The problem

Let \( y \) be a random sample obtained from the model \( f(y \mid \theta, \nu) \). We consider the hypothesis testing (model selection) problem

\[
H_1 : \theta = \theta_0 \quad \text{vs.} \quad H_2 : \theta \neq \theta_0,
\]

or equivalently

\[
H_1 : f(y \mid \theta = \theta_0) = f(y \mid \theta = \theta_0, \nu) = f(y \mid \theta = \theta_0, \nu) \quad \text{vs.} \quad H_2 : f(y \mid \theta = \theta_0, \nu).
\]

The Bayes factor in favor of \( H_1 \) and against \( H_2 \) is the quotient between the prior predictive marginal distributions evaluated at \( y \), that is, 

\[
\begin{align*}
\text{BF}_{12} &= \frac{m_1(y)}{m_2(y)} = \frac{\int f(y \mid \theta_0, \nu) d\theta_0}{\int f(y \mid \theta, \nu) d\theta}.
\end{align*}
\]

Goal: select in an objective Bayes fashion the priors \( Kass \) and \( Wasserberg, 1995; Raftery, 1998 \).

The theory

First case: \( \theta \) location, \( \nu \) known.

Only the prior \( \nu \) needs to be assigned. We propose two alternatives.

\[
\text{The sumDB prior: } \pi^{\text{sumDB}}_{\nu} \propto (1 + \frac{f(y \mid \theta, \nu)}{f(y \mid \theta_0, \nu)})^{-q},
\]

and

\[
\text{The minDB prior: } \pi^{\text{minDB}}_{\nu} \propto (1 + \frac{f(y \mid \theta_0, \nu)}{f(y \mid \theta, \nu)})^{-q},
\]

where \( q > 0 \) is such that the corresponding prior is proper. More on the election of \( q \) later.

The sumDB is generally our preferred prior although the minDB is an interesting alternative in certain scenarios. Properties: The sumDB priors are unimodal, symmetric w.r.t. \( \theta_0 \) with flat tails. In a neighborhood of \( \theta_0 \), behave as Student densities.

Second case: \( \nu \) known.

If \( \nu^{*}(\theta) = 1 \) think of 0 as location parameter (asymptotically behaves as location parameter) and use the ‘First case’ to derive the DB prior. In the general case (where \( \nu^{*}(\theta) \) doesn’t exist) we use the transformation formula giving rise to the definition of DB priors.

\[
\text{The sumDB prior: } \pi^{\text{sumDB}}_{\nu}(\theta) \propto (1 + \frac{f(y \mid \theta, \nu)}{f(y \mid \theta_0, \nu)})^{-q\nu(\theta)},
\]

and

\[
\text{The minDB prior: } \pi^{\text{minDB}}_{\nu}(\theta) \propto (1 + \frac{f(y \mid \theta_0, \nu)}{f(y \mid \theta, \nu)})^{-q\nu(\theta)},
\]

(more) Properties: Very easy to derive.

- Invariant under 1-1 transformations.
- Satisfy the efficiency principle.
- \( DB_1 = DB_2 \) (Correction factor (CF)), where the CF is a expectation with respect to the posterior (very convenient for computation in an MCMC scheme).

Third case: \( \nu \) unknown.

Let, possibly after reparameterization, \( \nu \) and \( \theta \) be orthogonal parameters, allowing one to assume that \( \nu = \nu_0 \). Define \( \nu^{*}(\theta) \) such that \( \pi^{\text{sumDB}}_{\nu}(\theta) = \nu^{*}(\theta) \pi^{\text{sumDB}}_{\nu}(\theta) \).

We propose using

\[
x_1(\nu) = \nu^{*}(\nu) \quad \text{and} \quad x_2(\nu) = \nu^{*}(\nu) \nu^{*}(\theta)
\]

where

\[
\begin{align*}
\text{The sumDB prior: } \pi^{\text{sumDB}}_{\nu}(\theta) \propto (1 + \frac{f(y \mid \theta, \nu)}{f(y \mid \theta_0, \nu)})^{-q\nu(\theta)},
\end{align*}
\]

and

\[
\begin{align*}
\text{The minDB prior: } \pi^{\text{minDB}}_{\nu}(\theta) \propto (1 + \frac{f(y \mid \theta_0, \nu)}{f(y \mid \theta, \nu)})^{-q\nu(\theta)},
\end{align*}
\]

(more) Properties: Indeterminate constants cancel out in Bayes factor.

- Invariance under transformations which conserve the orthogonality.

The selection of \( q \)

Let \( q = \inf\{q > 0 \mid \text{such that } \pi^{\text{sumDB}}_{\nu}(\theta) \text{ is proper} \} \).

If \( q < \infty \), then \( q \in (0, \infty) \) guarantees the propriety of the prior. We propose using

\[
q = g = \frac{1}{2},
\]

mainly inspired in the proposals of JZS of using a Cauchy in Normal problems. In general, this election gives rise to flat tailed priors and usually without moments.

Abstract

For the hypothesis testing problem and mainly based on measures of divergence between the competing models, we introduce a new type of proper prior distributions, called divergence based (DB) priors. Despite their simple form, the DB priors have some appealing properties like invariance and agreement with efficiency principle. In normal linear models, the DB priors exactly reproduce the proposals of Jeffreys-Zellner-Siow. In other standard scenarios the DB’s behave similarly to other existing proposals like the intrinsic priors. More interestingly is that for more challenging scenarios (like irregular and mixture models) where the other proposals fail, the DB priors are well defined, and provide sensible answers.

Selected examples

- In linear models the DB priors coincide with the proposals of Zellner and Siow (1980, 1984).
- In “traditional” problems (Bernoulli, one parameter exponential, Poisson or Normal models) the DB priors produce Bayes factors that are close to the Bayes factor obtained with AI and FI.
- In slightly more complicated problems, where the AI and FI priors do not seem to have a closed expression (Gelman model), the DB priors have a closed form due to derive. The Bayes factors obtained are close to the arithmetic intrinsic Bayes factors.
- In unbalanced scenarios (random effects models) the DB priors still have a closed expression.

Improper likelihoods: mixture models

- Let \( f(\theta \mid \theta_0, \nu) = p(\nu_0 \mid 1, 1) + (1 - \rho) p(\nu_0 \mid 1, 1) \). (p known)

The usual default Bayes factors cannot be defined. Berger and Pericchi (2001) recommend using \( \pi^{\text{DB}}(\theta) \propto \text{Cus}(\theta)^{p/2} \).

The minDB prior does not exist, while the sumDB:

\[
\pi^{\text{sumDB}}_{\nu}(\theta) = (1 + \bar{L}_j(\theta, \nu_0))^{-1}, \quad \text{where } \bar{L}_j(\theta, \nu_0) = (1 - \rho)^{p/2}.
\]

Hence

\[
\pi^{\text{sumDB}}_{\nu}(\theta) \propto \text{Cus}(\theta)^{p/2}.
\]

Graphical comparison

- \( \pi^{\text{sumDB}}_{\nu}(\theta) \) (solid) and Cus(\theta)^{p/2} (dashed) for \( \rho = 0.2 \) (left) and \( \rho = 0.70 \) (right). Right-Exp(\theta_0, 1).

Simulation:

\[
\begin{align*}
&\text{DF} = 0.5 \quad 1 \quad 2 \\
&\text{Inter} = 0.03 \quad 0.06 \quad 0.09
\end{align*}
\]

\[
\text{Note: Results quite similar, slightly better sumDB (large } \nu_0 \text{) under } M_1 \quad \text{and smaller } \nu_0 \text{ under } M_2 \text{ (} \theta = 0 \text{).}
\]

Irregular asymptotic: irregular models

- Let \( f(y \mid \theta) = \exp\{-|y - \theta|\} \), \( \gamma > 0 \), and the one-sided problem \( H_0 : \theta = \theta_0 \) vs. \( H_1 : \theta > \theta_0 \).

The Bayesian information criteria (BIC) is inapplicable. Besides, these models pose in problems to the AI (Berger and Pericchi, 2001). They ‘conjecture’ that the AI is \( \pi^{\text{AI}}(\theta) = (1 - \gamma \theta_0) \theta_0^{-1/2} \). \( \theta > \theta_0 \).

Besides, the fractional Bayes factors is unsatisfactory (favoring always \( M_2 \)).

The sumDB does not exist. The minDB is

\[
\pi^{\text{minDB}}_{\nu}(\theta) = (1 + 2\bar{L}_2)^{-1/2}, \quad \theta > \theta_0.
\]

Graphical comparison

- For \( \theta_0 = 0 \), \( \pi^{\text{minDB}}_{\nu}(\theta) \) and \( \pi^{\text{minDB}}_{\nu}(\theta) \) (dashed) for \( \gamma = 0.50 \) (left) and \( \gamma = 0.70 \) (right). Right-Exp(\theta_0, 1).

Analytical comparison

\[
\begin{align*}
\text{AI} &\approx 0.35 \quad 0.60 \quad 1.00 \\
\text{sumDB} &\approx 0.30 \quad 0.55 \quad 0.90 \\
\text{minDB} &\approx 0.30 \quad 0.55 \quad 0.90
\end{align*}
\]

Note: For small values of \( T \), the sumDB provides more support to the simpler model than AI prior. Both Bayes factors become similar as \( T \) grows.