Integral equation solutions as prior distributions for model selection

[1]Cano, J. A., Robert, [2]C. P. and [1]Salmerón D.
[1]University of Murcia. [2]Université Paris Dauphine.

June 21, 2005

Abstract

In many statistical problems we deal with more than one model. When the prior information on the parameters in the models is vague default priors are typically used. Unfortunately, these priors are usually improper provoking a calibration problem which precludes the comparison of the models. An attempt for solving this difficulty consists in using intrinsic priors introduced in Berger and Pericchi (1996) instead of the original default priors; however, there are situations where the class of intrinsic priors is too large.

Because of this we propose as prior distributions for model selection the solutions of a system of integral equations which is derived to calibrate the initial default priors. Under some assumptions our integral equations yield a unique solution. Some illustrative examples are provided.

Key words: Bayes factor, model selection, integral equations, intrinsic priors, expected posterior priors.

1 Introduction

The use of subjective prior distributions has largely been a major critic to Bayesian analysis. Because of this several objective o default methods have been proposed for both estimation and model selection problems. Among the most commonly used are Jeffreys (1961), Bernardo (1979) and Berger and Pericchi (1996); the first two for estimation problems and the last one for model selection problems.

Default methods usually yield an improper prior distribution $\pi^N(\theta) \propto h(\theta)$, where $h(\theta)$ is a function whose integral diverges. Therefore, $\pi^N(\theta)$ is determined up to a positive multiplicative constant c, that is $\pi^N(\theta) = c\, h(\theta)$. This is not a serious issue in estimation problems where the posterior is defined by the formal Bayes rule and therefore does not depend on c and it is usually justified with easy limiting arguments. On the other hand, in model selection problems we have two models M_i, $i = 1, 2$, the data $\mathbf{x}$ are related to the parameter θ_i by a density $f_i(\mathbf{x}|\theta_i)$, the default priors are $\pi_i^N(\theta_i) = c_i\, h_i(\theta_i)$, $i = 1, 2$, and the Bayes factor,

$$B_{21}^N(\mathbf{x}) = \frac{m_2^N(\mathbf{x})}{m_1^N(\mathbf{x})} = \frac{c_2 \int_{\Theta_2} f_2(\mathbf{x}|\theta_2) h_2(\theta_2) d\theta_2}{c_1 \int_{\Theta_1} f_1(\mathbf{x}|\theta_1) h_1(\theta_1) d\theta_1},$$

depends on the arbitrary ratio c_2/c_1. Therefore, we are left with two problems, first the determination of the ratio c_2/c_1, second the Bayes factor with $\pi_i^N(\theta_i)$ is not an actual Bayes factor and inference methods based on proper priors are preferable to those that are not (see Principle 1 in Berger and Pericchi, 1996). An attempt for solving these problems (Berger and Pericchi, 1996), consists in using intrinsic priors π_1^I and π_2^I which are the solutions to a system of two functional equations. Intrinsic priors provide a Bayes factor free of arbitrary constants but whether or not it is an actual Bayes factor or a limit of actual Bayes factors is a problem to be discussed (Berger and Pericchi, 1996).

An additional difficulty in considering intrinsic priors is that they might be not unique. In the case where model one is nested in model two, the system of functional equations reduces to a single equation with two *incognita* so that it is apparent that the solution is not unique. On the other hand, in the nonnested case the class of intrinsic priors may be very large too, for instance in Cano, Kessler and Moreno (2004) it is shown that when comparing the double exponential versus the normal location models any prior is intrinsic and the same prior has to be used for the location parameter of the two models. However, in the general nonnested case it is not clear enough how to choose a particular solution so that a robustness problem is likely to be present.

In this paper we put forward as prior distributions for model selection the generalized expected posterior priors, see Pérez and Berger (2002) for an introduction to this concept, which solve a system of integral equations. In section 2 we deal with the problem of the indetermination of the ratio c_2/c_1 introducing the above mentioned priors to eliminate it. Moreover, we prove that these priors provide an actual Bayes factor or a limit of actual Bayes factors and under some assumptions they are unique. Section 3 provides some illustrative examples and the conclusions are stated in section 4.

2 Integral equation solutions as priors for model selection problems

When we compare two models the main goal is to eliminate the indetermination c_2/c_1 occasioned by the use of noninformative priors. For it, we propose as prior distributions π_1 and π_2 to compare two models the unknown expected posterior priors which solve the system of integral equations

$$\pi_1(\theta_1) = \int_{\mathcal{X}} \pi_1^N(\theta_1 \mid x) m_2(x) dx \tag{1}$$

and

$$\pi_2(\theta_2) = \int_{\mathcal{X}} \pi_2^N(\theta_2 \mid x) m_1(x) dx, \tag{2}$$

where x is an imaginary minimal training sample and $m_i(x) = \int_{\Theta_i} f_i(x \mid \theta_i)\pi_i(\theta_i)d\theta_i$, $i = 1, 2$. We emphasize that in this system $\pi_i(\theta_i)$, $i = 1, 2$, are the *incognita*.

The argument to propose these equations is that *a priori* we suppose that the two models are equally valid and provided with ideal unknown priors $\pi_i(\theta_i)$, $i = 1, 2$, that yield true marginals allowing to balance each model with respect to the other one since the prior $\pi_i(\theta_i)$ is derived from the marginal $m_j(x)$, $j \neq i$, as an unknown expected posterior prior. This is nothing but the idea of being *a priori* neutral comparing the two models.

Setting $\pi_i^N(\theta_i) = c_i h_i(\theta_i)$, after simple algebraic computations equations (1) and (2) are converted into

$$\pi_1(\theta_1) = h_1(\theta_1) \int_{\Theta_2} \pi_2(\theta_2) g_1(\theta_1, \theta_2) d\theta_2 \tag{3}$$

and

$$\pi_2(\theta_2) = h_2(\theta_2) \int_{\Theta_1} \pi_1(\theta_1) g_2(\theta_1, \theta_2) d\theta_1 \tag{4}$$

where

$$g_i(\theta_1, \theta_2) = \int_{\mathcal{X}} \frac{f_1(x \mid \theta_1) f_2(x \mid \theta_2)}{m_i^{h_i}(x)} dx$$

and

$$m_i^{h_i}(x) = \int_{\Theta_i} f_i(x \mid \theta_i) h_i(\theta_i) d\theta_i$$

for $i = 1, 2$. Equations (3) and (4) together constitute a system of two integral equations with two incognita $\pi_1(\theta_1)$ and $\pi_2(\theta_2)$, which easily can be converted into a single integral equation for $\pi_1(\theta_1)$ plus an equation expressing $\pi_2(\theta_2)$ as a function of $\pi_1(\theta_1)$ or viceversa.

However, to look for the solutions of our integral equations, which from now on will be called integral priors, we go back to equations (1) and (2). First, let us state some basic properties on the way (1) and (2) operate.

Proposition 2.1 *If $h_1(\theta_1)$ is a probability density then equation (2) with $\pi_1(\theta_1) = h_1(\theta_1)$ provides a probability density $\pi_2(\theta_2)$. Alternatively, if $h_2(\theta_2)$ is a probability density then equation (1) provides a probability density $\pi_1(\theta_1)$.*

Proof. Let us prove the first statement, the second is similar. From (2) and Fubini's theorem

$$\int_{\Theta_2} \pi_2(\theta_2)d\theta_2 = \int_{\Theta_2}\int_{\mathcal{X}} \frac{f_2(x \mid \theta_2)h_2(\theta_2)}{\int_{\Theta_2} f_2(x \mid \theta_2)h_2(\theta_2)d\theta_2} \int_{\Theta_1} f_1(x \mid \theta_1)h_1(\theta_1)d\theta_1 dx d\theta_2 =$$

$$\int_{\Theta_1} h_1(\theta_1)d\theta_1 \int_{\mathcal{X}} f_1(x \mid \theta_1)dx \int_{\Theta_2} \frac{f_2(x \mid \theta_2)h_2(\theta_2)}{\int_{\Theta_2} f_2(x \mid \theta_2)h_2(\theta_2)d\theta_2} d\theta_2 = 1,$$

which proves the proposition.

An easy consequence of this proposition is that if $\pi_1(\theta_1)$ and $\pi_2(\theta_2)$ are integral priors then both of them are proper and integrate up to the same constant or both of them are improper.

Second, if $\pi_1(\theta_1)$ and $\pi_2(\theta_2)$ are a pair of solutions to the system they provide a Bayes factor. Concretely, we have

Proposition 2.2 *If $\pi_1(\theta_1)$ and $\pi_2(\theta_2)$ are integral priors and their marginals are finite, then its relative Bayes factor $B_{1,2}(\mathbf{x})$ is either an actual Bayes factor or a limit of actual Bayes factors.*

Proof. Suppose that the integral priors are not proper densities otherwise $B_{1,2}(\mathbf{x})$ is an actual Bayes factor. Consider a sequence of compact subsets $\Theta_1^l \subseteq \Theta_1$ such that $\Theta_1^l \nearrow \Theta_1$ as $l \to +\infty$ and define the following sequences of proper densities

$$\pi_{1,l}(\theta_1) = \frac{\pi_1(\theta_1)}{A_l} 1_{\Theta_1^l}(\theta_1)$$

and

$$\pi_{2,l}(\theta_2) = \int \pi_2^N(\theta_2 \mid x)\, m_{1,l}(x)\, dx$$

where $A_l = \int_{\Theta_1^l} \pi_1(\theta_1)\, d\theta_1$ and $m_{1,l}(x) = \int f_1(x \mid \theta_1)\, \pi_{1,l}(\theta_1)\, d\theta_1$. The sequence of actual Bayes factors using $\pi_{1,l}(\theta_1)$ and $\pi_{2,l}(\theta_2)$ as priors is given by

$$B_{1,2}^l(\mathbf{x}) = \frac{\int f_1(\mathbf{x} \mid \theta_1)\, \pi_{1,l}(\theta_1)\, d\theta_1}{\int f_2(\mathbf{x} \mid \theta_2)\, \pi_{2,l}(\theta_2)\, d\theta_2} = \frac{A_l^{-1} \int_{\Theta_1^l} f_1(\mathbf{x} \mid \theta_1)\, \pi_1(\theta_1)\, d\theta_1}{\int f_2(\mathbf{x} \mid \theta_2)\, \pi_2^N(\theta_2 \mid x)\, m_{1,l}(x)\, dx d\theta_2} =$$

$$\frac{\int_{\Theta_1^l} f_1(\mathbf{x} \mid \theta_1)\, \pi_1(\theta_1)\, d\theta_1}{\int f_2(\mathbf{x} \mid \theta_2)\, \pi_2^N(\theta_2 \mid x)\, f_1(x \mid \theta_1)\, \pi_1(\theta_1)\, 1_{\Theta_1^l}(\theta_1)\, d\theta_1 dx d\theta_2},$$

which by dominated convergence goes to

$$\frac{\int f_1(\mathbf{x} \mid \theta_1)\, \pi_1(\theta_1)\, d\theta_1}{\int f_2(\mathbf{x} \mid \theta_2)\, \pi_2^N(\theta_2 \mid x)\, f_1(x \mid \theta_1)\, \pi_1(\theta_1)\, d\theta_1 dx d\theta_2} =$$

$$\frac{\int f_1(\mathbf{x} \mid \theta_1)\, \pi_1(\theta_1)\, d\theta_1}{\int f_2(\mathbf{x} \mid \theta_2)\, \pi_2(\theta_2)\, d\theta_2}.$$

Third, the question of the existence and uniqueness of solutions to the integral equations has to be studied case by case. However, next we provide a general result.

Theorem 2.3 *Let us consider the continuous case, that is when the observations and the parameters in both models are continuous. Assume that the Markov chain with transition density $Q(\theta_1' \mid \theta_1)$ defined below is recurrent then there is a solution to the integral equations and it is unique up to a multiplicative constant.*

Proof. Consider the function

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

which is defined for $\theta_1, \theta_1' \in \Theta_1$, $\theta_2 \in \Theta_2$, and $x, x' \in \mathcal{X}$.

For a fix but arbitrary point $\theta_1 \in \Theta_1$ the function

$$(\theta_1', \theta_2, x, x') \mapsto g(\theta_1, \theta_1', \theta_2, x, x'),$$

is a continuous probability density, therefore

$$Q(\theta_1' \mid \theta_1) = \int g(\theta_1, \theta_1', \theta_2, x, x')\, dx dx' d\theta_2$$

is a transition density.

Because of the assumed recurrence there exists an invariant σ-finite measure $p\left(d\theta_1\right)$ satisfying

$$\int_A p\left(d\theta_1'\right) = \int_{\Theta_1} \left(\int_A Q\left(\theta_1' \mid \theta_1\right) d\theta_1' \right) p\left(d\theta_1\right),$$

for each measurable set A. Therefore $p\left(d\theta_1'\right)$ is absolutely continuous with respect to the Lebesgue measure. Let $\pi_1\left(\theta_1'\right) > 0$ denote the Radon-Nikodym derivative of $p\left(d\theta_1'\right)$ with respect to the Lebesgue measure, then

$$\pi_1\left(\theta_1'\right) = \int_{\Theta_1} Q\left(\theta_1' \mid \theta_1\right) \pi_1\left(\theta_1\right) d\theta_1 =$$

$$\int \pi_1^N\left(\theta_1' \mid x\right) f_2\left(x \mid \theta_2\right) \pi_2^N\left(\theta_2 \mid x'\right) f_1\left(x' \mid \theta_1\right) \pi_1\left(\theta_1\right) dx dx' d\theta_2 d\theta_1,$$

and taking $\pi_2\left(\theta_2\right) = \int \pi_2^N\left(\theta_2 \mid x'\right) \left(\int f_1\left(x' \mid \theta_1\right) \pi_1\left(\theta_1\right) d\theta_1 \right) dx'$, we conclude that $\pi_1\left(\theta_1\right)$ and $\pi_2\left(\theta_2\right)$ are integral priors.

On the other hand, let $\pi_1\left(\theta_1\right)$ and $\pi_2\left(\theta_2\right)$ be integral priors, then

$$\pi_1\left(\theta_1'\right) = \int_{\Theta_1} Q\left(\theta_1' \mid \theta_1\right) \pi_1\left(\theta_1\right) d\theta_1, \tag{5}$$

therefore $\pi_1\left(\theta_1\right) d\theta_1$ is an invariant σ-finite measure for $Q\left(\theta_1' \mid \theta_1\right)$ and because of the recurrence of the Markov chain $\pi_1\left(\theta_1\right)$ is the unique solution to (5) up to a multiplicative factor from where $\pi_1\left(\theta_1\right)$ and $\pi_2\left(\theta_2\right)$ are the unique integral priors up to a multiplicative factor.

By construction, the transition density $Q\left(\theta_1' \mid \theta_1\right)$ is essentially a transition density for the parameter θ_2 too. Therefore, there exists a parallel Markov chain having the same properties as the above defined; in particular, if one is (Harris) recurrent also is the other. Moreover, in the cases where recurrence is present but we are not able to find the unique pair of integral priors if the property of Harris recurrence is satisfied their corresponding Bayes factor can be approximated by using the ergodic theorem.

3 Examples

3.1 Point null hypothesis testing

Consider the point null hypothesis testing $H_0 : \theta = \theta^*$ versus $H_1 : \theta \neq \theta^*$, which is equivalent to consider the models

$$\begin{aligned} M_1 &: f\left(x \mid \theta^*\right), \\ M_2 &: f\left(x \mid \theta\right), \theta \in \Theta. \end{aligned}$$

The default priors are $\pi_1^N(\theta) = \delta_{\theta^*}(\theta)$ and $\pi_2^N(\theta) = \pi^N(\theta)$. Let x be a minimal training sample and consider the proper priors $\pi_1(\theta) = \delta_{\theta^*}(\theta)$ and $\pi_2(\theta) = \int \pi^N(\theta \mid x) f(x \mid \theta^*) dx$ for which the marginal densities are $m_1(x) = f(x \mid \theta^*)$ and $m_2(x) = \int f(x \mid \theta) \pi^N(\theta \mid y) f(y \mid \theta^*) dyd\theta$. It follows that

$$\int \pi_1^N(\theta \mid x) m_2(x) dx = \delta_{\theta^*}(\theta) \int m_2(x) dx = \delta_{\theta^*}(\theta) = \pi_1(\theta)$$

and

$$\int \pi_2^N(\theta \mid x) m_1(x) dx = \int \pi^N(\theta \mid x) f(x \mid \theta^*) dx = \pi_2(\theta)$$

and therefore $\pi_1(\theta)$ and $\pi_2(\theta)$ are integral priors. In this case the uniqueness of the Bayes factor is easily obtained. Let us suppose that $\tilde{\pi}_1(\theta)$ and $\tilde{\pi}_2(\theta)$ are integral priors. The prior $\tilde{\pi}_1(\theta)$ has to be $k\delta_{\theta^*}(\theta)$ for some $k > 0$ so that $\tilde{m}_1(x) = kf(x \mid \theta^*)$ and

$$\tilde{\pi}_2(\theta) = \int \pi_2^N(\theta \mid x) \tilde{m}_1(x) dx = k\pi_2(\theta)$$

and then $\{\tilde{\pi}_i(\theta)\}$ and $\{\pi_i(\theta)\}$ produce the same Bayes factor.

Remark 3.1 *In this case, integral priors and intrinsic priors coincide and are unique up to a multiplicative constant (see Moreno, Bertolino and Racugno, 1998).*

3.2 Location models

Consider two location models

$$\begin{aligned} M_1 &: f_1(x \mid \theta_1) = f_1(x - \theta_1), \theta_1 \in \mathbb{R}, \\ M_2 &: f_2(x \mid \theta_2) = f_2(x - \theta_2), \theta_2 \in \mathbb{R}. \end{aligned}$$

The default priors are $\pi_i^N(\theta_i) = c_i$, $i = 1, 2$, the minimal training sample size is one, and the marginal densities are $m_i^N(x) = c_i$, $i = 1, 2$. In this case $\pi_1^N(\theta_1)$ and $\pi_2^N(\theta_2)$ are integral priors when $c_1 = c_2$ since

$$\int \pi_1^N(\theta_1 \mid x) m_2^N(x) dx = \int c_2 f_1(x - \theta_1) dx = c_2$$

and

$$\int \pi_2^N(\theta_2 \mid x) m_1^N(x) dx = \int c_1 f_2(x - \theta_2) dx = c_1.$$

In fact, if the associated Markov chain is recurrent it follows that $\pi_1^N(\theta_1) = \pi_2^N(\theta_2) = c$ are the unique integral priors up to a multiplicative constant. On the other hand, in Cano, Kessler and Moreno (2004) is proved that these integral priors are intrinsic priors although the class of intrinsic priors could be very large.

3.3 Scale models

Consider two scale models

$$
\begin{aligned}
M_1 &: \; f_1(x \mid \sigma_1) = \frac{1}{\sigma_1} f_1\left(\frac{x}{\sigma_1}\right), \sigma_1 > 0, \\
M_2 &: \; f_2(x \mid \sigma_2) = \frac{1}{\sigma_2} f_2\left(\frac{x}{\sigma_2}\right), \sigma_2 > 0.
\end{aligned}
$$

The default priors are $\pi_i^N(\sigma_i) = \frac{c_i}{\sigma_i}$, $i = 1, 2$, the minimal training sample size is one and the marginal densities are $m_i^N(x) = \frac{c_i \alpha_i(x)}{|x|}$, $i = 1, 2$, where

$$
\alpha_i(x) = \begin{cases} \beta_i = \int_0^\infty f_i(y)\, dy, & \text{if} x > 0, \\ 1 - \beta_i, & \text{if} x < 0. \end{cases}
$$

The priors $\pi_1^N(\sigma_1)$ and $\pi_2^N(\sigma_2)$ are integral priors when $c_1 = c_2$, since

$$
\int \pi_1^N(\sigma_1 \mid x)\, m_2^N(x)\, dx = \int \frac{\sigma_1^{-2} f_1\left(\frac{x}{\sigma_1}\right)}{\alpha_1(x)} c_2 \alpha_2(x)\, dx =
$$

$$
c_2 \sigma_1^{-2} \left(\frac{1-\beta_2}{1-\beta_1} \int_{-\infty}^{0} f_1\left(\frac{x}{\sigma_1}\right) dx + \frac{\beta_2}{\beta_1} \int_0^{+\infty} f_1\left(\frac{x}{\sigma_1}\right) dx \right) =
$$

$$
c_2 \sigma_1^{-2} \left(\frac{1-\beta_2}{1-\beta_1} \sigma_1 (1-\beta_1) + \frac{\beta_2}{\beta_1} \sigma_1 \beta_1 \right) = c_2 \sigma_1^{-1}
$$

and

$$
\int \pi_2^N(\sigma_2 \mid x)\, m_1^N(x)\, dx = c_1 \sigma_2^{-1}.
$$

Also in this case if the associated Markov chain is recurrent it follows that $\pi_1^N(\sigma_1)$ and $\pi_2^N(\sigma_2)$ with $c_1 = c_2$ are the unique integral priors up to a multiplicative constant. On the other hand, in Cano, Kessler and Moreno (2004) is proved that when $\beta_1 = \beta_2$, these integral priors are intrinsic priors although there exist an infinite number of continuous intrinsic priors. Moreover, when

$$
\frac{\beta_1^2}{\beta_2} + \frac{(1-\beta_1)^2}{1-\beta_2} \neq 1,
$$

the prior $\pi_1^N(\sigma_1)$ is no longer an intrinsic prior.

3.4 Location-scale models

Consider two location-scale models

$$\begin{aligned} M_1 &: f_1(x \mid \mu_1, \sigma_1) = \frac{1}{\sigma_1} f_1\left(\frac{x-\mu_1}{\sigma_1}\right), \mu_1 \in \mathbb{R}, \sigma_1 > 0, \\ M_2 &: f_2(x \mid \mu_2, \sigma_2) = \frac{1}{\sigma_2} f_2\left(\frac{x-\mu_2}{\sigma_2}\right), \mu_2 \in \mathbb{R}, \sigma_2 > 0. \end{aligned}$$

The default priors are $\pi_i^N(\mu_i, \sigma_i) = \frac{c_i}{\sigma_i}$, $i = 1, 2$, the minimal training sample size is two and the marginal densities are $m_i^N(x_1, x_2) = \frac{c_i}{2|x_1 - x_2|}$, $i = 1, 2$. In this case $\pi_1^N(\mu_1, \sigma_1)$ and $\pi_2^N(\mu_2, \sigma_2)$ are integral priors when $c_1 = c_2$ since

$$\begin{aligned} \int \pi_1^N(\mu_1, \sigma_1 \mid x_1, x_2) m_2^N(x_1, x_2) dx_1 dx_2 &= \\ \int \pi_1^N(\mu_1, \sigma_1 \mid x_1, x_2) m_1^N(x_1, x_2) dx_1 dx_2 &= \pi_1^N(\mu_1, \sigma_1) \end{aligned}$$

and

$$\int \pi_2^N(\mu_2, \sigma_2 \mid x_1, x_2) m_1^N(x_1, x_2) dx_1 dx_2 = \pi_2^N(\mu_2, \sigma_2).$$

Again assuming recurrence we obtain that the priors $\pi_1^N(\mu_1, \sigma_1)$ and $\pi_2^N(\mu_2, \sigma_2)$ with $c_1 = c_2$ are the unique integral priors up to a multiplicative constant. On the other hand in Cano, Kessler and Moreno (2004) is proved that the above integral priors are intrinsic priors, but these are not necessarily unique.

In the general examples of subsections 3.2 to 3.4 stating the recurrence of the associated Markov chain yields the uniqueness of the integral priors up to a multiplicative constant. Of course, the recurrence has to be studied case by case. Finally, in the next subsection we present some specific examples where recurrence is found.

3.5 Some specific recurrent examples

3.5.1 The normal versus the double exponential model (location)

Consider the location models

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

The minimal training sample size is one, the posterior densities are

$$\pi_1^N (\theta \mid x) = N(x, 1) \text{ and } \pi_2^N (\lambda \mid x) = DE(x, 1),$$

and the transition $\theta \to \theta'$ of the associated Markov chain is thus made of the following steps :

1. $x' = \theta + \varepsilon_1$, $\varepsilon_1 \sim N(0, 1)$
2. $\lambda = x' + \varepsilon_2$, $\varepsilon_2 \sim DE(0, 1)$
3. $x = \lambda +$ ε_3, $\varepsilon_3 \sim DE(0, 1)$
4. $\theta' = x + \varepsilon_4$, $\varepsilon_4 \sim N(0, 1)$

Then, if we express the four moves at once, θ' given θ satisfies

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

and the transition is therefore a random walk in θ (it can also be checked that this is equally a random walk in λ). Therefore, the white noise in $\theta' = \theta + \varepsilon$ has a finite second moment. It follows from Meyn and Tweedie (1993) that the random walk is (null) recurrent. The resulting Lebesgue measures are thus invariant for θ and λ and $\pi_1(\theta) = 1$ and $\pi_2(\lambda) = 1$ are the unique solutions to the system of integral equations up to a multiplicative constant.

3.5.2 The normal versus the double exponential model (scale)

Consider the scale models

$$\begin{aligned} M_1 &: N(0, \sigma_1^2), \sigma_1 \in \mathbb{R}^+, \pi_1^N(\sigma_1) = \frac{c_1}{\sigma_1}, \\ M_2 &: DE(0, \sigma_2), \sigma_2 \in \mathbb{R}^+, \pi_2^N(\sigma_2) = \frac{c_2}{\sigma_2}. \end{aligned}$$

The minimal training sample size is one and the posterior densities are

$$\pi_1^N (\sigma_1 \mid x) \propto \frac{1}{\sigma_1^2} \exp\left(-\frac{x^2}{2\sigma_1^2}\right) \text{ and } \pi_2^N (\sigma_2 \mid x) \propto \frac{1}{\sigma_2^2} \exp\left(-\frac{|x|}{\sigma_2}\right),$$

which is equivalent to

$$\sigma_1^{-1} \mid x \sim HN^+\left(0, \frac{1}{x^2}\right) \text{ and } \sigma_2^{-1} \mid x \sim Exp(|x|),$$

where HN^+ denotes the half-normal distribution. The transition $\sigma_1 \to \sigma_1'$ of the associated Markov chain is thus made of the following steps:

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

Then, if we express the four moves at once, σ_1' given σ_1 satisfies

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

and the transition is therefore a random walk in $\log \sigma_1$ (it can also be checked that this is equally a random walk in $\log \sigma_2$) and again this random walk is (null) recurrent. The resulting Lebesgue measures are thus invariant for $\log \sigma_1$ and $\log \sigma_2$ and $\pi_1(\sigma_1) = 1/\sigma_1$ and $\pi_2(\sigma_2) = 1/\sigma_2$ are the unique solutions to the system of integral equations up to a multiplicative constant.

3.5.3 The normal versus the log-exponential model

Consider the models

$$\begin{aligned} M_1 &: N(\theta,1), \theta \in \mathbb{R}, \pi_1^N(\theta) = c_1, \\ M_2 &: \exp(X) \sim Exp(1/\lambda), \lambda \in \mathbb{R}^+, \pi_2^N(\lambda) = \frac{c_2}{\lambda}. \end{aligned}$$

The minimal training sample size is one and the posterior densities are

$$\pi_1^N(\theta \mid x) = N(x,1) \text{ and } \pi_2^N(\lambda \mid x) \propto \frac{1}{\lambda^2} \exp\left(-\frac{e^x}{\lambda}\right)$$

so that $\lambda^{-1} \mid x \sim Exp(e^x)$ and the transition $\theta \to \theta'$ of the associated Markov chain is thus made of the following steps:

1. Generate $x' \mid \theta \sim N(\theta,1)$
2. Generate $\lambda \mid x' \sim 1/Exp(e^{x'})$
3. Generate $x \mid \lambda \sim \log Exp(\lambda)$
4. Generate $\theta' \mid x \sim N(x,1)$

Then, if we express the four moves at once, θ' given θ satisfies

$$\theta' = \log(\lambda\varepsilon_3) + \varepsilon_4 = -\log(\varepsilon_2/e^{x'}) + \log\varepsilon_3 + \varepsilon_4 = \theta + \varepsilon_1 - \log\varepsilon_2 + \log\varepsilon_3 + \varepsilon_4,$$

where $\varepsilon_1, \varepsilon_4$ are iid $N(0,1)$ and $\varepsilon_2, \varepsilon_3$ are iid $Exp(1)$, and the transition is therefore a random walk in θ (it can also be checked that this is equally a random walk in $\log \lambda$).

Now, $\log(\varepsilon_3/\varepsilon_2)$ is a logistic variable. Indeed,

$$\eta = \varepsilon_3/\varepsilon_2 \sim \int_0^\infty \exp(-\varepsilon_2(1+\eta))\varepsilon_2 d\varepsilon_2 = (1+\eta)^{-2}$$

and $\zeta = \log(\eta) \sim e^\zeta/(1+e^\zeta)^2$.

Therefore, the white noise in $\theta' = \theta + \varepsilon$ has a finite second moment and again this random walk is (null) recurrent. The resulting Lebesgue measures are thus invariant for θ and $\log\lambda$ and $\pi_1(\theta) = 1$ and $\pi_2(\lambda) = 1/\lambda$ are the unique solutions to the system of integral equations up to a multiplicative constant.

4 Conclusions

A new methodology to deal with the problem of choosing calibrated default priors for model selection based on the solutions of a system of integral equations has been proposed.

The resulting priors have been called integral priors and the main advantage of the class of integral priors over the class of intrinsic priors proposed in Berger and Pericchi (1996) is that provided the Markov chain associated to the system of integral equations is recurrent the integral priors are unique. Specifically, in the examples we deal with in subsection 3.5 the integral priors are unique and they are explicitely obtained. On the other hand, in the cases where they are not explicitely obtained an approximation to the unique Bayes factor can be obtained using the ergodic theorem provided the associated Markov chain is (Harris) recurrent.

References

Berger, J.O. and Pericchi, L.R. (1996). The Intrinsic Bayes Factor for Model Selection and Prediction. *J. Amer. Statist. Assoc.* **91**, 109-122.

Bernardo, J. M. (1979). Reference Posterior Distribution for Bayesian Inference. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **41**, 113-147.

Cano, J. A., Kessler, M. and Moreno, E. (2004). On Intrinsic Priors for Nonnested Models. Test, **13-2**, 445-463.

Jeffreys, H. (1961). Theory of Probability. *Oxford University Press. London.*

Meyn, S. P. and Tweedie, R. L. (1993). Markov Chains and Stochastic Stability. Springer-Verlag. New York.

Moreno, E., Bertolino, F. and Racugno, W. (1998). An Intrinsic Limiting Procedure for Model Selection and Hypotheses Testing. *J. Amer. Statist.*

Assoc. **93**, 1451-1460.
Pérez, J. M. and Berger, J. (2002). Expected Posterior Prior Distributions for Model Selection. Biometrika, **89**, 491-512.