Equivariant HPD credible sets and MAP estimators

Pierre Druilhet, Jean-Michel Marin

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Equivariant HPD credible sets and MAP estimators

Pierre Druilhet — Jean-Michel Marin

N° 5921

Juin 2006

Thème COG

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--5923--FR+ENG

Equivariant HPD credible sets and MAP estimators

Pierre Druilhet * , Jean-Michel Marin †

Thème COG — Systèmes cognitifs
Projet Select

Rapport de recherche n° 5921 — Juin 2006 — 11 pages

Abstract: MAP estimators and HPD credible sets are often criticized in the literature because of paradoxical behaviour due to a lack of equivariance under reparametrization. In this paper, we propose a new version of MAP estimators and HPD credible sets that avoid this undesirable feature. Moreover, in the special case of non-informative prior, the new MAP estimators coincide with the equivariant frequentist ML estimators. We also propose several adaptations in the case of nuisance parameters.

Key-words: Bayesian statistics, HPD, MAP, Jeffreys measure, nuisance parameters, reference prior

* ENSAI, Rennes
† Projet Select, INRIA Futurs

Unité de recherche INRIA Futurs
Parc Club Orsay Université, ZAC des Vignes,
4, rue Jacques Monod, 91893 ORSAY Cedex (France)
Téléphone : +33 1 72 92 59 00 — Télécopie : +33 1 60 19 66 08

Régions HPD et estimateurs MAP équivariants

Résumé : Dans cet article, nous introduisons des régions HPD et des estimateurs MAP équivariants par reparamétrisation. Dans le cas non-informatif, les estimateurs MAP proposés sont égaux aux estimateurs du maximum de vraisemblance. Nous étudions le cas de paramètres de nuisance.

Mots-clés : statistique bayésienne, HPD, MAP, mesure de Jeffreys, paramètres de nuisance, loi a priori de référence

1 Introduction

The Maximum A Posteriori estimator (MAP) is defined to be the value (non necessary unique) that maximizes the posterior density w.r.t. the Lebesgue measure, denoted by λ. The MAP is the Bayesian equivalent to the frequentist Maximum Likelihood estimator (ML) and both coincide for the non-informative Laplace prior. Unlike MLs, MAPs are not equivariant under smooth reparametrization. Because of this undesirable feature, many authors do not recommend its use.
Consider the following example: $X|\theta \sim \mathcal{N}(\theta, \sigma^2)$ and $\theta \sim \mathcal{N}(\mu, \tau^2)$ where σ^2, μ and τ^2 are assumed to be known. The posterior distribution of θ is normal and

$$\mathrm{MAP}(\theta) = \frac{\tau^2}{\tau^2+\sigma^2}\, x + \frac{\sigma^2}{\tau^2+\sigma^2}\, \mu \,.$$

For the new parameterization $\alpha = e^\theta$, the posterior distribution of α is log-normal and

$$\mathrm{MAP}(\alpha) = e^{\frac{\tau^2}{\tau^2+\sigma^2}\, x + \frac{\sigma^2}{\tau^2+\sigma^2}\, \mu - \sqrt{\frac{\tau^2\sigma^2}{\tau^2+\sigma^2}}} \neq e^{\mathrm{MAP}(\theta)} \,.$$

The lack of equivariance for the MAP is mainly due to the choice of the Lebesgue measure as dominating measure. Indeed, the MAP is based only on the density of the posterior distribution and not on the exact distribution. Obviously, the MAP depends entirely on the choice of the dominating measure.
Consider a model given by the density $f(x|\theta)$, where the parameter θ lies in an open subset Θ of $\mathbb{R}^p$. We denote by Π the (possibly improper) continuous prior distribution on θ and by Π_x the posterior distribution of θ which is assumed to be proper. Denote respectively by $\pi_\lambda(\theta)$ and $\pi_\lambda(\theta|x) \propto f(x|\theta)\,\pi_\lambda(\theta)$ the prior and posterior density of θ w.r.t. λ. Consider a new dominating measure ν whose density w.r.t. λ is given by $g(\theta) > 0$. Similarly, denote respectively by $\pi_\nu(\theta)$ and $\pi_\nu(\theta|x) \propto f(x|\theta)\,\pi_\nu(\theta)$ the prior and posterior density of θ w.r.t. ν. We named MAP_ν the MAP based on the dominating measure ν, i.e.

$$\mathrm{MAP}_\nu(\theta) = \operatorname*{Argmax}_{\theta\in\Theta} \pi_\nu(\theta|x) = \operatorname*{Argmax}_{\theta\in\Theta} \left[\frac{\pi_\lambda(\theta|x)}{g(\theta)}\right] . \tag{1}$$

With this notation, $\mathrm{MAP} = \mathrm{MAP}_\lambda$. There is in fact no clear justification for the choice of the Lebesgue measure as dominating measure. In this paper, we discuss another choice for the MAP.

Another possibility to get information on θ through the posterior distribution is to use credible sets, regions in which θ belong a posteriori with a given probability. One way to choose such sets is to define Highest Probability Density credible sets (HPDs). Formally, a set $\mathrm{HPD}^\gamma(\theta) \subset \Theta$ is an HPD of level γ if there exists a constant k_γ such that

$$\{\theta : \pi_\lambda(\theta|x) > k_\gamma\} \subset \mathrm{HPD}^\gamma(\theta) \subset \{\theta : \pi_\lambda(\theta|x) \geq k_\gamma\} \quad \text{and}$$

$$\Pi_x\left(\{\theta : \pi_\lambda(\theta|x) > k_\gamma\}\right) \leq \gamma \leq \Pi_x\left(\{\theta : \pi_\lambda(\theta|x) \geq k_\gamma\}\right) .$$

If $\Pi_x(\{\theta : \pi_\lambda(\theta|x) = k_\alpha\}) = 0$ (this is not the case for example if the posterior distribution is uniform or is flat on some intervals), we simply write:

$$\mathrm{HPD}^\gamma(\theta) = \{\theta : \pi_\lambda(\theta|x) \geq k_\gamma\} \quad \text{and} \quad \Pi_x\left(\mathrm{HPD}^\gamma(\theta)\right) = \gamma\,. \tag{2}$$

From now on, to avoid unnecessary complicated writing, we assume that HPDs can always be written as in (2). It is well known that $\mathrm{HPD}^\gamma(\theta)$ minimizes the length (or volume in the multivariate case) among the credible sets of level greater or equal to γ, where the unit of length or volume is given by the Lebesgue measure λ. It is worth noting that when the MAP exists and is unique and when $\pi_\lambda(\theta|x)$ is regular (e.g. semi-continuous), the MAP can be obtained from HPDs by

$$\mathrm{MAP}(\theta) = \bigcap_{0<\gamma<1} \mathrm{HPD}^\gamma(\theta)\,. \tag{3}$$

From now on, we omit the subscript γ in HPD^γ. As MAPs, HPDs are criticized for their lack of equivariance under reparametrization leading to paradoxical behaviours.
Consider for example $X|\theta \sim \mathcal{B}(\theta)$ and $\theta \sim \mathcal{U}_{]0,1[}$. The HPD for θ is

$$\mathrm{HPD}(\theta) = \left\{\theta : \theta^x(1-\theta)^{1-x} \geq k_\gamma\right\}\,.$$

For the new parameterization $\alpha = \log(\theta/(1-\theta)) = \mathrm{logit}(\theta)$,

$$\mathrm{HPD}(\alpha) = \left\{\alpha : \frac{e^{(x+1)\alpha}}{(1+e^\alpha)^3} \geq k'_\gamma\right\}\,.$$

Figures 1 presents the posterior densities of θ and α when $x = 0$. The case $x = 1$ is similar. In the original parameterization by θ, the HPDs are one-sided, while for a monotonic reparametrization by α, the HPDs become two-sided and obviously, $\mathrm{HPD}(\alpha) \neq \mathrm{logit}(\mathrm{HPD}(\theta))$. For more examples and discussions of this conflict, see Lehmann and Romano (2005)[section 5.7] and Berger (1985)[section 4.3.2].

As MAPs, HPDs are defined only through the density of the posterior distribution and therefore depends on the arbitrary choice of the dominating measure, or equivalently the unit of length or volume. The implicit choice for the HPD is the Lebesgue measure. If we choose ν as dominating measure, we can define a new HPD region, named $\mathrm{HPD}_\nu(\theta)$, by

$$\mathrm{HPD}_\nu(\theta) = \{\theta : \pi_\nu(\theta|x) \geq k_\gamma\}\,. \tag{4}$$

With this notation, $\mathrm{HPD}(\theta) = \mathrm{HPD}_\lambda(\theta)$. Note that HPD_ν is the region of level γ with minimal length or volume where $\mathrm{length}_\nu(C) = \int_C \mathrm{d}\nu(\theta)$.
The aim of this paper is to discuss a new choice of dominating measure that make MAPs and HPDs equivariant under $1-1$ smooth reparametrization. Of course, the choice should only depend on the model $f(x|\theta)$. To make coherent the Bayesian and frequentist approaches, we impose that the new MAP and the ML coincide for a non informative prior. Under these conditions, it is natural to choose the Jeffreys measure as dominating measure. It is worth

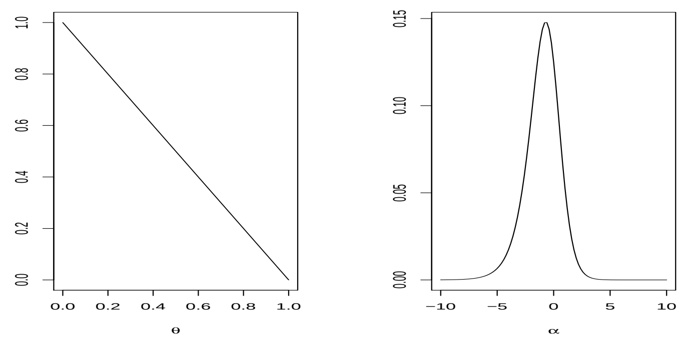

Figure 1: Posterior densities of θ and α when $x = 0$

noting that the choice of a dominating measure is not directly connected with the choice of a prior distribution which corresponds to a prior knowledge on the parameter. In Section 2, we discuss the implication of such a choice for the dominating measure. In Section 3, we adapt our approach to the delicate case of models with nuisance parameters.

2 Equivariant MAP and HPD

In this section, we assume the usual regularity conditions on the model given by $f(x|\theta)$ so that the Fisher information $I(\theta)$ is well defined. We also assume that $I(\theta)$ is positive definite for each $\theta \in \Theta$. The Jeffreys measure for θ, denoted by J_θ, is the measure with density $|I(\theta)|^{\frac{1}{2}}$ w.r.t. λ. We denote by $\mathrm{JMAP}(\theta) = \mathrm{MAP}_{J_\theta}(\theta)$ the MAP obtained by taking the Jeffreys measure as dominating measure. Similarly, we denote by $\mathrm{JHPD}(\theta) = \mathrm{HPD}_{J_\theta}(\theta)$ the HPD region with the Jeffreys measure as dominating measure. We have:

$$\mathrm{JMAP}(\theta) = \operatorname*{Argmax}_{\theta\in\Theta} \; f(x|\theta)\; |I(\theta)|^{-\frac{1}{2}}\; \pi_\lambda(\theta)\,, \tag{5}$$

and

$$\mathrm{JHPD}(\theta) = \{\theta \;:\; f(x|\theta)\; |I(\theta)|^{-\frac{1}{2}}\; \pi_\lambda(\theta) \geq k_\gamma\}\,. \tag{6}$$

The first motivation for using JHPDs and JMAPs is that they lead to equivariant inference under differentiable reparametrization. Rousseau and Robert (2005) briefly consider a similar idea in a discussion on a paper of Bernardo (2005). Consider now a differentiable reparametrization $\alpha = h(\theta)$, the posterior density of α w.r.t. to the Jeffreys measure for α, denoted by J_α, is:

$$\pi_{J_\alpha}(\alpha|x) = \frac{\pi_\lambda\left(h^{-1}(\alpha)|x\right)\; \left|\frac{d}{d\alpha}h^{-1}(\alpha)\right|}{|I(h^{-1}(\alpha))|^{-\frac{1}{2}}\; \left|\frac{d}{d\alpha}h^{-1}(\alpha)\right|} = \pi_{J_\theta}\left(h^{-1}(\alpha)|x\right)\,. \tag{7}$$

From Eq. (7), we obtain the functional equivariance properties of the JMAP and the JHPD:

$$\begin{aligned}
\mathrm{JMAP}(h(\theta)) &= \underset{\alpha\in h(\Theta)}{\mathrm{Argmax}}\, \pi_{J_\alpha}(\alpha|x) \\
&= \underset{\alpha\in h(\Theta)}{\mathrm{Argmax}}\, \pi_{J_\theta}\left(h^{-1}(\alpha)|x\right) \\
&= h(\mathrm{JMAP}(\theta))\,. \qquad (8)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{JHPD}(\alpha) &= \{\alpha : \pi_{J_\alpha}(\alpha|x) \geq k_\gamma\} \\
&= \left\{\alpha : \pi_{J_\theta}\left(h^{-1}(\alpha)|x\right) \geq k_\gamma\right\} \\
&= h\left(\mathrm{JHPD}(\theta)\right)\,. \qquad (9)
\end{aligned}$$

Let us consider the normal example of section 1, $X|\theta \sim \mathcal{N}(\theta,\sigma^2)$ and $\theta \sim \mathcal{N}(\mu,\tau^2)$. We have $I(\theta) \propto 1$ and $\mathrm{JMAP}(\theta) = \mathrm{MAP}_\lambda(\theta)$. For $\alpha = \exp(\theta)$, we can derive $\mathrm{JMAP}(\alpha)$ from $\mathrm{JMAP}(\theta)$ by

$$\mathrm{JMAP}(\alpha) = e^{\mathrm{JMAP}(\theta)}\,.$$

Consider now the Bernoulli example of Section 1, $X|\theta \sim \mathcal{B}(1,\theta)$ and $\theta \sim \mathcal{U}_{]0,1[}$. We have $I(\theta) = 1/(\theta(1-\theta))$ and

$$\mathrm{JHPD}(\theta) = \left\{\theta : \theta^{x+1/2}(1-\theta)^{1-x+1/2} \geq k_\gamma\right\}\,.$$

For $\alpha = h(\theta) = \log(\theta/(1-\theta))$, we have

$$\begin{aligned}
\mathrm{JHPD}(\alpha) &= \left\{\alpha : \frac{\exp((x+1/2)\alpha)}{(1+\exp(\alpha))^2} \geq k_\gamma\right\} \\
&= h(\mathrm{JHPD}(\theta))\,.
\end{aligned}$$

Figures 2 presents the posterior densities of θ and α w.r.t. the Jeffreys dominating measures. Contrary to the Lebesgue dominating measures case, we obtain two-sided regions.

The other motivation for the choice of J_θ as dominating measure is that the Jeffreys measure is a classical non-informative prior for θ (Jeffreys, 1961; Kass and Wasserman, 1996). Recall that, providing there are no nuisance parameters, Bernardo (1979) showed that the Jeffreys prior distribution minimizes the asymptotic expected Kullback-Leibler distance between the prior and the posterior distributions. Using our approach, if no prior knowledge is available on θ and if we accept the Jeffreys measure as noninformative prior, then the JMAP is equal to the frequentist ML whatever the parametrization is, provided the Fisher information is defined. Moreover, in this case, JHPDs can then be thought as credible sets "around" the ML, where the term around may be abusive in case of multimodal posterior distribution (multimodal is defined w.r.t. the Jeffreys measure).

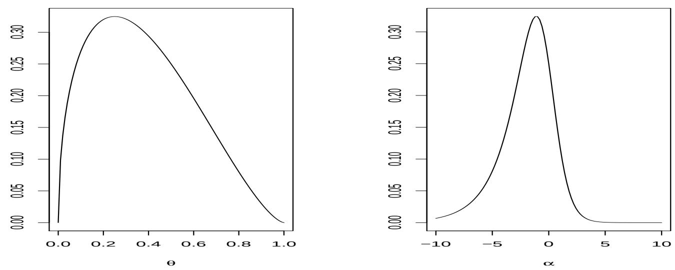

Figure 2: Posterior densities of θ and α w.r.t. the Jeffreys measure when $x = 0$

3 Models with nuisance parameter

In this section, we discuss several methods to derive an equivalent of JMAPs and JHPDs when nuisance parameter are present in the model. We assume that the parameter θ is split into two parts: $\theta = (\theta_1, \theta_2) \in \Theta_1 \otimes \Theta_2$ where θ_1 is the parameter of interest and θ_2 is the nuisance. We denote by $\pi_\nu(\theta_1|x)$ the density of the marginal posterior distribution of θ_1 w.r.t. the measure ν. The corresponding MAP_ν and HPD_ν are

$$\text{MAP}_\nu(\theta_1) = \underset{\theta_1 \in \Theta_1}{\text{Argmax}}\, \pi_\nu(\theta_1|x) \tag{10}$$

$$\text{and} \quad \text{HPD}_\nu(\theta_1) = \{\theta_1 : \pi_\nu(\theta_1|x) \geq k_\gamma\} \tag{11}$$

Because the Jeffreys prior does not distinguish between parameter of interest and nuisance parameter, Bernardo (1979) propose a new approach called reference prior approach. In this Section, we show how we can used this reference prior as dominating measure to define equivariant MAPs and HPDs, called by analogy with Section 2, JMAPs and JHPDs. Two cases are considered: the case where there is conditional subjective information for the nuisance parameter and the case where there is none.

3.1 A subjective conditional prior is available

Suppose that a subjective conditional prior is available for θ_2 given θ_1. We denote by $\pi_\lambda(\theta_2|\theta_1)$ its density w.r.t. λ. In this case, Sun and Berger (1998) propose two different approaches to derive reference priors for θ_1. We mimic their approaches and there is two reasonable options for finding a dominating measure on Θ_1.

Option 1. consider the marginal model $f(x|\theta_1) = \int f(x|\theta_1, \theta_2)\pi_\lambda(\theta_2|\theta_1)d\theta_2$. Denote by $I_m(\theta_1)$ the Fisher information matrix for θ_1 obtained from the marginal model. A dominating measure can be the Jeffreys measure on Θ_1 with density w.r.t. λ proportional to $|I_m(\theta_1)|^{1/2}$. In this case, the JMAP and the JHPD are defined by

$$\mathrm{JMAP}_1(\theta_1) = \operatorname*{Argmax}_{\theta_1 \in \Theta_1} \left[\frac{\pi_\lambda(\theta_1|x)}{|I_m(\theta_1)|^{1/2}} \right],$$

$$\mathrm{JHPD}_1(\theta_1) = \left\{ \theta_1 : \frac{\pi_\lambda(\theta_1|x)}{|I_m(\theta_1)|^{1/2}} \geq k_\gamma \right\}.$$

JMAP_1 and JHPD_1 are obviously equivariant for a $1-1$ smooth reparametrization on the parameter of interest. Unfortunately, the Fisher information matrix for $\int f(x|\theta_1,\theta_2)\pi_\lambda(\theta_2|\theta_1)d\theta_2$ is often difficult to compute. This difficulty motivates the introduction of another option.

Option 2. Following Bernardo (1979), Sun and Berger (1998) propose to maximize asymptotically the expected Kullback-Leibler divergence between the marginal posterior of θ_1 and the marginal prior of θ_1. This leads to the distribution on Θ_1 with density w.r.t. λ proportional to

$$\exp\left\{ \frac{1}{2} \int \pi_\lambda(\theta_2|\theta_1) \log\left(\frac{|I(\theta_1,\theta_2)|}{|I_c(\theta_2|\theta_1)|} \right) d\theta_2 \right\}$$

where $I(\theta_1,\theta_2)$ is the Fisher information matrix based on $f(x|\theta_1,\theta_2)$ and $I_c(\theta_2|\theta_1)$ is the Fisher information matrix based on the model $f(x|\theta_1,\theta_2)$ where θ_1 is known. This is essentially the solution used in Berger and Bernardo (1989, 1992), but here a subjective conditional prior for the nuisance parameter given the parameter of interest is used. We propose to use the previous distribution as dominating measure on Θ_1. In this case, the JMAP and the JHPD are defined by

$$\mathrm{JMAP}_2(\theta_1) = \operatorname*{Argmax}_{\theta_1 \in \Theta_1} \left[\frac{\pi_\lambda(\theta_1|x)}{\exp\left\{ \frac{1}{2} \int \pi_\lambda(\theta_2|\theta_1) \log\left(\frac{|I(\theta_1,\theta_2)|}{|I_c(\theta_2|\theta_1)|} \right) d\theta_2 \right\}} \right],$$

$$\mathrm{JHPD}_2(\theta_1) = \left\{ \theta_1 : \frac{\pi_\lambda(\theta_1|x)}{\exp\left\{ \frac{1}{2} \int \pi_\lambda(\theta_2|\theta_1) \log\left(\frac{|I(\theta_1,\theta_2)|}{|I_c(\theta_2|\theta_1)|} \right) d\theta_2 \right\}} \geq k_\gamma \right\}.$$

JMAP_2 and JHPD_2 are obviously equivariant for a $1-1$ smooth reparametrization on the parameter of interest.

Let us consider a sample $X_1, \ldots, X_n$ from a normal distribution with expectation $\theta_2 = \mu$ and standard deviation $\theta_1 = \sigma$, the parameter of interest. Suppose that the conditional prior distribution for μ given σ is normal with expectation m and variance τ^2. Applying proposition 2 of Sun and Berger (1998), Option 1 dominating measure has density w.r.t. λ proportional to $\left(\frac{1}{\sigma^2} + \frac{\sigma^2}{(n-1)(\sigma^2+n\tau^2)^2} \right)^{1/2}$, and Option 2 dominating measure has density w.r.t. λ proportional to $1/\sigma$. These two dominating measures are different. However, when $n \to \infty$, the second density converge uniformly to the first one.

Let us now consider the bivariate binomial model proposed by Crowder and Sweeting (1989) and revisited by Polson and Wasserman (1990):

$$\begin{aligned} f(x_1, x_2|\theta_1, \theta_2) &= \binom{m}{x_1} \theta_1^{x_1}(1-\theta_1)^{m-x_1} \binom{x_1}{x_2} \theta_2^{x_2}(1-\theta_2)^{x_1-x_2} \\ & \quad \mathbb{I}_{1,\dots,m}(x_1)\mathbb{I}_{1,\dots,x_1}(x_2)\,. \end{aligned}$$

where $\mathbb{I}_A$ is the indicator function on A and m is supposed to be known. Suppose that the conditional distribution of θ_2 given θ_1 is a Beta distribution with parameter a and b. For such a model, $|I(\theta_1, \theta_2)| = (1-\theta_1)^{-1}\theta_2^{-1}(1-\theta_2)^{-1}$ and $|I_c(\theta_2|\theta_1)| = \theta_1(\theta_2(1-\theta_2))^{-1}$. It is very easy to show that Option 1 and Option 2 dominating measures are the same and have density w.r.t. λ proportional to $\theta_1^{-1/2}(1-\theta_1)^{-1/2}$.

Sun and Berger (1998) considered the case where θ_1 and θ_2 are independent. For this other prior information, we can also mimic their approach to define a dominating measure on Θ_1.

3.2 No subjective conditional prior available

If no subjective conditional prior for θ_2 given θ_1 is available, we propose to mimic the reference prior approach of Berger and Bernardo (1989, 1992). This leads to the dominating measure on Θ_1 with density w.r.t. λ proportional to

$$\exp\left\{\frac{1}{2}\int |I_c(\theta_2|\theta_1)|^{1/2} \log\left(\frac{|I(\theta_1,\theta_2)|}{|I_c(\theta_2|\theta_1)|}\right) d\theta_2\right\}.$$

Often, the integral $\int |I_c(\theta_2|\theta_1)|^{1/2} \log\left(\frac{|I(\theta_1,\theta_2)|}{|I_c(\theta_2|\theta_1)|}\right) d\theta_2$ is not defined. The compact support argument that is typically used in the reference prior approach (Berger and Bernardo, 1992) may then be applied here. Choose a nested sequence $\Theta_1 \subset \Theta_2 \subset \dots$ of compact subsets of the parameter space Θ such that $\cup_i \Theta_i = \Theta$ and $|I_c(\theta_2|\theta_1)|^{1/2}$ has finite mass on $\Omega_i = \{\theta_2; (\theta_1, \theta_2) \in \Theta_i\}$ for all θ_1. Let $K_i(\theta_1) = \int_{\Omega_i} |I_c(\theta_2|\theta_1)|^{1/2} d\theta_2$ and

$$\pi_i(\theta_1) = \exp\left\{\frac{1}{2}\int |I_c(\theta_2|\theta_1)|^{1/2} \log\left(\frac{|I(\theta_1,\theta_2)|}{|I_c(\theta_2|\theta_1)|}\right) d\theta_2\right\}.$$

The dominating measure on Θ_1 has then density w.r.t. λ proportional to

$$\lim_{i\to\infty} \frac{K_i(\theta_1)\pi_i(\theta_1)}{K_i(\theta_1^{(0)})\pi_i(\theta_1^{(0)})}$$

where $\theta_1^{(0)}$ is any fixed point in Θ_1. Datta and Ghosh (1996) has established the invariance of this procedure under $1-1$ smooth reparametrization on θ_1. Therefore, the corresponding JMAP and JHPD are equivariant under a smooth reparametrization on the parameter of interest.

Let us consider again a sample $X_1, \ldots, X_n$ from a normal distribution with expectation μ and standard deviation σ, where σ is the parameter of interest. Applying the previous procedure, the dominating measure on σ has density w.r.t. λ proportional to $1/\sigma$, which is the invariant measure for scale models. We assume now that no prior information is available on σ. So, the non-informative reference prior is given by $\pi_\lambda(\sigma) = 1/\sigma$. Equivalently, if the parameter of interest is σ^2, the reference prior and dominating measure have density w.r.t. Lebesgue proportional to $\frac{1}{\sigma^2}$, which is again the corresponding the invariant measure. In that case $\mathrm{JMAP}(\sigma^2)$ is equal to the frequentist REML (REsidual Maximum Likelihood) estimator. This is a new interpretation of the REML estimator which also corresponds to the MAP under the Laplace prior for σ^2 (Harville, 1974). By equivariance of the JMAP, we have, $\mathrm{JMAP}(\sigma) = \sqrt{\mathrm{REML}}$. If we change σ into $\log(\sigma)$, then the reference prior is the Lebesgue measure. In that case, $\mathrm{JMAP}(\log(\sigma)) = \mathrm{MAP}(\log(\sigma)) = \frac{1}{2}\log(\mathrm{REML})$. Note that these results can be extended to more general variance components models.

4 Conclusion

The JMAPs and JHPDs proposed in this paper give a simple and coherent alternative to the usual MAPs and HPDs, avoiding peculiar behaviour under reparametrization. However, there are many important non-regular problems where the Jeffreys measure does not exist and some developments should be done in this direction.

Acknowledgments

We thank Pr. Robert for its helpful comments and a careful reading of a preliminary version of the paper. We also thank Pr. Foulley for helpful discussions on the Bayesian approach of variance component models.

References

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, 2 edition.

Berger, J. and Bernardo, J. (1989). Estimating a Product of Means: Bayesian Analysis With Reference Priors. *J. American Statist. Assoc.*, 84(405):200–207.

Berger, J. and Bernardo, J. (1992). Ordered Group reference Priors With Application to Multinomial and Variance Component Problems. *Biometrika*, 79:25–37.

Bernardo, J. (1979). Reference Posterior Distributions for Bayesian Inference. *J. Royal Statist. Soc. Series B*, 41(2):113–147.

Bernardo, J. (2005). Intrinsic Credible Regions:An Objective Bayesian Approach to Interval Estimation. *Test*, 14(2):317–384.

Crowder, M. and Sweeting, T. (1989). Bayesian inference for a bivariate binomial. *Biometrika*, 76:599–604.

Datta, G. and Ghosh, M. (1996). On invariance of noninformative priors. *Ann. Statist.*, 24(1):141–159.

Harville, D. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, 61:383–385.

Jeffreys, H. (1961). *Theory of Probability.* Oxford University Press, 3 edition.

Kass, R. and Wasserman, L. (1996). The selection of prior distributions by formal rules. *J. American Statist. Assoc.*, 431(91):1343–1370.

Lehmann, E. and Romano, J. (2005). *Testing Statistical Hypotheses.* Springer-Verlag, 3 edition.

Polson, N. and Wasserman, L. (1990). Prior distributino for the bivariate binomial. *Biometrika*, 77(4):901–904.

Rousseau, J. and Robert, C. (2005). Discussion on a paper of J. Bernardo: Intrinsic Credible Regions:An Objective Bayesian Approach to Interval Estimation. *Test*, 14(2):367–369.

Sun, D. and Berger, J. (1998). Reference priors with partial information. *Biometrika*, 77(4):901–904.

Unité de recherche INRIA Futurs
Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399