Merging experts’ opinions: A Bayesian hierarchical model with mixture of prior distributions

M. J. Rufo*, C. J. Pérez, J. Martín

Departamento de Matemáticas, Universidad de Extremadura, Avda. de la Universidad s/n, 10071 Cáceres, Spain.

Abstract

In this paper, a general approach is proposed to address a full Bayesian analysis for the class of quadratic natural exponential families in the presence of several expert sources of prior information. By expressing the opinion of each expert as a conjugate prior distribution, a mixture model is used by the decision maker to arrive at a consensus of the sources. A hyperprior distribution on the mixing parameters is considered and a procedure based on the expected Kullback-Leibler divergence is proposed to analytically calculate the hyperparameter values. Next, the experts’ prior beliefs are calibrated with respect to the combined posterior belief over the quantity of interest by using expected Kullback-Leibler divergences, which are estimated with a computationally low-cost method. Finally, it is remarkable that the proposed approach can be easily applied in practice, as it is shown with an application.

Key words: Bayesian analysis, conjugate prior distributions, exponential families, prior mixtures, Kullback-Leibler divergence.

* Corresponding author. Departamento de Matemáticas, Escuela Politécnica, Universidad de Extremadura, Avda. de la Universidad s/n, 10071 Cáceres, Spain.
Phone: +34 927257220 Fax: +34 927257203
Email addresses: mruf@unex.es (M. J. Rufo), carper@unex.es (C. J. Pérez), jrmartin@unex.es (J. Martín).
1 Introduction

The choice of suitable prior distributions is not a simple task where Bayesian methods are applied, particularly, when issues related to analysis of experts’ opinions and decision making are dealt with (see Korhonen et al. (1992), for a review of multiple criteria decision making problems). Often, the prior distribution is chosen to approximately reflect the initial expert’s opinion. In this context, a common choice is a conjugate prior distribution. However, in some situations, a single conjugate prior distribution may be inadequate to accurately reflect available prior knowledge.

Dalal and Hall (1983) and Diaconis and Ylvisaker (1985) showed that it is possible to extend these distributions through the use of mixtures of conjugate prior distributions (see also Lijoi (2003) for a more recent study). The main advantage is that mixtures of conjugate prior distributions can be sufficiently flexible (allowing, for example, multimodality), while they make simplified posterior calculations possible (since they are also conjugate families). Some interesting applications on prior mixtures can be found in Savchuk and Martz (1994), Liechty et al. (2004), and Atwood and Youngblood (2005).

This paper provides a general framework that allows to perform a full Bayesian analysis for natural exponential families with quadratic variance function (NEF-QVF) by using mixtures of conjugate prior distributions with unknown weights. These families have been considered because they contain distributions very commonly used in real applications, such as Poisson, binomial, negative-binomial, normal or gamma.

Throughout the paper, it is assumed that a decision maker consults several sources about a quantity of interest. Therefore, it is considered that the prior information comes from several sources such as experts. The opinion of each expert is elicited as a conjugate distribution over a quantity of interest (see, e.g., Szwed et al. (2006) for a particular case of prior distribution specification). Then, the decision maker combine the experts’ distributions by using a mixture model in order to represent a consensus of several experts. Chen and Pennock (2005) observed that the weights selection is an inconvenience of this
approach. Sometimes, the weights are fixed in advance. Here, the weights are considered as parameters and a suitable hyperprior distribution is specified. This fact leads to greater freedom and flexibility in the modeling of initial information. In order to obtain the hyperparameter values, a general procedure based on expected Kullback-Leibler divergences is proposed. An advantage is that the process is analytical. General expressions that allow a direct implementation for all distributions in these families are obtained. Nevertheless, other hyperparameter values can be chosen by the reader and used in the subsequent Bayesian analysis.

Finally, the expected discrepancies between the combined posterior belief over the quantity of interest and each expert’s prior belief are analyzed by using the expected Kullback-Leibler divergence between the mixture of the posterior distributions for this quantity and the prior distribution for each expert. A Monte Carlo-based approach is considered to estimate these values.

The outline of the paper is as follows. Section 2 presents the basic concepts and notation. In Section 3, a Bayesian analysis of NEF-QVF distributions by using mixtures of conjugate prior distributions is developed. In Section 4, the experts’ prior opinions are calibrated with respect to the combined posterior opinion over the quantity of interest by using expected Kullback-Leibler divergences. Section 5 shows a binomial application. Finally, a conclusion and a discussion including an alternative method for the choice of the hyperparameter values, are presented in Section 6.

2 Background

In this section, a short review of the natural exponential family and conjugate prior distributions is presented. Besides, the notation will be fixed for the rest of the paper.

Let \( \eta \) be a \( \sigma \)-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 \( \eta \) is:

\[
f_\theta(x) = \exp \{x \theta - M(\theta)\}, \quad \theta \in \Theta,
\]

where \( M(\theta) = \log \int \exp(x \theta) \eta(dx) \) and \( \Theta = \{\theta \in \mathbb{R} : M(\theta) < \infty\} \) is nonempty. \( \theta \) is called the natural parameter. Besides, it is satisfied \( E(X|\theta) = M'(\theta) = \mu \).

See Brown (1986) for a review on this family.

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 (1) and \( \Omega \) is the mean space. For 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 (see Morris (1982) and Gutiérrez-Peña (1997)).

Conjugate prior distributions as in Morris (1983) and Gutiérrez-Peña and Smith (1997) are considered. Therefore, the mean parameterization will be used throughout this paper. Let \( \mu_0 \in \Omega \) and \( m > 0 \), the conjugate prior distribution on \( \mu \) is:

\[
\pi(\mu) = K_0 \exp \{m \mu_0 \theta(\mu) - m M(\theta(\mu))\} V^{-1}(\mu),
\]

where \( K_0 \) is chosen to make \( \int_{\Omega} \pi(\mu) d\mu = 1 \) and \( \mu_0 \) is the prior mean.

3 Bayesian analysis

Let \( x_1, x_2, \ldots, x_n \) be a random sample drawn from density (1), then the likelihood function parameterized in terms of the mean is:

\[
l(\mu|x) = \exp \{n \pi \theta(\mu) - n M(\theta(\mu))\},
\]

with \( \overline{x} = n^{-1} \sum_{i=1}^{n} x_i \).
3.1 Prior distributions

Suppose that the prior information for \( \mu \) is provided by \( k \) experts as conjugate prior distributions:

\[
\pi_j(\mu) = K_{0j} \exp \{ \mu_0 m_j \theta(\mu) - m_j M(\theta(\mu)) \} V^{-1}(\mu), \ j = 1, 2, \ldots, k.
\]

The prior distributions can be mixed by different methods to form a combined prior distribution (see, e.g., Genest and Zidek (1986)). One of these methods is to use a mixture of prior distributions:

\[
\pi(\mu|\omega) = \sum_{j=1}^{k} \omega_j \pi_j(\mu),
\]

where \( \omega_j, j = 1, 2, \ldots, k \), are the mixture weights, which are non-negative and sum to unity.

Chen and Pennock (2005) observed that the weight selection is a possible inconvenience. Sometimes, the weights are chosen to reflect the relative importance of each expert. Here, the weight vector is considered as a random vector and a hyperprior distribution is proposed. The joint prior distribution for the parameters is expressed as:

\[
\pi(\omega, \mu) = \pi(\omega)\pi(\mu|\omega),
\]

where the weight vector is distributed as a Dirichlet \((\delta_1, \delta_2, \ldots, \delta_k)\) on the simplex \( \chi = \{ (\omega_1, \omega_2, \ldots, \omega_k) : \omega_j \geq 0, \sum_{j=1}^{k} \omega_j = 1 \} \), and a mixture of conjugate prior distributions is considered for \( \pi(\mu|\omega) \). Therefore:

\[
\pi(\omega, \mu) = \pi(\omega) \pi(\mu|\omega) = \left( z(\delta) \omega_1^{\delta_1-1} \cdots \omega_k^{\delta_k-1} \right) \left( \sum_{j=1}^{k} \omega_j \pi_j(\mu) \right),
\]

where \( z(\delta) = \Gamma(\sum_{l=1}^{k} \delta_l) / \prod_{l=1}^{k} \Gamma(\delta_l) \).

The hyperparameter values \( \delta_1, \delta_2, \ldots, \delta_k \) are chosen to assure that no expert has more prior influence than the others on the joint prior distribution \( \pi(\omega, \mu) \). This problem is solved in two steps. Firstly, the normalized vector, \( \delta^* = (\delta_1^*, \delta_2^*, \ldots, \delta_k^*) \) with \( \delta_j^* = \delta_j / \sum_{l=1}^{k} \delta_l, j = 1, 2, \ldots, k \), is obtained by
using the expected Kullback-Leibler divergence between the combined prior distribution \( \pi(\mu|\omega) \) and the component prior distribution \( \pi_l \) (see, e.g., Sun and Berger (1998) for the use of expected Kullback-Leibler divergence in a reference prior framework). Next, the values for \( \delta_1, \delta_2, \ldots, \delta_k \) are calculated by maximizing the resultant entropy.

For the first step, it is satisfied:

\[
E_\omega(KL(\pi||\pi_l)) = E_\omega \left( \int \pi(\mu|\omega) \log \left( \frac{\pi(\mu|\omega)}{\pi_l(\mu)} \right) \, d\mu \right) = \\
E_\omega \left( E_{\mu|\omega}(\log \pi(\mu|\omega)) \right) - E_\omega(E_{\mu|\omega}(\log \pi_l(\mu))),
\]

where \( E_\omega \) and \( E_{\mu|\omega} \) denote the expectations with respect to \( \pi(\omega) \) and \( \pi(\mu|\omega) \), respectively. The objective is to find \( \delta_1^*, \delta_2^*, \ldots, \delta_k^* \), such that:

\[
E_\omega(KL(\pi||\pi_1)) = E_\omega(KL(\pi||\pi_2)) = \ldots = E_\omega(KL(\pi||\pi_k)),
\]

(2)

with the constrain \( \sum_{j=1}^k \delta_j^* = 1 \) and \( \delta_j^* \geq 0 \). Therefore, the value of the expected discrepancy between the combined prior distribution and the prior distribution elicited by each expert, \( \pi_l(\mu) \), is the same for \( l = 1, 2, \ldots, k \).

The parameter values satisfying the previous equalities are the same that hold:

\[
E_\omega(E_{\mu|\omega}(\log \pi_h(\mu))) - E_\omega(E_{\mu|\omega}(\log \pi_1(\mu))) = 0, \text{ for } h = 2, 3, \ldots, k,
\]

(3)

with the same constraints. The previous addends satisfy (see Appendix A):

\[
E_\omega(E_{\mu|\omega}(\log \pi_l(\mu))) = \int \int \pi(\mu|\omega) \log \pi_l(\mu) d\mu \pi(\omega) d\omega = \\
= \sum_{j=1}^k \delta_j^* E_{\pi_j}(\log \pi_l(\mu)),
\]

where the expectation with respect to the experts’ prior distributions can be expressed as:

\[
E_{\pi_j}(\log \pi_l(\mu)) = \log K_{0l} + m_l \mu_{0l} E_{\pi_j}(\theta(\mu)) - m_l E_{\pi_j}(M(\theta(\mu))) + E_{\pi_j}(\log V^{-1}(\mu)).
\]

By taking into account the previous expressions, the solution for the normal-
ized vector, $\mathbf{\delta}^*$, can be obtained from the linear equation system given by:

$$
\sum_{j=1}^{k} \delta_j^* \left[ E_{\pi_j}(\theta(\mu))(m_h\mu_{0h} - m_1\mu_{01}) + E_{\pi_j}(M(\theta(\mu)))(m_1 - m_h) \right] = \\
\log K_{01} - \log K_{0h}, \ h = 2, 3, \ldots, k, \tag{4}
$$

with the constraints $\delta_j^* \geq 0$.

Observe that the normalized vector $\mathbf{\delta}^* = (\delta_1^*, \delta_2^*, \ldots, \delta_k^*)$ is considered because the equalities for the expected Kullback-Leibler divergences given in (3), are satisfied for any Dirichlet distribution with hyperparameters $(\delta_1, \delta_2, \ldots, \delta_k)$ such that $\delta_j/\sum_{l=1}^{k} \delta_l = \delta_j^*, j = 1, 2, \ldots, k$. In addition, the expectations $E_{\pi_j}(\theta(\mu))$ and $E_{\pi_j}(M(\theta(\mu)))$ can be analytically obtained for each family of the class (see Table 1).

Once the vector $\mathbf{\delta}^* = (\delta_1^*, \delta_2^*, \ldots, \delta_k^*)$ has been calculated, the hyperparameters for the Dirichlet distribution, $\delta_1, \delta_2, \ldots, \delta_k$, are obtained by maximizing the entropy:

$$
H(\pi(\omega)) = -E(\log(\pi(\omega))) \tag{5}
$$

with $\delta_j = \lambda^{\delta_j^*}, j = 1, 2, \ldots, k, \lambda > 0$. Thus, the least informative Dirichlet distribution, in the sense that it contains the least amount of information consistent with the given information, is obtained. Observe that, by using this procedure, the hyperparameters $\delta_1, \delta_2, \ldots, \delta_k$ also satisfy the equalities given in (2).

The objective is to find $\lambda > 0$ maximizing:

$$
H(\lambda) = -\log \Gamma\left(\sum_{l=1}^{k} \lambda \delta_l^*\right) + \sum_{l=1}^{k} \log \Gamma(\lambda \delta_l^*) - \\
- \sum_{j=1}^{k} \left[ (\lambda \delta_j^* - 1)(\Psi(\lambda \delta_j^*) - \Psi(\sum_{l=1}^{k} \lambda \delta_l^*)) \right],
$$

where $\Psi(\cdot) = \Gamma'(\cdot)/\Gamma(\cdot)$ represents the digamma function. Note that this is a one-dimensional continuous optimization problem and the solution can be easily obtained.
Families

<table>
<thead>
<tr>
<th>Families</th>
<th>$E_{\pi_j}(\theta(\mu))$</th>
<th>$E_{\pi_j}(M(\theta(\mu)))$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson $P(\mu)$</td>
<td>$\Psi(a) - \log(m_j)$</td>
<td>$\mu_{0j}$</td>
</tr>
<tr>
<td>Binomial $B(r,p)$</td>
<td>$\Psi(a) - \Psi(b-a)$</td>
<td>$r(\Psi(b) - \Psi(b-a))$</td>
</tr>
<tr>
<td>Negative-Binomial $NB(r,p)$</td>
<td>$\Psi(a) - \Psi(a+b+1)$</td>
<td>$-r(\Psi(b) - \Psi(a+b+1))$</td>
</tr>
<tr>
<td>Gamma $G(r,\lambda)$</td>
<td>$-\frac{(m_j+r+1)}{m_j\mu_{0j}}$</td>
<td>$-r(\log r + \Psi(b+1)+\log(ar))$</td>
</tr>
<tr>
<td>Normal $N(\mu,\sigma^2)$</td>
<td>$\frac{\mu_{0j}}{\sigma^2}$</td>
<td>$\frac{1}{2\sigma^2}\left(\frac{\sigma^2}{m_j} + \mu_{0j}^2\right)$</td>
</tr>
<tr>
<td>Hyperbolic-Secant $HS(\lambda)$</td>
<td>$-$</td>
<td>$\frac{1}{2}\left(\Psi\left(\frac{m_j+2}{2}\right) - \Psi\left(\frac{m_j+1}{2}\right)\right)$</td>
</tr>
</tbody>
</table>

Table 1

Expectations for NEF-QVF distributions ($a = m_j\mu_{0j}$, $b = m_jr$ for $j = 1, 2, \ldots, k$; $m_j > 0$ and $\mu_{0j} \in \Omega$ denote the prior parameters; $\Psi(\cdot)$ is the digamma function)

When $\delta_1^* = \delta_2^* = \ldots = \delta^*_k$, the optimal solution is achieved at $\lambda = k$, and therefore the distribution is the uniform one on the corresponding simplex.

Note that, the linear equation system given in (4) may have no solution satisfying the restrictions $\delta^*_j \geq 0$, $j = 1, 2, \ldots, k$. Hence, an alternative process should be used to obtain the hyperparameter values. An extension of the previous proposal is included in Appendix B.

3.2 Posterior distribution

By applying Bayes’ theorem, it follows that:

$$
\pi(\omega, \mu|x) = \frac{l(\mu|x) \pi(\omega) \pi(\mu|\omega)}{\int_{\Omega} \int_{\Omega} l(\mu|x) \pi(\omega) \pi(\mu|\omega) d\mu d\omega} = \sum_{j=1}^{k} \omega_j^* \pi_j(\mu|x),
$$

where $\pi_j(\mu|x)$ are the posterior distributions for each component (see, e.g., Bernardo and Smith (1994)), i.e.:

$$
\pi_j(\mu|x) = K_{1j} \exp\{(n\bar{x} + m_j\mu_{0j})\theta(\mu) - (n + m_j)M(\theta(\mu))\} V^{-1}(\mu),
$$

with $K_{1j}$ the normalizing constants for these distributions.
The updated weights become:

$$\omega_j^* = \frac{z(\delta) K_{0j} (\omega_j \prod_{l=1}^{k} \omega_l^{\delta_l-1})}{K_{1j} \left( \sum_{j=1}^{k} \delta_j K_{0j} \right) \left( \sum_{l=1}^{k} \delta_l K_{1j} \right)}^{-1}.$$  \hspace{1cm} (6)

In order to make inferences, the posterior distribution is factorized as:

$$\pi(\omega, \mu|x) = \pi(\omega|x) \pi(\mu|\omega, x).$$  \hspace{1cm} (7)

The conditional distribution for the mean is given by:

$$\pi(\mu|\omega, x) = \sum_{j=1}^{k} \omega_j K_{0j} K_{1j}^{-1} \pi_j(\mu|x) \left( \sum_{j=1}^{k} \omega_j K_{0j} K_{1j}^{-1} \right)^{-1},$$

thus, this conditional distribution is a mixture of the posterior distributions for each expert, \(\pi_j(\mu|x)\), and weights \(\omega_j K_{0j} K_{1j}^{-1} \left( \sum_{j=1}^{k} \omega_j K_{0j} K_{1j}^{-1} \right)^{-1}\).

The marginal posterior distribution of the weight vector can be expressed as:

$$\pi(\omega|x) = \sum_{j=1}^{k} \delta_j K_{0j} K_{1j}^{-1} \omega_j K_{0j} K_{1j}^{-1} \left( \sum_{j=1}^{k} \delta_j K_{0j} K_{1j}^{-1} \right)^{-1} \left( z(\delta) \omega_j \prod_{l=1}^{k} \omega_l^{\delta_l-1} \right) \left( \sum_{j=1}^{k} \delta_j K_{0j} K_{1j}^{-1} \right)^{-1},$$

with \(z(\delta) = \Gamma(k) \prod_{l=1}^{k} \Gamma(\delta_l)\). Note that this posterior distribution is a mixture of Dirichlet distributions with parameters \((\delta_1 + 1, \delta_2, \ldots, \delta_k), (\delta_1, \delta_2 + 1, \ldots, \delta_k), \ldots, (\delta_1, \delta_2, \ldots, \delta_k + 1)\) and weights \(\delta_j K_{0j} K_{1j}^{-1} \left( \sum_{j=1}^{k} \delta_j K_{0j} K_{1j}^{-1} \right)^{-1}\).

In order to generate random variates from \(\pi(\mu|\omega, x)\) and \(\pi(\omega|x)\), the composition method can be used (see, e.g., Devroye (1986)). Then, random samples from the posterior distribution can be obtained by using the relationship given in (7).

4 Calibration within the experts group

In some works dealing with expert opinions, the calibration focuses on the problem of how well the experts know their subject matter, i.e., the calibration measures the statistical likelihood that actual experimental results correspond
with the experts’ assessments. See, for example, Booker et al. (1993) and Roest (2002).

In this section, the calibration problem is addressed under a different point of view. The prior opinion of each expert is calibrated by comparing his/her prior belief with the posterior belief of the group. In order to do this, the expected discrepancy between the conditional distribution for $\mu$ and the prior distribution elicited by each expert is analyzed by using the expected Kullback-Leibler divergence between the mixture of the experts’ posterior distributions, $\pi(\mu|\omega, x)$, and the prior distribution of each one of them. Thus, the smaller the expected discrepancy, the closer the expert’s prior opinion is to the group opinion. Some related works consider a posterior distribution involving a reference prior for an expert’s calibration (see Bousquet (2006, 2008) and references therein).

By considering the factorized posterior distribution as in (7), the expected Kullback-Leibler divergence between the distribution $\pi(\mu|\omega, x)$ and the prior distribution $\pi_i(\mu)$ is:

\[
EKL_{\pi_i} = E_{\omega|x}(KL(\pi||\pi_i)) = E_{\omega|x} \left[ E_{\mu|\omega,x} \left( \log \frac{\pi(\mu|\omega,x)}{\pi_i(\mu)} \right) \right] = E_{\omega,x} \left( \log \frac{\pi(\mu|\omega,x)}{\pi_i(\mu)} \right) \tag{8}
\]

where $E_{\omega|x}$, $E_{\mu|\omega,x}$ and $E_{\mu,\omega|x}$ are the expectations with respect to the distributions $\pi(\omega|x)$, $\pi(\mu|\omega,x)$ and $\pi(\mu,\omega|x)$, respectively.

In order to estimate the expectation given in expression (8), a straightforward Monte Carlo-based approach can be performed. The corresponding Monte Carlo estimate is:

\[
\tilde{EKL}_{\pi_i} = \frac{1}{T} \sum_{t=1}^{T} \log \frac{\pi(\mu^{(t)}|\omega^{(t)}, x)}{\pi_i(\mu^{(t)})},
\]

where the vectors $(\omega^{(1)}, \mu^{(1)}), (\omega^{(2)}, \mu^{(2)}), \ldots, (\omega^{(T)}, \mu^{(T)})$ are generated from the posterior distribution $\pi(\omega, \mu|x)$, being $T$ the sample size. Note that $\tilde{EKL}_{\pi_1}, \tilde{EKL}_{\pi_2}, \ldots, \tilde{EKL}_{\pi_k}$ are obtained by using common random variates.
from $\pi(\omega, \mu|x)$. Thus, a computationally low-cost method is obtained. The Monte Carlo standard error estimate can be easily calculated.

By using this Monte Carlo-based approach, the expected discrepancies between the combined posterior distribution over the quantity of interest and the prior distribution elicited by each expert are estimated. This information allows to know the agreement degree between each expert’s opinion and the posterior group opinion over the quantity of interest. Subsequently, it can be used by the decision maker in the most appropriate way for his purpose (see, for example Valsecchi (2008)).

5 A binomial application

Gajewski and Mayo (2006) considered practical situations from a Bayesian perspective when there are two sources of information. A Binomial random variable $X$ with number of trials $r$ and probability of success $p$ is observed. The sources of information may come from conflicting clinical opinion or two previous trials with different results. The first source of information is optimistic and the second source of information is pessimistic towards drug success. Then, both informations are incorporated into the prior distribution by using a mixture of two Beta distributions.

Gajewski and Mayo (2006) indicated that a natural extension to their work would be to place a distribution on the mixing parameters. This is considered in this example. The application is based on a recent phase II clinical trial sought to determine the response to gemcitabine plus docetaxel among patients with leiomyosarcoma. Complete or partial response was observed in $x = 18$ patients out of $r = 34$. Then, the number of patients with positive response to the treatment, $X$, has a binomial distribution with parameters $r$ and $p$. Denoting by $\mu$ the mean number of patients with positive response to the treatment, the prior distribution for $\mu$ is then bimodal. It is modeled through a mixture of generalized beta prior distributions on $(0, r)$, with parameters
Thus, its expression is:

\[
\pi(\mu|\omega) = \sum_{j=1}^{2} \omega_j K_{0j} \exp\{m_j\mu_0j \log \left( \frac{\mu}{r - \mu} \right) - (m_j r - 1) \log \left( \frac{r}{r - \mu} \right) - \log \mu \},
\]

with unknown weights and such that \(\omega_1 + \omega_2 = 1\), \(\omega_j \geq 0\), \(j = 1, 2\). Additional information about the normalizing constants \(K_{0j}\) for NEF-QVF distributions is provided, for example, in Rufo et al. (2009).

The hyperprior distribution for the vector composed of the weights, \(\omega\), is a Dirichlet one with parameters \(\delta_1\) and \(\delta_2\). Therefore, the joint prior distribution is:

\[
\pi(\omega, \mu) = \pi(\omega)\pi(\mu|\omega).
\]

The parameters for the component prior distributions in the combined prior distribution (9) are the ones proposed by Gajewski and Mayo (2006), i.e.:

\[
m_1 = m_2 = 0.3823, \mu_{01} = 28.51, \mu_{02} = 5.5.
\]

In order to calculate the hyperparameters \(\delta_1\) and \(\delta_2\), firstly the linear equation system given in (4) is solved. By taking into account the expectations in Table 1 and substituting the parameter values, the following linear equation system is obtained:

\[
\begin{align*}
\alpha \delta_1^* - \alpha \delta_2^* &= 0 \\
\delta_1^* + \delta_2^* &= 1
\end{align*}
\]

constrained to \(\delta_j^* \geq 0\), \(j = 1, 2\) with \(\alpha = 8.8 (\Psi(2.1) - \Psi(10.9))\). The solution is \((\delta_1^*, \delta_2^*) = (0.5, 0.5)\). Now, the objective is to find \(\lambda > 0\) maximizing the function:

\[
H(\lambda) = -\log \Gamma(\lambda) + 2 \log \Gamma(0.5\lambda) - 2 [(0.5\lambda - 1)(\Psi(0.5\lambda) - \Psi(\lambda))] .
\]

It is obtained \(\lambda = 2\) and \((\delta_1, \delta_2) = (0.5\lambda, 0.5\lambda) = (1, 1)\). Note that this distribution is the uniform one on the simplex \(\chi\).

The posterior distribution is a mixture of generalized beta distributions with parameters \(n\bar{x} + m_j\mu_{0j}\) and \(m_j r + nr - n\bar{x} - m_j\mu_{0j}\), i.e.:
\[
\pi(\omega, \mu|x) = \sum_{j=1}^{2} \omega_j \pi_j(\mu|x) = \sum_{j=1}^{2} \omega_j K_{1j} \exp\{(n\bar{\pi} + m_j\mu_0) \log \left( \frac{\mu}{r - \mu} \right) - (nr + m_jr - 1) \log \left( \frac{r}{r - \mu} \right) - \log \mu \},
\]

where \(\omega_j^*\) is given by expression (6), i.e.:

\[
\omega_j^* = \frac{2\omega_j K_{0j}}{K_{1j}} \left( \sum_{j=1}^{k} \frac{K_{0j}}{K_{1j}} \right)^{-1}.
\]

Next, expected Kullback-Leibler divergences between the mixture of the experts’ posterior distributions, \(\pi(\mu|\omega, x)\), and each component prior distribution, \(\pi_l(\mu)\), \(l = 1, 2\), are analyzed. Following the development in Section 4, a random sample of size 10000 is drawn from the posterior distribution \(\pi(\omega, \mu|x)\) to estimate the expected Kullback-Leibler divergences. Figure 1 shows the histogram of the generated means \(\mu^{(t)}\), \(t = 1, 2, \ldots, 10000\), and the prior distributions elicited by each source.

![Histogram of the generated posterior means with the prior distributions \(\pi_1(\mu)\) (solid line) and \(\pi_2(\mu)\) (dotted line).](image)

The Monte Carlo estimates for the expected Kullback-Leibler divergences are \(\overline{EKL}_{\pi_1} = 2.8757\) and \(\overline{EKL}_{\pi_2} = 4.8249\), for the first and second source, respectively. Thus, the optimistic source provides prior information with more group opinion agreement given the data than the pessimistic source, since \(\overline{EKL}_{\pi_1}\)
is smaller than $\overline{EKL}_{\pi_2}$. The corresponding Monte Carlo error estimates are $\overline{se}(\overline{EKL}_{\pi_1}) = 0.0181$ and $\overline{se}(\overline{EKL}_{\pi_2}) = 0.0245$, respectively.

Next, it is examined which of the two sources is closer to the group opinion over the quantity of interest by considering the number of patients enrolled in the phase II trial. Thus, it could be reasonable to state when the treatment was or not successful. In order to do it, the previous expected Kullback-Leibler distances are also estimated for the remaining values of $x$. Figure 2 represents the Monte Carlo estimates, being $x = 0, 1, \ldots, 34$, the number of patients with response to the treatment.

![Fig. 2. Estimated expected divergences $\overline{EKL}_{\pi_1}$ (solid line) and $\overline{EKL}_{\pi_2}$ (dotted line).](image)

For $x = 0, 1, \ldots, 16$, the pessimistic source provides prior information with more group opinion agreement given the data than the optimistic source. The opposite happens when $x = 18, 19, \ldots, 34$. Note that, for $x = 17$, both sources have the same influence on the posterior decision over the quantity of interest ($\overline{EKL}_{\pi_1} = 3.7816$ and $\overline{EKL}_{\pi_2} = 3.7459$).
6 Conclusion

This paper presents a unified Bayesian framework for NEF-QVF distributions by using prior informations provided by several experts. This class of families includes many of the most commonly used distributions in statistical modeling. Mixture models with unknown weights are considered to combine the experts’ prior information. An advantage when using a hyperprior distribution on the weights is that the modeling of the available prior information is more flexible. Then, a general method to obtain the hyperparameters is proposed. The main advantage is that the procedure is analytical and general expressions allowing a direct implementation are obtained. Finally, a straightforward Monte Carlo-based approach is considered to estimate the expected discrepancies between the combined posterior distribution over the quantity of interest and the prior distribution elicited by each expert.

Moreover, it is important to observe that the general proposal to obtain the hyperparameter values is valid independently of the considered distribution. If the expectations can be analytically calculated, then a similar scheme can be followed. In a different case, simulation techniques could be used in an appropriate way, in order to obtain the required expectations to apply the proposal. The considered approach for the calibration problem could be applied for any distribution.

Acknowledgements

The authors thank the Associate Editor and the Reviewers 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 (Projects TSI2007-66706-C04-02 and TIN2008-06796-C04-03) and Junta de Extremadura (Project GRU09046).
A Theoretical results

This appendix shows that the following equality holds,

\[ E_\omega(E_{\mu|\omega}(\log \pi_l(\mu))) = \sum_{j=1}^{k} \delta_j^* E_{\pi_j}(\log \pi_l(\mu)), \]

where \( E_\omega(\cdot) \) denotes the expectation with respect to the Dirichlet prior distribution:

\[
\pi(\omega) = \frac{\Gamma(\sum_{l=1}^{k} \delta_l)}{\prod_{l=1}^{k} \Gamma(\delta_l)} \omega^{\delta_1 \ldots \delta_k} = z(\delta) \omega^{\delta_1 \ldots \delta_k},
\]

and \( E_{\mu|\omega}(\cdot) \) is the expectation with respect to the combined prior distribution:

\[ \pi(\mu|\omega) = \sum_{j=1}^{k} \omega_j \pi_j(\mu). \]

For any component prior distribution \( \pi_l(\mu) \), it is satisfied:

\[
E_\omega(E_{\mu|\omega}(\log \pi_l(\mu))) = \int \int_{\Omega} \pi(\mu|\omega) \log \pi_l(\mu) d\mu \pi(\omega) d\omega = \]
\[
= \int \int_{\Omega} \sum_{j=1}^{k} \omega_j \pi_j(\mu) \log \pi_l(\mu) d\mu \pi(\omega) d\omega = \]
\[
= \sum_{j=1}^{k} \int \int_{\Omega} \omega_j \pi_j(\mu) \log \pi_l(\mu) d\mu \pi(\omega) d\omega = \]
\[
= \sum_{j=1}^{k} \int_{\chi} \omega_j E_{\pi_j}(\log \pi_l(\mu)) \pi(\omega) d\omega = \]
\[
= \sum_{j=1}^{k} E_{\pi_j}(\log \pi_l(\mu)) z(\delta) \int_{\chi} \omega_j \omega_1^{\delta_1 \ldots} \omega_k^{\delta_k \ldots} d\omega.
\]

Now, for each component \( j = 1, 2, \ldots, k \), the last integral in the previous expression corresponds to a Dirichlet distributions, i.e.:

\[
\int_{\chi} \omega_j \omega_1^{\delta_1 \ldots} \omega_k^{\delta_k \ldots} d\omega = \frac{\Gamma(\delta_j + 1) \prod_{l=1, l\neq j}^{k} \Gamma(\delta_l)}{\Gamma(\sum_{l=1}^{k} \delta_l + 1)},
\]
where $\Gamma(\cdot)$ denotes the gamma function. Then, by taking into account the value of $z(\delta)$ and the general equality $\Gamma(a + 1) = a \Gamma(a)$, it is obtained:

$$z(\delta) \int x^j \omega_1^{\delta_1} \ldots \omega_k^{\delta_k} d\omega = \frac{\delta_j}{\sum_{l=1}^k \delta_l} = \delta_j^*,$$

and thus,

$$E_{\omega}(E_{\mu|\omega}(\log \pi_l(\mu))) = \sum_{j=1}^k \delta_j^* E_{\pi_j}(\log \pi_l(\mu)).$$

### B General hyperparameter choice method

A problem could arise when calculating the normalized hyperparameters, i.e., $\delta^* = (\delta_1^*, \delta_2^*, \ldots, \delta_k^*)$ might not belong to $S_k = \{(\delta_1^*, \delta_2^*, \ldots, \delta_k^*): \sum_{j=1}^k \delta_j^* = 1, \delta_j^* \geq 0, j = 1, 2, \ldots, k\}$.

In order to solve this problem, a new procedure, based on the idea presented in Subsection 3.1, is proposed. When it is not possible to find nonnegative real values: $\delta_1^*, \delta_2^*, \ldots, \delta_k^*$, such that the expected discrepancies satisfy:

$$E(KL(\pi||\pi_1)) - E(KL(\pi||\pi_h)) = 0$$

for $h = 2, 3, \ldots, k$,

the purpose is that the maximum difference among experts (regarding the expected Kullback-Leibler divergence between the combined prior distribution and each prior one) is the smallest possible. Therefore, the objective will be to find $\delta^* = (\delta_1^*, \delta_2^*, \ldots, \delta_k^*) \in S_k$, minimizing $u$ and satisfying:

$$|E(KL(\pi||\pi_l)) - E(KL(\pi||\pi_h))| \leq u$$

for $1 \leq l < h \leq k$,

where $u$ is a nonnegative real number.

Solving the previous problem is equivalent to solving the linear programming problem:

$$\min u$$
s. t.

\[ \sum_{j=1}^{k} \delta_j^* \left[ E_{x_j}(\theta (\mu))(m_{k\mu_{0h}} - m_{l\mu_{0l}}) + E_{x_j}(M(\theta (\mu)))(m_l - m_h) \right] - u \leq \log K_{0l} - \log K_{0h}, \]

\[ - \sum_{j=1}^{k} \delta_j^* \left[ E_{x_j}(\theta (\mu))(m_{k\mu_{0h}} - m_{l\mu_{0l}}) + E_{x_j}(M(\theta (\mu)))(m_l - m_h) \right] - u \leq \log K_{0h} - \log K_{0l}, \]

\[ \sum_{j=1}^{k} \delta_j^* = 1, \quad \delta_j^* \geq 0, \]

for \( 1 \leq l < h \leq k \).

Note that, when the optimal value is \( u = 0 \), the solution obtained is the same as the one obtained by solving the linear equation system presented in Subsection 3.1. Next, after maximizing the function given in (5), the hyperparameters \( \delta_1, \delta_2, \ldots, \delta_k \) are obtained.

References


