Bayesian Analysis (2012) **7**, Number 2, pp. 411–438

Log-Linear Pool to Combine Prior Distributions: A Suggestion for a Calibration-Based Approach

M. J. Rufo*, J. Martín† and C. J. Pérez‡

Abstract. An important issue involved in group decision making is the suitable aggregation of experts' beliefs about a parameter of interest. Two widely used combination methods are linear and log-linear pools. Yet, a problem arises when the weights have to be selected. This paper provides a general decision-based procedure to obtain the weights in a log-linear pooled prior distribution. The process is based on Kullback-Leibler divergence, which is used as a calibration tool. No information about the parameter of interest is considered before dealing with the experts' beliefs. Then, a pooled prior distribution is achieved, for which the expected calibration is the best one in the Kullback-Leibler sense. In the absence of other information available to the decision-maker prior to getting experimental data, the methodology generally leads to selection of the most diffuse pooled prior. In most cases, a problem arises from the marginal distribution related to the noninformative prior distribution since it is improper. In these cases, an alternative procedure is proposed. Finally, two applications show how the proposed techniques can be easily applied in practice.

Keywords: Bayesian analysis, Kullback-Leibler divergence, Pooled distribution

1 Introduction

The problem of aggregating experts' information is an important issue when dealing with decision making. For many years, this topic has received considerable attention in specialized literature. Abbas (2009) summarized the problem in a succinct manner as follows: A decision maker is interested in a certain quantity, X, described as a random variable. He/She consulted several experts who provided their information about a parameter, θ, as probability distributions. It is necessary to observe that the previous parameter, θ, is related to the random variable, X, through a statistical model. Furthermore, the decision maker needs to combine the experts' distributions to obtain an aggregated probability distribution. Notable reviews regarding this subject are provided by Genest and Zidek (1986), Ouchi (2004), and Clemen and Winkler (2007) and references therein. Clemen and Winkler (2007) classified the methods to get the aggregated opinion of a group of experts in two types: mathematical and behavioral methods. The classification of the mathematical methods were further divided as either axiomatic or Bayesian approaches. Two useful axiomatic approaches that combine

*Escuela Politécnica, Departamento de Matemáticas, Universidad de Extremadura, Cáceres, Spain, mrufo@unex.es

†Facultad de Ciencias, Departamento de Matemáticas, Universidad de Extremadura, Badajoz, Spain, jrmartin@unex.es

‡Facultad de Veterinaria, Departamento de Matemáticas, Universidad de Extremadura, Cáceres, Spain, carper@unex.es

 DOI:10.1214/12-BA714

individual experts' probability distributions to produce a joint probability distribution are the linear opinion pool (see Stone (1961) and Diaconis and Ylvisaker (1985)) and the logarithmic opinion pool (see Bacharach (1975), Genest (1984), Genest and Zidek (1984) and references therein). Opinion pooling can suffer from paradoxes (see O' Hagan et al. (2006) and references therein). Hence, for instance, an opinion pool might be expected to have the externally Bayesian property (see Garthwaite et al. (2004)). Except in trivial cases, the linear opinion pool fails to have this property. On the contrary, the logarithmic opinion pool is externally Bayesian when the weights sum up to a unity (see Kadane et al. (1999) and Genest (1984)). However, it does not have a second desirable property, which is invariance to event combination (see Garthwaite et al. (2005)). McConway (1981) shows that the linear opinion pool is the only combination which satisfies this marginalization criterion. Hence, it is not possible to find a formula for mathematical aggregation which satisfies both requirements. More recent papers dealing with both opinion pools in different areas are, among others, Poole and Raftery (2000), Abbas (2009) and Kascha and Ravazzolo (2010).

The determination of the weights is a problem which arises when using both linear pool and logarithmic-linear pool (see Genest and Zidek (1986)). In the literature, different methods have been proposed on how to choose them. Thereby, a frequently used strategy consists of assigning equal weights to each expert when there is nothing which suggests that an expert's opinion is better than any other one. Another classic way to obtain the weights is using Cooke's method. This method was designed to avoid uncertainty when the weights must be assigned. Following Cooke (1991), an expert is precise if he/she is well calibrated and his/her opinions are informative. At the same time, an expert is well calibrated when his/her assessed probabilities agree with actual observed frequencies. Cooke's method calculates an expert's weight using a score that is a combination of separate calibration and information scores. The design of the right set of parameters (seed variables) used to assign the weights is the main problem found in this method. See Cooke (1991) for additional information together with some examples and also see Clemen (2008). A method based on the Kullback-Leibler divergences to obtain the weights was proposed by Heskes (1998). Specifically, he classified the method behind his analysis as a supra-Bayesian method (see for instance Jacobs (1995) and Roback and Givens (2001)) and a simple heuristic method (see for instance Tversky and Kahneman (1974)). Finally, Kascha and Ravazzolo (2010) compared some common approaches for combining density forecasts. The two possible ways of aggregation are considered together with three different methods to obtain the weights. They are equal weights, recursive log score weights and (inverse) mean squared error weights. A full description and examples of these methods can be found in Kascha and Ravazzolo (2010).

Bousquet (2008) proposed a criterion based on Kullback-Leibler divergence to assess a possible conflict between the prior and the data. His work is focused on an industrial context where experts' opinions are frequently used. By using this divergence measure as a calibration tool, this work provides a general approach to assess the weights in a log-linear pooled prior distribution.

Throughout this paper, it is assumed that several experts supply prior information about a parameter, θ, as proper prior distributions. Then, the decision maker combines them through a log-linear pool. Merging these distributions, in this way, is proposed because it maintains unimodality and it is externally Bayesian. Such as is pointed out in Faria and Mubwandarikwa (2008), external Bayesianity preserves immunity of influence on decision making. This means that, when there is an agreed likelihood function, the opinion pool of the posterior distributions should coincide with the posterior distribution obtained from the opinion pool of the prior distributions (see Garthwaite et al. (2004)). Therefore, a decision maker would not have to deal with two different posterior distributions and a way of choosing between the two optimal decisions.

Next, the weights have to be assessed to obtain the full aggregated prior distribution. In order to do it, the problem is formulated as a decision problem. Therefore, a suitable loss function based on Kullback-Leibler divergence is defined. The proposal consists of finding an aggregated prior distribution which considers all experts' opinions. This prior distribution belongs to the set of alternatives in the decision-making process. The states of nature are the possible values, x, that the considered random variable X can take on. Besides, the decision maker assumes prior ignorance about the states of nature through a non informative prior distribution over θ. Thus, by considering the likelihood function, the distribution for each x is given by the marginal prior distribution $m(x)$. Then, the weights are obtained by minimizing the expected loss. A problem arises when the considered distribution $m(x)$ is not proper. Consequently, the optimization problem cannot be carried out. In this case, an alternative procedure is proposed, which adequately replaces the initial marginal distribution by a proper one.

The outline of the paper is as follows. In Section 2, the calibration measure is presented. Section 3 proposes the method to obtain the weights in the combined prior distribution. A modified technique based on the previous one is presented in Section 4. In Section 5, the developed methodology is applied to several distributions. For all of them, conjugate prior distributions are considered. A discussion is presented in Section 9. Finally, the appendices contain the theoretical results.

2 Background

The use of experts' opinions is usually found in industrial issues where data are collected with difficulty. Subjective perceptions are usually needed, and can sometimes be far from the objective information yielded by the data, for instance, when technical issues arise. Thus, the analyst can check the discrepancy or the agreement and use this information in an appropriate way. In this context, Bousquet (2008) provided a statistical criterion which indicates conflict or agreement between the prior distribution and the data. Specifically, he proposed to compute the ratio:

$$DAC^J\left(\pi|\boldsymbol{x}\right) = \frac{KL\left(\pi^J(.|\boldsymbol{x})||\pi\right)}{KL\left(\pi^J(.|\boldsymbol{x})||\pi^J\right)},$$

where π^J and π denote a *noninformative* prior distribution and the proposed prior distribution, respectively. $\pi^J(.|\boldsymbol{x})$ is the posterior distribution and $KL\left(f||g\right)$ denotes the Kullback-Leibler divergence between the distributions f and g, i.e.:

$$KL\left(f||g\right) = \int_{\Theta} f\left(\theta\right) \log\left(\frac{f\left(\theta\right)}{g\left(\theta\right)}\right) d\theta.$$

Thus, if $DAC^{J}\left(\pi|\boldsymbol{x}\right) \leq 1$, then prior and data are close enough and the proposed prior π is in agreement with the data $\boldsymbol{x}$. In other cases, a conflict is detected. In addition, the Data Agreement Criterion (DAC) can be used as a tool for the calibration of subjective prior distributions. See Bousquet (2008) for additional information.

3 The method.

Let X be a random variable distributed according to a density $f(x|\theta)$, and suppose that k experts provide prior information about the parameter θ. The opinion of each expert is elicited as a proper prior distribution $\pi_j\left(\theta\right)$. In this context, there are different methods available to achieve the aggregated opinion of a group of experts (see for instance Garthwaite et al. (2004, 2005)). One of them consists of using a logarithmic opinion pool:

$$\pi_{\boldsymbol{\omega}}(\theta) = t(\boldsymbol{\omega}) \prod_{j=1}^{k} \left(\pi_j\left(\theta\right)\right)^{\omega_j}, \tag{1}$$

where $t(\boldsymbol{\omega})$ is a normalizing constant i.e.:

$$t^{-1}(\boldsymbol{\omega}) = \int_{\Theta} \prod_{j=1}^{k} \left(\pi_j\left(\theta\right)\right)^{\omega_j} d\theta,$$

and the weights, ω_j, are nonnegative and sum up to one.

Combining prior distributions, in this way, is used because the resulting combined prior distribution is frequently unimodal and less dispersed than the one obtained through a linear combination (see Rufo et al. (2009) for a Bayesian analysis by using a linear combination). In consequence, it is more likely to indicate consensual values when decisions must be made (see Genest and Zidek (1986)). Note that, in linear opinion pooling the decision maker takes into account the full range of parameter values, whereas in logarithm opinion pooling he/she focuses on the common range of parameter values (see Kascha and Ravazzolo (2010) for an illustrative example showing the main differences between the two aggregation schemes). Moreover, for the logarithmic opinion pool, it is satisfied that if an expert gives zero probability to a certain set, then the pooled distribution must also assign zero probability to that set.

Weighting factors ω_j, $j = 1, 2, \ldots, k$ are introduced to indicate the reliability of each expert. Chen and Pennock (2005), among others, observed that an important question when opinion pool methods are used is how to choose the weights. A very frequent strategy consists of considering the same weights for all experts.

Here, no fixed values for the weights are initially considered and a general procedure to obtain them is proposed. This proposal uses the Kullback-Leibler divergence as a measure to calibrate the discrepancy from the pooled prior distribution to the posterior distribution for any random variable value $x \in X$. The purpose is to minimize this discrepancy.

Following Bousquet (2006, 2008), Jeffreys' priors for one-parameter distributions are used. They are denoted by $\pi^J(\theta)$, whereas the posterior distributions are $\pi^J(\theta|x)$. Furthermore, proper posterior distributions will be taken. Note that, for the considered parameters ($\theta \subseteq \mathbb{R}$), Jeffreys' priors and refererence priors are equivalent (see Liseo (1993)). The idea behind using this kind of prior distribution is that the posterior distribution, $\pi^J(\theta|x)$, is not greatly affected by this initial information, i.e., the influence of the prior distribution is minimized. Thus, by considering the set of all possible prior distributions supporting the experts' opinions, the obtained prior distribution, $\pi_{\boldsymbol{\omega}}(\theta)$, is the one in more accordance with the information provided by the value x through $\pi^J(\theta|x)$.

Then, for all possible weight vectors $\boldsymbol{\omega} \in \chi = \{(\omega_1, \omega_2, \ldots, \omega_k) : \sum_{j=1}^{k} \omega_j = 1, \omega_j \geq 0\}$ and each value $x \in X$, the loss function is defined by:

$$L(\boldsymbol{\omega}, x) = KL(\pi^J(\cdot|x)||\pi_{\boldsymbol{\omega}}(\cdot)), \tag{2}$$

where $KL(\pi^J(\cdot|x)||\pi_{\boldsymbol{\omega}}(\cdot))$ denotes the Kullback-Leibler divergence between the combined prior distribution $\pi_{\boldsymbol{\omega}}$ (see expression (1)) and the posterior distribution $\pi^J(\theta|x)$, i.e.:

$$KL(\pi^J(\cdot|x)||\pi_{\boldsymbol{\omega}}(\cdot)) = \int_{\Theta} \pi^J(\theta|x) \ \log\left(\frac{\pi^J(\theta|x)}{\pi_{\boldsymbol{\omega}}(\theta)}\right) d\theta. \tag{3}$$

Observe that other divergence measures in the Ali-Silvey class of information-theoretic measures (Ali and Silvey (1966)) can be used. Nevertheless, it will be observed throughout this paper that the use of Kullback-Leibler divergence provides analytical advantages as well as computational simplicity.

Now, a probability distribution for the states of nature is needed. The idea is to obtain this distribution through the likelihood function $l(\theta|x) = f(x|\theta)$. The previous prior distribution $\pi^J(\theta)$ is considered, since prior ignorance of $x \in X$ (states of nature) is assumed by the decision maker. Therefore, the probability distribution for each x is given by the predictive prior distribution:

$$m(x) = \int_{\Theta} f(x|\theta)\pi^J(\theta)d\theta.$$

The objective is to find the weight vector (decision), $\boldsymbol{\omega} \in \chi$, that minimizes the expected loss, i.e.:

$$\arg\min_{\boldsymbol{\omega} \in \chi} L(\boldsymbol{\omega}), \tag{4}$$

where $L(\boldsymbol{\omega}) = E_m(L(\boldsymbol{\omega}, x)) = E_m(KL(\pi^J(\cdot|x)||\pi_{\boldsymbol{\omega}}(\cdot)))$.

Note that, through this procedure, only the information provided by each expert is combined. Therefore, the combined prior distribution is not constructed based on a specific data set but under the assumption that this set is unknown. Relaxations of this assumption are considered in the Discussion Section. In addition, other methods to obtain the distribution $m(x)$, for $x \in X$, could be used.

By taking into account the expression of the Kullback-Leibler divergence (3), the function $L(\boldsymbol{\omega})$ can be written as:

$$L(\boldsymbol{\omega}) = E_m\left(E_{\pi^J}\left(\log \pi^J\left(\cdot|x\right)\right)\right) - E_m\left(E_{\pi^J}\left(\log \pi\left(\cdot\right)\right)\right). \tag{5}$$

The first summand on the right-hand side of the previous expression does not depend on $\boldsymbol{\omega} \in \chi$. Thus, the initial optimization problem is equivalent to the following one:

$$\max_{\boldsymbol{\omega}\in\chi} E_m\left(E_{\pi^J}\left(\log \pi\left(\cdot\right)\right)\right).$$

By considering the expression for the pooled prior distribution, $\pi_{\boldsymbol{\omega}}(\theta)$, it is satisfied:

$$E_{\pi^J}\left(\log \pi_{\boldsymbol{\omega}}(\theta)\right) = \log t(\boldsymbol{\omega}) + \sum_{j=1}^{k} \omega_j E_{\pi^J}\left(\log \pi_j\left(\theta\right)\right),$$

where the normalizing constant, $t(\boldsymbol{\omega})$, depends on the weights $\omega_j, j = 1, 2, \ldots, k$. Therefore, the optimization problem becomes:

$$\max_{\boldsymbol{\omega}\in\chi} \left\{ \log t(\boldsymbol{\omega}) + \sum_{j=1}^{k} \omega_j \; E_m(E_{\pi^J}\left(\log \pi_j\left(\theta\right)\right)) \right\}. \tag{6}$$

The following result holds:

Proposition 3.1. *There is only one aggregated prior distribution, $\pi_{\boldsymbol{\omega}}(\theta)$, which minimizes the expected loss $L(\boldsymbol{\omega})$.*

The proof is presented in Appendix A. Note that, it is shown that there exists only one aggregated distribution, but not only one weight vector $\boldsymbol{\omega}$, which minimizes the expected loss $L(\boldsymbol{\omega})$.

In summary, a statistical decision problem has been considered in which the decision space is the set of all possible weight vectors i.e.:

$$\chi = \{(\omega_1, \omega_2, \ldots, \omega_k) : \sum_{j=1}^{k} \omega_j = 1, \omega_j \geq 0\}.$$

The states of nature, S, is the set of all possible values, x, that the considered random variable X can take on. Hence, the loss function is given by (2). Thus, each possible decision $\boldsymbol{\omega} \in \chi$ leads to a certain loss $L(\boldsymbol{\omega}, x) \in \mathbb{R}$, depending on the state of nature $x \in S$.

Next, a probability distribution for x (states of nature) is needed, and then it is obtained through the likelihood function. Because the decision maker assumes prior ignorance hence, the Jeffreys' prior distribution $\pi^J(\theta)$ is taken and the probability distribution for each x is given by the predictive prior distribution $m(x)$. The minimum expected loss criterion is used in order to obtain the optimal decision $\boldsymbol{\omega}$.

Finally, some remarkable properties for the aggregated prior distribution, $\pi_{\boldsymbol{\omega}}$, are shown. They have been derived through the previous decision problem. It is satisfied that the Kullback-Leibler divergence between the posterior distribution and the combined prior distribution is never larger than the average Kullback-Leibler divergences between the posterior distribution and each expert's prior distribution (see Bousquet (2008)), that is:

$$KL(\pi^J(\cdot|x)||\pi_{\boldsymbol{\omega}}(\cdot)) \leq \sum_{j=1}^{k} \omega_j KL(\pi^J(\cdot|x)||\pi_j(\cdot)),$$

where $\pi_j(\theta)$ is the proper prior distribution provided by each expert. Hence, the aggregated prior distribution stays in agreement with $\pi^J(\theta|x)$, although some prior distributions could present higher discrepancies. An analogous result is satisfied for the linear opinion pool by taking into account Jensen's inequality (see Heskes (1998)).

By taking expectations in both sides of the previous inequality, it is satisfied:

$$E_m\left(\pi^J(\cdot|x)||\pi_{\boldsymbol{\omega}}(\cdot)\right) \leq \sum_{j=1}^{k} \omega_j E_m\left(KL(\pi^J(\cdot|x)||\pi_j(\cdot))\right),$$

thus, by using the previously described decision-making framework a pooled prior distribution is obtained for which the expected calibration is the best one in the Kullback-Leibler sense. Namely, the expected loss for this prior distribution is lower than for any other.

Note that, in expression (6), it is not always possible to calculate the expectation with respect to the predictive prior distribution because the considered prior distributions $\pi^J(\theta)$ could be improper. In the next Section, a solution to this problem is presented.

4 A modified method for improper distributions.

A problem arises when the predictive prior distribution is not proper since the optimization problem could not be carried out in a suitable way. It is usual that the considered Jeffreys' prior distributions, $\pi^J(\theta)$, are not proper and, therefore, improper predictive prior distributions are usually obtained.

Lancaster (2004) argued that a possible way to solve this problem, apart from using initial proper prior distributions, is to set aside a subset of the data (*training sample*) and to use it to build a proper predictive distribution for the rest. This is connected with the intrinsic prior as well as the expected posterior prior methodology (see among others Berger and Pericchi (1996), Moreno et al. (1996), and Pérez and Berger (2002) and references therein). Both methodologies are closely related and they were motivated by the need for using improper prior distributions in a model selection context. The key idea in both methodologies is to convert improper noninformative prior distributions into proper distributions. A relevant tool in this context is the *training sample* (see Berger and Pericchi (2004) for a recent publication on this topic). Suppose that $\pi^J(\theta)$ denotes a noninformative, possibly improper prior distribution for θ, x represents *imaginary* observations and $m^*(x)$ is a suitable marginal distribution for x. The smallest data set x making the posterior distribution $\pi^J(\theta|x)$ proper is calling a *minimal training sample*. So, the expected posterior prior distribution for θ is then defined as:

$$\pi^*(\theta) = \int \pi^J(\theta|x)m^*(x)dx. \tag{7}$$

Thus, it is the mean with respect to the selected marginal of the proper posterior distribution $\pi^J(\theta|x)$. Keeping in mind that the aim is to calculate the expectation $E_m(E_{\pi^J}(\log \pi_j(\theta))$, i.e.:

$$E_m(E_{\pi^J}(\log \pi_j(\theta)) = \int (E_{\pi^J}(\log \pi_j(\theta))\, m(x)dx, \tag{8}$$

then, as in expression (7), the proposal is to find an appropriate marginal distribution to determine the expectation (8).

Two proposals are considered to choose m^* (see Pérez and Berger (2002) for other alternatives). The first one consists of taking the *base-model* noninformative predictive, i.e.:

$$m^*(x) = \int_\Theta f(x|\theta)\pi^J(\theta)d\theta.$$

Note that, it is just the predictive prior distribution for the considered model. Therefore, it can be used when the prior distribution $\pi^J(\theta)$ is initially proper. If it is not proper, then another alternative is to use the *empirical* distribution given in Pérez and Berger (2002). Observe that, for the considered model in this paper (i.e. θ is one-dimensional), the minimal training sample is typically a replication of the random variable X (see for instance Moreno et al. (2003)). Therefore, given the data $x_1, x_2, \ldots, x_n$, the empirical marginal distribution is defined as:

$$m(\boldsymbol{x}) = \frac{1}{n}\sum_i I_{\{x_i\}}(\boldsymbol{x}),$$

where I_A denotes the indicator function for the set A.

Notice that, when empirical marginal distribution is used, then the considered data set is involved in the decision process (6). Thus, in this instance, not only the initial information provided by each expert is involved in the making-decision procedure, but also the data are taking part through the empirical predictive distribution. As it will be observed in the next Section, this scheme will favor the expert whose distribution better matches the data.

5 Applications.

In this Section, two applications are considered in order to illustrate the Bayesian approaches presented in the previous sections. Firstly, a general study is carried out for a class of natural exponential families. Next, a practical application for a distribution of this family is performed, specifically, for the binomial distribution. Finally, an example for the Weibull distribution is considered.

5.1 Natural exponential families.

Let η be a σ-finite positive measure on the Borel set of $\mathbb{R}$ not concentrated at a single point. A random variable X is distributed according to a natural exponential family if its density with respect to η is:

$$f_\theta(x) = b(x)\exp\{x\theta - M(\theta)\},\ \theta \in \Theta, \tag{9}$$

for some function $b(\cdot)$, where $M(\theta) = \log\int b(x)\exp(x\theta)\ d\eta(x)$ and $\Theta = \{\theta \in \mathbb{R}$: $M(\theta) < \infty\}$ is nonempty. θ is called the natural parameter. In addition, if the parametric space Θ is an open set, the exponential family is said to be regular. It is satisfied $E(X|\theta) = M^{'}(\theta) = \mu$ and $\mathrm{Var}(X|\theta) = M^{''}(\theta)$. See Brown (1986) for a review on these families.

The mapping $\mu = \mu(\theta) = M^{'}(\theta)$ is differentiable, with inverse $\theta = \theta(\mu)$. It provides an alternative parameterization for $f_\theta(x)$ called mean parameterization. The function $V(\mu) = M^{''}(\theta) = M^{''}(\theta(\mu))$, $\mu \in \Omega$, is the variance function of (9) and Ω is the mean space. For natural exponential families with quadratic variance function (NEF-QVF), this function has the expression: $V(\mu) = v_0 + v_1\mu + v_2\mu^2$, where $\mu \in \Omega$ and v_0, v_1 and v_2 are real constants. These families are all regulars. See Morris (1982, 1983) for a review about the properties of these families. A more recent study for these families is given in Morris and Lock (2009).

Conjugate prior distributions as in Morris (1983) and Gutiérrez-Peña and Smith (1997) are considered for each expert, in order to obtain analytical results. Let $\mu_{0j} \in \Omega$ and $m_j > 0$, the conjugate prior distributions on θ are:

$$\pi_j(\theta) = K(m_j\mu_{0j}, m_j)\exp\{m_j\mu_{0j}\theta - m_jM(\theta)\}, j = 1, 2, \ldots, k, \tag{10}$$

where $K(m_j\mu_{0j}, m_j)$ are chosen to make $\int_\Theta \pi_j(\theta)\, d\theta = 1$ and μ_{0j} are the prior means. These prior distributions are called DY-conjugate in Consonni and Veronese (1992) and Gutiérrez-Peña and Smith (1997).

Therefore the pooled prior distribution (1) is given by:

$$\pi(\theta) = t(\boldsymbol{\omega})\prod_{j=1}^{k}(\pi_j(\theta))^{\omega_j} = t(\boldsymbol{\omega})\prod_{j=1}^{k}\left(K(m_j\mu_{0j}, m_j)\exp\{m_j\mu_{0j}\theta - m_jM(\theta)\}\right)^{\omega_j},$$

where the normalizing constant $t(\boldsymbol{\omega})$ satisfies:

$$t^{-1}(\boldsymbol{\omega}) = \prod_{j=1}^{k} K\left(m_j \mu_{0j}, m_j\right)^{\omega_j} \;\; K^{-1}(\sum_{j=1}^{k} \omega_j m_j \mu_{0j}, \sum_{j=1}^{k} \omega_j m_j),$$

with $K\left(\sum_{j=1}^{k} \omega_j m_j \mu_{0j}, \sum_{j=1}^{k} \omega_j m_j\right)$ the normalizing constant for the distribution:

$$\exp\left\{\sum_{j=1}^{k} \omega_j \left(m_j \mu_{0j} \theta - m_j M\left(\theta\right)\right)\right\}.$$

Therefore, the pooled prior distribution becomes:

$$\pi\left(\theta\right) = \frac{K(\sum\limits_{j=1}^{k} \omega_j m_j \mu_{0j}, \sum_{j=1}^{k} \omega_j m_j)}{\prod\limits_{j=1}^{k} K\left(m_j \mu_{0j}, \mu_{0j}\right)^{\omega_j}} \prod_{j=1}^{k} \left(\pi_j\left(\theta\right)\right)^{\omega_j}.$$

Sometimes, it is preferred or desirable to use noninformative prior distributions reflecting prior ignorance in some sense. One of the most widely used prior distributions is the Jeffreys' prior.

Gutiérrez-Peña and Smith (1995, 1997) show that Jeffreys' prior is a conjugate prior as those described in (10) for NEF with quadratic variance function. In addition, Jeffreys' prior distribution can be expressed as:

$$\pi^J\left(\theta\right) \propto \exp\left\{\frac{1}{2} C_1 \theta - \frac{C_2}{2} M(\theta)\right\},$$

where C_1 and C_2 are real constants. Hence, the posterior distribution of θ under the previous prior distribution is given by:

$$\pi^J\left(\theta|x\right) = K\left(C_1, C_2\right) \exp\left\{\left(x + \frac{C_1}{2}\right)\theta - \left(1 + \frac{C_2}{2}\right) M(\theta)\right\}. \tag{11}$$

Before dealing with the constrained maximization problem given in (6), the expectation with respect to the posterior distribution, $E_{\pi^J}\left(\log \pi\left(\theta\right)\right)$ is calculated.

By taking into account the previous expression for $\pi\left(\theta\right)$ and $\pi_j\left(\theta\right)$, it is easily obtained:

$$\log\left(\pi\left(\theta\right)\right) = \log K\left(\sum_{j=1}^{k} \omega_j m_j \mu_{0j}, \sum_{j=1}^{k} \omega_j m_j\right) + \sum_{j=1}^{k} \omega_j m_j \mu_{0j} \theta - \sum_{j=1}^{k} \omega_j m_j M(\theta).$$

Hence, if expectations with respect to the posterior distribution are taken in both sides of the previous equality, then the problem reduces to calculate the expectations $E_{\pi^J}(\theta)$ and $E_{\pi^J}(M(\theta))$.

From now on, the attention is focused on the posterior distribution (11). Next, the expectations $E_{\pi^J}(\theta)$ and $E_{\pi^J}(M(\theta))$ are analytically obtained. In order to do it, the following function is considered:

$$\log \int_{\Theta} \exp\left\{\left(1+\frac{C_1}{2}\right)\theta - \left(1+\frac{C_2}{2}\right)M(\theta)\right\} d\theta = -\log\left(K\left(C_1, C_2\right)\right).$$

By deriving with respect to C_1 and C_2, it is obtained:

$$\partial_{C_1}\left(-\log\left(K\left(C_1,C_2\right)\right)\right) = \frac{1}{2}E_{\pi^J}(\theta),\ \partial_{C_2}\left(-\log\left(K\left(C_1,C_2\right)\right)\right) = -\frac{1}{2}E_{\pi^J}(M(\theta)),$$

and, as a consequence:

$$E_{\pi^J}(\theta) = 2\partial_{C_1}\left(-\log\left(K\left(C_1,C_2\right)\right)\right),\ E_{\pi^J}(M(\theta)) = -2\partial_{C_2}\left(-\log\left(K\left(C_1,C_2\right)\right)\right). \quad (12)$$

Finally, the optimization problem (6) becomes:

$$\max_{\boldsymbol{\omega}\in\chi}\left\{\log K\left(\sum_{j=1}^{k}\omega_j m_j \mu_{0j}, \sum_{j=1}^{k}\omega_j m_j\right) + \sum_{j=1}^{k}\omega_j m_j \mu_{0j}\right.$$
$$\left.\times\ E_m\left(2\partial_{C_1}\left(-\log\left(K(C_1,C_2)\right)\right)\right) - \sum_{j=1}^{k}\omega_j m_j\ E_m\left(-2\partial_{C_2}\left(-\log(K(C_1,C_2))\right)\right)\right\},$$

Next, a numerical application for the binomial distribution is shown. This distribution is considered because it belongs to the previous class. In addition, the first approach is implemented since it has a proper marginal prior distribution. A brief study of the main distributions included in this class is shown in Appendix B.

5.2 Direct application for the binomial distribution.

Savchuk and Martz (1994) considered practical situations in which multiple sources of prior information are available to be used in a Bayesian reliability framework for binomial sampling. The objective is to estimate the survival probability of a certain unit for which there have been $x = 9$ successes in $r = 10$ tests. Four experts supply partial initial information as Beta prior distributions, and this individual information is now combined by using a logarithmic opinion pooling. If the canonical parameterization is used (see Appendix B), then the aggregated prior distribution becomes:

$$\begin{aligned}\pi(\theta) &= \Gamma(r\sum_{j=1}^{4}m_j\omega_j)\left(\Gamma(r\sum_{j=1}^{4}m_j\omega_j - \sum_{j=1}^{4}m_j\mu_{0_j}\omega_j)\right)^{-1} \\ &\quad\times\left(\Gamma(\sum_{j=1}^{4}m_j\mu_{0_j}\omega_j)\right)^{-1}\prod_{j=1}^{4}\left(\exp\left\{m_j\mu_{0_j}\theta - m_j r\log\left(1+e^{\theta}\right)\right\}\right)^{\omega_j},\end{aligned}$$

where $\Gamma(.)$ denotes the gamma function. The parameters chosen by Savchuk and Martz (1994) for the component prior distributions are considered here, these are: $m_1 = 1.9055$, $\mu_{01} = 9.4988$, $m_2 = 0.4300$, $\mu_{02} = 8$, $m_3 = 0.9244$, $\mu_{03} = 9.0004$, $m_4 = 0.2828$ and $\mu_{04} = 7.0014$. The weights are obtained following the proposal in this paper. Figure 1 shows the prior distributions by using the usual parameterization.

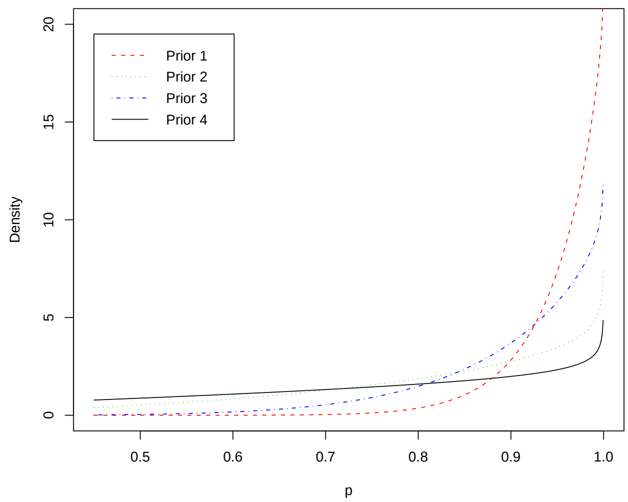

Figure 1: Prior distributions.

Notice that the second and fourth experts provide more diffuse prior distributions than the first and third experts, respectively. By considering the full development of the binomial distribution presented in Appendix B, the constrained optimization problem can be formulated in the following way:

$$\max_{\boldsymbol{\omega} \in \chi} \phi(\boldsymbol{\omega}), \tag{13}$$

where:

$$\begin{aligned} \phi(\boldsymbol{\omega}) \quad = \quad & \log\Gamma(r\sum_{j=1}^{4} m_j\omega_j) - \log\Gamma(\sum_{j=1}^{4} m_j\omega_j\mu_{0_j}) - \log\Gamma(\sum_{j=1}^{4} m_j\omega_j(r-\mu_{0_j})) \\ & + \left(E_m(\Psi(x+\frac{1}{2})) - E_m(\Psi(r-x+\frac{1}{2}))\right)\sum_{j=1}^{4} m_j\omega_j\mu_{0_j} \\ & + \left(E_m(\Psi(r-x+\frac{1}{2})) - r\Psi(r+1)\right)\sum_{j=1}^{4} m_j\omega_j. \end{aligned}$$

By taking into account the previous parameter values, the function to maximize is given by:

$$\begin{aligned}\phi\left(\boldsymbol{\omega}\right) \;=\;& \log\Gamma\left(19.055w_1+4.300w_2+9.244w_3+2.828w_4\right)-\log\Gamma(18.100w_1\\ &+3.440w_2+8.320w_3+1.980w_4)-\log\Gamma(0.955w_1+0.860w_2\\ &+0.924w_3+0.848w_4)-26.418w_1-5.961w_2-12.816w_3-3.921w_4.\end{aligned}$$

The following solution is found $(\omega_1,\omega_2,\omega_3,\omega_4)=(0,0,0,1)$. It has been obtained by using a pure random search algorithm, followed by a steepest descent method.

Thus, according to the proposal made in (6), an aggregated prior distribution that only considers the information provided by the fourth expert is enough to represent the group opinion of the four experts.

In order to explain this result, the Kullback-Leibler divergences between the posterior distribution $(\pi^J(\theta|x))$ and the prior distributions for each expert $(\pi_j(\theta))$ are examined for all possible number of successes in $r=10$ tests, i.e.: $x=0,1,2,\ldots,10$. Figure 2 shows these results.

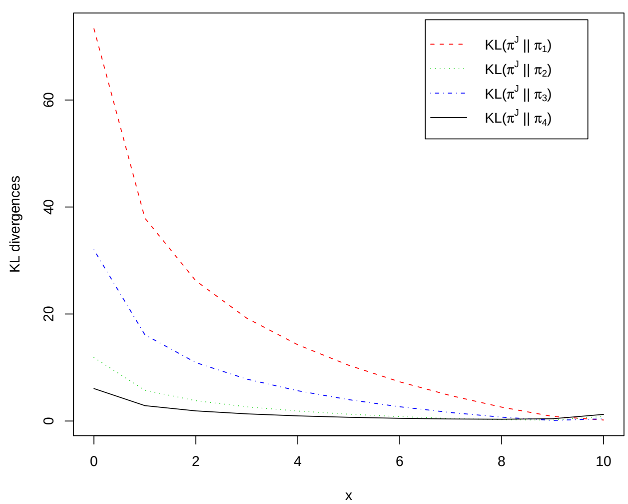

Figure 2: Values for the Kullback-Leibler divergences.

Note that, in this case, the lowest values for the Kullback-Leibler divergences are provided by the aggregated prior distribution when the data values are low (see Table 1). For high values in the data set, the lowest values for the Kullback-Leibler divergences are provided by the prior distributions $\pi_1(\theta)$ and $\pi_3(\theta)$, respectively. Therefore, the pooled prior distribution is overall the closest prior distribution, in the Kullback-Leibler sense, to the posterior distribution for any x (see Figure 2).

Consequently, based on the proposal in this paper, an expert who expresses a more diffuse prior distribution will be generally given a higher weight than one who expresses a less diffuse distribution.

Values	$x = 2$	$x = 3$	$x = 9$	$x = 10$
$KL(\pi^J \| \| \pi_1)$	26.2021	19.2111	0.8618	0.1886
$KL(\pi^J \| \| \pi_2)$	3.7971	2.6575	0.1733	0.8532
$KL(\pi^J \| \| \pi_3)$	10.9025	7.8194	0.1158	0.4101
$KL(\pi^J \| \| \pi_4)$	1.8940	1.3367	0.4277	1.2373

Table 1: Values for the Kullback-Leibler divergences.

Next, the empirical marginal distribution is considered in order to compare the previous result with those that are obtained when data sets are used. A more detailed study of this kind of distribution can be found in the next subsection.

Thus, random samples of size $n = 8$ from binomial distributions with parameters $r_1 = r_2 = r_3 = 10$, $p_1 = 0.9$, $p_2 = 0.8$ and $p_3 = 0.2$ respectively, are taken. Now, if the theoretical development made in Section 4 is taken into account, then the expectations involved in the problem are given by the expressions:

$$\begin{aligned} E_m\left(\Psi\left(x+\frac{1}{2}\right)\right) &= \frac{1}{n}\sum_{i=1}^{n}\Psi\left(x_i+\frac{1}{2}\right); \\ E_m\left(\Psi\left(r-x+\frac{1}{2}\right)\right) &= \frac{1}{n}\sum_{i=1}^{n}\Psi\left(r-x_i+\frac{1}{2}\right). \end{aligned}$$

The constrained optimization problem given in (13) is solved. Table 2 presents the obtained results. It can be observed how different they are to the previous result since the data sets are also involved in the decision-making process. Hence, the weighting scheme favors the experts whose distributions better match the data set.

Distributions	ω_1	ω_2	ω_3	ω_4
$B(10, 0.9)$	0.2440	0.0358	0.0012	0.7190
$B(10, 0.8)$	0.0008	0.0267	0.1647	0.8078
$B(10, 0.2)$	0.0010	0.0158	0.0068	0.9764

Table 2: Values for the weights.

Figure 3 shows the prior distributions and the values for the Kullback-Leibler divergences by taking the data set from the binomial distribution $B(10, 0.8)$. Notice that, the aggregated prior distribution gives the lowest values for the Kullback-Leibler divergences, on a whole. Thus, it is overall the closest to the posterior distribution for any x in the data set.

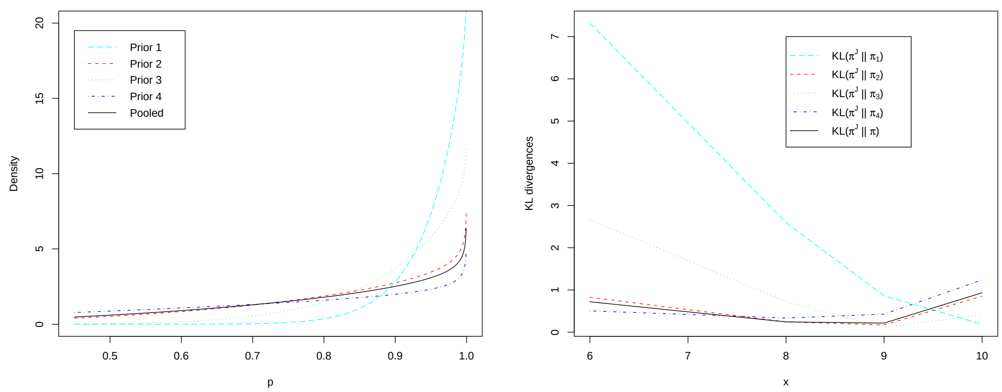

Figure 3: Prior distributions (left) and Kullback-Leibler divergences (right) when a data set from a binomial distribution $B(10, 0.8)$ is taken.

Therefore, in this case, working with observed data has given a less harsh result in calibration than the one obtained by using a proper predictive prior distribution. Nevertheless, in general, the obtained results will depend on both the prior distributions provided by the experts and the observed data.

5.3 Application for the Weibull distribution.

The approach presented in Section 4 is illustrated by considering a data set from a Weibull distribution, when the shape parameter is given. Thus, conjugate prior distributions can be selected for the scale parameter. Observe that, although Fink (1997) showed a prior distribution for Weibull shape and scale parameters, such that the posterior distribution has the same functional form as the prior, he argued that this prior distribution is not a true conjugate prior. Therefore, it is not considered here.

The Weibull distribution is often used in the reliability field (see Lawless (2003)).

It is assumed that data are coming from the density:

$$f\left(x|\lambda\right) = \alpha\lambda \exp\left(-\lambda x^{\alpha}\right) x^{\alpha-1} I_{(0,\infty)}\left(x\right),$$

where α, $\lambda > 0$ are the shape and scale parameters, respectively, and I is the indicator function. Note that, the same parameterization as in Kundu (2008) and Soland (1969) is considered here with the objective of simplifying the expressions in the subsequent analysis. In addition, this one-parameter distribution belongs to the subclass of the exponential family considered in Jozani et al. (2002). This subclass contains very useful distributions in reliability contexts and the prior distribution is usually a gamma distribution.

By assuming that k experts provide prior information over the scale parameter λ, as k conjugate prior distributions, the pooled prior distribution is given by:

$$\pi\left(\lambda\right)=t(\boldsymbol{\omega})\prod_{j=1}^{k}\left(\pi\left(\lambda_j\right)\right)^{\omega_j}, \tag{14}$$

where $\pi\left(\lambda_j\right)$ denotes a gamma density with parameters a_j and b_j, i.e:

$$\pi_j\left(\lambda\right)=\frac{b_j^{a_j}}{\Gamma\left(a_j\right)}\lambda^{a_j-1}\exp\left\{-b_j\lambda\right\}I_{(0,\infty)}\left(\lambda\right),$$

and the normalizing constant is given by:

$$t(\boldsymbol{\omega})=\frac{\left(\sum\limits_{j=1}^{k}b_j\omega_j\right)^{\sum\limits_{j=1}^{k}\omega_ja_j}}{\Gamma\left(\sum\limits_{j=1}^{k}\omega_ja_j\right)}\prod_{j=1}^{k}\left(\frac{\Gamma\left(a_j\right)}{b_j^{a_j}}\right)^{\omega_j}.$$

Those prior distributions are considered in order to obtain analytical results. The Jeffreys' prior distribution is:

$$\pi^J\left(\lambda\right)\propto\frac{1}{\lambda},$$

thus, the posterior distribution under the previous prior distribution is given by:

$$\pi^J\left(\lambda|x\right)=x^{\alpha}\exp\left(-\lambda x^{\alpha}\right), \tag{15}$$

which is a gamma distribution with shape parameter 1 and rate parameter x^{α}.

In order to obtain the function to maximize, firstly the expectation $E_{\pi^J}(\log\pi_j\left(\lambda\right))$ is calculated. By taking into account the following expression:

$$E_{\pi^J}(\log\pi_j\left(\lambda\right))=a_j\log b_j-\log\Gamma\left(a_j\right)+\left(a_j-1\right)E_{\pi^J}\left(\log\lambda\right)-b_jE_{\pi^J}\left(\lambda\right)$$

and that the next two equalities are satisfied:

$$E_{\pi^J}\left(\log\lambda\right)=\Psi\left(1\right)-\alpha\log x\ ;\ E_{\pi^J}\left(\lambda\right)=\frac{1}{x^{\alpha}},$$

then, the constrained optimization problem is formulated as:

$$\max_{\boldsymbol{\omega}\in\chi}\phi\left(\boldsymbol{\omega}\right),$$

where:

$$\begin{aligned}\phi\left(\boldsymbol{\omega}\right) &= \sum_{j=1}^{k}\omega_ja_j\log(\sum_{j=1}^{k}\omega_jb_j)-\log\Gamma(\sum_{j=1}^{k}\omega_ja_j)+\sum_{j=1}^{k}\omega_j\left(a_j-1\right)\\ &\quad\times\left[\Psi\left(1\right)-\alpha E_m\left(\log x\right)\right]-\sum_{j=1}^{k}\omega_jb_jE_m\left(x^{-\alpha}\right).\end{aligned}$$

Note that, the prior distribution $\pi^J(\lambda)$ is improper and the predictive prior distribution also. Therefore, the empirical marginal prior distribution has to be used. By considering the form of the posterior distribution, the minimal training samples are of size one and the empirical distribution is:

$$m(\boldsymbol{x}) = \frac{1}{n}\sum_i I_{\{x_i\}}(\boldsymbol{x}),$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$. Therefore, the expectations have the expressions:

$$E_m(\log x) = \frac{1}{n}\sum_{i=1}^{n}\log x_i \ ; \ E_m\left(x^{-\alpha}\right) = \frac{1}{n}\sum_{i=1}^{n} x_i^{-\alpha}.$$

In order to apply this proposal, the data in Linhart and Zucchini (1986) are considered. They represent failure times of the air conditioning system of an airplane. It is assumed that these data are coming from a Weibull distribution with shape parameter $\alpha = 0.8544$. It corresponds to the maximum likelihood estimate obtained in Gupta and Kundu (2001). Suppose that two experts supplied prior information over λ and that it is combined through expression (14). Thus, the function to maximize is:

$$\begin{aligned}\phi(\boldsymbol{\omega}) \quad = \quad & \sum_{j=1}^{2}\omega_j a_j \log(\sum_{j=1}^{2}\omega_j b_j) - \log\Gamma(\sum_{j=1}^{2}\omega_j a_j) - 3.449\sum_{j=1}^{2}\omega_j(a_j - 1) \\ & -0.113\sum_{j=1}^{2}\omega_j b_j,\end{aligned}$$

subject to $\boldsymbol{\omega} \in \chi$. In addition, suppose that both experts provide two Gamma prior distributions with parameters $a_1 = 0.1$, $a_2 = 0.5$ and $b_1 = b_2 = 2$. Then, the previous function is maximized by considering these parameter values. By using a pure random search algorithm, followed by a steepest descent method, the solution obtained is: $\omega_1 = 0.3133$ and $\omega_2 = 0.6867$.

Figure 4 shows the pooled and the prior distributions provided by each expert. In addition, Table 3 shows the estimated weights for some parameter sets.

As similarly done in the binomial application, the Kullback-Leibler divergences between the posterior distribution ($\pi^J(\theta|x)$) and the prior distributions for each expert ($\pi_j(\theta)$) are examined for the failure times of the air conditioning system of the airplane. Figure 5 shows these results.

Note that, in this case, the values of the Kullback-Leibler divergences for the aggregated prior distribution are always closer to the lowest value between those obtained by using the prior distribution of each expert. Also, for large values of x, this divergence value is lower than the minimum of the other two values. Thus, as in the previous application, the aggregated prior distribution gives the lowest values for the Kullback-Leibler divergences, on a whole.

Other prior distributions have been considered in order to know in which way the data set is affecting the prior distributions provided by the experts and consequently,

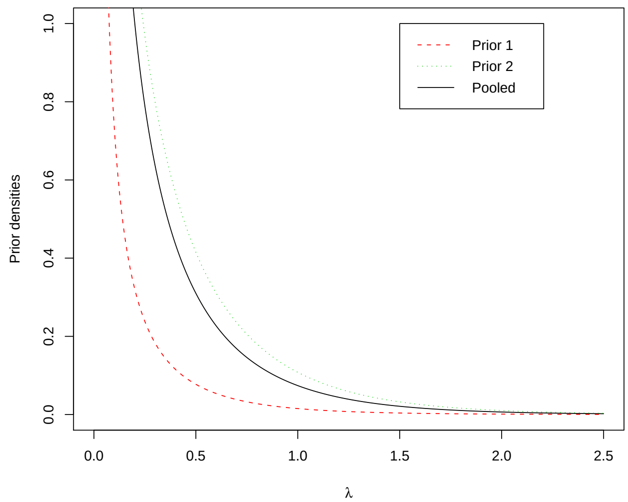

Figure 4: Prior distributions.

$a_2 = 0.5,\, b_2 = 2$			$a_1 = 0.05,\, b_1 = 0.15$		
	$a_1 = 0.09$	$a_1 = 0.11$		$a_2 = 0.45$	$a_2 = 0.55$
	$b_1 = 2$	$b_1 = 2$		$b_2 = 2$	$b_2 = 2$
ω_1	0.3056	0.3213	ω_1	0.2152	0.3896
ω_2	0.6944	0.6787	ω_2	0.7848	0.6104
	$a_1 = 0.05$	$a_1 = 0.05$		$a_2 = 0.5$	$a_2 = 0.5$
	$b_1 = 1.8$	$b_1 = 2.2$		$b_2 = 1.8$	$b_2 = 2.2$
ω_1	0.3105	0.3144	ω_1	0.3454	0.2806
ω_2	0.6895	0.6856	ω_2	0.6546	0.7194

Table 3: Estimated weights.

the aggregated prior distribution. The obtained results show that the influence on these prior distributions is given by the data set (considering each value one by one) and not by a summary statistic, such as the sample mean or any other statistics.

6 Discussion.

The aim of this paper is to develop a general decision making-based approach to assess the weights in an opinion pooling. Several experts' opinions about a parameter of interest are collected, and then, they are combined through a logarithmic opinion

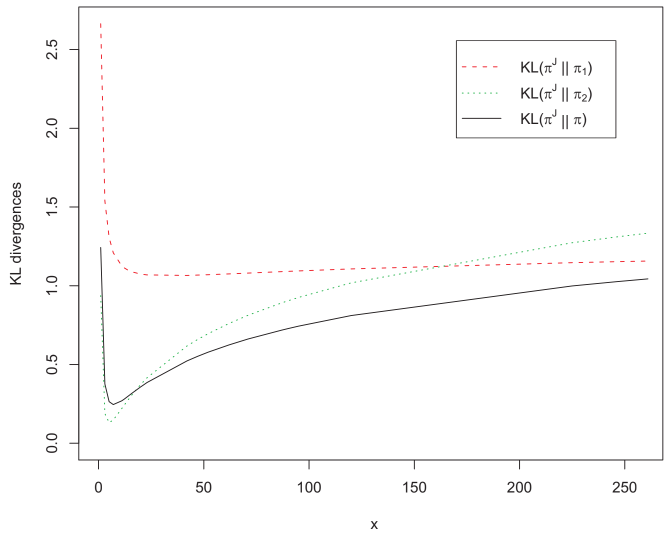

Figure 5: Values for the Kullback-Leibler divergences.

pool. Consequently, a prior distribution is considered, which mixes the experts' beliefs according to the features of this opinion pooling. Since the aggregated prior distribution depends on some parameters (i.e. weights), then they are calculated by taking into account both a calibration measure based on Kullback-Leibler divergence and the fact that there is no information from the decision maker. Thus, by considering the previous calibration measure, a probability distribution is proposed for which the pooled prior distribution has the least expected loss.

When no information from the decision maker about the states of nature is considered, then a general method is provided, which tends to result in the selection of the most diffuse pooled prior distribution. Observe that this prior distribution represents the information given by all experts. On the contrary, the belief of a diffuse expert might not be represented through a more informative combined prior distribution. Consequently, it would be very interesting to try to find a way of establishing some tests to analyze the local sensitivity of the method (see Pérez et al. (2006)).

A difficulty that arises from the assumption of prior ignorance for $x \in X$ (states of nature) by the decision maker is that the distribution, $m(x)$, could be improper. The proposal, in this case, is to use the empirical predictive distribution. Hence, the data set has to be included in the process. Perhaps, another way to obtain $m(x)$ is to assume prior ignorance on the weight vector. Therefore, a uniform Dirichlet prior distribution on these weights would be taken. However, the process would be non analytical.

Finally, an extension of this study could be performed, by considering that the decision maker has prior information about the states of nature independent of the experts. Some tests have been carried out, under the assumption that the information comes from historical data (or historical information). Observations have been made regarding matters related to the prior distributions elicited by the experts, the information provided by the decision maker through the predictive distribution and the obtained weights. Specifically, the obtained pooled distribution depends on the prior distributions given by both the experts and the decision maker. Thus, the proposed method does not always choose the less informative prior distribution. An issue open for further investigation can consider a sample, $\boldsymbol{x}$, of size n in order to obtain the posterior distribution $\pi^J(\theta|\boldsymbol{x})$. Under this assumption, some important questions to be analyzed are, among others, the choice of an appropriate sample size or the suitability of taking n tending to infinity. In addition, the computational difficulties that arise from the last assumption would have to be studied.

Appendix A. Proof of Proposition 1.

In order to deal with the existence of only one combined distribution, first, the convexity of the function to optimize is shown.

The optimization problem (4) is equivalent to the problem:

$$\min_{\boldsymbol{\omega}\in\chi}\left\{-\log t(\boldsymbol{\omega})-\sum_{j=1}^{k}\omega_j\ E_m(E_{\pi^J}\left(\log \pi_j\left(\theta\right)\right))\right\}.$$

Observe that the second term in the previous function is a linear combination over ω_j, thus the attention is focused on the first term:

$$\log t(\boldsymbol{\omega})=\log\left(\int_{\Theta}\prod_{j=1}^{k}\left(\pi_j\left(\theta\right)\right)^{\omega_j}d\theta\right).$$

Therefore, given $\boldsymbol{\omega}^1$, $\boldsymbol{\omega}^2\in\chi$ the following inequality must be proved:

$$\log t(\lambda\boldsymbol{\omega}^1+(1-\lambda)\boldsymbol{\omega}^2)\leq\lambda\log t(\boldsymbol{\omega}^1)+(1-\lambda)\log t(\boldsymbol{\omega}^2),\ \text{for all}\ \lambda\in[0,1]\,. \tag{16}$$

Since:

$$\begin{aligned}\log t(\lambda\boldsymbol{\omega}^1+(1-\lambda)\boldsymbol{\omega}^2) &= \log\left(\int_{\Theta}\prod_{j=1}^{k}\left(\pi_j\left(\theta\right)\right)^{\lambda\omega_j^1+(1-\lambda)\omega_j^2}d\theta\right) \qquad (17)\\ &= \log\left(\int_{\Theta}\left(\prod_{j=1}^{k}\left(\pi_j\left(\theta\right)\right)^{\omega_j^1}\right)^{\lambda}\left(\prod_{j=1}^{k}\left(\pi_j\left(\theta\right)\right)^{\omega_j^2}\right)^{(1-\lambda)}d\theta\right),\end{aligned}$$

the Hölder's inequality (see Cheung (2001)) is applied to the last integral in expression (17).

It is obtained the following inequality:

$$\log\left(\int_\Theta \prod_{j=1}^{k} \left(\pi_j\left(\theta\right)\right)^{\lambda\omega_j^1+(1-\lambda)\omega_j^2}\, d\theta\right)$$
$$\leq \lambda \log\left(\int_\Theta \prod_{j=1}^{k} \left(\pi_j\left(\theta\right)\right)^{\omega_j^1}\, d\theta\right) + (1-\lambda)\log\left(\int_\Theta \prod_{j=1}^{k} \left(\pi_j\left(\theta\right)\right)^{\omega_j^2}\, d\theta\right).$$

Therefore, the function to optimize: $\log t(\boldsymbol{\omega}) - \sum_{j=1}^{k} \omega_j \; E_m(E_{\pi^J}\left(\log \pi_j\left(\theta\right)\right)) = \log t(\boldsymbol{\omega}) - H(\boldsymbol{\omega})$ is convex because it is the sum of two convex functions and it has a minimum.

Nevertheless, as the previous function is a convex one on a compact set χ , if it has several minimum values then all of them take the same value. Let $\boldsymbol{\omega}^1$ and $\boldsymbol{\omega}^2$ be two weight vectors such that:

$$\log t(\boldsymbol{\omega}^1) - H(\boldsymbol{\omega}^1) = \log t(\boldsymbol{\omega}^2) - H(\boldsymbol{\omega}^2) = \min_{\boldsymbol{\omega}\in\chi}(\log t(\boldsymbol{\omega}) - H(\boldsymbol{\omega})).$$

Since $\lambda\boldsymbol{\omega}^1 + (1-\lambda)\boldsymbol{\omega}^2 \in \chi$ and χ is a compact set, then:

$$\log t(\lambda\boldsymbol{\omega}^1 + (1-\lambda)\boldsymbol{\omega}^2) - H(\lambda\boldsymbol{\omega}^1 + (1-\lambda)\boldsymbol{\omega}^2) = \min_{\boldsymbol{\omega}\in\chi}(\log t(\boldsymbol{\omega}) - H(\boldsymbol{\omega})). \tag{18}$$

It is satisfied:

$$\begin{aligned}
&\log t(\lambda\boldsymbol{\omega}^1 + (1-\lambda)\boldsymbol{\omega}^2) - H(\lambda\boldsymbol{\omega}^1 + (1-\lambda)\boldsymbol{\omega}^2)\\
&= \log t(\lambda\boldsymbol{\omega}^1 + (1-\lambda)\boldsymbol{\omega}^2) - \big(\lambda H(\boldsymbol{\omega}^1) - (1-\lambda)H(\boldsymbol{\omega}^2)\big)\\
&\leq \big(\lambda \log t(\boldsymbol{\omega}^1) + (1-\lambda)\log t(\boldsymbol{\omega}^2)\big) - \big(\lambda H(\boldsymbol{\omega}^1) - (1-\lambda)H(\boldsymbol{\omega}^2)\big)\\
&= \lambda\left(\log t(\boldsymbol{\omega}^1) - H(\boldsymbol{\omega}^1)\right) + (1-\lambda)\left(\log t(\boldsymbol{\omega}^2) - H(\boldsymbol{\omega}^2)\right)\\
&= \min_{\boldsymbol{\omega}\in\chi}(\log t(\boldsymbol{\omega}) - H(\boldsymbol{\omega})),
\end{aligned}$$

thus, by considering the expression (18), the equality in the previous expression is satisfied and, in consequence:

$$\log t(\lambda\boldsymbol{\omega}^1 + (1-\lambda)\boldsymbol{\omega}^2) = \lambda \log t(\boldsymbol{\omega}^1) + (1-\lambda)\log t(\boldsymbol{\omega}^2).$$

However, the previous equality holds, when the equality in Hölder's inequality holds. Following Ash (1972) (from Lemma 2.4.3 to Lemma 2.4.5), it occurs when the following equality is satisfied, up to a null measure set :

$$\frac{\prod\limits_{j=1}^{k}\left(\pi_j\left(\theta\right)\right)^{\omega_j^1}}{\int_\Theta \prod\limits_{j=1}^{k}\left(\pi_j\left(\theta\right)\right)^{\omega_j^1} d\theta} = \frac{\prod\limits_{j=1}^{k}\left(\pi_j\left(\theta\right)\right)^{\omega_j^2}}{\int_\Theta \prod\limits_{j=1}^{k}\left(\pi_j\left(\theta\right)\right)^{\omega_j^2} d\theta} \text{ for all } \theta\in\Theta.$$

Hence,

$$\log t(\boldsymbol{\omega}^1) = \log t(\boldsymbol{\omega}^2),$$

and it follows that there exists a unique aggregated prior which minimizes the expected loss.

Appendix B. Predictive prior distributions.

Poisson distribution. Consider the exponential family with densities of the form:

$$f\left(x|\lambda\right) = e^{-\lambda}\,\frac{\lambda^x}{x!}\ \ \left(\lambda > 0,\ x = 0, 1\ldots\right).$$

The canonical representation is given by (9), where $\theta = \log\left(\lambda\right), \Theta = \left(-\infty, \infty\right), M\left(\theta\right) = e^{\theta}$ and $b\left(x\right) = 1/x!$. The Jeffreys' prior on θ is:

$$\pi^J\left(\theta\right) \propto \exp\left\{\frac{\theta}{2}\right\},$$

thus, $C_1 = 1$ and $C_2 = 0$. In order to obtain $E_{\pi^J}(\theta)$ and $E_{\pi^J}(M(\theta))$, the partial derivatives $\partial_{C_1}\left(-\log(K(C_1, C_2))\right)$ and $\partial_{C_2}\left(-\log(K(C_1, C_2))\right)$ are calculated by considering the general case:

$$K(C_1, C_2) = \frac{\left(1 + \frac{C_2}{2}\right)^{x + \frac{C_1}{2}}}{\Gamma\left(x + \frac{C_1}{2}\right)}.$$

By substituting the values of C_1 and C_2, it is finally obtained:

$$E_{\pi^J}(\theta) = \Psi\left(x + \frac{1}{2}\right)\ ;\ E_{\pi^J}(M(\theta)) = x + \frac{1}{2}.$$

The Jeffreys' prior distribution is improper and the predictive prior too. Therefore, the empirical marginal prior has to be used. For this case, the minimal training samples are of size one and the empirical distribution is:

$$m\left(\boldsymbol{x}\right) = \frac{1}{n}\sum_i I_{\{x_i\}}\left(\boldsymbol{x}\right), \tag{19}$$

where $\boldsymbol{x} = \left(x_1, x_2, \ldots, x_n\right)$. Hence,

$$E_m\left(\Psi\left(x + \frac{1}{2}\right)\right) = \frac{1}{n}\sum_{i=1}^{n}\Psi\left(x_i + \frac{1}{2}\right)\ ;\ E_m\left(x_i + \frac{1}{2}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i + \frac{1}{2}\right).$$

Binomial distribution. Consider the exponential family with densities of the form:

$$f\left(x|p\right) = \binom{r}{x}\,p^x\left(1 - p\right)^{r - x}\ \ \left(r = 1, 2, \ldots, 0 < p < 1,\ x = 0, 1, \ldots, r\right).$$

The canonical representation is given by (9), where $\theta = \log\left(\frac{p}{1-p}\right), \Theta = (-\infty, \infty)$, $M(\theta) = r\log\left(1+e^{\theta}\right)$ and $b(x) = \binom{r}{x}$. The Jeffreys' prior on θ is:

$$\pi^J(\theta) \propto \exp\left\{\frac{\theta}{2} - \log\left(1+e^{\theta}\right)\right\},$$

thus, $C_1 = 1$ and $C_2 = \frac{2}{r}$. The partial derivatives $\partial_{C_1}(-\log(K(C_1, C_2)))$ and $\partial_{C_2}(-\log(K(C_1, C_2)))$ are calculated by considering the general case:

$$K(C_1, C_2) = \frac{\Gamma\left(r + \frac{r\,C_2}{2}\right)}{\Gamma\left(x + \frac{C_1}{2}\right)\Gamma\left(r + \frac{r\,C_2}{2} - x - \frac{C_1}{2}\right)}.$$

By substituting the values of C_1 and C_2, it is finally obtained:

$$E_{\pi^J}(\theta) = \Psi\left(x + \frac{1}{2}\right) - \Psi\left(r - x + \frac{1}{2}\right);$$

$$E_{\pi^J}(M(\theta)) = r\Psi(r+1) - r\Psi\left(r - x + \frac{1}{2}\right).$$

The predictive prior distribution is given by:

$$m(x) = \binom{r}{x}\frac{\Gamma(1)}{\left(\Gamma\left(\frac{1}{2}\right)\right)^2\,\Gamma(r+1)}\,\Gamma\left(x+\frac{1}{2}\right)\,\Gamma\left(r - x + \frac{1}{2}\right),$$

that is a Binomial-Beta $\sim \left(\frac{1}{2}, \frac{1}{2}, r\right)$ (see Bernardo and Smith (1994)). Therefore,

$$E_m\left(\Psi\left(x+\frac{1}{2}\right)\right) = \frac{\Gamma(1)}{\left(\Gamma\left(\frac{1}{2}\right)\right)^2)}\sum_{x=0}^{r}\frac{\Gamma\left(x+\frac{1}{2}\right)\Gamma\left(r-x+\frac{1}{2}\right)}{\Gamma(x+1)\Gamma(r-x+1)}\Psi\left(x+\frac{1}{2}\right).$$

The expectation $E_m\left(\Psi\left(r - x + \frac{1}{2}\right)\right)$ is obtained in a similar way as the previous one.

Negative-Binomial distribution. Consider the exponential family with densities of the form:

$$f(x|p) = \binom{x+r-1}{r-1}p^x(1-p)^r \quad (r = 1, 2, \ldots, 0 < p < 1,\ x = 0, 1, \ldots).$$

The canonical representation is given by (9), where $\theta = \log(p), \Theta = (-\infty, 0), M(\theta) = -r\log\left(1-e^{\theta}\right)$ and $b(x) = \binom{x+r-1}{r-1}$. The Jeffreys' prior on θ is:

$$\pi^J(\theta) \propto \exp\left\{\frac{\theta}{2} - \log\left(1-e^{\theta}\right)\right\},$$

thus $C_1 = 1$ and $C_2 = -\frac{2}{r}$. The partial derivatives $\partial_{C_1}(-\log(K(C_1, C_2)))$ and $\partial_{C_2}(-\log(K(C_1, C_2)))$ are calculated by considering the general case:

$$K(C_1, C_2) = \frac{\Gamma\left(x + \frac{C_1}{2} + r + \frac{C_2 r}{2} + 1\right)}{\Gamma\left(x + \frac{C_1}{2}\right)\Gamma\left(r + \frac{r\,C_2}{2} + 1\right)}.$$

By substituting the values of C_1 and C_2, it is finally obtained:

$$E_{\pi^J}(\theta) = \Psi\left(x+\frac{1}{2}\right) - \Psi\left(x+r+\frac{1}{2}\right);$$
$$E_{\pi^J}(M(\theta)) = -r\Psi(r) + r\Psi\left(x+r+\frac{1}{2}\right).$$

The expectations with respect to the empirical prior distributions (19) are:

$$E_m\left(\Psi\left(x+\frac{1}{2}\right)\right) = \frac{1}{n}\sum_{i=1}^{n}\Psi\left(x_i+\frac{1}{2}\right);$$
$$E_m\left(\Psi\left(x+r+\frac{1}{2}\right)\right) = \frac{1}{n}\sum_{i=1}^{n}\Psi\left(x_i+r+\frac{1}{2}\right).$$

Gamma distribution. Consider the exponential family with densities of the form:

$$f(x|\lambda) = \frac{\lambda^r}{\Gamma(r)}\ x^{r-1}\ e^{-\lambda x}\ \ (r,\ \lambda \text{ and } x>0).$$

The canonical representation is given by (9), where $\theta = -\lambda, \Theta = (-\infty, 0)\,, M(\theta) = -r$ $\log(-\theta)$ and $b(x) = \frac{x^{r-1}}{\Gamma(r)}$. The Jeffreys' prior on θ is:

$$\pi^J(\theta) \propto \exp\{-\log(-\theta)\},$$

thus, $C_1 = 0$ and $C_2 = -\frac{2}{r}$. The partial derivatives $\partial_{C_1}(-\log(K(C_1,\ C_2)))$ and $\partial_{C_2}(-\log(K(C_1,\ C_2)))$ are calculated by considering the general case:

$$K(C_1, C_2) = \frac{\Gamma\left(x+\frac{C_1}{2}\right)^{r+\frac{rC_2}{2}+1}}{\Gamma\left(r+\frac{rC_2}{2}+1\right)}.$$

By substituting the values of C_1 and C_2, it is finally obtained:

$$E_{\pi^J}(\theta) = -\frac{r}{x}\ ;\ E_{\pi^J}(M(\theta)) = r(\log(x) - \Psi(r)).$$

The expectations with respect to the empirical prior distributions (19) are:

$$E_m\left(-\frac{r}{x}\right) = -\frac{r}{n}\sum_{i=1}^{n}\frac{1}{x_i}\ ;\ E_m(r\log x - r\Psi(r)) = \frac{r}{n}\sum_{i=1}^{n}\log(x_i) - r\Psi(r).$$

Normal distribution. Consider the exponential family with densities of the form:

$$f(x|\lambda) = \frac{1}{\sqrt{2\pi\sigma^2}}\ \exp\left\{-\frac{1}{2\sigma^2}(x-\lambda)^2\right\}\ \left(-\infty < x,\ \lambda < \infty \text{ and } \sigma^2 > 0\right).$$

The canonical representation is given by (1), where $\theta = \lambda/\sigma^2, \Theta = (-\infty, \infty)\,, M(\theta) = \sigma^2\theta^2/2$ and $b(x) = \exp\left\{-(1/2\sigma^2)x^2\right\}/\sqrt{2\pi\sigma^2}$. The Jeffreys' prior on θ is:

$$\pi^J(\theta) \propto 1,$$

thus, $C_1 = C_2 = 0$. The partial derivatives $\partial_{C_1}(-\log(K(C_1, C_2)))$and $\partial_{C_1}(-\log(K(C_1, C_2)))$ are calculated by considering the general case:

$$K^B_{pos} = \sqrt{\frac{(C_2+2)\,\sigma^2}{4\pi}} \exp\left\{-\frac{(2x+C_1)^2}{4\sigma^2\,(C_2+2)}\right\}.$$

By substituting the values of C_1 and C_2, it is finally obtained:

$$E_{\pi^J}(\theta) = \frac{x}{\sigma^2}\ ;\ E_{\pi^J}(M(\theta)) = \frac{\sigma^2 + x^2}{2\sigma^2}.$$

The expectations with respect to the empirical prior distributions (19) are:

$$E_m\left(\frac{x}{\sigma^2}\right) = \frac{1}{n\sigma^2}\sum_{i=1}^{n} x_i\ ;\ E_m\left(\frac{\sigma^2 + x^2}{2\sigma^2}\right) = \frac{1}{2} + \frac{1}{2\sigma^2 n}\sum_{i=1}^{n} x_i^2.$$

Acknowledgments

The authors thank the Associate Editor and the three Referees for comments and suggestions which have substantially improved the readability and the content of this paper. This research has been supported by *Ministerio de Educación y Ciencia*, Spain (Project TIN2008-06796-C04-03) and *Junta de Extremadura* (Projects GRU10100 and PDT09A009).

References

Abbas, A. E. (2009). "A Kullback–Leibler View of Linear and Log–Linear Pools." *Decision Analysis*, 6: 25–37. 411, 412

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: 131–142. 415

Ash, R. B. (1972). *Real Analysis and Probability*. Academic Press, INC, London. 431

Bacharach, M. (1975). "Group decisions in the face of differences of opinion." *Management Science*, 22: 182–191. 412

Berger, J. O. and Pericchi, L. R. (1996). "The Intrinsic Bayes factor for model selection and prediction." *Journal of the American Statistical Association*, 91: 109–122. 417

— (2004). "Training samples in objective Bayesian model selection." *The Annals of Statistics*, 32(3): 841–869. 418

Bernardo, J. M. and Smith, A. (1994). *Bayesian Theory*. John Wiley & Sons. 433

Bousquet, N. (2006). "Subjective Bayesian statistics: agreement between prior and data." Technical Report RR–5900, Institut National de Recherche en Informatique et en Automatique. 415

— (2008). "Diagnostics of prior-data agreement in applied Bayesian analysis." *Journal of Applied Statistics*, 35(9): 1011–1029. 412, 413, 414, 415, 417

Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. Lecture Notes, 9. Hayward: Institute of Mathematical Statistics. 419

Chen, Y. and Pennock, D. M. (2005). "Information Markets vs Opinion Pools: An Empirical Comparison." *Proceedings of the 6th ACM Conference on Electronic Commerce*, 58–67. 414

Cheung, W.-S. (2001). "Generalizations of Hölder's inequality." *International Journal of Mathematics and Mathematical Sciences*, 26: 7–10. 430

Clemen, R. T. (2008). "A comment on Cooke's classical method." *Reliability Engineering and System Safety*, 93(5): 760–765. 412

Clemen, R. T. and Winkler, R. (2007). "Aggregating probability distributions." In W. Edwards, R. M. and von Winterfeldt, D. (eds.), *Advances in Decision Analysis: From Foundations to Applications*, 154–176. Cambridge, UK: Cambridge University Press. 411

Consonni, G. and Veronese, P. (1992). "Conjugate priors for exponential families having quadratic variance functions." *Journal of the American Statistical Association*, 87(420): 1123–1127. 419

Cooke, R. M. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford University Press, New York. 412

Diaconis, P. and Ylvisaker, D. (1985). "Quantifying prior opinion." In Bernardo, J. M., Berger, J. O., Dawid, A. P., Lindley, D., and Smith, A. F. M. (eds.), *Bayesian Statistics 2*, 133–156. Elsevier/North-Holland. 412

Faria, A. E. and Mubwandarikwa, E. (2008). "The geometric combination of Bayesian forecasting models." *Journal of Forecasting*, 27(6): 519–535. 413

Fink, D. (1997). "A compendium of conjugate priors." Technical report, Environmental Statistical group, Department of Biology, Montana State University, USA. 425

Garthwaite, P. H., Kadane, J. B., and O' Hagan, A. (2004). "Elicitation." Technical Report 04/01, Department of Statistics, The Open University (UK). 412, 413, 414

— (2005). "Eliciting probability distributions." *Journal of American Statistical Association*, 100(470): 680–700. 412, 414

Genest, C. (1984). "A Characterization theorem for externally Bayesian groups." *The Annals of Statistics*, 16(3): 1100–1105. 412

Genest, C. and Zidek, J. V. (1984). "Aggregating opinions through logarithmic pooling." *Theory and decision*, 17(1): 61–70. 412

— (1986). "Combining Probability distributions: A Critique and an Annotated Bibliography." *Statistical Science*, 1(1): 114–148. 411, 412, 414

Gupta, R. D. and Kundu, D. (2001). "Exponentiated Exponential Family: An Alternative to Gamma and Weibull distributions." *Biometrical Journal*, 43(1): 117–130. 427

Gutiérrez-Peña, E. and Smith, A. F. M. (1995). "Conjugate parameterisations for natural exponential families." *Journal of the American Statistical Association*, 90: 1347–1356. 420

— (1997). "Exponential and Bayesian conjugate families: Review and extensions (with discussion)." *Test*, 6: 1–90. 419, 420

Heskes, T. (1998). "Selecting weighting factors in logarithmic opinion pools." In *Proceedings of the 1997 conference on Advances in Neural Information Processing Systems 10*, 266–272. The MIT Press. 412, 417

Jacobs, R. A. (1995). "Methods for combining expert's probability assessments." *Neuronal Computation*, 7: 867–888. 412

Jozani, M. J., Nematollahi, N., and Shafie, K. (2002). "An admissible minimax estimator of a bounded scale-parameter in a subclass of the exponential family under scale-invariant squared-error loss." *Statistics and Probability Letters*, 60: 437–444. 425

Kadane, J. B., Schervish, M. J., and Seidenfeld, T. (1999). *Rethinking the foundations of Statistics*. Cambridge University Press. 412

Kascha, C. and Ravazzolo, F. (2010). "Combining inflation density forecasts." *Journal of Forecasting*, 29: 231–250. 412, 414

Kundu, D. (2008). "Bayesian Inference and Life Testing Plan for the Weibull distribution in presence of Progressive Censoring." *Technometrics*, 50(2): 144–154. 425

Lancaster, T. (2004). *An Introduction to modern Bayesian Econometrics*. Blackwell Publishing. 417

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data, Second Edition*. Wiley Series in Probability and Statistics. 425

Linhart, H. and Zucchini, W. (1986). *Model Selection*. Wiley, New York. 427

Liseo, B. (1993). "Elimination of nuisance parameters with reference priors." *Biometrika*, 80: 295–304. 415

McConway, K. J. (1981). "Marginalization and linear opinion pools." *Journal of the American Statistical Association*, 76: 410–414. 412

Moreno, E., Bertolino, F., and Racugno, W. (1996). "The intrinsic priors in model selection and hypotesis testing." Technical report, University of Granada. 417

— (2003). "Bayesian Inference under partial prior information." *Scandinavian Journal of Statistics*, 30: 565–580. 418

Morris, C. N. (1982). "Natural exponential families with quadratic variance functions." *The Annals of Statistics*, 10: 65–80. 419

— (1983). "Natural exponential families with quadratic variance functions: Statistical theory." *Annals of Statistics*, 11: 515–529. 419

Morris, C. N. and Lock, K. F. (2009). "Unifying the named natural exponential families and their relatives." *The American Statistician*, 63(3): 247–253. 419

O' Hagan, A., Buck, C., Daneshkhah, A., Eiser, J., Garthwaite, P., Jenkinson, D., Oakley, J., and Rakow, T. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*. Wiley. 412

Ouchi, F. (2004). "A literature review on the use of expert opinion in probabilistic risk analysis." Technical Report 3201, World Bank, Washington, D.C. 411

Pérez, C. J., Martín, J., and Rufo, M. J. (2006). "MCMC-based local parametric sensitivity estimations." *Computational Statistics and Data Analysis*, 51: 823–835. 429

Pérez, J. M. and Berger, J. O. (2002). "Expected–posterior prior distributions for model selection." *Biometrika*, 83(3): 491–511. 417, 418

Poole, D. and Raftery, A. (2000). "Inference for Deterministic Simulation Models: The Bayesian Melding Approach." *Jounal of the American Statistical Association*, 95(452): 1244–1255. 412

Roback, P. J. and Givens, G. H. (2001). "Supra-Bayesian Pooling of priors linked by a deterministic simulation model." *Communications in Statistics: Simulation and Computation*, 30(3): 447–476. 412

Rufo, M. J., Martín, J., and Pérez, C. J. (2009). "Inference on exponential families with mixture of prior distributions." *Computational Statistics and Data Analysis*, 53(9): 3271–3280. 414

Savchuk, V. P. and Martz, H. F. (1994). "Bayes Reliability Estimation Using Multiple Sources of Prior Information: Binomial Sampling." *IEEE Transactions on Reliability*, 43(1): 138–144. 421, 422

Soland, R. (1969). "Bayesian Analysis of the Weibull Process with unknown Scale and Shape parameters." *IEEE Transactions on Reliability Analysis*, 18: 181–184. 425

Stone, M. (1961). "The opinion pool." *Annals of Mathematical Statistics*, 32: 1339–1342. 412

Tversky, A. and Kahneman, D. (1974). "Judgment under uncertainty: Heuristics and Biases." *Science*, 185: 1124–1131. 412