Bayes factors for linear regression

Rui Wang

Sunday 3rd March, 2019

1 Introduction

Bayes factor is proposed by Jeffreys. This note gives a review for Bayes factors for linear regression.

2 Bayes factor

2.1 How it works

Suppose that M models are proposed for the data $\mathbf{x} = (x_1, \ldots, x_n)$. Under model $\mathcal{M}_i, i = 0, \ldots, M$, the data are related to parameter θ_i by a density $f_i(\mathbf{x}|\theta_i)$. Often, the null model $\mathcal{M}_0$ is nested in all other models. In this case, we can write $\theta_0 = \boldsymbol{\alpha}$, $\theta_i = (\boldsymbol{\alpha}, \boldsymbol{\beta}_i)$, $i = 1, \ldots, M$, where $\boldsymbol{\alpha}$ is the common parameter and $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_M$ are model specific parameters.

Let $\Pr(\mathcal{M}_i)$ be the prior probability of model $\mathcal{M}_i$. Under $\mathcal{M}_i$, let $\pi_i(\theta_i)$ be the prior density for θ_i. The marginal density under model $\mathcal{M}_i$ is

$$m_i(\mathbf{x}) = \int_{\Theta_i} f_i(\mathbf{x}|\theta_i)\pi_i(\theta_i)d\theta_i$$

Hence the posterior probability of $\mathcal{M}_i$ is

$$\Pr(\mathcal{M}_i|\mathbf{x}) = \frac{m_i(\mathbf{x}) \Pr(\mathcal{M}_i)}{\sum_{i=0}^{M} m_i(\mathbf{x}) \Pr(\mathcal{M}_i)} = \frac{B_{ij}(\mathbf{x}) \Pr(\mathcal{M}_i)}{\Pr(\mathcal{M}_j) + \sum_{i \neq j} B_{ij}(\mathbf{x}) \Pr(\mathcal{M}_i)},$$

where $B_{ij}(\mathbf{x}) = m_i(\mathbf{x})/m_j(\mathbf{x})$ is the Bayes factor of $\mathcal{M}_i$ to $\mathcal{M}_j$. If $\mathcal{M}_0$ is the null model, then the null-Based approach takes $j = 0$. It can be seen that the posterior model probability only depends on the Bayes factor. Note that

$$\frac{\Pr(\mathcal{M}_i|\mathbf{x})}{\Pr(\mathcal{M}_j|\mathbf{x})} = B_{ij}(\mathbf{x}) \frac{\Pr(\mathcal{M}_i)}{\Pr(\mathcal{M}_j)}.$$

Hence the Bayes factor is the ratio of the posterior odds to its prior odds. The logarithm of Bayes factor is called *weight of evidence*. See Kass and Raftery (1995) for a review for the interpretation, computation and applications of Bayes factor.

Consider hypothesis testing problem, that is, to compare two models $\mathcal{M}_i$ and $\mathcal{M}_j$. Bayesian hypothesis testing utilize Bayes factor, and reject the $\mathcal{M}_i$ if $B_{ji}(\mathbf{x})$ is larger than certain threshold, say, 20; see Kass and Raftery (1995). Often, the Bayes factor is consistent in the sense that as $n \to \infty$,

$$B_{ji}(\mathbf{x}) \xrightarrow{P} 0 \quad \text{if the true distribution belongs to } \mathcal{M}_j,$$
$$B_{ji}(\mathbf{x}) \xrightarrow{P} +\infty \quad \text{if the true distribution belongs to } \mathcal{M}_i.$$

If the Bayes factor is consistent, then Bayesian test criterion has the property that the frequentist type I error rate and type II error rate both tend to 0. This property is very different from frequentist significant tests which fix the type I error rate to the significant level, say, 0.05. With this property, the Bayesian hypothesis testing method can be easily generalized to multiple hypothesis testing and model selection. In fact, if the Bayes factor $B_{ij}(\mathbf{x})$ is consistent for all i, j, then the posterior model probability has the following property

$$\Pr(\mathcal{M}_i|\mathbf{x}) \xrightarrow{P} 1 \quad \text{if the true distribution belongs to } \mathcal{M}_j,$$
$$\Pr(\mathcal{M}_i|\mathbf{x}) \xrightarrow{P} 0 \quad \text{if the true distribution belongs to } \mathcal{M}_i.$$

There are many papers consider the consistency of Bayes factor for specefic models.

2.2 Difficulties

Berger and Pericchi (2001) listed several reasons to use Bayes factor. Also, they illustrated several difficulties with Bayes factor. One of the most serious difficulties is how to choose the prior distributions of parameters.

In the field of objective Bayesian analysis, "noninformative" (or "default", "automatic") priors are used. The most commonly used noninformative priors are Jeffreys priors and reference priors. See Bernardo (2005) for a good introduction to reference priors. These noninformative priors are typically improper. However, improper noninformative priors yield indeterminate answers. To see this, suppose we use improper priors π_i and π_j to compute Bayes factor $B_{ij}(\mathbf{x})$. Since priors are improper, one could have just as well used $c_i\pi_i$ and $c_j\pi_j$, which results in the Bayes factor $(c_i/c_j)B_{ij}(\mathbf{x})$. Thus, "default" priors can not be used as default to do model selection.

2.3 Objective Bayesian model selection

2.3.1 Conventional prior approach

(Jeffreys, 2003, Chapter 5) dealt with the issue of indeterminacy of noninformative priors by (i) using noninformative priors only for common (orthogonal) parameters in the models, so that the arbitrary multiplicative constant for the priors would cancel in all Bayes factors, and (ii) using default proper priors for parameters that would occur in one model but not the other. This line of development has been successfully followed by many others.

Example 1. *Suppose $X_1, \ldots, X_n$ are iid $\mathcal{N}(\mu, \sigma_2^2)$ under $\mathcal{M}_2$. Under $\mathcal{M}_1$, the X_i are $\mathcal{N}(0, \sigma_1^2)$. In this situation the mean and variance are orthogonal parameters, in which case Jeffreys argues that σ_1^2 and σ_2^2 do have the same meanning across models and can be identified as $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Because of this identification, Jeffreys suggests that the variances can be assigned the same (improper) noninformative prior $\pi^J(\sigma) = 1/\sigma$, since the indeterminate multiplicative constant for the prior would cancel in the Bayes factor. As the unknown mean μ occurs in only $\mathcal{M}_2$, it needs to assigned a proper prior. Through a series of ingenious arguments, Jeffreys obtains the following desiderata that this proper prior should satisfy: i) it should be centered at zero (i.e., centered at $\mathcal{M}_1$); ii) have scale σ; iii) be symmetric around zero; and iv) have no moments. He argues that the simplest distribution that satisfies these conditions is the Cauchy$(0, \sigma^2)$. In summary, Jeffreys's conventional prior for this problem is:*

$$\pi_1^J(\sigma_1) = \frac{1}{\sigma_1}, \quad \pi_2^J(\mu, \sigma_2) = \frac{1}{\sigma_2} \cdot \frac{1}{\pi\sigma_2(1 + \mu^2/\sigma_2^2)}.$$

Although this solution appears to be rather ad hoc, it is quite reasonable; choosing the scale of the prior for μ to be σ_2 and centering it at $\mathcal{M}_1$ are natural choices, and Cauchy priors are known to be robust in various ways.

Example 2. *Suppose $\mathbf{Y} \in \mathbb{R}^n$ is generated from the model*

$$\mathcal{M}_\gamma : \mathbf{Y} = \mathbf{1}_n\alpha + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{X} \in \mathbb{R}^{n\times p}$ and $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \phi^{-1}\mathbf{I}_n)$. Let $\mathbf{X}_\gamma \in \mathbb{R}^{n\times p_\gamma}$ be a submatrix of $\mathbf{X}$. Then the submodel $\mathcal{M}_\gamma$ is defined as

$$\mathcal{M}_\gamma : \mathbf{Y} = \mathbf{1}_n\alpha + \mathbf{X}_\gamma\boldsymbol{\beta}_\gamma + \boldsymbol{\varepsilon}.$$

The null model $\mathcal{M}_N$ is

$$\mathcal{M}_N : \mathbf{Y} = \mathbf{1}_n\alpha + \boldsymbol{\varepsilon}.$$

We would like to compare $\mathcal{M}_\gamma$ with $\mathcal{M}_N$. Without loss of generality, we assume $\mathbf{1}_n^\top\mathbf{X}_\gamma = 0$.

Zellner and Siow (1980) proposed the following priors. Under $\mathcal{M}_N$,

$$p(\alpha|\phi, \mathcal{M}_N) \propto 1, \quad p(\phi|\mathcal{M}_N) = \frac{1}{\phi}.$$

Under $\mathcal{M}_\gamma$,

$$\pi(\boldsymbol{\beta}_\gamma|\phi, \mathcal{M}_\gamma) \propto \frac{\Gamma(p_\gamma)}{\pi^{p_\gamma/2}} \left|\frac{\mathbf{X}_\gamma^\top\mathbf{X}_\gamma}{n/\phi}\right|^{1/2} \left(1 + \boldsymbol{\beta}_\gamma^\top \frac{\mathbf{X}_\gamma^\top\mathbf{X}_\gamma}{n/\phi}\boldsymbol{\beta}_\gamma\right)^{-p_\gamma/2}, \quad p(\alpha|\phi, \mathcal{M}_\gamma) \propto 1, \quad p(\phi|\mathcal{M}_\gamma) = \frac{1}{\phi}.$$

(Here need to be further confirmed.) Thus, the improper priors of the "common" (α, ϕ) are assumed to be the same for the two models (again justifiable by orthogonality), while the conditional prior of the parameter β_2, given σ, is assumed to be the (proper) p-dimensional Cauchy distirbution, with location at 0 (so that it is 'centered' at $\mathcal{M}_1$) and covariance matrix $\phi\mathbf{X}_\gamma^\top\mathbf{X}_\gamma/n$, "... a matrix suggested by the form of the information matrix," to quote Zellner and Siow (1980).

Example 3 (g-priors)**.** *Under $\mathcal{M}_N$, the g prior is*

$$p(\alpha, \phi | \mathcal{M}_N) = \frac{1}{\phi}.$$

Under $\mathcal{M}_\gamma$, the g prior is

$$\boldsymbol{\beta}_\gamma | \phi, \mathcal{M}_\gamma \sim \mathcal{N}(0, \frac{g}{\phi}(\mathbf{X}_\gamma^\top \mathbf{X}_\gamma)^{-1}), \quad p(\alpha|\phi, \mathcal{M}_\gamma) \propto 1, \quad p(\phi|\mathcal{M}_\gamma) = \frac{1}{\phi}.$$

The joint pdf is

$$\begin{aligned} & p(\mathbf{Y}, \alpha, \boldsymbol{\beta}_\gamma, \phi | \mathcal{M}_\gamma) = p(\mathbf{Y}|\alpha, \boldsymbol{\beta}_\gamma, \phi, \mathcal{M}_\gamma) p(\boldsymbol{\beta}_\gamma | \phi, \mathcal{M}_\gamma) p(\alpha|\phi, \mathcal{M}_\gamma) p(\phi|\mathcal{M}_\gamma) \\ =& (2\pi)^{-(n+p_\gamma)/2} g^{-p_\gamma/2} \phi^{(n+p_\gamma)/2-1} |\mathbf{X}_\gamma^\top \mathbf{X}_\gamma|^{1/2} \exp\left\{ -\frac{n\phi}{2}(\bar{\mathbf{Y}} - \alpha)^2 \right\} \\ & \exp\left\{ -\frac{\phi(g+1)}{2g} \left\| \mathbf{X}_\gamma \left(\boldsymbol{\beta}_\gamma - \frac{g}{g+1} \hat{\boldsymbol{\beta}}_\gamma \right) \right\|^2 - \frac{\phi}{2(g+1)} \left\| \mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma \right\|^2 - \frac{\phi}{2} \left\| \mathbf{Y} - \mathbf{1}_n \bar{\mathbf{Y}} - \mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma \right\|^2 \right\}, \end{aligned}$$

where $\bar{\mathbf{Y}} = n^{-1} \mathbf{1}_n^\top \mathbf{Y}$, $\hat{\boldsymbol{\beta}}_\gamma = (\mathbf{X}_\gamma^\top \mathbf{X}_\gamma)^{-1} \mathbf{X}_\gamma^\top \mathbf{Y}$.

Direct calculation yields

$$p(\mathbf{Y}|\mathcal{M}_\gamma, g) = \frac{\Gamma((n-1)/2)}{\pi^{(n-1)/2}\sqrt{n}} \left\| \mathbf{Y} - \mathbf{1}_n \bar{\mathbf{Y}} \right\|^{-(n-1)} \frac{(1+g)^{(n-p_\gamma-1)/2}}{[1 + g(1 - R_\gamma^2)]^{(n-1)/2}},$$

where $R_\gamma^2 = 1 - \|\mathbf{Y} - \mathbf{1}_n \bar{\mathbf{Y}} - \mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma\|^2 / \|\mathbf{Y} - \mathbf{1}_n \bar{\mathbf{Y}}\|^2$. Also, we have

$$p(\mathbf{Y}|\mathcal{M}_N) = \frac{\Gamma((n-1)/2)}{\pi^{(n-1)/2}\sqrt{n}} \left\| \mathbf{Y} - \mathbf{1}_n \bar{\mathbf{Y}} \right\|^{-(n-1)}.$$

Thus,

$$BF[\mathcal{M}_\gamma : \mathcal{M}_N] = (1+g)^{(n-p_\gamma-1)/2} [1 + g(1-R_\gamma^2)]^{-(n-1)/2}.$$

Example 4. *Choices of g.*

Local empirical Bayes. *The local EB estimates a separate g for each model $\mathcal{M}_\gamma$.*

$$\hat{g}_\gamma^{EBL} = \arg\max_{g \geq 0} p(\mathbf{Y}|\mathcal{M}_\gamma, g) = \arg\max_{g \geq 0} \frac{(1+g)^{(n-p_\gamma-1)/2}}{[1 + g(1-R_\gamma^2)]^{(n-1)/2}} = \max\{F_\gamma - 1, 0\},$$

where

$$F_\gamma = \frac{R_\gamma^2 / p_\gamma}{(1 - R_\gamma^2)/(n - 1 - p_\gamma)}$$

is the usual F statistic for testing $\boldsymbol{\beta}_\gamma = 0$.

Global empirical Bayes. *The global EB procedure assumes one common g for all models.*

$$\hat{g}_\gamma^{EBG} = \arg\max_{g \geq 0} \sum_\gamma p(\mathcal{M}_\gamma) p(\mathbf{Y}|\mathcal{M}_\gamma, g) = \arg\max_{g \geq 0} \sum_\gamma p(\mathcal{M}_\gamma) \frac{(1+g)^{(n-p_\gamma-1)/2}}{[1 + g(1-R_\gamma^2)]^{(n-1)/2}}.$$

In general, this marginal likelihood is not tractable and does not provide a closed-form solution for $\hat{g}_\gamma^{EBG}$. It can be computed by an EM algorithm, which is based on treating both the model indicator and the precision ϕ as latent data.

Example 5 (Mixtures of g priors)**.** *Under $\mathcal{M}_\gamma$, the mixtures of g prior take the form*

$$\boldsymbol{\beta}_\gamma|g,\phi,\mathcal{M}_\gamma \sim \mathcal{N}(0, \frac{g}{\phi}(\mathbf{X}_\gamma^\top \mathbf{X}_\gamma)^{-1}), \quad \pi(g), \quad p(\alpha|\phi,\mathcal{M}_\gamma) \propto 1, \quad p(\phi|\mathcal{M}_\gamma) = \frac{1}{\phi}.$$

Zellner-Siow Priors

$$\pi(\boldsymbol{\beta}_\gamma|\phi) \propto \frac{\Gamma(p_\gamma)}{\pi^{p_\gamma/2}} \left| \frac{\mathbf{X}_\gamma^\top \mathbf{X}_\gamma}{n/\phi} \right|^{1/2} \left(1 + \boldsymbol{\beta}_\gamma^\top \frac{\mathbf{X}_\gamma^\top \mathbf{X}_\gamma}{n/\phi} \boldsymbol{\beta}_\gamma\right)^{-p_\gamma/2}$$

The Zellner-Siow priors can be represented as a mixture of g priors with an Inv-Gamma(1/2,n/2) prior on g, namely,

$$\phi(\boldsymbol{\beta}_\gamma|\phi) \propto \int \mathcal{N}(0, \frac{g}{\phi}(\mathbf{X}_\gamma^\top \mathbf{X}_\gamma)^{-1})\pi(g)dg,$$

with

$$\pi(g) = \frac{(n/2)^{1/2}}{\Gamma(1/2)} g^{-3/2} e^{-n/(2g)}.$$

Hyper-g priors

$$\pi(g) = \frac{a-2}{2}(1+g)^{-a/2}\mathbf{1}_{(0,\infty)}(g), \quad a > 2.$$

Equivalently,

$$\frac{g}{1+g} \sim Beta(1, \frac{a}{2} - 1).$$

The null-based Bayes factor is

$$\begin{aligned} BF[\mathcal{M}_\gamma : \mathcal{M}_N] =& \frac{a-2}{2} \int_0^\infty (1+g)^{(n-1-p_\gamma-a)/2}[1+(1-R_\gamma^2)g]^{-(n-1)/2} dg \\ =& \frac{a-2}{p_\gamma + a - 2} \times {}_2F_1\left(\frac{n-1}{2}, 1; \frac{p_\gamma + a}{2}; R_\gamma^2\right), \end{aligned}$$

where ${}_2F_1(a,b;c;z)$ is the Gaussian hypergeometric function defined as

$${}_2F_1(a,b;c;z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}}{(1-tz)^a} dt.$$

Beta prime prior *Maruyama and George (2011) proposed to use the beta prime prior for g:*

$$\pi(g) = \frac{g^b(1+g)^{-a-b-2}}{B(a+1,b+1)} \mathbf{1}_{(0,\infty)}(g),$$

where $a > -1$, $b > -1$. Equivalently,

$$\frac{1}{1+g} \sim Be(a+1, b+1).$$

They observed that the Bayes factor has a closed form if we take

$$b = \frac{n - p_\gamma - 5}{2} - a.$$

Bayarri et al. (2012) proposed a "robust prior" on g, which is a class of priors including the three considered by Liang et al. (2008) and some other related priors.

Example 6 (High-dimensional setting). *The asymptotic behaviors of the Bayes factors with g priors in high dimensional setting have been investigated by Mukhopadhyay et al. (2014), Wang and Maruyama (2017), Wang et al. (2016), Wang and Sun (2014), Wang (2017), Xiang et al. (2016), Mukhopadhyay and Samanta (2016).*

Generalization of g priors to $p > n$: Maruyama and George (2011), Shang and Clayton (2011).

2.3.2 Intrinsic prior

Intrinsic prior, introduced by Berger and Pericchi (1996) and further developed by Moreno et al. (1998), is a method for objective Bayes hypothesis testing.

A noninformative prior for θ_i is denoted by $\pi_i^N(\theta_i)$, $i = 1, 2$.

But the conventional Bayes factor suffers from arbitrary normalizing constant. To solve this problem, Berger and Pericchi (1996) proposed the intrinsic Bayes factor.

The intrinsic Bayes factor is based on training samples. This idea is to split the sample $\mathbf{x}$ into two parts as $\mathbf{x} = (x(l), x(n-l))$, where part $x(l)$, the training sample, is utilized to convert $\pi_i^N(\theta_i)$ into proper distributions,

$$\pi_i(\theta_i|x(l)) = \frac{f_i(x(l)|\theta_i)\pi_i^N(\theta_i)}{m_i^N(x(l))},$$

where $m_i^N(x(l)) = \int f_i(x(l)|\theta_i)\pi_i^N(\theta_i)d\theta_i$. With the remaining portion of the data $x(n-l)$, the Bayes factor is computed using the foregoing $\pi_i(\theta_i|x(l))$ as priors. The resulting partial Bayes factor is

$$B_{21}(x(n-l)|x(l)) := B_{21}(l) = B_{21}^N(\mathbf{x}) \cdot B_{12}^N(x(l)),$$

where

$$B_{12}^N(x(l)) = \frac{m_1^N(x(l))}{m_2^N(x(l))}.$$

Note that $B_{12}^N(l)$ does not depend on the arbitrary constants in $\pi_i^N(\theta_i)$. In addition, it is well defiend only if $x(l)$ is such that $0 < m_i^N(x(l)) < \infty$, $i = 1, 2$. If there is no subsample of $x(l)$ for which $0 < m_i^N(x(l)) < \infty$, $i = 1, 2$, then $x(l)$ is called a minimal training sample.

Berger and Pericchi (1996) suggested using a minimal training sample to compute $B_{21}(l)$ and to take an averate over all of the minumal training samples contained in the sample. This gives the arithmetic intrinsic Bayes factor (AIBF) of M_2 against M_1 as

$$B_{21}^{AI}(\mathbf{x}) = B_{21}^N(\mathbf{x}) \frac{1}{L} \sum_{i=1}^{L} B_{12}^N(x(l)),$$

where L is the number of minimal training samples $x(l)$ contained in $\mathbf{x}$.

Other averaging methods can also be used. The geometric intrinsic Bayes factor (GIBF) is defined by

$$B_{21}^{GI}(\mathbf{x}) = B_{21}^N(\mathbf{x}) \left(\prod_{i=1}^{L} B_{12}^N(x(l)) \right)^{1/L}.$$

However, IBF is not an actual Bayes factor and is not coherent in many aspect. An important question about the AIBF is to know whether it corresponds to an actual Bayes factor for sensible priors. Such a prior, if it exists, is called an intrinsic prior. Berger and Pericchi (1996) define intrinsic priors by using an (asymptotic) imaginary training sample.

Let $\pi_1(\theta_1)$ and $\pi_2(\theta_2)$ be certain priors. The corresponding Bayes factor is

$$B_{21}(\mathbf{x}) = \frac{\int_{\Theta_2} f_2(\mathbf{x}|\theta_2)\pi_2(\theta_2)d\theta_2}{\int_{\Theta_1} f_1(\mathbf{x}|\theta_1)\pi_1(\theta_1)d\theta_1}.$$

The following approximation is valid in the standard situation.

$$B_{21} = B_{21}^N \cdot \frac{\pi_2(\hat{\theta}_2)\pi_1^N(\hat{\theta}_1)}{\pi_2^N(\hat{\theta}_2)\pi_1(\hat{\theta}_1)} \cdot (1 + o_P(1)),$$

where $\hat{\theta}_i$ are the MLE's under M_i, $i = 1, 2$. Equating B_{21} and $B_{21}^{AI}(\mathbf{x})$, we have

$$\frac{\pi_2(\hat{\theta}_2)\pi_1^N(\hat{\theta}_1)}{\pi_2^N(\hat{\theta}_2)\pi_1(\hat{\theta}_1)} = \frac{1}{L}\sum_{i=1}^{L} B_{12}^N(x(l)) \quad \text{or} \quad \left(\prod_{i=1}^{L} B_{12}^N(x(l))\right)^{1/L}.$$

Suppose M_1 is the true model and θ_1 is the true parameter. Letting $n \to \infty$, we have

$$\frac{\pi_2(\psi_2(\theta_1))\pi_1^N(\theta_1)}{\pi_2^N(\psi_2(\theta_1))\pi_1(\theta_1)} = \mathrm{E}_{\theta_1}^{M_1} B_{12}^N(x(l)) \quad \text{or} \quad \exp\left(\mathrm{E}_{\theta_1}^{M_1} \log B_{12}^N(x(l))\right), \tag{1}$$

where $\psi_2(\theta_1)$ is the limiting MLE of θ_2. Similarly,

$$\frac{\pi_2(\theta_2)\pi_1^N(\psi_1(\theta_2))}{\pi_2^N(\theta_2)\pi_1(\psi_1(\theta_2))} = \mathrm{E}_{\theta_2}^{M_2} B_{12}^N(x(l)) \quad \text{or} \quad \exp\left(\mathrm{E}_{\theta_2}^{M_2} \log B_{12}^N(x(l))\right). \tag{2}$$

If M_1 is nested in M_2, then (1) is implicit in (2). A natural solution is given by

$$\pi_1^I(\theta_1) = \pi_1^N(\theta_1), \quad \pi_2^I(\theta_2) = \pi_2^N(\theta_2)\,\mathrm{E}_{\theta_2}^{M_2} B_{12}^N(x(l)).$$

Equivalently,

$$\pi_2^I(\theta_2|\theta) = \pi_2^N(\theta_2)\,\mathrm{E}_{\theta_2}^{M_2} \frac{f_1(x(l)|\theta_1)}{m_2^N(x(l))}.$$

2.4 Linear model

Casella and Moreno (2006) proposed a fully automatic Bayesian procedure for variable selection in normal regression models. The posterior probabilities are computed using intrinsic priors. Consider the standard normal regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\varepsilon},$$

where $\mathbf{y} = (y_1, \ldots, y_n)^\top$ is the vector of observations, $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_k]$ is the $n \times k$ design matrix, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)^\top$ is the $k \times 1$ column vector of the regression coefficients, and $\boldsymbol{\varepsilon}$ is an error vector distributed as $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(0, \sigma^2\mathbf{I}_n)$. This is the full model for $\mathbf{y}$ and is denoted by $\mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)$.

Let $\boldsymbol{\gamma}$ denote a vector of length k with components equal to either 0 or 1, and let $\mathbf{Q}_{\boldsymbol{\gamma}}$ denote a $k \times k$ diagonal matrix with the elements of $\boldsymbol{\gamma}$ on the leading diagonal and 0 elsewhere. Because we want to include the intercept in every model, the first component of each $\boldsymbol{\gamma}$ is equal to 1. We let γ denote the set of 2^{k-1} different configurations of $\boldsymbol{\gamma}$.

A submodel is written as $\mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2_{\boldsymbol{\gamma}}\mathbf{I}_n)$, where $\boldsymbol{\beta}_{\boldsymbol{\gamma}} = \mathbf{Q}_{\boldsymbol{\gamma}}\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ is a configuration to be interpreted as $\gamma_i = 0$ if $\alpha_i = 0$ and 1 otherwise.

We have the Bayesian model

$$M_{\boldsymbol{\gamma}} : \{\mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2_{\boldsymbol{\gamma}}\mathbf{I}_n), \pi(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}})\}, \quad \boldsymbol{\gamma} \in \Gamma.$$

They used the full model method. The Bayes factor of a generic model $M_{\boldsymbol{\gamma}}$, when compared with the full model $M_{\mathbf{1}}$, is given by the ratio of marginal distributions

$$B_{\boldsymbol{\gamma}\mathbf{1}}(\mathbf{y}, \mathbf{X}) = \frac{m_{\boldsymbol{\gamma}}(\mathbf{y}, \mathbf{X})}{m_{\mathbf{1}}(\mathbf{y}, \mathbf{X})} = \frac{\int \mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2_{\boldsymbol{\gamma}}\mathbf{I}_n)\pi(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}})d\boldsymbol{\beta}_{\boldsymbol{\gamma}}d\sigma_{\boldsymbol{\gamma}}}{\int \mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)\pi(\boldsymbol{\alpha}, \sigma)d\boldsymbol{\alpha}d\sigma}.$$

Casella and Moreno (2006) considered the standard default prior on parameter $(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}})$, giving the Bayesian model

$$M_{\boldsymbol{\gamma}} : \{\mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2_{\boldsymbol{\gamma}}\mathbf{I}_n), \pi^N(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}) = c_{\boldsymbol{\gamma}}/\sigma^2_{\boldsymbol{\gamma}}\}, \quad \boldsymbol{\gamma} \in \Gamma.$$

We first take an arbitrary but fixed point $(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}})$ in the null space, and then find the intrinsic prior for $(\boldsymbol{\alpha}, \sigma)$ conditional on $(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}})$. To do this, we note that a theoretical minimal training sample for this problem is a random vector $\mathbf{y}^{ts}$ of dimension $k+1$ such that it is $\mathcal{N}_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2_{\boldsymbol{\gamma}}\mathbf{I}_n)$ distributed under the null model and is $\mathcal{N}_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)$ distributed under the full model. Here $\mathbf{Z}^{ts}$ represents a $(k+1) \times k$ unknown design matrix associated with $\mathbf{y}^{ts}$.

Therefore,

$$\pi^I(\boldsymbol{\alpha}, \sigma|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\sigma}_{\boldsymbol{\gamma}}) = \pi^N(\boldsymbol{\alpha}, \sigma) \times \mathrm{E}_{\mathbf{y}^{ts}|\boldsymbol{\alpha},\sigma} \frac{N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2_{\boldsymbol{\gamma}}\mathbf{I}_n)}{\int N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)\pi^N(\boldsymbol{\alpha}, \sigma)d\boldsymbol{\alpha}d\sigma}.$$

Note that

$$\int N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)\pi^N(\boldsymbol{\alpha}, \sigma)d\boldsymbol{\alpha}d\sigma = \int \left(\int N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)d\boldsymbol{\alpha}\right)\frac{c}{\sigma^2}d\sigma.$$

We have

$$\begin{aligned}
&\int N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)d\boldsymbol{\alpha} \\
=&\frac{1}{(2\pi)^{(k+1)/2}\sigma^{k+1}}\exp\left\{-\frac{1}{2\sigma^2}\left\|\left(I - \mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top}\right)\mathbf{y}^{ts}\right\|^2\right\}\cdot \\
&\int \exp\left\{-\frac{1}{2\sigma^2}\left\|\mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top}\mathbf{y}^{ts} - \mathbf{Z}^{ts}\boldsymbol{\alpha}\right\|^2\right\}d\boldsymbol{\alpha} \\
=&\frac{1}{(2\pi)^{(k+1)/2}\sigma^{k+1}}\exp\left\{-\frac{1}{2\sigma^2}\left\|\left(I - \mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top}\right)\mathbf{y}^{ts}\right\|^2\right\}\cdot(2\pi)^{k/2}\sigma^k\left|\mathbf{Z}^{ts\top}\mathbf{Z}^{ts}\right|^{-1/2} \\
=&\frac{1}{(2\pi)^{1/2}\sigma\left|\mathbf{Z}^{ts\top}\mathbf{Z}^{ts}\right|^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}\left\|\left(I - \mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top}\right)\mathbf{y}^{ts}\right\|^2\right\}.
\end{aligned}$$

Thus,

$$
\begin{aligned}
&\int N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\alpha},\sigma^2\mathbf{I}_n)\pi^N(\boldsymbol{\alpha},\sigma)d\boldsymbol{\alpha}d\sigma\\
=&\int \frac{c}{(2\pi)^{1/2}\sigma^3\left|\mathbf{Z}^{ts\top}\mathbf{Z}^{ts}\right|^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}\left\|\left(I-\mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top}\right)\mathbf{y}^{ts}\right\|^2\right\}d\sigma\\
(\phi:=\sigma^{-2})=&\int \frac{c\phi^{3/2}}{(2\pi)^{1/2}\left|\mathbf{Z}^{ts\top}\mathbf{Z}^{ts}\right|^{1/2}}\exp\left\{-\frac{\phi}{2}\left\|\left(I-\mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top}\right)\mathbf{y}^{ts}\right\|^2\right\}\frac{1}{2}\phi^{-3/2}d\phi\\
=&\frac{c}{2(2\pi)^{1/2}\left|\mathbf{Z}^{ts\top}\mathbf{Z}^{ts}\right|^{1/2}}\int \exp\left\{-\frac{\phi}{2}\left\|\left(I-\mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top}\right)\mathbf{y}^{ts}\right\|^2\right\}d\phi\\
=&\frac{c}{(2\pi)^{1/2}\left|\mathbf{Z}^{ts\top}\mathbf{Z}^{ts}\right|^{1/2}\left\|\left(I-\mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top}\right)\mathbf{y}^{ts}\right\|^2}.
\end{aligned}
$$

With the above expression, we have

$$
\begin{aligned}
&\mathrm{E}_{\mathbf{y}^{ts}|\boldsymbol{\alpha},\sigma}\frac{N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\beta}_{\boldsymbol{\gamma}},\sigma_{\boldsymbol{\gamma}}^2\mathbf{I}_n)}{\int N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\alpha},\sigma^2\mathbf{I}_n)\pi^N(\boldsymbol{\alpha},\sigma)d\boldsymbol{\alpha}d\sigma}\\
=&c^{-1}(2\pi)^{1/2}\left|\mathbf{Z}^{ts\top}\mathbf{Z}^{ts}\right|^{1/2}\mathrm{E}_{\mathbf{y}^{ts}|\boldsymbol{\alpha},\sigma}\left\|\left(I-\mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top}\right)\mathbf{y}^{ts}\right\|^2 N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\beta}_{\boldsymbol{\gamma}},\sigma_{\boldsymbol{\gamma}}^2\mathbf{I}_n)\\
=&\frac{(2\pi)^{1/2}}{c}\left|\mathbf{Z}^{ts\top}\mathbf{Z}^{ts}\right|^{1/2}\int\left\|\left(I-\mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top}\right)\mathbf{y}^{ts}\right\|^2 N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\beta}_{\boldsymbol{\gamma}},\sigma_{\boldsymbol{\gamma}}^2\mathbf{I}_n)N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\alpha},\sigma^2\mathbf{I}_n)d\mathbf{y}^{ts}.
\end{aligned}
$$

To compute this integral, we use the following lemma.

Lemma 1.

$$
\int_{\mathbb{R}^n}\left(\mathbf{y}^\top\mathbf{K}\mathbf{y}\prod_{i=1}^{2}\mathcal{N}_n(\mathbf{y}|\mathbf{X}\theta_i,\sigma_i^2\mathbf{I}_n)\right)d\mathbf{y}=\frac{\sigma_2^2\,\mathrm{tr}(\mathbf{K})|\mathbf{X}^\top\mathbf{X}|^{-1/2}}{(2\pi\sigma_1^2)^{(n-k)/2}(1+\sigma_2^2/\sigma_1^2)^{(n-k+2)/2}}\mathcal{N}_k(\theta_2|\theta_1,(\sigma_1^2+\sigma_2^2)(\mathbf{X}^\top\mathbf{X})^{-1}),
$$

where $\mathbf{K}$ *is an* $n\times n$ *symmetric matrix,* $\mathbf{X}$ *is an* $n\times k$ *matrix of rank* k *such that* $\mathbf{K}\mathbf{X}=0$.

Proof.

$$
\begin{aligned}
&\int_{\mathbb{R}^n} \left(\mathbf{y}^\top \mathbf{K} \mathbf{y} \prod_{i=1}^{2} \mathcal{N}_n(\mathbf{y}|\mathbf{X}\theta_i, \sigma_i^2 \mathbf{I}_n) \right) d\mathbf{y} \\
=&\frac{1}{(2\pi)^n \sigma_1^n \sigma_2^n} \int_{\mathbb{R}^n} \mathbf{y}^\top \mathbf{K} \mathbf{y} \exp\left\{ -\frac{1}{2\sigma_1^2} \|\mathbf{y} - \mathbf{X}\theta_1\|^2 - \frac{1}{2\sigma_2^2} \|\mathbf{y} - \mathbf{X}\theta_2\|^2 \right\} d\mathbf{y} \\
=&\frac{1}{(2\pi)^n \sigma_1^n \sigma_2^n} \exp\left\{ -\frac{1}{2\sigma_1^2} \theta_1^\top \mathbf{X}^\top \mathbf{X} \theta_1 - \frac{1}{2\sigma_2^2} \theta_2^\top \mathbf{X}^\top \mathbf{X} \theta_2 \right\} \cdot \\
&\int_{\mathbb{R}^n} \mathbf{y}^\top \mathbf{K} \mathbf{y} \exp\left\{ -\frac{1}{2} \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 \sigma_2^2} \left(\|\mathbf{y}\|^2 - 2\mathbf{y}^\top \mathbf{X} \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \theta_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \theta_2 \right) \right) \right\} d\mathbf{y} \\
=&\frac{1}{(2\pi)^n \sigma_1^n \sigma_2^n} \exp\left\{ -\frac{1}{2\sigma_1^2} \theta_1^\top \mathbf{X}^\top \mathbf{X} \theta_1 - \frac{1}{2\sigma_2^2} \theta_2^\top \mathbf{X}^\top \mathbf{X} \theta_2 \right\} \cdot \\
&\int_{\mathbb{R}^n} \mathbf{y}^\top \mathbf{K} \mathbf{y} \exp\left\{ -\frac{1}{2} \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 \sigma_2^2} \left(\left\| \mathbf{y} - \mathbf{X} \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \theta_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \theta_2 \right) \right\|^2 - \left\| \mathbf{X} \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \theta_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \theta_2 \right) \right\|^2 \right) \right\} d\mathbf{y} \\
=&\frac{1}{(2\pi)^n \sigma_1^n \sigma_2^n} \exp\left\{ -\frac{1}{2\sigma_1^2} \theta_1^\top \mathbf{X}^\top \mathbf{X} \theta_1 - \frac{1}{2\sigma_2^2} \theta_2^\top \mathbf{X}^\top \mathbf{X} \theta_2 + \frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1^2 \sigma_2^2} \left\| \mathbf{X} \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \theta_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \theta_2 \right) \right\|^2 \right\} \cdot \\
&\int_{\mathbb{R}^n} \mathbf{y}^\top \mathbf{K} \mathbf{y} \exp\left\{ -\frac{1}{2} \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 \sigma_2^2} \left\| \mathbf{y} - \mathbf{X} \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \theta_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \theta_2 \right) \right\|^2 \right\} d\mathbf{y} \\
=&\frac{1}{(2\pi)^n \sigma_1^n \sigma_2^n} \exp\left\{ -\frac{1}{2\sigma_1^2} \theta_1^\top \mathbf{X}^\top \mathbf{X} \theta_1 - \frac{1}{2\sigma_2^2} \theta_2^\top \mathbf{X}^\top \mathbf{X} \theta_2 + \frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1^2 \sigma_2^2} \left\| \mathbf{X} \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \theta_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \theta_2 \right) \right\|^2 \right\} \cdot \\
&(2\pi)^{n/2} \left(\frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \right)^{n/2} \int_{\mathbb{R}^n} \mathbf{y}^\top \mathbf{K} \mathbf{y} \mathcal{N}_n \left(\mathbf{X} \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \theta_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \theta_2 \right), \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \mathbf{I}_n \right) d\mathbf{y} \\
=&\frac{1}{(2\pi)^n \sigma_1^n \sigma_2^n} \exp\left\{ -\frac{1}{2\sigma_1^2} \theta_1^\top \mathbf{X}^\top \mathbf{X} \theta_1 - \frac{1}{2\sigma_2^2} \theta_2^\top \mathbf{X}^\top \mathbf{X} \theta_2 + \frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1^2 \sigma_2^2} \left\| \mathbf{X} \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \theta_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \theta_2 \right) \right\|^2 \right\} \cdot \\
&(2\pi)^{n/2} \left(\frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \right)^{n/2+1} \operatorname{tr}(\mathbf{K}) \\
=&\frac{\sigma_1^2 \sigma_2^2 \operatorname{tr}(\mathbf{K})}{(2\pi)^{n/2} (\sigma_1^2 + \sigma_2^2)^{n/2+1}} \exp\left\{ -\frac{1}{2\sigma_1^2} \theta_1^\top \mathbf{X}^\top \mathbf{X} \theta_1 - \frac{1}{2\sigma_2^2} \theta_2^\top \mathbf{X}^\top \mathbf{X} \theta_2 + \frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1^2 \sigma_2^2} \left\| \mathbf{X} \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \theta_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \theta_2 \right) \right\|^2 \right\} \\
=&\frac{\sigma_1^2 \sigma_2^2 \operatorname{tr}(\mathbf{K})}{(2\pi)^{n/2} (\sigma_1^2 + \sigma_2^2)^{n/2+1}} \exp\left\{ -\frac{1}{2(\sigma_1^2 + \sigma_2^2)} (\theta_1 - \theta_2)^\top \mathbf{X}^\top \mathbf{X} (\theta_1 - \theta_2) \right\} \\
=&\frac{\sigma_1^2 \sigma_2^2 \operatorname{tr}(\mathbf{K})}{(2\pi)^{n/2} (\sigma_1^2 + \sigma_2^2)^{n/2+1}} \cdot (2\pi)^{k/2} (\sigma_1^2 + \sigma_2^2)^{k/2} |\mathbf{X}^\top \mathbf{X}|^{-1/2} \mathcal{N}_n \left(\theta_2 | \theta_1, (\sigma_1^2 + \sigma_2^2)(\mathbf{X}^\top \mathbf{X})^{-1} \right) \\
=&\frac{\sigma_1^2 \sigma_2^2 \operatorname{tr}(\mathbf{K}) |\mathbf{X}^\top \mathbf{X}|^{-1/2}}{(2\pi)^{(n-k)/2} (\sigma_1^2 + \sigma_2^2)^{(n-k)/2+1}} \mathcal{N}_n \left(\theta_2 | \theta_1, (\sigma_1^2 + \sigma_2^2)(\mathbf{X}^\top \mathbf{X})^{-1} \right).
\end{aligned}
$$

□

Using the above Lemma, we have

$$\int \left\| \left(I - \mathbf{Z}^{ts} (\mathbf{Z}^{ts\top} \mathbf{Z}^{ts})^{-1} \mathbf{Z}^{ts\top} \right) \mathbf{y}^{ts} \right\|^2 N_{k+1}(\mathbf{y}^{ts} | \mathbf{Z}^{ts} \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}^2 \mathbf{I}_n) N_{k+1}(\mathbf{y}^{ts} | \mathbf{Z}^{ts} \boldsymbol{\alpha}, \sigma^2 \mathbf{I}_n) d\mathbf{y}^{ts}$$
$$= \frac{\sigma^2 |\mathbf{Z}^{ts\top} \mathbf{Z}^{ts}|^{-1/2}}{(2\pi\sigma_{\boldsymbol{\gamma}}^2)^{1/2} (1 + \sigma^2/\sigma_{\boldsymbol{\gamma}}^2)^{3/2}} \mathcal{N}_k(\boldsymbol{\alpha} | \boldsymbol{\beta}_{\boldsymbol{\gamma}}, (\sigma^2 + \sigma_{\boldsymbol{\gamma}}^2)(\mathbf{Z}^{ts\top} \mathbf{Z}^{ts})^{-1}).$$

It follows that

$$\mathrm{E}_{\mathbf{y}^{ts}|\boldsymbol{\alpha},\sigma} \frac{N_{k+1}(\mathbf{y}^{ts} | \mathbf{Z}^{ts} \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}^2 \mathbf{I}_n)}{\int N_{k+1}(\mathbf{y}^{ts} | \mathbf{Z}^{ts} \boldsymbol{\alpha}, \sigma^2 \mathbf{I}_n) \pi^N(\boldsymbol{\alpha}, \sigma) d\boldsymbol{\alpha} d\sigma}$$
$$= \frac{(2\pi)^{1/2}}{c} \left| \mathbf{Z}^{ts\top} \mathbf{Z}^{ts} \right|^{1/2} \int \left\| \left(I - \mathbf{Z}^{ts} (\mathbf{Z}^{ts\top} \mathbf{Z}^{ts})^{-1} \mathbf{Z}^{ts\top} \right) \mathbf{y}^{ts} \right\|^2 N_{k+1}(\mathbf{y}^{ts} | \mathbf{Z}^{ts} \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}^2 \mathbf{I}_n) N_{k+1}(\mathbf{y}^{ts} | \mathbf{Z}^{ts} \boldsymbol{\alpha}, \sigma^2 \mathbf{I}_n) d\mathbf{y}^{ts}$$
$$= \frac{\sigma^2}{c \sigma_{\boldsymbol{\gamma}} (1 + \sigma^2/\sigma_{\boldsymbol{\gamma}}^2)^{3/2}} \mathcal{N}_k(\boldsymbol{\alpha} | \boldsymbol{\beta}_{\boldsymbol{\gamma}}, (\sigma^2 + \sigma_{\boldsymbol{\gamma}}^2)(\mathbf{Z}^{ts\top} \mathbf{Z}^{ts})^{-1}).$$

Thus,

$$\pi^I(\boldsymbol{\alpha}, \sigma | \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\sigma}_{\boldsymbol{\gamma}})$$
$$= \pi^N(\boldsymbol{\alpha}, \sigma) \times \mathrm{E}_{\mathbf{y}^{ts}|\boldsymbol{\alpha},\sigma} \frac{N_{k+1}(\mathbf{y}^{ts} | \mathbf{Z}^{ts} \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}^2 \mathbf{I}_n)}{\int N_{k+1}(\mathbf{y}^{ts} | \mathbf{Z}^{ts} \boldsymbol{\alpha}, \sigma^2 \mathbf{I}_n) \pi^N(\boldsymbol{\alpha}, \sigma) d\boldsymbol{\alpha} d\sigma}$$
$$= \frac{1}{\sigma_{\boldsymbol{\gamma}} (1 + \sigma^2/\sigma_{\boldsymbol{\gamma}}^2)^{3/2}} \mathcal{N}_k(\boldsymbol{\alpha} | \boldsymbol{\beta}_{\boldsymbol{\gamma}}, (\sigma^2 + \sigma_{\boldsymbol{\gamma}}^2)(\mathbf{Z}^{ts\top} \mathbf{Z}^{ts})^{-1}).$$

Proposition 1. *The conditional intrinsic prior is*

$$\pi^I(\boldsymbol{\alpha}, \sigma | \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\sigma}_{\boldsymbol{\gamma}}) = \frac{1}{\sigma_{\boldsymbol{\gamma}} (1 + \sigma^2/\sigma_{\boldsymbol{\gamma}}^2)^{3/2}} \mathcal{N}_k(\boldsymbol{\alpha} | \boldsymbol{\beta}_{\boldsymbol{\gamma}}, (\sigma^2 + \sigma_{\boldsymbol{\gamma}}^2) \mathbf{W}^{-1}),$$

where $\mathbf{W} = \mathbf{Z}^{ts\top} \mathbf{Z}^{ts}$.

A way of assesing $\mathbf{W}^{-1}$ is to use the original idea of the arithmetic intrinsic Bayes factor. This entails averaging over all possible training samples of minimal size contained in the sample. This would give the matrix

$$\mathbf{W}^{-1} = \frac{1}{L} \sum_{l=1}^{L} (\mathbf{Z}^\top(l) \mathbf{Z}(l))^{-1},$$

where $\{\mathbf{Z}(l), l = 1, \ldots, L\}$ is the set of all submatrices of $\mathbf{X}$ of order $(k+1) \times k$ of rank k.

For the data $(\mathbf{y}, \mathbf{X})$, the Bayes factor for comparing models $M_{\boldsymbol{\gamma}}$ and $M_{\mathbf{1}}$ with the intrinsic priors $\{\pi^N(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\beta}}), \pi^I(\boldsymbol{\alpha}, \sigma)\}$ has the formal expression

$$B_{\boldsymbol{\gamma}\mathbf{1}}(\mathbf{y}, \mathbf{X}) = \frac{\int \mathcal{N}_n(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}^2 \mathbf{I}_n) \pi^N(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}) d\boldsymbol{\beta}_{\boldsymbol{\gamma}} d\sigma_{\boldsymbol{\gamma}}}{\int \mathcal{N}_n(\mathbf{y} | \mathbf{X} \boldsymbol{\alpha}, \sigma^2 \mathbf{I}_n) \pi^I(\boldsymbol{\alpha}, \sigma | \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}) \pi^N(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}) d\boldsymbol{\alpha} d\sigma d\boldsymbol{\beta}_{\boldsymbol{\gamma}} d\sigma_{\boldsymbol{\gamma}}}.$$

In what follows we partition the design matrix $\mathbf{X}$ as $\mathbf{X} = (\mathbf{X}_{0\boldsymbol{\gamma}} | \mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}})$, where $\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}$ contains the column j of $\mathbf{X}$ if the configuration $\boldsymbol{\gamma}$ is such that $\gamma_j = 1$. Therefore, the dimension of $\mathbf{X}_{1\boldsymbol{\gamma}}$ is $n \times k_{\boldsymbol{\gamma}}$, where $k_{\boldsymbol{\gamma}} = \sum_{i=1}^{k} \gamma_i$.

Proposition 2. *The Bayes factor is given by*

$$B_{\boldsymbol{\gamma}\mathbf{1}}(\mathbf{y},\mathbf{X}) = \left(|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top \mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}|^{1/2}(\mathbf{y}^\top(\mathbf{I}_n - \mathbf{H}_{\boldsymbol{\gamma}})\mathbf{y})^{(n-k_{\boldsymbol{\gamma}}+1)/2} I_{\boldsymbol{\gamma}}\right)^{-1},$$

where $\mathbf{H}_{\boldsymbol{\gamma}} = \mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}(\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top \mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}})\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top$,

$$\begin{aligned}
I_{\boldsymbol{\gamma}} &= \int_0^{\pi/2} \frac{d\varphi}{|\mathbf{A}_{\boldsymbol{\gamma}}(\varphi)|^{1/2}|\mathbf{B}(\varphi)|^{1/2} E_{\boldsymbol{\gamma}}(\varphi)^{(n-k_{\boldsymbol{\gamma}}+1)/2}},\\
\mathbf{B}(\varphi) &= (\sin^2\varphi)\mathbf{I}_n + \mathbf{X}\mathbf{W}^{-1}\mathbf{X}^\top,\\
\mathbf{A}_{\boldsymbol{\gamma}}(\varphi) &= \mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top \mathbf{B}^{-1}(\varphi)\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}},\\
E_{\boldsymbol{\gamma}}(\varphi) &= \mathbf{y}^\top\left(\mathbf{B}^{-1}(\varphi) - \mathbf{B}^{-1}(\varphi)\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}\mathbf{A}_{\boldsymbol{\gamma}}^{-1}(\varphi)\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top \mathbf{B}^{-1}(\varphi)\right)\mathbf{y}.
\end{aligned}$$

Proof. For the numerator, we have

$$\begin{aligned}
&\int \mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}^2\mathbf{I}_n)\pi^N(\boldsymbol{\beta}_{\boldsymbol{\gamma}},\sigma_{\boldsymbol{\gamma}})d\boldsymbol{\beta}_{\boldsymbol{\gamma}}d\sigma_{\boldsymbol{\gamma}}\\
=&\int \frac{1}{(2\pi)^{n/2}\sigma_{\boldsymbol{\gamma}}^n}\exp\left\{-\frac{1}{2\sigma_{\boldsymbol{\gamma}}^2}\|\mathbf{y}-\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}\|^2\right\}\frac{c_{\boldsymbol{\gamma}}}{\sigma_{\boldsymbol{\gamma}}^2}d\boldsymbol{\beta}_{1\boldsymbol{\gamma}}d\sigma_{\boldsymbol{\gamma}}\\
=&\int \frac{1}{(2\pi)^{n/2}\sigma_{\boldsymbol{\gamma}}^n}\exp\left\{-\frac{1}{2\sigma_{\boldsymbol{\gamma}}^2}\|(\mathbf{I}_n-\mathbf{H}_{\boldsymbol{\gamma}})\mathbf{y}\|^2\right\}\exp\left\{-\frac{1}{2\sigma_{\boldsymbol{\gamma}}^2}\|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}(\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}-\hat{\boldsymbol{\beta}}_{\mathbf{1}\boldsymbol{\gamma}})\|^2\right\}\frac{c_{\boldsymbol{\gamma}}}{\sigma_{\boldsymbol{\gamma}}^2}d\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}d\sigma_{\boldsymbol{\gamma}}\\
=&\int \frac{1}{(2\pi)^{n/2}\sigma_{\boldsymbol{\gamma}}^n}(2\pi)^{k_{\boldsymbol{\gamma}}/2}\sigma_{\boldsymbol{\gamma}}^{k_{\boldsymbol{\gamma}}}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}|^{-1/2}\exp\left\{-\frac{1}{2\sigma_{\boldsymbol{\gamma}}^2}\|(\mathbf{I}_n-\mathbf{H}_{\boldsymbol{\gamma}})\mathbf{y}\|^2\right\}\frac{c_{\boldsymbol{\gamma}}}{\sigma_{\boldsymbol{\gamma}}^2}d\sigma_{\boldsymbol{\gamma}}\\
=&\int \frac{\phi^{(n-k_{\boldsymbol{\gamma}})/2}}{(2\pi)^{(n-k_{\boldsymbol{\gamma}})/2}}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}|^{-1/2}\exp\left\{-\frac{\phi}{2}\|(\mathbf{I}_n-\mathbf{H}_{\boldsymbol{\gamma}})\mathbf{y}\|^2\right\}c_{\boldsymbol{\gamma}}\phi\left(\frac{1}{2\phi^{3/2}}\right)d\phi\\
=&\int \frac{c_{\boldsymbol{\gamma}}}{2}\frac{1}{(2\pi)^{(n-k_{\boldsymbol{\gamma}})/2}}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}|^{-1/2}\phi^{(n-k_{\boldsymbol{\gamma}}+1)/2-1}\exp\left\{-\frac{\phi}{2}\|(\mathbf{I}_n-\mathbf{H}_{\boldsymbol{\gamma}})\mathbf{y}\|^2\right\}d\phi\\
=&\frac{c_{\boldsymbol{\gamma}}}{2}\frac{1}{(2\pi)^{(n-k_{\boldsymbol{\gamma}})/2}}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}|^{-1/2}\Gamma((n-k_{\boldsymbol{\gamma}}+1)/2)\left(\frac{2}{\|(\mathbf{I}_n-\mathbf{H}_{\boldsymbol{\gamma}})\mathbf{y}\|^2}\right)^{(n-k_{\boldsymbol{\gamma}}+1)/2}\\
=&\frac{c_{\boldsymbol{\gamma}}}{\sqrt{2}}\frac{1}{(\pi)^{(n-k_{\boldsymbol{\gamma}})/2}}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}|^{-1/2}\Gamma((n-k_{\boldsymbol{\gamma}}+1)/2)\|(\mathbf{I}_n-\mathbf{H}_{\boldsymbol{\gamma}})\mathbf{y}\|^{-(n-k_{\boldsymbol{\gamma}}+1)}.
\end{aligned}$$

Now we deal with the denominator

$$\int \mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\alpha},\sigma^2\mathbf{I}_n)\pi^I(\boldsymbol{\alpha},\sigma|\boldsymbol{\beta}_{\boldsymbol{\gamma}},\sigma_{\boldsymbol{\gamma}})\pi^N(\boldsymbol{\beta}_{\boldsymbol{\gamma}},\sigma_{\boldsymbol{\gamma}})d\boldsymbol{\alpha}d\sigma d\boldsymbol{\beta}_{\boldsymbol{\gamma}}d\sigma_{\boldsymbol{\gamma}}.$$

We have

$$\begin{aligned}
&\int \mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\alpha},\sigma^2\mathbf{I}_n)\pi^I(\boldsymbol{\alpha},\sigma|\boldsymbol{\beta}_{\boldsymbol{\gamma}},\sigma_{\boldsymbol{\gamma}})d\boldsymbol{\alpha}d\sigma\\
=&\int \mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\alpha},\sigma^2\mathbf{I}_n)\mathcal{N}_k(\boldsymbol{\alpha}|\boldsymbol{\beta}_{\boldsymbol{\gamma}},(\sigma^2+\sigma_{\boldsymbol{\gamma}}^2)\mathbf{W}^{-1})\frac{1}{\sigma_{\boldsymbol{\gamma}}(1+\sigma^2/\sigma_{\boldsymbol{\gamma}}^2)^{3/2}}d\boldsymbol{\alpha}d\sigma\\
=&\int \mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}_{\boldsymbol{\gamma}},(\sigma^2+\sigma_{\boldsymbol{\gamma}}^2)\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^\top+\sigma^2\mathbf{I}_n)\frac{1}{\sigma_{\boldsymbol{\gamma}}(1+\sigma^2/\sigma_{\boldsymbol{\gamma}}^2)^{3/2}}d\sigma.
\end{aligned}$$

Thus,

$$
\begin{aligned}
&\int \mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)\pi^I(\boldsymbol{\alpha}, \sigma|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}})\pi^N(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}})d\boldsymbol{\alpha}d\sigma d\boldsymbol{\beta}_{\boldsymbol{\gamma}}d\sigma_{\boldsymbol{\gamma}}\\
=&\int \mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}_{\boldsymbol{\gamma}}, (\sigma^2+\sigma_{\boldsymbol{\gamma}}^2)\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^\top + \sigma^2\mathbf{I}_n)\frac{1}{\sigma_{\boldsymbol{\gamma}}(1+\sigma^2/\sigma_{\boldsymbol{\gamma}}^2)^{3/2}}\pi^N(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}})d\sigma d\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}d\sigma_{\boldsymbol{\gamma}}\\
=&\int \left(\int \mathcal{N}_n\left(\mathbf{y}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}, (\sigma^2+\sigma_{\boldsymbol{\gamma}}^2)\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^\top + \sigma^2\mathbf{I}_n\right)d\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}\right)\frac{1}{\sigma_{\boldsymbol{\gamma}}(1+\sigma^2/\sigma_{\boldsymbol{\gamma}}^2)^{3/2}}\frac{c_{\boldsymbol{\gamma}}}{\sigma_{\boldsymbol{\gamma}}^2}d\sigma d\sigma_{\boldsymbol{\gamma}}\\
=&\int \left(\int \mathcal{N}_n\left(\mathbf{y}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}, (\sigma^2+\sigma_{\boldsymbol{\gamma}}^2)\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^\top + \sigma^2\mathbf{I}_n\right)d\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}\right)\frac{c_{\boldsymbol{\gamma}}}{(\sigma_{\boldsymbol{\gamma}}^2+\sigma^2)^{3/2}}d\sigma d\sigma_{\boldsymbol{\gamma}}.
\end{aligned}
$$

Let $\tilde{\boldsymbol{\Sigma}} = (\sigma^2+\sigma_{\boldsymbol{\gamma}}^2)\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^\top + \sigma^2\mathbf{I}_n$. Then

$$
\begin{aligned}
&\int \mathcal{N}_n\left(\mathbf{y}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}, \tilde{\boldsymbol{\Sigma}}\right)d\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}\\
=&\frac{1}{(2\pi)^{n/2}|\tilde{\boldsymbol{\Sigma}}|^{1/2}}\int \exp\left\{-\frac{1}{2}(\mathbf{y}-\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}})^\top\tilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}-\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}})\right\}d\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}\\
=&\frac{1}{(2\pi)^{n/2}|\tilde{\boldsymbol{\Sigma}}|^{1/2}}\exp\left\{-\frac{1}{2}\mathbf{y}^\top\left(\tilde{\boldsymbol{\Sigma}}^{-1}-\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}(\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}})^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\right)\mathbf{y}\right\}\cdot\\
&\int \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}-(\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}})^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{y}\right)^\top\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}\left(\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}-(\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}})^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{y}\right)\right\}d\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}\\
=&\frac{1}{(2\pi)^{n/2}|\tilde{\boldsymbol{\Sigma}}|^{1/2}}\exp\left\{-\frac{1}{2}\mathbf{y}^\top\left(\tilde{\boldsymbol{\Sigma}}^{-1}-\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}(\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}})^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\right)\mathbf{y}\right\}\cdot\\
&(2\pi)^{k_{\boldsymbol{\gamma}}/2}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}|^{-1/2}\\
=&\frac{1}{(2\pi)^{(n-k_{\boldsymbol{\gamma}})/2}|\tilde{\boldsymbol{\Sigma}}|^{1/2}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}|^{1/2}}\exp\left\{-\frac{1}{2}\mathbf{y}^\top\left(\tilde{\boldsymbol{\Sigma}}^{-1}-\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}(\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}})^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\right)\mathbf{y}\right\}.
\end{aligned}
$$

Let $\sigma_{\boldsymbol{\gamma}} = \rho\cos\varphi$, $\sigma = \rho\sin\varphi$, where $\rho\in(0,+\infty)$, $\varphi\in(0,\pi/2)$. Then

$$
\begin{aligned}
\tilde{\boldsymbol{\Sigma}} &= \rho^2(\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^\top + \sin^2\varphi\mathbf{I}_n) = \rho^2\mathbf{B}(\varphi),\\
\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}} &= \rho^{-2}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\mathbf{B}(\varphi)^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}} = \rho^{-2}\mathbf{A}_{\boldsymbol{\gamma}}(\varphi),\\
\mathbf{y}^\top\left(\tilde{\boldsymbol{\Sigma}}^{-1}-\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}(\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}})^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\right)\mathbf{y} &= \rho^{-2}E_{\boldsymbol{\gamma}}(\varphi).
\end{aligned}
$$

Then

$$
\int \mathcal{N}_n\left(\mathbf{y}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}, \tilde{\boldsymbol{\Sigma}}\right)d\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}} = \frac{1}{(2\pi)^{(n-k_{\boldsymbol{\gamma}})/2}\rho^{n-k_{\boldsymbol{\gamma}}}|\mathbf{B}(\varphi)|^{1/2}|\mathbf{A}_{\boldsymbol{\gamma}}(\varphi)|^{1/2}}\exp\left\{-\frac{1}{2\rho^2}E_{\boldsymbol{\gamma}}(\varphi)\right\}.
$$

It follows that

$$
\begin{aligned}
&\int \mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)\pi^I(\boldsymbol{\alpha}, \sigma|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}})\pi^N(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}})d\boldsymbol{\alpha}d\sigma d\boldsymbol{\beta}_{\boldsymbol{\gamma}}d\sigma_{\boldsymbol{\gamma}}\\
=&\int \left(\int \mathcal{N}_n\left(\mathbf{y}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}, (\sigma^2+\sigma^2_{\boldsymbol{\gamma}})\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^\top + \sigma^2\mathbf{I}_n\right) d\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}\right) \frac{c_{\boldsymbol{\gamma}}}{(\sigma^2_{\boldsymbol{\gamma}}+\sigma^2)^{3/2}} d\sigma d\sigma_{\boldsymbol{\gamma}}\\
=&\int \frac{1}{(2\pi)^{(n-k_{\boldsymbol{\gamma}})/2}\rho^{n-k_{\boldsymbol{\gamma}}}|\mathbf{B}(\varphi)|^{1/2}|\mathbf{A}_{\boldsymbol{\gamma}}(\varphi)|^{1/2}} \exp\left\{-\frac{1}{2\rho^2}E_{\boldsymbol{\gamma}}(\varphi)\right\} \frac{c_{\boldsymbol{\gamma}}}{\rho^2} d\rho d\varphi\\
=&\int \frac{c_{\boldsymbol{\gamma}}}{(2\pi)^{(n-k_{\boldsymbol{\gamma}})/2}\rho^{n-k_{\boldsymbol{\gamma}}+2}|\mathbf{B}(\varphi)|^{1/2}|\mathbf{A}_{\boldsymbol{\gamma}}(\varphi)|^{1/2}} \exp\left\{-\frac{1}{2\rho^2}E_{\boldsymbol{\gamma}}(\varphi)\right\} d\rho d\varphi\\
=&\int \frac{c_{\boldsymbol{\gamma}}\phi^{(n-k_{\boldsymbol{\gamma}}+2)/2}}{(2\pi)^{(n-k_{\boldsymbol{\gamma}})/2}|\mathbf{B}(\varphi)|^{1/2}|\mathbf{A}_{\boldsymbol{\gamma}}(\varphi)|^{1/2}} \exp\left\{-\frac{\phi}{2}E_{\boldsymbol{\gamma}}(\varphi)\right\} \frac{1}{2\phi^{3/2}} d\phi d\varphi\\
=&\int \frac{c_{\boldsymbol{\gamma}}}{2(2\pi)^{(n-k_{\boldsymbol{\gamma}})/2}|\mathbf{B}(\varphi)|^{1/2}|\mathbf{A}_{\boldsymbol{\gamma}}(\varphi)|^{1/2}} \phi^{(n-k_{\boldsymbol{\gamma}}+1)/2-1} \exp\left\{-\frac{\phi}{2}E_{\boldsymbol{\gamma}}(\varphi)\right\} d\phi d\varphi\\
=&\int \frac{c_{\boldsymbol{\gamma}}}{2(2\pi)^{(n-k_{\boldsymbol{\gamma}})/2}|\mathbf{B}(\varphi)|^{1/2}|\mathbf{A}_{\boldsymbol{\gamma}}(\varphi)|^{1/2}} \Gamma((n-k_{\boldsymbol{\gamma}}+1)/2) \left(\frac{2}{E_{\boldsymbol{\gamma}}(\varphi)}\right)^{(n-k_{\boldsymbol{\gamma}}+1)/2} d\varphi\\
=&\frac{c_{\boldsymbol{\gamma}}\Gamma((n-k_{\boldsymbol{\gamma}}+1)/2)}{\sqrt{2}(\pi)^{(n-k_{\boldsymbol{\gamma}})/2}} \int \frac{d\varphi}{|\mathbf{B}(\varphi)|^{1/2}|\mathbf{A}_{\boldsymbol{\gamma}}(\varphi)|^{1/2}E_{\boldsymbol{\gamma}}(\varphi)^{(n-k_{\boldsymbol{\gamma}}+1)/2}}\\
=&\frac{c_{\boldsymbol{\gamma}}\Gamma((n-k_{\boldsymbol{\gamma}}+1)/2)}{\sqrt{2}(\pi)^{(n-k_{\boldsymbol{\gamma}})/2}} I_{\boldsymbol{\gamma}}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
B_{\boldsymbol{\gamma}\mathbf{1}}(\mathbf{y}, \mathbf{X}) =& \frac{\frac{c_{\boldsymbol{\gamma}}}{\sqrt{2}}\frac{1}{(\pi)^{(n-k_{\boldsymbol{\gamma}})/2}}|\mathbf{X}^\top_{\mathbf{1}\boldsymbol{\gamma}}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}|^{-1/2}\Gamma((n-k_{\boldsymbol{\gamma}}+1)/2)\|(\mathbf{I}_n-\mathbf{H}_{\boldsymbol{\gamma}})\mathbf{y}\|^{-(n-k_{\boldsymbol{\gamma}}+1)}}{\frac{c_{\boldsymbol{\gamma}}\Gamma((n-k_{\boldsymbol{\gamma}}+1)/2)}{\sqrt{2}(\pi)^{(n-k_{\boldsymbol{\gamma}})/2}} I_{\boldsymbol{\gamma}}}\\
=& \left(|\mathbf{X}^\top_{\mathbf{1}\boldsymbol{\gamma}}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}|^{1/2}\|(\mathbf{I}_n-\mathbf{H}_{\boldsymbol{\gamma}})\mathbf{y}\|^{n-k_{\boldsymbol{\gamma}}+1} I_{\boldsymbol{\gamma}}\right)^{-1}.
\end{aligned}
$$

This completes the proof.

□

3 Mixture of g prior

This section is adapted from Liang et al. (2008).

4 Fractional Bayes factor

Fractional Bayes factor is proposed by O'Hagan (1995). Fractional intrinsic Bayes factor is proposed by De Santis and Spezzaferri (1997). See Santis and Spezzaferri (1999) for a review. Divergence-based (DB) priors are proposed by Bayarri and Garca-Donato (2008).

5 Expected-posterior priors

Expected-posterior prior is proposed by Perez (2002).

6 Normal-inverse-gamma (NIG) prior

Zhou and Guan (2018)

Consider the testing problem in linear regression with independent normal errors:

$$H_0 : \mathbf{Y}|\mathbf{a}, \tau \sim \mathcal{N}(\mathbf{W}\mathbf{a}, \tau^{-1}\mathbf{I}_n),$$
$$H_1 : \mathbf{Y}|\mathbf{a}, \mathbf{b}, \tau \sim \mathcal{N}(\mathbf{W}\mathbf{a} + \mathbf{L}\mathbf{b}, \tau^{-1}\mathbf{I}_n),$$

where $\mathbf{W}$ is a full-rank $n \times q$ matrix representing the nuisance covariates, including a column of $\mathbf{1}_n$. $\mathbf{L}$ is an $n \times p$ matrix representing the covariates of interest.

NIG prior:

$$\mathbf{a}|\tau \sim \mathcal{N}(0, \tau^{-1}\mathbf{V}_a),$$
$$\mathbf{b}|\tau \sim \mathcal{N}(0, \tau^{-1}\mathbf{V}_b),$$
$$\tau \sim \text{Gamma}(\kappa_1/2, 2/\kappa_2).$$

Here

$$\pi(\tau) = \frac{(\kappa_2/2)^{\kappa_1/2}}{\Gamma(\kappa_1/2)}\tau^{\kappa_1/2-1}\exp\left\{-\frac{\kappa_2\tau}{2}\right\}$$

Then

$$f(\mathbf{Y}|\mathbf{a}, \mathbf{b}, \tau)\pi(\mathbf{a}|\tau)\pi(\mathbf{b}|\tau)\pi(\tau)$$
$$=\frac{(\kappa_2/2)^{\kappa_1/2}\tau^{(n+p+q+\kappa_1)/2-1}}{(2\pi)^{(n+p+q)/2}|\mathbf{V}_a|^{1/2}|\mathbf{V}_b|^{1/2}\Gamma(\kappa_1/2)}\exp\left\{-\frac{\tau}{2}\left(\|\mathbf{Y}-\mathbf{W}\mathbf{a}-\mathbf{L}\mathbf{b}\|^2+\mathbf{a}^\top\mathbf{V}_a^{-1}\mathbf{a}+\mathbf{b}^\top\mathbf{V}_b^{-1}\mathbf{b}+\kappa_2\right)\right\}.$$

7 Nonnested linear models

Moreno and Girón (2007) said:

"There are two natural ways of encompassing: one way is to encompass all models into the model containing all possible regressors, and the other is to encompass the model containing only the intercept into any other. "

8 High-dimensional setting

Armagan et al. (2013) investigated the posterior consistency in linear models. Their focus is on shrinkage priors, including Laplace prior, Student's t, Generalized double Pareto, and hourseshoe-type priors. Bai and Ghosh (2018) investigated the posterior consistency under the global-local shrinkage priors.

8.1 Nonlocal priors

Nonlocal priors are proposed by Johnson and Rossell (2010) in the context of Bayesian hypothesis testing. Johnson and Rossell (2012) and Johnson (2013) considered using nonlocal priors to solve model selection problem. A more recent work is Shin et al. (2018).

Estimation: Rossell and Telesca (2017).

8.2 Intrinsic priors

The asymptotic behaviors of the Bayes factors with intrinsic priors in high dimensional setting have been investigated by Casella et al. (2009), Girón et al. (2010) and Moreno et al. (2010) and Moreno et al. (2015).

8.3 $p > n$ case

Laplace approximation Barber et al. (2016).

9 A fractional intrinsic Bayes factor for linear model in high-dimensional setting

Suppose we would like to compare models $\mathcal{M}_0$ and $\mathcal{M}_1$.

$$\begin{aligned}\mathcal{M}_0 : \mathbf{y} = \mathbf{X}_0\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}_n(0, \phi^{-1}\mathbf{I}_n),\\ \mathcal{M}_1 : \mathbf{y} = \mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}_n(0, \phi^{-1}\mathbf{I}_n).\end{aligned}$$

Here $\boldsymbol{\beta}_0$ is p_0 dimensional and $\boldsymbol{\beta}_1$ is p_1 dimensional. We assume that as n tends to infinity, p_0 is fixed while p_1 may diverge. This assumption is reasonable. In practice, p_0 is often 1 and $\mathbf{X}_0$ is $\mathbf{1}_n$.

There have been several extensions of g-priors to $p > n$ case: Maruyama and George (2011), Shang and Clayton (2011).

Under $\mathcal{M}_0$, we impose the reference prior $\pi_0(\boldsymbol{\beta}_0, \phi) = c_0/\phi$. Note that the posterior corresponding to the referece prior is proper only if $n > p_0 + p_1$? That is, the minimal training sample size is $p_0 + p_1 + 1$. So we cannot impose the reference prior under $\mathcal{M}_1$ provided $p_0 + p_1 + 1 > n$. We impose the conditional prior $\boldsymbol{\beta}_1|\boldsymbol{\beta}_0, \phi \sim \mathcal{N}_{p_1}(0, \kappa\phi^{-1}\mathbf{I}_{p_1})$. Following the heuristic device of Kass and Wasserman (1995), we choose κ such that the amount of information about the parameter equal to the amount of information contained in one observation. Kass and Wasserman (1995) used Fisher information to define "amount of information". In the $p_1 > n$ setting, if $\mathbf{X}_1$ is a fixed design, the Fisher information is not invertible which invalidate Kass and Wasserman (1995)'s method. To overcome this difficulty, we temporarily assume that the rows of $\mathbf{X}_1$ are iid $\mathcal{N}_p(0, c\mathbf{I}_{p_1})$ random vectors. Then the block of Fisher information matrix corresponding to $\boldsymbol{\beta}_1$ is $c\phi\mathbf{I}_{p_1}$. Then κ should satisfy

$$\kappa\phi^{-1}\mathbf{I}_{p_1} = (c\phi\mathbf{I}_{p_1})^{-1}.$$

That is, κ should equal c^{-1}. However, c is unknown. Note that $\|\mathbf{X}_1\|_F^2/c \sim \chi^2(np_1)$. An estimator of c is $\|\mathbf{X}_1\|_F^2/(np_1)$. So we put $\kappa = np_1/\|\mathbf{X}_1\|_F^2$. Thus, under $\mathcal{M}_1$, we put prior

$$\pi_1(\boldsymbol{\beta}_1|\boldsymbol{\beta}_0, \phi) = \frac{(\phi\|\mathbf{X}_1\|_F^2)^{p_1/2}}{(2\pi np_1)^{p_1/2}} \exp\left\{-\frac{\phi\|\mathbf{X}_1\|_F^2}{2np_1}\|\boldsymbol{\beta}_1\|^2\right\}, \quad \pi_1(\boldsymbol{\beta}_0, \phi) = \frac{c_1}{\phi}.$$

9.1 Fractional intrinsic priors

Suppose we are comparing two models based on n iid observations, $\mathcal{M}_i : f_i(x|\theta_i)$, $i = 0, 1$, where $f_0(x|\theta_0)$ is nested in $f_1(x|\theta_1)$. Suppose prior $\pi_i(\theta_i)$ is imposed under $\mathcal{M}_i$, $i = 0, 1$. Bayes factors suffers from some paradox. Several remedies have been proposed. Fractional Bayes factor (O'Hagan (1995)) is defined as

$$B_{10}^F = \frac{\int \prod_{i=1}^n f_1(x_i|\theta_1)\pi_1(\theta_1)d\theta_1}{\int \prod_{i=1}^n f_0(x|\theta_0)\pi_0(\theta_0)d\theta_0} \cdot \frac{\int (\prod_{i=1}^n f_0(x|\theta_0))^{m/n}\pi_0(\theta_0)d\theta_0}{\int (\prod_{i=1}^n f_1(x|\theta_1))^{m/n}\pi_1(\theta_1)d\theta_1},$$

where $1 \le m \le n$ is the training sample size. Although Fractional Bayes factor has good properties, it is not a real Bayes factor. Intrinsic fractional prior is proposed by De Santis and Spezzaferri (1997). The Bayes factor derived from intrinsic fractional prior is asymptotically equivalent to the fractional Bayes factor. We can take $\pi_0^I(\theta_0) = \pi_0(\theta_0)$ and $\pi_1^I(\theta_1)$ satisfies

$$B_{10}^{IF} := \frac{\int \prod_{i=1}^n f_1(x_i|\theta_1)\pi_1^I(\theta_1)d\theta_1}{\int \prod_{i=1}^n f_0(x|\theta_0)\pi_0^I(\theta_0)d\theta_0} \approx \frac{\int \prod_{i=1}^n f_1(x_i|\theta_1)\pi_1(\theta_1)d\theta_1}{\int \prod_{i=1}^n f_0(x|\theta_0)\pi_0(\theta_0)d\theta_0} \cdot \frac{\int (\prod_{i=1}^n f_0(x|\theta_0))^{m/n}\pi_0(\theta_0)d\theta_0}{\int (\prod_{i=1}^n f_1(x|\theta_1))^{m/n}\pi_1(\theta_1)d\theta_1}.$$

Suppose $f_1(x|\theta^*)$ is the true model which generates the data. Then

$$\frac{\int (\prod_{i=1}^n f_0(x|\theta_0))^{m/n}\pi_0(\theta_0)d\theta_0}{\int (\prod_{i=1}^n f_1(x|\theta_1))^{m/n}\pi_1(\theta_1)d\theta_1} = \frac{\int (\prod_{i=1}^n \frac{f_0(x|\theta_0)}{f_1(x|\theta^*)})^{m/n}\pi_0(\theta_0)d\theta_0}{\int (\prod_{i=1}^n \frac{f_1(x|\theta_1)}{f_1(x|\theta^*)})^{m/n}\pi_1(\theta_1)d\theta_1} \approx \frac{\int \exp\{-m\mathrm{KL}(f_1(x|\theta^*)||f_0(x|\theta_0))\}\,\pi_0(\theta_0)d\theta_0}{\int \exp\{-m\mathrm{KL}(f_1(x|\theta^*)||f_1(x|\theta_1))\}\,\pi_1(\theta_1)d\theta_1}.$$

Thus,

$$\begin{aligned}\frac{\int \prod_{i=1}^n f_1(x_i|\theta_1)\pi_1^I(\theta_1)d\theta_1}{\int \prod_{i=1}^n f_0(x|\theta_0)\pi_0^I(\theta_0)d\theta_0} &\approx \frac{\int \prod_{i=1}^n f_1(x_i|\theta_1)\pi_1(\theta_1)d\theta_1}{\int \prod_{i=1}^n f_0(x|\theta_0)\pi_0(\theta_0)d\theta_0} \cdot \frac{\int \exp\{-m\mathrm{KL}(f_1(x|\theta^*)||f_0(x|\theta_0))\}\,\pi_0(\theta_0)d\theta_0}{\int \exp\{-m\mathrm{KL}(f_1(x|\theta^*)||f_1(x|\theta_1))\}\,\pi_1(\theta_1)d\theta_1} \\ &\approx \frac{\int \prod_{i=1}^n f_1(x_i|\theta_1)\pi_1^I(\theta_1)d\theta_1}{\int \prod_{i=1}^n f_0(x|\theta_0)\pi_0^I(\theta_0)d\theta_0} \cdot \frac{\pi_1(\theta^*)}{\pi_1^I(\theta^*)} \cdot \frac{\int \exp\{-m\mathrm{KL}(f_1(x|\theta^*)||f_0(x|\theta_0))\}\,\pi_0(\theta_0)d\theta_0}{\int \exp\{-m\mathrm{KL}(f_1(x|\theta^*)||f_1(x|\theta_1))\}\,\pi_1(\theta_1)d\theta_1}.\end{aligned}$$

Thus,

$$\pi_1^I(\theta^*) = \pi_1(\theta^*) \cdot \frac{\int \exp\{-m\mathrm{KL}(f_1(x|\theta^*)||f_0(x|\theta_0))\}\,\pi_0(\theta_0)d\theta_0}{\int \exp\{-m\mathrm{KL}(f_1(x|\theta^*)||f_1(x|\theta_1))\}\,\pi_1(\theta_1)d\theta_1}.$$

In the linear model case, we have

$$\pi_0^I(\boldsymbol{\beta}_0, \phi) = c_0/\phi.$$

Note that

$$\begin{aligned}\mathrm{KL}(f_1(x|\theta^*)||f_0(x|\theta_0)) =&\mathrm{KL}\left(\mathcal{N}_n(\mathbf{X}_0\boldsymbol{\beta}_0^* + \mathbf{X}_1\boldsymbol{\beta}_1^*, \phi^{*-1}\mathbf{I}_n)||\mathcal{N}_n(\mathbf{X}_0\boldsymbol{\beta}_0, \phi^{-1}\mathbf{I}_n)\right) \\ =&\frac{1}{2}\left(n\frac{\phi}{\phi^*} + \phi\|\mathbf{X}_0(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0^*) - \mathbf{X}_1\boldsymbol{\beta}_1^*\|^2 - n - n\log\frac{\phi}{\phi^*}\right).\end{aligned}$$

$$\begin{aligned}\mathrm{KL}(f_1(x|\theta^*)||f_1(x|\theta_1)) =&\mathrm{KL}\left(\mathcal{N}_n(\mathbf{X}_0\boldsymbol{\beta}_0^* + \mathbf{X}_1\boldsymbol{\beta}_1^*, \phi^{*-1}\mathbf{I}_n)||\mathcal{N}_n(\mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{X}_1\boldsymbol{\beta}_1, \phi^{-1}\mathbf{I}_n)\right) \\ =&\frac{1}{2}\left(n\frac{\phi}{\phi^*} + \phi\|\mathbf{X}_0(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0^*) + \mathbf{X}_1(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^*)\|^2 - n - n\log\frac{\phi}{\phi^*}\right).\end{aligned}$$

Note that $m = p_0 + 1$.

We have

$$
\begin{aligned}
&\int \exp\left\{-m\mathrm{KL}(f_1(x|\theta^*)||f_0(x|\theta_0))\right\}\pi_0(\theta_0)d\theta_0 \\
=&\int \exp\left\{-\frac{m}{2}\left(n\frac{\phi}{\phi^*} + \phi\|\mathbf{X}_0(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0^*) - \mathbf{X}_1\boldsymbol{\beta}_1^*\|^2 - n - n\log\frac{\phi}{\phi^*}\right)\right\}\frac{c_0}{\phi}d\boldsymbol{\beta}_0 d\phi \\
=&\int \exp\left\{-\frac{mn\phi}{2\phi^*} - \frac{m\phi}{2}\|\mathbf{X}_0(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0^*) - \mathbf{X}_1\boldsymbol{\beta}_1^*\|^2 + \frac{mn}{2}\right\}\left(\frac{\phi}{\phi^*}\right)^{mn/2}\frac{c_0}{\phi}d\boldsymbol{\beta}_0 d\phi \\
=&c_0\exp\left\{\frac{mn}{2}\right\}\phi^{*-mn/2}\int \exp\left\{-\frac{mn\phi}{2\phi^*} - \frac{m\phi}{2}\|\mathbf{X}_0(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0^*) - \mathbf{X}_1\boldsymbol{\beta}_1^*\|^2\right\}\phi^{mn/2-1}d\boldsymbol{\beta}_0 d\phi \\
=&c_0\exp\left\{\frac{mn}{2}\right\}\phi^{*-mn/2}\int \exp\left\{-\frac{mn\phi}{2\phi^*} - \frac{m\phi}{2}\|\mathbf{X}_0(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0^* - (\mathbf{X}_0^\top\mathbf{X}_0)^{-1}\mathbf{X}_0^\top\mathbf{X}_1\boldsymbol{\beta}_1^*)\|^2 - \frac{m\phi}{2}\|(\mathbf{I}_n - \mathbf{H}_0)\mathbf{X}_1\boldsymbol{\beta}_1^*\|^2\right\}\phi \\
=&c_0\exp\left\{\frac{mn}{2}\right\}\phi^{*-mn/2}(2\pi)^{p_0/2}m^{-p_0/2}|\mathbf{X}_0^\top\mathbf{X}_0|^{-1/2}\int \exp\left\{-\frac{mn\phi}{2\phi^*} - \frac{m\phi}{2}\|(\mathbf{I}_n - \mathbf{H}_0)\mathbf{X}_1\boldsymbol{\beta}_1^*\|^2\right\}\phi^{(mn-p_0)/2-1}d\phi \\
=&c_0\exp\left\{\frac{mn}{2}\right\}\phi^{*-mn/2}(2\pi)^{p_0/2}m^{-p_0/2}|\mathbf{X}_0^\top\mathbf{X}_0|^{-1/2}\Gamma((mn-p_0)/2)\left(\frac{mn}{2\phi^*} + \frac{m}{2}\|(\mathbf{I}_n - \mathbf{H}_0)\mathbf{X}_1\boldsymbol{\beta}_1^*\|^2\right)^{-(mn-p_0)/2}
\end{aligned}
$$

On the other hand,

$$\int \exp\left\{-m\mathrm{KL}(f_1(x|\theta^*)||f_1(x|\theta_1))\right\}\pi_1(\theta_1)d\theta_1$$

$$=\int \exp\left\{-\frac{m}{2}\left(n\frac{\phi}{\phi^*}+\phi\|\mathbf{X}_0(\boldsymbol{\beta}_0-\boldsymbol{\beta}_0^*)+\mathbf{X}_1(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*)\|^2-n-n\log\frac{\phi}{\phi^*}\right)\right\}$$

$$\frac{(\phi\|\mathbf{X}_1\|_F^2)^{p_1/2}}{(2\pi np_1)^{p_1/2}}\exp\left\{-\frac{\phi\|\mathbf{X}_1\|_F^2}{2np_1}\|\boldsymbol{\beta}_1\|^2\right\}\frac{c_1}{\phi}d\boldsymbol{\beta}_1 d\boldsymbol{\beta}_0 d\phi$$

$$=c_1\exp\left\{\frac{mn}{2}\right\}\phi^{*-mn/2}\int\exp\left\{-\frac{mn\phi}{2\phi^*}-\frac{m\phi}{2}\|\mathbf{X}_0(\boldsymbol{\beta}_0-\boldsymbol{\beta}_0^*)+\mathbf{X}_1(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*)\|^2\right\}\phi^{mn/2-1}$$

$$\frac{(\phi\|\mathbf{X}_1\|_F^2)^{p_1/2}}{(2\pi np_1)^{p_1/2}}\exp\left\{-\frac{\phi\|\mathbf{X}_1\|_F^2}{2np_1}\|\boldsymbol{\beta}_1\|^2\right\}d\boldsymbol{\beta}_1 d\boldsymbol{\beta}_0 d\phi$$

$$=c_1\exp\left\{\frac{mn}{2}\right\}\phi^{*-mn/2}\frac{(\|\mathbf{X}_1\|_F^2)^{p_1/2}}{(2\pi np_1)^{p_1/2}}$$

$$\int\exp\left\{-\frac{mn\phi}{2\phi^*}-\frac{m\phi}{2}\|\mathbf{X}_0(\boldsymbol{\beta}_0-\boldsymbol{\beta}_0^*)+\mathbf{X}_1(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*)\|^2-\frac{\phi\|\mathbf{X}_1\|_F^2}{2np_1}\|\boldsymbol{\beta}_1\|^2\right\}\phi^{(mn+p_1)/2-1}d\boldsymbol{\beta}_1 d\boldsymbol{\beta}_0 d\phi$$

$$=c_1\exp\left\{\frac{mn}{2}\right\}\phi^{*-mn/2}\frac{(\|\mathbf{X}_1\|_F^2)^{p_1/2}}{(2\pi np_1)^{p_1/2}}$$

$$\int\exp\left\{-\frac{mn\phi}{2\phi^*}-\frac{m\phi}{2}\left\|\mathbf{X}_0\left(\boldsymbol{\beta}_0-\boldsymbol{\beta}_0^*+(\mathbf{X}_0^\top\mathbf{X}_0)^{-1}\mathbf{X}_0^\top\mathbf{X}_1(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*)\right)\right\|^2-\frac{m\phi}{2}\|(\mathbf{I}_n-\mathbf{H}_0)\mathbf{X}_1(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*)\|^2-\frac{\phi\|\mathbf{X}_1\|}{2np_1}$$

$$=c_1\exp\left\{\frac{mn}{2}\right\}\phi^{*-mn/2}\frac{(\|\mathbf{X}_1\|_F^2)^{p_1/2}}{(2\pi np_1)^{p_1/2}}(2\pi)^{p_0/2}m^{-p_0/2}|\mathbf{X}_0^\top\mathbf{X}_0|^{-1/2}$$

$$\int\exp\left\{-\frac{mn\phi}{2\phi^*}-\frac{m\phi}{2}\left(\|(\mathbf{I}_n-\mathbf{H}_0)\mathbf{X}_1(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*)\|^2+\frac{\|\mathbf{X}_1\|_F^2}{nmp_1}\|\boldsymbol{\beta}_1\|^2\right)\right\}\phi^{(mn+p_1-p_0)/2-1}d\boldsymbol{\beta}_1 d\phi$$

$$=c_1\exp\left\{\frac{mn}{2}\right\}\phi^{*-mn/2}\frac{(\|\mathbf{X}_1\|_F^2)^{p_1/2}}{(2\pi np_1)^{p_1/2}}(2\pi)^{p_0/2}m^{-p_0/2}|\mathbf{X}_0^\top\mathbf{X}_0|^{-1/2}$$

$$\int\exp\left\{-\frac{mn\phi}{2\phi^*}-\frac{m\phi}{2}\left((\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*)^\top\mathbf{X}_1^\top(\mathbf{I}_n-\mathbf{H}_0)\mathbf{X}_1(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*)+\frac{\|\mathbf{X}_1\|_F^2}{nmp_1}\boldsymbol{\beta}^\top\boldsymbol{\beta}_1\right)\right\}\phi^{(mn+p_1-p_0)/2-1}d\boldsymbol{\beta}_1 d\phi$$

$$=c_1\exp\left\{\frac{mn}{2}\right\}\phi^{*-mn/2}\frac{(\|\mathbf{X}_1\|_F^2)^{p_1/2}}{(2\pi np_1)^{p_1/2}}(2\pi)^{p_0/2}m^{-p_0/2}|\mathbf{X}_0^\top\mathbf{X}_0|^{-1/2}$$

$$\int\exp\left\{-\frac{mn\phi}{2\phi^*}\right\}\phi^{(mn+p_1-p_0)/2-1}\cdot(2\pi)^{p_1/2}(m\phi)^{-p_1/2}\left|\mathbf{X}_1^\top(\mathbf{I}_n-\mathbf{H}_0)\mathbf{X}_1+\frac{\|\mathbf{X}_1\|_F^2}{nmp_1}\mathbf{I}_{p_1}\right|^{-1/2}$$

$$\exp\left\{-\frac{m\phi}{2}\left(\boldsymbol{\beta}_1^{*\top}\mathbf{X}_1^\top(\mathbf{I}-\mathbf{H}_0)\mathbf{X}_1\boldsymbol{\beta}_1^*-\boldsymbol{\beta}_1^{*\top}\mathbf{X}_1^\top(\mathbf{I}-\mathbf{H}_0)\mathbf{X}_1(\mathbf{X}_1^\top(\mathbf{I}-\mathbf{H}_0)\mathbf{X}_1+\frac{\|\mathbf{X}_1\|_F^2}{nmp_1}\mathbf{I}_{p_1})^{-1}\mathbf{X}_1^\top(\mathbf{I}-\mathbf{H}_0)\mathbf{X}_1\boldsymbol{\beta}_1^*\right)\right\}d$$

$$=c_1\exp\left\{\frac{mn}{2}\right\}\phi^{*-mn/2}\frac{(\|\mathbf{X}_1\|_F^2)^{p_1/2}}{(2\pi np_1)^{p_1/2}}(2\pi)^{(p_0+p_1)/2}m^{-(p_0+p_1)/2}|\mathbf{X}_0^\top\mathbf{X}_0|^{-1/2}\left|\mathbf{X}_1^\top(\mathbf{I}_n-\mathbf{H}_0)\mathbf{X}_1+\frac{\|\mathbf{X}_1\|_F^2}{nmp_1}\mathbf{I}_{p_1}\right|^{-1/2}$$

$$\int\phi^{(mn-p_0)/2-1}\cdot$$

$$\exp\left\{-\frac{m\phi}{2}\left(\frac{n}{\phi^*}+\boldsymbol{\beta}_1^{*\top}\mathbf{X}_1^\top(\mathbf{I}-\mathbf{H}_0)\mathbf{X}_1\boldsymbol{\beta}_1^*-\boldsymbol{\beta}_1^{*\top}\mathbf{X}_1^\top(\mathbf{I}-\mathbf{H}_0)\mathbf{X}_1(\mathbf{X}_1^\top(\mathbf{I}-\mathbf{H}_0)\mathbf{X}_1+\frac{\|\mathbf{X}_1\|_F^2}{nmp_1}\mathbf{I}_{p_1})^{-1}\mathbf{X}_1^\top(\mathbf{I}-\mathbf{H}_0)\mathbf{X}_1\boldsymbol{\beta}\right.$$

$$=c_1\exp\left\{\frac{mn}{2}\right\}\phi^{*-mn/2}\frac{(\|\mathbf{X}_1\|_F^2)^{p_1/2}}{(2\pi np_1)^{p_1/2}}(2\pi)^{(p_0+p_1)/2}m^{-(p_0+p_1)/2}|\mathbf{X}_0^\top\mathbf{X}_0|^{-1/2}\left|\mathbf{X}_1^\top(\mathbf{I}_n-\mathbf{H}_0)\mathbf{X}_1+\frac{\|\mathbf{X}_1\|_F^2}{nmp_1}\mathbf{I}_{p_1}\right|^{-1/2}$$

$$\Gamma((mn-p_0)/2)\left[\frac{m}{2}\left(\frac{n}{\phi^*}+\boldsymbol{\beta}_1^{*\top}\mathbf{X}_1^\top(\mathbf{I}-\mathbf{H}_0)\mathbf{X}_1\boldsymbol{\beta}_1^*-\boldsymbol{\beta}_1^{*\top}\mathbf{X}_1^\top(\mathbf{I}-\mathbf{H}_0)\mathbf{X}_1(\mathbf{X}_1^\top(\mathbf{I}-\mathbf{H}_0)\mathbf{X}_1+\frac{\|\mathbf{X}_1\|_F^2}{nmp_1}\mathbf{I}_{p_1})^{-1}\mathbf{X}_1^\top(\mathbf{I}-\right.\right.$$

10 A Bayesian-motivated:

References

Armagan, A., Dunson, D. B., Lee, J., Bajwa, W. U., and Strawn, N. (2013). Posterior consistency in linear models under shrinkage priors. *Biometrika*, 100(4):1011–1018.

Bai, R. and Ghosh, M. (2018). High-dimensional multivariate posterior consistency under global-local shrinkage priors. *Journal of Multivariate Analysis*, pages –.

Barber, R. F., Drton, M., and Tan, K. M. (2016). Laplace approximation in high-dimensional bayesian regression. In Frigessi, A., Bühlmann, P., Glad, I. K., Langaas, M., Richardson, S., and Vannucci, M., editors, *Statistical Analysis for High-Dimensional Data*, pages 15–36. Springer International Publishing.

Bayarri, M. J., Berger, J. O., Forte, A., and Garca-Donato, G. (2012). Criteria for bayesian model choice with application to variable selection. *Ann. Statist.*, 40(3):1550–1577.

Bayarri, M. J. and Garca-Donato, G. (2008). Generalization of jeffreys divergence-based priors for bayesian hypothesis testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5):981–1003.

Berger, J. O. and Pericchi, L. R. (1996). The intrinsic bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 91(433):109–122.

Berger, J. O. and Pericchi, L. R. (2001). *Objective Bayesian Methods for Model Selection: Introduction and Comparison*, volume Volume 38 of *Lecture Notes–Monograph Series*, pages 135–207. Institute of Mathematical Statistics, Beachwood, OH.

Bernardo, J. M. (2005). Reference analysis. In Dey, D. and Rao, C., editors, *Bayesian Thinking*, volume 25 of *Handbook of Statistics*, pages 17 – 90. Elsevier.

Casella, G., Girón, F. J., Martínez, M. L., and Moreno, E. (2009). Consistency of bayesian procedures for variable selection. *The Annals of Statistics*, 37(3):1207–1228.

Casella, G. and Moreno, E. (2006). Objective bayesian variable selection. *Journal of the American Statistical Association*, 101(473):157–167.

De Santis, F. and Spezzaferri, F. (1997). Alternative bayes factors for model selection. *Canadian Journal of Statistics*, 25(4):503–515.

Girón, F. J., Moreno, E., Casella, G., and Martínez, M. L. (2010). Consistency of objective bayes factors for nonnested linear models and increasing model dimension. *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas*, 104(1):57–67.

Jeffreys, H. (2003). *Theory of probability.* Oxford University Press, reprint of 3rd ed. edition.

Johnson, V. E. (2013). On numerical aspects of bayesian model selection in high and ultrahigh-dimensional settings. *Bayesian Analysis*, 8(4):741–758.

Johnson, V. E. and Rossell, D. (2010). On the use of nonlocal prior densities in bayesian hypothesis tests. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(2):143–170.

Johnson, V. E. and Rossell, D. (2012). Bayesian model selection in high-dimensional settings. *Journal of the American Statistical Association*, 107(498):649–660.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

Kass, R. E. and Wasserman, L. (1995). A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. *Journal of the American Statistical Association*, 90(431):928–934.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g priors for bayesian variable selection. *Journal of the American Statistical Association*, 103(481):410–423.

Maruyama, Y. and George, E. I. (2011). Fully bayes factors with a generalized g -prior. *Ann. Statist.*, 39(5):2740–2765.

Moreno, E., Bertolino, F., and Racugno, W. (1998). An intrinsic limiting procedure for model selection and hypotheses testing. *Journal of the American Statistical Association*, 93(444):1451–1460.

Moreno, E. and Girón, F. J. (2007). Comparison of bayesian objective procedures for variable selection in linear regression. *TEST*, 17(3):472–490.

Moreno, E., Girn, F. J., and Casella, G. (2010). Consistency of objective bayes factors as the model dimension grows. *Ann. Statist.*, 38(4):1937–1952.

Moreno, E., Girn, J., and Casella, G. (2015). Posterior model consistency in variable selection as the model dimension grows. *Statist. Sci.*, 30(2):228–241.

Mukhopadhyay, M. and Samanta, T. (2016). A mixture of g-priors for variable selection when the number of regressors grows with the sample size. *TEST*, 26(2):377–404.

Mukhopadhyay, M., Samanta, T., and Chakrabarti, A. (2014). On consistency and optimality of bayesian variable selection based on \$\$g\$\$ g -prior in normal linear regression models. *Annals of the Institute of Statistical Mathematics*, 67(5):963–997.

O'Hagan, A. (1995). Fractional bayes factors for model comparison. 57:99–138.

Perez, J. M. (2002). Expected-posterior prior distributions for model selection. *Biometrika*, 89(3):491–512.

Rossell, D. and Telesca, D. (2017). Nonlocal priors for high-dimensional estimation. *Journal of the American Statistical Association*, 112(517):254–265.

Santis, F. and Spezzaferri, F. (1999). Methods for default and robust bayesian model comparison: the fractional bayes factor approach. *International Statistical Review*, 67(3):267–286.

Shang, Z. and Clayton, M. K. (2011). Consistency of bayesian linear model selection with a growing number of parameters. *Journal of Statistical Planning and Inference*, 141(11):3463–3474.

Shin, M., Bhattacharya, A., and Johnson, V. E. (2018). Scalable bayesian variable selection using nonlocal prior densities in ultrahigh-dimensional settings. *Statistica Sinica*, 28:1053–1078.

Wang, M. (2017). Mixtures of g-priors for analysis of variance models with a diverging number of parameters. *Bayesian Anal.*, 12(2):511–532.

Wang, M. and Maruyama, Y. (2017). Posterior consistency of g-prior for variable selection with a growing number of parameters. *Journal of Statistical Planning and Inference*.

Wang, M. and Sun, X. (2014). Bayes factor consistency for nested linear models with a growing number of parameters. *Journal of Statistical Planning and Inference*, 147:95 – 105.

Wang, T., Berthet, Q., and Samworth, R. J. (2016). Statistical and computational trade-offs in estimation of sparse principal components. *Ann. Statist.*, 44(5):1896–1930.

Xiang, R., Ghosh, M., and Khare, K. (2016). Consistency of bayes factors under hyper g-priors with growing model size. *Journal of Statistical Planning and Inference*, 173:64 – 86.

Zellner, A. and Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. *Trabajos de Estadistica Y de Investigacion Operativa*, 31(1):585–603.

Zhou, Q. and Guan, Y. (2018). On the null distribution of bayes factors in linear regression. *Journal of the American Statistical Association*, 113(523):1362–1371.