Subjective Bayesian statistics: agreement between prior and data

Nicolas Bousquet

▶ To cite this version:

Nicolas Bousquet. Subjective Bayesian statistics: agreement between prior and data. [Research Report] RR-5900, INRIA. 2006. inria-00071367

HAL Id: inria-00071367
https://hal.inria.fr/inria-00071367

Submitted on 23 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Subjective Bayesian statistics: agreement between prior and data

Nicolas Bousquet

N° 5900

Mai 2006

Thème COG

Rapport
de recherche

ISRN INRIA/RR--5900--FR+ENG
ISSN 0249-6399

Subjective Bayesian statistics: agreement between prior and data

Nicolas Bousquet*

Thème COG — Systèmes cognitifs
Projet SELECT

Rapport de recherche n° 5900 — Mai 2006 — 23 pages

Abstract: When Bayesian inference is required to estimate the parameter of a decision-making model, the prior modelling must be specified with great care, especially when subjective knowledge is used. Indeed, Bayesian inference is jeopardized when there is a too large discrepancy between prior knowledge and observed data. From a theoretical point of view, prior and likelihood constitute the complete Bayesian model and it seems relevant to study the agreement of the data to this model. In the fact, the detection of a possible conflict between the prior and the observed likelihood remains a significant preliminary to subjective Bayesian inference, particularly in industrial reliability, and seems not to have been much studied before. In the present report, we propose to use a criterion $\mathcal{C}oh(\pi; \mathbf{X_n})$ measuring the agreement of the available prior π and the data $\mathbf{X_n}$ information with a ratio of Kullback-Leibler distances between the proposed prior and benchmark weakly or non-informative prior distributions. If $\mathcal{C}oh(\pi; \mathbf{X_n}) \leq 1$ then both informations are declared in agreement. To use it we have to define a class of minimaly informative proper prior distributions from data. Therefore we propose a general approach using *minimal training samples* and *posterior priors*, overcoming some difficulties to model ignorance. This approach is applied to the exponential and Weibull models, the most used distributions in lifetime problems. Finally we discuss the use of the $\mathcal{C}oh$ criterion as a tool of calibration for subjective prior modelling.

Key-words: Industrial Reliability; Bayesian Inference; Expert Opinion; Reference Prior; Kullback-Leibler distance; Noninformative Priors; Default Priors; *Minimal Training Samples*; Entropy; Prior Calibration; Exponential; Weibull.

* INRIA Futurs, Département de Mathématiques, Université Paris-Sud, 91405 Orsay Cedex,email : nicolas.bousquet@math.u-psud.fr

Mesure de cohérence entre information *a priori* et données en statistique bayésienne subjective

Résumé : Quand le paradigme bayésien doit être mis en œuvre pour estimer le paramètre d'un modèle de décision, la modélisation *a priori* de ce paramètre doit être calibrée très soigneusement, particulièrement dans un cadre de fiabilitié industrielle où la connaissance subjective est très utilisée. Si la vraisemblance des données observées et la connaissance *a priori* se révèlent en trop grand désaccord sur la zone de confiance qu'elles induisent pour ce paramètre, l'inférence bayésienne peut conduire à de mauvaises décisions. La détection d'un tel conflit revêt un aspect important - et curieusement sous-estimé - en statistique appliquée à des problèmes concrets. Dans ce rapport nous proposons l'emploi d'un critère $\mathcal{C}oh(\pi; \mathbf{X_n})$ mesurant la comparabilité entre la modélisation *a priori* π et l'information apportée par les données $\mathbf{X_n}$, par un rapport de distances de Kullback-Leibler. Si $\mathcal{C}oh(\pi; \mathbf{X_n}) \leq 1$ ces deux informations sont déclarées cohérentes. Pour l'utiliser de manière concrète, nous devons définir une classe d'*a priori* propres, faiblement ou non informatifs, représentant l'apport d'une connaissance minimale cohérente avec les données. Nous proposons pour ce faire une démarche fondée sur l'utilisation d'échantillons d'apprentissage minimaux et plus généralement sur des lois *a priori a posteriori*. Enfin, une discussion générale clôture ce rapport, introduisant le critère $\mathcal{C}oh$ comme un outil de calibration d'un *a priori* subjectif. L'utilisation de ce crit'ere sera illustrée par l'application aux lois exponentielles et de Weibull, certainement les plus utilisés en fiabilité pour modéliser des durées de vie.

Mots-clés : fiabilité industrielle ; inférence bayésienne ; opinion d'expert ; modélisation *a priori* ; distance de Kullback-Leibler ; *a priori* non informatifs ; *a priori* objectifs ; échantillons d'apprentissage minimaux ; entropie ; calibration *a priori* ; modèle exponentiel; modèle de Weibull

1 Motivation

Among the articles on subjective Bayesian inference to estimate the parameter θ of a decision-making model $\mathcal{M}(\theta)$, many of them insist on the displeasing behavior of the posterior odds when the prior information is badly conditionned. By this term we encompass a wrong information given by expert opinions (or any other source of subjective information) and a prior distribution $\pi(\theta)$ far from the frequentist confidence region on parameter θ brought by observed data $\mathbf{X_n} = (X_1, \ldots, X_n)$. The objective knowledge of the reality given by the likelihood $\mathcal{L}(\mathbf{X_n}, \theta)$ can thus be drown by the choice of a wrong prior and the Bayesian analyst can take a posterior decision with unwelcome consequences. In most cases, these articles are centered on a reinforcement of the posterior robustness. However, one first difficulty is just considering that the avalaible knowledge is not self-contradictory.

Note that in case of highly censored data and small samples, when the Bayesian informative approach is very recommended (Robert, 2001), the heterogeneity of the data can have the same effects as a wrong prior with correct data. A prior can be very well chosen but the posterior threatens to lead to bad decisions. For instance, imagine an industrial system whose failures have been noted down in the past. After some amendments the system is currently working. A material expert will give probable failure times in the future which can be right-shifted from the data, because of the technical evolution. In this case, some of the formest data may be considered as polluting the objective knowledge. Such a situation must be noticed before all posterior acceptation.

Then, a prior knowledge and a real knowledge can have symmetric roles in a conflict. Especially in an industrial context, the prior calibration answers to the shakeout of the observed data (by removing outliers, burn-in data, etc.). Curiously, this subject seems not to have been much studied although it constitutes, from our point of view, an important preliminary to Bayesian subjective inference. As far as we know, it does not exist simple statistic tools for defining the agreement between a prior distribution and an objective knowledge brought by observed data.

Notice however that in the strict observation of the Bayesian paradigm, prior distribution and likelihood constitute a whole decision model and it seems at first sight not relevant to determine degrees of discrepancy between them, from a theoretical point of view. That is why many Bayesian articles are centered on studying the agreement of the data with the Bayesian model choice and the detection of *surprise*. See for instance Bayarri and Berger (1997,1998). To our knowledge, there are few articles whose subject is the detection of conflict between informative priors and data. In former works, indicators are usually built to allow to modify the prior or the likelihood by detecting and removing outliers or too influential data. Another issue is to elicit robust priors with negligible posterior impact when a conflict is detected. Main references are De Finetti (1961), Dawid (1982), Hill (1974) Lucas (1993) and especially O'Hagan (1979, 1988, 1990, 2003) and Angers (2000). In industrial studies, Idée et *al.* (2001) have suggested to use a Fisher test between prior and empiric mesures of uncertainty and an *ad hoc* approach has been proposed by Usureau (2001) without statistical justification.

We propose the following criterion for detecting a potential conflict between prior and data. Denote π the proposed prior distribution on θ. Suppose being able to define π^{MIA} the density of a *minimaly informative data-agreed* (MIA) proper prior distribution, which ideally represents the minimal knowledge from a prior opinion, in agreement with the data. Then compute the ratio

$$\mathcal{C}oh(\pi;\mathbf{X_n}) \quad = \quad \frac{KL\left(\pi^{MIA}(.|\mathbf{X_n}) \mid\mid \pi\right)}{KL\left(\pi^{MIA}(.|\mathbf{X_n}) \mid\mid \pi^{MIA}\right)}$$

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

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

$\mathcal{C}oh(\pi;\mathbf{X_n})$ is regarded as a data-agreement criterion. If $\mathcal{C}oh(\pi;\mathbf{X_n}) \leq 1$, prior and data-given confidence regions for θ are close enough and the prior proposal π is in agreement with the objective data knowledge. Else a conflict is detected.

In Section 2 of this article, we detail the arguments which lead to the choice of this criterion. One of the main difficulty raised by its definition is giving a precise sense to the MIA distributions. Section 3 is thus entirely devoted to considerations on the choice of MIA candidates and estimation of $\mathcal{C}oh(\pi;\mathbf{X_n})$. Some relevant difficulties are evoked and links with existing noninformative priors are described. The main point of this section will be the use of *minimal training samples*, defined in the context of Bayesian model selection by Berger and Perrichi (1996, 2002), and more generally of the notion of *posterior prior* developed by Pérez (1998) and Pérez and Berger (2002). Afterwards Section 4 is dealing with the application of previously defined rules on the examples of exponential and Weibull models. The conclusion evokes the use of the $\mathcal{C}oh$ statistic to the calibration of subjective prior distributions. Open issues are proposed in this way.

2 Bayesian agreement between prior and data

In this section we are interested in finding a criterion allowing us to predict a symmetric incoherence for the comparison of a prior distribution $\pi(\theta)$ and data for a parametric model $\mathcal{M}(\theta)$. The choice of a subjective prior distribution $\pi(\theta)$ can be judged by the comparison between $\pi(\theta)$ and other priors considered as in agreement with the observed data (agreement "benchmarks"), for the model $\mathcal{M}(\theta)$. Thus well known tools of comparison between distributions can be used. This is the starting point of our main idea.

Let $\mathbf{X_n} = (X_1,\ldots,X_n)$ be independently and identically distributed real-or-vector-valued random variables in the sample space S with a probability density function (pdf)

$p(x|\theta)$ with respect to a dominating measure μ, and $\theta \in \Theta \subset I\!R^p$ a $p-$dimensional parameter vector. Let Π be the set of all prior measures on θ and $\Pi^I \subset \Pi$ the subset of all proper (or *informative* priors on Θ (i.e. $\Pi^I = \{\pi \in \Pi, \int_\Theta \pi(d\theta) = 1\}$). Note that we confound sometimes the notation π of a prior density with the canonic notation of the distribution when no confusion can result). The notion of information for a prior density is related to a measure of uncertainty. A prior π_1 is intuitively more informative than a prior π_2 if the uncertainty of π_1 is smaller than the π_2's one. One of the most used measure of uncertainty for a density π is the differential entropy $-\int_\Theta \pi(\theta)\log\pi(\theta)\ d\theta$.

Suppose having a rule to define $\pi^{MIA} \in \Pi^I$ as modelling the less informative available prior opinion on θ, in agreement with the data. A discussion about such a rule and the selection of π^{MIA} is done in the next section. Assuming Θ to be a compact subset of $I\!R^p$ whose bounds depend on extreme values of $\mathbf{X_n}$, a good example is a uniform prior on Θ, which maximizes the entropy.

In few words, this rule can be introduced as following: π^{MIA} reflects so few information that the posterior density $\pi^{MIA}(.|\mathbf{X_n})$ is very close to the posterior density $\pi^J(.|\mathbf{X_n})$ when π^J is a noninformative prior. This postulate seems to be a reasonable and intuitive outset, assuming π^J gets posterior coverage properties with an accurate frequentist validity.

Then, a subjective prior opinion is perfectly in accordance with the objective knowledge given by the data on θ if the following hypothesis is verified.

HYPOTHESIS 1. *The prior modelling is perfectly in agreement with the observed data* $\mathbf{X_n}$ *if*

$$\pi(\theta) \quad = \quad \pi^{MIA}(\theta|\mathbf{X_n}) \propto \mathcal{L}(\mathbf{X_n},\theta)\pi^{MIA}(\theta).$$

In subjective Bayesian inference, one of the main issues is to control the information translated in the prior π. Defining $\pi^{MIA}(.|\mathbf{X_n})$ as the *best informative* prior choice, we are interested in measuring the *informative regret*

$$KL\left(\pi^{MIA}(.|\mathbf{X_n})\ ||\ \pi\right) \tag{1}$$

and selecting densities π producing weak values of this regret. Indeed, $KL(\pi_1||\pi_2)$ states the regret due to the choice of π_2 when the true distribution is π_1. Well defined notions of regret with information-theoretic tools as the Kullback-Leibler distance can be found in Cover & Thomas (1991) and Sweeting, Datta & Ghosh (2005).

A premise of this idea to define $\pi^{MIA}(.|\mathbf{X_n})$ as the best possible prior has been notably introduced by Bernardo (1997). Then it is necessary to choose a constant C such as if $KL\left(\pi^{MIA}(.|\mathbf{X_n})\ ||\ \pi\right) > C$ then π loses its data-agreement. This upper bound is reached when $\pi = \pi^{MIA}$. Finally, the data-agreement criterion and rule are given in Definition 1.

DEFINITION 1. *Knowing π^{MIA}, define the data-agreement criterion by*

$$\mathcal{C}oh(\pi; \mathbf{X_n}) = \frac{KL\left(\pi^{MIA}(.|\mathbf{X_n}) \mid\mid \pi\right)}{KL\left(\pi^{MIA}(.|\mathbf{X_n}) \mid\mid \pi^{MIA}\right)} \quad (2)$$

and accept π as in agreement with the likelihood of $\mathbf{X_n}$ if $\mathcal{C}oh(\pi; \mathbf{X_n}) \leq 1$.

Clearly, Definition 1 is not restricted to the Kullback-Leibler distance and another choice in the Ali-Silvey class of information-theoretic measures (Ali and Silvey, 1966) can be made. Justifications for using it preferentially can be found in Hartigan (1998) and Sinanović and Johnson (2003). Note that Definition 1 can be put in relation with the *information transfer ratio*, defined by last authors. According to intuition, the data-agreement criterion should have to accept any nondegenerate prior when $n \to \infty$, because of the preeminence of data. This result can be proved using the asymptotic normality of the posterior density $\pi(.|\mathbf{X_n})$, under classical conditions of regularity for π, π^{MIA} and $p(x|\theta)$. Main references are Johnson (1970), Clarke & Barron (1990), Clarke (1999) and Hartigan's book (Hartigan, 1983). In Appendix, this behavior is briefly studied in an exponential case.

3 Eliciting MIA prior distributions

In this section we try to define prior proper distributions which represent the minimal knowledge of a subjective opinion in accordance with the data. That is to say π^{MIA} is a density which is vague enough to model the most fuzzy confidence on θ, available from the data. Firstly, we will examine the elicitation of some MIA candidates. Their set will always be noted Π^{MIA}. Then we will propose rules of selection. In parallel, an arithmetic version of the data-agreement statistic is defined, for general use when no MIA candidate can be easily selected.

3.1 MIA candidates

MIA candidates have to be chosen for their posterior closeness with a noninformative prior π^J. Kass & Wasserman (1996) have given a noteworthy overview of the understanding and elicitation of noninformative priors by structuring rules, which presents some of the concepts used in this section.

Ideally, $\pi^{MIA} = \pi^J$ and the choice of π^J as the Jeffreys prior ensures invariance of criterion (2) by parametrization change. However, except in discrete cases, the Jeffreys prior is improper and cannot be used since the denominator of the criterion is defined up to an unknown additive constant.

Thus some default π^J have to be selected for the posterior comparison. Our strategy in § 3.2 will be to consider their *posterior coverage matching* properties. Such priors have been especially studied by Peers (1965), Ghosh & Mukerjee (1993), Datta & Ghosh (1995) and Datta (1996). They provide orders of frequentist validity in that sense that the posterior confidence area matches with the frequentist coverage on Θ. In a one-dimensional case ($p = 1$), π^J is Jeffreys. When $p = 2$, the reference priors are the best matching priors (Berger and Bernardo, 1992). Note Π^J the set of selected noninformative priors.

Posterior priors. A natural, simple way of creating MIA candidates is well known in Bayesian statistics under the term of *posterior prior* elicitation, and corresponds to an adjustment of improper priors into proper priors.

This methodology is linked to the notion of *minimal training samples* (MTS), introduced and extended in the context of Bayesian model selection by Berger and Perrichi in a series of articles since 1996. The main idea is to use parts of the data as training samples. Let $\pi^J(\theta)$ be an improper prior. Let $X(l)$ denote a part of the data $\mathbf{X_n}$ such as $\pi^J(\theta|X(l))$ is a proper pdf. Noting $m^J(X(l))$ the marginal distribution $\int p(X(l)|\theta)\ \pi^J(\theta)\ d\theta$, a training sample is called *proper* if $0 < m^J(x(l)) < \infty$ and *minimal* if it is proper and no subset is proper. A MTS is thus the minimal quantity of data for which all parameters in the model are identifiable.

Conditioning π^J to a MTS give also a proper posterior, considered as a *posterior prior*. Using means or medians on MTS combinations, in the context of model selection, leads to the definition of *intrinsic priors*. Such proper priors have the same asymptotic influence on Bayes factors than improper priors π^J and avoid the presence of a unknown constant in calculations. They can also be seen as reasonable conventional priors. Pérez (1998) and Pérez and Berger (2002) extended this approach in the same context of model selection. They introduce a general measure (often seen as a marginal distribution) on a MTS space χ for integrating $\pi^J(\theta|X(l))$ with respect to $X(l)$. Such posterior priors are called *expected posterior priors*. For more precision about the MTS and the expected posterior prior approach, see Andrieu et *al.* (2001) and Berger and Perrichi (2002).

In our context of prior modelling, we retain the main idea of using minimal quantity of data for eliciting several MIA candidates from $\{\pi_1^J, \ldots, \pi_q^J\}$. In few words, we can summarize the vision of *posterior priors* by the choice

1. of a noninformative prior π^J in Π^J;
2. of an empirical or imaginary MTS space χ, depending on data $\mathbf{X_n}$;
3. of a measure g_i on χ. Any element of χ will be called $X(l)$ in the following.

We decide of the proper π_i^{MIA} candidate such as the mean with respect to g_i

$$\pi_i^{MIA}(\theta) \quad \propto \quad \int_\chi \pi^J(\theta|X(l))\ g_i(dX(l)). \tag{3}$$

Then, from one π^J, many MIA candidates can be elicited.

We propose in Definition 2 to consider the *arithmetic data-agreement criteria*. In the following of this article, we compare the numerical results of this estimation with the ones derived from explicit MIA priors. Note that data are used twice in criterion (4) to overcome the lack of information. The second proposal (5) appeared to be less stable in simulations when few data are available.

DEFINITION 2. *Let π^J be an improper prior in Π^J, and let $(X(l_1),\ldots,X(l_L))$ be L MTS for π^J. The arithmetic data-agreement criterion is*

$$\mathcal{C}oh_A(\pi;\mathbf{X_n}) \quad = \quad \frac{1}{L}\sum_{i=1}^{L}\frac{KL\left\{\pi^J\left(.|X(l_i),\mathbf{X_n}\right)\ ||\ \pi\right\}}{KL\left\{\pi^J\left(.|X(l_i),\mathbf{X_n}\right)\ ||\ \pi^J\left(.|X(l_i)\right)\right\}}. \tag{4}$$

A "leave-one-MTS-out" version of criterion (4) is

$$\widetilde{\mathcal{C}oh}_A(\pi;\mathbf{X_n}) \quad = \quad \frac{1}{L}\sum_{i=1}^{L}\frac{KL\left\{\pi^J\left(.|\mathbf{X_n}\right)\ ||\ \pi\right\}}{KL\left\{\pi^J\left(.|\mathbf{X_n}\right)\ ||\ \pi^J\left(.|X(l_i)\right)\right\}}. \tag{5}$$

Remark: Some other possibilities are evoked in Bayesian literature for the use of proper priors. Box and Tiao (1973, p.23) *locally uniform* priors have density which slowly varies over the region in which the likelihood function is concentrated, and can be regarded as good candidates. Hartigan's *maximum likelihood priors* (Hartigan, 1998) have good reasons to be MIA candidates too. If their densities exist, they are defined as priors for which corresponding Bayes estimates are asymptotically negligibly different from the maximum likelihood estimate.

3.2 MIA prior selection

Note $\Pi^{MIA}=\{\pi_i^{MIA}, i\in I\}\subset\Pi^I$ a set of MIA candidates and let $\pi_i^{MIA}\in\Pi^{MIA}$ be a MIA candidate elicited from π^J and a measure g_i by rule (3). Some rules of selection in Π^{MIA} are to be given. The best candidate should be "the less informative as possible". The two following points of view make sense to precise this vague statement.

1. Firstly, $\pi^{MIA}(.|\mathbf{Y_n})$ has to be the closest density of $\pi^J(.|\mathbf{Y_n})$ among all candidates in Π^{MIA}, with respect to the MIA marginal density $m_i(x)=\int_\Theta \pi_i^{MIA}(\theta)p(x|\theta)\ d\theta$, where $\mathbf{Y_n}=Y_1,\ldots,Y_n$ are i.i.d. random variables defined in $S^n=S\times\ldots\times S$ with n large enough. Define the *expected posterior regret* as

$$\mathcal{R}_n(i) \quad = \quad \int_{S^n} KL\left\{\pi^J(.|\mathbf{Y_n})\ ||\ \pi_i^{MIA}(.|\mathbf{Y_n})\right\}\ m_i(\mathbf{Y_n})\ d\mathbf{Y_n}.$$

For reasons of asymptotic posterior normality, under classical regularity conditions, $\mathcal{R}_n(i) \to 0$ when $n \to \infty$ (cf. Clarke 1999). It leads to select

$$\pi^{MIA} = \arg\min_{i\in I} \lim_{n\to\infty} \sum_{j\in I, j\neq i} \frac{\mathcal{R}_n(i)}{\mathcal{R}_n(j)}. \qquad (C_1)$$

2. Secondly, similarly to Bernardo (1979), define the *expected gain in information* provided by the data, where the expectation is with respect to $m_i(x)$:

$$\mathcal{K}_n(i) = \int_{S^n} KL\left\{\pi_i^{MIA}(.|\mathbf{Y_n}) \mid\mid \pi_i^{MIA}\right\} \, m_i(\mathbf{Y_n}) \, d\mathbf{Y_n}$$

and select

$$\pi^{MIA} = \arg\min_{i\in I} \lim_{n\to\infty} \sum_{j\in I, j\neq i} \frac{\mathcal{K}_n(j)}{\mathcal{K}_n(i)}. \qquad (C_2)$$

4 Application to exponential and Weibull models

In this section applications of the approach are proposed on exponential and Weibull models. These distributions are among the most used for modelling the lifetime of a industrial component and estimating their parameters is of primary interest in many reliability studies. Especially in typical cases of censored, small-sized data, the use of Bayesian inference is desirable and the modelling of subjective prior distributions is an intensive domain of research. See for instance Singpurwalla and Mao (1988), and Berger and Sun (1993, 1994).

4.1 The exponential model

We suppose to have $\mathbf{X_n} = (X_1, \ldots, X_n)$ data with pdf $p(x|\theta) = \theta \exp(-\theta x) \, 1_{\{x>0\}}$. Π^J is restricted to the Jeffreys prior $\pi^J(\theta) \propto \theta^{-1}$. A MTS for the exponential model, in the uncensored case, is a single data. Note that the right-censored case can alternatively be solved using special Jeffreys priors defined by De Santis et *al.* (2001) or using *sequential minimal training samples* (SMTS) including censored observations, defined by Berger and Perrichi (2002). For reasons of simplicity this case will not be considered here.

MIA candidates. Note that the Jeffreys posterior is the $\mathcal{G}(n, \sum_{i=1}^n X_i)$ distribution, where $\mathcal{G}(a,b)$ is the Gamma distribution with mean a/b and variance a/b^2. From π^J we propose three MIA candidates depending from data.

1. *Sufficient Conjugate Prior (SCP).* Choosing g_1 as the Dirac distribution concentrated at the sufficient statistic in rule (3), a natural prior is the exponential conjugate $\pi_1^{MIA}(\theta) = \mathcal{G}(1, n^{-1}\sum_{i=1}^n X_i)$ centered on the maximum likelihood estimator (MLE)

$\hat{\theta}_n = n^{-1}\sum_{i=1}^{n} X_i$. This approach can easily be adapted to censored data. Then criterion (2), $\mathcal{R}_n(1)$ and $\mathcal{K}_n(1)$ are explicit. In particular, it can be easily proved that $\mathcal{R}_n(1) \sim (3n)^{-1}$ for large values of n.

2. *Empirical Posterior Prior (EmPP).* Choose L a number of MTS to be sampled from data $\mathbf{X_n}$. Formally, $L = n$ if data are i.i.d. (applied in the following numerical experiments). See Berger and Perrichi (2002) and Varshavsky (1995) for more precisions. Then define

$$\pi_2^{MIA}(\theta) = \frac{1}{L}\sum_{l=1}^{L} \pi^J(\theta|X_l),$$

each $\pi^J(\theta|X_l)$ being an exponential $\mathcal{G}(1, X_l)$ distribution.

3. *Expected Posterior Prior (ExPP).* The third candidate is a common distribution originally elicited as an intrinsic prior (Berger and Perrichi, 2002). Noting that it is elicited using $g_3(x) = p(x|\hat{\theta}_n)$ in rule (3), define

$$\pi_3^{MIA}(\theta) = \frac{\hat{\theta}_n}{\left(\hat{\theta}_n + \theta\right)^2}.$$

MIA selection. On Figure 1 is displayed the evolution of quantities $\mathcal{R}_n(i)$ and $\mathcal{K}_n(i)$ with respect to size n. Marginal densities in selection rules (C_1) and (C_2) depend twice on data since MIA candidates are data-dependent. Then computations of $\mathcal{R}_n(i)$ and $\mathcal{K}_n(i)$ have to be done using importance sampling from independent prior distributions. Kullback-Leibler distances have been computed by Monte Carlo integration with 30,000 particles.

Convergence to 0 and divergence plots on Figure 1 respectively say that both criteria (C_1) and $C_2)$ will fortunately select the conjugate SCP prior as the less informative MIA candidate. However, criterion $C_2)$ can select a MIA candidate far from the MLE since the posterior-prior distance increases with n. It explains the differences between hierarchies. Indeed central values of the EmPP prior are far from $\hat{\theta}_n$, unlike SCP and ExPP priors. This leads us to prefer criterion (C_1).

Data-agreement of a conjugate prior. We sampled a 10-sized data-set $\mathcal{E}$ from an exponential distribution with parameter $\theta_0 = 150^{-1}$:

$$\mathcal{E} = (142.76, 142.99, 470.3, 419.09, 185.20, 84.41, 8.13, 27.15, 573.17, 17.12)$$

The MLE $\hat{\theta}_n = 207^{-1}$ underestimates the real value. We give in Table 1 the values of the data-agreement criterion (2) and its arithmetic version (4) with respect to the variations of a conjugate prior $\pi(\theta) = \mathcal{G}(a, aX_e)$ where X_e is a central value given by an expert and a is directing the prior variance $\text{Var}[\theta] = (a.X_e^2)^{-1}$.

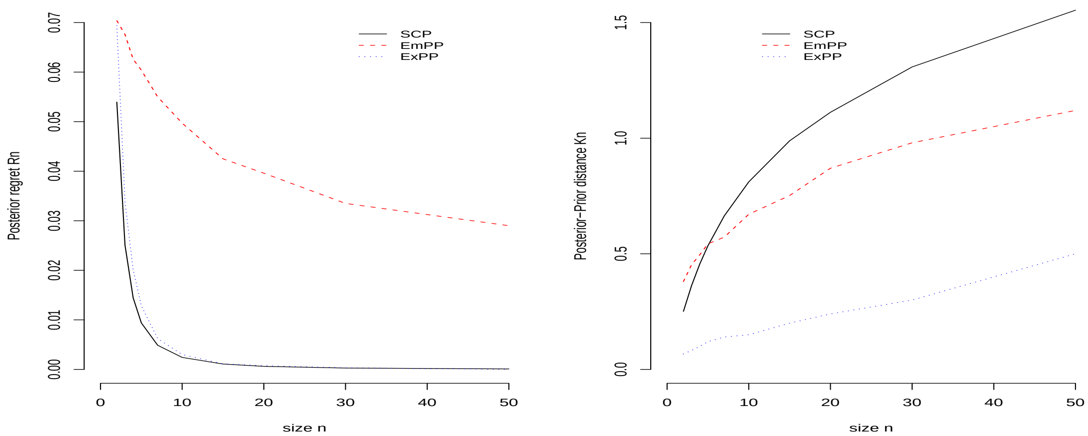

Figure 1: Evolutions of $\mathcal{R}_n(i)$ (left) and $\mathcal{K}_n(i)$ (right) for the MIA candidates (SCP: $i = 1$, EmPP: $i = 2$ and ExPP: $i = 3$) with respect to size n.

The integer a is equivalent to the size of a fictitious exponential sample $\tilde{\mathbf{X}}_{\mathbf{a}}$ with mean X_e. Since $\pi(\theta) \propto \pi^J(\theta)\mathcal{L}(\tilde{\mathbf{X}}_{\mathbf{a}}, \theta)$ where $\mathcal{L}(X_1, \ldots, X_n, \theta)$ is the exponential likelihood, thus π can be seen as a posterior prior too. Hyperparameter X_e is made varying from 10 to 500, with best choice at $\bar{X} = 207$, while a is chosen to correspond to fictitious sizes 5 and 10. It means that the prior translates as much information as the half and the complete data sample, respectively.

From these results some interesting points can be highlighted.

1. We note the proximity of results between criterion (2) and (4) in domains of data-agreement. The SCP selected prior is slightly more restrictive. When detection of conflict is the main aim, the MIA elicitation and selection can be avoided using the arithmetic criterion (4).

2. The evolution of statistics follows an intuitive and predictible vision. The better data-agreement follows the decreasing of $|X_e - \bar{X}|$. With the increasing of a (so the decreasing of the prior variance), the size of the data-agreement interval is decreasing. Criteria (2) and (4) are minimized in $a = 10$ (the size of the real sample) when the expert opinion is perfectly in agreement with data ($X_e = 207$).

3. Even a good expert value can be finally rejected if the prior is too informative (see Fig.2). In other words, the data-agreement criteria discard unbiased priors with small variances and very biased priors with large variances.

		SCP	EmPP	ExPP	
Prior proposals π		Values of $\mathcal{C}oh_i(\pi;\mathbf{X_n})$			$\mathcal{C}oh_A(\pi;\mathbf{X_n})$
$a=5$	$X_e=10$	12.97	8.750	9.823	8.805
	$X_e=150$	0.444	0.364	0.235	0.316
	$X_e=207$	0.156	0.097	0.015	0.118
	$X_e=300$	0.638	0.282	0.402	0.437
	$X_e=500$	3.444	1.827	2.543	2.323
$a=10$	$X_e=10$	25.63	17.24	17.95	17.40
	$X_e=150$	0.580	0.560	0.402	0.424
	$X_e=207$	0	0.021	0.004	0.003
	$X_e=300$	0.966	0.386	0.679	0.665
	$X_e=500$	6.580	3.481	4.645	4.436

Table 1: Values of criteria $\mathcal{C}oh_i$ and $\mathcal{C}oh_A$ for prior choices $\pi(\theta)=\mathcal{G}(a,a.X_e)$ for data set $\mathcal{E}$.

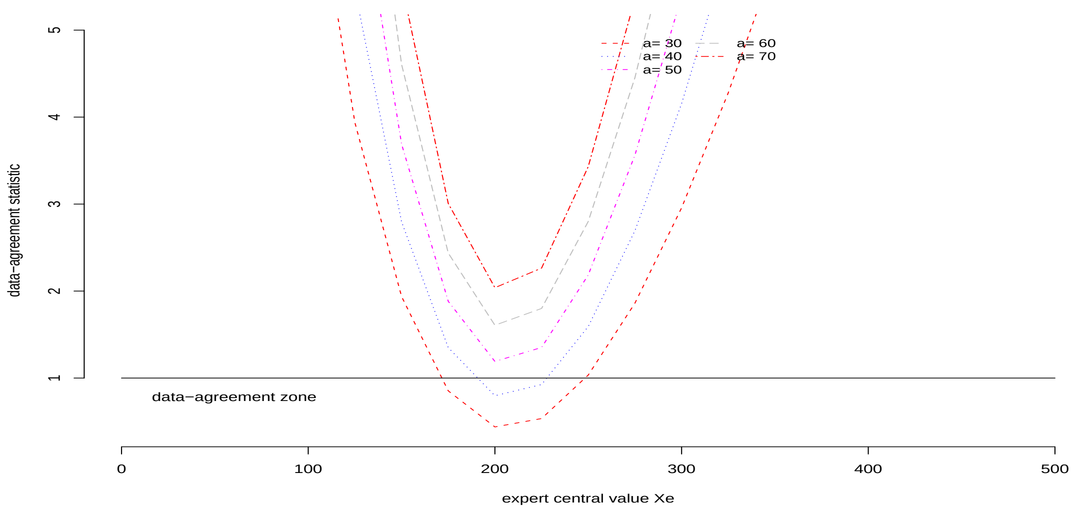

Figure 2: Evolution of $\mathcal{C}oh_A(\pi;\mathbf{X_n})$ in function of X_e around the best expert choice ($\simeq 207$), for several increasing variances, for data set $\mathcal{E}$.

4.2 The Weibull model

Bayesian inference with the Weibull distribution raises some complications, since it does not accept conjugate priors. The specification of any prior leads to use numerical tools as acceptation-rejection algorithms, MCMC or importance sampling technics for computing the posterior distribution. In the following, calculations of integration coefficients and posterior sampling will be done using Population Monte Carlo algorithms (Celeux et *al.* 2003, Cappé et *al.* 2004, Douc et *al.* 2005), which provide adaptative mixtures of densities converging to the posterior distribution in terms of Kullback divergence.

We suppose to have $\mathbf{X_n} = (X_1, \ldots, X_n)$ data with pdf $p(x|\theta) = \beta\eta^{-\beta}x^{\beta-1}\exp(-\eta^{-\beta}x^{\beta})$. Following Sun (1997), we consider the second-order matching reference prior $\pi^J(\eta,\beta) \propto (\eta\beta)^{-1}$. A MTS for the Weibull model, in the uncensored case, is a couple $(X_i, X_j) > 1$, $(X_i \neq X_i)$. Note that we define no hierarchy between the two parameters, i.e. we do not consider the reference prior in the case of a nuisance parameter.

4.2.1 An explicit MIA candidate.

In absence of convenience priors, the posterior prior approach seems a natural way of elicitation and the most intuitive idea is to use the EmPP candidate. Note that for any MTS (X_i, X_j) the posterior of the reference prior is (cf. Berger et *al.* 1998)

$$\pi^{ij}(\eta,\beta) \quad = \quad (2(X_iX_j)|\log X_i/X_j|)^{-1}\,(X_iX_j)^{\beta-1}\,\beta\eta^{-2\beta-1}\exp\left(-\eta^{-\beta}\left(X_i^{\beta}+X_j^{\beta}\right)\right)$$

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 improper prior is $\pi^J(\mu,\beta) \propto (\mu\beta^2)^{-1}$. Thus $\pi^{ij}(\mu,\beta) = \pi^{ij}(\mu|\beta)\,\pi^{ij}(\beta)$ with

$$\begin{aligned}
\pi^{ij}(\mu|\beta) &= \mathcal{G}\left(2, X_i^{\beta}+X_j^{\beta}\right), \\
\pi^{ij}(\beta) &= \frac{(X_iX_j)^{\beta-2}}{2|\log X_i/X_j|\left(X_i^{\beta}+X_j^{\beta}\right)^2}.
\end{aligned}$$

Deciding of numbers L_1 and L_2 such that $L = L_1L_2$ is the total number of possible MTS, the reparametrized EmPP prior and posterior are

$$\begin{aligned}
\pi_1^{MIA}(\mu,\beta) &= \frac{1}{L_1L_2}\sum_{i=1}^{L_1}\sum_{i=1}^{L_2}\pi^{ij}(\mu|\beta)\,\pi^{ij}(\beta), \\
\pi_1^{MIA}(\mu,\beta|\mathbf{X_n}) &= \sum_{i=1}^{L_1}\sum_{j=1}^{L_2}\alpha_{ij}\,\pi^{ij}(\mu|\beta,\mathbf{X_n})\,\pi^{ij}(\beta|\mathbf{X_n})
\end{aligned}$$

with

$$\pi^{ij}(\mu|\beta,\mathbf{X_n}) = \mathcal{G}\left(n+2,\sum_{k=1}^{n}X_k^\beta+X_i^\beta+X_j^\beta\right),$$

$$\pi^{ij}(\beta|\mathbf{X_n}) \propto \beta^n \frac{(X_iX_j)^\beta\left(\prod_{k=1}^{n}X_k\right)^\beta}{\left(\sum_{k=1}^{n}X_k^\beta+X_i^\beta+X_j^\beta\right)^{n+2}}$$

and

$$\alpha_{ij} = \frac{\nu_{ij}}{\sum_{p=1}^{L_1}\sum_{q=1}^{L_2}\nu_{pq}},$$

$$\nu_{ij} = \left\{2|\log X_i/X_j|\left(\prod_{k=1}^{n}X_k\right)(X_iX_j)^2\right\}^{-1}\int_0^\infty \beta^n \frac{(X_iX_j)^\beta\left(\prod_{k=1}^{n}X_k\right)^\beta}{\left(\sum_{k=1}^{n}X_k^\beta+X_i^\beta+X_j^\beta\right)^{n+2}}.$$

The EmPP prior is interesting since it is explicit. Moreover, the choice of the new parametrization makes the posterior sampling easier. A unique sampling algorithm with Gibbs sampling step is needed for the whole computation of EmPP and arithmetic data-agreement criteria.

4.2.2 Tests of data-agreement.

In Table 2 right-censored lifetime data $X_1,\ldots,X_n$ $(n=18)$ from EDF nuclear devices (used in secondary water circuit) are given. Data are given in years and multiplied by 10 (for readibility). For physical reasons and according to a large consensus, those data are assumed to come from a Weibull distribution. 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 unusual in reliability problems and pleads (like the large estimated standard deviation) for a Bayesian estimation. Indeed, $\beta>2$ characterizes a system whose aging failure rate is accelerated (see for instance Lawless 1982). When aging is assumed, the variation domain $D_\beta=[1,5]$ is usually considered as reasonable.

Two prior opinions are available on X, given by independent experts $\mathcal{E}_1$ and $\mathcal{E}_2$. They are summarized in Table 3. The $\mathcal{E}_1$ opinion is largely more informative than $\mathcal{E}_2$ and both are right-shifted with respect to data (note that censored data may hide a real data-agreement for the priors). This is an example of incoherence of time between past data and expert opinions taking account of technical evolution, as introduced in Section 1. Moreover they are

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 2: Lifetime data from EDF nuclear device.

not interrogated at the same level of precision. $\mathcal{E}_1$ is a nuclear operator and speaks for one particular component while $\mathcal{E}_2$ can be seen as a component producer whose opinion takes account of a variety of running conditions. So the width of $\mathcal{E}_2$ opinion is due to explicative variables as environmental constraints.

	(5%,95%) Interval	Median value
Expert $\mathcal{E}_1$	(200,300)	250
Expert $\mathcal{E}_2$	(100,500)	250

Table 3: Expert opinions on Weibull observable X.

Prior modelling. We consider several modellings which have close variances and means to observe the consistency properties of data-agreement criteria. Independence between parameters is considered for simplicity.

- *Shape parameter* β. Let β_0 be the mean and α_β directing the evolution of priors

$$\pi_1(\beta|\beta_0,\alpha_\beta) = \mathcal{G}(\alpha_\beta, \alpha_\beta\beta_0^{-1}), \qquad \pi_2(\beta|\beta_0,\alpha_\beta) = \mathcal{N}^+\left(\beta_0, \sigma_\beta^2 = \frac{\beta_0^2}{\alpha_\beta}\right)$$

  where $\mathcal{N}^+$ is a normal distribution truncated in 0. Its mode being smaller than its mean, a gamma prior favorizes the lowest values of β and constitutes a versatile prior on the shape parameter.

- *Scale parameter* η. Since η is the 63th order percentile of the Weibull distribution and is more tractable from the expert opinion than β, we use the Weibull mean formula, choosing $\tilde{\beta} = 3$ (the middle of D_β), for obtaing prior domains of Table 4. In the same Table, hyperparameter values (α_η, t) are proposed to calibrate following priors

$$\begin{aligned}\pi_1(\eta|t,\alpha_\eta) &= \mathcal{G}\left(\alpha_\eta, \alpha_\eta\frac{\Gamma(1+1/\tilde{\beta})}{t}\right),\\ \pi_2(\eta|t,\alpha_\eta) &= \mathcal{N}^+\left(\frac{t}{\Gamma(1+1/\tilde{\beta})}, \sigma_\eta^2 = \frac{t^2}{\alpha_\eta\ \Gamma^2(1+1/\tilde{\beta})}\right).\end{aligned}$$

	(5%,95%) Interval	Median value	t	α_η	σ_η
Expert $\mathcal{E}_1$	(224,336)	280	252	60	36.4
Expert $\mathcal{E}_2$	(112,560)	280	269	4.5	142.0

Table 4: Prior domains and hyperparameters values for $\pi(\eta)$ in the Weibull distribution.

Prior comparisons. We compare in Figure 3 the impact on $\mathcal{C}oh$ and $\mathcal{C}oh_A$ of variations on prior distributions $\pi_{ij}(\eta,\beta) = \pi_i(\eta)\pi_j(\beta)$, assuming variations of β_0 and a constant $\sigma_\beta = 1.5$. $L = 30$ randomized MTS are used from uncensored data. These results exemplify a slightly more stable behavior of the arithmetic data-agreemeent criterion (4). For each expert opinion, we present in Figures 4 and 5 the evolutions of $\mathcal{C}oh_A$ with respect to several values of β_0 and σ_β. The data-agreement domains logically grow with σ_β.

Results allows to conclude that the discrepancy due to the time incoherence between expert and data is detected, especially for the $\mathcal{E}_1$ expert opinion. This conflict between $\mathcal{E}_1$ prior and data is not acceptable since the reasonable choice $\beta_0 = 1$ (which means a constant failure rate) and $\sigma_\beta \leq 2.5$ does not lead to an agreement. Work has to be done for explaining the origin of data or toning down those strong expert beliefs.

5 Conclusion

Reliability is a domain in which data-agreement notions are natural and useful. Using expert opinions is frequent in industrial issues where lifetime data are collected with difficulty. One of the most interesting properties of the data-agreement criterion is to accept only prior distributions with reasonable variance and bias. In particular, the arithmetic version (4) provides a stable estimation of the potential discrepancy between prior and data. It can be generalized to all models for which an noninformative prior π^J can be defined from arguments given in Section 3. The selection of MIA candidates can be seen as a tool of modelling benchmark proper densities, as informative as vague expert opinions. Thus the criterion will accept only ***sufficiently influent*** priors. An open issue is to obtain large ranges of these reasonable candidates.

We believe that the data-agreement test constitutes a useful preliminary to subjective Bayesian inference, especially when no former judgment on experts is available. The detection of a potential conflict must help to correct subjective beliefs, turn down outliers or judging the joint trust into data and expert opinions. Moreover, the use of posterior priors as benchmark densities allows to calibrate the subjective prior information. When data are assumed to be correct, the criterion can work as a simple and practical tool of calibration. Indeed, choose $\pi(\theta) = \pi(\theta|\omega,\theta_e)$ where θ_e is a central value of θ (elicited from the expert viewpoint) and ω directs the prior uncertainty ; then select ω such as $\mathcal{C}oh(\omega;\mathbf{X_n}) = 1$.

Choosing ω in such a way allows to get a compromise between prior variance and bias which makes sense.

6 Acknowledgments

The author would like to thank his advisor Gilles Celeux, from INRIA, and Jean-Michel Marin, from CEREMADE, INRIA and Université Paris Dauphine, for many fruitful discussions and advices. The idea of this study has been proposed from experimental issues by François Billy and Emmanuel Remy from EDF Research & Devt division. Many thanks to Christian P. Robert, Anthony O'Hagan, Guido Consonni and Piero Veronese for their advices and questionings.

7 Appendix: asymptotics for the exponential model

As Hartigan says (1983): *"the prior density does not affect the asymptotic distribution of θ in the terms of $O(1)$ or $O(n^{-1/2})$"*. Under classical regularity conditions on MIA densities and likelihood, the asymptotic normality of $\pi^{MIA}(.|\mathbf{X_n})$ should make the criterion (2) accept any reasonable prior. Applying the Central Limit Theorem (CLT) to (2), we prove in an illustrative purpose the following theorem of asymptotic data-agreement for the exponential model $p(x|\theta) = \theta\exp(-\theta x)\ 1_{\{x>0\}}$ then explain briefly our meaning of "reasonable".

THEOREM 1. *Denote $\theta_0 > 0$ the true value of θ. Suppose that $\pi^{MIA}(\theta)$ is the $\mathcal{G}\left(1, n^{-1}\sum_{i=1}^n X_i\right)$ (SCP) prior distribution. When $\pi(\theta)$ is a conjugate $\mathcal{G}(a,b)$ distribution, such as $a\log b\theta_0 + b\theta_0 - \log\Gamma(a) + 1 > 0$, then*

$$\lim_{n\to\infty} P\left(\{\mathcal{C}oh(\pi;\mathbf{X_n}) - 1\} \leq 0\right) = 1.$$

Proof. Note $Y_n = n^{-1}\sum_{i=1}^n X_i$. With $X \rightsquigarrow \mathcal{E}(\theta_0)$, $Y_n \rightsquigarrow \mathcal{G}(n, n\theta_0^{-1})$.
By CLT, $\sqrt{n}\left(Y_n - \theta_0^{-1}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\theta_0^{-2})$. Note E_n the mean with respect to $\pi^{MIA}(.|\mathbf{X_n})$. The criterion (2) can be written as $\mathcal{C}oh(\pi;\mathbf{X_n}) = 1 - \Delta_n^{-1}\mathrm{E}_n[\log\pi(\theta)/\pi^{MIA}(\theta)]$ where

$$\begin{aligned}\Delta_n &= KL\left(\pi^{MIA}(.|\mathbf{X_n})\ ||\ \pi^{MIA}\right),\\ &= \frac{1}{2}\log n - \frac{1}{2}(\log 2\pi - 1) + \frac{5}{6n} + o(n^{-1})\end{aligned}$$

from Penny (2001) and Abramowitz and Stegun (1972, p.258-259)). Moreover,

$$\mathrm{E}_n[\log\pi(\theta)/\pi^{MIA}(\theta)] = g(a,b) + t_a(n) - U_n$$

where $g(a,b) = a\log b - \log\Gamma(a) + 1$, $t_a(n) = (a-1)[\Psi(n+1) - \log(n+1)] = O(n^{-1})$ and $U_n = \varphi_{a,b}(Y_n)$ with $\varphi_{a,b}(x) = a\log x - b/x$. Ψ is the digamma function. Since $\varphi_{a,b}$ is

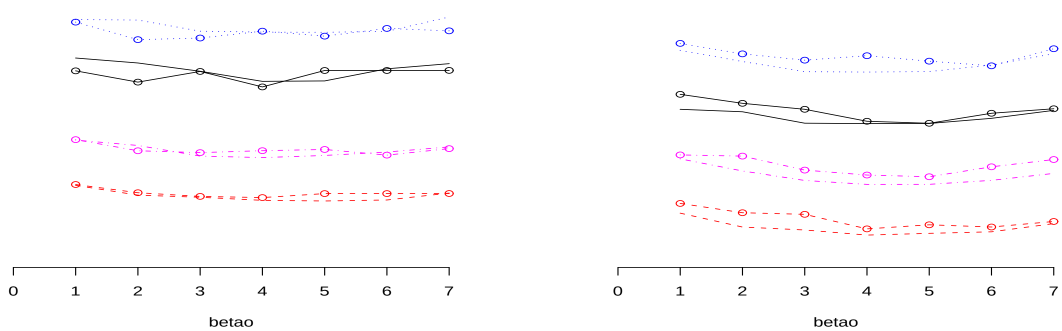

Figure 3: Parallel evolutions of $\mathcal{C}oh$ *(-o-)* and $\mathcal{C}oh_A$ according to β_0 variations. Left side: $\mathcal{E}_1$ opinion modelling (resp. $\mathcal{E}_2$). From top to bottom: succesive choices of $\pi_2(\beta)\pi_1(\eta)$, $\pi_1(\beta)\pi_1(\eta)$, $\pi_1(\beta)\pi_2(\eta)$ and $\pi_2(\beta)\pi_2(\eta)$. Scales on $y-$axis are shifted and removed for readability.

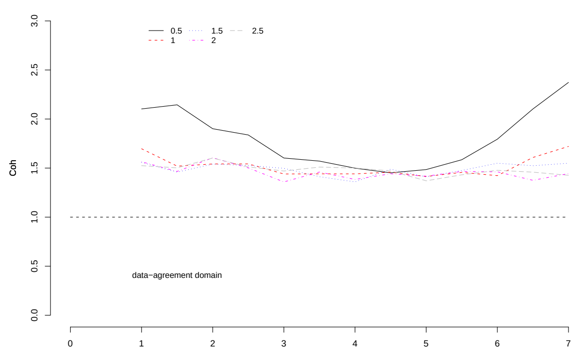

Figure 4: Evolution of $\mathcal{C}oh_A(\pi; \mathbf{X_n})$ for $\mathcal{E}_1$ prior modellings $\pi_1(\eta)\pi_(\beta)$ in function of β_0 and σ_β.

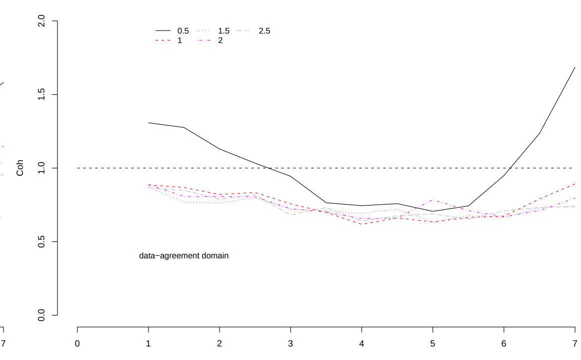

Figure 5: Evolution of $\mathcal{C}oh_A(\pi; \mathbf{X_n})$ for $\mathcal{E}_2$ prior modellings $\pi_1(\eta)\pi_(\beta)$ in function of β_0 and σ_β.

differentiable on $x > 0 \; \forall (a,b) > 0$,

$$\sqrt{n}\left\{U_n - \varphi_{a,b}(\theta_0^{-1})\right\} \quad \xrightarrow{\mathcal{L}} \quad \mathcal{N}\left(0, \left[\theta_0^{-1}\varphi'_{a,b}(\theta_0^{-1})\right]^2\right)$$

so $$\Delta_n\sqrt{n}\;\left[\{\mathcal{C}oh(\pi;\mathbf{X_n}) - 1\} - \Delta_n^{-1}\left\{\varphi_{a,b}(\theta_0^{-1}) - g(a,b)\right\}\right] \quad \xrightarrow{\mathcal{L}} \quad \mathcal{N}\left(0, \left[\theta_0^{-1}\varphi'_{a,b}(\theta_0^{-1})\right]^2\right).$$

Finally, denoting Φ the standard normal cumulative distribution, the probability of data-agreement when n is large enough is

$$\begin{aligned} P\left(\{\mathcal{C}oh(\pi;\mathbf{X_n}) - 1\} \leq 0\right) \quad &= \quad \Phi\left(\frac{\left\{g(a,b) - \varphi_{a,b}(\theta_0^{-1})\right\}\sqrt{n}}{\theta_0^{-1}|\varphi'_{a,b}(\theta_0^{-1})|}\right), \\ &= \quad \Phi\left(\frac{\{a\log b\theta_0 + b\theta_0 - \log\Gamma(a) + 1\}\sqrt{n}}{a + b\theta_0}\right), \\ &\xrightarrow{n\to\infty} \quad 1 \quad \text{if} \;\; a\log b\theta_0 + b\theta_0 - \log\Gamma(a) + 1 > 0. \;\; \square \end{aligned}$$

Let choose hyperparameters $(a, b = a\theta_e^{-1})$ (as in § 4.1) where θ_E embodies a central prior value elicited from an expert opinion. Then $P(\{\mathcal{C}oh(\pi;\mathbf{X_n}) - 1\} \leq 0) \to 1 \Leftrightarrow \tau(a, \theta_0/\theta_e) > 0$ where

$$\tau(a,x) \quad = \quad x + \log x + \log a + a^{-1}(1 - \log\Gamma(a)).$$

From contour plots in Figure 6 we notice that τ is nonpositive when π is less informative than the benchmark density π^{MIA} or extremely biased with respect to θ_0. This brief asymptotic study highlights that π^{MIA} must be selected as few informative as possible.

References

Abramowitz, M. and Stegun, I. A. (Eds.). (1972). "Bernoulli and Euler Polynomials and the Euler-Maclaurin Formula." in *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, *J.Roy.Stat.Soc. Ser.B*, 28, 131-142.

Andrieu, C., Doucet, A., Fitzgerald, W.J. and Pérez, J.M. (2001). *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, 81-108.

Bayarri, M.J. and Berger, J.O. (1997).*Measures of Surprise in Bayesian Analysis*, ISDS Discussion Paper 97-46, Duke University.

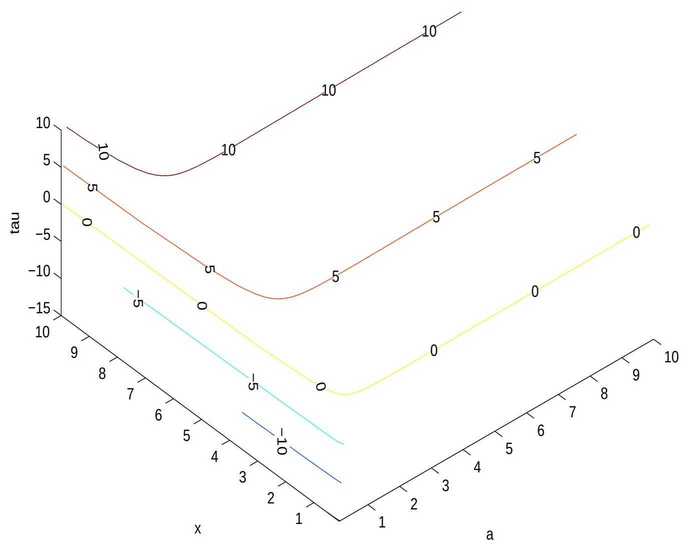

Figure 6: Contour plots of τ with respect to a and x.

Bayarri, M.J. and Berger, J.O. (1998).Quantifying Surprise in the Data and Model verification (with discussion), *Bayesian Statistics 6*, J.M. Bernardo, et.al., eds., 53-82, Oxford University Press.

Berger, J.O and Bernardo, J.M. (1992). On the development of reference priors (with discussion). In: J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, Eds., *Bayesian Statistics*, Vol.4, Oxford University Press, Oxford, 35-60.

Berger, J.O. and Perrichi, L.R. (1996). The Intrinsic Bayes Factor for Model Selection and Prediction, *Journal of the Americal Statistical Association*, Vol.91, No.433, Theory and Methods.

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

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

Berger, J.O. and Sun, D. (1993). Bayesian analysis for the Poly-Weibull Distribution, *Journal Amer. Statis. Assoc.*, 88, 1412-1418.

Berger, J.O. and Sun, D. (1994). Bayesian sequential reliability for Weibull and related distributions, *Ann. Inst. Statis. Math.*, 46, 221-249.

Bernardo, J.M. (1979). Reference Posterior Distributions for Bayesian Inference, *J. R. Statist. Soc.*, B, 41, 113-147.

Bernardo, J.M. (1997). "Noninformative Priors Do Not Exist: A Discussion", *J. Statis. Planning and Inference*, **65**, 159-189 (with discusssion).

Box, G.E.P. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*, Reading, MA:Addison-Wesley.

Cappé, O., Guillin, A., Marin, J.M. and Robert, C.P. (2004). Population Monte Carlo, *Journal of Computational and Graphical Statistics*, 13, 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. (1999). *Asymptotic Normality of the Posterior in Relative Entropy*, IEEE Trans. Infor. Theory, 45,1,165-176.

Clarke B.S. et Barron A.R. (1990). *Information-Theoretic Asymptotics of Bayes Methods*, IEEE Trans.Infor.Theory,36,3,453-471.

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

Datta, G.S. and Ghosh, J.K. (1995). On priors providing frequentist validity for Bayesian inference, *Biometrika*, 82, pp.37-45.

Datta, G.S. (1996). On priors providing frequentist validity for Bayesian inference for multiple parametric functions, *Biometrika*, **83**, 287-298.

Dawid, A.P. (1982). The Well-calibrated Bayesian, *Journal of the Americal Statistical Association*, Vol.77, No.379, Theory and Methods, 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, 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, 193-209.

Douc, R., Guillin, A., Marin, J.M. and Robert, C.P. (2005). Convergence of adaptive sampling schemes, Research report n.5485, INRIA Futurs.

Ghosh, J.K. and Mukerjee, R. (1993). On priors that match posterior and frequentist distribution functions, *Can. J. Statis*, **21**, 89-96.

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

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

Johnson R.A. (1970). *Asymptotic expansions associated with posterior distributions*, Ann. Math. Statist., 41,851-854.

Kass R.E. and Wasserman L. (1996). The selection of prior distributions by formal rules, *J. Amer. Stat. Ass.*, Vol.91,No.435, 1343-1370.

Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data*, Wiley series in probability and mathematical statistics.

Lucas, W. (1993). When is Conflict Normal ?, *J. Amer. Stat. Ass.*, Vol.88, No.424, 1433-1437.

O'Hagan, A. (1979). On outlier rejection phenonema in Bayes inference, *J. Roy. Statist. Soc. B*, **41**, 358-367.

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

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

O'Hagan, A. (1998). Eliciting expert beliefs in substantial practical applications, *The Statistician*, **47**, Part.1, 21-35.

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), 423-453. Oxford University Press.

Peers, H.W. (1965). On confidence sets and Bayesian probability points in the case of several parameters, *J. Roy. Statis. Soc. Ser. B*, **27**, 9-16.

Pérez, J.M. (1998). *Development of Conventional Prior Distributions for Model Comparisons*, Ph.D. Thesis, Purdue University.

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

Sinanović, S. and Johnson, D.H. (2003). Towards a Theory of Information Processing, J. Franklin Institute.(Submitted).

Singpurwalla, N.D. and Mao, S.S. (1988) "Reliability Analysis using Weibull Lifetime Data and Expert Opinion.", *IEEE Transactions on Reliability*, 3, 340-7.

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

Sweeting, T.J., Datta, G.S. and Ghosh M. (2005). *Nonsubjective priors via predicitive relative entropy regret*, Research Report No.253, Dpt of Statistical Science, University College London.

Usureau, E. (2001). *Application des méthodes bayésiennes pour l'optimisation des coûts de développement des produits nouveaux*, Ph.D. Thesis n°413, Institut des Sciences et Techniques d'Angers.

Varshavsky, J. (1995). *On the Development of Intrinsic Bayes Factors*, Ph.D. thesis, Purdue University.

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