Diagnostics of prior-data agreement in applied Bayesian analysis

Nicolas Bousquet

INRIA Futurs and Paris XI University (France) and Laval University (Canada)

nicolas.bousquet.1@ulaval.ca

Summary. This article focused on the definition and the study of a binary Bayesian criterion which measures a statistical agreement between a subjective prior and data information. The setting of this work is concrete Bayesian studies. It is an alternative and a complementary tool to the method recently proposed by Evans and Moshonov (2006). Both methods try to help the work of the Bayesian analyst in preliminary to the posterior computation. Our criterion is defined as a ratio of Kullback-Leibler divergences ; two of its main features are to make easy the check of a hierarchical prior and be used as a default calibration tool to obtain flat but proper priors in applications. Discrete and continuous distributions exemplify the approach and an industrial case-study in reliability, involving the Weibull distribution, is highlighted.

Keywords. Prior-data conflict, Expert opinion, Subjective prior, Objective prior, Kullback-Leibler divergence, discrete distributions, lifetime distributions.

1. Motivation

Among the large number of industrial case-studies involving, in subjective Bayesian frameworks, the posterior distribution of the parameter $\theta \in \Theta$ of a decision-making model $\mathcal{M}(\theta)$, many of them mention its undesirable behavior when the prior information threatens to be *conflicting* with the information brought by the observed data $\mathbf{y_n} = (y_1, \ldots, y_n) \sim \mathcal{M}(\theta)$ (Evans and Moshonov 2006, 2007, Bonnevialle and Billy 2006). The term "conflicting", introduced by the first authors, has originally been specified as follows: the prior distribution favors region of Θ far from the frequentist confidence region brought by $\mathbf{y_n}$. The objective knowledge yielded by the likelihood $\mathcal{L}(\mathbf{y_n};\theta)$ can thus be littered by the choice of a wrong prior and the Bayesian analyst can take a posterior decision with unwelcome consequences. In case of highly censored data and small sample sizes, when the Bayesian informative approach is highly recommended (Robert 2001), the heterogeneity of the data or a high level of censoring can have the same effect as a poor prior with correct data. Thus prior knowledge and objective data knowledge may have symmetric roles in a conflict. Hence detecting a conflict is a warning for the Bayesian analyst: at least one of the two sources of information has to be carefully checked through a meta-analysis. Then it can be rejected or the conflict can be ignored. In decision chains where inferences are automated, like Bayesian networks, such warnings appear as possible guide rails. Besides, when industrial data are costly to collect and there is a consensus between expert opinions, the number of additional data necessary to reduce a conflict is a relevant concern.

Curiously, to our knowledge, this subject seems not to have been much studied in Bayesian statistics, although it constitutes an important preliminary to the inference. Notice however that in the strict observation of the Bayesian paradigm, the prior distribution and the sampling model constitute a whole decision model. Therefore it seems not relevant, from a theoretical point of view, to determine degrees of discrepancy between them. Thus the context of such a study remains clearly applied. Although focusing on the pre-inferential Bayesian framework, this concern is however to connect with the numerous works led on a reinforcement of the posterior robustness. Thus, functional structures of the prior and the sampling model have been studied to obtain negligible posterior influence in response to an increasing discrepancy. Main references are De Finetti (1961), Dawid (1982), Hill (1974), Lucas (1993) and especially O'Hagan (1979, 1988, 1990, 2003), Andrade and O'Hagan (2006) and Angers (2000). Some authors as Gelman *et al.* (1996) consider the simultaneous effects of the prior and the sampling model on a possible discrepancy with the data. Closer to our matter of concern, numerous authors have developed nonparametrical scoring rules to compare the accordance of the empirical repartition of data with various marginal distributions. See Gneiting *et al.* (2007) and Gneiting and Raftery (2007) for a review.

In a parametrical Bayesian framework, Idée *et al.* (2001) suggested using a Fisher test between prior and empiric measures of uncertainty in industrial studies, and an *ad hoc* approach using convolution products was proposed by Usureau (2001) without statistical justification. Finally, Evans and Moshonov (2006) summarized the two major steps of a "good" prior elicitation. First, one has to check the goodness of the model $\mathcal{M}(\theta)$ with respect to $\mathbf{y_n}$. Appropriate Bayesian methods can be found in Bayarri and Berger (2000). In most industrial cases this check is often missed because the parametric model is historically well-tried and Bayesian technics are used to overcome precision limits of the frequentist approach. Second, if no evidence of error is obtained, one has to check a possible prior-data conflict. Computing the $p-$value of an observed sufficient statistic with respect to its predictive distribution, Evans and Moshonov (2006) gave a first solution (called EMo in the following) to this requirement.

One issue raised by their approach is the absence of a binary answer for the Bayesian analyst: how to define a clear threshold of conflict or agreement? We propose to use the following criterion, which settles this issue. Compute the ratio

$$\mathrm{DAC}^J(\pi|\mathbf{y_n}) \quad = \quad \frac{KL\left(\pi^J(.|\mathbf{y_n}) \;||\; \pi\right)}{KL\left(\pi^J(.|\mathbf{y_n}) \;||\; \pi^J\right)} \tag{1}$$

where $KL(\pi_1||\pi_2)$ is the Kullback-Leibler divergence between distributions π_1 and π_2

$$KL(\pi_1||\pi_2) = \int_\Theta \pi_1(\theta) \log \frac{\pi_1(\theta)}{\pi_2(\theta)} \, d\theta$$

and π^J is a privileged noninformative prior for the study. Acronym DAC means *data-agreement criterion.* If $\mathrm{DAC}^J(\pi|\mathbf{y_n}) \leq 1$, prior and data-given confidence regions for θ are close enough and the prior proposal π is in agreement with $\mathbf{y_n}$. Else a conflict is defined. This new definition is easily transposable on any function of interest $g(\theta)$ with associated prior $\pi(g(\theta))$.

The remainder of the paper is organized as follows. The EMo approach is briefly analyzed in Section 3. In Section 4, we detail the arguments which lead to the choice of this criterion and give some properties when the prior is hierarchically specified or defined as a combination of priors. When π^J is proper, some ideal examples are treated and relevant features of DAC are highlighted. When π^J is improper, DAC cannot be computed and an *intrinsic adaptation* is defined in Section 5, similarly to the pseudo-Bayes factor in Bayesian model selection. We show how these two areas can be connected to justify this adaptation. Comparisons are done between the ideal and the approximated form of the criterion and some improvements are pointed up. In Section 6, we briefly focus on using DAC as a calibration tool subjective priors, providing objective variance bounds or default values. Finally, some research avenues are suggested in a discussion section. Along the paper, numerical examples are considered to illustrate the behavior of DAC and compare the information provided by DAC and EMo. Especially, applications to the Weibull lifetime model are highlighted, using real data and expert opinions. The two criteria are shown to lead to complementary diagnostics that are useful for the Bayesian analyst to improve the prior elicitation.

2. Notation

Here we introduce some general notation that will be used in the article. For $n \in I\!N^*$, let $\mathbf{X_n} = X_1, \ldots, X_n \sim \mathcal{M}(\theta)$ be independently and identically distributed (i.i.d.) real-or-vector-valued random variables in the sample space χ^n with a probability density function (pdf) $f(x|\theta)$ and a distribution function $F(x|\theta)$. Denote $S(x|\theta) = 1 - F(x|\theta)$. The pdf is defined with respect to a dominating (unwritten) measure which is usually Lebesgue. However, some of our examples will use discrete measures. Let $\theta \in \Theta$ and denote $d = \dim \Theta < \infty$. In Section 7 we will consider a case where the observed sample $\mathbf{y_n}$ contains r uncensored i.i.d. data $\mathbf{x_r} = (x_1, \ldots, x_r)$ following $\mathcal{M}(\theta)$ and $n - r$ fixed (type-I) right-censored values, denoted $\mathbf{c_{n-r}} = (c_1, \ldots, c_{n-r})$. Thus, the observed likelihood can be written as

$$\mathcal{L}(\mathbf{y_n}; \theta) = \prod_{i=1}^{r} f(x_i|\theta) \prod_{j=1}^{n-r} S(c_i|\theta).$$

Similarly to f, any prior or posterior measure on θ will be denoted π and is dominated by a discrete or continuous reference (unwritten) measure on Θ. Remind that π (and by extension the prior distribution) is said *proper* if and only if π is a density, i.e. $\int_\Theta \pi(\theta) d\theta = 1$.

3. The EMo procedure

To our knowledge, Evans and Moshonov's work (2006) seems to be the first one dedicated to check prior-data conflicts that lays on statistical foundations rather than rules of thumb. They consider the marginal prior distribution M_T of a *minimal sufficient statistic* T with sampling density $f_T(t|\theta)$. This distribution has density

$$m_T(t) = \int_\Theta \pi(\theta) f_T(t|\theta)\ d\theta.$$

If the observed statistic $t_o = t(\mathbf{y_n})$ is a surprising value for M_T, namely when the marginal $p-$value

$$F_T(t_o) = M_T \left\{ m_T(t) \leq m_T(t_o) \right\}$$

is extreme, then a conflict is detected. This typically occurs when $F_T(t_o) \leq 0.05$ or $F_T(t_o) \geq 0.95$. In this case, as these authors say, "the data provide little or no support to those values of θ where the prior places its support". A difficulty however occurs when a component U of T is an ancillary statistic, namely a statistic whose distribution does not depend on θ (Ghosh *et al.* 2007). Thus no prior-data conflict concerning U can be highlighted and the marginal distribution threatens to reflect more the behavior of the sampling model $f_T(.|\theta)$ than the prior distribution. Hence it is necessary to compute the $p-$value of the conditional marginal distribution

$$m_T(t|U) = \int_\Theta \pi(\theta) f_T(t|\theta, U)\ d\theta.$$

When the prior is hierarchised (Evans and Moshonov 2007), similar conditional checks have to be carefully set up. Despite the difficulty to choose good sufficient and ancillary statistics, and the computational complexity which rises when no sufficient statistic exists (when $\mathcal{M}(\theta)$ does not belong to the natural exponential family), this method is an intuitive, powerful tool, whose performance is shown throughout numerous examples in the two articles previously cited.

There is however a difficulty, for the Bayesian analyst, to work with $p-$values. A wrong but common idea is to consider them as probabilities of conflict between $\theta \sim \pi$ and $\mathbf{y_n} \sim \mathcal{M}(\theta)$. Seen as a decision tool in a test, the $p-$value is a random variable following a uniform distribution under the null hypothesis. As Bayarri and Berger (2000, 2003) recommend, $p-$values must be carefully used; the understanding of the result could thus be mistaken for industrial analysts that are not statisticians. Another issue is that a binary definition of a conflict can be preferred in applied studies (Bonnevialle and Billy 2006), especially when data can be removed or added in sensitivity studies. Choosing a couple of $p-$values threshold can be difficult: why should we choose $(5\%, 95\%)$ percentiles rather than $(2\%, 98\%)$? We propose in next section another definition of conflict which settles this issue. When the EMo conflict is uniquely defined in term of location in the sample space through the choice of a sufficient statistic, which induces indirectly a conflict in the parameter space, our conflict is directly defined in term of location and uncertainty in this same parameter space.

4. A criterion of prior-data conflict

4.1. Definition and first examples

Our motivation here is to define what could be a Bayesian conflict in the parameter space Θ, when information comes from independent subjective (experts) and objective (data) sources. First we make the following assumption.

Assumption A. *There is always a unique noninformative benchmark prior π^J for the inference problem.*

This assumption can appear somewhat vague, but a large amount of work has been dedicated to the elicitation and the choice of noninformative priors in applications (Kass and Wasserman 1996). The choice between noninformative priors lays on criteria like invariance to reparametrization or group actions, entropy or missing information maximization. The coverage matching properties of the priors (Ghoshal 1999) allow to discriminate between alternative candidates (Robert 2001 chap. 2 and 8). Again, because the Bayesian approach is often used in industry to overcome the precision limits of the frequentist approach, the best regularizer of the frequentist results should be considered as an intuitive benchmark. From this point of view, choosing coverage matching priors of maximal order seems logical. But more generally, in applied studies, it seems reasonable to assume that a convenient or intuitive prior measure $\pi^J(\theta)$ can always be defined when no expert opinion is available (we consider implicitly that $\pi^J(\theta|\mathbf{y_n})$ is proper).

It seems reasonable too that π^J, as a noninformative prior, should always stay not conflicting with any observed data $\mathbf{y_n} \sim \mathcal{M}(\theta)$, although the sense of a conflict remains fuzzy yet. Furthermore, Bernardo (1997) informally proposed to use $\pi^J(\theta|\mathbf{y_n})$ as a benchmark prior for the study of subjective assessments. Actually, it can be regarded as the prior density on θ of a *fictitious expert perfectly in agreement with* $\mathbf{y_n}$. This idea of "an ideal expert" goes in the same sense than the "ideal forecaster" who enables to elicit calibrating criteria in predictive assessments (Gneiting *et al.* 2007). Our view can be formalized in Assumption B.

Assumption B. *An indicator of conflict between an informative prior and observed data, which increases with the level of conflict, is minimal when* $\pi(.) \equiv \pi^J(.|\mathbf{y_n})$.

Then observing a conflict boils down to observe a large distance between $\pi^J(\theta|\mathbf{y_n})$ and $\pi(\theta)$, independently of any parametrization choice. An *informative regret* between the ideal prior $\pi^J(.|\mathbf{y_n})$ and the assessed prior π can be defined by $KL(\pi^J(.|\mathbf{y_n}) \mid\mid \pi)$. Indeed, $KL(\pi_1||\pi_2)$ states the regret due to the choice of π_2 when the true distribution is π_1 (Cover and Thomas 1991). If π is such that the regret is large, $\pi(\theta)$ will be considered as too far from the data information on θ. Now it is necessary to choose a constant C such that if $KL\left(\pi^J(.|\mathbf{y_n}) \mid\mid \pi\right) > C$, π is declared in conflict with the data. The rationale for choosing C is as follows: assume that $KL(\pi^J(.|\mathbf{y_n}) \mid\mid \pi) > KL(\pi^J(.|\mathbf{y_n}) \mid\mid \pi^J)$. This is possible only

a) when π is less informative than π^J, which is false under Assumption A; or
b) when π is informative and favors regions of Θ that are far from data-confidence regions, or when the prior information on θ is considerably more precise than the data information and directs the posterior behavior. This latter case, in subjective Bayesian inference, leads to chiefly subjective decision-taking and should be avoided.

Consequently, it leads to choose $C = KL\left(\pi^J(.|\mathbf{y_n}) \mid\mid \pi^J\right)$. We obtain finally the normalized criterion (1) and the following rule: π is said *conflicting* with $\mathbf{y_n}$ if $\text{DAC}^J(\pi|\mathbf{y_n}) > 1$. The superscript J reflects the choice of π^J for the problem. In the following, we omit sometimes this notation for simplicity. Notice that, clearly, Expression 1 is not restricted of using the KL-divergence and another choice in the Ali-Silvey class of information-theoretic measures (Ali and Silvey 1966) can be made. However a significant feature of the KL-divergence is its invariance to reparametrization. Other justifications can be found in Hartigan (1998) and Sinanović and Johnson (2007), from both computational and analytic viewpoints.

Requirements. Expression (1) is well defined when π^J is proper. This is true when Θ is discrete or bounded. Notice that bounded priors are common in industrial settings because of running constraints (physical impossibilities, etc.). Section 5 is dedicated to the cases where π^J remains improper and DAC must be adapted.

Example 1. *Location normal model.*

Let $\mathbf{y_n}$ be an i.i.d $n-$sample from a $\mathcal{N}(\theta, 1)$ distribution with $\theta \in D = [T_l, T_m]$. Denote θ_0 the real value of θ. We place on θ a $\mathcal{N}(\mu_0, \sigma_0^2)$ prior ; π^J is chosen as the uniform prior $\pi^J(\theta) = (T_m - T_l)^{-1}\mathbb{1}_D(\theta)$. Then $\pi^J(\theta|\mathbf{y_n})$ is the $\mathcal{N}(\bar{y}_n, 1/n)$ density restricted on D. Choosing $\theta_0 = 0$, $n = 5$, $T_m = -T_l = 15$ and $\sigma_0 = 1$, we consider the evolution of DAC with respect to μ_0 in Figure 1, for several values of σ_0. Results are averaged on 30 simulated samples. [*Temporarily, for readability, all figures and tables have been placed in Section 10.*] A symmetric evolution around θ_0 appears natural, while the length of the agreement domain decreases when σ_0 increases (i.e., when the prior becomes more and more informative). DAC is minimal when $\pi \equiv \pi^J(.|\mathbf{y_n})$. ■

Example 2. *Bernoulli model.*

Let $\mathbf{y_n}$ be an i.i.d. $n-$sample from a Bernoulli distribution $\mathcal{B}_r(\theta)$. We assume on θ a prior Beta distribution $\mathcal{B}_e(\alpha, \beta)$ on $[0, 1]$. Note $\delta_n = \sum_{i=1}^n y_i$. In this one-dimensional case, the Jeffreys prior is the most common choice of a benchmark prior (Clarke 1996). Then π^J is the $\mathcal{B}_e(1/2, 1/2)$ density. Hence $\pi^J(\theta|\mathbf{y_n})$ is the $\mathcal{B}_e(\delta_n + 1/2, n - \delta_n + 1/2)$ density. KL-divergences are explicit and can be found in Penny (2001). Like Evans and Moshonov (2006, ex.2), we choose $(\alpha, \beta) = (5, 20)$ (so that $\mathrm{E}[\theta] = 0.2$) then we generate a sample of size $n = 10$ from the $\mathcal{B}_r(\theta_0 = 0.9)$ distribution: we obtain $\mathrm{DAC}^J(\alpha, \beta|\mathbf{y_n}) = 1.102$ and we conclude to a prior-data conflict ; modifying $\theta_0 = 0.25$, we obtain $\mathrm{DAC}^J(\alpha, \beta|\mathbf{y_n}) = 0.1026$ which means an agreement (as expected). Modifying π to be uniform ($\alpha = \beta = 1$), we found a global agreement of any data set. All these results are similar to the EMo results (using usual threshold $p-$values).

However, EMo is less restrictive than DAC is this sense it rejects less priors. This behavior was noticed on all tested models. We set $\theta_0 = 0.7$ and two sizes $n = 10$ and $n = 5$. Assessing the prior standard deviation to 0.2, we compute the range of values for the prior mean $\mu_0 = \alpha/(\alpha + \beta)$ such that π is not conflicting. We obtain $\mu_0 \in [0.2, 0.96]$ for $n = 10$ and $\mu_0 \in [0.23, 0.94]$ for $n = 5$

using DAC. Using EMo, the 10%-90% percentile domain is $\mu_0 \in [0.054, 0.928]$ for $n = 10$ and $\mu_0 \in [0.056, 0.925]$ for $n = 5$. The 5%-95% domain remains close to $[0.048, 0.95]$. ■

4.2. Main features

Hierarchically specified priors. One can desire to check a potential conflict with the data when the prior is hierarchically specified. A key but uncomfortable requirement for the use of EMo is the existence of statistics which are ancillary for parts of the parameter. Defining a conflict with DAC, the following proposition makes easy the separate checks of the hierarchical levels and links them to the check of the whole prior. Thus, the agreement of the full prior is not a sufficient condition to obtain the agreement of any hierarchical prior. The opposite is clearly true. The proof is obvious and not reported here. An application is given in § 7.

PROPOSITION 1. *Assume that* π *and* π^J *can be hierarchically written* $\pi(\theta) = \pi(\theta_1|\theta_2)\pi(\theta_2)$ *and* $\pi^J(\theta) = \pi^J(\theta_1|\theta_2)\pi^J(\theta_2)$. *Denote* $\tilde{\pi}_1(\theta) = \pi(\theta_1|\theta_2)\pi^J(\theta_2)$ *and* $\tilde{\pi}_2(\theta) = \pi^J(\theta_1|\theta_2)\pi(\theta_2)$. *Then*

$$DAC(\pi|\mathbf{y_n}) \quad = \quad DAC(\tilde{\pi}_1|\mathbf{y_n}) + DAC(\tilde{\pi}_2|\mathbf{y_n}) - 1.$$

Combination of multiple priors. Suppose to have checked m priors $\pi_1, \ldots, \pi_m$. Next proposition indicates that any geometric weighted combination of the priors π_i does not need to be checked if all priors are in agreement with the data. Possibly, some priors can be conflicting while the combination stays in agreement. Such a combination is typically used to obtain a global prior when several independent experts are available (Budescu and Rantilla 2000). The proof is simple, using generalized Hölder inequalities, and can be found in Bousquet (2006).

PROPOSITION 2. *Let* $\pi_1, \ldots, \pi_m$ *be* m *priors on* Θ, *and let* $\alpha_1, \ldots, \alpha_m$ *be* m *weights such that* $0 < \alpha_i < 1$, $\sum_{i=1}^m \alpha_i = 1$. *Denote* $\pi(\theta) \propto \prod_{i=1}^m \pi_i(\theta)^{\alpha_i}$. *Then*

$$DAC^J(\pi|\mathbf{y_n}) \quad \leq \quad \sum_{i=1}^m \alpha_i \; DAC^J(\pi_i|\mathbf{y_n}).$$

Asymptotic behavior. The asymptotic behavior (however of limited interest in subjective Bayesian statistics) is usually roped in to evaluate the coherence of the approach. Under classical regularity conditions on prior densities and likelihood, the asymptotic normality of $\pi^J(.|\mathbf{y_n})$ (Hartigan 1983) should make numerator and denominator of DAC take close values. Then DAC could be expected to tend to 1 when n takes large values (asymptotic agreement). In term of decision-making, since the posterior influence of the prior becomes negligible, this result appears reassuring for the Bayesian analyst. A more precise result follows, in the case of a binomial model (Ex. 3). Beyond the scope of this paper, general results should come out of strong arguments like asymptotic uniform integrability (van der Vaart 1998).

Example 3. *Bernoulli model.*
Let us take back the frame of Ex. 2. Denoting $0 < \theta_0 < 1$ the true value of the parameter, we obtain the following proposition (proof given in Appendix).

PROPOSITION 3. $\forall(\alpha,\beta) > 0$, *for any* $0 < q < 1$, *if* $\mathbf{X_n} \overset{i.i.d.}{\sim} \mathcal{B}_r(\theta)$, *then*

$$(\log n)^q \, E\left[DAC^J(\alpha,\beta|\mathbf{X_n}) - 1\right] \quad \xrightarrow{n\to\infty} \quad 0. \;\blacksquare \tag{2}$$

5. An intrinsic adaptation of DAC

Here we focus especially on the case where π^J is improper and defined up to an unknown multiplicative constant c_j. This constant has an additive effect in the denominator of DAC

$$KL\left(\pi^J(.|\mathbf{y_n}) \mid\mid \pi^J\right) \quad = \quad -\log c_j + KL\left(\pi^J(.|\mathbf{y_n}) \mid\mid \pi'^J\right)$$

where $\pi'^J(\theta) \propto \pi^J(\theta)$. Thus DAC^J is not well defined. In this section, we link this difficulty to a typical issue of Bayesian model selection. Some proposals are done to get around this issue, including the *intrinsic heuristic*. Thus we propose an intrinsic version DAC^{AIJ} of the criterion. This adaptation is compared to the ideal case and EMo through two examples. Issues raised by the adaptation and possible improvements are discussed too.

5.1. *The intrinsic heuristic*

First we show how the additive constant c_j, can be regarded as the multiplicative constant of a Bayes factor. Formally, we must have

$$\exp\left\{KL\left(\pi^J(.|\mathbf{y_n}) \mid\mid \pi\right) - KL\left(\pi^J(.|\mathbf{y_n}) \mid\mid \pi^J\right)\right\} \quad \leq \quad 1. \tag{3}$$

Denoting the marginal measures $m^J(x) = \int_\Theta f(x|\theta)\pi^J(\theta)\, d\theta$ and $m(x) = \int_\Theta f(x|\theta)\pi(\theta)\, d\theta$, the left member of inequation (3) takes the form

$$m^J(\mathbf{y_n}) \exp\left\{\int_\Theta \pi^J(\theta|\mathbf{y_n}) \log \frac{\pi^J(\theta|\mathbf{y_n})}{\pi(\theta) f(\mathbf{y_n}|\theta)} \; d\theta\right\} \tag{4}$$

which does not need the computation of the posterior $\pi(.|\mathbf{y_n})$ density. Then we can write the criterion under the alternative form

$$\text{DAC}_2^J(\pi|\mathbf{y_n}) = B_{J,\pi}(\mathbf{y_n}) \exp\left(KL \left\{ \pi^J(.|\mathbf{y_n}) \mid\mid \pi(.|\mathbf{y_n}) \right\} \right) \tag{5}$$

where $B_{J,\pi}(\mathbf{y_n})$ is the Bayes factor

$$B_{J,\pi}(\mathbf{y_n}) = \frac{m(\mathbf{y_n})}{m^J(\mathbf{y_n})}.$$

Thus, $\theta \sim \pi$ is not conflicting with $\mathbf{y_n} \sim \mathcal{M}(\theta)$ if $\text{DAC}_2^J(\pi|\mathbf{y_n}) \leq 1$. There is some interest to intepret c_j as the multiplicative unknown constant of a Bayes factor. Indeed, it is a relevant issue in objective model selection about which a huge literature is dedicated. See Andrieu *et al.* (2001) for a review. Numerous approaches have been proposed to obtain default Bayes factors. The most famous is defining *intrinsic Bayes factors* (IBF, Berger and Pericchi 1996). This methodology is deeply linked to the notion of *minimal training samples* (MTS) took among the observed data: a data subset $x(l) = (x_1, \ldots, x_q) \in \chi^q$ is a MTS if the *posterior prior* (Pérez and Berger 2002) $\pi^J(\theta|x(l))$ is proper and no subset of $x(l)$ leads to a proper posterior. A MTS is thus the minimal quantity of data for which all parameters in the model are identifiable. Then the IBF is defined as the Bayes factor conditional to a MTS $x(l)$ by

$$B_{J,\pi}^{IBF} = B_{J,\pi}(\mathbf{y_n}) B_{\pi,J}(x(l))$$

where $B_{\pi,J}(x(l)) = m_J(x(l))/m(x(l))$. By construction, $B_{J,\pi}^{IBF}$ removes the arbitrariness in the choice of c_j. To reduce the dependence on MTS, using the *arithmetic* and *expected arithmetic* IBF,

$$B_{J,\pi}^{AI}(\mathbf{y_n}) = B_{J,\pi}(\mathbf{y_n}) \frac{1}{L} \sum_{l=1}^{L} B_{\pi,J}(x(l)) \quad \text{and} \quad B_{J,\pi}^{EAI}(\mathbf{y_n}) = B_{J,\pi}(\mathbf{y_n}) \frac{1}{L} \sum_{l=1}^{L} \text{E}_{\hat{\theta}_n} \left[B_{\pi,J}(x(l)) \right],$$

makes sense (with $\hat{\theta}_n$ the MLE and $\text{E}_\theta[.]$ the expectation under $\mathcal{M}(\theta)$). Other averages may be used as the *geometric* IBF or the *median* IBF (Berger and Pericchi 1998). Finally, a nice property of the arithmetic IBF is its asymptotic equivalence with a "proper" Bayes factor arising from neutral *intrinsic priors*. This is a strong justification of the heuristic. For more precisions see Dass (2001).

Thus the intrinsic heuristic is based on the use of small quantities of training data, which are chosen among the observed data (explaining the term *intrinsic*), to redefine a statistic which is formally valid but remains, in fact, difficult to assess. In the following, we are more interested in adapting DAC^J than adapting DAC_2^J. Firstly because our context is not model selection, secondly because we want to preserve the intuitive sense of the criterion and its useful features listed in § 4.2.

5.2. Adapting DAC

Denote $x(l)$ a MTS taken among the data $\mathbf{y_n}$ and denote $\mathbf{y_n}(l) = \mathbf{y_n}/x(l)$. Denote L the number of available MTS. The divergence

$$KL\left\{\pi^J(.|\mathbf{y_n}(l)) \mid\mid \pi^J(.|x(l))\right\}$$

is now the maximal value for the divergence between $\pi^J(.|\mathbf{y_n}(l))$ and the assessed prior π conditioned to the learning information $x(l)$. This conditioning must be done in the numerator as in the denominator of DAC to attenuate the impact of the disturbing information yielded by the MTS. In order to reduce the dependence on $x(l)$, we use a cross-validation argument which leads to define the *intrinsic (arithmetic) DAC criterion* by

$$\mathrm{DAC}^{AIJ}(\pi|\mathbf{y_n}) \quad = \quad \frac{1}{L}\sum_{i=1}^{L} \frac{KL\left\{\pi^J(.|\mathbf{y_n}(l)) \mid\mid \pi(.|x(l))\right\}}{KL\left\{\pi^J(.|\mathbf{y_n}(l)) \mid\mid \pi^J(.|x(l))\right\}}. \tag{6}$$

Then, if the size of the MTS remains small with respect to n, DAC^{AIJ} remains small when $\pi(\theta)$ is close to $\pi^J(\theta|\mathbf{y_n})$. When π is very little informative and is arbitrary chosen close to π^J, $\mathrm{DAC}^{AIJ}(\pi|\mathbf{y_n}) \simeq 1$, similarly to DAC^J. It is easy to show that the features described in § 4.2 are preserved.

Example 4. *Bernoulli model.*

Let us take back the frame of Ex. 2 again. We choose $\theta_0 = 0.7$ as the true value of θ. Since the Jeffreys prior is proper here, we can compare DAC^J and DAC^{AIJ} through the choice of the prior mean $\mu_0 = \alpha/(\alpha+\beta)$, fixing the prior standard deviation at 0.2. Figures 2 and 3 show the evolution of both DAC criteria, averaged on 30 simulated samples of sizes $n = 20$ (Figure 2) and $n = 10$ (Figure 3). They illustrate that DAC^{AIJ} and DAC^J can produce close agreement domains. ■

Example 5. *Exponential model.*

A 10-sized i.i.d. dataset $\mathbf{y_n}$ from an exponential distribution with rate $\lambda_0 = 150^{-1}$ has been sampled: $\mathbf{y_n} = (142.76, 142.99, 470.3, 419.09, 185.20, 84.41, 8.13, 27.15, 573.17, 17.12)$. The usual minimal sufficient statistic is the maximum likelihood estimator (MLE) $\bar{y} = \hat{\lambda}_n^{-1} = 207$. Choosing $\pi^J(\lambda) \propto \lambda^{-1}$ as the standard Jeffreys prior, a MTS is a single value. We choose the conjugate prior

$$\lambda \quad \sim \quad \mathcal{G}\left(a, a x_e\right)$$

in which a embodies the size of a *virtual* sample of mean x_e (Robert 2001, chap.3) and $\mathcal{G}(\alpha,\beta)$ is the gamma distribution with mean α/β. Notice π is perfectly in agreement with $\mathbf{y_n}$ if $a = n = 10$ and $x_e = \bar{y} = \hat{\lambda}_n^{-1} = 207$. We give in Table 1 the values of DAC^{AIJ} for several prior scenarios. The approximation rejects informative priors for which x_e is far from the data summarized by $\bar{y}$. Evolutions of DAC^{AIJ} are displayed on Figures 4 and 5. However, a major difference with EMo appears: using an information-theoretic distance between benchmark priors and the assessed prior makes the criterion reject priors that remain close to the data but are disproportionately informative with respect to $\pi^J(.|\mathbf{y_n})$ (cf Fig. 5 and § 4.1). In other words, DAC^{AIJ} as DAC^J discard unbiased priors with very small variances and very biased priors with reasonable variances. Moreover, we provide in Table 2 the agreement domains for x_e computed by DAC^{AIJ} and EMo. Explicit writing of the $p-$value is given in Bousquet (2006) in a general censored case. The agreement domain is the $(5\%, 95\%)$ percentile interval of the marginal distribution of the usual minimal sufficient statistic. We observe, similarly to Example 2, that EMo accepts biased priors that are strongly rejected by DAC. ■

5.3. Issues and possible improvements

In an industrial context where n can be small and the data can be censored, the number L of available MTS of size q may be problematic since DAC^{AIJ} can remain strongly dependent on some MTS containing outliers. Hence it is desirable to increase the number of MTS. Berger and Pericchi (2002) propose several ideas in this sense (especially when sufficient statistics can summarize the data information) and introduce the notion of *sequential* MTS (SMTS), including censored data.

Using the special "censored" Jeffreys prior π_c^J defined by De Santis *et al.* (2001), instead of a standard noninformative prior π^J, is a practical alternative for simple models as the exponential distribution: a censored data can become a MTS. However, the size of a MTS can remain high with respect to n, especially when $\dim\Theta$ increases. Bousquet and Celeux (2006) proposed to modify the posterior priors into *pseudoposterior priors* using the fractional likelihoods defined by O'Hagan (1995, 1997). The noisy information carried through such priors can be considerably lower than the information carried through simple posterior priors. Applications have been done on lifetime models but more work is required to be generalized.

Notice the computational cost of the intrinsic adaptation: except in conjugate cases, L posterior samplings of $\pi^J(\theta|\mathbf{y_n}(l))$ are needed to obtain Monte Carlo estimates of DAC^{AIJ}. Notorious MCMC algorithms (Robert and Casella 2004) can be used but importance sampling methods (Cappé *et al.* 2004) are more appropriate since the instrumental sampling function and the importance weights can be reused, provided $\pi^J(\theta|\mathbf{y_n}(l_i))$ stays not far from $\pi^J(\theta|\mathbf{y_n}(l_j))$ $(i \neq j)$. Then the computational cost can be reduced.

Finally, summarizing our experiments using discrete models, what seemed a reasonable approximation in 1- and 2-dimensional cases (less than 10% L^2 relative error between agreement domains computed using DAC^J and DAC^{AIJ}, the prior variance being fixed) needed at least $L \geq 10$ and $n > 5q$. Such empirical results have to be refined in a large variety of models.

6. Help to prior calibration

In previous examples we considered that π was entirely assessed. Its dispersion was set and a prior pointwise estimation θ_e of θ (a central value as the prior mean) was checked with respect to the data location. Thus, we obtained an agreement domain for θ_e. When elicited from an expert subjective opinion, θ_e reflects a personal viewpoint and is usually easy to assess (Daneskhah 2004). But common prior uncertainty measures are more difficult to set. A prior elicited from the expert opinion, without critical work from the Bayesian analyst, can be strongly and dangerously informative (Garthwaite *et al.* 2005). Besides it can happen that *no credible information* is available on the expert opinion. In those cases, a default rule for setting or limiting the prior uncertainty in a proper way is desirable. DAC can help to answer this question. Denoting ω the prior hyperparameter, default or boundary values for ω can be found such that

$$\mathrm{DAC}(\omega|\mathbf{y_n}) \quad = \quad 1$$

under the constraint $g(\omega) = \theta_e$ where, typically, $g(\omega) = \mathrm{E}[\theta]$. Acting in such a way, we choose the most informative prior in accordance with the expert opinion and not conflicting with the data. This trade-off was first noticed in Ex. 5.

Example 6. *Exponential model.*
Let us take back the frame of Ex. 5. Consider again the prior $\lambda \sim \mathcal{G}(a, ax_e)$ where a is the size of a virtual sample with mean x_e and variance ax_e^2. Thus a embodies the expert uncertainty: calibrating π is choosing a. Denote $\hat{a}$ the strictly positive value such that $\mathrm{DAC}^{AIJ}(\hat{a}|\mathbf{y_n}) = 1$. Existence and unicity of $\hat{a}$ can be proved using convexity arguments (Bousquet 2006). The variations of $\hat{a}$ in function of x_e, using the data from Ex. 5, are displayed in Figure 6. A boundary line is placed at $a = n = 10$ since n is a natural upper bound for a such that the posterior distribution stays mainly directed by objective data information. Combining both limits gives to the Bayesian analyst a more precise view of the expert "reasonableness". Thus, if x_e is far from $\bar{y} = 207$, the analyst can select default vague priors. ■

7. A recapitulative example

We consider the right-censored real lifetime data $\mathbf{y_n}$ ($n = 18$) from Table 3. They correspond to failure times or stopping times collected on some similar devices belonging to the secondary water circuit of nuclear plants (Bousquet and Celeux 2006). For physical reasons and according to a large consensus, those data are assumed to arise from a Weibull distribution $\mathcal{W}(\eta, \beta)$ with density

$$f(x|\eta,\beta) = \frac{\beta}{\eta}\left(\frac{x}{\eta}\right)^{\beta-1}\exp\left\{-\left(\frac{x}{\eta}\right)^{\beta}\right\}\mathbb{1}_{\{x\geq 0\}}.$$

The MLE is $(\hat{\eta}_n, \hat{\beta}_n) = (140.8, 4.51)$ with estimated standard deviations $\hat{\sigma}_n = (7.3, 1.8)$. The high value of $\hat{\beta}_n$ is unexpected because it reflects an unreasonable aging of the device (Dodson 2006). Two prior opinions on the lifetime are available, given by independent experts $\mathcal{E}_1$ and $\mathcal{E}_2$. They are summarized in Table 4. $\mathcal{E}_1$'s opinion is much more informative than $\mathcal{E}_2$'s and both are right-shifted with respect to the data. Moreover the experts are not questioned at the same precision level. $\mathcal{E}_1$ is a nuclear operator and speaks for a particular component while $\mathcal{E}_2$ can be seen as a component producer whose opinion takes into account a variety of running conditions. Since the Weibull distribution does not admit conjugate continuous prior (Soland 1969), the posterior computation needs numerical approximations (Singpurwalla and Song 1988). In our applications, we used adaptive importance sampling dedicated to missing data problems (Celeux *et al.* 2003).

We consider the priors

$$\begin{cases} \eta & \sim & \mathcal{G}(a,b), \\ \beta & \sim & \mathcal{G}(c,d). \end{cases}$$

Assuming that the device is submitted to aging, an usual domain of main variations for the values of β is $D_\beta = [1,5]$ (Bacha *et al.* 1998). Since η is the 63rd percentile of the distribution, it is more tractable from the expert opinion than β. We translate approximatively the percentiles on X into the percentiles on η using the Weibull pdf, fixing $\beta = 3$. This translated knowledge and the corresponding values of a and b (assessed by least squares regression) are given in Table 4.

In estimation, Sun (1997) recommended to use the reference prior π^J whether one or both parameters are of interest (especially in small samples cases). Besides, when both parameters are of interest, π^J is the unique second-order coverage matching prior. Since π^J is improper, we have to compute DAC^{AIJ}. An uncensored MTS $x(l)$ is a couple of values (x_i, x_j) such that $x_i \neq x_j$ and $x_i > 1$, $x_j > 1$. Fortunately, $\pi^{ij}(\eta,\beta) = \pi^J(\theta|x_i,x_j)$ is explicit, which considerably simplifies the computation. From Berger *et al.* (1998),

$$\pi^{ij}(\eta,\beta) = \left(2(x_i x_j)|\log x_i/x_j|\right)^{-1}(x_i x_j)^{\beta-1}\beta\eta^{-2\beta-1}\exp\left(-\eta^{-\beta}\left(x_i^{\beta}+x_j^{\beta}\right)\right).$$

Then consider the new parametrization $\eta \to \mu = \eta^{-\beta}$, $\beta \to \beta$ with Jacobian $J(\mu,\beta) = \beta\mu^{1+1/\beta}$. The corresponding noninformative prior is $\pi^J(\mu,\beta) \propto (\mu\beta^2)^{-1}$. Thus $\pi^{ij}(\mu,\beta) = \pi^{ij}(\mu|\beta)\ \pi^{ij}(\beta)$ with

$$\begin{array}{rcl} \mu|\beta & \sim & \mathcal{G}\left(2, x_i^\beta + x_j^\beta\right), \\ \pi^{ij}(\beta) & = & \dfrac{(x_i x_j)^{\beta-2}}{2|\log x_i/x_j|\left(x_i^\beta + x_j^\beta\right)^2}. \end{array}$$

The computation of DAC^{AIJ} needs the posterior density of π^J conditionally to $\mathbf{y}_{\mathbf{n}(ij)} = (y_{(ij)1},\ldots,y_{(ij)n})$ (the sample $\mathbf{y_n}$ whose components x_i and x_j have been removed). Denote similarly $\mathbf{x}_{\mathbf{r}(ij)}$ the subsample of uncensored data. The posterior densities are

$$\pi^{ij}(\mu,\beta|\mathbf{y}_{\mathbf{n}(ij)}) \quad = \quad \pi^{ij}(\mu|\beta,\mathbf{y}_{\mathbf{n}(ij)})\ \pi^{ij}(\beta|\mathbf{y}_{\mathbf{n}(ij)})$$

with

$$\begin{array}{rcl} \mu|\beta,\mathbf{y}_{\mathbf{n}(ij)} & \sim & \mathcal{G}\left(r, \displaystyle\sum_{k=1}^{n} y_{(ij)k}^\beta\right), \\ \pi(\beta|\mathbf{y}_{\mathbf{n}(ij)}) & \propto & \beta^r \dfrac{\left(\prod_{k=1}^{r} x_{(ij)k}\right)^\beta}{\left(\sum_{k=1}^{n} y_{(ij)k}^\beta\right)^r}. \end{array}$$

When the MTS contain censored data, we use the special Jeffreys prior introduced by De Santis *et al.* (1998) and elicited by Bousquet (2006b). Denote

$$\begin{array}{rcl} \tilde{\pi}(\eta,\beta) & = & \pi(\eta)\pi^J(\beta) \ \propto \ \dfrac{1}{\beta}\pi(\eta), \\ \tilde{\tilde{\pi}}(\eta,\beta) & = & \pi^J(\eta)\pi(\beta) \ \propto \ \dfrac{1}{\eta}\pi(\beta). \end{array}$$

From (2) we have

$$\mathrm{DAC}^{AIJ}(\pi|\mathbf{y_n}) \quad = \quad \mathrm{DAC}^{AIJ}(a,b|\mathbf{y_n}) + \mathrm{DAC}^{AIJ}(c,d|\mathbf{y_n}) - 1. \tag{7}$$

where

$$\begin{array}{rcl} \mathrm{DAC}^{AIJ}(a,b|\mathbf{y_n}) & = & \dfrac{1}{L}\displaystyle\sum_{i=1}^{L} \dfrac{KL\left\{\pi^J(.|\mathbf{y_n}(l)) \ || \ \tilde{\pi}(.|x(l))\right\}}{KL\left\{\pi^J(.|\mathbf{y_n}(l)) \ || \ \pi^J(.|x(l))\right\}}, \\ \mathrm{DAC}^{AIJ}(c,d|\mathbf{y_n}) & = & \dfrac{1}{L}\displaystyle\sum_{i=1}^{L} \dfrac{KL\left\{\pi^J(.|\mathbf{y_n}(l)) \ || \ \tilde{\tilde{\pi}}(.|x(l))\right\}}{KL\left\{\pi^J(.|\mathbf{y_n}(l)) \ || \ \pi^J(.|x(l))\right\}}. \end{array}$$

Thus, we can check $\pi(\eta)$ separately from $\pi(\beta)$. We chose $L = 30$ (among $n(n-1)/2$ possible uncensored and censored MTS). For expert $\mathcal{E}_1$, we obtained $\mathrm{DAC}^{AIJ}(a,b|\mathbf{y_n}) = 3.41$. For expert $\mathcal{E}_2$, we obtained $\mathrm{DAC}^{AIJ}(a,b|\mathbf{y_n}) = 1.76$. Thus, we detect a conflict between the data and the experts on the lifetime scale. Notice that the gamma prior on η, for expert $\mathcal{E}_1$, is very peaked and can be well approximated with a normal distribution (since $a > 30$). From (7), it is visible that no choice of $\pi(\beta)$, even a flat prior, allows the complete prior π to be not conflicting. In an industrial context, such a situation must be noticed before the inference ; this discrepancy reflects a deep disparity between data and expert information.

The second expert opinion is not in this case. The scale parameter is affected by a similar conflict, but it remains possible to ensure that the complete prior is not conflicting: one must elicit $\pi(\beta)$ such that

$$\mathrm{DAC}^{AIJ}(c,d|\mathbf{y_n}) \quad \leq \quad 0.24.$$

From the analyst viewpoint, the experts are optimistic with respect to the data. So they seem to favor a soft aging of the device $\sum$ (a simple reason is that they integrate some knowledge about the technical evolution in their opinions). For this reason, the Bayesian analyst should choose the expectation c/d of $\pi(\beta)$ in $[1,2]$. For instance, the analyst selects $\mathrm{E}[\beta] = 1.5$. Then the second expert opinion is not conflicting for values c such that $\mathrm{Var}[\beta] \geq 0.683$, so for $c \leq 0.65$.

8. Discussion

In this paper, we provide a characterization of the conflict between prior subjective knowledge and data information for the Bayesian decision-maker. We suggested two features for this definition. A) In the same idea as Evans and Moshonov (2006), both information can favor regions of the parameter space Θ that are far from each other. This is for instance the case, in reliability, where there is a time discrepancy between data that are formerly collected on a old device, and prospective expert opinions that take account of technical evolution. B) The subjective information introduced throughout the prior into the inference has not to overwhelm the data information, otherwise the Bayesian decision-making suffers of a lack of objectivity and threatens to lost its justification. The DAC criterion, based on Kullback-Leibler divergences between benchmark objective priors and the assessed prior, enables the Bayesian analyst to respect both point of views and check all floors of a hierarchised prior elicitation.

Since DAC is a binary criterion, it leads to a first, understandable diagnostic which can be statistically refined using the EMo procedure, uniquely based on the parameter location: a $p-$value close to 0.5 will discriminate between a conflict in location and a conflict in information uncertainty.

However, DAC indicates threshold values for the prior hyperparameters but such values remain undecidable for EMo. Thus a procedure of prior rejection or prior calibration based on DAC is devoid of the uncomfortable choice of a significance level.

There remain some difficulties to use DAC. When π^J is improper, the intrinsic adaptation can suffer of the small size n and the high dimension of the model. Possible improvements have been highlighted, which are deeply linked to objective Bayesian model selection issues; future improvements of DAC adaptation should probably follow from improvements in this area.

Finally we think that the construction principle of DAC is an interesting alternative to the EMo procedure and a helpful complementary method to place in the toolkit of the Bayesian analyst. In function of the available information about the conditions of the experiment and the expert credibility, he or she could correct the subjective beliefs or ask for other experiments to understand a detected discrepancy. An open issue could be detecting some outliers or too influential data in the sample by sequential computations of DAC, increasing or randomizing the dataset.

9. Acknowledgements

The author thanks Gilles Celeux (INRIA) and Jean-Michel Marin (INRIA) for many fruitful discussions and advices. This study has been proposed from experimental issues by François Billy and Emmanuel Remy (EDF). Many thanks to Profs. Christian P. Robert, Guido Consonni, Piero Veronese, Nozer D. Singpurwalla and Michael Evans for their advices, questionings and comments. Finally, the author would like to thank an unknown reviewer for several comments and critics which greatly help to improve the paper.

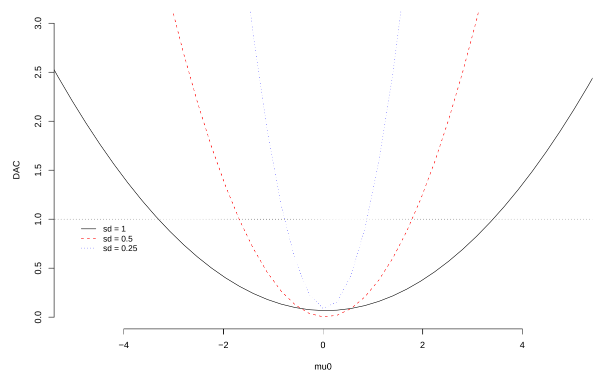

Fig. 1. Mean evolution of DAC^J in function of the prior mean μ_0 and standard deviation σ_0 (labeled by "sd" on the graphics). (Ex. 1).

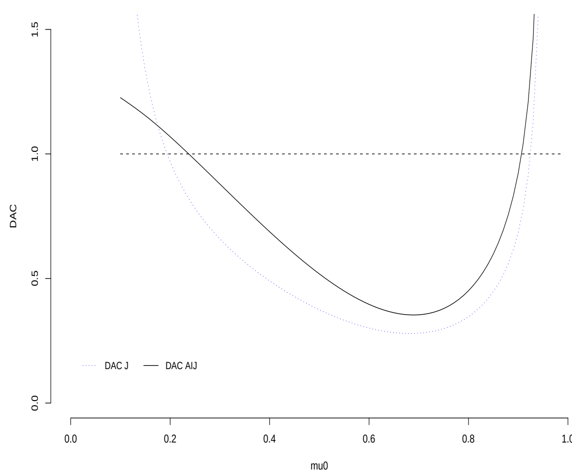

Fig. 2. Mean evolution of DAC^J and DAC^{AIJ} in function of the prior mean μ_0 (Ex.4, $n = 20$).

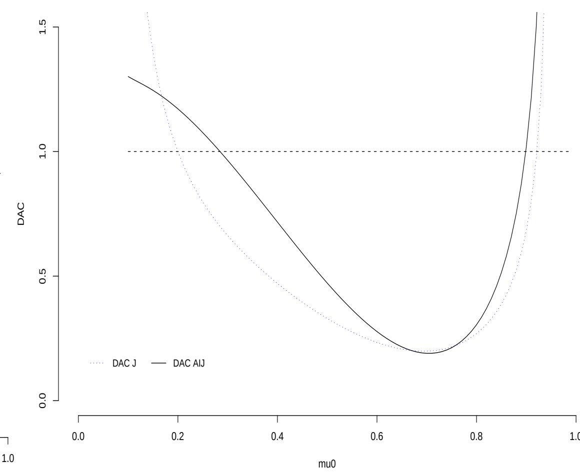

Fig. 3. Mean evolution of DAC^J and DAC^{AIJ} in function of the prior mean μ_0 (Ex.4, $n = 10$).

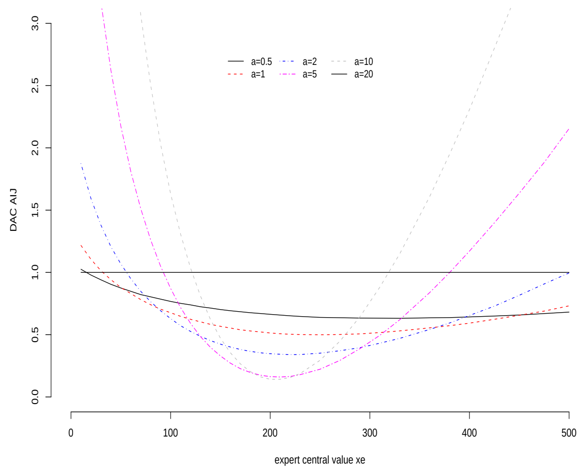

Fig. 4. Evolutions of DAC^{AIJ} w.r.t. prior mean x_e and virtual size a (Ex.5)

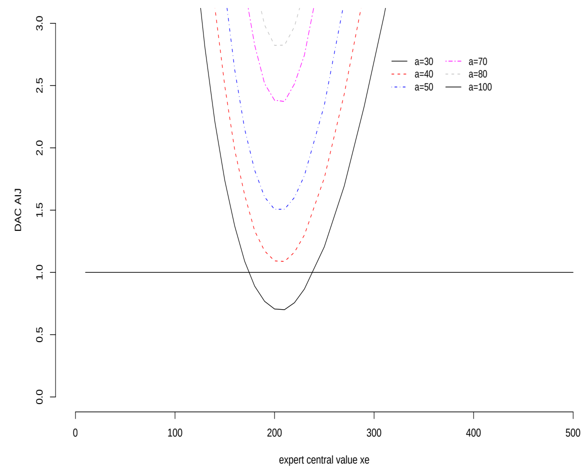

Fig. 5. Evolutions of DAC^{AIJ} w.r.t. prior mean x_e and virtual size a (Ex.5).

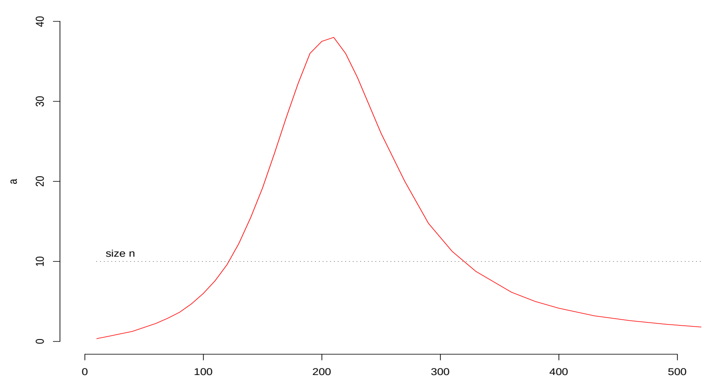

Fig. 6. Evolutions of the limiting virtual size $\hat{a}$ in function of the prior mean x_e in a neighborhood of $\bar{y} = 207$ (Ex. 6).

10. Tables and figures

a	x_e				
	10	150	200	300	500
5	4.92	0.33	0.16	0.43	2.16
1	1.22	0.57	0.52	0.51	0.73
0.5	1.02	0.70	0.66	0.63	0.68
0.25	0.98	0.80	0.78	0.74	0.72

Table 1. Values of DAC^{AIJ} (Ex.5).

a	DAC^{AIJ}	$p-$value proc. $(5\% - 95\%)$
2	55 - 500	37 - 625
5	90 - 383	41 - 510
7	102 - 355	53 - 480
10	118 - 320	65 - 460
15	137 - 290	77 - 440

Table 2. Agreement domains for x_e (Ex.5).

real failure times:	134.9, 152.1, 133.7, 114.8, 110.0, 129.0, 78.7, 72.8, 132.2, 91.8
right-censored times :	70.0, 159.5, 98.5, 167.2, 66.8, 95.3, 80.9, 83.2

Table 3. Lifetimes (months) of nuclear components from secondary water circuits (§ 7).

	expert knowledge (on X)		translated knowledge (on η)		a	b
	(5%,95%) interval	median value	(5%,95%) interval	median value		
Expert $\mathcal{E}_1$	(200,300)	250	(224,336)	280	66.3	0.23
Expert $\mathcal{E}_2$	(100,500)	250	(112,560)	280	4.6	0.015

Table 4. Prior domains on X and η and hyperparameter values for $\pi_1(\eta)$ (§ 7).

Appendix: proof of proposition 3

Denote $\xrightarrow{\mathcal{L}}$ the convergence in distribution. By the central limit theorem, under $\mathcal{B}_r(\theta_0)$,

$$\frac{\delta_n - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}} \quad \xrightarrow{\mathcal{L}} \quad \mathcal{N}(0,1).$$

Denote $\delta_n = n\theta_0 + U_n\sqrt{n}$ where $U_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,\theta_0(1-\theta_0))$. Denote Ψ the digamma function (the derivative of the log-gamma function). After some heavy calculations using asymptotic following developments

$$\begin{aligned}
\Psi(n+1) &= \log n + \frac{1}{2n} - \frac{1}{12n^2} + o(n^{-3}), \\
\log\Gamma(n+1) &= \frac{1}{2}\log 2\pi + \left(n+\frac{1}{2}\right)\log n - n + \frac{\alpha}{12n}, \quad \text{where } 0<\alpha<1
\end{aligned}$$

which can be derived from Abramowitz and Stegun (1972, p. 258-260) and Artin (1964, p.24), respectively, we obtain for $n > \max\{1/2\theta_0, 1/2(1-\theta_0)\}$ that

$$\begin{aligned}
KL\left\{\pi^J(.|\mathbf{X_n}) \mid \pi^J\right\} &= (n+1/2)\log n - n\Psi(1/2) + U_n\sqrt{n}\left\{4-2\Psi(1/2)\right\} + K_{\theta_0}(1/2,1/2) + o(1), \\
KL\left\{\pi^J(.|\mathbf{X_n}) \mid \pi\right\} &= (n+3/2-\alpha-\beta)\log n - n\left(\Psi(\beta) + \theta_0\left\{\Psi(\alpha)-\Psi(\beta)\right\}\right) + U_n\sqrt{n}\left\{4-\Psi(\alpha)-\Psi(\beta)\right\} \\
&\quad + K_{\theta_0}(\alpha,\beta) + o(1) \\
\text{where} \quad K_{\theta_0}(\alpha,\beta) &= \log\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} - \log\sqrt{2\pi} + \left(\alpha-\frac{1}{2}\right)\log\frac{\Psi(\alpha)}{\theta_0} + \left(\beta-\frac{1}{2}\right)\log\frac{\Psi(\beta)}{1-\theta_0}.
\end{aligned}$$

Then the asymptotic development of DAC^J gives

$$\begin{aligned}
\mathrm{DAC}^J(\alpha,\beta|\mathbf{X_n}) &= 1 + \frac{A_{\theta_0}(\alpha,\beta)}{\log n}\left\{1-\frac{\Psi(1/2)}{\log n}\right\} + U_n\frac{B(\alpha,\beta)}{\sqrt{n}\log n}\left\{1-\frac{\Psi(1/2)}{\log n}\right\} \\
&\quad - U_n\frac{C_{\theta_0}(\alpha,\beta)}{\sqrt{n}\,(\log n)^2} + \frac{D(\alpha,\beta)}{n} + o(n^{-1})
\end{aligned} \tag{8}$$

$$\begin{aligned}
\text{where} \quad A_{\theta_0}(\alpha,\beta) &= \Psi(1/2) - \Psi(\beta) + \theta_0\left\{\Psi(\beta)-\Psi(\alpha)\right\}, \\
C_{\theta_0}(\alpha,\beta) &= A_{\theta_0}(\alpha,\beta)\left\{4-2\Psi(1/2)\right\},
\end{aligned}$$

and $B(\alpha,\beta) = 2\Psi(1/2) - \Psi(\alpha) - \Psi(\beta)$ and $D(\alpha,\beta) = 1-\alpha-\beta$. Note that at least one term in (8) is nonzero, except when $\pi \equiv \pi^J$ ($\Leftrightarrow \alpha=\beta=1/2 \Leftrightarrow \mathrm{DAC}^J = 1$). Indeed,

$$\left\{\begin{array}{lcl} A_{\theta_0}(\alpha,\beta) &=& 0 \\ B(\alpha,\beta) &=& 0 \\ D(\alpha,\beta) &=& 0 \end{array}\right. \quad\Leftrightarrow\quad \left\{\begin{array}{lcl} \Psi(\alpha) &=& \Psi(\beta) \\ \alpha+\beta &=& 1 \end{array}\right. \quad\Leftrightarrow\quad \alpha=\beta=1/2.$$

To prove (2) for any $0<q<1$, it is enough to control $\mathrm{E}[V_n]$ where $V_n = U_n\left[n(\log n)^q\right]^{-1}$. A sufficient condition is to show that $\mathrm{E}[V_n] \to 0$ when n increases. This can be done as follows.

Denote $Z_n = \delta_n/n$. With

$$V_n \quad = \quad \frac{(Z_n - \theta_0)}{(\log n)^q},$$

we obtain by Markov 's inequality, for any $M > 0$,

$$\begin{aligned}
\mathrm{E}\left[|V_n|\mathbb{1}_{\{|V_n|\geq M\}}\right] &\leq M^{-1}\mathrm{E}\left[|V_n|^2\right], \\
&\leq M^{-1}\frac{\mathrm{E}[Z_n^2] + 2\theta_0\mathrm{E}[Z_n] + \theta_0^2}{(\log n)^{2q}}, \\
&\leq M^{-1}\frac{\theta_0\left(1 - \theta_0 + 2n\theta_0\right)}{n\left(\log n\right)^{2q}}
\end{aligned}$$

which obviously tends to 0 when $n \to \infty$ followed by $M \to \infty$. This result ensures that V_n is asymptotically uniformly integrable. Then, from van der Vaart (1998, Theorem 2.20), we have

$$\lim_{n\to\infty} \mathrm{E}[V_n] = \lim_{n\to\infty} \frac{\mathrm{E}[U]}{n(\log n)^q}$$

where $U \sim \mathcal{N}(0, \theta_0(1-\theta_0))$. Then $\mathrm{E}[V_n] \to 0$ and the statement of the proposition follows.

References

Abramowitz, M. and Stegun, I. A. (Eds.). (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover.

Ali, S.M. and Silvey, D. (1966). A general class of coefficients of divergence of one distribution from another, *Journal of the Royal Statistical Society: Series B*, **28**, pp. 131-142.

Andrade, J.A.A., and O'Hagan, A. (2006). Bayesian robustness modeling using regularly varying distributions, *Bayesian Analysis*, **1**, pp. 169-188.

Andrieu, C., Doucet, A., Fitzgerald, W.J. and Pérez, J.M. (2001). Bayesian Computational Approaches to Model Selection, *Nonlinear and Non Gaussian Signal Processing*, Smith, R.L., Young, P.C. and Walkden. A. (Eds), Cambridge University Press.

Angers, J.F. (2000). Credence and Robustness Behavior, *Metron*, **58**, pp. 81-108.

Artin, E. (1964). *The Gamma function*, Holt Rinehart Winston, New York.

Bacha, M., Celeux, G., Idée, E., Lannoy, A. and Vasseur, D. (1998). *Estimation de modèles de durées de vie fortement censurées*, Eyrolles.

Bayarri, M.J. and Berger, J.O. (2000). P-values for composite null models, *Journal of the American Statistical Association*, **95**, pp. 1127-1142.

Bayarri, M.J. and Berger, J.O. (2003). *The interplay of Bayesian and Frequentist Analysis*, Technical Report of the University of Valencia and Duke University.

Berger, J.O. and Pericchi, L.R. (1996). The Intrinsic Bayes Factor for Model Selection and Prediction, *Journal of the Americal Statistical Association*, **91**, pp. 109-122.

Berger, J.O., Pericchi, L.R. and Varshavsky, J.A. (1998). Bayes Factors and Marginal Distributions in invariant situations, *Sankhya: The Indian Journal of Statistics*, **60**, pp. 307-321.

Berger, J.O. and Pericchi, L.R. (1998). Accurate and stable Bayesian Model Selection: the Median Intrinsic Bayes Factor, *Sankhyā: the Indian Journal of Statistics*, **60**, pp.1-18.

Berger, J.O., and Pericchi, L.R. (2002). Training Samples in Objective Bayesian Model Selection,ISDS Discussion Paper 02-14.

Bernardo, J.M. (1997). Noninformative Priors Do Not Exist: A Discussion, *Journal of Statistical Planning and Inference*, **65**, pp. 159-189 .

Bousquet, N. (2006). Subjective Bayesian statistics: agreement between prior and data, Research report HAL-INRIA 00115528.

Bousquet, N., and Celeux, G. (2006). Bayesian agreement between prior and data, *Proceedings of the ISBA congress*, Benidorm, Spain.

Bonnevialle, A.-M., and Billy, F. (2006). Reupdating FED reliability data: feasibility of a subjective Bayesian method (Réactualisation de données de fiabilité issues du REX : faisabilité d'une méthode bayésienne subjective), *Lambda-Mu Proceedings* (french), Lille.

Budescu, D.V. and Rantilla, A. K. (2000). Confidence in aggregation of expert opinions. *Acta Psychologica*, **104**, pp. 371-398.

Cappé, O., Guillin, A., Marin, J.M. and Robert, C.P. (2004). Population Monte Carlo, *Journal of Computational and Graphical Statistics*, **13**, pp. 907-909.

Celeux, G., Marin, J.M. and Robert, C.P. (2003). Iterated importance sampling in missing data problems, *Computational Statistics and Data Analysis* (to appear).

Clarke B.S. (1996). Implications of reference priors for prior information and for sample size, *Journal of the American Statistical Association*, **91**, pp. 173-184.

Cover, T.M. and Thomas, J.A. (1991). *Elements of Information Theory.* New York: Wiley.

Daneshkhah, A.R. (2004). *Psychological Aspects Influencing Elicitation of Subjective Probability*, Research Report, University of Sheffield.

Dass, S.C. (2001). "Propriety of Intrinsic Priors in Invariant Testing Situations", *Journal of Statistical Planning and Inference*, **92**, pp. 147-162.

Dawid, A.P. (1982). The Well-calibrated Bayesian, *Journal of the American Statistical Association*, **77**, pp. 605-613.

De Finetti, B. (1961). "The Bayesian Approach to the Rejection of Outliers", in *Proceedings of the Fourth Berkeley Symposium on Probability and Statistics* (Vol.1), Berkeley: University of California Press, pp. 199-210.

De Santis,F., Mortera, J. and Nardi, A. (2001). Jeffreys priors for survival models with censored data, *Journal of Statistical Planning and Inference*, **99**, pp. 193-209.

Dodson, B. (2006). *The Weibull Analysis Handbook, second edition*, ASQ Quality Press, Milwaukee.

Evans, M., and Moshonov, H. (2006). Checking for Prior-Data conflict, *Bayesian Analysis*, **1**, pp. 893-914.

Evans, M., and Moshonov, H. (2007). Checking for Prior-Data conflict with Hierarchically Specified Priors, in *Bayesian Statistics and its Applications*, eds. A.K. Upadhyay, U. Singh, D. Dey, Anamaya Publishers, New Delhi, pp. 145-159.

Garthwaite, P.H., Kadane, J.B. and O'Hagan, A. (2005). Statistical methods for eliciting probability distributions, *Journal of the American Statistical Association*, **100**, pp. 680-701.

Gelman, A., Meng, X. and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies, *Statistica Sinica*, **6**, pp. 733-808.

Ghoshal, S. (1999). Probability matching priors for non-regular cases, *Biometrika*, **86**, pp. 956-964.

Ghosh, M., Reid, N. and Fraser, D.A.S. (2007) Ancillary statistics: a review, *submitted*.

Gneiting, T., Balabdaoui, F. and Raftery, A.E. (2007) Probabilistic forecasts, calibration and sharpness, *Journal of the Royal Statistical Society: Series B*, **69** (2), pp. 243-268.

Gneiting, T. and Raftery, A.E. (2007) Strictly Proper Scoring Rules, Prediction, and Estimation, *JASA*, **102**, pp. 359-378.

Hartigan, J.A. (1983). *Bayes' Theory*, New York: Springer-Verlag.

Hartigan, J.A. (1998). The Maximum Likelihood Prior, *The Annals of Statistics*, **26**, pp. 2083-2103.

Hill, B.M. (1974). "On Coherence, Inadmissibility and Inference About Many Parameters in the Theory of Least Squares", in *Studies in Bayesian Econometrics and Statistics*, eds. S.E. Fienberg and A. Zellner, Amsterdam: North-Holland, pp. 555-584.

Idée, E., Lannoy, A. and Meslin, T. (2001). *Estimation of a lifetime law for equipment on the basis of a highly right multicensored sample and expert assessments*, preprint of LAMA team, 01-10b, Université de Savoie.

Kass, R.E. and Wasserman, L. (1996). The selection of prior distributions by formal rules, *Journal of the American Statistical Association*, **91**, pp. 1343-1370.

Lucas, W. (1993). When is Conflict Normal ?, *Journal of the American Statistical Association*, **88**, pp. 1433-1437.

O'Hagan, A. (1979). On outlier rejection phenonema in Bayes inference, *Journal of the Royal Statistical Society: Series B*, **41**, pp. 358-367.

O'Hagan, A. (1988). *Modelling with heavy tails.* J. M. Bernardo et al (Eds.), *Bayesian Statistics 3*,Oxford University Press, pp. 345-359.

O'Hagan, A. (1990). On outliers and credence for location parameter inference. *Journal of the American Statistical Association*, **85**, pp. 172- 176.

O'Hagan, A. (1995). Fractional Bayes factors for model comparisons, *Journal of the Royal Statistical Society: Series B*, **57**, pp. 99-138.

O'Hagan, A. (1997). Properties of intrinsic and fractional Bayes factors, *Test*, **6**, pp. 101-118.

O'Hagan, A. (2003). HSSS model criticism (with discussion). In Highly Structured Stochastic Systems, P. J. Green, N. L. Hjort and S. T. Richardson (eds),Oxford University Press, pp. 423-453.

Pérez, J.M., and Berger, J. (2002). Expected posterior prior distributions for model selection, *Biometrika*, **89**, pp. 491-512.

Penny, W.D. (2001). *KL-Divergences of normal, gamma, Dirichlet and Wishart densities*, Technical Report, Wellcome Dpt of Cognitive Neurology, University College London.

Robert, C.P. (2001). *The Bayesian Choice. From Decision-Theoretic Motivations to Computational Implementation. (second edition)*, Springer-Verlag: New York.

Robert, C.P., and Casella G. (2004). *Monte Carlo Statistical Methods (second edition)*, Springer-Verlag: New York.

Sinanović, S. and Johnson, D.H. (2007). Towards a Theory of Information Processing, *Signal Processing*, **87**, pp. 1326-1344.

Singpurwalla, N.D. and Song, M.S. (1988) "Reliability Analysis using Weibull Lifetime Data and Expert Opinion.", *IEEE Transactions on Reliability*, **37**, pp. 340-347.

Soland, R. (1969). Bayesian analysis of the Weibull process with unknown scale and shape parameters, *IEEE Transactions on Reliability*, **18**, pp. 181-184.

Sun, D. (1997). A note on noninformative priors for Weibull distributions, *Journal of Statistical Planning and Inference*, **61**, pp. 319-338.

Usureau, E. (2001). *Application des méthodes bayésiennes pour l'optimisation des coûts de développement des produits nouveaux*, Ph.D. Thesis n413, Institut des Sciences et Techniques d'Angers.

van der Vaart, A.W. (1998). *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics.