priors cannot be used since the BFs depend on the undefined normalizing constants of the priors.

2 Prior Specification

Prior on the model space

- Uniform prior on the model space

$$f(M_\ell) = \frac{1}{|\mathcal{M}|}.$$

In variable selection $\rightarrow$ it is equivalent of assuming that each covariate has prior inclusion probability $\pi_j = 0.5$ to enter in the model.

- Beta-Binomial hierarchical prior on the model size d_ℓ

$$d_\ell \sim \text{Binomial}(\pi, p) \text{ and } \pi \sim Beta(\alpha, \beta),$$

where π is the probability of including one covariate in the model and p is the total number of covariates under consideration.

If $\alpha = \beta = 1$ then we have a uniform prior on the model size.

Prior on model parameters

- Proper prior distributions (conjugate if available).
  - For example in the case of the Gaussian regression models a popular choice is the **Zellner's g-prior** (Zellner, 1986).
  - Main issue: Specification of hyperparameter g that controls the prior variance.
  - Large values of $g \rightarrow$ Bartlett's paradox (e.g., Bartlett, 1957 Biometrika).
  - For $g = n \Rightarrow$ **unit information prior** (Kass & Wasserman, 1995, JASA).
  - Beta prior on $\frac{g}{g+1} \rightarrow$ Hyper-g prior (e.g. Liang *et al.* , 2008, JASA).
- Non-local priors (e.g. Johnson and Rossell, 2010, RSSS B).
  - They have zero mass for values of the parameter under the null hypothesis.
  - Products of independent normal moment priors.

Prior on model parameters (cont.)

- Shrinkage priors.
  - E.g. Bayesian Lasso (Park and Casella, 2008, JASA), horseshoe prior (Carvalho *et al.* , 2010, Biometrika), etc.
- Improper (reference) priors (**defined up to arbitrary constants**).
  - Objectivity.
  - Jeffreys prior.
  - Bayes factors cannot be determined.
- Priors defined via **imaginary data**.
  - Power prior (Ibrahim & Chen, 2000, Statistical Science).
  - Expected-Posterior prior (Pérez & Berger, 2002, Biometrika).
- Intrinsic priors.

3 Expected-Posterior Priors (EPP)

Pérez & Berger (2002, Biometrika) developed **expected-posterior prior** (EPP).

Suitable for model comparison, using imaginary training samples.

The EPP is the posterior distribution of a parameter vector for a given model, averaged over all possible imaginary samples $\boldsymbol{y}^*$ coming from a "suitable" predictive distribution $m^*(\boldsymbol{y}^*)$.

$$\pi_\ell^E(\boldsymbol{\theta}_\ell) = \int \pi_\ell^N(\boldsymbol{\theta}_\ell|\boldsymbol{y}^*)\, m^*(\boldsymbol{y}^*)\, d\boldsymbol{y}^*\,, \tag{1}$$

where $\pi_\ell^N(\boldsymbol{\theta}_\ell|\boldsymbol{y}^*)$ is the posterior of $\boldsymbol{\theta}_\ell$ for model M_ℓ using a baseline prior $\pi_\ell^N(\boldsymbol{\theta}_\ell)$ and data $\boldsymbol{y}^*$.

Specification of the predictive distribution

Select m^* to be the predictive distribution $m_0^N(\boldsymbol{y}^*)$ of a "reference" model M_0 under the baseline prior $\pi_0^N(\boldsymbol{\theta}_0)$.

In the **variable-selection** the constant model is clearly a good reference model since it is nested in all the models under consideration.

- It supports a-priori the parsimony principle assuming no causal structure for the data.
- Simplifies calculations.
- Makes EPP approach equivalent to the arithmetic intrinsic Bayes factor approach of Berger and Pericchi (1996, JASA).

An attractive property

EPPs can avoid the impropriety of the baseline priors which cause problems in Bayesian inference.

Impropriety in m^* also does not cause indeterminacy, because m^* is common to the EPPs for all models.

[nevertheless EPP looses its nice interpretation as an average over all imaginary samples coming from the predictive distribution.]

3.1 EPPs for variable selection in Gaussian linear models

We consider models M_ℓ (for $\ell = 0, 1$) with

Parameters: $\boldsymbol{\theta}_\ell = (\boldsymbol{\beta}_\ell\,, \sigma^2_\ell)$

Likelihood:

$$(\boldsymbol{Y}|\mathrm{X}_\ell, \boldsymbol{\beta}_\ell, \sigma^2_\ell, M_\ell) \sim N_n(\mathrm{X}_\ell\,\boldsymbol{\beta}_\ell\,, \sigma^2_\ell\,\mathrm{I}_n)\,, \tag{2}$$

$\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ is a vector of the responses,

X_ℓ is an $n \times d_\ell$ data/design matrix of the explanatory variables

I_n is the $n \times n$ identity matrix,

$\boldsymbol{\beta}_\ell$ is a vector of length d_ℓ of the model coefficients and

σ^2_ℓ is the error variance for model M_ℓ.

Additionally, suppose we **an imaginary/training data set $\boldsymbol{y}^*$**, of size n^*, and design matrix X^*.

Training sample

Generally, **EPP does not depend on the training sample $\boldsymbol{y}^*$** of the response variable Y since this is averaged over all possible samples coming from the reference predictive distribution.

Nevertheless, **in variable selection, EPP depends on the training sample** X^* of the explanatory variables $\Rightarrow$ creates additional computational difficulties.

EPP also **depends on the size of the training sample** n^*.

Proposed solutions/approaches: Selection of a **minimal training sample** $\Rightarrow$ makes the information induced by the prior as small as possible.

Minimal training sample

Selection of a **minimal training sample** ⇒ makes the information induced by the prior as small as possible.

We select a sample sufficiently large to specify all the estimated parameters of the models under consideration.

- Specification in terms of the largest model in every pairwise comparison
  - ⇒ the prior changes in every comparison
  - ⇒ overall variable-selection procedure incoherent.
- Specification in terms of the full model for all pairwise comparisons,
  - ⇒ Inference within the current data set is coherent.
  - ⇒ Prior should change if additional covariates are included later in the study(?)
  - ⇒ Influential prior for cases with n close to p.

- The problem of choosing a training sample still remains. Possible solutions:
  - $\Rightarrow$ The arithmetic mean of the Bayes factors over all possible training samples
    - $\Rightarrow$ This approach can be computationally infeasible for large dataset.
  - $\Rightarrow$ Calculate BFs for a random sample minimal training samples $\Rightarrow$ This adds an extraneous layer of Monte-Carlo noise to the model-comparison process

Training sample (cont.)

A solution was proposed by researchers working with intrinsic priors (e.g. Giròn *et al*. 2006, *Scandinavian Journal of Statistics*).

1. They proved that the intrinsic prior depends on X_k^* only through the expression $\mathrm{W}_k^{-1} = (\mathrm{X}_k^{*T}\mathrm{X}_k^*)^{-1}$;
   where X_k^* *is the imaginary design matrix of dimension* $(d_k+1)\times d_k$ *for a minimal training sample of size* (d_k+1).

2. They propose to replace W_k^{-1} with its average over all possible training samples of minimal size. *This idea is driven by the use of the arithmetic intrinsic Bayes factor.*

3. This average is equal to $\frac{n}{d_k+1}\left(\mathrm{X}_k^T\mathrm{X}_k\right)^{-1}$. Here X_k refers to the design matrix of the largest model in each pairwise comparison.

Training sample (cont.)

The solution proposed by researchers working with intrinsic priors (e.g. Giròn *et al*. 2006, *Scandinavian Journal of Statistics*).

- Seems intuitively sensible and dispenses with the extraction of the submatrices from X_k.
- It is unclear if the procedure retains its intrinsic interpretation, i.e., whether it is equivalent to the arithmetic intrinsic Bayes factor.
- The resulting prior can be influential when n is not much larger than p (in contrast to the prior we propose here, which has a unit-information interpretation).

4 Motivation

AIM

1. Produce a less influential EPP . This will extremely helpful especially in cases when n is not much larger than p.

2. Diminish the effect of training samples.

Ingredients

We combine ideas from the **power prior approach** and **unit information prior approach**.

Characteristics

- The likelihood involved in the EPP is raised to the power of $1/\delta$.
- For $\delta = n^*$ $\rightarrow$ prior with information equivalent to one data point.
- The method is sufficiently insensitive to the size of n^*.
  - $\Rightarrow$ We consider $n^* = n$ (and therefore $\mathrm{X}^* = \mathrm{X}$) and dispense with training samples altogether.
  - $\Rightarrow$ This both removes the instability arising from the random choice of training samples and greatly reduces computing time.

Baseline prior choices

1. The independence Jeffreys prior (improper).
   Usual choice of improper prior among researchers developing objective variable-selection methods.

2. The g-prior (proper).
   Usual choice of proper prior among researchers developing variable-selection methods.

Further comments

1. The BFs of the first baseline-prior choice can be considered as a limiting case of the BFs using the second prior.

2. Due to its (conditional) conjugacy, the second approach is easier to calculate. Hence using the 2nd approach to estimate the BFs of the 1st, considerably decreases the computational time.

5 Power-Expected-Posterior Prior

We denote by $\pi_\ell^N(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \mathrm{X}_\ell^*)$ the baseline prior for model parameters $\boldsymbol{\beta}_\ell$ and σ_ℓ^2, for any model $M_\ell \in \mathcal{M}$.

The *power-expected-posterior* (PEP) prior is defined as:

$$\pi_\ell^{PEP}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 \,|\, \mathrm{X}_\ell^*, \delta) = \int f(\boldsymbol{\beta}_\ell, \sigma_\ell^2 \,|\boldsymbol{y}^*, M_\ell\,; \mathrm{X}_\ell^*, \delta) m_0^N(\boldsymbol{y}^* | \,\mathrm{X}_0^*, \delta)\, d\boldsymbol{y}^*\,, \tag{3}$$

where

$$f(\boldsymbol{\beta}_\ell, \sigma_\ell^2 \,|\boldsymbol{y}^*, M_\ell\,; \mathrm{X}_\ell^*, \delta) = \frac{f(\boldsymbol{y}^* | \,\boldsymbol{\beta}_\ell, \sigma_\ell^2, M_\ell\,; \mathrm{X}_\ell^*, \delta)\pi_\ell^N(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \mathrm{X}_\ell^*)}{m_\ell^N(\boldsymbol{y}^* | \,\mathrm{X}_\ell^*, \delta)}$$

can be considered as a posterior with likelihood equal to the original likelihood raised to the power of $\frac{1}{\delta}$ and density-normalized, i.e.,

$$f(\boldsymbol{y}^* | \,\boldsymbol{\beta}_\ell, \sigma_\ell^2, M_\ell\,; \mathrm{X}_\ell^*, \delta) = \frac{f(\boldsymbol{y}^* | \boldsymbol{\beta}_\ell, \sigma_\ell^2, M_\ell\,; \mathrm{X}_\ell^*)^{\frac{1}{\delta}}}{\int f(\boldsymbol{y}^* | \boldsymbol{\beta}_\ell, \sigma_\ell^2, M_\ell\,; \mathrm{X}_\ell^*)^{\frac{1}{\delta}}\, d\boldsymbol{y}^*} = f_{N_{n^*}}(\boldsymbol{y}^*\,; \mathrm{X}_\ell^* \boldsymbol{\beta}_\ell\,, \delta\, \sigma_\ell^2 \mathrm{I}_{n^*})\,. \tag{4}$$

5 Power-Expected-Posterior Prior

We denote by $\pi_\ell^N(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \mathrm{X}_\ell^*)$ the baseline prior for model parameters $\boldsymbol{\beta}_\ell$ and σ_ℓ^2, for any model $M_\ell \in \mathcal{M}$.

The *power-expected-posterior* (PEP) prior is defined as:

$$\pi_\ell^{PEP}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 \,|\, \mathrm{X}_\ell^*, \delta) = \int f(\boldsymbol{\beta}_\ell, \sigma_\ell^2 \,|\boldsymbol{y}^*, M_\ell\,; \mathrm{X}_\ell^*, \delta) m_0^N(\boldsymbol{y}^* | \, \mathrm{X}_0^*, \delta) \, d\boldsymbol{y}^* \,, \tag{5}$$

The distribution $m_\ell^N(\boldsymbol{y}^* | \, \mathrm{X}_\ell^*, \delta)$ appearing in (5) is the prior predictive distribution, evaluated at $\boldsymbol{y}^*$, of model M_ℓ with the power likelihood defined in (4) under the baseline prior $\pi_\ell^N(\boldsymbol{\beta}_\ell, \sigma_\ell^2 \,|\, \mathrm{X}_\ell^*)$, i.e.,

$$m_\ell^N(\boldsymbol{y}^* | \, \mathrm{X}_\ell^*, \delta) = \int\!\!\int f_{N_{n^*}}(\boldsymbol{y}^* \,;\, \mathrm{X}_\ell^* \boldsymbol{\beta}_\ell \,, \delta\, \sigma_\ell^2\, \mathrm{I}_{n^*})\, \pi_\ell^N(\boldsymbol{\beta}_\ell \,, \sigma_\ell^2 \,|\, \mathrm{X}_\ell^*)\, d\boldsymbol{\beta}_\ell \, d\sigma_\ell^2 \,. \tag{6}$$

An alternative expression of PEP prior

This expression is closer to the intrinsic variable selection approach:

$$\pi_\ell^{PEP}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 \,|\, \mathrm{X}_\ell^*, \delta) = \pi_\ell^N(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \mathrm{X}_\ell^*) \int \frac{m_0^N(\boldsymbol{y}^* | \mathrm{X}_0^*, \delta)}{m_\ell^N(\boldsymbol{y}^* | \mathrm{X}_\ell^*, \delta)} f(\boldsymbol{y}^* | \boldsymbol{\beta}_\ell, \sigma_\ell^2, M_\ell; \mathrm{X}_\ell^*, \delta)\, d\boldsymbol{y}^*, \tag{7}$$

The posterior distribution

Under the PEP prior distribution (5), the posterior distribution of the model parameters $(\boldsymbol{\beta}_\ell, \sigma^2_\ell)$ is

$$\pi^{PEP}_\ell(\boldsymbol{\beta}_\ell, \sigma^2_\ell | \boldsymbol{y}; \mathrm{X}_\ell, \mathrm{X}^*_\ell, \delta) \propto \int \pi^N_\ell(\boldsymbol{\beta}_\ell, \sigma^2_\ell | \boldsymbol{y}, \boldsymbol{y}^*; \mathrm{X}_\ell, \mathrm{X}^*_\ell, \delta) \times m^N_\ell(\boldsymbol{y}|\boldsymbol{y}^*; \mathrm{X}_\ell, \mathrm{X}^*_\ell, \delta)\, m^N_0(\boldsymbol{y}^*|\mathrm{X}^*_0, \delta)\, d\boldsymbol{y}^*, \quad (8)$$

5.1 J-PEP: PEP-prior using the Jeffreys prior as baseline

Baseline prior:

$$\pi_\ell^N(\boldsymbol{\beta}_\ell, \sigma^2 \,|\, \mathrm{X}_\ell^*) = \frac{c_\ell}{\sigma_\ell^2}, \tag{9}$$

– c_ℓ is an unknown normalizing constant.

5.1 J-PEP: PEP-prior using the Jeffreys prior as baseline

Baseline prior:

$$\pi_\ell^N(\boldsymbol{\beta}_\ell\,, \sigma^2\,|\,\mathrm{X}_\ell^*) = \frac{c_\ell}{\sigma_\ell^2}\,, \tag{9}$$

– c_ℓ is an unknown normalizing constant.

J-PEP prior:

$$\begin{aligned}\pi_\ell^{J\text{-}PEP}(\boldsymbol{\beta}_\ell, \sigma_\ell^2|\mathrm{X}_\ell^*\,, \delta) \quad = \quad & \int f_{N_{d_\ell}}\Big[\boldsymbol{\beta}_\ell\,; \widehat{\boldsymbol{\beta}}_\ell^*, \delta\,(\mathrm{X}_\ell^{*\,T}\mathrm{X}_\ell^*)^{-1}\sigma_\ell^2\Big] \times \\ & f_{IG}\Big(\sigma_\ell^2\,; \frac{n^* - d_\ell}{2}, \frac{RSS_\ell^*}{2\delta}\Big)\, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*\,, \delta)\, d\boldsymbol{y}^*\,, \end{aligned} \tag{10}$$

5.1 J-PEP: PEP-prior using the Jeffreys prior as baseline

Baseline prior:

$$\pi_\ell^N(\boldsymbol{\beta}_\ell\,, \sigma^2\,|\, \mathrm{X}_\ell^*) = \frac{c_\ell}{\sigma_\ell^2}\,, \tag{9}$$

– c_ℓ is an unknown normalizing constant.

J-PEP prior:

$$\pi_\ell^{J\text{-}PEP}(\boldsymbol{\beta}_\ell, \sigma_\ell^2|\mathrm{X}_\ell^*\,, \delta) = \int f_{N_{d_\ell}}\Big[\boldsymbol{\beta}_\ell\,;\,\widehat{\boldsymbol{\beta}}_\ell^*, \delta\,({\mathrm{X}_\ell^*}^T\mathrm{X}_\ell^*)^{-1}\sigma_\ell^2\Big] \times f_{IG}\Big(\sigma_\ell^2\,;\,\frac{n^*-d_\ell}{2}, \frac{RSS_\ell^*}{2\delta}\Big)\, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*\,, \delta)\, d\boldsymbol{y}^*\,, \tag{10}$$

– $\widehat{\boldsymbol{\beta}}_\ell^* = ({\mathrm{X}_\ell^*}^T\mathrm{X}_\ell^*)^{-1}{\mathrm{X}_\ell^*}^T\boldsymbol{y}^*$: is the MLE of $\boldsymbol{\beta}_\ell$ with response $\boldsymbol{y}^*$ and design matrix X_ℓ^*,

– RSS_ℓ^* is the residual sum of squares using $(\boldsymbol{y}^*, \mathrm{X}_\ell^*)$ as data.

5.1 J-PEP: PEP-prior using the Jeffreys prior as baseline

Baseline prior:

$$\pi_\ell^N(\boldsymbol{\beta}_\ell\,,\sigma^2\,|\,X_\ell^*) = \frac{c_\ell}{\sigma_\ell^2}\,, \tag{9}$$

- c_ℓ is an unknown normalizing constant.

J-PEP prior:

$$\begin{aligned}\pi_\ell^{J\text{-}PEP}(\boldsymbol{\beta}_\ell,\sigma_\ell^2|X_\ell^*,\delta) \quad = \quad & \int f_{N_{d_\ell}}\left[\boldsymbol{\beta}_\ell\,;\,\widehat{\boldsymbol{\beta}}_\ell^*,\delta\,({X_\ell^*}^T X_\ell^*)^{-1}\sigma_\ell^2\right]\times \\ & f_{IG}\left(\sigma_\ell^2\,;\,\frac{n^*-d_\ell}{2},\frac{RSS_\ell^*}{2\delta}\right) m_0^N(\boldsymbol{y}^*|X_0^*,\delta)\,d\boldsymbol{y}^*\,, \end{aligned} \tag{10}$$

- $\widehat{\boldsymbol{\beta}}_\ell^* = ({X_\ell^*}^T X_\ell^*)^{-1}{X_\ell^*}^T\boldsymbol{y}^*$: is the MLE of $\boldsymbol{\beta}_\ell$ with response $\boldsymbol{y}^*$ and design matrix X_ℓ^*,
- RSS_ℓ^* is the residual sum of squares using $(\boldsymbol{y}^*,X_\ell^*)$ as data.

Prior predictive density: The prior predictive distribution of a model M_ℓ with power likelihood defined in (4) under the baseline prior (9) is given by

$$m_\ell^N(\boldsymbol{y}^*\,|\,X_\ell^*\,,\delta) = c_\ell\,\pi^{\frac{1}{2}(d_\ell-n^*)}\,|{X_\ell^*}^T X_\ell^*|^{-\frac{1}{2}}\,\Gamma\left(\frac{n^*-d_\ell}{2}\right) {RSS_\ell^*}^{-\left(\frac{n^*-d_\ell}{2}\right)}\,. \tag{11}$$

Posterior distribution using J-PEP:

$$\pi_\ell^{J\text{-}PEP}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \boldsymbol{y}; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \propto \int f_{N_{d_\ell}}(\boldsymbol{\beta}_\ell\,;\, \widetilde{\boldsymbol{\beta}}^N,\, \widetilde{\Sigma}^N \sigma_\ell^2)\, f_{IG}(\sigma_\ell^2\,;\, \widetilde{a}_\ell^N, \widetilde{b}_\ell^N) \times m_\ell^N(\boldsymbol{y}|\boldsymbol{y}^*;\, \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)\, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta)\, d\boldsymbol{y}^* \,. \quad (12)$$

Posterior distribution using J-PEP:

$$\pi_\ell^{J\text{-}PEP}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \boldsymbol{y}; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \propto \int f_{N_{d_\ell}}(\boldsymbol{\beta}_\ell \,;\, \widetilde{\boldsymbol{\beta}}^N, \, \widetilde{\Sigma}^N \sigma_\ell^2)\, f_{IG}(\sigma_\ell^2 \,;\, \widetilde{a}_\ell^N, \widetilde{b}_\ell^N) \times m_\ell^N(\boldsymbol{y}|\boldsymbol{y}^*; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)\, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta)\, d\boldsymbol{y}^* \,. \quad (12)$$

– $\pi_\ell^N(\boldsymbol{\beta}_\ell|\sigma_\ell^2, \boldsymbol{y}, \boldsymbol{y}^*; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) = f_{N_{d_\ell}}(\boldsymbol{\beta}_\ell \,;\, \widetilde{\boldsymbol{\beta}}^N, \, \widetilde{\Sigma}^N \sigma_\ell^2)$ is the conditional posterior of $\boldsymbol{\beta}_\ell|\sigma_\ell^2$ under the actual likelihood (and data), the power likelihood (and the imaginary data) and the Jeffreys baseline prior.

Posterior distribution using J-PEP:

$$\pi_\ell^{J\text{-}PEP}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \boldsymbol{y}; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \propto \int f_{N_{d_\ell}}(\boldsymbol{\beta}_\ell \,;\, \widetilde{\boldsymbol{\beta}}^N ,\, \widetilde{\Sigma}^N \sigma_\ell^2)\, f_{IG}(\sigma_\ell^2 \,;\, \widetilde{a}_\ell^N, \widetilde{b}_\ell^N) \times m_\ell^N(\boldsymbol{y}|\boldsymbol{y}^*; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)\, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta)\, d\boldsymbol{y}^* \,. \qquad (12)$$

– $\pi_\ell^N(\boldsymbol{\beta}_\ell|\sigma_\ell^2, \boldsymbol{y}, \boldsymbol{y}^*; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) = f_{N_{d_\ell}}(\boldsymbol{\beta}_\ell \,;\, \widetilde{\boldsymbol{\beta}}^N ,\, \widetilde{\Sigma}^N \sigma_\ell^2)$ is the conditional posterior of $\boldsymbol{\beta}_\ell|\sigma_\ell^2$ under the actual likelihood (and data), the power likelihood (and the imaginary data) and the Jeffreys baseline prior.

Posterior mean: $\widetilde{\boldsymbol{\beta}}^N = \widetilde{\Sigma}^N(\mathrm{X}_\ell^T \boldsymbol{y} + \delta^{-1} {\mathrm{X}_\ell^*}^T \boldsymbol{y}^*)$.

Posterior variance: $\widetilde{\Sigma}^N = \left[\mathrm{X}_\ell^T \mathrm{X}_\ell + \delta^{-1} {\mathrm{X}_\ell^*}^T \mathrm{X}_\ell^*\right]^{-1}$.

Posterior distribution using J-PEP:

$$\pi_\ell^{J\text{-}PEP}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \boldsymbol{y}; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \quad \propto \quad \int f_{N_{d_\ell}}(\boldsymbol{\beta}_\ell\,;\, \widetilde{\boldsymbol{\beta}}^N,\ \widetilde{\Sigma}^N \sigma_\ell^2)\, f_{IG}(\sigma_\ell^2\,;\, \widetilde{a}_\ell^N, \widetilde{b}_\ell^N) \times$$
$$m_\ell^N(\boldsymbol{y}|\boldsymbol{y}^*;\, \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)\, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta)\, d\boldsymbol{y}^* \,. \qquad (12)$$

– $\pi_\ell^N(\boldsymbol{\beta}_\ell|\sigma_\ell^2, \boldsymbol{y}, \boldsymbol{y}^*; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) = f_{N_{d_\ell}}(\boldsymbol{\beta}_\ell\,;\, \widetilde{\boldsymbol{\beta}}^N,\ \widetilde{\Sigma}^N \sigma_\ell^2)$ is the conditional posterior of $\boldsymbol{\beta}_\ell|\sigma_\ell^2$ under the actual likelihood (and data), the power likelihood (and the imaginary data) and the Jeffreys baseline prior.

Posterior mean: $\widetilde{\boldsymbol{\beta}}^N = \widetilde{\Sigma}^N(\mathrm{X}_\ell^T \boldsymbol{y} + \delta^{-1} {\mathrm{X}_\ell^*}^T \boldsymbol{y}^*)$.

Posterior variance: $\widetilde{\Sigma}^N = \left[\mathrm{X}_\ell^T \mathrm{X}_\ell + \delta^{-1} {\mathrm{X}_\ell^*}^T \mathrm{X}_\ell^*\right]^{-1}$.

– Similarly, $\pi_\ell^N(\sigma_\ell^2|\boldsymbol{y}, \boldsymbol{y}^*; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) = f_{IG}(\sigma_\ell^2\,;\, \widetilde{a}_\ell^N, \widetilde{b}_\ell^N)$ is the corresponding posterior of σ_ℓ^2 under the actual and power likelihood and the Jeffreys baseline prior.

Posterior distribution using J-PEP:

$$\pi_\ell^{J\text{-}PEP}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \boldsymbol{y}; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \propto \int f_{N_{d_\ell}}(\boldsymbol{\beta}_\ell\,;\, \widetilde{\boldsymbol{\beta}}^N,\, \widetilde{\Sigma}^N \sigma_\ell^2)\, f_{IG}(\sigma_\ell^2\,;\, \widetilde{a}_\ell^N, \widetilde{b}_\ell^N) \times m_\ell^N(\boldsymbol{y}|\boldsymbol{y}^*;\, \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)\, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta)\, d\boldsymbol{y}^* \,. \tag{12}$$

- $\pi_\ell^N(\boldsymbol{\beta}_\ell|\sigma_\ell^2, \boldsymbol{y}, \boldsymbol{y}^*; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) = f_{N_{d_\ell}}(\boldsymbol{\beta}_\ell\,;\, \widetilde{\boldsymbol{\beta}}^N,\, \widetilde{\Sigma}^N \sigma_\ell^2)$ is the conditional posterior of $\boldsymbol{\beta}_\ell|\sigma_\ell^2$ under the actual likelihood (and data), the power likelihood (and the imaginary data) and the Jeffreys baseline prior.

  Posterior mean: $\widetilde{\boldsymbol{\beta}}^N = \widetilde{\Sigma}^N(\mathrm{X}_\ell^T\boldsymbol{y} + \delta^{-1}{\mathrm{X}_\ell^*}^T\boldsymbol{y}^*)$.

  Posterior variance: $\widetilde{\Sigma}^N = \left[\mathrm{X}_\ell^T\mathrm{X}_\ell + \delta^{-1}{\mathrm{X}_\ell^*}^T\mathrm{X}_\ell^*\right]^{-1}$.

- Similarly, $\pi_\ell^N(\sigma_\ell^2|\boldsymbol{y}, \boldsymbol{y}^*; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) = f_{IG}(\sigma_\ell^2\,;\, \widetilde{a}_\ell^N, \widetilde{b}_\ell^N)$ is the corresponding posterior of σ_ℓ^2 under the actual and power likelihood and the Jeffreys baseline prior.

  Posterior parameters: $\widetilde{a}_\ell^N = \frac{1}{2}(n + n^* - d_\ell)$, $\widetilde{b}_\ell^N = \frac{1}{2}(SS_\ell^N + \delta^{-1}RSS_\ell^*)$ and

$$SS_\ell^N = (\boldsymbol{y} - \mathrm{X}_\ell\,\widehat{\boldsymbol{\beta}}_\ell^*)^T \left[\mathrm{I}_n + \delta\,\mathrm{X}_\ell({\mathrm{X}_\ell^*}^T\mathrm{X}_\ell^*)^{-1}\mathrm{X}_\ell^T\right]^{-1} (\boldsymbol{y} - \mathrm{X}_\ell\,\widehat{\boldsymbol{\beta}}_\ell^*).$$

Posterior distribution using J-PEP:

$$\pi_\ell^{J\text{-}PEP}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \boldsymbol{y}; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \propto \int f_{N_{d_\ell}}(\boldsymbol{\beta}_\ell ; \widetilde{\boldsymbol{\beta}}^N, \widetilde{\Sigma}^N \sigma_\ell^2)\, f_{IG}(\sigma_\ell^2 ; \widetilde{a}_\ell^N, \widetilde{b}_\ell^N) \times m_\ell^N(\boldsymbol{y}|\boldsymbol{y}^*; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)\, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta)\, d\boldsymbol{y}^* . \quad (12)$$

– $\pi_\ell^N(\boldsymbol{\beta}_\ell|\sigma_\ell^2, \boldsymbol{y}, \boldsymbol{y}^*; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) = f_{N_{d_\ell}}(\boldsymbol{\beta}_\ell ; \widetilde{\boldsymbol{\beta}}^N, \widetilde{\Sigma}^N \sigma_\ell^2)$ is the conditional posterior of $\boldsymbol{\beta}_\ell|\sigma_\ell^2$ under the actual likelihood (and data), the power likelihood (and the imaginary data) and the Jeffreys baseline prior.

Posterior mean: $\widetilde{\boldsymbol{\beta}}^N = \widetilde{\Sigma}^N(\mathrm{X}_\ell^T \boldsymbol{y} + \delta^{-1} {\mathrm{X}_\ell^*}^T \boldsymbol{y}^*)$.

Posterior variance: $\widetilde{\Sigma}^N = \left[\mathrm{X}_\ell^T \mathrm{X}_\ell + \delta^{-1} {\mathrm{X}_\ell^*}^T \mathrm{X}_\ell^*\right]^{-1}$.

– Similarly, $\pi_\ell^N(\sigma_\ell^2|\boldsymbol{y}, \boldsymbol{y}^*; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) = f_{IG}(\sigma_\ell^2 ; \widetilde{a}_\ell^N, \widetilde{b}_\ell^N)$ is the corresponding posterior of σ_ℓ^2 under the actual and power likelihood and the Jeffreys baseline prior.

Posterior parameters: $\widetilde{a}_\ell^N = \frac{1}{2}(n + n^* - d_\ell)$, $\widetilde{b}_\ell^N = \frac{1}{2}(SS_\ell^N + \delta^{-1} RSS_\ell^*)$ and

$$SS_\ell^N = (\boldsymbol{y} - \mathrm{X}_\ell \widehat{\boldsymbol{\beta}}_\ell^*)^T \left[\mathrm{I}_n + \delta\, \mathrm{X}_\ell ({\mathrm{X}_\ell^*}^T \mathrm{X}_\ell^*)^{-1} \mathrm{X}_\ell^T\right]^{-1} (\boldsymbol{y} - \mathrm{X}_\ell \widehat{\boldsymbol{\beta}}_\ell^*).$$

– $m_\ell^N(\boldsymbol{y}|\boldsymbol{y}^*; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) = f_{St_n}\left\{\boldsymbol{y} ; n^* - d_\ell, \mathrm{X}_\ell \widehat{\boldsymbol{\beta}}_\ell^*, \frac{RSS_\ell^*}{\delta(n^* - d_\ell)}\left[\mathrm{I}_n + \delta\, \mathrm{X}_\ell ({\mathrm{X}_\ell^*}^T \mathrm{X}_\ell^*)^{-1} \mathrm{X}_\ell^T\right]\right\}$.

5.2 Z-PEP: PEP Prior using the Zellner's g-prior as baseline

Baseline prior:

$$\pi_\ell^N(\boldsymbol{\beta}_\ell|\sigma_\ell^2\,;\,\mathrm{X}_\ell^*) = f_{N_{d_\ell}}\left[\boldsymbol{\beta}_\ell\,;\,\mathbf{0}, g\,({\mathrm{X}_\ell^*}^T\mathrm{X}_\ell^*)^{-1}\sigma_\ell^2\right] \text{ and } \pi_\ell^N(\sigma_\ell^2) = f_{IG}\left(\sigma_\ell^2\,;\,a_\ell, b_\ell\right)\,. \qquad (13)$$

5.2 Z-PEP: PEP Prior using the Zellner's g-prior as baseline

Baseline prior:

$$\pi_\ell^N(\boldsymbol{\beta}_\ell|\sigma_\ell^2\,;\,\mathrm{X}_\ell^*) = f_{N_{d_\ell}}\left[\boldsymbol{\beta}_\ell\,;\,\mathbf{0}, g\,({\mathrm{X}_\ell^*}^T\mathrm{X}_\ell^*)^{-1}\sigma_\ell^2\right] \text{ and } \pi_\ell^N(\sigma_\ell^2) = f_{IG}\left(\sigma_\ell^2\,;\,a_\ell, b_\ell\right)\,. \tag{13}$$

Z-PEP prior:

$$\begin{aligned}\pi_\ell^{Z\text{-}PEP}(\boldsymbol{\beta}_\ell, \sigma_\ell^2|\mathrm{X}_\ell^*, \delta) \quad = \quad & \int f_{N_{d_\ell}}\left[\boldsymbol{\beta}_\ell\,;\, w\,\widehat{\boldsymbol{\beta}}_\ell^*, w\,\delta\,({\mathrm{X}_\ell^*}^T\mathrm{X}_\ell^*)^{-1}\sigma_\ell^2\right]\times \\ & f_{IG}\left(\sigma_\ell^2\,;\,a_\ell + \frac{n^*}{2}, b_\ell + \frac{SS_\ell^*}{2}\right) m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta)\,d\boldsymbol{y}^*\,,\end{aligned} \tag{14}$$

5.2 Z-PEP: PEP Prior using the Zellner's g-prior as baseline

Baseline prior:

$$\pi_\ell^N(\boldsymbol{\beta}_\ell|\sigma_\ell^2\,;\,\mathrm{X}_\ell^*) = f_{N_{d_\ell}}\left[\boldsymbol{\beta}_\ell\,;\,\mathbf{0}, g\,({\mathrm{X}_\ell^*}^T\mathrm{X}_\ell^*)^{-1}\sigma_\ell^2\right] \text{ and } \pi_\ell^N(\sigma_\ell^2) = f_{IG}\left(\sigma_\ell^2\,;\,a_\ell, b_\ell\right)\,. \tag{13}$$

Z-PEP prior:

$$\begin{aligned}\pi_\ell^{Z\text{-}PEP}(\boldsymbol{\beta}_\ell,\sigma_\ell^2|\mathrm{X}_\ell^*,\delta) \quad &= \quad \int f_{N_{d_\ell}}\left[\boldsymbol{\beta}_\ell\,;\,w\,\widehat{\boldsymbol{\beta}}_\ell^*, w\,\delta\,({\mathrm{X}_\ell^*}^T\mathrm{X}_\ell^*)^{-1}\sigma_\ell^2\right]\times \\ & f_{IG}\left(\sigma_\ell^2\,;\,a_\ell+\frac{n^*}{2}, b_\ell+\frac{SS_\ell^*}{2}\right) m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*,\delta)\,d\boldsymbol{y}^*\,,\end{aligned} \tag{14}$$

- $w = \frac{g}{g+\delta}$ is a shrinkage weight, $\widehat{\boldsymbol{\beta}}_\ell^*$ is the MLE of $\boldsymbol{\beta}_\ell$ with response $\boldsymbol{y}^*$ and design matrix X_ℓ^*,
- $SS_\ell^* = {\boldsymbol{y}^*}^T\Lambda_\ell^*\,\boldsymbol{y}^*$ is a posterior sum of squares,
- ${\Lambda_\ell^*}^{-1} = \delta\left[\mathrm{I}_{n^*} - \frac{g}{g+\delta}\mathrm{X}_\ell^*\left({\mathrm{X}_\ell^*}^T\mathrm{X}_\ell^*\right)^{-1}{\mathrm{X}_\ell^*}^T\right]^{-1} = \delta\,\mathrm{I}_{n^*} + g\,\mathrm{X}_\ell^*\left({\mathrm{X}_\ell^*}^T\mathrm{X}_\ell^*\right)^{-1}{\mathrm{X}_\ell^*}^T\,.$

5.2 Z-PEP: PEP Prior using the Zellner's g-prior as baseline

Baseline prior:

$$\pi_\ell^N(\boldsymbol{\beta}_\ell|\sigma_\ell^2\,;\,\mathrm{X}_\ell^*) = f_{N_{d_\ell}}\left[\boldsymbol{\beta}_\ell\,;\,\mathbf{0}, g\,({\mathrm{X}_\ell^*}^T\mathrm{X}_\ell^*)^{-1}\sigma_\ell^2\right] \text{ and } \pi_\ell^N(\sigma_\ell^2) = f_{IG}\left(\sigma_\ell^2\,;\,a_\ell, b_\ell\right)\,. \tag{13}$$

Z-PEP prior:

$$\pi_\ell^{Z\text{-}PEP}(\boldsymbol{\beta}_\ell, \sigma_\ell^2|\mathrm{X}_\ell^*, \delta) = \int f_{N_{d_\ell}}\left[\boldsymbol{\beta}_\ell\,;\,w\,\widehat{\boldsymbol{\beta}}_\ell^*, w\,\delta\,({\mathrm{X}_\ell^*}^T\mathrm{X}_\ell^*)^{-1}\sigma_\ell^2\right]\times f_{IG}\left(\sigma_\ell^2\,;\,a_\ell+\frac{n^*}{2}, b_\ell+\frac{SS_\ell^*}{2}\right) m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta)\,d\boldsymbol{y}^*\,, \tag{14}$$

- $w = \frac{g}{g+\delta}$ is a shrinkage weight, $\widehat{\boldsymbol{\beta}}_\ell^*$ is the MLE of $\boldsymbol{\beta}_\ell$ with response $\boldsymbol{y}^*$ and design matrix X_ℓ^*,
- $SS_\ell^* = {\boldsymbol{y}^*}^T\Lambda_\ell^*\,\boldsymbol{y}^*$ is a posterior sum of squares,
- ${\Lambda_\ell^*}^{-1} = \delta\left[\mathrm{I}_{n^*} - \frac{g}{g+\delta}\mathrm{X}_\ell^*\left({\mathrm{X}_\ell^*}^T\mathrm{X}_\ell^*\right)^{-1}{\mathrm{X}_\ell^*}^T\right]^{-1} = \delta\,\mathrm{I}_{n^*} + g\,\mathrm{X}_\ell^*\left({\mathrm{X}_\ell^*}^T\mathrm{X}_\ell^*\right)^{-1}{\mathrm{X}_\ell^*}^T$.

The prior predictive density: The prior predictive distribution under the baseline prior is

$$m_\ell^N(\boldsymbol{y}^*\,|\,\mathrm{X}_\ell^*, \delta) = f_{St_{n^*}}\left(\boldsymbol{y}^*\,;2\,a_\ell, \mathbf{0}, \frac{b_\ell}{a_\ell}{\Lambda_\ell^*}^{-1}\right). \tag{15}$$

The prior mean vector and covariance matrix of $\boldsymbol{\beta}_\ell$, and the prior mean and variance of σ^2_ℓ, can be calculated analytically.

Posterior distribution using Z-PEP:

$$
\begin{aligned}
\pi_\ell^{Z\text{-}PEP}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \boldsymbol{y}; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \quad \propto \quad & \int f_{N_{d_\ell}}(\boldsymbol{\beta}_\ell \, ; \widetilde{\boldsymbol{\beta}}^N, \, \widetilde{\Sigma}^N \sigma_\ell^2) \, f_{IG}(\sigma_\ell^2 \, ; \widetilde{a}_\ell^N, \widetilde{b}_\ell^N) \times \\
& m_\ell^N(\boldsymbol{y} | \boldsymbol{y}^*; \, \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \, m_0^N(\boldsymbol{y}^* | \mathrm{X}_0^*, \delta) \, d\boldsymbol{y}^* \qquad (16)
\end{aligned}
$$

Posterior distribution using Z-PEP:

$$\pi_\ell^{Z\text{-}PEP}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \boldsymbol{y}; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \propto \int f_{N_{d_\ell}}(\boldsymbol{\beta}_\ell\,;\, \widetilde{\boldsymbol{\beta}}^N,\, \widetilde{\Sigma}^N \sigma_\ell^2)\, f_{IG}(\sigma_\ell^2\,;\, \widetilde{a}_\ell^N, \widetilde{b}_\ell^N) \times m_\ell^N(\boldsymbol{y}|\boldsymbol{y}^*;\, \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)\, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta)\, d\boldsymbol{y}^* \quad (16)$$

– Parameters of the normal posterior of $\boldsymbol{\beta}_\ell$ given σ_ℓ^2:

$$\widetilde{\boldsymbol{\beta}}^N = \widetilde{\Sigma}^N (\mathrm{X}_\ell^T \boldsymbol{y} + \delta^{-1} {\mathrm{X}_\ell^*}^T \boldsymbol{y}^*),\ \widetilde{\Sigma}^N = \left[\mathrm{X}_\ell^T \mathrm{X}_\ell + (w\,\delta)^{-1} {\mathrm{X}_\ell^*}^T \mathrm{X}_\ell^*\right]^{-1}$$

Posterior distribution using Z-PEP:

$$\pi_\ell^{Z\text{-}PEP}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \boldsymbol{y}; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \propto \int f_{N_{d_\ell}}(\boldsymbol{\beta}_\ell\,;\, \widetilde{\boldsymbol{\beta}}^N,\, \widetilde{\Sigma}^N \sigma_\ell^2)\, f_{IG}(\sigma_\ell^2\,;\, \widetilde{a}_\ell^N, \widetilde{b}_\ell^N) \times m_\ell^N(\boldsymbol{y}|\boldsymbol{y}^*;\, \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)\, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta)\, d\boldsymbol{y}^* \quad (16)$$

– Parameters of the normal posterior of $\boldsymbol{\beta}_\ell$ given σ_ℓ^2:

$$\widetilde{\boldsymbol{\beta}}^N = \widetilde{\Sigma}^N (\mathrm{X}_\ell^T \boldsymbol{y} + \delta^{-1} {\mathrm{X}_\ell^*}^T \boldsymbol{y}^*),\ \widetilde{\Sigma}^N = \left[\mathrm{X}_\ell^T \mathrm{X}_\ell + (w\,\delta)^{-1} {\mathrm{X}_\ell^*}^T \mathrm{X}_\ell^*\right]^{-1}$$

– Parameters of the inverse gamma posterior of σ_ℓ^2:

$$\widetilde{a}_\ell^N = \frac{n+n^*}{2} + a_\ell\,,\ \widetilde{b}_\ell^N = \frac{SS_\ell^N + SS_\ell^*}{2} + b_\ell\,. \quad (17)$$

with $SS_\ell^N = (\boldsymbol{y} - w\,\mathrm{X}_\ell\,\widehat{\boldsymbol{\beta}}_\ell^*)^T \left[\mathrm{I}_n + \delta\, w\, \mathrm{X}_\ell ({\mathrm{X}_\ell^*}^T \mathrm{X}_\ell^*)^{-1} \mathrm{X}_\ell^T\right]^{-1} (\boldsymbol{y} - w\,\mathrm{X}_\ell\,\widehat{\boldsymbol{\beta}}_\ell^*)$ being a posterior sum of squares.

Posterior distribution using Z-PEP:

$$\pi_\ell^{Z\text{-}PEP}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \boldsymbol{y}; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \quad \propto \quad \int f_{N_{d_\ell}}(\boldsymbol{\beta}_\ell \,;\, \widetilde{\boldsymbol{\beta}}^N, \, \widetilde{\Sigma}^N \sigma_\ell^2) \, f_{IG}(\sigma_\ell^2 \,;\, \widetilde{a}_\ell^N, \widetilde{b}_\ell^N) \times m_\ell^N(\boldsymbol{y}|\boldsymbol{y}^*;\, \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta) \, d\boldsymbol{y}^* \qquad (16)$$

– Parameters of the normal posterior of $\boldsymbol{\beta}_\ell$ given σ_ℓ^2:

$$\widetilde{\boldsymbol{\beta}}^N \quad = \quad \widetilde{\Sigma}^N (\mathrm{X}_\ell^T \boldsymbol{y} + \delta^{-1} {\mathrm{X}_\ell^*}^T \boldsymbol{y}^*), \; \widetilde{\Sigma}^N = \left[\mathrm{X}_\ell^T \mathrm{X}_\ell + (w\,\delta)^{-1} {\mathrm{X}_\ell^*}^T \mathrm{X}_\ell^*\right]^{-1}$$

– Parameters of the inverse gamma posterior of σ_ℓ^2:

$$\widetilde{a}_\ell^N \quad = \quad \frac{n+n^*}{2} + a_\ell \,, \; \widetilde{b}_\ell^N = \frac{SS_\ell^N + SS_\ell^*}{2} + b_\ell \,. \qquad (17)$$

with $SS_\ell^N = (\boldsymbol{y} - w\,\mathrm{X}_\ell\, \widehat{\boldsymbol{\beta}}_\ell^*)^T \left[\mathrm{I}_n + \delta\, w\, \mathrm{X}_\ell ({\mathrm{X}_\ell^*}^T \mathrm{X}_\ell^*)^{-1} \mathrm{X}_\ell^T\right]^{-1} (\boldsymbol{y} - w\,\mathrm{X}_\ell\, \widehat{\boldsymbol{\beta}}_\ell^*)$ being a posterior sum of squares.

– $m_\ell^N(\boldsymbol{y}|\boldsymbol{y}^*; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) = f_{St_n}\left\{\boldsymbol{y}\,;\, 2\,a_\ell + n^*, w\,\mathrm{X}_\ell \widehat{\boldsymbol{\beta}}_\ell^*, \, \frac{2b_\ell + SS_\ell^*}{2\,a_\ell + n^*}\left[\mathrm{I}_n + w\,\delta\, \mathrm{X}_\ell ({\mathrm{X}_\ell^*}^T \mathrm{X}_\ell^*)^{-1} \mathrm{X}_\ell^T\right]\right\}$.

Specification of hyper-parameters

– The normal baseline prior parameter g is set equal to $\delta\, n^*$. If $\delta = n^* \Rightarrow g = (n^*)^2$.

– This choice makes the contribution of g-prior to be equal to approximately equal to one data point within the posterior $\pi_\ell^N(\boldsymbol{\beta}_\ell\,, \sigma_\ell^2 | \boldsymbol{y}^*;\, \mathrm{X}_\ell^*\,, \delta)$.

– The entire Z-PEP prior contribution is equal to $\left(1 + \frac{1}{\delta}\right)$ data points.

– We set $a_\ell = b_\ell = 0.01$ in the Inverse-Gamma baseline prior (prior mean of 1 and variance of 100).

Connection between the J-PEP and Z-PEP distributions

The two approaches coincide in terms of posterior inference for:

- large g (and therefore $w \approx 1$),
- $a_\ell = -\frac{d_\ell}{2}$ and $b_\ell = 0$.

Therefore, the posterior results using the Jeffreys prior as baseline can be obtained as a special (limiting) case of the results using the g-prior as baseline.

This is beneficial for the computation of the posterior distribution.

5.3 Marginal-likelihood computation

It is straightforward to show that the marginal likelihood of any model $M_\ell \in \mathcal{M}$ can be re-written as

$$m_\ell^{PEP}(\boldsymbol{y}|\mathrm{X}_\ell\,,\mathrm{X}_\ell^*\,,\delta) = m_\ell^N(\boldsymbol{y}|\mathrm{X}_\ell\,,\mathrm{X}_\ell^*\,) \int \frac{m_\ell^N(\boldsymbol{y}^*|\boldsymbol{y},\mathrm{X}_\ell,\mathrm{X}_\ell^*\,,\delta)}{m_\ell^N(\boldsymbol{y}^*|\mathrm{X}_\ell^*\,,\delta)}\, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*\,,\delta)\, d\boldsymbol{y}^*\,. \tag{18}$$

5.3 Marginal-likelihood computation

It is straightforward to show that the marginal likelihood of any model $M_\ell \in \mathcal{M}$ can be re-written as

$$m_\ell^{PEP}(\boldsymbol{y}|\mathrm{X}_\ell\,,\mathrm{X}_\ell^*\,,\delta) = m_\ell^N(\boldsymbol{y}|\mathrm{X}_\ell\,,\mathrm{X}_\ell^*\,) \int \frac{m_\ell^N(\boldsymbol{y}^*|\boldsymbol{y},\mathrm{X}_\ell,\mathrm{X}_\ell^*\,,\delta)}{m_\ell^N(\boldsymbol{y}^*|\mathrm{X}_\ell^*\,,\delta)}\, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*\,,\delta)\, d\boldsymbol{y}^*\,. \tag{18}$$

$m_\ell^N(\boldsymbol{y}|\mathrm{X}_\ell\,,\mathrm{X}_\ell^*)$ is the marginal likelihood of M_ℓ for the actual data under the baseline prior.

5.3 Marginal-likelihood computation

It is straightforward to show that the marginal likelihood of any model $M_\ell \in \mathcal{M}$ can be re-written as

$$m_\ell^{PEP}(\boldsymbol{y}|\mathrm{X}_\ell\,,\mathrm{X}_\ell^*\,,\delta) = m_\ell^N(\boldsymbol{y}|\mathrm{X}_\ell\,,\mathrm{X}_\ell^*\,) \int \frac{m_\ell^N(\boldsymbol{y}^*|\boldsymbol{y},\mathrm{X}_\ell,\mathrm{X}_\ell^*\,,\delta)}{m_\ell^N(\boldsymbol{y}^*|\mathrm{X}_\ell^*\,,\delta)}\, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*\,,\delta)\, d\boldsymbol{y}^*\,. \tag{18}$$

$m_\ell^N(\boldsymbol{y}|\mathrm{X}_\ell\,,\mathrm{X}_\ell^*)$ is the marginal likelihood of M_ℓ for the actual data under the baseline prior.

– Under the baseline g-prior (13), is given by

$$m_\ell^N(\boldsymbol{y}|\mathrm{X}_\ell\,,\mathrm{X}_\ell^*) = f_{St_n}\left\{\boldsymbol{y}\,;2\,a_\ell,\mathbf{0},\frac{b_\ell}{a_\ell}\left[\mathrm{I}_n + g\,\mathrm{X}_\ell\left({\mathrm{X}_\ell^*}^T\mathrm{X}_\ell^*\right)^{-1}{\mathrm{X}_\ell}^T\right]\right\}\,. \tag{19}$$

5.3 Marginal-likelihood computation

It is straightforward to show that the marginal likelihood of any model $M_\ell \in \mathcal{M}$ can be re-written as

$$m_\ell^{PEP}(\boldsymbol{y}|\mathrm{X}_\ell\,, \mathrm{X}_\ell^*\,, \delta) = m_\ell^N(\boldsymbol{y}|\mathrm{X}_\ell\,, \mathrm{X}_\ell^*\,) \int \frac{m_\ell^N(\boldsymbol{y}^*|\boldsymbol{y}, \mathrm{X}_\ell, \mathrm{X}_\ell^*\,, \delta)}{m_\ell^N(\boldsymbol{y}^*|\mathrm{X}_\ell^*\,, \delta)}\, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*\,, \delta)\, d\boldsymbol{y}^*\,. \quad (18)$$

$m_\ell^N(\boldsymbol{y}|\mathrm{X}_\ell\,, \mathrm{X}_\ell^*)$ is the marginal likelihood of M_ℓ for the actual data under the baseline prior.

– Under the baseline g-prior (13), is given by

$$m_\ell^N(\boldsymbol{y}|\mathrm{X}_\ell\,, \mathrm{X}_\ell^*) = f_{St_n}\left\{\boldsymbol{y}\,; 2\,a_\ell, \mathbf{0}, \frac{b_\ell}{a_\ell}\left[\mathrm{I}_n + g\,\mathrm{X}_\ell \left(\mathrm{X}_\ell^{*T}\mathrm{X}_\ell^*\right)^{-1}\mathrm{X}_\ell{}^T\right]\right\}. \quad (19)$$

– Under the Jeffreys baseline prior (9), $m_\ell^N(\boldsymbol{y}|\mathrm{X}_\ell\,, \mathrm{X}_\ell^*)$ is given by (11) with data $(\boldsymbol{y}, \mathrm{X}_\ell)$ (it is improper).

Estimation of the marginal likelihood

Two possible Monte-Carlo estimates.

(1) Generate $\boldsymbol{y}^{*(t)}$ $(t = 1, \ldots, T)$ from $m_\ell^N(\boldsymbol{y}^*|\boldsymbol{y}, \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)$ and estimate the marginal likelihood by

$$\hat{m}_\ell^{PEP}(\boldsymbol{y}|\mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) = m_\ell^N(\boldsymbol{y}|\mathrm{X}_\ell, \mathrm{X}_\ell^*)\left[\frac{1}{T}\sum_{t=1}^{T}\frac{m_0^N(\boldsymbol{y}^{*(t)}|\mathrm{X}_0^*, \delta)}{m_\ell^N(\boldsymbol{y}^{*(t)}|\mathrm{X}_\ell^*, \delta)}\right]. \qquad (20)$$

Estimation of the marginal likelihood

Two possible Monte-Carlo estimates.

(1) Generate $\boldsymbol{y}^{*(t)}$ $(t = 1, \ldots, T)$ from $m_\ell^N(\boldsymbol{y}^*|\boldsymbol{y}, \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)$ and estimate the marginal likelihood by

$$\hat{m}_\ell^{PEP}(\boldsymbol{y}|\mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) = m_\ell^N(\boldsymbol{y}|\mathrm{X}_\ell, \mathrm{X}_\ell^*) \left[\frac{1}{T} \sum_{t=1}^{T} \frac{m_0^N(\boldsymbol{y}^{*(t)}|\mathrm{X}_0^*, \delta)}{m_\ell^N(\boldsymbol{y}^{*(t)}|\mathrm{X}_\ell^*, \delta)} \right]. \quad (20)$$

(2) Generate $\boldsymbol{y}^{*(t)}$ $(t = 1, \ldots, T)$ from $m_\ell^N(\boldsymbol{y}^*|\,\boldsymbol{y}; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)$ and estimate the marginal likelihood by

$$\hat{m}_\ell^{PEP}(\boldsymbol{y}|\mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) =$$

$$= m_0^N(\boldsymbol{y}|\mathrm{X}_0, \mathrm{X}_0^*) \left[\frac{1}{T} \sum_{t=1}^{T} \frac{m_\ell^N(\boldsymbol{y}|\,\boldsymbol{y}^{*(t)}; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)}{m_0^N(\boldsymbol{y}|\,\boldsymbol{y}^{*(t)}; \mathrm{X}_0, \mathrm{X}_0^*, \delta)} \frac{m_0^N(\boldsymbol{y}^{*(t)}|\,\boldsymbol{y}; \mathrm{X}_0, \mathrm{X}_0^*, \delta)}{m_\ell^N(\boldsymbol{y}^{*(t)}|\,\boldsymbol{y}; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)} \right]. \quad (21)$$

Some Comments for the MC schemes

Some Comments for the MC schemes

– Monte-Carlo schemes (1) and (2) generate imaginary data from the posterior predictive distribution of the model under consideration.

Some Comments for the MC schemes

- Monte-Carlo schemes (1) and (2) generate imaginary data from the posterior predictive distribution of the model under consideration.
- Thus we expect them to be relatively accurate.

Some Comments for the MC schemes

– Monte-Carlo schemes (1) and (2) generate imaginary data from the posterior predictive distribution of the model under consideration.

– Thus we expect them to be relatively accurate.

– For Monte-Carlo scheme 2, we only need to evaluate posterior predictive distributions when we estimate Bayes factors. These are available even in the case of improper baseline priors such as the Jeffrey baseline used in J=PEP.

Some Comments for the MC schemes

– Monte-Carlo schemes (1) and (2) generate imaginary data from the posterior predictive distribution of the model under consideration.

– Thus we expect them to be relatively accurate.

– For Monte-Carlo scheme 2, we only need to evaluate posterior predictive distributions when we estimate Bayes factors. These are available even in the case of improper baseline priors such as the Jeffrey baseline used in J=PEP.

– The marginal likelihoods for J-PEP and Z-PEP result in the same posterior odds and model probabilities for $g \to \infty$, $a_\ell = -\frac{d_\ell}{2}$ and $b_\ell = 0$.

6 Consistency of the J-PEP Bayes factor

Theorem 1. *For any two models M_ℓ, $M_k \in \mathcal{M} \setminus \{M_0\}$ and for large n, we have that*

$$-2 \log BF^{J\text{-}PEP}_{\ell\, k} \approx n \log \frac{RSS_\ell}{RSS_k} + (d_\ell - d_k) \log n = BIC_\ell - BIC_k \,. \qquad (22)$$

Therefore the J-PEP approach has the same asymptotic behavior as the BIC-based variable-selection procedure.

6 Consistency of the J-PEP Bayes factor

Theorem 1. *For any two models M_ℓ, $M_k \in \mathcal{M} \setminus \{M_0\}$ and for large n, we have that*

$$-2 \log BF_{\ell\, k}^{J\text{-}PEP} \approx n \log \frac{RSS_\ell}{RSS_k} + (d_\ell - d_k) \log n = BIC_\ell - BIC_k \,. \tag{22}$$

Therefore the J-PEP approach has the same asymptotic behavior as the BIC-based variable-selection procedure.

Lemma 1. *Let $M_\ell \in \mathcal{M}$ be a normal regression model of type (2) such that*

$$\lim_{n \to \infty} \frac{\mathrm{X}_T \big(\mathrm{I}_n - \mathrm{X}_\ell (\mathrm{X}_\ell^T \mathrm{X}_\ell)^{-1} \mathrm{X}_\ell^T \big) \mathrm{X}_T}{n} \text{ is a positive semidefinite matrix,}$$

with X_T being the design matrix of the true data generating regression model $M_T \neq M_j$. Then, the variable selection procedure based on J-PEP Bayes factor is consistent since $BF_{jT}^{J-PEP} \to 0$ as $n \to \infty$.

7 Simulated Example

- We use the simulation scheme used in Nott & Kohn (2005, Biometrika).
- We generate data-sets of size $n = 50$ observations and $p = 15$ covariates.
- For $i = 1, \ldots, n$, We generate covariates using the following scheme:

$$\begin{aligned} X_{ij} &\sim N(\mu_{ij}, 1) \text{ with} \\ \mu_{ij} &= 0 \text{ for } j = 1, \ldots, 10 \text{ and} \\ \mu_{ij} &= 0.3X_{i1} + 0.5X_{i2} + 0.7X_{i3} + 0.9X_{i4} + 1.1X_{i5} \text{ for } j = 11, \ldots, 15; \end{aligned}$$

while the response is generated from

$$Y_i \sim N\Big(\, 4 + 2X_{i1} - X_{i5} + 1.5X_{i7} + X_{i11} + 0.5X_{i13},\ 2.5^2\, \Big).$$

- Full enumeration is feasible for all $2^{15} = 32,768$ models.

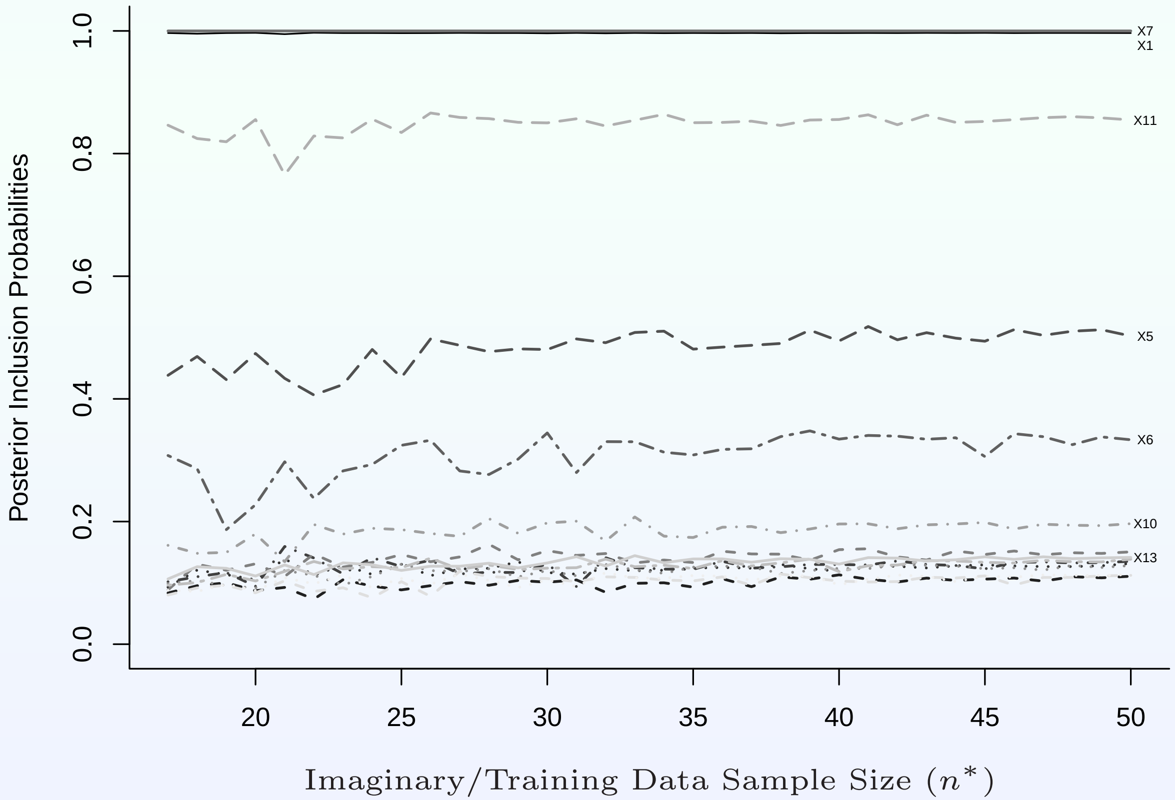
PEP prior results

Table 1: *Posterior model probabilities for the best models, together with Bayes factors for the Z-PEP MAP model* (M_1) *against* $M_j, j = 2, \ldots, 7$, *for the Z-PEP and J-PEP prior methodologies.*

		Z-PEP		J-PEP		
M_j	Predictors	Posterior Model Probability	Bayes Factor	Rank	Posterior Model Probability	Bayes Factor
1	$X_1 + X_5 + X_7 + X_{11}$	0.0783	1.00	(2)	0.0952	1.00
2	$X_1 + X_7 + X_{11}$	0.0636	1.23	(1)	0.1054	0.90
3	$X_1 + X_5 + X_6 + X_7 + X_{11}$	0.0595	1.32	(3)	0.0505	1.88
4	$X_1 + X_6 + X_7 + X_{11}$	0.0242	3.23	(4)	0.0308	3.09
5	$X_1 + X_7 + X_{10} + X_{11}$	0.0175	4.46	(5)	0.0227	4.19
6	$X_1 + X_5 + X_7 + X_{10} + X_{11}$	0.0170	4.60	(9)	0.0146	6.53
7	$X_1 + X_5 + X_7 + X_{11} + X_{13}$	0.0163	4.78	(10)	0.0139	6.87

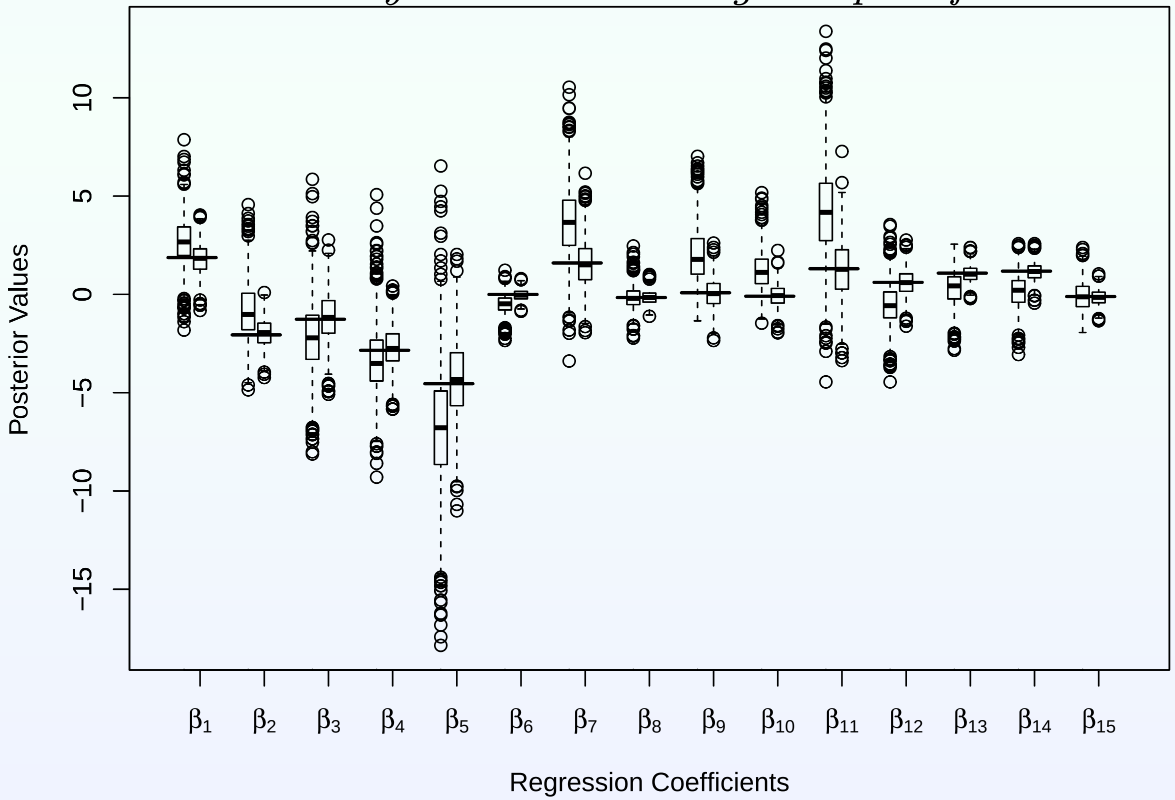
Sensitivity analysis on imaginary sample size

Figure 1: *Posterior marginal inclusion probabilities, for n^* values from 17 to $n = 50$, with the Z-PEP prior methodology.*



Sensitivity analysis on imaginary sample size (cont.)

Figure 2: *Boxplots of the posterior distributions of the regression coefficients. For each coefficient, the left-hand boxplot summarizes the EPP results and the right-hand boxplot displays the Z-PEP posteriors; solid lines in both posteriors identify the MLEs. We used the first 20 observations from the simulated data-set and a randomly selected training sample of size* $n^* = 17$*.*



Comparisons with IBF and J-EPP approaches

– We compare the Bayes factors between the two best models ($X_1 + X_5 + X_7 + X_{11}$ versus $X_1 + X_5 + X_7$) for J-PEP, ZPEP, J-EPP and IBF.

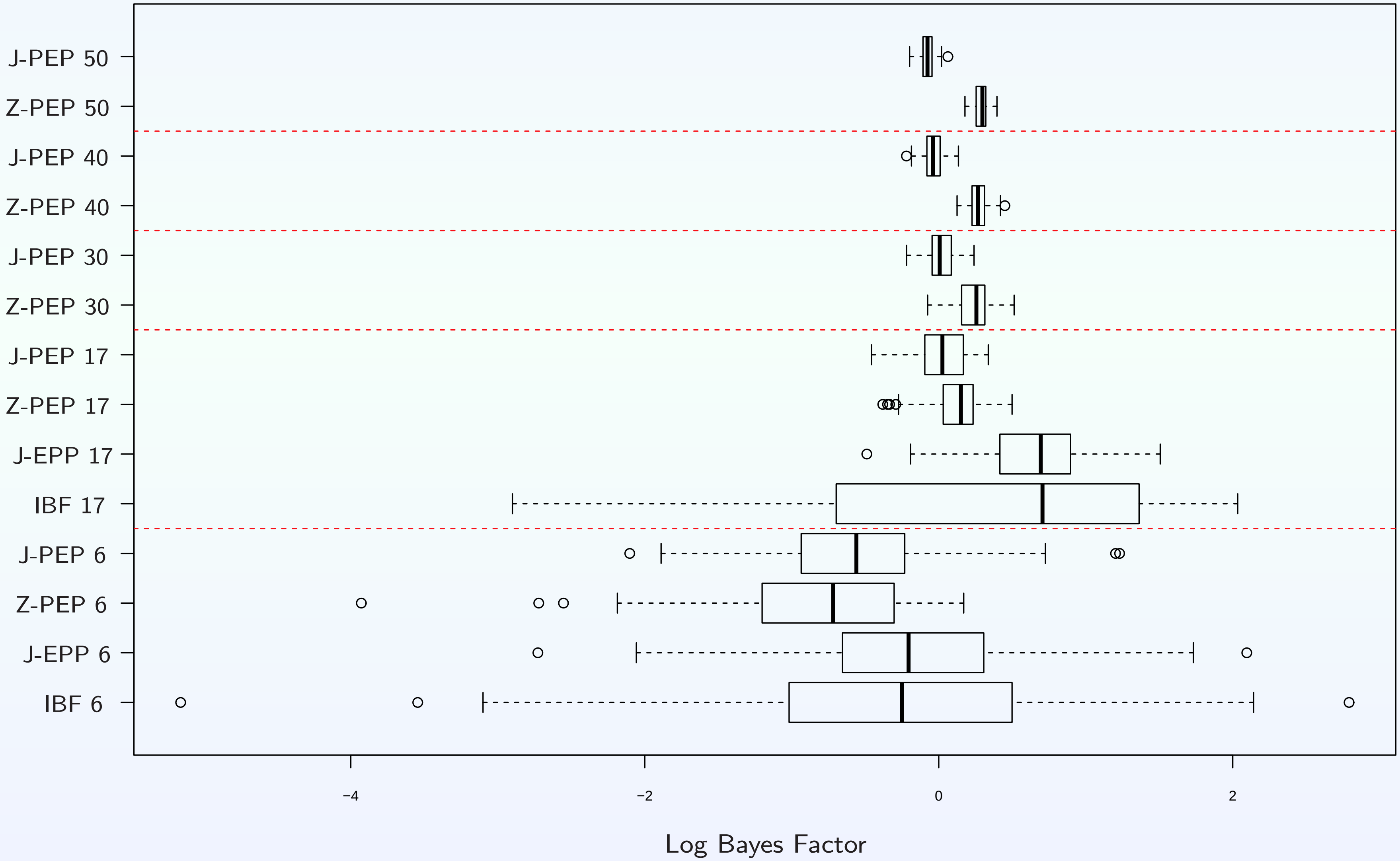
– For IBF and J-EPP $\Rightarrow$ 100 randomly selected training samples of size:

- $n^* = 6$ (minimal training samples for these two models) and
- $n^* = 17$ (minimal training sample for the full model with $p = 15$ covariates),

– For PEP we randomly select 100 training samples of sizes $n^* = 6,\ 17, 20, 25, 30, 35, 40, 45, 50.$

– Each marginal likelihood estimate is obtained with 1000 iterations.

Comparisons with IBF and J-EPP approaches (cont.)



8 Real Life Example: Ozone data

– Source: Breinman & Friedman (1985, JASA).

– Response: The logarithm of the **ozone concentration** variable of the original data set.

– 56 covariates: 9 main effects, 9 quadratic terms, 2 cubic terms, and 36 two-way interactions.

• The main effects we considered are the following:

X_1	Day of Year
X_2	Wind speed (mph) at LAX
X_3	500 mb pressure height (m) at VAFB
X_4	Humidity (%) at LAX
X_5	Temperature (°F) at Sandburg
X_6	Inversion base height (feet) at LAX
X_7	Pressure gradient (mm Hg) from LAX to Daggett
X_8	Inversion base temperature (°F) at LAX
X_9	Visibility (miles) at LAX

– All main effects and the response where standardised.

Searching the model space

(1) **Large model space with $2^{56} = 7.2057610^{16}$ models**

- Run MC^3 to approximate posterior marginal inclusion probabilities $P(\gamma_j = 1|\boldsymbol{y})$.
- We created a reduced model space with covariates having marginal inclusion probabilities ≥ 0.3.

(2) **Reduced model space:** Run again MC^3 to accurately estimate:

- posterior marginal inclusion probabilities.
- posterior model probabilities and odds ratios.

MCMC details

– 100,000 iterations for MC^3 for Z-PER and EIBF (arithmetic mean of IBFs over different minimal training samples).

– 30 randomly-selected minimal training samples for size $n^* = 58$ for EIBF

– For the threshold posterior inclusion probability value of 0.3,

p: $56 \rightarrow 22$ covariates and

the number of models under consideration: $7.2057610^{16} \rightarrow 4,194,304$.

Reduced space

Variables common in all three analyses were: $X_1 + X_2 + X_8 + X_9 + X_{10} + X_{15} + X_{16} + X_{18} + X_{43}$

J-PEP

J-PEP	Z-PEP	EIBF	Additional Variables	# of Covariates	PO_{1k}
1	(>5)	(>5)		9	1.00
2	(1)	(5)	$X_7 + X_{12} + X_{13} + X_{20}$	13	1.29
3	(>5)	(>5)	$X_7 + X_{13} + X_{20}$	12	1.46

Z-PEP

Z-PEP	J-PEP	EIBF	Additional Variables	# of Covariates	PO_{1k}
1	(2)	(5)	$X_7 + X_{12} + X_{13} + X_{20}$	13	1.00
2	(>5)	(>5)	$X_5 + X_7 + X_{12} + X_{13} + X_{20}$	14	1.19
3	(>5)	(3)	$X_5 + X_7 + X_{12} + X_{13} + X_{20} + X_{42}$	15	1.77

EIBF

EIBF	J-PEP	Z-PEP	Additional Variables	# of Covariates	PO_{1k}
1	(>5)	(4)	$X_7 + X_{12} + X_{13} + X_{20} + X_{42}$	14	1.00
2	(>5)	(>5)	$X_5 + X_7 + X_{12} + X_{13} + X_{20} + X_{26} + X_{42}$	16	1.17
3	(>5)	(3)	$X_5 + X_7 + X_{12} + X_{13} + X_{20} + X_{42}$	15	1.30

Comparison of the predictive performance

– We evaluate the out-of-sample predictive performance of the two highest a-posteriori models.

– We consider 50 randomly selected half-splits.

– For each split, we generate an MCMC sample of T iterations from the model of interest M_ℓ and then calculate the average root mean square error by

$$ARMSE_\ell = \frac{1}{T}\sum_{t=1}^{T} RMSE_\ell^{(t)} \quad \text{with} \quad RMSE_\ell^{(t)} = \sqrt{\frac{1}{n_V}\sum_{i\in\mathcal{V}}\left(y_i - \widehat{y}_{i|M_\ell}^{(t)}\right)^2}.$$

- $RMSE_\ell^{(t)} \Rightarrow$ root mean square error for the validation dataset V of size n_V calculated for the t-iteration of the MCMC
- $\widehat{y}_{i|M_\ell}^{(t)} = \mathrm{X}_{\ell(i)}\boldsymbol{\beta}_\ell^{(t)}$ are the expected values of y_i under model M_ℓ for iteration t
- $\boldsymbol{\beta}_\ell^{(t)}$ is the vector of the model parameters for iteration t and
- $\mathrm{X}_{\ell(i)}$ is the i-th row of matrix X_ℓ of model M_ℓ.

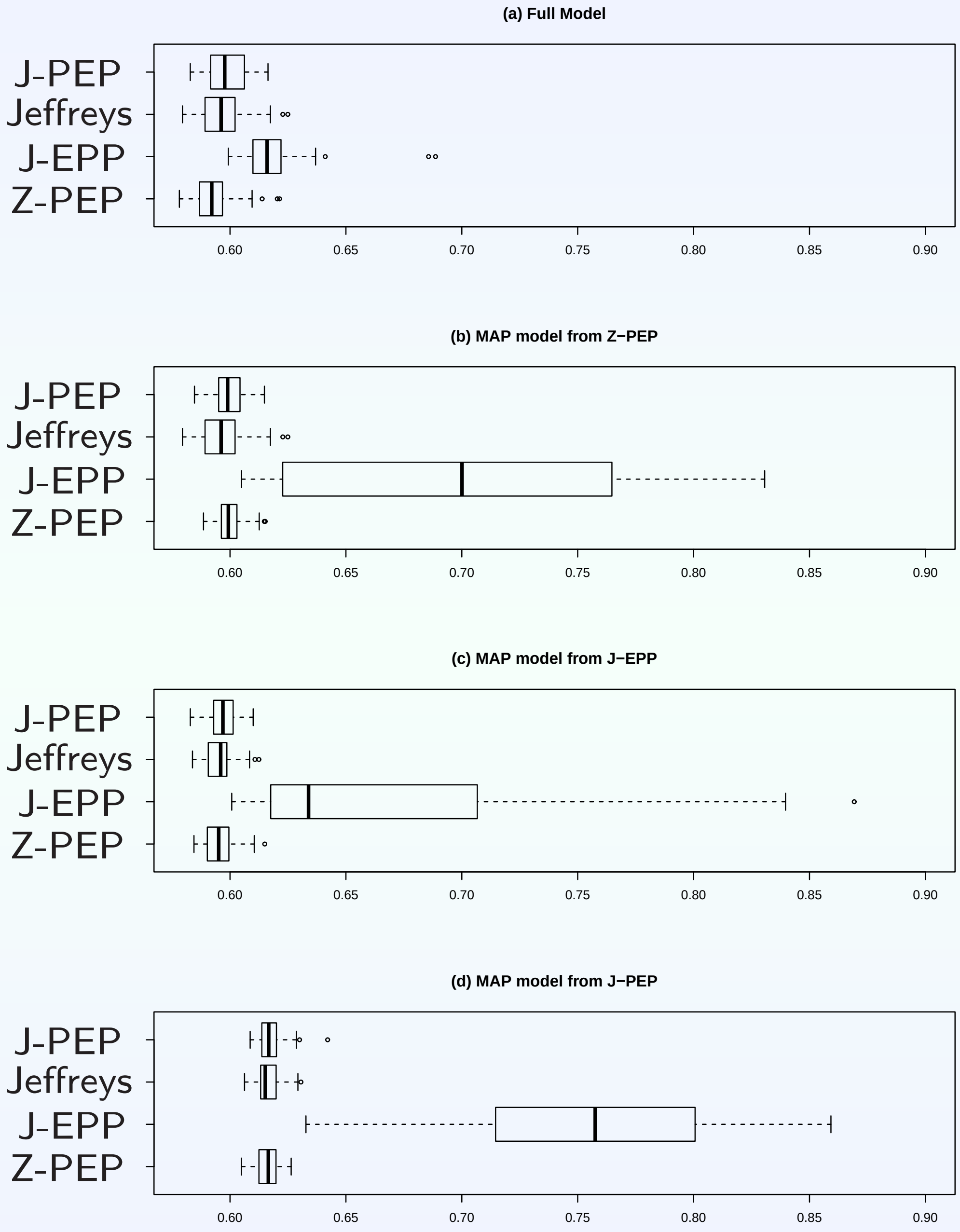
Comparison of the predictive performance (cont.)

Model	d_ℓ	R^2	R^2_{adj}	*RMSE** J-PEP	Z-PEP	J-EPP	Jeffreys Prior
Full	22	0.8500	0.8392	0.5988	0.5935	0.6194	0.5972
				(0.0087)	(0.0097)	(0.0169)	(0.0104)
J-PEP MAP	9	0.8070	0.8016	0.5975	0.6161	0.7524	0.6165
				(0.0063)	(0.0051)	(0.0626)	(0.0052)
Z-PEP MAP	13	0.8370	0.8303	0.5994	0.5999	0.6982	0.5994
				(0.0071)	(0.0060)	(0.0734)	(0.0049)
EIBF MAP	14	0.8398	0.8326	0.6182	0.5961	0.6726	0.5958
				(0.0066)	(0.0072)	(0.0800)	(0.0061)

Comparison with the full model (percentage changes)

Model	d_ℓ	R^2	R^2_{adj}	*RMSE* J-PEP	Z-PEP	J-EPP	Jeffreys Prior
J-PEP MAP	−59%	−5.06%	−4.48%	−0.22%	+3.81%	+21.5%	+3.23%
Z-PEP MAP	−41%	−1.50%	−1.06%	+0.10%	+1.01%	+12.7%	+0.37%
EIBF MAP	−36%	−1.20%	−0.78%	+3.24%	+0.44%	+10.9%	−0.23%

Note: *Mean (standard deviation) over 50 different split-half out-of-sample evaluations.


(a) Full Model
J-PEP
Jeffreys
J-EPP
Z-PEP
0.60 0.65 0.70 0.75 0.80 0.85 0.90
(b) MAP model from Z−PEP
J-PEP
Jeffreys
J-EPP
Z-PEP
0.60 0.65 0.70 0.75 0.80 0.85 0.90
(c) MAP model from J−EPP
J-PEP
Jeffreys
J-EPP
Z-PEP
0.60 0.65 0.70 0.75 0.80 0.85 0.90
(d) MAP model from J−PEP
J-PEP
Jeffreys
J-EPP
Z-PEP
0.60 0.65 0.70 0.75 0.80 0.85 0.90

9 Discussion

Major contribution:

Simultaneously produce a minimally-informative prior and sharply diminish the effect of training samples on previously-studied expected-posterior-prior (EPP) methodology.

(a)
- Generally, in the EPP approach the training data y^* are generated directly from the prior predictive distribution of a reference model.
- Nevertheless, the choice of the training sub-samples for the covariates remains open in the regression set-up.
- Using our approach, we can work with training-samples of size equal to the size of the full data set. Hence, we avoid the selection of such subsamples by choosing $X^* = X$.

(b)
- The full model is usually specifies the size of the minimal training sample.
- Thus, for large $p \rightarrow n$, the effect of the minimal training sample will be large $\Rightarrow$ informative priors.

Some further conclusions based on empirical evidence

- is systematically more parsimonious (under either baseline prior choice) than the J-EPP approach;
- is robust to the size of the training sample, thus supporting the use of the entire data set as a "training sample" and thereby promoting stability and fast computation;
- good out-of-sample predictive performance for the selected maximum a-posteriori model;
- has low impact on the posterior distribution even when n is not much larger than p.

Some further conclusions based on empirical evidence

- is systematically more parsimonious (under either baseline prior choice) than the J-EPP approach;
- is robust to the size of the training sample, thus supporting the use of the entire data set as a "training sample" and thereby promoting stability and fast computation;
- good out-of-sample predictive performance for the selected maximum a-posteriori model;
- has low impact on the posterior distribution even when n is not much larger than p.


THE
END