Integral Prior Distributions for linear models and multiple comparison

Salmerón D. and Cano J.A.

Universidad de Murcia

- Integral priors for model selection and testing
- The Markov chains associated with the integral priors
- Computation of Bayes factors with integral priors
- Multiple comparison
- Variable selection

The problem

- Two models

$$M_i : f_i(\cdot \mid \theta_i), \ \ \theta_i \in \Theta_i, \ \ i = 1, 2$$

are under consideration to explain the data $\mathbf{x}$

- The Bayes factor

$$B_{21} = \frac{m_2(\mathbf{x})}{m_1(\mathbf{x})} = \frac{\int f_2(\mathbf{x} \mid \theta_2)\pi_2(\theta_2)\mathrm{d}\theta_2}{\int f_1(\mathbf{x} \mid \theta_1)\pi_1(\theta_1)\mathrm{d}\theta_1}$$

requires specification of $\pi_1(\theta_1)$ and $\pi_2(\theta_2)$

Estimation priors do not work

- Default priors, $\pi_i^N(\theta_i)$ (Jeffreys or reference priors) are often used for estimation
- Usually improper priors

$$\pi_i^N(\theta_i) = c_i h_i(\theta_i), \;\; c_i > 0, \;\; \int h_i(\theta_i) \mathrm{d}\theta_i = +\infty$$

- The Bayes factor is not well-defined

$$B_{21}^N = \frac{\int f_2(\mathbf{x} \mid \theta_2) \pi_2^N(\theta_2) \mathrm{d}\theta_2}{\int f_1(\mathbf{x} \mid \theta_1) \pi_1^N(\theta_1) \mathrm{d}\theta_1} = \frac{c_2}{c_1} \frac{\int f_2(\mathbf{x} \mid \theta_2) h_2(\theta_2) \mathrm{d}\theta_2}{\int f_1(\mathbf{x} \mid \theta_1) h_1(\theta_1) \mathrm{d}\theta_1}$$

because c_2/c_1 is arbitrary

Proposals for model selection priors

The Intrinsic Bayes Factor for Model Selection and Prediction

James O. Berger and Luis R. Pericchi

Expected-posterior prior distributions for model selection

By JOSÉ M. PÉREZ

Centro de Estadística y Software Matemático, Universidad Simón Bolívar, Aptdo. 89000, Caracas 1080A, Venezuela

jperez@cesma.usb.ve

and JAMES O. BERGER

Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina 27708-0251, U.S.A.

berger@stat.duke.edu

An Intrinsic Limiting Procedure for Model Selection and Hypotheses Testing

Elías Moreno, Francesco Bertolino, and Walter Racugno

Proposals for model selection priors

Integral equation solutions as prior distributions for Bayesian model selection

J.A. Cano · D. Salmerón · C.P. Robert

Generalization of Jeffreys divergence-based priors for Bayesian hypothesis testing

M. J. Bayarri
University of Valencia, Spain

and G. García-Donato
University of Castilla-La Mancha, Albacete, Spain

CRITERIA FOR BAYESIAN MODEL CHOICE WITH APPLICATION TO VARIABLE SELECTION[1]

BY M. J. BAYARRI, J. O. BERGER, A. FORTE AND G. GARCÍA-DONATO

Proposals for variable selection priors

Zellner's g-priors (1986) and Mixtures (Liang *et al.* (2008))

Robust prior (Bayarri *et al.* (2012))

Power-expected-posterior priors (Fouskakis *et al.* (2015))

Intrinsic priors for nested models

$\{\pi_1^I(\theta_1), \pi_2^I(\theta_2)\}$

$$\pi_2^I(\theta_2) = \int \pi_2^I(\theta_2 \mid \theta_1)\pi_1^I(\theta_1)\mathrm{d}\theta_1$$

$\pi_2^I(\theta_2 \mid \theta_1) = \int \pi_2^N(\theta_2 \mid x) f_1(x \mid \theta_1)\mathrm{d}x$

x is an imaginary minimal training sample

Intrinsic priors for nested models

$\{\pi_1^I(\theta_1), \pi_2^I(\theta_2)\}$

$$\pi_2^I(\theta_2) = \int \pi_2^I(\theta_2 \mid \theta_1)\pi_1^I(\theta_1)\mathrm{d}\theta_1$$

$\pi_2^I(\theta_2 \mid \theta_1) = \int \pi_2^N(\theta_2 \mid x) f_1(x \mid \theta_1)\mathrm{d}x$

x is an imaginary minimal training sample

$\pi_1^I(\theta_1)$ is free!!!

Usually $\pi_1^I(\theta_1) := \pi_1^N(\theta_1)$ (Moreno *et al.* (1998))

Expected posterior priors

$$\pi_1^E(\theta_1) := \int \pi_1^N(\theta_1 \mid x) m(x) \mathrm{d}x$$

$$\pi_2^E(\theta_2) := \int \pi_2^N(\theta_2 \mid x) m(x) \mathrm{d}x$$

where x is an imaginary minimal training sample and $m(x)$ can be any predictive distribution, proper or not

Two proposals for $m(x)$ are the empirical distribution of the data, and the predictive distribution of the simplest model

Expected posterior priors for nested models

- If M_1 is nested in M_2, and $m(x) := m_1^N(x)$, then

$$\pi_i^E(\theta_i) = \pi_i^I(\theta_i), \quad i = 1, 2$$

- The expected posterior priors can be seen as a generalization of intrinsic priors

Integral priors

Integral priors

Integral priors are the solutions $\pi_1(\theta_1)$ and $\pi_2(\theta_2)$ to the system of integral equations

$$\pi_1(\theta_1) = \int \pi_1^N(\theta_1 \mid x) m_2(x) \mathrm{d}x$$
$$\pi_2(\theta_2) = \int \pi_2^N(\theta_2 \mid x) m_1(x) \mathrm{d}x$$

where

$$m_i(x) = \int f_i(x \mid \theta_i) \pi_i(\theta_i) \mathrm{d}\theta_i, \quad i = 1, 2,$$

and x is an imaginary minimal training sample

Integral priors

Because of $m_i(x) = \int f_i(x \mid \theta_i)\pi_i(\theta_i)\mathrm{d}\theta_i$

$$\pi_1(\theta_1) = \int \pi_1^N(\theta_1 \mid x) f_2(x \mid \theta_2)\pi_2(\theta_2)\mathrm{d}x\mathrm{d}\theta_2$$

$$\pi_2(\theta_2) = \int \pi_2^N(\theta_2 \mid x) f_1(x \mid \theta_1)\pi_1(\theta_1)\mathrm{d}x\mathrm{d}\theta_1$$

Therefore we have a system of two integral equations, and the integral priors are its solution

Justification

Model selection priors should be *close* to the initial default priors

Any prior $\pi_1(\theta_1)$ satisfies

$$\pi_1(\theta_1) = \int \pi_1(\theta_1 \mid x) m_1(x) \mathrm{d}x,$$

Justification

Model selection priors should be *close* to the initial default priors

Any prior $\pi_1(\theta_1)$ satisfies

$$\pi_1(\theta_1) = \int \pi_1(\theta_1 \mid x) m_1(x) \mathrm{d}x,$$

and a sensible way to get a prior $\pi_1(\theta_1)$ *close* to $\pi_1^N(\theta_1)$ is by means of

$$\pi_1(\theta_1) = \int \pi_1^N(\theta_1 \mid x) m_1(x) \mathrm{d}x$$

Justification

The predictive distributions $m_i(x) = \int f_i(x \mid \theta_i)\pi_i(\theta_i)\mathrm{d}\theta_i$, $i = 1, 2$, should be as *close* as possible.

Any prior $\pi_1(\theta_1)$ satisfies

$$\pi_1(\theta_1) = \int \pi_1(\theta_1 \mid x) m_1(x) \mathrm{d}x,$$

Justification

The predictive distributions $m_i(x) = \int f_i(x \mid \theta_i)\pi_i(\theta_i)\mathrm{d}\theta_i$, $i = 1, 2$, should be as *close* as possible.

Any prior $\pi_1(\theta_1)$ satisfies

$$\pi_1(\theta_1) = \int \pi_1(\theta_1 \mid x) m_1(x) \mathrm{d}x,$$

and a sensible way to get $m_1(x)$ and $m_2(x)$ to be *close* is by means of

$$\pi_1(\theta_1) = \int \pi_1(\theta_1 \mid x) m_2(x) \mathrm{d}x$$

$$\pi_1(\theta_1) = \int \pi_1(\theta_1 \mid x) m_1(x) \mathrm{d}x$$

$\pi_1(\theta_1)$ close to $\pi_1^N(\theta_1)$

$m_1(x)$ and $m_2(x)$ should be as close as possible

$$\pi_1(\theta_1) = \int \pi_1^N(\theta_1 \mid x) m_1(x) \mathrm{d}x$$

$$\pi_1(\theta_1) = \int \pi_1(\theta_1 \mid x) m_2(x) \mathrm{d}x$$

$$\pi_1(\theta_1) = \int \pi_1(\theta_1 \mid x) m_1(x) \mathrm{d}x$$

$\pi_1(\theta_1)$ close to $\pi_1^N(\theta_1)$

$m_1(x)$ and $m_2(x)$ should be as close as possible

$$\pi_1(\theta_1) = \int \pi_1^N(\theta_1 \mid x) m_1(x) \mathrm{d}x$$

$$\pi_1(\theta_1) = \int \pi_1(\theta_1 \mid x) m_2(x) \mathrm{d}x$$

$$\pi_1(\theta_1) = \int \pi_1^N(\theta_1 \mid x) m_2(x) \mathrm{d}x$$

In summary, a sensible way to get priors

- *close* to the initial default priors, and
- with predictive distributions as *close* as possible (**predictive matching**, see Berger and Pericchi (2001), and Bayarri *et al.* (2012))

is by means of

$$\pi_1(\theta_1) = \int \pi_1^N(\theta_1 \mid x) m_2(x) \mathrm{d}x$$
$$\pi_2(\theta_2) = \int \pi_2^N(\theta_2 \mid x) m_1(x) \mathrm{d}x$$

and these are the integral priors!!!

Intrinsic and Integral priors

- If M_1 is nested in M_2, then the intrinsic priors satisfy

$$\pi_2(\theta_2) = \int \pi_2^I(\theta_2 \mid \theta_1)\pi_1(\theta_1)\mathrm{d}\theta_1$$

- If we add the symmetrical equation

$$\pi_1(\theta_1) = \int \pi_1^I(\theta_1 \mid \theta_2)\pi_2(\theta_2)\mathrm{d}\theta_2$$

Intrinsic and Integral priors

- If M_1 is nested in M_2, then the intrinsic priors satisfy

$$\pi_2(\theta_2) = \int \pi_2^I(\theta_2 \mid \theta_1)\pi_1(\theta_1)\mathrm{d}\theta_1$$

- If we add the symmetrical equation

$$\pi_1(\theta_1) = \int \pi_1^I(\theta_1 \mid \theta_2)\pi_2(\theta_2)\mathrm{d}\theta_2$$

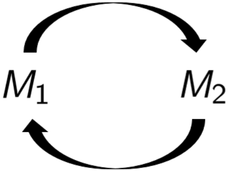
- Again we have the integral priors!!!

The Markov chains associated with the integral priors

The associated Markov chains

The integral prior $\pi_1(\theta_1)$ is the invariant σ-finite measure of the Markov chain with transition $\theta_1 \to \theta_1'$ defined by the following four steps

1. $z_2 \sim f_1(z_2 \mid \theta_1)$
2. $\theta_2 \sim \pi_2^N(\theta_2 \mid z_2)$
3. $z_1 \sim f_2(z_1 \mid \theta_2)$
4. $\theta_1' \sim \pi_1^N(\theta_1' \mid z_1)$



- If this Markov chain is **Harris recurrent**, then the integral prior $\pi_1(\theta_1)$ **can be approximated by simulation**
- Therefore, the transition $\theta_1 \to \theta_1'$ and the integral priors are essentially the same thing
- There exists a parallel Markov chain for θ_2 with the same properties; in particular, if one is (Harris) recurrent then so is the other

The first example

One-sided testing for the exponential distribution

$$M_1 : \mathcal{E}xp(\theta_1),\ \theta_1 < 1$$

$$M_2 : \mathcal{E}xp(\theta_2),\ \theta_2 > 1$$

$$\pi_1^N(\theta_1) \propto \theta_1^{-1} 1_{(0,1)}(\theta_1)$$

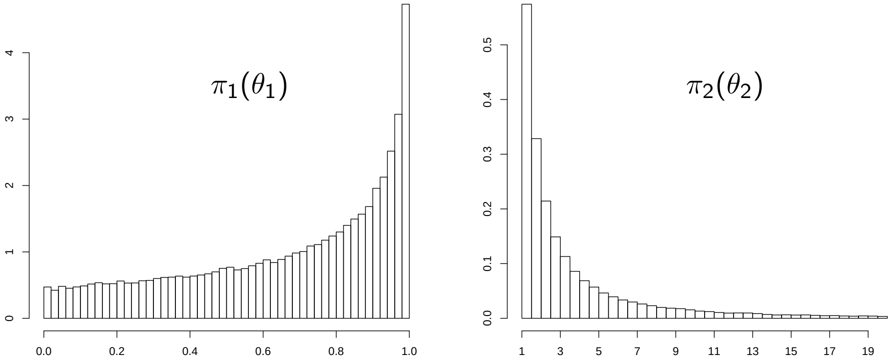
$$\pi_2^N(\theta_2) \propto \theta_2^{-1} 1_{(1,+\infty)}(\theta_2)$$

Markov chain for θ_1

1. $x' = -\theta_1 \log u_1$
2. $\theta_2 = -x' / \log(u_2(1 - e^{-x'}) + e^{-x'})$
3. $x = -\theta_2 \log u_3$
4. $\theta_1' = (1 - \frac{1}{x} \log u_4)$

$u_1, u_2, u_3, u_4 \sim U(0, 1)$

$M_1 : \theta_1 < 1$ $\qquad$ $M_2 : \theta_2 > 1$



- Integral priors can be applied to nested and non-nested situations
- Priors *close* to the initial default priors, and with predictive distributions as *close* as possible
- The integral prior for each model takes into account the existence of the other model

Group invariance

- An important situation is when M_1 and M_2 have the same group invariance structure
- In this situation right-Haar priors are exact predictive matching for minimal training samples (Berger, Pericchi and Varshavsky (1998))
- Right-Haar priors are Integral priors when these priors are the initial default priors

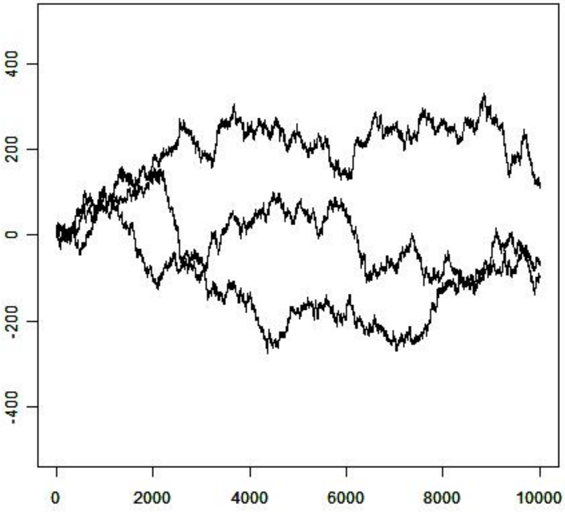
Location models

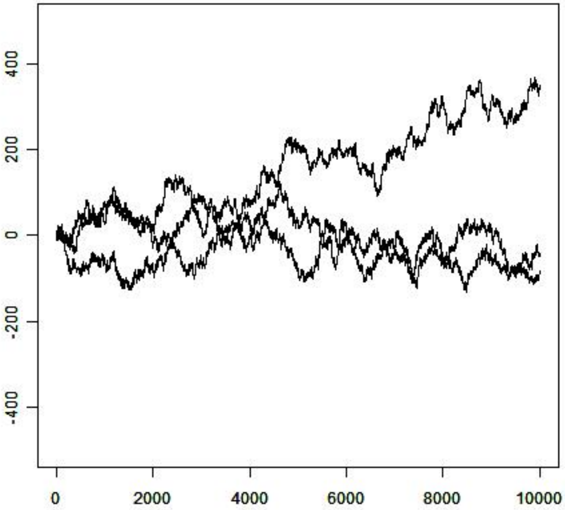
$M_1: N(\theta, 1), \;\; \pi_1^N(\theta) = c_1$

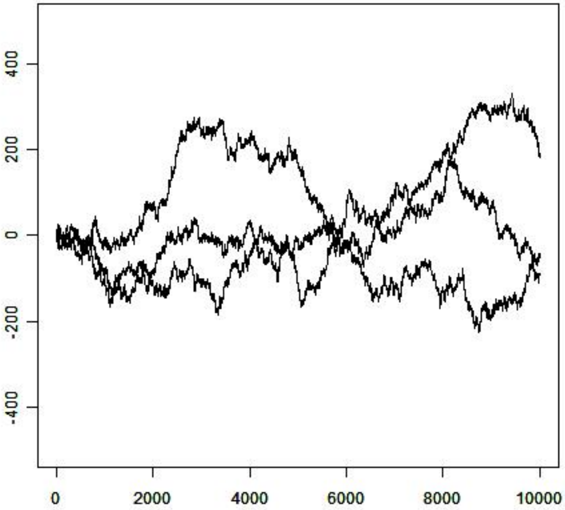
$M_2: DE(\lambda, 1), \;\; \pi_2^N(\lambda) = c_2$

$\pi_1(\theta) = 1$ and $\pi_2(\lambda) = 1$ are the integral priors

Because these priors are improper, we expect a lack of stability in their associated Markov chains


400
200
0
-200
-400
0
2000
4000
6000
8000
10000


400
200
0
-200
-400
0
2000
4000
6000
8000
10000


400
200
0
-200
-400
0
2000
4000
6000
8000
10000

Bayesian Analysis (2013) 8, Number 2, pp. 361–380

Integral Priors and Constrained Imaginary Training Samples for Nested and Non-nested Bayesian Model Comparison

Juan Antonio Cano * and Diego Salmerón †‡

1. $x' \sim f_1(x' \mid \theta_1)$
2. $\theta_2 \sim \pi_2^N(\theta_2 \mid x')$
3. $x \sim f_2(x \mid \theta_2)$
4. $\theta_1' \sim \pi_1^N(\theta_1' \mid x)$

Bayesian Analysis (2013) 8, Number 2, pp. 361–380

Integral Priors and Constrained Imaginary Training Samples for Nested and Non-nested Bayesian Model Comparison

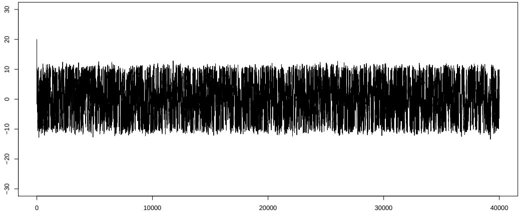
Juan Antonio Cano * and Diego Salmerón †‡

1. $x' \sim f_1(x' \mid \theta_1)$
2. $\theta_2 \sim \pi_2^N(\theta_2 \mid x')$
3. $x \sim f_2(x \mid \theta_2)$
4. $\theta_1' \sim \pi_1^N(\theta_1' \mid x)$

1. $x' \sim f_1^A(x' \mid \theta_1) \propto f_1(x' \mid \theta_1)\mathbb{I}_A(x)$
2. $\theta_2 \sim \pi_2^N(\theta_2 \mid x')$
3. $x \sim f_2^A(x \mid \theta_2) \propto f_2(x' \mid \theta_2)\mathbb{I}_A(x)$
4. $\theta_1' \sim \pi_1^N(\theta_1' \mid x)$

$M_1 : N(\theta, 1), \ \pi_1^N(\theta) = c_1$ and $M_2 : DE(\lambda, 1), \ \pi_2^N(\lambda) = c_2$

The constraint $x \in A = [-10, 10]$ on the imaginary trainig samples prevents the *explosion* of the chain



Our recommendation is keeping the imaginary training samples within an interval $\pm 5s$ about the sample mean

The only thing one needs to apply this methodology is

- To simulate **minimal** training samples from $f_i(x \mid \theta_i)$, which seems easy to do, and
- To simulate from the posteriors $\pi_i^N(\theta_i \mid x)$, which usually is also easy to do, or it can be done using MCMC

The one way heteroscedastic ANOVA

$$M_1 : \mu_1 = \mu_2 = \cdots = \mu_k = \mu$$
$$M_2 : \text{all the } \mu_i{}'s \text{ are not equal}$$

$$\pi_1^N(\mu, \sigma_1, \ldots, \sigma_k) \propto (\sigma_1 \cdots \sigma_k)^{-1}$$
$$\pi_2^N(\mu_1, \ldots, \mu_k, \sigma_1, \ldots, \sigma_k) \propto (\sigma_1 \cdots \sigma_k)^{-1}$$

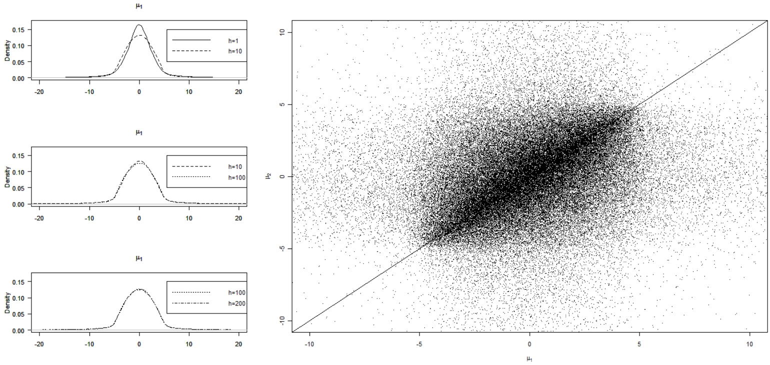
Here the simulation from the posterior $\pi_1^N(\theta_1 \mid x)$ can not be performed directly

$$\pi_1^N(\mu, \sigma_1, \ldots, \sigma_k \mid x) \propto \prod_{i=1}^{k} \sigma_i^{-3} \exp\left(-\frac{(x_{i1}-\mu)^2 + (x_{i2}-\mu)^2}{2\sigma_i^2}\right).$$

We use Gibbs sampling within this step

1. $x' \sim f_1^A(x' \mid \theta_1) \propto f_1(x' \mid \theta_1)\mathbb{I}_A(x')$
2. $\theta_2 \sim \pi_2^N(\theta_2 \mid x')$
3. $x \sim f_2^A(x \mid \theta_2) \propto f_2(x \mid \theta_2)\mathbb{I}_A(x)$
4. $\theta_1' \sim \pi_1^N(\theta_1' \mid x)$: **Gibbs sampling** with $h \geq 1$ iterations

Four populations. 100,000 iterations for the Markov chain and $h = 1, 10, 100, 200$ iterations of the Gibbs sampling



- There are no differences from $h = 10$ to $h = 100$ or larger, so $h = 10$ is enough for the Gibbs algorithm
- Integral prior for model M_2 concentrates mass in favor of model M_1

Cano, J. A., Kessler, M. and Salmerón, D. (2007a). Integral priors for the one way random effects model. Bayesian Analysis, 2-1, 59–68.

Cano, J. A., Kessler, M. and Salmerón, D. (2007b). A synopsis of integral priors for the one way random effects model. Bayesian Statistics, 8, 577–582. Oxford University Press.

Cano, J. A., Salmerón, D. and Robert, C. P. (2008). Integral equation solutions as prior distributions for Bayesian model selection. Test, 17-3, 493–504.

Cano, J. A. and Salmerón, D. (2013). Integral Priors and Constrained Imaginary Training Samples for Nested and Non-nested Bayesian Model Comparison. Bayesian Analysis, 8-2, 361–380.

Salmerón, D., Cano, J. A. and Robert, C. P. (2015). Objective Bayesian hypothesis testing in binomial regression models with integral prior distributions. Statistica Sinica, 25-3, 1009–1023.

Cano, J.A. and Salmerón, D. (2016). A Review of the Developments on Integral Priors for Bayesian Model Selection. BEIO, 32-2, 96-111.

Cano, J.A., Iniesta M. and Salmerón, D. Integral priors for Bayesian model selection: How they operate from simple to complex cases. Submitted.

Computation of Bayes factors with integral priors

- Monte Carlo
- Laplace approximation
- Importance sampling

Monte Carlo

- The Markov chain $\theta_i^{(1)}, \theta_i^{(2)}, \dots$ for $\pi_i(\theta_i)$

$$\lim_{L \to +\infty} \frac{1}{L} \sum_{t=1}^{L} f_i(\boldsymbol{x} \mid \theta_i^t) = m_i(\boldsymbol{x}) = \int f_i(\boldsymbol{x} \mid \theta_i) \pi_i(\theta_i) \mathrm{d}\theta_i$$

- Very large values of L are needed if $f_i(\boldsymbol{x} \mid \theta_i)$ is concentrated relative to $\pi_i(\theta_i)$

Laplace approximation

$m_i(\boldsymbol{x}) = \int f_i(\boldsymbol{x} \mid \theta_i)\pi_i(\theta_i)\mathrm{d}\theta_i$

$\hat{\pi}_i$ is a nonparametric estimate of the integral prior π_i

$$\hat{m}_i(\boldsymbol{x}) = (2\pi)^{\frac{dim(\Theta_i)}{2}} |\hat{\Sigma}_i|^{1/2} f_i(\boldsymbol{x} \mid \hat{\theta}_i)\hat{\pi}_i(\hat{\theta}_i)$$

$\hat{\theta}_i = MLE$

$\hat{\Sigma}_i^{-1}$ observed information matrix under M_i

Importance sampling I

$\hat{\pi}_i$ is a nonparametric estimate of the integral prior π_i

$$m_i(\boldsymbol{x}) = \int f_i(\boldsymbol{x} \mid \theta_i)\pi_i(\theta_i)\mathrm{d}\theta_i \approx \int f_i(\boldsymbol{x} \mid \theta_i)\hat{\pi}_i(\theta_i)\mathrm{d}\theta_i$$

$$= \int \frac{f_i(\boldsymbol{x} \mid \theta_i)\hat{\pi}_i(\theta_i)}{p(\theta_i \mid \boldsymbol{x})} p(\theta_i \mid \boldsymbol{x})\mathrm{d}\theta_i$$

$p(\theta_i \mid \boldsymbol{x})$ the importance density

Importance sampling II

$$m_2(\boldsymbol{x}) = \int f_2(\boldsymbol{x} \mid \theta_2)\pi_2(\theta_2)\mathrm{d}\theta_2 = \int f_2(\boldsymbol{x} \mid \theta_2)\pi_2^N(\theta_2 \mid z_2)m_1(z_2)\mathrm{d}z_2\mathrm{d}\theta_2$$

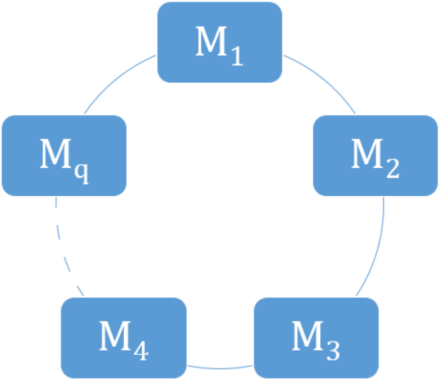
$$= \int \frac{f_2(\boldsymbol{x} \mid \theta_2)\pi_2^N(\theta_2 \mid z_2)}{p(\theta_2 \mid \boldsymbol{x}, z_2)} p(\theta_2 \mid \boldsymbol{x}, z_2)m_1(z_2)\mathrm{d}z_2\mathrm{d}\theta_2$$

$p(\theta_2 \mid \boldsymbol{x}, z_2)$ the importance density

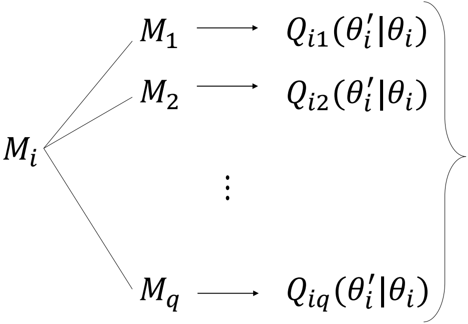
The simulation of the Markov chain gives us simulations from $m_1(z_2)$

We need to evaluate $\pi_2^N(\theta_2 \mid z_2)$, where z_2 is a minimal training sample

Multiple comparison



Integral priors for Multiple comparison



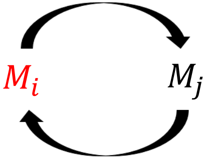
$$Q_i(\theta_i'|\theta_i) = \frac{\sum_{j \neq i} Q_{ij}(\theta_i'|\theta_i)}{q-1}$$

Integral priors for Multiple comparison

Definition: $\pi_i(\theta_i)$ for M_i

The integral prior $\pi_i(\theta_i)$ is the invariant σ-finite measure of the Markov chain with transition $\theta_i \to \theta_i'$ defined by the following four steps

1. $x_j \sim f_i(\cdot \mid \theta_i)$
2. $\theta_j \sim \pi_j^N(\cdot \mid x_j)$
3. $x_i \sim f_j(\cdot \mid \theta_j)$
4. $\theta_i' \sim \pi_i^N(\cdot \mid x_i)$



$j \sim \mathcal{U}\{1, 2, \ldots, i-1, i+1, \ldots, q\}$

Testing for the exponential distribution

$$M_1 : \mathcal{E}xp(\theta_1), \;\; \theta_1 \in I_1 = (0,1)$$

$$M_2 : \mathcal{E}xp(\theta_2), \;\; \theta_2 \in I_2 = (1,+\infty)$$

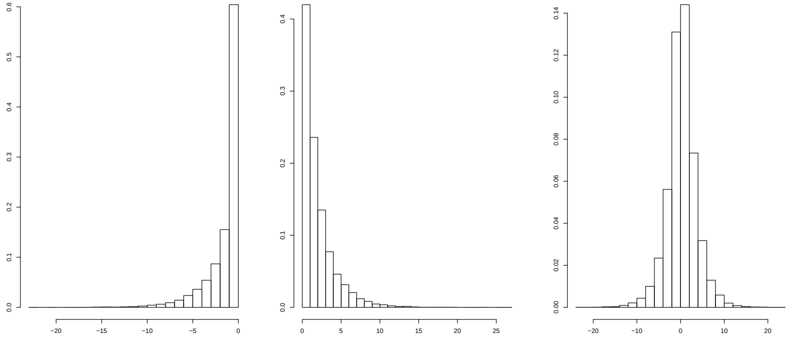
$$M_3 : \mathcal{E}xp(\theta_3), \;\; \theta_3 \in I_3 = (0,+\infty)$$

$$\pi_i^N(\theta_i) \propto \theta_i^{-1} 1_{I_i}(\theta_i)$$

$$\xi_i = \log \theta_i$$

$$i = 1,2,3$$

$\xi_1 < 0$ $\xi_2 > 0$ $\xi_3 \in \mathbb{R}$



Variable selection

Variable selection for the linear regression model

Full model

$$\boldsymbol{y} = X\beta + \varepsilon, \;\; \varepsilon \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I})$$

$$\pi^N(\beta, \sigma) \propto 1/\sigma$$

$\beta \in \mathbb{R}^k$, $\sigma > 0$

$X = [x_1, \dots, x_k]$ an $n \times k$ full rank matrix and $n > k$

$x_j = (x_{1j}, \dots, x_{nj})'$, $j = 1, \dots, k$

Usually $x_1 = \mathbf{1}_n$

Submodels

The full model is represented by the matrix $X = (x_{ij}) \in \mathbb{R}^{n \times k}$

$\mathcal{R}$ a subsequence of $\mathcal{I} = \{1, \ldots, n\}$ representing rows of X

$\mathcal{C}$ a subsequence of $\mathcal{J} = \{1, \ldots, k\}$ representing columns of X

$X_{\mathcal{R},\mathcal{C}} = (x_{ij})_{i \in \mathcal{R}, j \in \mathcal{C}}$ and $X_{\mathcal{C}} = X_{\mathcal{I},\mathcal{C}}$

The submodel $M_{\mathcal{C}}$ is represented by the matrix $X_{\mathcal{C}}$

$$M_{\mathcal{C}} : \boldsymbol{y} = X_{\mathcal{C}} \beta_{\mathcal{C}} + \boldsymbol{\varepsilon}_{\mathcal{C}}, \;\; \boldsymbol{\varepsilon}_{\mathcal{C}} \sim N_n(\mathbf{0}, \sigma_{\mathcal{C}}^2 \mathbf{I})$$

$$\pi^N(\beta_{\mathcal{C}}, \sigma_{\mathcal{C}}) \propto 1/\sigma_{\mathcal{C}}$$

$M_{\mathcal{C}_1}$ versus $M_{\mathcal{C}_2}$

$(\beta_{\mathcal{C}_1}, \sigma_{\mathcal{C}_1}) \to (\beta'_{\mathcal{C}_1}, \sigma'_{\mathcal{C}_1})$

Select random sequences, $\mathcal{R}$ and $\mathcal{S}$, from $\mathcal{I} = \{1, \ldots, n\}$, with $|\mathcal{R}| = |\mathcal{C}_2| + 1$ and $|\mathcal{S}| = |\mathcal{C}_1| + 1$, such that $X_{\mathcal{R}\mathcal{C}_2}$ and $X_{\mathcal{S}\mathcal{C}_1}$ be full rank matrices.

1. Simulate a training sample $y_2 \sim N(X_{\mathcal{R}\mathcal{C}_1}\beta_{\mathcal{C}_1}, \sigma^2_{\mathcal{C}_1}\mathbf{I})$
2. Simulate the posterior $\pi^N(\beta_{\mathcal{C}_2}, \sigma_{\mathcal{C}_2} \mid y_2, X_{\mathcal{R}\mathcal{C}_2})$
3. Simulate a training sample $y_1 \sim N(X_{\mathcal{S}\mathcal{C}_2}\beta_{\mathcal{C}_2}, \sigma^2_{\mathcal{C}_2}\mathbf{I})$
4. Simulate the posterior $\pi^N(\beta'_{\mathcal{C}_1}, \sigma'_{\mathcal{C}_1} \mid y_1, X_{\mathcal{S}\mathcal{C}_1})$

The caterpillar dataset: $2^{10} = 1024$ models

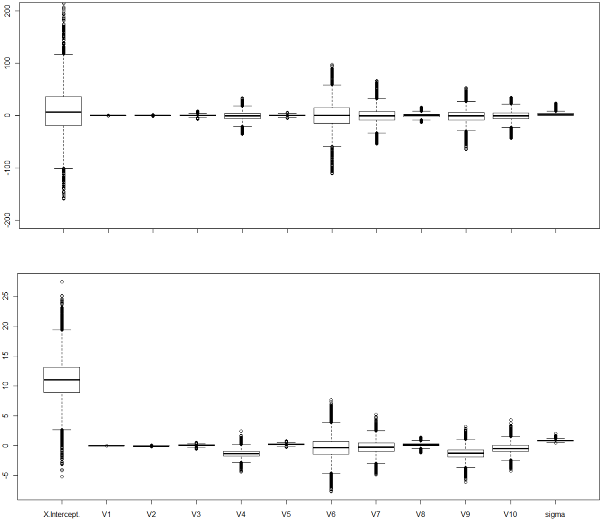
$Y =$ log of the average number of nests of caterpillars per tree in an area

$k = 10$ potential explanatory variables defined on $n = 33$ areas

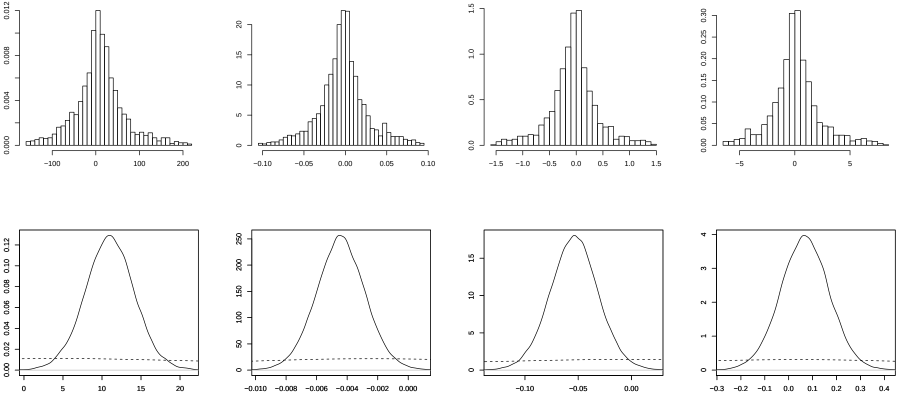
x_1 altitude
x_2 slope
x_3 number of pines in the area
$\vdots$

Bayesian core: a practical approach to computational Bayesian statistics. Jean-Michel Marin and Christian P. Robert. Springer.

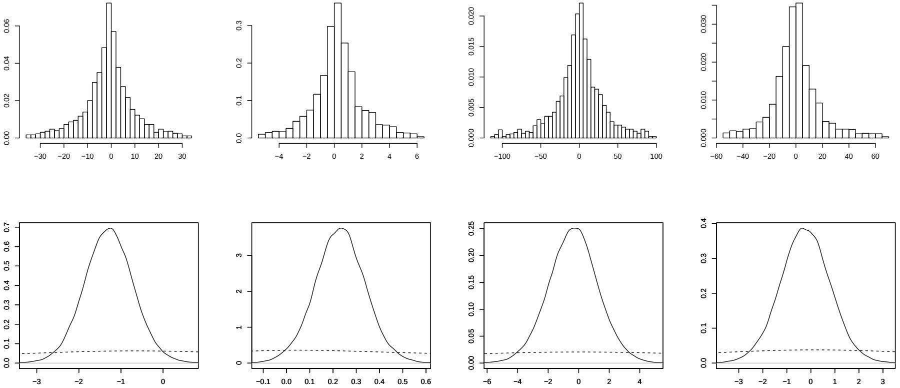
Full model: Integral prior and $\pi^N(\theta \mid \mathbf{y})$



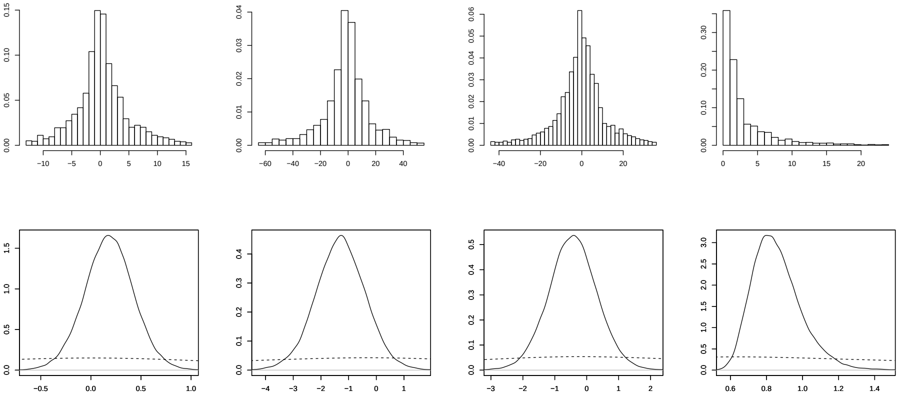
Full model: Integral prior and $\pi^N(\theta \mid \mathbf{y})$



Full model: Integral prior and $\pi^N(\theta \mid \mathbf{y})$



Full model: Integral prior and $\pi^N(\theta \mid \mathbf{y})$



The marginal distributions with integral priors

$$m(\boldsymbol{y}) = \int f(\boldsymbol{y} \mid \theta)\pi(\theta)\mathrm{d}\theta = \int f(\boldsymbol{y} \mid \theta)\pi^N(\theta)\frac{\pi(\theta)}{\pi^N(\theta)}\mathrm{d}\theta$$

$$\approx \int f(\boldsymbol{y} \mid \theta)\pi^N(\theta)\frac{\hat{\pi}(\theta)}{\pi^N(\theta)}\mathrm{d}\theta \approx \frac{\hat{\pi}(\hat{\theta})}{\pi^N(\hat{\theta})}\int f(\boldsymbol{y} \mid \theta)\pi^N(\theta)\mathrm{d}\theta$$

$$m(\boldsymbol{y}) \approx \frac{\hat{\pi}(\hat{\theta})}{\pi^N(\hat{\theta})} m^N(\boldsymbol{y})$$

Variables in the model	Posterior probability (%)
(0,9)	21.8
(0,1,9)	11.2
(0,3)	6.4
(0,8)	5.6
(0,6)	3.9
(0,2,9)	3.4
(0,4,9)	2.2
(0,1,2,4,5)	2.2
(0,1,8)	2.2
(0,1)	2
(0,1,2,9)	1.9
(0,1,2)	1.7
(0,1,4,5)	1.5
(0,5,9)	1.1
(0,7,9)	1.1

Variables	Posterior probability (%)
V1	35.7
V2	23.5
V3	12.1
V4	17.9
V5	13.8
V6	10.6
V7	5.6
V8	15.5
V9	53.8
V10	4.3

For two Binomial regression models with a general link function

Statistica Sinica **25** (2015), 1009-1023
doi:http://dx.doi.org/10.5705/ss.2013.338

OBJECTIVE BAYESIAN HYPOTHESIS TESTING IN BINOMIAL REGRESSION MODELS WITH INTEGRAL PRIOR DISTRIBUTIONS

D. Salmerón, J. A. Cano and C. P. Robert

CIBER Epidemiología y Salud Pública (CIBERESP), Universidad de Murcia and PSL, Université Paris-Dauphine

Variable selection for Generalized linear models

For two Binomial regression models with a general link function

Statistica Sinica **25** (2015), 1009-1023
doi:http://dx.doi.org/10.5705/ss.2013.338

OBJECTIVE BAYESIAN HYPOTHESIS TESTING IN BINOMIAL REGRESSION MODELS WITH INTEGRAL PRIOR DISTRIBUTIONS

D. Salmerón, J. A. Cano and C. P. Robert

CIBER Epidemiología y Salud Pública (CIBERESP), Universidad de Murcia and PSL, Université Paris-Dauphine

1. Simulate a training sample y_2
2. Simulate the posterior given y_2
3. Simulate a training sample y_1
4. Simulate the posterior given y_1

It can be done for multiple comparison!

Variable selection for Nonlinear regression models

$$\boldsymbol{y} = \boldsymbol{g}(X, \beta) + \varepsilon, \;\; \varepsilon \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})$$

It can be done!

$$(\beta_{\mathcal{C}_1}, \sigma_{\mathcal{C}_1}) \to (\beta'_{\mathcal{C}_1}, \sigma'_{\mathcal{C}_1})$$

1. Simulate a training sample y_2:

Variable selection for Nonlinear regression models

$$\boldsymbol{y} = \boldsymbol{g}(X, \beta) + \varepsilon, \ \ \varepsilon \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})$$

It can be done!

$$(\beta_{\mathcal{C}_1}, \sigma_{\mathcal{C}_1}) \to (\beta'_{\mathcal{C}_1}, \sigma'_{\mathcal{C}_1})$$

1. Simulate a training sample y_2: it is easy
2. Simulate the posterior given y_2:

Variable selection for Nonlinear regression models

$$\boldsymbol{y} = \boldsymbol{g}(X, \beta) + \varepsilon, \;\; \varepsilon \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I})$$

It can be done!

$$(\beta_{\mathcal{C}_1}, \sigma_{\mathcal{C}_1}) \to (\beta'_{\mathcal{C}_1}, \sigma'_{\mathcal{C}_1})$$

1. Simulate a training sample y_2: it is easy
2. Simulate the posterior given y_2: MCMC <u>if it is necessary</u>
3. Simulate a training sample y_1:

Variable selection for Nonlinear regression models

$$\boldsymbol{y} = \boldsymbol{g}(X, \beta) + \varepsilon, \ \ \varepsilon \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I})$$

It can be done!

$$(\beta_{\mathcal{C}_1}, \sigma_{\mathcal{C}_1}) \rightarrow (\beta'_{\mathcal{C}_1}, \sigma'_{\mathcal{C}_1})$$

1. Simulate a training sample y_2: it is easy
2. Simulate the posterior given y_2: MCMC <u>if it is necessary</u>
3. Simulate a training sample y_1: it is easy
4. Simulate the posterior given y_1:

Variable selection for Nonlinear regression models

$$\boldsymbol{y} = \boldsymbol{g}(X, \beta) + \varepsilon, \;\; \varepsilon \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I})$$

It can be done!

$$(\beta_{\mathcal{C}_1}, \sigma_{\mathcal{C}_1}) \rightarrow (\beta'_{\mathcal{C}_1}, \sigma'_{\mathcal{C}_1})$$

1. Simulate a training sample y_2: it is easy
2. Simulate the posterior given y_2: MCMC if it is necessary
3. Simulate a training sample y_1: it is easy
4. Simulate the posterior given y_1: MCMC if it is necessary

Conclusions and oncoming research

- The application of integral priors just needs to simulate imaginay minimal training samples z from the involved models, and their posterior distributions $\pi_i^N(\theta_i \mid z)$

Conclusions and oncoming research

- The application of integral priors just needs to simulate imaginay minimal training samples z from the involved models, and their posterior distributions $\pi_i^N(\theta_i \mid z)$
- This methodology can directly be applied to the comparison of nonnested models, that is a common restriction in other methodologies

Conclusions and oncoming research

- The application of integral priors just needs to simulate imaginay minimal training samples z from the involved models, and their posterior distributions $\pi_i^N(\theta_i \mid z)$
- This methodology can directly be applied to the comparison of nonnested models, that is a common restriction in other methodologies
- Integral priors are obtained by simulation, and therefore the predictive distribution and the Bayes factors have not a closed form in general

Conclusions and oncoming research

- The application of integral priors just needs to simulate imaginay minimal training samples z from the involved models, and their posterior distributions $\pi_i^N(\theta_i \mid z)$
- This methodology can directly be applied to the comparison of nonnested models, that is a common restriction in other methodologies
- Integral priors are obtained by simulation, and therefore the predictive distribution and the Bayes factors have not a closed form in general
- Computation of Bayes factors with integral priors is work in progress

Gracias por vuestra atención