

# Invariant HPD credible sets and MAP estimators

Pierre Druilhet\* and Jean-Michel Marin†

**Abstract.** MAP estimators and HPD credible sets are often criticized in the literature because of paradoxical behaviour due to a lack of invariance 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 invariant frequentist ML estimators. We also propose several adaptations in the case of nuisance parameters.

**Keywords:** Bayesian statistics, HPD, MAP, Jeffreys measure, nuisance parameters, reference prior

## 1 Introduction

The Maximum A Posteriori estimator (MAP) is defined to be the value (not necessarily unique) that maximizes the posterior density w.r.t. the Lebesgue measure, denoted by  $\lambda$ . 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 invariant under smooth reparametrization. Because of this undesirable feature, many authors do not recommend their use.

Consider the following example:  $X|\theta \sim \mathcal{N}(\theta, \sigma^2)$  and  $\theta \sim \mathcal{N}(\mu, \tau^2)$  where  $\sigma^2$ ,  $\mu$  and  $\tau^2$  are assumed to be known. The posterior distribution of  $\theta$  is normal and

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

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

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

The lack of invariance for the MAP is mainly due to the choice of the Lebesgue measure as dominating measure, see [Lehmann and Romano \(2005\)](#) [section 5.7] and [Berger \(1985\)](#) [section 4.3.2]. 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  $\theta$  lies in an open subset  $\Theta$  of  $\mathbb{R}^p$ . We denote by  $\Pi$  the (possibly improper) continuous prior distribution

\*ENSAI, Rennes, France <mailto:druilhet@ensai.fr>

†Université d'Orsay, Laboratoire de Mathématiques (Bât. 425), INRIA FUTURS, Projet SELECT, 91405 Orsay, France [Jean-Michel.Marin@inria.fr](mailto:Jean-Michel.Marin@inria.fr)

on  $\theta$  and by  $\Pi_x$  the posterior distribution of  $\theta$  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  $\theta$  w.r.t.  $\lambda$ . Consider a new dominating measure  $\nu$  whose density w.r.t.  $\lambda$  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  $\theta$  w.r.t.  $\nu$ . We named  $\text{MAP}_\nu$  the MAP based on the dominating measure  $\nu$ , i.e.

$$\text{MAP}_\nu(\theta) = \underset{\theta \in \Theta}{\text{Argmax}} \pi_\nu(\theta|x) = \underset{\theta \in \Theta}{\text{Argmax}} \left[ \frac{\pi_\lambda(\theta|x)}{g(\theta)} \right]. \quad (1)$$

With this notation,  $\text{MAP} = \text{MAP}_\lambda$ . 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  $\theta$  through the posterior distribution is to use credible sets, regions to which  $\theta$  belongs with a given posterior probability. One way to choose such sets is to define Highest Probability Density credible sets (HPDs). Formally, a set  $\text{HPD}^\gamma(\theta) \subset \Theta$  is an HPD of level  $\gamma$  if there exists a constant  $k_\gamma$  such that

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

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:

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

From now on, to avoid unnecessary complicated writing, we assume that HPDs can always be written as in (2). It is well known that  $\text{HPD}^\gamma(\theta)$  minimizes the length (or volume in the multivariate case) among the credible sets of level greater or equal to  $\gamma$ , where the unit of length or volume is given by the Lebesgue measure  $\lambda$ . 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

$$\text{MAP}(\theta) = \bigcap_{0 < \gamma < 1} \text{HPD}^\gamma(\theta). \quad (3)$$

From now on, we omit the subscript  $\gamma$  in  $\text{HPD}^\gamma$ . As MAPs, HPDs are criticized for their lack of invariance 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  $\theta$  is

$$\text{HPD}(\theta) = \{\theta : \theta^x(1-\theta)^{1-x} \geq k_\gamma\}.$$

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

$$\text{HPD}(\alpha) = \left\{ \alpha : \frac{e^{(x+1)\alpha}}{(1+e^\alpha)^3} \geq k_\gamma \right\}.$$

Moreover, for the parametrization  $\beta = \theta/(1 - \theta) = \exp(\alpha) = \text{ODD}(\theta)$ ,

$$\text{HPD}(\beta) = \left\{ \beta : \frac{\beta^x}{(1 + \beta)^3} \geq k_\gamma \right\}.$$

Figure 1 presents the posterior densities of  $\theta$ ,  $\alpha$  and  $\beta$  when  $x = 1$ . The case  $x = 0$  is similar. In the original parameterization by  $\theta$ , the HPDs are one-sided, while for a monotonic reparametrization by  $\alpha$ , the HPDs become two-sided and obviously,  $\text{HPD}(\alpha) \neq \text{logit}(\text{HPD}(\theta))$ . For  $x = 1$ ,  $\text{MAP}(\theta) = 1$  which corresponds to  $\alpha = +\infty$ , whereas  $\text{MAP}(\alpha) = \log(2)$  which corresponds to  $\theta = 2/3$ . The case of the ODD parametrization is much more interesting. Indeed, for  $x = 1$ ,  $\text{MAP}(\beta) = 1/2$  (which corresponds to  $\theta = 1/3$ ) whereas one expected a MAP larger than 1. This is due to the fact that the ODD parametrization breaks the natural symmetry, around 1/2 for  $\theta$  and around 0 for  $\alpha$ . The Lebesgue measure as dominating measure does not take into account this change. As MAPs, HPDs are defined only through the density of the

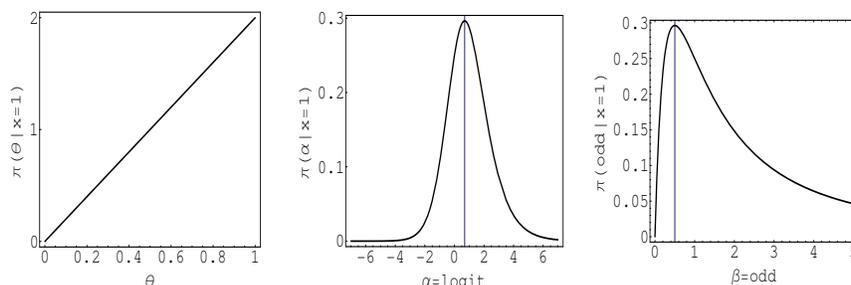


Figure 1: Posterior densities of  $\theta$ ,  $\alpha$  and  $\beta$  when  $x = 1$

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  $\nu$  as dominating measure, we can define a new HPD region, named  $\text{HPD}_\nu(\theta)$ , by

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

With this notation,  $\text{HPD}(\theta) = \text{HPD}_\lambda(\theta)$ . Note that  $\text{HPD}_\nu$  is the region of level  $\gamma$  with minimal length or volume where  $\text{length}_\nu(C) = \int_C d\nu(\theta)$ .

The aim of this paper is to discuss a new choice of dominating measure that makes MAPs and HPDs invariant 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 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 Invariant 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  $\theta$ , denoted by  $J_\theta$ , is the measure with density  $|I(\theta)|^{\frac{1}{2}}$  w.r.t.  $\lambda$ . We denote by  $\text{JMAP}(\theta) = \text{MAP}_{J_\theta}(\theta)$  the MAP obtained by taking the Jeffreys measure as dominating measure. Similarly, we denote by  $\text{JHPD}(\theta) = \text{HPD}_{J_\theta}(\theta)$  the HPD region with the Jeffreys measure as dominating measure. We have:

$$\text{JMAP}(\theta) = \underset{\theta \in \Theta}{\text{Argmax}} f(x|\theta) |I(\theta)|^{-\frac{1}{2}} \pi_\lambda(\theta), \quad (5)$$

and

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

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

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

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

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

and

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

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  $\text{JMAP}(\theta) = \text{MAP}_\lambda(\theta)$ . For  $\alpha = \exp(\theta)$ , we can derive  $\text{JMAP}(\alpha)$  from  $\text{JMAP}(\theta)$  by

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

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

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

For  $\alpha = h_1(\theta) = \log(\theta/(1 - \theta))$ , we have

$$\begin{aligned} \text{JHPD}(\alpha) &= \left\{ \alpha : \frac{\exp((x + 1/2)\alpha)}{(1 + \exp(\alpha))^2} \geq k_\gamma \right\} \\ &= h_1(\text{JHPD}(\theta)). \end{aligned}$$

For  $\beta = h_2(\theta) = \theta/(1 - \theta)$ , we have

$$\begin{aligned} \text{JHPD}(\beta) &= \left\{ \beta : \frac{\beta^{x+1/2}}{(1 + \beta)^2} \geq k_\gamma \right\} \\ &= h_2(\text{JHPD}(\theta)). \end{aligned}$$

Figure 2 presents the posterior densities of  $\theta$ ,  $\alpha$  and  $\beta$  w.r.t. the Jeffreys dominating measures. Contrary to the Lebesgue dominating measures case, we obtain two-sided regions. Moreover, we have  $\text{JMAP}(\theta) = 3/4$ ,  $\text{JMAP}(\alpha) = 3$  and  $\text{JMAP}(\beta) = \log(3)$ .

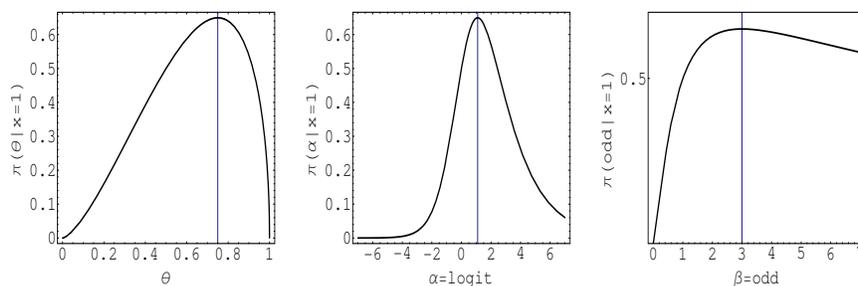


Figure 2: Posterior densities of  $\theta$ ,  $\alpha$  and  $\beta$  w.r.t. the Jeffreys measure when  $x = 1$

As a new example, let us consider a sample  $x_1, \dots, x_n$  ( $n > 2$ ) from an exponential distribution with mean  $1/\theta$  and denote by  $\bar{x}$  the sample mean. Let us suppose that the prior distribution on  $\theta$  is a Gamma distribution with parameter  $(a, b)$  (mean  $b/a$ ). It is very easy to see that

$$\text{MAP}(\theta) = \frac{n + a - 1}{b + n\bar{x}} \quad ; \quad \text{JMAP}(\theta) = \frac{n + a - 3}{b + n\bar{x}} \quad ;$$

and, if  $\beta = 1/\theta$ ,

$$\begin{aligned}\text{MAP}(\beta) &= \frac{b + n\bar{x}}{n + a + 1} \neq 1/\text{MAP}(\theta) \quad ; \\ \text{JMAP}(\beta) &= \frac{b + n\bar{x}}{n + a - 3} = 1/\text{JMAP}(\theta).\end{aligned}$$

We also consider a sample of size  $N$  from a right censored exponential distribution. The censoring occurs at time  $x_0$  and the number of observations greater than  $x_0$  is known and equal to  $m < N$ . Let  $n > 0$  be the number of observations less than  $x_0$  and denote these observations by  $x_1, \dots, x_n$ . Letting  $\theta$  be the failure rate, the likelihood function of  $\theta$  is given by

$$N!m!n!\theta^n \exp\left(-\theta \sum_{i=1}^n x_i - \theta m x_0\right).$$

It is well-known (see [Deemer and Votaw \(1955\)](#)) that the maximum likelihood of  $\theta$  is

$$\hat{\theta} = n \left( m x_0 + \sum_{i=1}^n x_i \right)^{-1}.$$

Moreover, for such a model, the Fisher information is such that

$$I(\theta) \propto \theta^{-2}(1 - \exp(-\theta x_0))$$

(see [Halperin \(1952\)](#)). Therefore, if we choose the Jeffreys measure as non-informative prior distribution for  $\theta$ ,  $\pi_\lambda(\theta) \propto \theta^{-1} \sqrt{1 - \exp(-\theta x_0)}$  and

$$\text{JMAP}(\theta) = \hat{\theta} = n \left( m x_0 + \sum_{i=1}^n x_i \right)^{-1}.$$

If we change the parametrization and consider the new parameter  $\beta = 1/\theta$  (where  $\beta$  corresponds to the mean of the exponential distribution)

$$\text{JMAP}(\beta) = 1/\hat{\theta} = \left( m x_0 + \sum_{i=1}^n x_i \right) / n.$$

The other motivation for the choice of  $J_\theta$  as dominating measure is that the Jeffreys measure is a classical non-informative prior for  $\theta$  ([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  $\theta$  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, when the posterior distribution is unimodal w.r.t. the Jeffreys measure, JHPDs can then be thought as credible sets “around” the ML.

### 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  $\theta$  is split into two parts:  $\theta = (\theta_1, \theta_2) \in \Theta_1 \otimes \Theta_2$  where  $\theta_1$  is the parameter of interest and  $\theta_2$  is the nuisance parameter. We denote by  $\pi_\nu(\theta_1|x)$  the density of the marginal posterior distribution of  $\theta_1$  w.r.t. the measure  $\nu$ . The corresponding  $\text{MAP}_\nu$  and  $\text{HPD}_\nu$  are

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

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

Because the Jeffreys prior does not distinguish between parameter of interest and nuisance parameter, [Bernardo \(1979\)](#) proposed a new approach called reference prior approach. In this section, we show how we can use this reference prior as dominating measure to define invariant 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  $\theta_2$  given  $\theta_1$ . We denote by  $\pi_\lambda(\theta_2|\theta_1)$  its density w.r.t.  $\lambda$ . In this case, [Sun and Berger \(1998\)](#) proposed two different approaches to derive reference priors for  $\theta_1$ . We mimic their approaches and there are two reasonable options for finding a dominating measure on  $\Theta_1$ .

*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  $\theta_1$  obtained from the marginal model. A dominating measure can be the Jeffreys measure on  $\Theta_1$  with density w.r.t.  $\lambda$  proportional to  $|I_m(\theta_1)|^{1/2}$ . In this case, the JMAP and the JHPD are defined by

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

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

$\text{JMAP}_1$  and  $\text{JHPD}_1$  are obviously invariant 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) proposed to maximize asymptotically the expected Kullback-Leibler divergence between the marginal posterior of  $\theta_1$  and the marginal prior of  $\theta_1$ . This leads to the distribution on  $\Theta_1$  with density w.r.t.  $\lambda$  proportional to

$$\exp \left\{ \frac{1}{2} \int \pi_\lambda(\theta_2|\theta_1) \log \left( \frac{|I(\theta_1, \theta_2)|}{|I_c(\theta_2|\theta_1)|} \right) d\theta_2 \right\} \quad (12)$$

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  $\theta_1$  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 distribution defined by Equation (12) as dominating measure on  $\Theta_1$ . In this case, the JMAP and the JHPD are defined by

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

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

$\text{JMAP}_2$  and  $\text{JHPD}_2$  are obviously invariant for a 1 – 1 smooth reparametrization on the parameter of interest.

Let us consider a sample  $X_1, \dots, X_n$  from a normal distribution with expectation  $\theta_2 = \mu$  and standard deviation  $\theta_1 = \sigma$ , the parameter of interest. This example is denoted as the normal nuisance example. Suppose that the conditional prior distribution for  $\mu$  given  $\sigma$  is normal with expectation  $m$  and variance  $\tau^2$ . Applying proposition 2 of Sun and Berger (1998), Option 1 dominating measure has density w.r.t.  $\lambda$  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.  $\lambda$  proportional to  $1/\sigma$ . These two dominating measures are different. However, when  $n \rightarrow \infty$ , the first density converges uniformly to the second one.

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

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

where  $\mathbb{I}_A$  is the indicator function on  $A$  and  $m$  is supposed to be known. Suppose that the conditional distribution of  $\theta_2$  given  $\theta_1$  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.  $\lambda$  proportional to  $\theta_1^{-1/2} (1-\theta_1)^{-1/2}$ .

Sun and Berger (1998) also considered the case where  $\theta_1$  and  $\theta_2$  are independent. For this other prior information, we can also mimic their approach to define a dominating measure on  $\Theta_1$ .

### 3.2 No subjective conditional prior available

If no subjective conditional prior for  $\theta_2$  given  $\theta_1$  is available, we propose to mimic the reference prior approach of Berger and Bernardo (1989, 1992). This leads to the dominating measure on  $\Theta_1$  with density w.r.t.  $\lambda$  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  $\Theta$  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  $\theta_1$ . 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  $\Theta_1$  has then density w.r.t.  $\lambda$  proportional to

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

where  $\theta_1^{(0)}$  is any fixed point in  $\Theta_1$ . Datta and Ghosh (1996) established the invariance of this procedure under 1–1 smooth reparametrization on  $\theta_1$ . Therefore, the corresponding JMAP and JHPD are invariant under a smooth reparametrization on the parameter of interest.

Let us consider again the normal nuisance example. Applying the previous procedure, the dominating measure on  $\sigma$  has density w.r.t.  $\lambda$  proportional to  $1/\sigma$ , which is the invariant measure for scale models. We assume now that no prior information is available on  $\sigma$ . So, the non-informative reference prior is given by  $\pi_\lambda(\sigma) = 1/\sigma$ . Equivalently, if the parameter of interest is  $\sigma^2$ , the reference prior and dominating measure have density w.r.t. Lebesgue proportional to  $\frac{1}{\sigma^2}$ , which is again the corresponding invariant measure. In that case  $\text{JMAP}(\sigma^2)$  is equal to the frequentist REML (REstricted Maximum Likelihood) estimator. This is a new interpretation of the REML estimator which also corresponds to the MAP under the Laplace prior for  $\sigma^2$  (Harville 1974). By invariance of the JMAP, we have,  $\text{JMAP}(\sigma) = \sqrt{\text{REML}}$ . If we change  $\sigma$  into  $\log(\sigma)$ , then the reference prior is the Lebesgue measure. In that case,  $\text{JMAP}(\log(\sigma)) = \text{MAP}(\log(\sigma)) = \frac{1}{2} \log(\text{REML})$ . 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.

## References

- Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York: Springer, 2 edition. 681
- Berger, J. and Bernardo, J. (1989). “Estimating a Product of Means: Bayesian Analysis With Reference Priors.” *J. American Statist. Assoc.*, 84(405): 200–207. 688, 689
- (1992). “Ordered Group reference Priors With Application to Multinomial and Variance Component Problems.” *Biometrika*, 79: 25–37. 688, 689
- Bernardo, J. (1979). “Reference Posterior Distributions for Bayesian Inference.” *J. Royal Statist. Soc. Series B*, 41(2): 113–147. 686, 687, 688
- (2005). “Intrinsic Credible Regions: An Objective Bayesian Approach to Interval Estimation.” *Test*, 14(2): 317–384. 684
- Crowder, M. and Sweeting, T. (1989). “Bayesian inference for a bivariate binomial.” *Biometrika*, 76: 599–604. 688
- Datta, G. and Ghosh, M. (1996). “On invariance of noninformative priors.” *Ann. Statist.*, 24(1): 141–159. 689
- Deemer, W. and Votaw, D. (1955). “Estimation of parameters of truncated or censored exponential distributions.” *Ann. Math. Statist.*, 26: 498–504. 686
- Halperin, M. (1952). “Maximum likelihood estimation in truncated samples.” *Ann. Math. Statistics*, 23: 226–238. 686
- Harville, D. (1974). “Bayesian inference for variance components using only error contrasts.” *Biometrika*, 61: 383–385. 689
- Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press, 3 edition. 686
- Kass, R. and Wasserman, L. (1996). “The selection of prior distributions by formal rules.” *J. American Statist. Assoc.*, 431(91): 1343–1370. 686
- Lehmann, E. and Romano, J. (2005). *Testing Statistical Hypotheses*. New York: Springer, 3 edition. 681
- Polson, N. and Wasserman, L. (1990). “Prior distribution for the bivariate binomial.” *Biometrika*, 77(4): 901–904. 688

- 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. [684](#)
- Sun, D. and Berger, J. (1998). “Reference priors with partial information.” *Biometrika*, 77(4): 901–904. [687](#), [688](#)

**Acknowledgments**

We thank Professor Christian Robert for his helpful comments and a careful reading of a preliminary version of the paper. We also thank Doctor Jean-Louis Foulley for helpful discussions on the Bayesian approach of variance component models.

