Integral priors para la selección de modelos bayesianos

Cano Sánchez, J.A. and Salmerón, D.

Universidad de Murcia. Departamento de Estadística e Investigación Operativa, Spain
Universidad de Murcia. Departamento de Ciencias SocioSanitarias, Spain
CIBER Epidemiología y Salud Pública (CIBERESP)

Madrid - Noviembre 2011

Part I: The problem of bayesian model selection

Part II: Our proposal. Integral priors

Part III: How the methodology operates

The problem of bayesian model selection

Model selection

Components

- We consider two models, M_1 and M_2, to explain the data $\mathbf{x}$.

Model selection

Components

- We consider two models, M_1 and M_2, to explain the data $\mathbf{x}$.

- Under model M_i the data $\mathbf{x}$ are related to the parameter θ_i by a density $f_i(\mathbf{x}|\theta_i)$, $i = 1, 2$.

Model selection

Components

- We consider two models, M_1 and M_2, to explain the data $\mathbf{x}$.
- Under model M_i the data $\mathbf{x}$ are related to the parameter θ_i by a density $f_i(\mathbf{x}|\theta_i)$, $i = 1, 2$.
- If no prior information is available, default priors π_i^N, $i = 1, 2$, are often used for estimation.

Model selection

Components

- We consider two models, M_1 and M_2, to explain the data $\mathbf{x}$.
- Under model M_i the data $\mathbf{x}$ are related to the parameter θ_i by a density $f_i(\mathbf{x}|\theta_i)$, $i = 1, 2$.
- If no prior information is available, default priors π_i^N, $i = 1, 2$, are often used for estimation.
- Default priors for estimation: Jeffreys' prior (1961), reference prior (Bernardo, 1979)

Model selection. Bayes factors.

Indetermination of Bayes factors

Very often default priors are improper: $\pi_i^N(\theta_i) = c_i h_i(\theta_i)$, $i = 1, 2$.
and therefore

Model selection. Bayes factors.

Indetermination of Bayes factors

Very often default priors are improper: $\pi_i^N(\theta_i) = c_i h_i(\theta_i)$, $i = 1, 2$.
and therefore

$$B_{21}^N(\mathbf{x}) = \frac{m_2^N(\mathbf{x})}{m_1^N(\mathbf{x})} = \frac{c_2}{c_1} \frac{\int f_2(\mathbf{x}|\theta_2) h_2(\theta_2) d\theta_2}{\int f_1(\mathbf{x}|\theta_1) h_1(\theta_1) d\theta_1}$$

and

$$P(M_2|\mathbf{x}) = \frac{P(M_2) B_{21}^N(\mathbf{x})}{P(M_1) + P(M_2) B_{21}^N(\mathbf{x})}$$

depend on the arbitrary ratio c_2/c_1.

The problem

- Improper priors produce arbitrary answers
- Proper priors with very large variance (very often used in BUGS) are not a satisfactory solution.
- Nowadays to choose objective priors for Bayesian model selection is an open problem

Our proposal

In this presentation we focus on the development of default (automatic) priors called

Integral priors

for Bayesian model selection.

Our proposal. Integral priors

Other approaches

Intrinsic priors

- Among the many attempts for solving the problem of using improper priors in Bayesian model selection, Berger and Pericchi (1996) introduced the intrinsic priors, later justified by Moreno *et al.* (1998).

Other approaches

Intrinsic priors

- Among the many attempts for solving the problem of using improper priors in Bayesian model selection, Berger and Pericchi (1996) introduced the intrinsic priors, later justified by Moreno *et al.* (1998).

- A particular choice of intrinsic priors has proved to behave well in nested problems (Casella and Moreno, 2006). However the class of intrinsic priors for NONNESTED problems can be very large (Cano *et al.* 2004), and it is not clear enough how to choose a particular solution.

Other approaches

Expected posterior priors (EPP).Pérez and Berger (2002)

- For an arbitrary density $m^*(x)$ for the imaginary trainig sample x

$$\pi_i^*(\theta_i) = \int \pi_i^N(\theta_i|x)m^*(x)dx$$

- A trouble with this approach is the choice of $m^*(x)$.

Other approaches

Some proposal for $m^*(x)$ are:

1. The predictive density derived from a model at least as simple as the others under consideration, however

Other approaches

Some proposal for $m^*(x)$ are:

1. The predictive density derived from a model at least as simple as the others under consideration, however
   - It is difficult to precise when we consider nonnested models
   - It is not guaranted that $\pi_i^*(\theta_i)$ are well defined

Other approaches

Some proposal for $m^*(x)$ are:

1. The predictive density derived from a model at least as simple as the others under consideration, however
   - It is difficult to precise when we consider nonnested models
   - It is not guaranted that $\pi_i^*(\theta_i)$ are well defined

2. The empirical distribution of x based on the observed data, however

Other approaches

Some proposal for $m^*(x)$ are:

1. The predictive density derived from a model at least as simple as the others under consideration, however
   - It is difficult to precise when we consider nonnested models
   - It is not guaranted that $\pi_i^*(\theta_i)$ are well defined
2. The empirical distribution of x based on the observed data, however
   - The resulting priors tend to favour the more complex model
   - For some applications like regression models, the empirical distribution can be an inaccurate approximation

Integral priors

Integral priors

With the aim of solve the problems with the intrinsic priors and the EPP, Cano, Salmerón and Robert (2008) have proposed the integral priors for model selection, defined as the solution to the following system of integral equations

$$\pi_1(\theta_1) = \int \pi_1^N(\theta_1|x)m_2(x)dx$$

$$\pi_2(\theta_2) = \int \pi_2^N(\theta_2|x)m_1(x)dx$$

where $m_i(x) = \int f_i(x|\theta_i)\pi_i(\theta_i)d\theta_i$, $i = 1, 2$, and x is an imaginary training sample.

Integral priors

Integral priors can be seen as generalised expected posterior priors

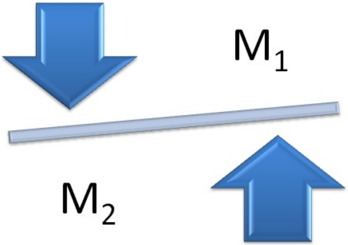
- $\pi_1(\theta_1)$ is the EPP derived from $m^*(x) = m_2(x)$
- $\pi_2(\theta_2)$ is the EPP derived from $m^*(x) = m_1(x)$
- $m_i(x) = \int f_i(x|\theta_i)\pi_i(\theta_i)d\theta_i,\ i = 1, 2$

The method is a symmetrization of the EPP, but it does not requiere any predictive density $m^*(x)$.

Integral priors - Motivation

Being a priori neutral comparing two models

The models M_1 and M_2 are equally valid and provided with ideal unknown priors (the integral priors) that yield true marginals allowing to balance each model with respect to the other one.



Properness of integral priors

Theorem. Proper distributions

If $\pi_1(\theta_1)$ is a proper integral prior, then

$$\pi_2(\theta_2) = \int \pi_2^N(\theta_2|x)m_1(x)dx$$

is a proper prior.

Coherence of integral priors

Theorem. Actual Bayes factor

If $\pi_1(\theta_1)$ and $\pi_2(\theta_2)$ are integral priors and $m_i(\mathbf{x}) < \infty$, $i = 1, 2$, then

$$B_{12}(\mathbf{x}) = m_1(\mathbf{x})/m_2(\mathbf{x})$$

is either an actual Bayes factor or a limit of actual Bayes factors.

Integral priors - existence/uniqueness

Theorem. Asociated Markov chain

Assume that observations and parameters in both models are continuous. If the Markov chain on Θ_1 with transition

$$Q(\theta_1'|\theta_1) = \int g(\theta_1, \theta_1', \theta_2, x, x')dxdx'd\theta_2$$

where

$$g(\theta_1, \theta_1', \theta_2, x, x') = \pi_1^N(\theta_1'|x)f_2(x|\theta_2)\pi_2^N(\theta_2|x')f_1(x'|\theta_1),$$

is recurrent, then there exists a solution $\{\pi_1(\theta_1), \pi_2(\theta_2)\}$ to the integral equations, unique up to a multiplicative constant, and $\pi_1(\theta_1)$ is the invariant measure of the Markov chain.

Integral priors - existence/uniqueness

- When the associated Markov chain is positive recurrent there exists a unique pair of proper integral priors.

Integral priors - existence/uniqueness

- When the associated Markov chain is positive recurrent there exists a unique pair of proper integral priors.
- There exists a parallel Markov chain on Θ_2 with the same properties; if one is (Harris) recurrent then so is the other.

Integral priors - existence/uniqueness

- When the associated Markov chain is positive recurrent there exists a unique pair of proper integral priors.
- There exists a parallel Markov chain on Θ_2 with the same properties; if one is (Harris) recurrent then so is the other.
- This duality property can be found both in the MCMC literature and in the decision theory (Diebolt and Robert, 1992; Eaton, 1992)

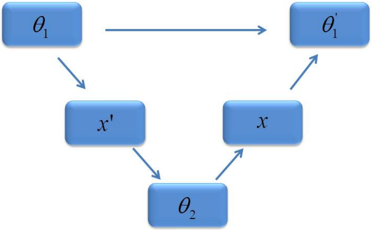
Integral priors - existence/uniqueness

- When the associated Markov chain is positive recurrent there exists a unique pair of proper integral priors.
- There exists a parallel Markov chain on Θ_2 with the same properties; if one is (Harris) recurrent then so is the other.
- This duality property can be found both in the MCMC literature and in the decision theory (Diebolt and Robert, 1992; Eaton, 1992)
- If Harris recurrence holds but the integral priors cannot be obtained, the Bayes factor can be approximated by MCMC simulation.

Simulation of the Markov chain

The transition $\theta_1 \to \theta_1'$ of the associated Markov chain is made of the following four steps

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



Some initial examples

- Point null hypothesis testing
- Location models
- Scale models
- The one way random effects model

Point null hypothesis testing

Testing $H_0 : \theta = \theta^*$ *versus* $H_1 : \theta \neq \theta^*$ is equivalent to consider the models

$$M_1 : f(x|\theta^*) \qquad vs \qquad M_2 : f(x|\theta), \theta \in \Theta$$

Point null hypothesis testing

Testing $H_0 : \theta = \theta^*$ *versus* $H_1 : \theta \neq \theta^*$ is equivalent to consider the models

$$M_1 : f(x|\theta^*) \qquad vs \qquad M_2 : f(x|\theta), \theta \in \Theta$$

The integral priors are $\pi_1(\theta_1) = \delta_{\theta^*}(\theta_1)$ and

$$\pi_2(\theta_2) = \int \pi_2^N(\theta_2|x) f_1(x|\theta^*) dx.$$

(=Intrinsic prios)

Location models - a nonnested case

$$M_1 : f_1(x|\theta_1) = f_1(x-\theta_1), \theta_1 \in \mathbb{R}$$
$$M_2 : f_2(x|\theta_2) = f_2(x-\theta_2), \theta_2 \in \mathbb{R}$$

The initial default priors are $\pi_i^N(\theta_i) = c_i$, $i = 1, 2$ and the minimal trainig sample size is one.

Location models - a nonnested case

$$M_1 : f_1(x|\theta_1) = f_1(x - \theta_1), \theta_1 \in \mathbb{R}$$
$$M_2 : f_2(x|\theta_2) = f_2(x - \theta_2), \theta_2 \in \mathbb{R}$$

The initial default priors are $\pi_i^N(\theta_i) = c_i$, $i = 1, 2$ and the minimal trainig sample size is one.

The priors $\pi_1(\theta_1) = \pi_2(\theta_2) = 1$ are integral priors.

Recurrence: case by case.

Location models - a nonnested case

The normal versus the double exponential model

$$M_1 : N(\theta, 1),\ \theta \in \mathbb{R},\ \pi_1^N(\theta_1) = c_1$$
$$M_2 : DE(\lambda, 1),\ \lambda \in \mathbb{R},\ \pi_2^N(\lambda) = c_2$$

Location models - a nonnested case

The normal versus the double exponential model

$$M_1 : N(\theta, 1),\ \theta \in \mathbb{R},\ \pi_1^N(\theta_1) = c_1$$
$$M_2 : DE(\lambda, 1),\ \lambda \in \mathbb{R},\ \pi_2^N(\lambda) = c_2$$

1. $x' = \theta + \varepsilon_1,\ \varepsilon_1 \sim N(0,1)$
2. $\lambda = x' + \varepsilon_2,\ \varepsilon_2 \sim DE(0,1)$
3. $x = \lambda + \varepsilon_3,\ \varepsilon_3 \sim DE(0,1)$
4. $\theta_1' = x + \varepsilon_4,\ \varepsilon_4 \sim N(0,1)$

Location models - a nonnested case

The normal versus the double exponential model

$$M_1 : N(\theta, 1),\ \theta \in \mathbb{R},\ \pi_1^N(\theta_1) = c_1$$
$$M_2 : DE(\lambda, 1),\ \lambda \in \mathbb{R},\ \pi_2^N(\lambda) = c_2$$

1. $x' = \theta + \varepsilon_1,\ \varepsilon_1 \sim N(0, 1)$
2. $\lambda = x' + \varepsilon_2,\ \varepsilon_2 \sim DE(0, 1)$
3. $x = \lambda + \varepsilon_3,\ \varepsilon_3 \sim DE(0, 1)$
4. $\theta_1' = x + \varepsilon_4,\ \varepsilon_4 \sim N(0, 1)$

Expressing the four moves at one

$$\theta' = \theta + \varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon_4,$$

the Markov chain is a null recurrent random walk, and $\pi_i(\theta_i) = 1$ are the integral priors.

Scale models - a nonnested case

$$M_1 : f_1(x|\sigma_1) = \frac{1}{\sigma_1} f_1(x/\sigma_1), \sigma_1 > 0$$
$$M_2 : f_2(x|\sigma_2) = \frac{1}{\sigma_2} f_2(x/\sigma_2), \sigma_2 > 0$$

The initial default priors are $\pi_i^N(\sigma_i) = c_i/\sigma_i$, $i = 1, 2$ and the minimal trainig sample size is one.

Scale models - a nonnested case

$$M_1 : f_1(x|\sigma_1) = \frac{1}{\sigma_1} f_1(x/\sigma_1), \sigma_1 > 0$$
$$M_2 : f_2(x|\sigma_2) = \frac{1}{\sigma_2} f_2(x/\sigma_2), \sigma_2 > 0$$

The initial default priors are $\pi_i^N(\sigma_i) = c_i/\sigma_i$, $i = 1, 2$ and the minimal trainig sample size is one.

The priors $\pi_1(\sigma_1) = 1/\sigma_1$ and $\pi_2(\sigma_2) = 1/\sigma_2$ are integral priors.

Recurrence: case by case.

Scale models - a nonnested case

The normal versus the double exponential model

$$M_1 : N(0, \sigma_1^2),\ \sigma_1 \in \mathbb{R}^+,\ \pi_1^N(\sigma_1) = c_1/\sigma_1$$
$$M_2 : DE(0, \sigma_2),\ \sigma_2 \in \mathbb{R}^+,\ \pi_2^N(\sigma_2) = c_2/\sigma_2$$

Scale models - a nonnested case

The normal versus the double exponential model

$$M_1 : N(0,\sigma_1^2),\ \sigma_1 \in \mathbb{R}^+,\ \pi_1^N(\sigma_1) = c_1/\sigma_1$$
$$M_2 : DE(0,\sigma_2),\ \sigma_2 \in \mathbb{R}^+,\ \pi_2^N(\sigma_2) = c_2/\sigma_2$$

1. $x' = \sigma_1\varepsilon_1,\ \varepsilon_1 \sim N(0,1)$
2. $\sigma_2 = |x'|/\varepsilon_2,\ \varepsilon_2 \sim Exp(1)$
3. $x = \sigma_2\varepsilon_3,\ \varepsilon_3 \sim DE(0,1)$
4. $\sigma_1' = |x|/\varepsilon_4,\ \varepsilon_4 \sim N(0,1)$

Scale models - a nonnested case

The normal versus the double exponential model

$$M_1 : N(0, \sigma_1^2),\ \sigma_1 \in \mathbb{R}^+,\ \pi_1^N(\sigma_1) = c_1/\sigma_1$$
$$M_2 : DE(0, \sigma_2),\ \sigma_2 \in \mathbb{R}^+,\ \pi_2^N(\sigma_2) = c_2/\sigma_2$$

1. $x' = \sigma_1 \varepsilon_1,\ \varepsilon_1 \sim N(0,1)$
2. $\sigma_2 = |x'|/\varepsilon_2,\ \varepsilon_2 \sim Exp(1)$
3. $x = \sigma_2 \varepsilon_3,\ \varepsilon_3 \sim DE(0,1)$
4. $\sigma_1' = |x|/\varepsilon_4,\ \varepsilon_4 \sim N(0,1)$

Expressing the four moves at one

$$\sigma_1' = \sigma_1 \frac{|\varepsilon_1 \varepsilon_3|}{\varepsilon_2 |\varepsilon_4|},$$

the Markov chain is a null recurrent random walk in $\log \sigma_1$, and $\pi_i(\sigma_i) = 1/\sigma_i$ are the integral priors.

The one way random effects model

We consider the model

$$y_{ij} = \mu + a_i + e_{ij},\ i = 1, \ldots, k;\ j = 1, \ldots, n,$$

where $e_{ij} \sim N(0, \sigma^2)$ and $a_i \sim N(0, \sigma_a^2)$ are independent.

The one way random effects model

We consider the model

$$y_{ij} = \mu + a_i + e_{ij},\ i = 1, ..., k;\ j = 1, ..., n,$$

where $e_{ij} \sim N(0, \sigma^2)$ and $a_i \sim N(0, \sigma_a^2)$ are independent.

We are interested in the selection problem between the models with parameters

$$M_1 : \theta_1 = (\mu_1, \sigma_1, 0) \text{ and } M_2 : \theta_2 = (\mu_2, \sigma_2, \sigma_a)$$

The one way random effects model

We consider the model

$$y_{ij} = \mu + a_i + e_{ij},\ i = 1, ..., k;\ j = 1, ..., n,$$

where $e_{ij} \sim N(0, \sigma^2)$ and $a_i \sim N(0, \sigma_a^2)$ are independent.

We are interested in the selection problem between the models with parameters

$$M_1 : \theta_1 = (\mu_1, \sigma_1, 0) \text{ and } M_2 : \theta_2 = (\mu_2, \sigma_2, \sigma_a)$$

$\pi_1^N(\theta_1) = c_1/\sigma_1$ and $\pi_2^N(\theta_2) = c_2 \sigma_2^{-2} (1 + (\sigma_a/\sigma_2)^2)^{-3/2}$

The one way random effects model - the Markov chain

The transition $\theta_1 \to \theta_1'$ of the Markov chain associated with the integral priors for this example, can be written as

$$\begin{aligned} \mu_1' &= \mu_1 + \sigma_1 \alpha \\ \sigma_1' &= \sigma_1 \beta \end{aligned}$$

where α and β are random variables with a complex distribution but easy to simulate.

The one way random effects model - the Markov chain

Proposition

1. The reference priors $\pi_1(\theta_1) = 1/\sigma_1$ and

$$\pi_2(\theta_2) = \sigma_2^{-2}(1 + (\sigma_a/\sigma_2)^2)^{-3/2}$$

are integral priors.

The one way random effects model - the Markov chain

Proposition

1. The reference priors $\pi_1(\theta_1) = 1/\sigma_1$ and
$$\pi_2(\theta_2) = \sigma_2^{-2}(1 + (\sigma_a/\sigma_2)^2)^{-3/2}$$
are integral priors.

2. If $g(\theta_1) = \phi(\sigma_1)$ is an invariant measure for the Markov chain on Θ_1, then $g(\theta_1) = \pi_1(\theta_1) = 1/\sigma_1$ (up to a multiplicative constant).

How the methodology operates

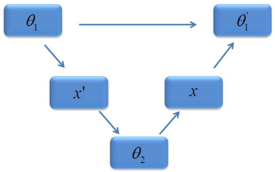
Integral priors from simulation

- For the above examples, integral priors have been found explicitly and most of the times they were the initial default priors after the adjustment of the constants c_i.

Integral priors from simulation

- For the above examples, integral priors have been found explicitly and most of the times they were the initial default priors after the adjustment of the constants c_i.

- However, when we are not able to find the integral priors we can use the simulation of the associated Markov chain



to approximate the Bayes factor.

- A toy example. Testing a normal mean with known variance
- A not so toy example. One-sided testing for the exponential distribution

- Constrained imaginary training samples $\Rightarrow$ Existence and uniqueness of proper integral priors.
    - Testing a normal mean with unknown variance using constrained imaginary training samples
- Testing in Binomial regression models

A toy example. Testing a normal mean with known variance

With this example we explain how our methodology works.

A toy example. Testing a normal mean with known variance

With this example we explain how our methodology works.

Suppose that $\mathbf{x} = (x_1, ..., x_m)$ is a random sample form $N(\theta, \sigma^2)$, with σ known.

We consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.

The integral priors are $\pi_1(\theta_1) = \delta_{\theta_0}(\theta_1)$ and $\pi_2(\theta_2) = N(\theta_0, 2\sigma^2)$.

A toy example. Testing a normal mean with known variance

The transition of the Markov chain on Θ_2 now is made of two steps

1. $x' = \theta_0 + \varepsilon_1$, $\varepsilon_1 \sim N(0, \sigma^2)$
2. $\theta_2' = x' + \varepsilon_2$, $\varepsilon_2 \sim N(0, \sigma^2)$

A toy example. Testing a normal mean with known variance

The transition of the Markov chain on Θ_2 now is made of two steps

1. $x' = \theta_0 + \varepsilon_1$, $\varepsilon_1 \sim N(0, \sigma^2)$
2. $\theta_2' = x' + \varepsilon_2$, $\varepsilon_2 \sim N(0, \sigma^2)$

The Bayes factor $B_{21}(\bar{\mathbf{x}})$ is

$$\frac{1}{\sqrt{2m+1}} \exp\left(\frac{m^2(\bar{\mathbf{x}} - \theta_0)^2}{(2m+1)\sigma^2}\right)$$

On the other hand, we can simulate the Markov chain (θ_2^t) and

$$B_{21}(\bar{\mathbf{x}}) \approx \frac{\sum_{t=1}^{L} f(\bar{\mathbf{x}}|\theta_2^t)/L}{f(\bar{\mathbf{x}}|\theta_0)}$$

A toy example. Testing a normal mean with known variance

$\theta_0 = 0$
$m = 1, 5, 10, 20, 30, 50$
$\sigma = 1, 2, 3$

A toy example. Testing a normal mean with known variance

$\theta_0 = 0$
$m = 1, 5, 10, 20, 30, 50$
$\sigma = 1, 2, 3$

We have generated samples of size m from $N(\theta, \sigma^2)$, ranging θ from -1 to 1 step equal to 0.005.

Exact and approximate posterior probabilities (Markov chain with length 10000)

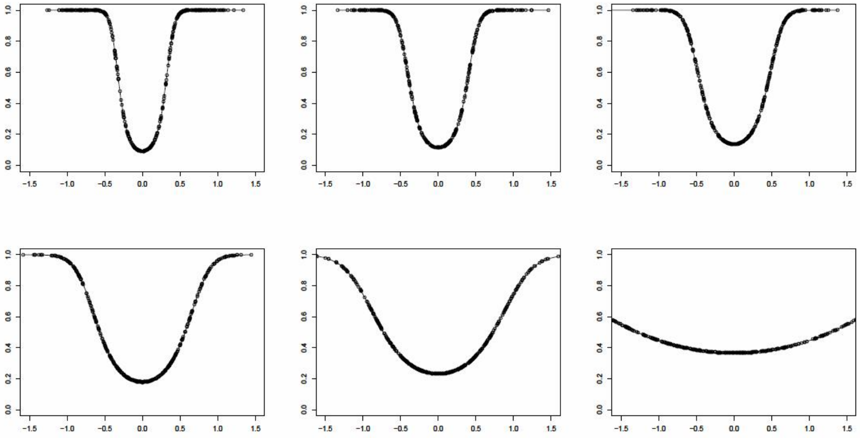


Figure 1: Approximate (solid) and exact (dotted) probabilities of the complex model for $\sigma = 1$ and several values of m.

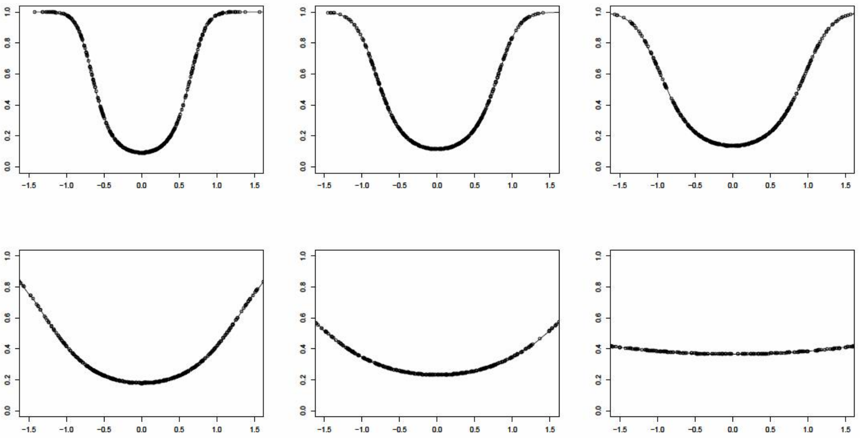


Figure 2: Approximate (solid) and exact (dotted) probabilities of the complex model for $\sigma = 2$ and several values of m.

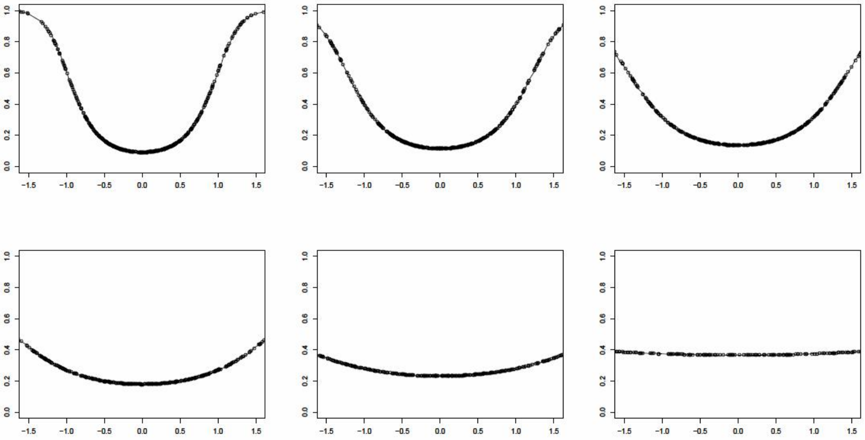


Figure 3: Approximate (solid) and exact (dotted) probabilities of the complex model for $\sigma = 3$ and several values of m.

A not so toy example.
One-sided testing for the exponential distribution

Let $\mathbf{x} = (x_1, ..., x_m)$ be a sample from the exponential distribution with mean θ.

$$H_0 : \theta \in (0, 1) \quad vs \quad H_1 : \theta > 1$$

A not so toy example.
One-sided testing for the exponential distribution

Let $\mathbf{x} = (x_1, ..., x_m)$ be a sample from the exponential distribution with mean θ.

$$H_0 : \theta \in (0, 1) \quad vs \quad H_1 : \theta > 1$$

$$M_1 : f_1(x|\theta_1) = \frac{1}{\theta_1} \exp(-x/\theta_1),\ \pi_1(\theta_1) = \frac{c_1}{\theta_1} 1_{(0,1)}(\theta_1)$$
$$M_2 : f_2(x|\theta_2) = \frac{1}{\theta_2} \exp(-x/\theta_2),\ \pi_2(\theta_2) = \frac{c_2}{\theta_2} 1_{(1,+\infty)}(\theta_2)$$

A not so toy example.
One-sided testing for the exponential distribution

- No intrinsic priors exist for this problem
- The typical encompassing approach does not give an actual Bayes factor
- Moreno (2005) has proposed an alternative solution
- The methodology of the integral priors works

A not so toy example.
One-sided testing for the exponential distribution

Integral priors - Markov chain

The transition of the asocciated Markov chain is made of the following steps:

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. $\theta_1' = (1 - \frac{1}{x} log\ u_4)^{-1}$

where u_i are i.i.d $\sim U(0, 1)$

A not so toy example.
One-sided testing for the exponential distribution

- The transition density is bounded

$$
\begin{aligned}
Q(\theta_1'|\theta_1) &\geq \int \pi_1^N(\theta_1'|x) f_2(x|\theta_2) \left(\int \theta_2^{-2} e^{-x'/\theta_2} f_1(x'|\theta_1) dx' \right) dx d\theta_2 \\
&= \int \pi_1^N(\theta_1'|x) f_2(x|\theta_2) \frac{1}{\theta_2(\theta_1+\theta_2)} dx d\theta_2 \\
&\geq \int \frac{\pi_1^N(\theta_1'|x) f_2(x|\theta_2)}{\theta_2(1+\theta_2)} dx d\theta_2 =: q(\theta_1')
\end{aligned}
$$

where $0 < \int_0^1 q(\theta_1') d\theta_1' \leq 1$.

A not so toy example.
One-sided testing for the exponential distribution

- The transition density is bounded

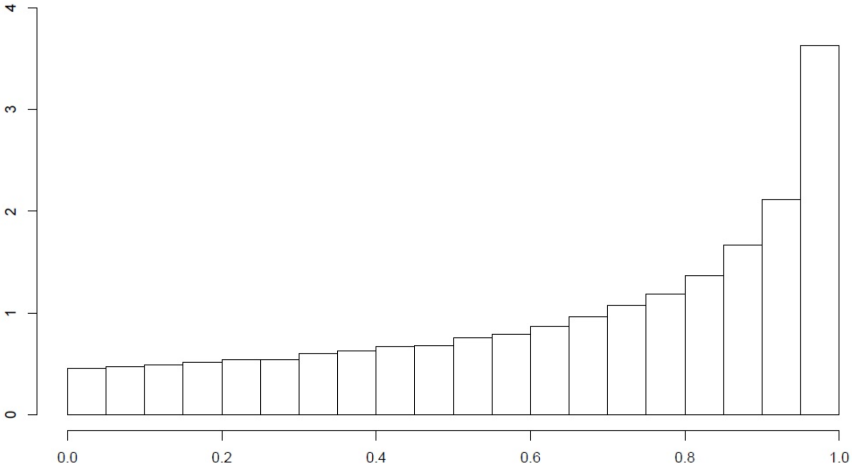
$$\begin{aligned} Q(\theta_1'|\theta_1) &\geq \int \pi_1^N(\theta_1'|x) f_2(x|\theta_2) \left(\int \theta_2^{-2} e^{-x'/\theta_2} f_1(x'|\theta_1) dx' \right) dx d\theta_2 \\ &= \int \pi_1^N(\theta_1'|x) f_2(x|\theta_2) \frac{1}{\theta_2(\theta_1 + \theta_2)} dx d\theta_2 \\ &\geq \int \frac{\pi_1^N(\theta_1'|x) f_2(x|\theta_2)}{\theta_2(1+\theta_2)} dx d\theta_2 =: q(\theta_1') \end{aligned}$$

  where $0 < \int_0^1 q(\theta_1') d\theta_1' \leq 1$.

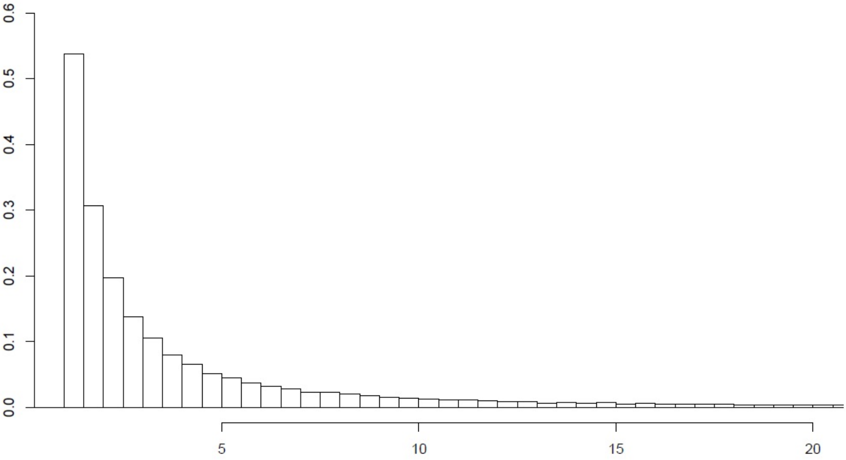
- Therefore the Markov chain satisfies the Doeblin condition $\Rightarrow$ integral priors are unique and proper priors.

- Integral priors can be obtained simulating the Markov chain.

Histogram of $\pi_1(\theta_1)$ by simulation of the Markov chain



Histogram of $\pi_2(\theta_2)$ by simulation of the Markov chain



Posterior probability of the null: Integral priors - Intrinsic priors (Moreno (2005))

$\bar{x}$	$m = 5$	$m = 10$	$m = 15$	$m = 20$
0.1	1.00, 1.00	1.00, 1.00	1.00, 1.00	1.00, 1.00
0.6	0.83, 0.81	0.92, 0.91	0.96, 0.96	0.98, 0.98
1.4	0.39, 0.19	0.30, 0.12	0.23, 0.07	0.17, 0.05
1.9	0.15, 0.05	0.04, 0.01	0.01, 0.00	0.00, 0.00

Table 3: Posterior probability of the null hypothesis using the integral priors (left) and the intrinsic priors proposed in Moreno (2005) (right).

Constrained imaginary trainig samples

Constrained imaginary trainig samples

- The recurrence of the associated Markov chain is of fundamental importance for the application of integral priors

Constrained imaginary trainig samples

- The recurrence of the associated Markov chain is of fundamental importance for the application of integral priors
- However it can be difficult to asses for complex models

Constrained imaginary trainig samples

- The recurrence of the associated Markov chain is of fundamental importance for the application of integral priors
- However it can be difficult to asses for complex models
- We propose using a constrain on the imaginary training samples space to ensure that the associated Markov chain is positive recurrent, and therefore the existence and the uniqueness of proper integral priors.

Constrained imaginary trainig samples

Let A be a subset of the imaginary training samples space.

Constrained imaginary trainig samples

Let A be a subset of the imaginary training samples space.
The constrain is applied in steps 1 and 3 of the transition $\theta_1 \to \theta_1'$.

Constrained imaginary trainig samples

Let A be a subset of the imaginary training samples space.
The constrain is applied in steps 1 and 3 of the transition $\theta_1 \to \theta_1'$.

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

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

Constrained imaginary trainig samples

Let A be a subset of the imaginary training samples space.
The constrain is applied in steps 1 and 3 of the transition $\theta_1 \to \theta_1'$.

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

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

The idea behind this is that the constrain on the imaginary training samples prevents the Markov chain from escaping to infinity and therefore guarantees the existence and the uniqueness of an invariant probability measure

Theorem

If the set A is chosen such that the function

$$K_A(x|x^*) = \mathbb{I}_A(x^*) \int f_1^A(x|\tilde{\theta}_1)\pi_1^N(\tilde{\theta}_1|\tilde{x}) f_2^A(\tilde{x}|\theta_2')\pi_2^N(\theta_2'|x^*) d\theta_2' d\tilde{x} d\tilde{\theta}_1$$

satisfies the minorizing condition $K_A(x|x^*) \geq g_A(x)$, for some function $g_A(x)$ with $\beta = \int g_A(x)dx > 0$, then there exists a unique invariant probability for the Markov chain with imaginary training samples space A

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

Corollary

If A is a compact set and the model M_2 is regular enough to satisfy

$$\inf\{\pi_2^N(\theta_2'|x_1') : x_1' \in A\} > 0 \ \forall \theta_2',$$

then there exists a unique invariant probability for the Markov chain with imaginary training samples space A.

If A is a compact set and the model M_1 is regular enough to satisfy

$$\inf\{\pi_1^N(\theta_1'|x_2') : x_2' \in A\} > 0 \ \forall \theta_1',$$

then there exists a unique invariant probability for the Markov chain with imaginary training samples space A.

Testing a normal mean with unknown variance using constrained imaginary training samples

Suppose the data $\mathbf{x}$ are i.i.d. $N(\mu, \sigma^2)$ and we consider testing $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$. A Bayesian setting for this problem is that of choosing between the models

$$M_1 : N(\mathbf{x}|\mathbf{0}, \sigma_1^2\mathbf{I})$$

and

$$M_2 : N(\mathbf{x}|\mu_2\mathbf{1}, \sigma_2^2\mathbf{I}).$$

Testing a normal mean with unknown variance using constrained imaginary training samples

Suppose the data $\mathbf{x}$ are i.i.d. $N(\mu, \sigma^2)$ and we consider testing $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$. A Bayesian setting for this problem is that of choosing between the models

$$M_1 : N(\mathbf{x}|\mathbf{0}, \sigma_1^2 \mathbf{I})$$

and

$$M_2 : N(\mathbf{x}|\mu_2 \mathbf{1}, \sigma_2^2 \mathbf{I}).$$

Here a reasonable choice for the compact set is

$$A = \{(x_1, x_2) \in \mathbb{R}^2 : \ |x_1| \leq b, \ |x_2| \leq b\}$$

with $b > 0$.

The Markov chain with imaginary training samples space A

1. x_i is simulated from the density proportional to $N(x_i|0,\sigma_1^2)\mathbb{I}_{[-b,b]}(x_i)$, $i = 1, 2$, that is, a truncated normal density.

The Markov chain with imaginary training samples space A

1. x_i is simulated from the density proportional to $N(x_i|0, \sigma_1^2)\mathbb{I}_{[-b,b]}(x_i)$, $i = 1, 2$, that is, a truncated normal density.
2. $$\sigma_2^2 = \frac{\overline{x^2} - \overline{x}^2}{v} \text{ and } \mu_2 \sim N(\overline{x}, \sigma_2^2/2),$$

   with v simulated from a gamma density with shape $1/2$ and scale 1.

The Markov chain with imaginary training samples space A

1. x_i is simulated from the density proportional to $N(x_i|0, \sigma_1^2)\mathbb{I}_{[-b,b]}(x_i)$, $i = 1, 2$, that is, a truncated normal density.

2. $$\sigma_2^2 = \frac{\overline{x^2} - \overline{x}^2}{v} \text{ and } \mu_2 \sim N(\overline{x}, \sigma_2^2/2),$$

   with v simulated from a gamma density with shape $1/2$ and scale 1.

3. x_i' is simulated from the density proportional to $N(x_i'|\mu_2, \sigma_2^2)\mathbb{I}_{[-b,b]}(x_i')$, $i = 1, 2$.

The Markov chain with imaginary training samples space A

1. x_i is simulated from the density proportional to $N(x_i|0, \sigma_1^2)\mathbb{I}_{[-b,b]}(x_i)$, $i = 1, 2$, that is, a truncated normal density.
2. $$\sigma_2^2 = \frac{\overline{x^2} - \bar{x}^2}{v} \text{ and } \mu_2 \sim N(\bar{x}, \sigma_2^2/2),$$
   with v simulated from a gamma density with shape $1/2$ and scale 1.
3. x_i' is simulated from the density proportional to $N(x_i'|\mu_2, \sigma_2^2)\mathbb{I}_{[-b,b]}(x_i')$, $i = 1, 2$.
4. $\sigma_1' = \sqrt{\frac{x_1'^2 + x_2'^2}{2w}}$, where $w \sim Exp(1)$.

- For a sample size of $n = 10$ we approximate the Bayes factor $B_{12}^{A}(\overline{\mathbf{x}}, \overline{\mathbf{x}^2})$.
- The imaginary training samples spaces A we are used are the ones defined for $b = 10, 25, 50$ and 100, respectively.
- The results are based on 100000 transitions of the associated Markov chain.
- We compare our results with the ones obtained using intrinsic priors.

$\overline{\mathbf{x}^2}$	$\overline{\mathbf{x}}$	b=10	b=25	b=50	b=100	Intrinsic	$\overline{x} \pm 3\widehat{\sigma}$
1	0	0.814	0.809	0.817	0.812	0.789	(-3.2,3.2)
	0.2	0.786	0.782	0.788	0.785	0.757	(-2.9,3.3)
	0.4	0.675	0.672	0.677	0.676	0.635	(-2.5,3.3)
	0.6	0.395	0.398	0.397	0.401	0.351	(-1.9,3.1)
	0.8	0.058	0.058	0.056	0.058	0.049	(-1.1,2.7)
	1	0.000	0.000	0.000	0.000	0.000	(1.0,1.0)
10	0	0.828	0.820	0.810	0.809	0.789	(-10.0,10.0)
	0.2	0.826	0.816	0.807	0.806	0.786	(-9.8,10.2)
	0.4	0.818	0.808	0.798	0.798	0.777	(-9.5,10.3)
	0.6	0.804	0.793	0.783	0.784	0.761	(-9.2,10.4)
	0.8	0.783	0.770	0.761	0.762	0.736	(-8.9,10.5)
	1	0.752	0.737	0.728	0.731	0.701	(-8.5,10.5)

Table: Posterior probabilities of the simple model for differents values of $\overline{x}$, $\overline{x^2}$ and b and for the intrinsic priors

$\overline{\mathbf{x}^2}$	$\overline{\mathbf{x}}$	b=10	b=25	b=50	b=100	Intrinsic	$\overline{x} \pm 3\widehat{\sigma}$
50	0	0.806	0.826	0.814	0.807	0.789	(-22.4,22.4)
	0.2	0.806	0.826	0.813	0.806	0.788	(-22.2,22.6)
	0.4	0.804	0.825	0.811	0.804	0.786	(-21.9,22.7)
	0.6	0.802	0.822	0.809	0.801	0.783	(-21.7,22.9)
	0.8	0.798	0.819	0.805	0.797	0.779	(-21.4,23.0)
	1	0.793	0.814	0.799	0.792	0.774	(-21.1,23.1)

Table: Posterior probabilities of the simple model for differents values of $\overline{x}$, $\overline{x^2}$ and b and for the intrinsic priors

Testing in binomial regression models

Testing in binomial regression models

- Binomial regression models are used very often to investigate associations and risks in epidemiological studies where the goal is to asses the effect of specific exposure factors.

- We apply our methodology to binomial regression models

- Logistic regression (link=logit) is one of the main techniques in analytical epidemilogy, but other link functions are possible (probit, complementary log-log, Cauchit).

literature

- The literature on objective prior distributions, we mean automatic or near it, for testing in binomial regression models is very limited.
- Intrinsic priors for binomial regression models with a general link function has not been developed.
- Leon-Novelo *et al.* (2011) have applied the intrinsic priors to the problem of variable selection in the probit regression model using the relation between the probit model and the normal regression model.
- Integral priors can be directly applied to other link functions.
- Sabanés and Held (2011) have developed an extension of the Zellner's g-prior for generalized linear models, however this extension need the specification of the hyperprior distribution for g.

The model

Notation

Suppose $\{(y_i, x_i); i = 1, \ldots, n\}$ are independent observations
$y_i \sim Ber(p_i)$, $x_i = (x_{i1}, \ldots, x_{ik})$ vector of covariates

The model

Notation

Suppose $\{(y_i, x_i); i = 1, \ldots, n\}$ are independent observations
$y_i \sim Ber(p_i)$, $x_i = (x_{i1}, \ldots, x_{ik})$ vector of covariates
X the matrix with rows $x_1, \ldots, x_n$

The model

Notation

Suppose $\{(y_i, x_i); i = 1, ..., n\}$ are independent observations
$y_i \sim Ber(p_i)$, $x_i = (x_{i1}, ..., x_{ik})$ vector of covariates
X the matrix with rows $x_1, ..., x_n$
Link: $g(p_i) = x_i\beta$, $i = 1, ..., n$

The model

Notation

Suppose $\{(y_i, x_i); i = 1, ..., n\}$ are independent observations
$y_i \sim Ber(p_i)$, $x_i = (x_{i1}, ..., x_{ik})$ vector of covariates
X the matrix with rows $x_1, ..., x_n$
Link: $g(p_i) = x_i\beta$, $i = 1, ..., n$
$\beta = (\beta_1, ..., \beta_k)^T \in \Theta$ the vector of regression coefficients

The model

Notation

Suppose $\{(y_i, x_i); i = 1, ..., n\}$ are independent observations
$y_i \sim Ber(p_i)$, $x_i = (x_{i1}, ..., x_{ik})$ vector of covariates
X the matrix with rows $x_1, ..., x_n$
Link: $g(p_i) = x_i\beta$, $i = 1, ..., n$
$\beta = (\beta_1, ..., \beta_k)^T \in \Theta$ the vector of regression coefficients

Testing: for a fix given value k_0 we consider the hypothesis testing

$$H_0 : \beta_1 = ... = \beta_{k_0} = 0$$
$$H_1 : \exists\, k^* \in \{1, ..., k_0\} \,\text{such that}\, \beta_{k^*} \neq 0$$

As a model selection problem

$$
\begin{aligned}
M_1: \quad & y_i|x_i,\theta_1 \sim Ber(p_i),\ g(p_i) = x_i\theta_1,\ i = 1,...,n \\
& \theta_1 = (\theta_{11},...,\theta_{1k})^T \in \Theta_1 \subseteq \mathbb{R}^k,\ \theta_{1j} = 0\, \forall j = 1,...,k_0
\end{aligned}
$$

$$
\begin{aligned}
M_2: \quad & y_i|x_i,\theta_2 \sim Ber(p_i),\ g(p_i) = x_i\theta_2,\ i = 1,...,n \\
& \theta_2 = (\theta_{21},...,\theta_{2k})^T \in \Theta_2 \subseteq \mathbb{R}^k
\end{aligned}
$$

θ_1 and θ_2 are vectors of dimension k. The numbers of unknown parameters is $k - k_0$ in model M_1 and k in model M_2

Integral priors

Some integrals involved in the definition of integral priors become sums. The transition of the associated Markov chain is

$$Q(\theta_2'|\theta_2) = \sum_{z_1,z_2} \int \pi_2^N(\theta_2'|z_2) f_1(z_2|\theta_1) \pi_1^N(\theta_1|z_1) f_2(z_1|\theta_2) d\theta_1$$

$$= \sum_{z_1,z_2} \pi_2^N(\theta_2'|z_2) H(z_2|z_1) f_2(z_1|\theta_2),$$

where $H(z_2|z_1) = \int f_1(z_2|\theta_1)\pi_1^N(\theta_1|z_1)d\theta_1$

Integral priors

The function $H(z_2|z_1)$ reaches its minimum in some point (z_1^*, z_2^*). Moreover $H(z_2^*|z_1^*) > 0$ since $H(z_2^*|z_1^*) = 0$ yields

$$\int f_1(z_2^*|\theta_1)\pi_1^N(\theta_1|z_1^*)d\theta_1 = 0.$$

Therefore

$$Q(\theta_2'|\theta_2) \geq H(z_2^*|z_1^*)\sum_{z_2}\pi_2^N(\theta_2'|z_2)\sum_{z_1} f_2(z_1|\theta_2)$$

$$= H(z_2^*|z_1^*)\sum_{z_2}\pi_2^N(\theta_2'|z_2),$$

which means that the Doeblin condition is satisfied and the Markov chain has a unique invariant distribution that can be obtained by simulation.

Imaginary trainig sample

Transition $\theta_2 \to \theta_2'$

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

To generate the Markov chain associated with the integral priors two things are required

- generate imaginary training samples
- simulate from the corresponding posteriors

Imaginary trainig sample

Training samples are subsets of the data such that the corresponding posteriors are proper.

Imaginary trainig sample

Training samples are subsets of the data such that the corresponding posteriors are proper.

If $\tilde{y} = (\tilde{y}_1, ..., \tilde{y}_k)$ is a subset of the data and the submatrix $\tilde{X}$ with rows $\tilde{x}_1, ..., \tilde{x}_k$ of X associated to $\tilde{y}$ is of full rank, then Jeffreys prior , $\pi^N(\beta|\tilde{X})$, and the corresponding posterior, $\pi^N(\beta|\tilde{y}, \tilde{X})$, are proper (Ibrahim and Laud (1991)).

Imaginary trainig sample

Training samples are subsets of the data such that the corresponding posteriors are proper.

If $\tilde{y} = (\tilde{y}_1, ..., \tilde{y}_k)$ is a subset of the data and the submatrix $\tilde{X}$ with rows $\tilde{x}_1, ..., \tilde{x}_k$ of X associated to $\tilde{y}$ is of full rank, then Jeffreys prior , $\pi^N(\beta|\tilde{X})$, and the corresponding posterior, $\pi^N(\beta|\tilde{y}, \tilde{X})$, are proper (Ibrahim and Laud (1991)).

The dimension of the training samples will be $k_1 = k - k_0$ and k, respectively, and the corresponding $\tilde{X}$ should be of full rank.

Imaginary trainig sample

- Casella and Moreno (2009), Berger and Pericchi (2004), Consonni *et al.* (2011), among others, applying intrinsic priors , have found convenient to increase the size of the imaginary training samples when the data come from a binomial distribution.

Imaginary trainig sample

- Casella and Moreno (2009), Berger and Pericchi (2004), Consonni *et al.* (2011), among others, applying intrinsic priors , have found convenient to increase the size of the imaginary training samples when the data come from a binomial distribution.

- One way to achieve this in the case of binomial regression models is to take more than a Bernoulli variable $\tilde{y}_i$ for each row $\tilde{x}_i$.

Imaginary trainig sample

- Casella and Moreno (2009), Berger and Pericchi (2004), Consonni *et al.* (2011), among others, applying intrinsic priors , have found convenient to increase the size of the imaginary training samples when the data come from a binomial distribution.

- One way to achieve this in the case of binomial regression models is to take more than a Bernoulli variable $\tilde{y}_i$ for each row $\tilde{x}_i$.

- We propose that the number of Bernoulli variables be a discrete uniform random variable between 1 and the number $N(x)$ of times that each row x is repeated in the matrix X

Imaginary trainig sample

- Casella and Moreno (2009), Berger and Pericchi (2004), Consonni *et al*. (2011), among others, applying intrinsic priors , have found convenient to increase the size of the imaginary training samples when the data come from a binomial distribution.

- One way to achieve this in the case of binomial regression models is to take more than a Bernoulli variable $\tilde{y}_i$ for each row $\tilde{x}_i$.

- We propose that the number of Bernoulli variables be a discrete uniform random variable between 1 and the number $N(x)$ of times that each row x is repeated in the matrix X

- When a covariate is continuous, we can work with a discretized version to compute $N(x)$. Note that discretization of a continuous variable is a very common strategy.

Algorithm to run the Markov chain

Step 1. Simulation of z_1.

- Randomly select $k_1 = k - k_0$ rows of the matrix X: $\tilde{x}_1, \ldots, \tilde{x}_{k_1}$, with the condition that if R_1 is the submatrix of X with these rows, then $|R_2| \neq 0$ where R_2 is the submatrix of R_1 with the columns $k_0 + 1, \ldots, k$.

Algorithm to run the Markov chain

Step 1. Simulation of z_1.

- Randomly select $k_1 = k - k_0$ rows of the matrix X: $\tilde{x}_1, ..., \tilde{x}_{k_1}$, with the condition that if R_1 is the submatrix of X with these rows, then $|R_2| \neq 0$ where R_2 is the submatrix of R_1 with the columns $k_0 + 1, ..., k$.

- Simulate $q_i \sim U\{1, ..., N_1(\tilde{x}_i)\}$, $i = 1, ..., k_1$, where $N_1(\tilde{x}_i)$ is the number of times that the vector with the columns $k_0 + 1, ..., k$ of $\tilde{x}_i$ appears in the design matrix of model M_1.

Algorithm to run the Markov chain

Step 1. Simulation of z_1.

- Randomly select $k_1 = k - k_0$ rows of the matrix X: $\tilde{x}_1, \dots, \tilde{x}_{k_1}$, with the condition that if R_1 is the submatrix of X with these rows, then $|R_2| \neq 0$ where R_2 is the submatrix of R_1 with the columns $k_0 + 1, \dots, k$.
- Simulate $q_i \sim U\{1, \dots, N_1(\tilde{x}_i)\}$, $i = 1, \dots, k_1$, where $N_1(\tilde{x}_i)$ is the number of times that the vector with the columns $k_0 + 1, \dots, k$ of $\tilde{x}_i$ appears in the design matrix of model M_1.
- Independently simulate $\tilde{y}_i^t \sim Ber(g^{-1}(\tilde{x}_i \theta_2))$, $t = 1, \dots, q_i$, $i = 1, \dots, k_1$, and take $z_1 = (\tilde{y}_1, \dots, \tilde{y}_{k_1})$ where $\tilde{y}_i = (\tilde{y}_i^1, \dots, \tilde{y}_i^{q_i})$.

Algorithm to run the Markov chain

Step 2. Simulation of θ_1.

Simulate $\tilde{p}_i \sim Beta\left(\tilde{p}_i | q_i \overline{\tilde{y}_i} + 1/2, q_i\left(1 - \overline{\tilde{y}_i}\right) + 1/2\right)$, $i = 1, ..., k_1$, and compute

$$v = R_2^{-1}(g(\tilde{p}_1), ..., g(\tilde{p}_{k_1}))^T.$$

Take $\theta_1 = (0, ..., 0, v^T)^T$.

Algorithm to run the Markov chain

Step 3. Simulation of z_2.

- Randomly select k rows of the matrix X: $\tilde{x}_1, ..., \tilde{x}_k$, with the condition that if S is is the submatrix of X with these rows, then $|S| \neq 0$.

Algorithm to run the Markov chain

Step 3. Simulation of z_2.

- Randomly select k rows of the matrix X: $\tilde{x}_1, ..., \tilde{x}_k$, with the condition that if S is is the submatrix of X with these rows, then $|S| \neq 0$.

- Simulate $q_i \sim U\{1, ..., N_2(\tilde{x}_i)\}$, $i = 1, ..., k$, where $N_2(\tilde{x}_i)$ is the number of times that $\tilde{x}_i$ appears in the design matrix of model M_2.

Algorithm to run the Markov chain

Step 3. Simulation of z_2.

- Randomly select k rows of the matrix X: $\tilde{x}_1, ..., \tilde{x}_k$, with the condition that if S is is the submatrix of X with these rows, then $|S| \neq 0$.

- Simulate $q_i \sim U\{1, ..., N_2(\tilde{x}_i)\}$, $i = 1, ..., k$, where $N_2(\tilde{x}_i)$ is the number of times that $\tilde{x}_i$ appears in the design matrix of model M_2.

- Independently simulate $\tilde{y}_i^t \sim Ber(g^{-1}(\tilde{x}_i \theta_1))$, $t = 1, ..., q_i$, $i = 1, ..., k$, and take $z_2 = (\tilde{y}_1, ..., \tilde{y}_k)$ where $\tilde{y}_i = (\tilde{y}_i^1, ..., \tilde{y}_i^{q_i})$.

Algorithm to run the Markov chain

Step 4. Simulation of θ_2'.

Simulate $\tilde{p}_i \sim Beta\left(\tilde{p}_i | q_i \overline{\tilde{y}_i} + 1/2, q_i\left(1 - \overline{\tilde{y}_i}\right) + 1/2\right)$, $i = 1, ..., k$, and compute

$$v = S^{-1}(g(\tilde{p}_1), ..., g(\tilde{p}_k))^T.$$

Take $\theta_2' = v$.

Computing the integral Bayes factor

- To compute the Bayes factor $B_{21}(\mathbf{y})$ associated to the integral priors we can use the simulation of the Markov chain.

- Actually with this procedure we obtain two parallel Markov chains $(\theta_1^t)_t$ and $(\theta_2^t)_t$, with stationary probability distributions the integral priors.
  Then
  $$\lim_{T\to\infty} \frac{\sum_{t=1}^{T} f_2(\mathbf{y}|\theta_2^t)}{\sum_{t=1}^{T} f_1(\mathbf{y}|\theta_1^t)} = B_{21}(\mathbf{y})$$
  and this result can be used to compute the Bayes factor.

Computing the integral Bayes factor

- The major difficulty with this approach is that when the likelihood is in conflict with the integral prior, most of the simulations θ_i^t will have small likelihood values, which means that the approximation procedure can be inefficient.

Computing the integral Bayes factor

- The major difficulty with this approach is that when the likelihood is in conflict with the integral prior, most of the simulations θ_i^t will have small likelihood values, which means that the approximation procedure can be inefficient.

- This problem can be solved by importance sampling and a nonparametric density estimation of the integral priors

Computing the integral Bayes factor

- Concretely, if $\hat{\pi}_i(\theta_i)$ is a nonparametric density estimation of $\pi_i(\theta_i)$, and $G_i(\theta_i)$ is a normal approximation to the posteriori density, then

$$\int f_i(\mathbf{y}|\theta_i)\pi_i(\theta_i)d\theta_i \approx \int \frac{f_i(\mathbf{y}|\theta_i)\hat{\pi}_i(\theta_i)}{G_i(\theta_i)} G_i(\theta_i)d\theta_i.$$

Computing the integral Bayes factor

- Concretely, if $\hat{\pi}_i(\theta_i)$ is a nonparametric density estimation of $\pi_i(\theta_i)$, and $G_i(\theta_i)$ is a normal approximation to the posteriori density, then

$$\int f_i(\mathbf{y}|\theta_i)\pi_i(\theta_i)d\theta_i \approx \int \frac{f_i(\mathbf{y}|\theta_i)\hat{\pi}_i(\theta_i)}{G_i(\theta_i)} G_i(\theta_i)d\theta_i.$$

- Simulating $G_i(\theta_i)$ and evaluating $f_i(\mathbf{y}|\theta_i)$, $\hat{\pi}_i(\theta_i)$ and $G_i(\theta_i)$, we can approximate the Bayes factor.

Example. Breast cancer mortality. Logistic regression.

Table 3 presents data on the relation of receptor level and stage to survival in a cohort of women with breast cancer.

Stage	ReceptorLevel	Deaths	Total
1	1	2	12
1	2	5	55
2	1	9	22
2	2	17	74
3	1	12	14
3	2	9	15

Table: Data relating receptor level and stage to 5-year breast cancer mortality.

Example. Breast cancer mortality. Logistic regression.

Table 3 presents data on the relation of receptor level and stage to survival in a cohort of women with breast cancer.

Stage	ReceptorLevel	Deaths	Total
1	1	2	12
1	2	5	55
2	1	9	22
2	2	17	74
3	1	12	14
3	2	9	15

Table: Data relating receptor level and stage to 5-year breast cancer mortality.

First, we are going to compare the model with the intercept and the stage *versus* the full model. From the classical logistic regression perspective we find an association between the receptor level and mortality, with 2.51 as the estimation for the *OR* and a p-value of 0.02.

- $P(M_2|\mathbf{y})$: importance sampling based on the normal distribution centered at the maximum likelihood estimator $\hat{\theta}_i$ and covariance $2\hat{V}_i$ where $\hat{V}_i$ is the estimated covariance of $\hat{\theta}_i$.

- $P(M_2|\mathbf{y})$: importance sampling based on the normal distribution centered at the maximum likelihood estimator $\hat{\theta}_i$ and covariance $2\hat{V}_i$ where $\hat{V}_i$ is the estimated covariance of $\hat{\theta}_i$.
- $\pi_1(\theta_1)$ and $\pi_2(\theta_2)$: simulation of the Markov chain and kernel density estimation. For the values $T = 1000$, 5000 and 10000 we have run 50 Markov chains of length T and the importance sampling has been carried out with T simulations too.

- $P(M_2|\mathbf{y})$: importance sampling based on the normal distribution centered at the maximum likelihood estimator $\hat{\theta}_i$ and covariance $2\hat{V}_i$ where $\hat{V}_i$ is the estimated covariance of $\hat{\theta}_i$.
- $\pi_1(\theta_1)$ and $\pi_2(\theta_2)$: simulation of the Markov chain and kernel density estimation. For the values $T = 1000$, 5000 and 10000 we have run 50 Markov chains of length T and the importance sampling has been carried out with T simulations too.
- The mean and the standard deviation of the 50 estimations of $P(M_2|\mathbf{y})$ appears in table 4.

	$T = 1000$	$T = 5000$	$T = 10000$
Mean	0.710	0.722	0.726
SD	(0.020)	(0.010)	(0.008)

Table: Estimations of the posterior probability of the model M_2 running 50 Markov chains of length T and importance sampling based on T simulations.

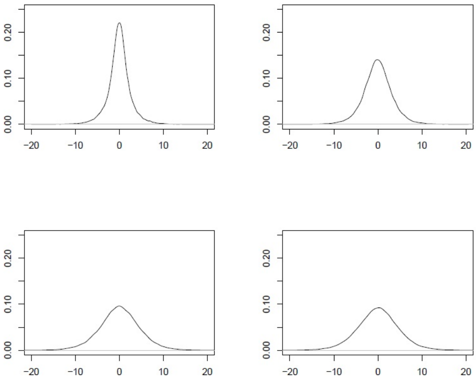


Figure: Integral priors obtained based on 50000 iterations of the associated Markov chain

In the first row there are the priors for the coefficient of the receptor level and the intercept; the second row corresponds to the stage.

Example. Low birth weight. Logistic regression.

The birthwt data frame has 189 rows and 10 columns (see the object birthwt from the statistical software R).

Data were collected at the Baystate Medical Center, Springfield, Massachusetts during 1986 to attempt to identify which factors contributed to an increased risk of low birth weight infants.

Information was recorded for 189 women of whom 59 had low birth weight infants.

Example. Low birth weight. Logistic regression.

We have studied the association between the low birth weight and smoking (two levels), race (three levels), previous premature labours (two levels) and age (five levels, defined taking as included the endpoints, 18, 20 25, and 30, respectively).

We have considered as the reduced model the one without the variable smoking. The p-value associated to smoking is 0.014 ($OR = 2.62$).

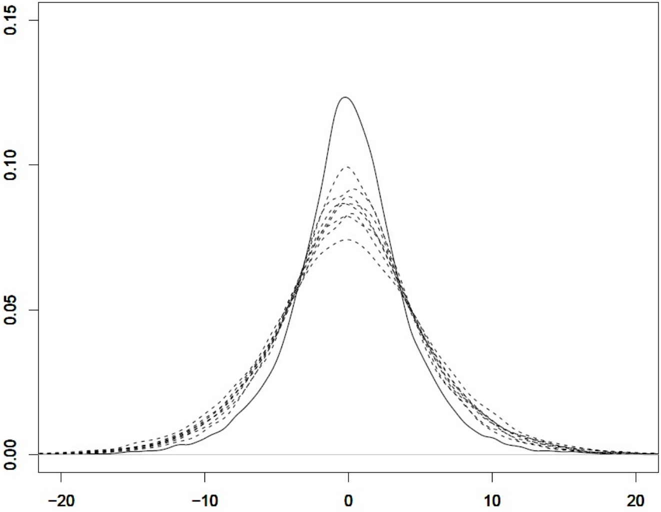
Example. Low birth weight. Logistic regression.

The Bayesian results are based on 30000 iterations of the Markov chain and 10000 simulations for the importance sampling

The posterior probability of smoking having effect over the low birth weight was 0.67

In the next figure appear the integral prior distributions of the 9 regression coefficients. The integral priors of all regression coefficients under model M_2 are very similar except the one for the smoking coefficient, this prior is more concentrated about the null hypothesis.

Example. Low birth weight. Logistic regression.



Thank you very much
for your attention