Electronic Journal of Statistics
Vol. 6 (2012) 2362–2382
ISSN: 1935-7524
DOI: 10.1214/12-EJS752

A Bayesian approach to aggregate experts' initial information

María Jesús Rufo[*], **Carlos J. Pérez**

Department of Mathematics, University of Extremadura
Cáceres, Spain
e-mail: mrufo@unex.es; carper@unex.es

and

Jacinto Martín

Department of Mathematics, University of Extremadura
Badajoz, Spain
e-mail: jrmartin@unex.es

Abstract: This paper provides a Bayesian procedure to aggregate experts' information in a group decision making context. The belief of each expert is elicited as a multivariate prior distribution. Then, linear and logarithmic combination methods are used to represent a consensus distribution. Anyway, the choice of the appropriate strategy will depend on the decision maker's judgements. A significant task when using opinion pooling is to find the optimal weights. In order to carry it out, a criterion based on Kullback-Leibler divergence is proposed. Furthermore, based on the previous idea, an alternative procedure is presented when a solution cannot be found. The theoretical foundations are discussed in detail for each aggregation scheme. In particular, it is shown that a general unified method is achieved when they are applied to multivariate natural exponential families. Finally, two illustrative examples show that the proposed techniques can be easily applied in practice and their usefulness for decision making under the described situations.

AMS 2000 subject classifications: Primary 62C10; secondary 62H99.
Keywords and phrases: Bayesian analysis, group decision, Kullback-Leibler divergence, multivariate exponential families, opinion pooling.

Received March 2012.

1. Introduction

An important issue involved in decision problems is the aggregation of experts' opinions. See, for instance, [17, 25], or [11]. Decision makers often consult experts before making their choices about a topic. In this sense, it could be necessary or desirable to obtain a joint probability distribution of the quantities of interest, which includes the beliefs of the experts (see, e.g., [24]).

Belief aggregation of subjective distributions has been of great interest in several fields of knowledge. [11] classified the methods to get the aggregated opinion

[*]Corresponding author.

of a group of experts in two types: mathematical and behavioral methods. In the first case, each expert provides his/her information as a probability distribution over the quantities of interest and independently of the others. Then, the decision maker needs to combine these distributions to obtain a single distribution (see, e.g, [1]). A widely used mathematical technique to get this joint distribution is opinion pooling. The main idea is to find an aggregated distribution that satisfies a set of reasonable axioms (see, e.g., [17, 10] and [24]). The two most common and used combined distributions are the linear and logarithmic opinion pools (see, e.g., [28] and [3]). For the first of them, the probability distributions of the experts are averaged whereas, for logarithmic opinion pool, these distributions are multiplied and renormalized. Anyway, the decision maker should consistently decide on the way to combine the distributions elicited by the experts.

A point of controversy when these combined distributions are used, is how to choose the experts' weights. Namely, the decision maker needs to reflect his/her beliefs about the expertise of the experts through the weights. Some related works can be referred to [17, 12, 9, 26] and references therein, and [21], among others.

[27] presented a general Bayesian procedure to estimate a set of appropriate weights in a mixture of prior distributions. The proposed procedure was theoretically formulated for the class of the natural exponential families with quadratic variance.

The aim of this paper is to extend the previous procedure to multivariate settings by considering both linear and also including logarithmic opinion pools. Although two ways to combine experts' probability distributions are discussed, the paper does not pretend to compare the merits of linear and logarithmic opinion pools when the proposed approach is applied. Thus, the focus of this paper is the choice of the weights when using either of the two aggregation methods. The fact of increasing the dimensionality of the model could lead to a computationally complex procedure which, at the same time, involves the number of experts.

Firstly, a theoretical development is made for each aggregation technique. It is shown how neither the dimensionality nor the number of experts represent difficulties in the final process. Next, a general unified analysis of distributions that belong to the natural exponential families is made. These families are considered because they include distributions frequently used in practice. In addition, a straightforward method that allows to obtain the multivariate expectations involved in the process is presented. Therefore, the entire procedure for those distributions is analytical. It is important to emphasize that the reader could apply the technique for a particular distribution. Generally, numerical methods must be used.

The outline of the paper is as follows. In Section 2, we introduce the notation for multivariate natural exponential families. Section 3 shows the general procedure by differentiating between the two schemes to get the joint prior distribution. In addition, we present a unified framework for the previously considered multivariate distributions in both cases. In Section 4, we apply the developed methodology to the multinomial distribution and to the bivariate normal distribution, as examples of a discrete case and a continuous case, respectively. The conclusion is presented in Section 5.

2. Multivariate natural exponential families

Let η be a σ-finite positive measure on the Borel sets of $\mathbb{R}^d$. Suppose η is not concentrated on an affine subspace of $\mathbb{R}^d$. A random vector $\boldsymbol{X}$ is distributed according to a natural exponential family if its density with respect to η satisfies:

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = \exp\left\{\boldsymbol{x}^T\boldsymbol{\theta} - M\left(\boldsymbol{\theta}\right)\right\},\ \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathcal{N}, \tag{2.1}$$

where $M\left(\boldsymbol{\theta}\right) = \log \int \exp\{\boldsymbol{x}^T\boldsymbol{\theta}\} d\eta\left(\boldsymbol{x}\right)$, $\mathcal{N} = \{\boldsymbol{\theta} \in \mathbb{R}^d : M\left(\boldsymbol{\theta}\right) < \infty\}$ and $\boldsymbol{\Theta} =$ Interior$(\mathcal{N})$ is nonempty. $\mathcal{N}$ is called the natural parameter space and $M\left(\boldsymbol{\theta}\right)$ the cumulant generating function. See [4] and [7] for a description of these families.

If $\boldsymbol{X}$ is a random vector distributed according to (2.1), then:

$$E(\boldsymbol{X}|\boldsymbol{\theta}) = \nabla M\left(\boldsymbol{\theta}\right) = \boldsymbol{\mu} \text{ and } \mathrm{Var}(\boldsymbol{X}|\boldsymbol{\theta}) = H_M(\boldsymbol{\theta}),$$

where $\nabla M(\boldsymbol{\theta}) = \frac{\partial M(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ and $H_M\left(\boldsymbol{\theta}\right) = \frac{\partial^2 M(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}}$ denote the gradient vector and the Hessian matrix of $M(\boldsymbol{\theta})$, respectively.

Conjugate prior distributions as in [13] and [19] are considered. Let $\boldsymbol{\mu}_0 = \left(\mu_{0_1}, \mu_{0_2}, \ldots, \mu_{0_d}\right)^T \in \boldsymbol{\Omega}$ (the mean space) and $m > 0$, the conjugate prior distributions for $\boldsymbol{\theta}$ are:

$$\pi\left(\boldsymbol{\theta}\right) = K(m, \boldsymbol{\mu}_0^T) \exp\left\{m\boldsymbol{\mu}_0^T\boldsymbol{\theta} - mM\left(\boldsymbol{\theta}\right)\right\}, \tag{2.2}$$

where $K(m, \boldsymbol{\mu}_0^T)$ is chosen to make $\int_{\boldsymbol{\Theta}} \pi\left(\boldsymbol{\theta}\right) d\boldsymbol{\theta} = 1$.

3. The method

Let $\boldsymbol{X}$ be a random vector distributed according to a density $f(\boldsymbol{x}|\boldsymbol{\theta})$, and suppose that k experts provide prior information about the quantities of interest $\boldsymbol{\theta}$. Then, the opinion of each expert is elicited as a proper prior distribution $\pi_j(\boldsymbol{\theta})$. In this context, there are different methods to achieve the aggregated opinion of a group of experts (see, e.g., [14, 15]). One of them consists of using a linear opinion pool:

$$\pi_m\left(\boldsymbol{\theta}\right) = \sum_{j=1}^{k} \omega_j \pi_j\left(\boldsymbol{\theta}\right), \tag{3.1}$$

where $\omega_j,\ j = 1, 2, \ldots, k$ are the mixture weights, which are nonnegative and sum up to one (see, e.g., [16]).

An alternative approach to the linear opinion pool is the logarithmic opinion pool, which has the expression:

$$\pi_l\left(\boldsymbol{\theta}\right) = t \prod_{j=1}^{k} \left(\pi_j\left(\boldsymbol{\theta}\right)\right)^{\omega_j}, \tag{3.2}$$

where t is the normalizing constant, i.e.:

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

and the weights ω_j satisfy the previously specified conditions.

The linear and logarithmic opinion pools lead to quite different aggregated prior distributions (see, e.g.,[24]). In general, a logarithmic opinion pool results in a combined prior distribution which is frequently unimodal and less dispersed than the one obtained through a linear combination. In consequence, it is more likely to indicate consensual values when decisions must be made (see [17]). Note that, in a linear opinion pool the decision maker takes into account the full range of parameter values, that is, the range of values of $\boldsymbol{\theta}$ supported by each expert, whereas in a logarithm opinion pool he/she focuses on the common range of parameter values. Illustrative examples that show the main differences between the two aggregation schemes are presented in [24] and [21]. 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 assigns zero probability to that set.

In any case, the preference for a particular combination will depend on the the decision maker's judgements about the parameter of interest, which have to support the previous information provided by the experts.

In expressions (3.1) and (3.2), ω_j, $j = 1, 2, \ldots, k$ are the specified weights according to the experts' beliefs and they should be chosen to indicate the relative reliability of each expert. One of the most common practical strategies consists of placing an equal prior weight over each component (see, [23] where several examples are discussed). Then, the information provided for each expert is equally reliable. Thus, all experts participate equally in the initial aggregated distribution.

Here, the weights are not fixed from the beginning and, a general procedure to obtain them is proposed. The objective is to choose the weights such that no expert has more prior influence than the others on the combined prior distributions $\pi_m\left(\boldsymbol{\theta}\right)$ and $\pi_l\left(\boldsymbol{\theta}\right)$, respectively. In order to do it, the Kullback-Leibler divergence is used as a general measure to describe the discrepancy from each component distribution to the combined distribution. Observe that other divergence measures can be used (see, for instance, [2] for other information-theoretic measures). Nevertheless, it will be observed throughout the paper that the use of Kullback-Leibler divergence provides analytical advantages as well as computational simplicity. See [6] and [5] for the use of this divergence measure.

In the next two Subsections, the procedure is developed when linear and logarithmic opinion pools are selected. Subsequently, the previous process is applied to natural exponential families.

3.1. Linear opinion pool

The Kullback-Leibler divergence between the combined prior distribution π_m and the component prior distribution π_j is defined as:

$$KL(\pi_m||\pi_j) = \int_{\mathbf{\Theta}} \pi_m(\boldsymbol{\theta}) \log\left(\frac{\pi_m(\boldsymbol{\theta})}{\pi_j(\boldsymbol{\theta})}\right) d\boldsymbol{\theta} = E_{\pi_m}(\log \pi_m(\boldsymbol{\theta})) - E_{\pi_m}(\log \pi_j(\boldsymbol{\theta})).$$

The objective is to find $\omega_1, \omega_2, \ldots, \omega_k$, such that:

$$KL(\pi_m||\pi_1) = KL(\pi_m||\pi_2) = \cdots = KL(\pi_m||\pi_k), \tag{3.3}$$

with the constrains $\sum_{j=1}^{k} \omega_j = 1$ and $\omega_j \geq 0$.

From the expression for the linear opinion pool and the definition of the Kullback-Leibler divergence, it is obtained:

$$E_{\pi_m}\left(\log \pi_s\left(\boldsymbol{\theta}\right)\right) = \sum_{j=1}^{k} \omega_j E_{\pi_j}\left(\log \pi_s\left(\boldsymbol{\theta}\right)\right),$$

and the solution can be obtained from the linear equation system:

$$\begin{aligned} &\sum_{j=1}^{k} \omega_j \left(E_{\pi_j}\left(\log \pi_h\left(\boldsymbol{\theta}\right)\right) - E_{\pi_j}\left(\log \pi_1\left(\boldsymbol{\theta}\right)\right)\right) = 0,\ \ h = 2, 3, \ldots, k, \\ &\sum_{j=1}^{k} \omega_j = 1, \end{aligned} \tag{3.4}$$

being $\omega_j \geq 0$, $j = 1, 2, \ldots, k$. Observe that the equalities given in (3.3) are equivalent to the previous linear equation system.

Note that the main difficulty in the previous procedure is to obtain the expectations involve in the process. In particular, a unified analytical process can be performed when prior distributions as those in Section 2 are considered.

In order to do it, firstly, the expression for the conjugate prior distribution (2.2) is considered. Then, the expectations are given by:

$$E_{\pi_j}\left(\log \pi_s(\boldsymbol{\theta})\right) = \log K(m_s, \boldsymbol{\mu}_{0s}^T) + m_s \boldsymbol{\mu}_{0s}^T E_{\pi_j}(\boldsymbol{\theta}) - m_s E_{\pi_j}(M\left(\boldsymbol{\theta}\right)),$$

and the linear equation system (3.4) has the following expression:

$$\begin{aligned} &\sum_{j=1}^{k} \omega_j \left[(m_h \boldsymbol{\mu}_{0h}^T - m_1 \boldsymbol{\mu}_{01}^T) E_{\pi_j}(\boldsymbol{\theta}) + (m_1 - m_h)\ E_{\pi_j}(M(\boldsymbol{\theta}))\right] \\ &\quad = \log K\left(m_1, \boldsymbol{\mu}_{01}^T\right) - \log K\left(m_h, \boldsymbol{\mu}_{0h}^T\right), \text{ for } h = 2, 3, \ldots, k, \\ &\qquad\qquad \sum_{j=1}^{k} \omega_j = 1, \end{aligned} \tag{3.5}$$

with $\omega_j \geq 0$, $j = 1, 2, \ldots, k$.

Next, the expectations $E_{\pi_j}(\boldsymbol{\theta})$ and $E_{\pi_j}(M(\boldsymbol{\theta}))$ are analytically obtained for this class of prior distributions. Then, the following equality is considered:

$$\log \int_{\boldsymbol{\Theta}} \exp\left\{m_j \boldsymbol{\mu}_{0j}^T \boldsymbol{\theta} - m_j M(\boldsymbol{\theta})\right\} d\boldsymbol{\theta} = -\log K(m_j, \boldsymbol{\mu}_{0j}^T),$$

and the partial derivatives $\partial_{\boldsymbol{\mu}_{0j}}(-\log K(m_j, \boldsymbol{\mu}_{0j}^T))$ and $\partial_{m_j}(-\log K(m_j, \boldsymbol{\mu}_{0j}^T))$ are calculated.

Observe that

$$\partial_{\boldsymbol{\mu}_{0j}}(-\log K(m_j, \boldsymbol{\mu}_{0j}^T)) = \Big(\partial_{\mu_{0_1 j}}(-\log K(m_j, \boldsymbol{\mu}_{0j}^T)), \ldots, \partial_{\mu_{0_d} j}(-\log K(m_j, \boldsymbol{\mu}_{0j}^T))\Big)^T.$$

Then, the partial derivatives with respect to $\mu_{0_t j}$, $t = 1, 2, \ldots, d$, are given by:

$$\partial_{\mu_{0_t j}}(-\log K(m_j, \boldsymbol{\mu}_{0j}^T)) = \frac{\int_{\boldsymbol{\Theta}} \exp\left\{m_j \boldsymbol{\mu}_{0j}^T \boldsymbol{\theta} - m_j M(\boldsymbol{\theta})\right\} m_j \theta_t d\boldsymbol{\theta}}{\int_{\boldsymbol{\Theta}} \exp\left\{m_j \boldsymbol{\mu}_{0j}^T \boldsymbol{\theta} - m_j M(\boldsymbol{\theta})\right\} d\boldsymbol{\theta}} = m_j E_{\pi_j}(\theta_t).$$

Thus,

$$\partial_{\boldsymbol{\mu}_{0j}}(-\log K(m_j, \boldsymbol{\mu}_{0j}^T)) = \left(m_j E_{\pi_j}(\theta_1), m_j E_{\pi_j}(\theta_2), \ldots, m_j E_{\pi_j}(\theta_d)\right)^T.$$

The partial derivatives with respect to m_j, $j = 1, 2, \ldots, k$, have the expression:

$$\begin{aligned} \partial_{m_j}(-\log K(m_j, \boldsymbol{\mu}_{0j}^T)) &= \frac{\int_{\boldsymbol{\Theta}} \exp\left\{m_j \boldsymbol{\mu}_{0j}^T \boldsymbol{\theta} - m_j M(\boldsymbol{\theta})\right\} \left(\boldsymbol{\mu}_{0j}^T \boldsymbol{\theta} - M(\boldsymbol{\theta})\right) d\boldsymbol{\theta}}{\int_{\boldsymbol{\Theta}} \exp\left\{m_j \boldsymbol{\mu}_{0j}^T \boldsymbol{\theta} - m_j M(\boldsymbol{\theta})\right\} d\boldsymbol{\theta}} \\ &= \boldsymbol{\mu}_{0j}^T E_{\pi_j}(\boldsymbol{\theta}) - E_{\pi_j}(M(\boldsymbol{\theta})). \end{aligned}$$

Finally, the expectations are given by the following expressions:

$$\begin{aligned} E_{\pi_j}(\boldsymbol{\theta}) &= \frac{1}{m_j}\Big(\partial_{\mu_{0_1 j}}(-\log K(m_j, \boldsymbol{\mu}_{0j}^T)),\ \ldots, \partial_{\mu_{0_d} j}(-\log K(m_j, \boldsymbol{\mu}_{0j}^T)\Big)^T, \\ E_{\pi_j}(M(\boldsymbol{\theta})) &= \boldsymbol{\mu}_{0j}^T E_{\pi_j}(\boldsymbol{\theta}) - \partial_{m_j}(-\log K(m_j, \boldsymbol{\mu}_{0j}^T)). \end{aligned} \tag{3.6}$$

A difficulty arising from the previous procedure is that, it is not always possible to find a solution satisfying the constrains $\omega_j \geq 0$, $j = 1, 2, \ldots, k$, i.e., on the simplex $S_k = \{(\omega_1, \omega_2, \ldots, \omega_k) : \sum_{j=1}^k \omega_j = 1, \omega_j \geq 0, \ j = 1, 2, \ldots, k\}$. The multinomial application presented in Subsection 4.1 shows this case in a clear way. In order to solve this problem, the following alternative process can be considered. The procedure consists of finding the weight vector $(\omega_1, \omega_2, \ldots, \omega_k)$ that minimizes u, satisfying:

$$|KL(\pi_m||\pi_s) - KL(\pi_m||\pi_h)| \leq u \ \text{ for } 1 \leq s < h \leq k, \tag{3.7}$$

where u is a nonnegative real number. Finally, this problem is equivalent to the linear optimization problem:

minimize u, such that:

$$\begin{gathered}\sum_{j=1}^{k}\omega_j\left[E_{\pi_j}\left(\log\left(\pi_h\left(\boldsymbol{\theta}\right)\right)\right)-E_{\pi_j}\left(\log\left(\pi_s\left(\boldsymbol{\theta}\right)\right)\right)\right]-u\leq 0,\\ -\sum_{j=1}^{k}\omega_j\left[E_{\pi_j}\left(\log\left(\pi_h\left(\boldsymbol{\theta}\right)\right)\right)-E_{\pi_j}\left(\log\left(\pi_s\left(\boldsymbol{\theta}\right)\right)\right)\right]-u\leq 0,\\ \sum_{j=1}^{k}\omega_j=1,\omega_j\geq 0,\end{gathered}\tag{3.8}$$

for $1\leq s<h\leq k$.

Note that, by using this procedure, the maximum difference among experts is the smallest possible in the Kullback-Leibler sense. When $u=0$, the procedures (3.4) and (3.8) are equivalent. Thus, the obtained solution is the same. In addition, for the conjugate prior distributions (2.2), the previous linear optimization problem has the expression:

minimize u such that:

$$\begin{gathered}\sum_{j=1}^{k}\omega_j\left[\left(m_h\boldsymbol{\mu}_{0h}^T-m_s\boldsymbol{\mu}_{0s}^T\right)E_{\pi_j}\left(\boldsymbol{\theta}\right)+\left(m_s-m_h\right)E_{\pi_j}M\left(\left(\boldsymbol{\theta}\right)\right)\right]-u\\ \leq\log K\left(m_s,\boldsymbol{\mu}_{0s}^T\right)-\log K\left(m_h,\boldsymbol{\mu}_{0h}^T\right),\\ -\sum_{j=1}^{k}\omega_j\left[\left(m_h\boldsymbol{\mu}_{0h}^T-m_s\boldsymbol{\mu}_{0s}^T\right)E_{\pi_j}\left(\boldsymbol{\theta}\right)+\left(m_s-m_h\right)E_{\pi_j}M\left(\left(\boldsymbol{\theta}\right)\right)\right]-u\\ \leq\log K\left(m_h,\boldsymbol{\mu}_{0h}^T\right)-\log K\left(m_s,\boldsymbol{\mu}_{0s}^T\right),\\ \sum_{j=1}^{k}\omega_j=1,\omega_j\geq 0,\end{gathered}\tag{3.9}$$

for $1\leq s<h\leq k$ and where the expectations are given by (3.6).

3.2. Logarithmic opinion pool

Now, a logarithmic opinion pooling is chosen to obtain the aggregated prior distribution. If the same reasoning, as in the previous Subsection is applied, then the solution can be found by solving the nonlinear equation system:

$$\begin{gathered}E_{\pi_l}\left(\log\pi_h\left(\boldsymbol{\theta}\right)\right)-E_{\pi_l}\left(\log\pi_1\left(\boldsymbol{\theta}\right)\right)=0,\ h=2,3,\ldots,k,\\ \sum_{j=1}^{k}\omega_j=1,\end{gathered}\tag{3.10}$$

with the constrains $\omega_j\geq 0$, $j=1,2,\ldots,k$ and, where π_l denotes the combined prior distribution (3.2).

In order to solve it, the expectations have to be calculated. Once again, a general study could be addressed when conjugate prior distributions (2.2) are considered.

For these distributions, the aggregated prior distribution (3.2) is given by:

$$\pi_l(\boldsymbol{\theta}) = t\prod_{j=1}^{k}(\pi_j(\boldsymbol{\theta}))^{\omega_j} = t\prod_{j=1}^{k}\left(K(m_j,\boldsymbol{\mu}_{0j}^T)\exp\left\{m_j\boldsymbol{\mu}_{0j}^T\boldsymbol{\theta} - m_jM(\boldsymbol{\theta})\right\}\right)^{\omega_j}$$

where the normalizing constant t satisfies:

$$t^{-1} = \prod_{j=1}^{k}K(m_j,\boldsymbol{\mu}_{0j}^T)^{\omega_j}K^{-1}(\sum_{j=1}^{k}\omega_j m_j\boldsymbol{\mu}_{0j}^T,\sum_{j=1}^{k}\omega_j m_j),$$

being $K(\sum_{j=1}^{k}\omega_j m_j\boldsymbol{\mu}_{0j}^T,\sum_{j=1}^{k}\omega_j m_j)$, the normalizing constant for the distribution:

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

Thus, the pooled prior distribution is:

$$\begin{aligned}\pi_l(\boldsymbol{\theta}) &= \frac{K(\sum_{j=1}^{k}\omega_j m_j\boldsymbol{\mu}_{0j}^T,\sum_{j=1}^{k}\omega_j m_j)}{\prod_{j=1}^{k}K(m_j,\boldsymbol{\mu}_{0j}^T)^{\omega_j}}\prod_{j=1}^{k}(\pi_j(\boldsymbol{\theta}))^{\omega_j} \qquad (3.11)\\ &= K(\sum_{j=1}^{k}\omega_j m_j\boldsymbol{\mu}_{0j}^T,\sum_{j=1}^{k}\omega_j m_j)\exp\left\{\sum_{j=1}^{k}\omega_j m_j(\boldsymbol{\mu}_{0j}^T\boldsymbol{\theta} - M(\boldsymbol{\theta}))\right\}.\end{aligned}$$

It is satisfied:

$$E_{\pi_l}(\log\pi_s(\boldsymbol{\theta})) = \log K(m_s,\boldsymbol{\mu}_{0s}^T) + m_s\boldsymbol{\mu}_{0s}^t E_{\pi_l}(\boldsymbol{\theta}) - m_sE_{\pi_l}(M(\boldsymbol{\theta})),$$

and, in consequence, the nonlinear equation system (3.10) has the following expression:

$$\begin{aligned}&(m_h\boldsymbol{\mu}_{0h}^t - m_1\boldsymbol{\mu}_{01}^t)E_{\pi_l}(\boldsymbol{\theta}) + (m_1 - m_h)E_{\pi_l}(M(\boldsymbol{\theta}))\\ &\quad = \log K(m_1,\boldsymbol{\mu}_{01}^t) - \log K(m_h,\boldsymbol{\mu}_{0h}^t),\ \text{for } h = 2,3,\ldots,k, \qquad (3.12)\\ &\qquad\sum_{j=1}^{k}\omega_j = 1,\end{aligned}$$

being $\omega_j \geq 0$, $j = 1,2,\ldots,k$.

Now, the main aim is to obtain the expectations $E_{\pi_l}(\boldsymbol{\theta})$ and $E_{\pi_l}(M(\boldsymbol{\theta}))$, respectively. Observe that, the pooled prior distribution (3.11) belongs to the same class as the prior distributions (2.2). Therefore it has, in particular, the general expression:

$$\pi(\boldsymbol{\theta}) = K(m,\boldsymbol{\mu}_0^T)\exp\left\{m\boldsymbol{\mu}_0^T\boldsymbol{\theta} - mM(\boldsymbol{\theta})\right\}, \qquad (3.13)$$

where the parameters, between the expressions (3.13) and (3.11), can be identified as:

$$m = \sum_{j=1}^{k}\omega_j m_j \text{ and } m\boldsymbol{\mu}_0^T = \sum_{j=1}^{k}\omega_j m_j\boldsymbol{\mu}_{0j}^T.$$

Hence, in order to obtain the expectations $E_{\pi_l}(\boldsymbol{\theta})$ and $E_{\pi_l}(M(\boldsymbol{\theta}))$, the same scheme as the one used in the linear case could be followed.

Observe that, it is not always possible to find solution on the simplex $S_k = \{(\omega_1, \omega_2, \ldots, \omega_k) : \sum_{j=1}^{k} \omega_j = 1, \omega_j \geq 0, j = 1, 2, \ldots, k\}$. In order to solve this problem the procedure in the previous Subsection could be applied. Now, the following nonlinear optimization problem is obtained:

minimize u such that:

$$\begin{gathered}\left(m_h \boldsymbol{\mu}_{0h}^T - m_l \boldsymbol{\mu}_{0s}^T\right) E_{\pi_l}\left(\boldsymbol{\theta}\right) + \left(m_l - m_h\right) E_{\pi_l} M\left(\left(\boldsymbol{\theta}\right)\right) - u \\ \leq \log K\left(m_s, \boldsymbol{\mu}_{0s}^T\right) - \log K\left(m_h, \boldsymbol{\mu}_{0h}^T\right), \\ -\left[\left(m_h \boldsymbol{\mu}_{0h}^T - m_s \boldsymbol{\mu}_{0s}^T\right) E_{\pi_l}\left(\boldsymbol{\theta}\right) + \left(m_s - m_h\right) E_{\pi_l} M\left(\left(\boldsymbol{\theta}\right)\right)\right] - u \\ \leq \log K\left(m_h, \boldsymbol{\mu}_{0h}^T\right) - \log K\left(m_s, \boldsymbol{\mu}_{0s}^T\right), \\ \sum_{j=1}^{k} \omega_j = 1, \omega_j \geq 0,\end{gathered} \tag{3.14}$$

for $1 \leq s < h \leq k$.

4. Illustrative examples

In this Section two examples are presented to illustrate the proposed Bayesian approaches. Firstly, a practical application for the multinomial distribution is shown. Next, an example for the bivariate normal distribution is considered. Observe that, in both cases, bivariate random variables are considered in order to graphically illustrate both the component prior distributions and the joint prior distributions.

4.1. Linear opinion pool: Multinomial sampling

A discrete random vector $\boldsymbol{x} = (x_1, x_2)^T$ has a multinomial distribution of dimension 2, if its probability mass function is

$$f\left(\boldsymbol{x}|\boldsymbol{p}, r\right) = \frac{r!}{\prod\limits_{t=1}^{2} x_t!(r - \sum\limits_{t=1}^{2} x_t)!} \prod_{t=1}^{2} p_t^{x_t} (1 - \sum_{t=1}^{2} p_t)^{r - \sum_{t=1}^{2} x_t},$$

where $\sum_{t=1}^{2} x_t \leq r$, $r = 1, 2, \ldots$, $0 < p_t < 1$, $\sum_{t=1}^{2} p_t < 1, t = 1, 2$.

The canonical representation (see, e.g., [8]) is given by (2.1), where the natural parameter is $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$, and

$$\theta_t = \log\left(\frac{p_t}{1 - p_1 - p_2}\right),\ t = 1, 2,$$

thus, $\boldsymbol{\Theta} = \mathbb{R}^2$ and $M\left(\boldsymbol{\theta}\right)$ can be expressed as:

$$M\left(\boldsymbol{\theta}\right) = -r \log(1 - p_1 - p_2) = r \log(1 + e^{\theta_1} + e^{\theta_2}).$$

Suppose that three experts supply prior information over $\boldsymbol{p} = (p_1, p_2)^T$ as three Dirichlet distributions with parameters $m_j\mu_{01j}$, $m_j\mu_{02j}$, $m_j(r-\mu_{01j}-\mu_{02j})$, $j = 1, 2, 3$, where $m_j > 0$ and $\mu_{0tj} \in (0, r)$, $t = 1, 2$. If the canonical parameterization is considered, then the combined prior distributions over $\boldsymbol{\theta}$ is:

$$\pi_m(\boldsymbol{\theta}) = \sum_{j=1}^{3} \omega_j \pi_j(\boldsymbol{\theta}) = \sum_{j=1}^{3} \omega_j K(m_j, \boldsymbol{\mu}_{0j}^T) \exp\left\{m_j \boldsymbol{\mu}_{0j}^T \boldsymbol{\theta} - m_j M(\boldsymbol{\theta})\right\}, \quad (4.1)$$

where the normalizing constant is given by the expression:

$$K(m_j, \boldsymbol{\mu}_{0j}^T) = \frac{\Gamma(rm_j)}{\prod_{t=1}^{2} \Gamma(m_j\mu_{0tj})\, \Gamma(m_j(r - \sum_{t=1}^{2} \mu_{0tj}))}, j = 1, 2, 3.$$

Next, in order to obtain the weights, the expectations $E_{\pi_j}(\boldsymbol{\theta})$ and $E_{\pi_j}(M(\boldsymbol{\theta}))$ must be calculated. It is satisfied:

$$-\log K(m_j, \boldsymbol{\mu}_{0j}^T) = \sum_{t=1}^{2} \log\Gamma(m_j\mu_{0tj}) + \log\Gamma(m_j(r - \sum_{t=1}^{2} \mu_{0tj})) - \log\Gamma(rm_j).$$

Thus, the partial derivatives are given by:

$$\begin{aligned}
\partial_{\boldsymbol{\mu}_{0j}}(-\log K(m_j, \boldsymbol{\mu}_{0j}^T)) &= \Bigg(m_j(\Psi(m_j\mu_{01j}) - \Psi(m_j(r - \sum_{t=1}^{2} \mu_{0tj}))), m_j \\
&\quad \times(\Psi(m_j\mu_{02j}) - \Psi(m_j(r - \sum_{t=1}^{2} \mu_{0tj})))\Bigg)^T, \\
\partial_{m_j}(-\log K(m_j, \boldsymbol{\mu}_{0j}^T)) &= \sum_{t=1}^{2} \mu_{0tj}\Psi(m_j\mu_{0tj}) + (r - \sum_{t=1}^{2} \mu_{0tj}) \\
&\quad \times\Psi(m_j(r - \sum_{t=1}^{2} \mu_{0tj})) - r\Psi(rm_j),
\end{aligned}$$

where $\Psi(\cdot) = \Gamma'(\cdot) \big/ \Gamma(\cdot)$ denotes the digamma function.

From both previous expressions, it is obtained:

$$E_{\pi_j}(\boldsymbol{\theta}) = \left(\Psi(m_j\mu_{01j}) - \Psi(m_j(r - \sum_{t=1}^{2} \mu_{0tj})), \Psi(m_j\mu_{02j}) - \Psi(m_j(r - \sum_{t=1}^{2} \mu_{0tj}))\right)^T,$$

and

$$E_{\pi_j}(M(\boldsymbol{\theta})) = r\left(\Psi(rm_j) - \Psi(m_j(r - \sum_{t=1}^{2} \mu_{0tj}))\right).$$

Therefore, both the linear equation system (3.5) and the linear optimization problem (3.9) can be easily obtained by considering the previous expressions.

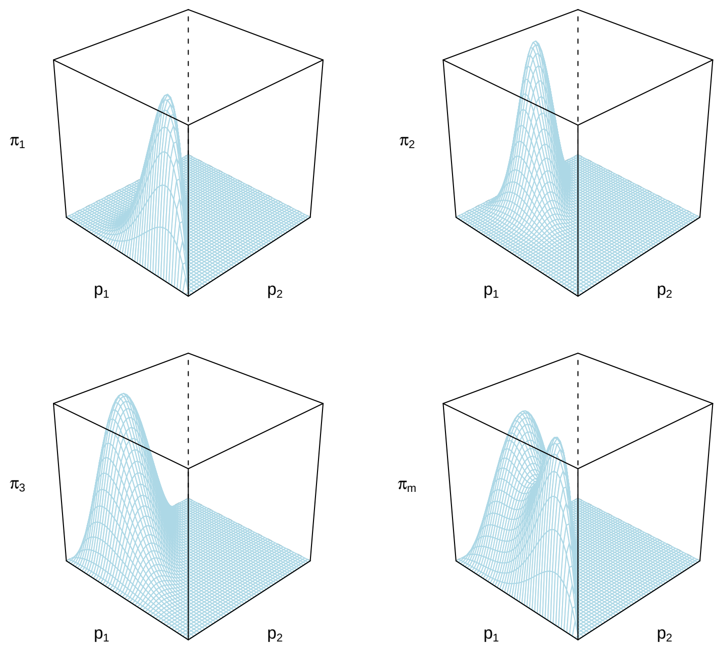

Fig 1. *Prior distributions and mixture prior distributions.*

In order to illustrate the previous development, it is assumed that the three experts are providing prior information over $\boldsymbol{\theta}$ in $r = 10$ trials. This individual information is combined by using the mixture of prior distributions (4.1). The selected parameters are $m_1 = 1$, $m_2 = 1.5$, $m_3 = 0.9$, $\boldsymbol{\mu}_{01} = (6, 2, 2)^T$, $\boldsymbol{\mu}_{02} = (2, 14/3, 10/3)^T$ and $\boldsymbol{\mu}_{03} = (20/9, 10/3, 40/9)^T$. These parameters have been chosen with the aim of presenting how the proposed procedures behave. Now, the weights are calculated by solving the linear equation system given in (3.5). By using the previous parameters, the following linear equation system is obtained:

$$\begin{aligned} -12.9951\omega_1 - 2.2575\omega_2 - 3.5891\omega_3 &= -6.7259, \\ -3.3043\omega_1 + 3.8682\omega_2 + 3.8845\omega_3 &= -0.1055, \\ \omega_1 + \omega_2 + \omega_3 &= 1, \end{aligned}$$

constrained to $\omega_j \geq 0$, $j = 1, 2, 3$. The solution obtained is $(\omega_1, \omega_2, \omega_3) = (0.5516,\ 1.5403,\ -1.0919)$, which does not belong to S_3. Therefore the alternative approach must be used. The linear optimization problem (3.9) is solved, leading to the solution $(\omega_1, \omega_2, \omega_3) = (0.4389, 0.1184, 0.4428)$ and to the optimal value $u = 0.8333$.

Figure 1 shows the prior distributions provided by each expert together with the aggregated prior distribution by using the obtained weights. The usual parameterization is used to represent all prior distributions since they can be more easily visualized.

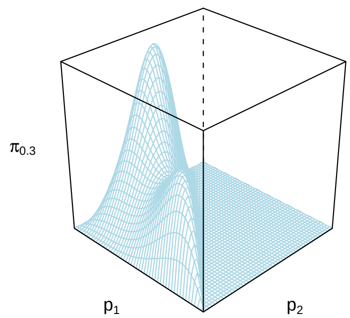

Fig 2. *Mixture prior distributions by using the same weights for all experts.*

Observe that, through this procedure, the best weights combination is obtained, in the sense that it provides the lowest value for the differences between the corresponding Kullback-Leibler divergences. Therefore, the obtained aggregated distribution is the one with the highest agreement with the information provided by the experts. Consequently, any other combination is worse than the previous one since the obtained value for the proposed differences is higher. Figure 2 shows the combined distribution, $\pi_{0.3}$, by using the prior distributions described in the first setting and by taking the same values for the weights ω_j, $j = 1, 2, 3$.

It can be graphically observed, that the the first expert's information is less represented in the density $\pi_{0.3}$ (see Figure 2) than in the combined prior distribution, π_m, displayed in Figure 1. Thus, the value for the differences $|KL(\pi_{0.3}||\pi_h) - KL(\pi_{0.3}||\pi_s)|$, $1 \leq s < h \leq 3$ are higher than those obtained by using the proposed procedure. Consequently, taking equal weights is not the best combination when the criterion in this paper is used.

4.2. Logarithmic opinion pool: Multinomial sampling

Under the initial conditions presented in the previous Subsection, a logarithmic opinion pooling is considered to get the following aggregated prior distribution:

$$\pi\left(\boldsymbol{\theta}\right) = K(\sum_{j=1}^{3} \omega_j m_j \boldsymbol{\mu}_{0j}^T, \sum_{j=1}^{3} \omega_j m_j) \exp\left\{\sum_{j=1}^{3} \omega_j m_j (\boldsymbol{\mu}_{0j}^T \boldsymbol{\theta} - M\left(\boldsymbol{\theta}\right))\right\},$$

where the normalizing constant is given by:

$$K(\sum_{j=1}^{3} \omega_j m_j \boldsymbol{\mu}_{0j}^T, \sum_{j=1}^{3} \omega_j m_j) = \frac{\Gamma(r \sum_{j=1}^{3} \omega_j m_j)}{\prod_{t=1}^{2} \Gamma(\sum_{j=1}^{3} \omega_j m_j \mu_{0_t j}) \Gamma(\sum_{j=1}^{3} \omega_j m_j (r - \sum_{t=1}^{2} \mu_{0_t j}))}.$$

By taking into account the previous development for the linear combination and the development in Subsection 3.2, it is followed that:

$$\begin{aligned} E_{\pi_l}(\boldsymbol{\theta}) &= \Bigg(\Psi(\sum_{j=1}^{3}\omega_j m_j \mu_{0_1 j}) - \Psi(\sum_{j=1}^{3}\omega_j m_j (r - \sum_{t=1}^{2}\mu_{0_t j}), \Psi(\sum_{j=1}^{3}\omega_j m_j \mu_{0_2 j}) \\ &\quad - \Psi(\sum_{j=1}^{3}\omega_j m_j (r - \sum_{t=1}^{2}\mu_{0_t j})\Bigg)^T, \end{aligned}$$

and

$$E_{\pi_l}(M(\boldsymbol{\theta})) = r\left(\Psi(\sum_{j=1}^{3} r\omega_j m_j) - \Psi(\sum_{j=1}^{3}\omega_j m_j (r - \sum_{t=1}^{2}\mu_{0_t j})\right).$$

Therefore, for the considered parameter values $m_1 = 1$, $m_2 = 1.5$, $m_3 = 0.9$, $\boldsymbol{\mu}_{01} = (6,2,2)^T$, $\boldsymbol{\mu}_{02} = (2,14/3,10/3)^T$ and $\boldsymbol{\mu}_{03} = (20/9,10/3,40/9)^T$, the nonlinear equation system is given by:

$$\begin{gathered} -3\Psi(6\omega_1 + 3\omega_2 + 2\omega_3) + 5\Psi(2\omega_1 + 7\omega_2 + 3\omega_3) - 5\Psi(10\omega_1 + 15\omega_2 + 9\omega_3) \\ +3\Psi(2\omega_1 + 5\omega_2 + 4\omega_3) = -6.7259, \\ -4\Psi(6\omega_1 + 3\omega_2 + 2\omega_3) + \Psi(2\omega_1 + 7\omega_2 + 3\omega_3) + \Psi(10\omega_1 + 15\omega_2 + 9\omega_3) \\ +2\Psi(2\omega_1 + 5\omega_2 + 4\omega_3) = -0.1055 \\ \omega_1 + \omega_2 + \omega_3 = 1. \end{gathered}$$

As in the linear case, it is not possible to find a solution that satisfies $\omega_j \geq 0$, $j = 1,2,3$, i.e., on the simplex S_3. Hence, the nonlinear optimization problem (3.14) is solved. The following solution is found: $(\omega_1, \omega_2, \omega_3) = (0.5406, 0.4594, 0)$, being the optimal value $u = 0.3596$.

Thus, according to the proposal made in (3.7), a combined prior distribution that considers only the information provided by the first and second experts, is enough to represent the initial group opinion. In addition, this combination is the best one under the criterion exposed in this paper. Figure 3 shows the prior distributions provided by each expert together with the aggregated prior distribution by using the calculated weights.

In contrast to the linear opinion pool in the previous Subsection, it can be observed that a narrower prior distribution is obtained with the logarithmic opinion pool, i.e., it is more concentrated on the common range of parameter values.

Figure 4 shows the combined distribution, $\pi_{0.3}$, by using the prior distributions described in the first setting and by taking the same values for the weights ω_j, $j = 1,2,3$. By considering the comment in the previous paragraph, it is visually complicated to observe the differences between both combined prior distributions. Thus, Figure 4 also exhibits the logarithmic pools under the two weight vectors. The third expert has more impact on the aggregated prior distribution $\pi_{0.3}$ than on the prior distribution π_l. Hence, under the proposal in this paper, the expert 3 has to be excluded in order to obtain the best weight combination.

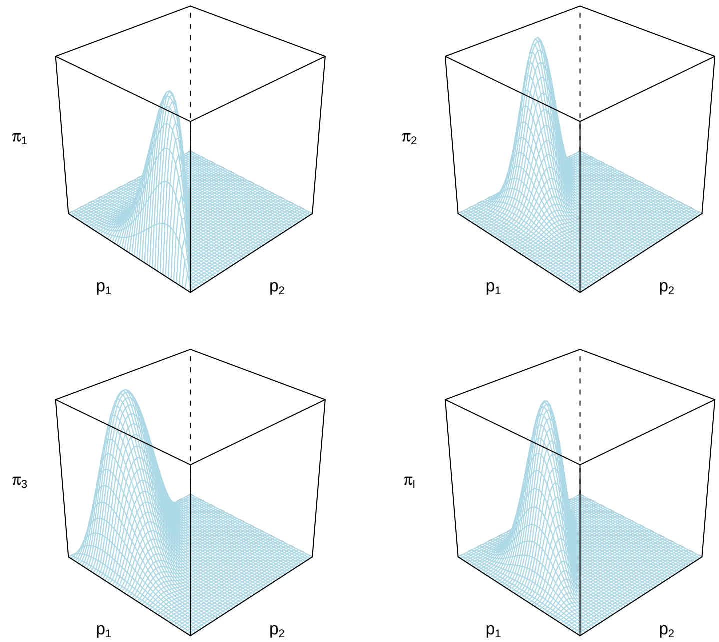

FIG 3. *Prior distributions and logarithmic aggregated prior distribution.*

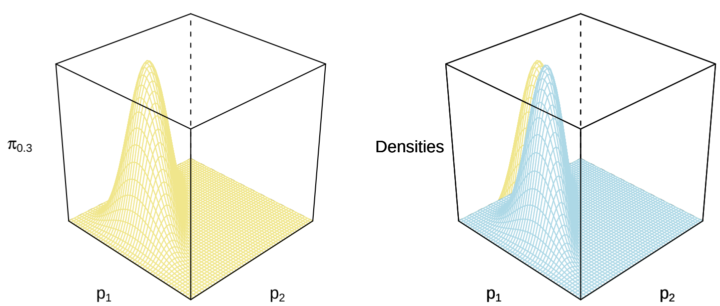

FIG 4. *Logarithmic aggregated prior distributions*

4.3. Linear opinion pool: Bivariate normal distribution

A continuous random vector $\boldsymbol{x} = (x_1, x_2)^T$ has a bivariate normal distribution, if its density function is

$$f(\boldsymbol{x}|\lambda) = \frac{1}{2\pi |\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\lambda})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\lambda})\right\},$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^T$ is the two-dimensional mean vector and $\boldsymbol{\Sigma}$ denotes the non-singular covariance matrix, i.e.:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{x_1}^2 & \rho\sigma_{x_1}\sigma_{x_2} \\ \rho\sigma_{x_1}\sigma_{x_2} & \sigma_{x_2}^2 \end{pmatrix},$$

with ρ being the correlation between the random variables X_1 and X_2. It is assumed that this matrix is fixed.

The canonical representation is given by (2.1), where $\boldsymbol{\theta} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}$, $\boldsymbol{\Theta} = \mathbb{R}^2$, $M(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\lambda}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda} = \frac{1}{2}\boldsymbol{\theta}^T\boldsymbol{\Sigma}\boldsymbol{\theta}$ and $b(\boldsymbol{x}) = \frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}}\exp\{-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}\}$ (see, e.g., [22]).

Suppose that two experts provide prior information over $\boldsymbol{\lambda}$ as two normal prior distributions $N(\boldsymbol{\mu}_{0j}, \boldsymbol{\Sigma}/m_j)$, $j = 1, 2$. If the canonical parameterization is used, the prior distributions over $\boldsymbol{\theta} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}$ supplied by each expert are $N(\boldsymbol{\mu}_{0j}\boldsymbol{\Sigma}^{-1}, \boldsymbol{\Sigma}^{-1}/m_j)$, $j = 1, 2$.

Then, the mixture of prior distributions is given by:

$$\pi_m(\boldsymbol{\theta}) = \sum_{j=1}^{2} \omega_j K(m_j, \boldsymbol{\mu}_{0j}^T) \exp\left\{ m_j\boldsymbol{\mu}_{0j}^T\boldsymbol{\theta} - m_j\frac{1}{2}\boldsymbol{\theta}^T\boldsymbol{\Sigma}\boldsymbol{\theta} \right\},$$

where $\omega_1 + \omega_2 = 1$, $\omega_j \geq 0$ and

$$K(m_j, \boldsymbol{\mu}_{0j}^T) = \frac{m_j^{1/2}}{2\pi\left|\boldsymbol{\Sigma}^{-1}\right|^{1/2}} \exp\left\{ -\frac{1}{2}\boldsymbol{\mu}_{0j}^T m_j \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_{0j} \right\},$$

with

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_{x_1}^2\sigma_{x_2}^2(1-\rho^2)} \begin{pmatrix} \sigma_{x_2}^2 & -\rho\sigma_{x_1}\sigma_{x_2} \\ -\rho\sigma_{x_1}\sigma_{x_2} & \sigma_{x_1}^2 \end{pmatrix}.$$

In order to obtain the linear equation system given in (3.5), the expectations $E_{\pi_j}(\boldsymbol{\theta})$ and $E_{\pi_j}(M(\boldsymbol{\theta}))$ must be calculated. Firstly, the partial derivatives $\partial_{\boldsymbol{\mu}_{0j}}$ $(-\log\ K(m_j,\ \boldsymbol{\mu}_{0j}^T))$ and $\partial_{m_j}(-\log K(m_j, \boldsymbol{\mu}_{0j}^T))$ are obtained. By taking into account the following equality:

$$-\log K\left(m_j, \boldsymbol{\mu}_{0j}^T\right) = -\frac{1}{2}\log m_j + \log(2\pi) + \frac{1}{2}\log\left|\boldsymbol{\Sigma}^{-1}\right| + \frac{1}{2}\boldsymbol{\mu}_{0j}^T m_j\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_{0j},$$

it is obtained:

$$\begin{aligned} \partial_{\boldsymbol{\mu}_{0j}}\left(-\log K\left(m_j, \boldsymbol{\mu}_{0j}^T\right)\right) \quad = \quad & \left(\frac{m_j}{2\left|\boldsymbol{\Sigma}\right|}\left(2\mu_{0_1j}\sigma_{x_2}^2 - 2\rho\sigma_{x_1}\sigma_{x_2}\mu_{0_2j}\right),\right. \\ & \left.\frac{m_j}{2\left|\boldsymbol{\Sigma}\right|}\left(-2\rho\sigma_{x_1}\sigma_{x_2}\mu_{0_1j} + 2\mu_{0_2j}\sigma_{x_1}^2\right)\right)^T. \end{aligned}$$

Therefore:

$$E_{\pi_j}(\boldsymbol{\theta}) = \left(\frac{1}{2\left|\boldsymbol{\Sigma}\right|}\left(2\mu_{0_1j}\sigma_{x_2}^2 - 2\rho\sigma_{x_1}\sigma_{x_2}\mu_{0_2j}\right), \frac{1}{2\left|\boldsymbol{\Sigma}\right|}\left(-2\rho\sigma_{x_1}\sigma_{x_2}\mu_{0_1j} + 2\mu_{0_2j}\sigma_{x_1}^2\right)\right)^T.$$

On the other hand, it is satisfied:

$$\partial_{m_j}\left(-\log K\left(m_j, \boldsymbol{\mu}_{0j}^T\right)\right) = -\frac{1}{2m} + \frac{1}{2}\boldsymbol{\mu}_{0j}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_{0j}.$$

Then,

$$E_{\pi_j}\left(M\left(\boldsymbol{\theta}\right)\right) = \frac{1}{2m_j} - \frac{\rho\sigma_{x_1}\sigma_{x_2}\mu_{0_1 j}\mu_{0_2 j}}{|\boldsymbol{\Sigma}|} + \frac{\mu_{0_1 j}^2\sigma_{x_2}^2 + \mu_{0_2 j}^2\sigma_{x_1}^2}{2\,|\boldsymbol{\Sigma}|}.$$

Thus, it is finally obtained the following linear equation system:

$$\begin{gathered}
\sum_{j=1}^{2}\omega_j\left[\left(m_2\boldsymbol{\mu}_{02}^T - m_1\boldsymbol{\mu}_{01}^T\right)E_{\pi_j}\left(\boldsymbol{\theta}\right) + \left(m_1 - m_2\right)E_{\pi_j}M\left(\left(\boldsymbol{\theta}\right)\right)\right] \\
= \log K\left(m_1, \boldsymbol{\mu}_{01}^T\right) - \log K\left(m_2, \boldsymbol{\mu}_{02}^T\right), \\
\sum_{j=1}^{k}\omega_j = 1
\end{gathered} \tag{4.2}$$

where $E_{\pi_j}(\boldsymbol{\theta})$ and $E_{\pi_j}(M(\boldsymbol{\theta}))$ are given by the previous expressions and $\omega_j \geq 0$, $j = 1, 2$.

[18] consider Bayesian inference when prior distributions and likelihood functions are both available for inputs and outputs of a deterministic simulation model. This problem is related to the issue of aggregating experts' opinions. They studied alternative strategies for aggregation, then describe computational approaches for implementing pooled inference for simulation models. The following particular situation was considered:

Let $M(\boldsymbol{\phi}) = (\frac{1}{2}\phi_1 - \frac{1}{2}\phi_2 + 2,\ \frac{1}{4}\phi_1 + \frac{1}{8}\phi_2 - 1) = (\theta_1, \theta_2) = \boldsymbol{\theta}^T$ be a model linking an input $\boldsymbol{\phi}$ to an output $\boldsymbol{\theta}$. Suppose an expert on $\boldsymbol{\phi}$ (Expert 1) and an expert on $\boldsymbol{\theta}$ (Expert 2) are independently consulted, and they each offer a prior distribution which describes their beliefs and corresponding uncertainty about reasonable values of their parameter of expertise. Given $M(\boldsymbol{\phi})$, Expert 2 implicitly professes an opinion about reasonable values of $\boldsymbol{\phi}$. Similarly, Expert 1 implicitly professes an opinion about $\boldsymbol{\theta}$. This occurs even if each expert has no knowledge of the other's field. Assume that the solicited prior distributions are two-dimensional prior distributions (see Figure 5):

$$\pi_1\left(\boldsymbol{\theta}\right) = N_2\left[\begin{pmatrix} 3 \\ -\frac{13}{8}\end{pmatrix}, \begin{pmatrix} 1 & .8 \\ .8 & 2\end{pmatrix}\right],\ \pi_2\left(\boldsymbol{\theta}\right) = N_2\left[\begin{pmatrix} 1 \\ 0\end{pmatrix}, \begin{pmatrix} 2 & 1.6 \\ 1.6 & 4\end{pmatrix}\right].$$

Then, the two available prior distributions for $\boldsymbol{\theta}$, $\pi_1(\boldsymbol{\theta})$ and $\pi_2(\boldsymbol{\theta})$ are pooled through (3.1) to form a single prior $\pi(\boldsymbol{\theta})$.

Next, the weights must be calculated such that the mixture of prior distributions represents a consensus between the beliefs of the experts. In order to do it, the linear equation system given by (4.2) is solved for this particular case. The solution is $(\omega_1, \omega_2) = (0.66522, 0.33478)$, which belongs to S_2. Figure 6 shows the mixtures of prior distributions by using the previously calculated weights and $\omega_1 = \omega_2 = 0.5$, respectively. This last pooled distribution is denoted by $\pi_{0.5}$.

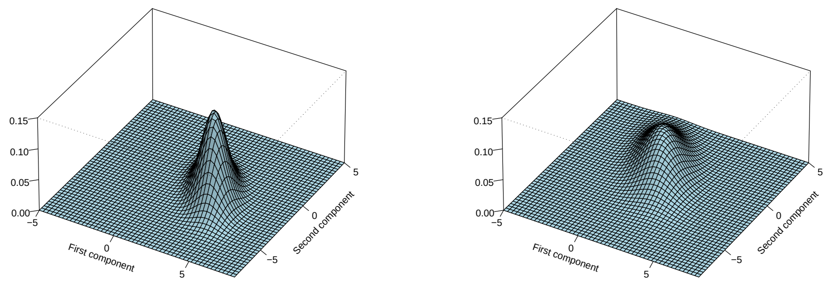

Fig 5. *From left to right: Prior distributions* $\pi_1(\boldsymbol{\theta})$ *and* $\pi_2(\boldsymbol{\theta})$.

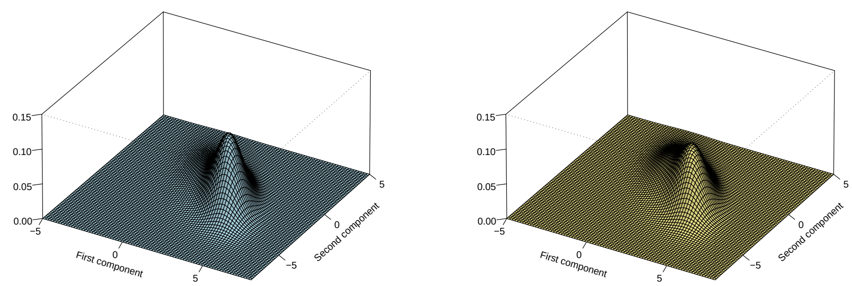

Fig 6. *From left to right: Mixture prior distributions* $\pi_m(\boldsymbol{\theta})$ *and* $\pi_{0.5}(\boldsymbol{\theta})$.

Given the number of experts as well as the prior distributions elicited by them, it is difficult to made a comparison between the two combined prior distributions. Nevertheless, it can be observed that when the obtained weights are used, then the prior distribution π_m is more peak-shaped than the aggregated prior distribution $\pi_{0.5}$. Therefore, the information supplied by the first expert is less represented in the joint prior distribution $\pi_{0.5}$ than in the prior distribution π_m.

Consequently, whereas the density π_m is providing equal reliability to the information provided by the experts, i.e., $KL(\pi_m||\pi_1) = KL(\pi_m||\pi_2)$, the joint prior distribution, $\pi_{0.5}$, satisfies $KL(\pi_{0.5}||\pi_2) < KL(\pi_{0.5}||\pi_1)$.

4.4. Logarithmic opinion pool: Bivariate normal distribution

When a logarithmic opinion pooling is used, it is obtained:

$$\pi\left(\boldsymbol{\theta}\right) = K(\sum_{j=1}^{2} \omega_j m_j \boldsymbol{\mu}_{0j}^T, \sum_{j=1}^{2} \omega_j m_j) \exp\left\{\sum_{j=1}^{2} \omega_j m_j \boldsymbol{\mu}_{0j}^T \boldsymbol{\theta} - m_j M\left(\boldsymbol{\theta}\right))\right\},$$

which, is a normal distribution: $N\left(\frac{\sum_{j=1}^{2}\omega_j m_j \boldsymbol{\mu}_{0j}}{\sum_{j=1}^{2}\omega_j m_j}\boldsymbol{\Sigma}^{-1}, \frac{\boldsymbol{\Sigma}^{-1}}{\sum_{j=1}^{2}\omega_j m_j}\right)$. Thus, the normalizing constant has the expression:

$$
\begin{aligned}
K\left(\sum_{j=1}^{2}\omega_j m_j \boldsymbol{\mu}_{0j}^{T}, \sum_{j=1}^{2}\omega_j m_j\right) &= \frac{\left(\sum\limits_{j=1}^{2}\omega_j m_j\right)^{1/2}}{2\pi\left|\Sigma^{-1}\right|^{1/2}} \times \exp\left\{\left(\frac{\sum\limits_{j=1}^{2}\omega_j m_j \boldsymbol{\mu}_{0j}^{T}}{\sum\limits_{j=1}^{2}\omega_j m_j}\right)\right. \\
&\quad \left. \times \left(\sum_{j=1}^{2}\omega_j m_j\right)\Sigma^{-1}\left(\frac{\sum\limits_{j=1}^{2}\omega_j m_j \boldsymbol{\mu}_{0j}}{\sum\limits_{j=1}^{2}\omega_j m_j}\right)\right\}.
\end{aligned}
$$

Next, the expectations $E_{\pi_l}(\boldsymbol{\theta})$ and $E_{\pi_l}(M(\boldsymbol{\theta}))$ have to be calculated. By taking into account the previous development for the linear combination and the comments in Subsection 3.2, it is followed that:

$$
\begin{aligned}
E_{\pi_l}(\boldsymbol{\theta}) &= \left(\frac{1}{|\Sigma|\sum\limits_{j=1}^{2}\omega_j m_j}\left(\sum_{j=1}^{2}\omega_j m_j \mu_{0_1 j}\sigma_{x_2}^2 - \sum_{j=1}^{2}\omega_j m_j \mu_{0_2 j}\rho\sigma_{x_1}\sigma_{x_2}\right),\right. \\
&\quad \left. \frac{1}{|\Sigma|\sum\limits_{j=1}^{2}\omega_j m_j}\left(\sum_{j=1}^{2}\omega_j m_j \mu_{0_2 j}\sigma_{x_1}^2 - \sum_{j=1}^{2}\omega_j m_j \mu_{0_1 j}\rho\sigma_{x_1}\sigma_{x_2}\right)\right)^{T},
\end{aligned}
$$

and

$$
\begin{aligned}
E_{\pi_l}(M(\boldsymbol{\theta})) = &-\frac{1}{2\sum\limits_{j=1}^{2}\omega_j m_j} + \frac{1}{2|\Sigma|\left(\sum\limits_{j=1}^{2}\omega_j m_j\right)^2}\left[\left(\sum_{j=1}^{2}\omega_j m_j \mu_{0_1 j}\right)^2\sigma_{x_2}^2\right. \\
&\left. + \left(\sum_{j=1}^{2}\omega_j m_j \mu_{0_2 j}\right)^2\sigma_{x_1}^2\right] - \frac{\left(\sum\limits_{j=1}^{2}\omega_j m_j \mu_{0_1 j}\right)\left(\sum\limits_{j=1}^{2}\omega_j m_j \mu_{0_2 j}\right)}{|\Sigma|\left(\sum\limits_{j=1}^{2}\omega_j m_j \mu_{0_1 j}\right)^2}\rho\sigma_{x_1}\sigma_{x_2}.
\end{aligned}
$$

Now, the scenario described in [18] is considered. Hence, the prior distributions (see Figure 5) supplied by the two experts are:

$$
\pi_1(\boldsymbol{\theta}) = N_2\left[\begin{pmatrix} 3 \\ -\frac{13}{8} \end{pmatrix}, \begin{pmatrix} 1 & .8 \\ .8 & 2 \end{pmatrix}\right],\ \pi_2(\boldsymbol{\theta}) = N_2\left[\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 1.6 \\ 1.6 & 4 \end{pmatrix}\right].
$$

Given these prior distributions, the nonlinear equation system given in (3.12) is solved The solution obtained is $(\omega_1, \omega_2) = (0.33558, 0.66442)$.

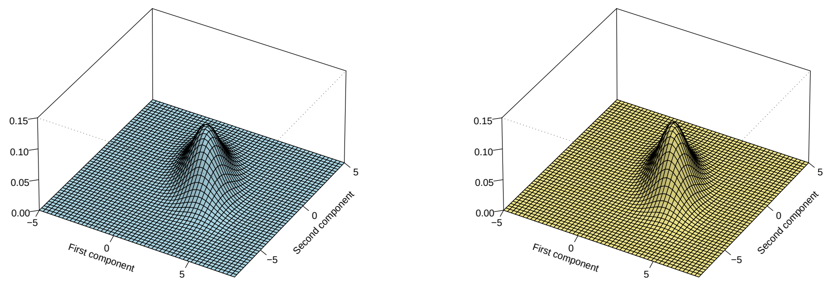

FIG 7. *From left to right: Mixture prior distributions* $\pi_l(\boldsymbol{\theta})$ *and* $\pi_{0.5}(\boldsymbol{\theta})$.

Figure 7 presents the mixtures of prior distributions by using the previously calculated weights and $\omega_1 = \omega_2 = 0.5$, respectively. This last pooled distribution is denoted by $\pi_{0.5}$.

Contrary to the linear combination in the previous Subsection, it can be noted that the logarithmic opinion pool produces a prior distribution more concentrated on the common range of parameter values. Thus, depending on what the decision maker needs, the obtained aggregated prior distributions have different characteristics.

In relation to the joint prior distribution $\pi_{0.5}$, Figure 7 shows that it is more peak-shaped than the density π_l. Thus, the information provided by the first expert has more impact on the aggregated prior distribution $\pi_{0.5}$ than on the prior density π_l. In consequence, it satisfies $KL(\pi_{0.5}||\pi_1) < KL(\pi_{0.5}||\pi_2)$ and the weight values $(\omega_1, \omega_2) = (0.5, 0.5)$ are not the best ones, under the proposed criterion.

5. Conclusion

The main novelty of this paper is to build a combined prior distribution by considering two axiomatic approaches in multivariate settings. Firstly, a theoretical development is performed under the two axiomatic methods. This process leads to different optimization problems and, in consequence, to weights according to the features of the used formulation. This last fact is illustrated with examples, where it is shown how the combined prior distributions behave with the obtained weights and equal weights. The proposed methodology is applied to the natural exponential families. Thus, it is presented how is possible a direct implementation for several distributions widely used in practice. Hence, despite the fact that everything becomes computationally much more difficult in multidimensional situations, the proposed approaches are reduced to problems based on simple theory that can be easily solved.

Finally, it is important to observe that the general proposal in Subsections 3.1 and 3.2, is valid independently of the considered distributions. In particular, the use of prior predictive distributions as the experts' distributions would be very useful in many applications (see, for instance, [20] and references therein). Thus, if the expectations involved in the process can be analytically calculated, then a similar scheme as the one developed for the natural exponential families could be followed. Otherwise, simulation techniques could be adequately used, in order to obtain the expectations required to apply the proposal.

Acknowledgements

This research has been supported by *Ministerio de Ciencia e Innovación* and FEDER, Spain (Project MTM2011-28983-C03-02) and *Junta de Extremadura* (Project GRU10110).

References

[1] Abbas, A. E. (2009), 'A Kullback–Leibler view of linear and log–linear pools', *Decision Analysis* **6**, 25–37.

[2] 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. MR0196777

[3] Bacharach, M. (1975), 'Group decisions in the face of differences of opinion', *Management Science* **22**, 182–191. MR0401078

[4] Barndorff-Nielsen, O. (1978), *Information and Exponential Families in Statistical Theory*, Wiley. MR0489333

[5] Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis*, Springer. MR0804611

[6] Bernardo, J. M. and Smith, A. (1994), *Bayesian Theory*, John Wiley & Sons. MR1274699

[7] Brown, L. D. (1986), *Fundamental of Statistical Exponential Families with Applications in Statistical Decision Theory*, Lecture Notes, 9. Hayward: Institute of Mathematical Statistics. MR0882001

[8] Casalis, M. (1996), 'The $2d + 4$ simple quadratic natural exponential families on $\mathbb{R}^d$', *The Annals of Statistics* **24**(4), 1828–1854. MR1416663

[9] Chen, Y. and Pennock, D. M. (2005), 'Information markets vs opinion pools: An Empirical comparison', *Proceedings of the 6th ACM Conference on Electronic Commerce* pp. 58–67.

[10] Clemen, R. T. and Winkler, R. (1999), 'Combining probability distributions from experts in risk analysis', *Risk Analysis* **19**(2), 187–203.

[11] Clemen, R. T. and Winkler, R. (2007), Aggregating probability distributions, *in* R. M. W. Edwards and D. von Winterfeldt, eds, 'Advances in Decision Analysis: From Foundations to Applications', Cambridge, UK: Cambridge University Press, pp. 154–176.

[12] Cook, R. M. (1991), *Experts in Uncertainty: Opinion and Subjective Probabiliy in Science*, Oxford University Press, New York. MR1136548

[13] Diaconis, P. and Ylvisaker, D. (1979), 'Conjugate priors for exponential families', *The Annals of Statistics* **7**, 269–281. MR0520238

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

[15] Garthwaite, P. H., Kadane, J. B. and O' Hagan, A. (2005), 'Eliciting probability distributions', *Journal of American Statistical Association* **100**(470), 680–700. MR2170464

[16] Genest, C. (1984), 'A characterization theorem for externally Bayesian groups', *The Annals of Statistics* **16**(3), 1100–1105. MR0751297

[17] Genest, C. and Zidek, J. V. (1986), 'Combining probability distributions: A critique and an annotated bibliography', *Statistical Science* **1**(1), 114–148. MR0833278

[18] Givens, G. H. and Roback, P. J. (1999), 'Logarithmic pooling of priors linked by a deterministic simulation model', *Journal of Computational and Graphical Statistics* **30**, 438–457.

[19] Gutiérrez-Peña, E. and Smith, A. F. M. (1997), 'Exponential and Bayesian conjugate families: Review and extensions (with discussion)', *Test* **6**, 1–90. MR1466433

[20] Kadane, J. B. and Wolfson, L. J. (1998), 'Experiences in elicitation', *Journal of the Royal Statistical Society. Series D (The Statistician)* **47**(1), 3–19.

[21] Kascha, C. and Ravazzolo, F. (2010), 'Combining inflation density forecast', *Journal of Forecasting* **29**, 231–250. MR2752012

[22] Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000), *Continous Multivariate distributions, Volume 1, Models ans applications, 2nd Edition*, Wiley Series in Probability and Statistics.

[23] Meyer, M. A. and Booker, J. M. (2001), *Eliciting and Analyzing Expert Judgement: A Practical Guide*, Society of Industrial and Applied Mathematics (SIAM), Philadelphia. MR1823794

[24] 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.

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

[26] Regan, H. M., Colyvan, M. and Markovchick-Nicholls, L. (2006), 'A formal model for consensus and negotiation in environmental management', *Journal of Environmental Management* **80**(2), 167–176.

[27] 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. MR2751136

[28] Stone, M. (1961), 'The opinion pool', *Annals of Mathematical Statistics* **32**, 1339–1342. MR0135190