
BANFF
CANADA


AMERICAN STATISTICAL ASSOCIATION
AWARDS THE HONORARY RANK OF
Fellow
TO
Luis Raul Pericchi Guerra
University of Puerto Rico, Rio Piedras Campus, San Juan

Reflections on Robust and Intrinsic Priors

Jim Berger

Duke University

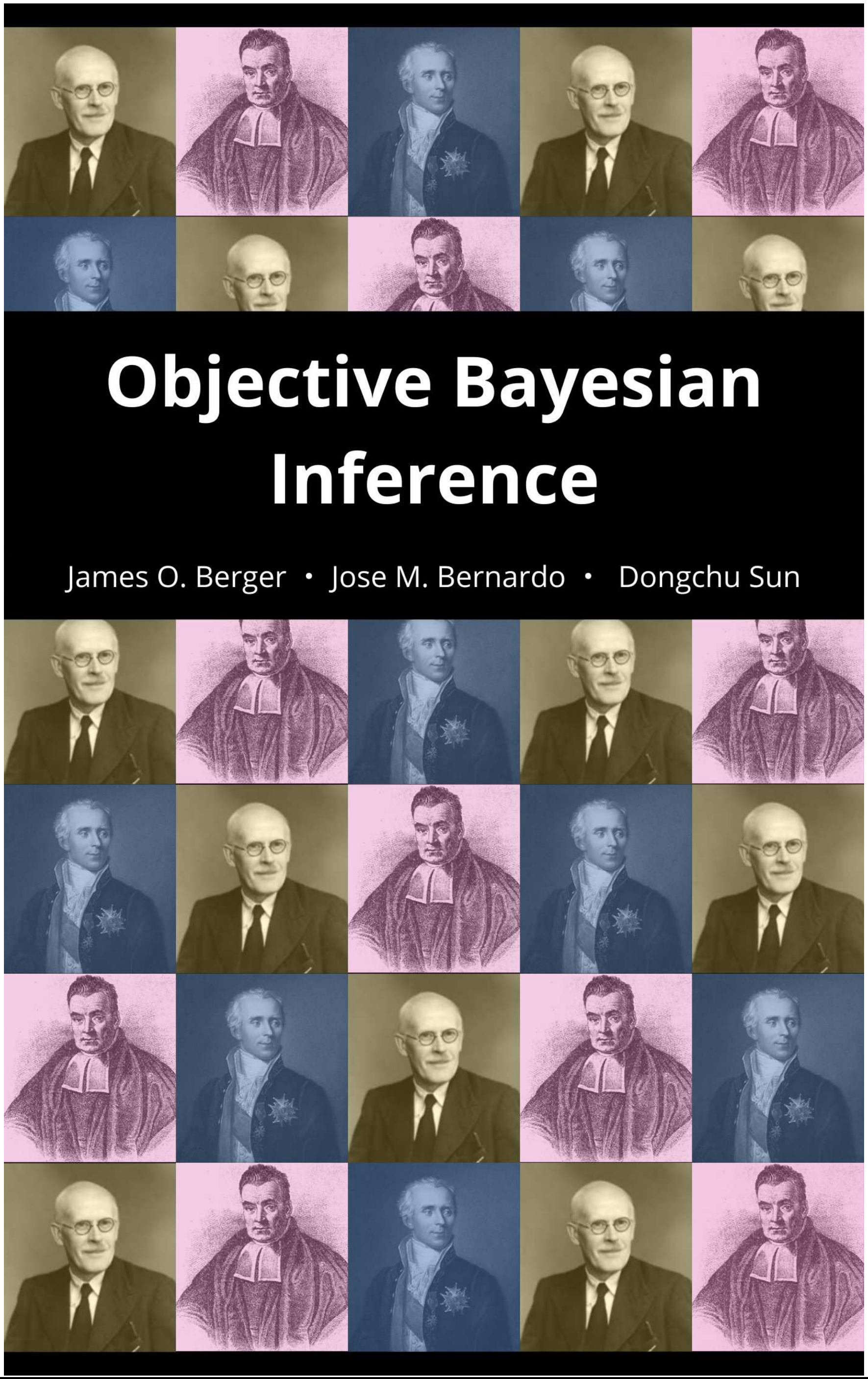
from the in-progress book *Bayesian Testing and Model Uncertainty*
with **Elías Moreno, Gonzalo García-Donato**, and **Luis Pericchi**

OBayes 2022, UC Santa Cruz

Dedicated to Luis Pericchi's 70^{th} *Birthday*

September 8, 2022


A Bridge between Cross-validation Bayes Factors and Geometric Intrinsic Bayes Factors
Yekun Wang, Luis Pericchi
Department of Mathematics, University of Puerto Rico
Introductions
Normal Means Problem
LOO-CVBF with GIBF and CVBF
Expectations
Simulations
Bridge Rule
Butterfly


Objective Bayesian Inference
James O. Berger • Jose M. Bernardo • Dongchu Sun

The intrinsic prior Bayes factor for testing a precise hypothesis

Consider the test of $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}$ is a k-dimensional unknown parameter, based on i.i.d. data $\boldsymbol{x} = (x_1, \ldots, x_n)$ having density $f(\boldsymbol{x} \mid \boldsymbol{\theta})$. Suppose $\pi^O(\boldsymbol{\theta})$ is a standard objective estimation prior (it could be just constant). Then the intrinsic prior Bayes factor for the test is

$$
\begin{aligned}
B_{01} &= \frac{f(\boldsymbol{x} \mid \boldsymbol{\theta}_0)}{\int [\int f(\boldsymbol{x} \mid \boldsymbol{\theta}) \pi^O(\boldsymbol{\theta} \mid \boldsymbol{x}^*) \, d\boldsymbol{\theta}] f(\boldsymbol{x}^* \mid \boldsymbol{\theta}_0) \, d\boldsymbol{x}^*} \\
&= \frac{f(\boldsymbol{x} \mid \boldsymbol{\theta}_0)}{\int m^O(\boldsymbol{x} \mid \boldsymbol{x}^*) f(\boldsymbol{x}^* \mid \boldsymbol{\theta}_0) \, d\boldsymbol{x}^*} \,,
\end{aligned}
$$

when $\boldsymbol{x}^* = (x_1^*, \ldots, x_k^*)$ is an imaginary minimal training sample (more formally, one wants that $m^O(\boldsymbol{x}^*) = \int f(\boldsymbol{x}^* \mid \boldsymbol{\theta}) \pi^O(\boldsymbol{\theta}) \, d\boldsymbol{\theta} < \infty$ but $m^O(\boldsymbol{x}^*)$ be infinite for a smaller sample), $\pi^O(\boldsymbol{\theta} \mid \boldsymbol{x}^*) = f(\boldsymbol{x}^* \mid \boldsymbol{\theta}) \pi^O(\boldsymbol{\theta}) / m^O(\boldsymbol{x}^*)$, and $m^O(\boldsymbol{x} \mid \boldsymbol{x}^*) = m^O(\boldsymbol{x}, \boldsymbol{x}^*) / m^O(\boldsymbol{x}^*)$.

This is the Bayes factor arising from the intrinsic prior

$$\pi^I(\boldsymbol{\theta}) = \int \pi^O(\boldsymbol{\theta} \mid \boldsymbol{x}^*) f(\boldsymbol{x}^* \mid \boldsymbol{\theta}_0) \, d\boldsymbol{x}^* \,,$$

using the *expected posterior prior* formulation of an intrinsic prior.

Example (Higgs Boson): Test $H_0 : \theta = 0$ versus $H_0 : \theta > 0$, based on i.i.d. $x_i \sim f(x_i \mid \theta) = (\theta + b) \exp\{-(\theta + b) x_i\}$, $i = 1, \ldots, n$, where θ is the mass of the Higgs boson and b is a known background mean rate.

- The usual objective estimation prior for θ would be $\pi^O(\theta) = 1/(\theta + b)$.
- A minimal sample size for the resulting posterior to be proper is $k = 1$.
- Computation then yields $\pi^I(\theta) = \int \pi^O(\theta \mid x_1^*) f(x_1^* \mid 0) dx_1^* = b/(\theta + b)^2$ and

$$B_{01} = \frac{b^n \exp\{-bn\bar{x}\}}{\int_0^\infty (\theta + b)^n \exp\{-(\theta + b)n\bar{x}\} b(\theta + b)^{-2} \, d\theta} = \left[(n-2)! \sum_{i=2}^{n} \frac{(bn\bar{x})^{1-i}}{(n-i)!} \right]^{-1} .$$

Variable Selection in the Normal Linear Model

- The *full model:* observe independent $y_1, y_2, \ldots, y_n$, where

$$y_i = [x_{0,i1}\beta_{01} + \cdots + x_{0,ik_0}\beta_{0k_0}] + x_{i1}\beta_1 + \cdots + x_{ip}\beta_p + \epsilon_i \, ,$$

  – the $x_{0,ij}$ and x_{ij} being given covariates;

  – the β's being unknown;

  – the ϵ_i being independent $N(\epsilon \mid 0, \sigma^2)$ errors, with σ^2 unknown.

- Defining $\boldsymbol{y} = (y_1, \ldots, y_n)'$, $\boldsymbol{X}_0$ as the $n \times k_0$ matrix with elements $x_{0,ij}$, $\boldsymbol{X}$ as the $n \times p$ matrix with elements x_{ij}, $\boldsymbol{\beta}_0 = (\beta_{01}, \ldots, \beta_{0k_0})'$, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$, this model can be written

$$M_F : \boldsymbol{y} \sim N_n(\boldsymbol{y} \mid \boldsymbol{X}_0\boldsymbol{\beta}_0 + \boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}) \, .$$

- The simplest model is assumed to be

$$M_0 : \boldsymbol{y} \sim N_n(\boldsymbol{y} \mid \boldsymbol{X}_0\boldsymbol{\beta}_0, \sigma^2\boldsymbol{I})\,;$$

with $\boldsymbol{X}_0$ consisting of the covariates that are to be included in all models (e.g., the intercept in ordinary linear regression).

- Between M_0 and M_F are $2^p - 2$ other models M_i, each additionally including a non-null subset of k_i of the remaining p covariates:

$$M_i : \boldsymbol{y} \sim N_n(\boldsymbol{y} \mid \boldsymbol{X}_0\boldsymbol{\beta}_0 + \boldsymbol{X}_i\boldsymbol{\beta}_i, \sigma^2\boldsymbol{I})\,;$$

$\boldsymbol{X}_i$ is the $n \times k_i$ matrix consisting of the chosen covariates, i.e., the chosen columns of $\boldsymbol{X}$; the corresponding vector of unknown parameters is denoted $\boldsymbol{\beta}_i$.

- $(\boldsymbol{\beta}_0, \sigma^2)$ are the *common parameters* in all models,
- $\pi_i(\boldsymbol{\beta}_0, \boldsymbol{\beta}_i, \sigma^2)$ is the prior distribution of the parameters in M_i,
- $Pr(M_i)$ is the prior probability of model M_i. (We use the recommendation of Jeffreys in examples.)

Bayes model selection

- Is based on posterior probabilities for each model:

$$Pr(M_i \mid \boldsymbol{y}) = \frac{m_i(\boldsymbol{y}) Pr(M_i)}{\sum_{j=1}^{N} m_j(\boldsymbol{y}) Pr(M_j)},$$

  where

$$m_j(\boldsymbol{y}) = \int N_n(\boldsymbol{y} \mid \boldsymbol{X}_0 \boldsymbol{\beta}_0 + \boldsymbol{X}_j \boldsymbol{\beta}_j, \sigma^2 \boldsymbol{I})\, \pi_j(\boldsymbol{\beta}_0, \boldsymbol{\beta}_j, \sigma^2)\, d\boldsymbol{\beta}_0\, d\boldsymbol{\beta}_j\, d\sigma^2$$

  is the marginal likelihood of M_j, quantifying how likely the observed data is under that model.

- $B_{ji} = \dfrac{m_j(\boldsymbol{y})}{m_i(\boldsymbol{y})}$ is the Bayes factor of M_j to M_i.

- $p_i = \sum_{\{M_k \text{ that contain } x_i\}} Pr(M_k \mid \boldsymbol{y})$ is the posterior inclusion probability of the covariate x_i.

The 'Robust' conventional priors for variable selection

(Bayarri, Berger, Forte and Garcia-Donato, 2012):

Defining $\boldsymbol{\Sigma}_i = \sigma^2\,(\boldsymbol{V}_i^{'}\boldsymbol{V}_i)^{-1}$, where $\boldsymbol{V}_i = (\boldsymbol{I}_n - \boldsymbol{X}_0(\boldsymbol{X}_0^{'}\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0^{'})\boldsymbol{X}_i$,

$$\begin{aligned}
\pi_0^R(\boldsymbol{\beta}_0, \sigma^2) &= \frac{1}{\sigma^2}, \\
\pi_i^R(\boldsymbol{\beta}_0, \boldsymbol{\beta}_i, \sigma^2) &= \frac{1}{\sigma^2} \int_0^1 N_{k_i}\left(\boldsymbol{\beta}_i \,\middle|\, \mathbf{0}, \left(\frac{1+n}{\lambda(k_0+k_i)} - 1\right)\boldsymbol{\Sigma}_i\right) \frac{1}{2\sqrt{\lambda}}\,d\lambda\,.
\end{aligned}$$

"Although this prior is not closed form, it gives closed form marginal likelihoods, and closed form Bayes factors

$$B_{i0} = \left[\frac{n+1}{k_i+k_0}\right]^{-\frac{k_i}{2}} \frac{Q_{i0}^{-(n-k_0)/2}}{k_i+1}\, {}_2F_1\left[\frac{k_i+1}{2}; \frac{n-k_0}{2}; \frac{k_i+3}{2}; \frac{(1-Q_{i0}^{-1})(k_i+k_0)}{(1+n)}\right],$$

where ${}_2F_1$ is the standard (Gauss) hypergeometric function and

$$Q_{i0} = SSE_i/SSE_0$$

is the ratio of the sum of squared errors of models M_i and M_0."

New expression:

$$B_{i0} = \frac{C_{n,k_i}}{2}\, Q_{i0}^{-(n-k_0-k_i-1)/2}(1-Q_{i0})^{-(k_i+1)/2} Beta\left(\frac{k_i+1}{2}, \frac{n-k_0-k_i-1}{2}\right)$$
$$\times F_{Beta(\frac{k_i+1}{2},\frac{n-k_0-k_i-1}{2})}\left(\left(1+C_{n,k_i}^2\frac{Q_{i0}}{1-Q_{i0}}\right)^{-1}\right),$$

where $C_{n,k_i} = \frac{\sqrt{1+n}}{\sqrt{k_0+k_i}}$, $Beta(\cdot,\cdot)$ is the Beta function and $F_{Beta(a,b)}(\cdot)$ is the CDF of the $Beta(a,b)$ distribution.

Example: *Bayesian t-test.* Here $k_0 = 0$ and $k_1 = 1$. It is easy to see that $Q_{10} = (1+\frac{t^2}{n-1})^{-1}$, where t is the usual t-statistic. Computation yields that

$$B_{10} = \frac{\sqrt{n+1}}{(n-2)}\frac{(n-1)}{t^2}\left(1+\frac{t^2}{n-1}\right)^{n/2}\left(1-\left[1+\frac{t^2}{n^2-1}\right]^{-(n-2)/2}\right).$$

As $t \to 0$, $B_{10} \to 1/[2\sqrt{n+1}]$. For $n=2$, this is to be interpreted as

$$B_{10} = \frac{\sqrt{3}}{2}\cdot\left(1+\frac{1}{t^2}\right)\cdot\log\left(1+\frac{t^2}{3}\right).$$

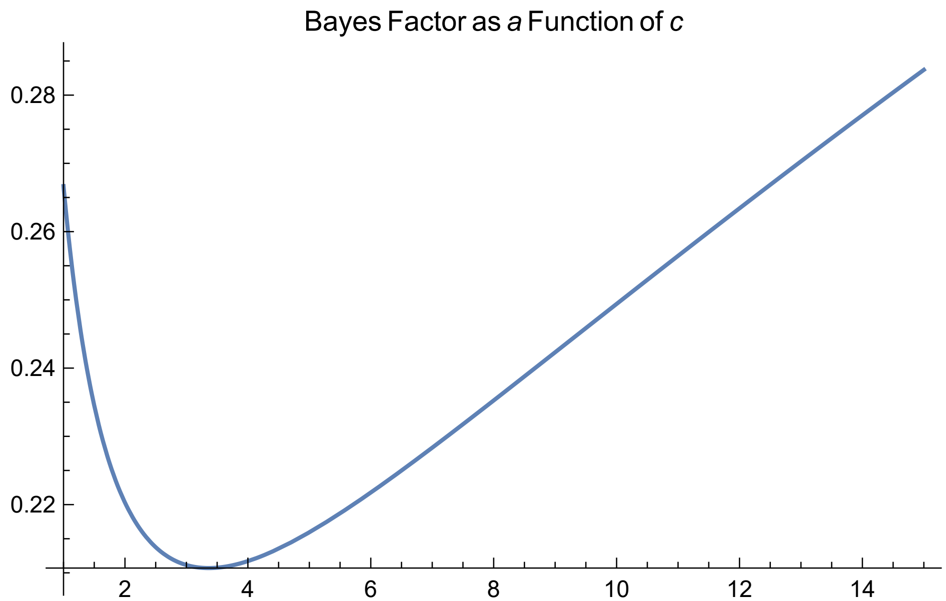
To study robustness with respect to the scale of the prior, consider the prior

$$\pi_c^R(\beta,\sigma^2) = \frac{1}{\sigma^2}\int_0^1 N\left(\beta \mid 0, \left[\frac{c}{\lambda}-1\right]\frac{\sigma^2}{n}\right)\frac{1}{2\sqrt{\lambda}}d\lambda\,,$$

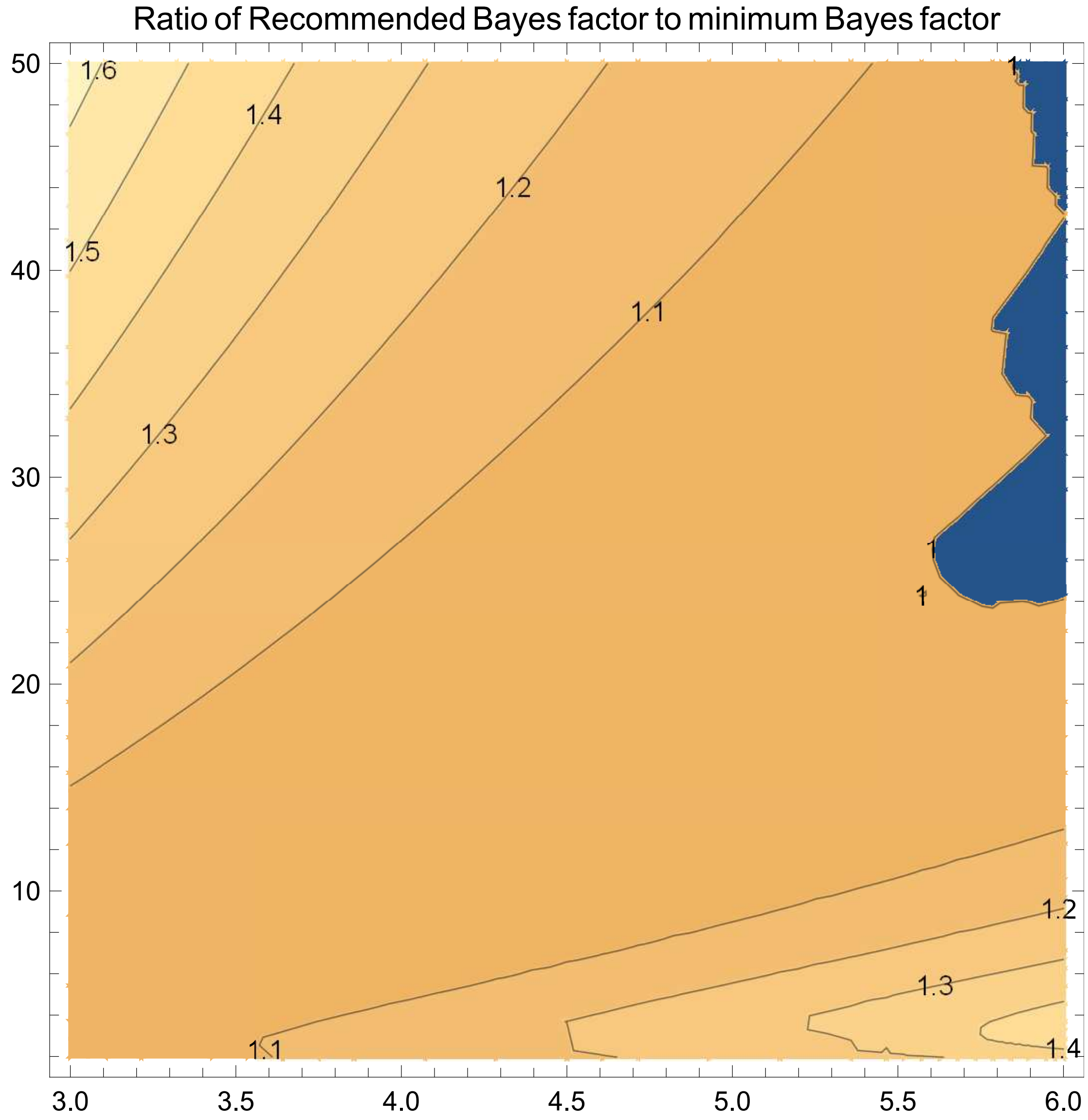
with resulting Bayes factor (of M_0 to M_1)

$$B_{01}(c) = \frac{1}{\sqrt{c}}\left(\frac{n-2}{n-1}\right)t^2\left(1+\frac{t^2}{n-1}\right)^{-\frac{n}{2}}\left[1-\left(1+\frac{t^2}{c(n-1)}\right)^{-(\frac{n}{2}-1)}\right]^{-1}.$$

If $n=10$ and $t=3$ (p-value=0.015), $B_{01}(c)$ as a function of $c \geq 1$ is



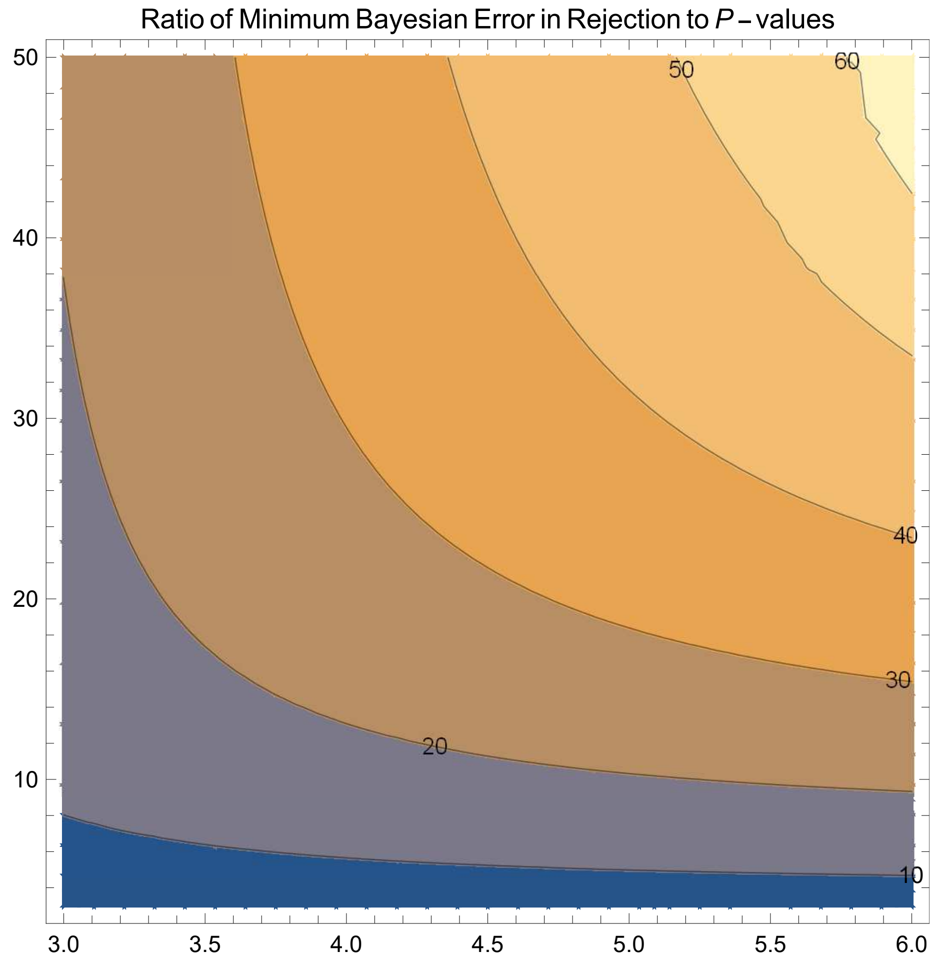
Contour plot of the ratio of the recommended Bayes factor to the minimum Bayes factor for $2 \leq n \leq 50$ and $3 \leq t \leq 6$.



Comparison with p-values

For a given n and t, compute the p-value $p(n,t)$ and consider the contours of

$$R(n,t) = \frac{\inf_c [B_{01}(c)/(1+B_{01}(c)]}{p(n,t)}\,, \qquad \text{for } 2 \leq n \leq 50 \text{ and } 3 \leq t \leq 6\,.$$



Full inferences for the robust prior:

- Model specific posterior means and variances are available in closed form.
- Predictive means and variances (through model averaging) are available in closed form.
- Given λ, all means and variances and easy to sample (just normal distribution inferences). One can obtain *independent* samples from the posterior distribution of λ as follows:
  - Draw u from Uniform $(0, F_{Beta(\frac{k_i+1}{2}, \frac{n-k_0-k_i-1}{2})}((1 + C^2_{n,k_i} \frac{Q_{i0}}{1-Q_{i0}})^{-1}))$,
  - Set $v = F^{-1}_{Beta(\frac{k_i+1}{2}, \frac{n-k_0-k_i-1}{2})}(u)$,
  - Let $\lambda = C^2_{n,k_i} \frac{Q_{i0} v}{(1-Q_{i0})(1-v)}$.

The 'Intrinsic' conventional priors for variable selection

The intrinsic prior for the parameters in model M_i, when M_0 is the *base model* and minimal training sample size $m_i = k_0 + k_i + 1$ is used, is given by

$$\begin{aligned}\pi_0^I(\boldsymbol{\beta}_0, \sigma^2) &= \frac{1}{\sigma^2}, \\ \pi_i^I(\boldsymbol{\beta}_0, \boldsymbol{\beta}_i, \sigma^2) &= \frac{1}{\sigma^2} \int_0^1 N_{k_i}\left(\boldsymbol{\beta}_i \mid \mathbf{0}, \frac{(n-k_0)}{\lambda(k_i+1)} \boldsymbol{\Sigma}_i\right) \frac{1}{\pi \lambda^{\frac{1}{2}} (1-\lambda)^{\frac{1}{2}}}\, d\lambda\,.\end{aligned}$$

The resulting Bayes factor of M_i to M_0 is (recalling that $Q_{i0} = SSE_i/SSE_0$)

$$B_{i0} = \int_0^1 \left(1 + \frac{(n-k_0)}{\lambda(k_i+1)} Q_{i0}\right)^{-\frac{(n-k_0)}{2}} \left(1 + \frac{(n-k_0)}{\lambda(k_i+1)}\right)^{\frac{(n-k_0-k_i)}{2}} \frac{1}{\pi \lambda^{\frac{1}{2}} (1-\lambda)^{\frac{1}{2}}}\, d\lambda\,.$$

Two Innovations:

1. The derivation does not require $\boldsymbol{X}_0$ and $\boldsymbol{X}_i$ to be orthogonal.

2. Obtaining $\boldsymbol{\Sigma}_i$ as an average over training samples is done for more than just a null base model.

Choice of the Base Model for Intrinsic Priors

The null model and 'intercept only' model are often chosen to be the base model. *Proposal:* the base model should be the smallest *plausible* model.

- The intrinsic prior arises from 'imaginary training samples' from the base model. If the base model is extremely implausible, then the intrinsic prior is being trained by training samples that are not at all like the real data.
- The *conditional Lindley paradox* (Som et al. (2016)) can arise when an implausible base model is used. Consider the following three models, with $\boldsymbol{\varepsilon} \sim N_n(\boldsymbol{\varepsilon} \mid 0, I)$: $\quad M_0 : \mathbf{y} = \beta_0 \mathbf{1} + \sigma\boldsymbol{\varepsilon}, \quad M_1 \ : \ \mathbf{y} = \beta_0\mathbf{1} + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \sigma\boldsymbol{\varepsilon}\,,$

  $$M_2 \ : \ \mathbf{y} = \beta_0\mathbf{1} + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \sigma\boldsymbol{\varepsilon}\,,$$

  and suppose $||\boldsymbol{\beta}_1|| \to \infty$. If M_0 is used as the base model, then $B_{21} \to 0$, even if $\boldsymbol{\beta}_2$ is significatively different from zero; indeed, while M_2 is the true model, the posterior probability of M_1 will go to one!
- Simulations in Casella and Moreno (2006) show that using the null model as the base model can give inferior results.

Recommendation: Choose the base model in two stages.

Stage 1. Perform the intrinsic prior analysis using the intercept model (or possibly the null model) as the base model and compute the posterior inclusion probabilities (the overall posterior probability that the variable occurs in models) of all the variables.

Stage 2. If any variables have extremely high posterior inclusion probability (e.g., 0.99), include them in the base model, yielding a new base model M_0^* (with covariates $(\beta_1, \ldots, \beta_{k_0^*})$), and derive the intrinsic priors with respect to this new base model.

The simplest ensuing analysis is to

- only consider M_0^* and larger models;
- assign these models re-weighted prior probabilities;
- complete the model uncertainty analysis with the new intrinsic priors.

This overcomes the problems of using an implausible base model and avoids the need for complications such as using different mixing parameters for different variables.

Here is a result (from Berger et al. (2022)) that shows the discarded models have negligible total posterior probability.

Lemma. Suppose the variable corresponding to β_i has posterior inclusion probability p_i . Let $\mathcal{M}^e$ denote the set of models excluded by the above process. Then the posterior probability of $\mathcal{M}^e$ is less than $\sum_{i=1}^{k_0^*}(1-p_i)$.

Example *Children obesity dataset (OBICE study)* has $n = 996$ and $k = 16$ possible covariates, listed in the first column of Table 1. The original base model was the intercept-only model. The table entries are the posterior inclusion probabilities of the variables. IN refers to variables that were included at Stage 2 in the new base model. Two different prior distributions are considered: the intrinsic prior and the Zellner g-prior.

The Stage 2 analysis restricts the model space to M_0^* (intercept and the IN variables) and larger models. Thus the model space goes from 2^{16} models at Stage 1 to 2^{10} models, for the intrinsic prior, and 2^{11} for the Zellner g-prior at Stage 2. So the vast majority of models are excluded but the upper bounds on the total posterior probability of all excluded models are, respectively,

$$0 + 0.007 + 0 + 0 + 0.002 + 0 = 0.009 \quad \text{and} \quad 0 + 0 + 0 + 0.009 + 0 = 0.009.$$

	Intrinsic prior		Zellner g-prior	
Variable	Stage 1	Stage 2	Stage 1	Stage 2
Age	1.000	IN	1.00	IN
Weight at birth	0.993	IN	0.849	0.845
Height at birth	0.987	0.987	0.817	0.812
Sex	0.567	0.503	0.173	0.172
The father is obese	1.000	IN	1.000	IN
The mother is obese	1.000	IN	1.000	IN
...has 5 daily meals	0.998	IN	0.991	IN
...eats vegetables	0.388	0.326	0.064	0.064
...eats fruit	0.348	0.289	0.053	0.053
...afternoon snacks.	0.841	0.793	0.431	0.423
...was breastfed	0.442	0.376	0.082	0.082
...practices sports	0.885	0.860	0.676	0.672
...watches TV	1.000	IN	1.000	IN
...plays electronic	0.409	0.343	0.066	0.065
...sleeps	0.445	0.380	0.084	0.084
Daily candy consumption	0.987	0.982	0.930	0.928

Table 1: Posterior inclusion probabilities of variables at Stage 1 and Stage 2, under the intrinsic prior and the Zellner g-prior.

The empirical geometric expected posterior prior in linear models

For model M of dimension m (the empirical expected posterior priors do not depend on any other models, so we can drop all the model indices), the empirical density of any training sample $\boldsymbol{y}_m(l)$, of size m from the actual data, is just $1/L = 1/\begin{pmatrix} n \\ m \end{pmatrix}$, so the empirical expected posterior prior, starting with the objective prior $\pi^O(\boldsymbol{\beta}, \sigma^2) = 1/\sigma^2$, is

$$\pi^{EEP}(\boldsymbol{\beta}, \sigma^2) = \sum_l \pi^O(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}_m(l)) \frac{1}{L} \,.$$

The empirical geometric expected posterior prior is, where $\boldsymbol{X}(l)$ denotes the covariates corresponding to $\boldsymbol{y}_m(l)$,

$$\pi^{EGEP}(\boldsymbol{\beta}, \sigma^2) \propto \prod_l [\pi^O(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}_m(l))]^{1/L} \propto \sigma^{-2} \prod_l N_m(\boldsymbol{y}_m(l) \mid \boldsymbol{X}(l), \boldsymbol{\beta}, \sigma^2)^{1/L} \,,$$

since the normalizing constants in the posteriors are just constants. Thus

$$\begin{aligned}\pi^{EGEP}(\boldsymbol{\beta},\sigma^2) &\propto \sigma^{-(2+m)}\exp\left\{-\frac{1}{2\sigma^2\ L}\sum_l |\boldsymbol{y}_m(l)-\boldsymbol{X}(l)\boldsymbol{\beta})|^2\right\}\\ &\propto \sigma^{-(2+m)}\exp\left\{-\frac{m}{2\sigma^2 n}|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})|^2\right\}\\ &\propto \sigma^{-(2+m)}\exp\left\{-\frac{m}{2\sigma^2 n}\left[(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})+S^2\right]\right\},\end{aligned}$$

which is the recommended version of O'Hagan's fractional prior. Note that

$$\pi^{EGEP}(\boldsymbol{\beta}) \propto \left[\frac{(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{S^2}+1\right]^{-m/2}.$$

- This is clearly a 'non-local' prior; in contrast, the intrinsic prior was a 'local' prior.
- In general, intrinsic priors are local priors (favoring smaller models) and empirical expected posterior priors are non-local priors (favoring larger models). Try both and, if the answers are similar, great!
- While centered at $\widehat{\boldsymbol{\beta}}$, this has very flat tails.

Luis and I recommended weighting training samples by their information content $|\boldsymbol{X}(l)'\boldsymbol{X}(l)|$. Using these in geometric averaging with the intrinsic prior (null model as base model and given λ and σ^2) yields the prior

$$N_{k_i}\left(\boldsymbol{\beta} \,\middle|\, \mathbf{0}, \frac{n\sigma^2}{\lambda(k_i+1)}\boldsymbol{\Omega}\right), \quad \boldsymbol{\Omega}^{-1} = \sum_l \frac{|\boldsymbol{X}(l)'\boldsymbol{X}(l)|}{|\boldsymbol{X}'\boldsymbol{X}|}(\boldsymbol{X}(l)'\boldsymbol{X}(l)).$$

Example: Suppose $\boldsymbol{X}' = \begin{pmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \end{pmatrix} \equiv \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \end{pmatrix}$ and $k_i = 2$.

Minimal training samples are of size $m = 2$. It can be shown that

$$\begin{aligned} \boldsymbol{\Omega}^{-1} &= \frac{1}{(|\boldsymbol{x}_1|^2|\boldsymbol{x}_2|^2 - 2\boldsymbol{x}_1'\boldsymbol{x}_2)}\left[|\boldsymbol{x}_1|^2 \begin{pmatrix} \boldsymbol{x}_1^{(2)'}\boldsymbol{x}_2^{(2)} & \boldsymbol{x}_1'\boldsymbol{x}_2^{(3)} \\ \boldsymbol{x}_1'\boldsymbol{x}_2^{(3)} & \boldsymbol{x}_2^{(2)'}\boldsymbol{x}_2^{(2)} \end{pmatrix}\right. \\ &+ \left. |\boldsymbol{x}_2|^2 \begin{pmatrix} \boldsymbol{x}_1^{(2)'}\boldsymbol{x}_1^{(2)} & \boldsymbol{x}_2'\boldsymbol{x}_1^{(3)} \\ \boldsymbol{x}_2'\boldsymbol{x}_1^{(3)} & \boldsymbol{x}_1^{(2)'}\boldsymbol{x}_2^{(2)} \end{pmatrix} - 2(\boldsymbol{x}_1'\boldsymbol{x}_2) \begin{pmatrix} \boldsymbol{x}_2'\boldsymbol{x}_1^{(3)} & \boldsymbol{x}_1^{(2)'}\boldsymbol{x}_2^{(2)} \\ \boldsymbol{x}_1^{(2)'}\boldsymbol{x}_2^{(2)} & \boldsymbol{x}_1'\boldsymbol{x}_2^{(3)} \end{pmatrix}\right], \end{aligned}$$

where $\boldsymbol{v}^{(j)} = (v_1^j, v_2^j, \dots, v_n^j)'$.

Thanks all and
HAPPY BIRTHDAY LUIS!

References

Berger, J., G. García-Donato, E. Moreno, and L. Pericchi (2022). The intrinsic hyper-g prior for normal linear models. *in preparation*.

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

Som, A., C. Hans, and S. N. MacEachern (2016). A conditional Lindley paradox in Bayesian linear models. *Biometrika 103*(4), 993–999.