European Journal of Operational Research 207 (2010) 284–289

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

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

ARTICLE INFO

Article history:
Received 15 December 2008
Accepted 7 April 2010
Available online 13 April 2010

Keywords:
Bayesian analysis
Conjugate prior distributions
Exponential families
Prior mixtures
Kullback–Leibler divergence

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.

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.

* Corresponding author. Address: Departamento de Matemáticas, Escuela Politécnica, Universidad de Extremadura, Avda. de la Universidad s/n, 10071 Cáceres, Spain. Tel.: +34 927257220; fax: +34 927257203.

E-mail addresses: mrufo@unex.es (M.J. Rufo), carper@unex.es (C.J. Pérez), jrmartin@unex.es (J. Martín).

0377-2217/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2010.04.005

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 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:

$$f_\theta(x) = \exp\{x\theta - M(\theta)\}, \quad \theta \in \Theta, \tag{1}$$

where $M(\theta) = \log \int \exp(x\theta)\eta(dx)$ and $\Theta = \{\theta \in \mathbb{R} : M(\theta) < \infty\}$ is nonempty. θ 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 Ω 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 μ is:

$$\pi(\mu) = K_0 \exp\{m\mu_0\theta(\mu) - mM(\theta(\mu))\}V^{-1}(\mu),$$

where K_0 is chosen to make $\int_\Omega \pi(\mu)d\mu = 1$ and μ_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:

$$l(\mu|\mathbf{x}) = \exp\{n\bar{x}\theta(\mu) - nM(\theta(\mu))\},$$

with $\bar{x} = n^{-1}\sum_{i=1}^n x_i$.

3.1. Prior distributions

Suppose that the prior information for μ is provided by k experts as conjugate prior distributions:

$$\pi_j(\mu) = K_{0j} \exp\{\mu_{0j}m_j\theta(\mu) - m_jM(\theta(\mu))\}V^{-1}(\mu), \quad j = 1, 2, \ldots, k.$$

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:

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

where ω_j, $j = 1, 2, \ldots, k$, are the mixture weights, which are nonnegative 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:

$$\pi(\boldsymbol{\omega}, \mu) = \pi(\boldsymbol{\omega})\pi(\mu|\boldsymbol{\omega}),$$

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

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

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

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 $\pi(\boldsymbol{\omega}, \mu)$ This problem is solved in two steps. Firstly, the normalized vector, $\boldsymbol{\delta}^* = (\delta_1^*, \delta_2^*, \ldots, \delta_k^*)$ 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 $\pi(\mu|\boldsymbol{\omega})$ and the component prior distribution π_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:

$$E_{\boldsymbol{\omega}}(KL(\pi\|\pi_l)) = E_{\boldsymbol{\omega}}\left(\int_\Omega \pi(\mu|\boldsymbol{\omega}) \log\left(\frac{\pi(\mu|\boldsymbol{\omega})}{\pi_l(\mu)}\right)d\mu\right) = E_{\boldsymbol{\omega}}(E_{\mu|\boldsymbol{\omega}}(\log \pi(\mu|\boldsymbol{\omega}))) - E_{\boldsymbol{\omega}}(E_{\mu|\boldsymbol{\omega}}(\log \pi_l(\mu))),$$

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

$$E_{\boldsymbol{\omega}}(KL(\pi\|\pi_1)) = E_{\boldsymbol{\omega}}(KL(\pi\|\pi_2)) = \cdots = E_{\boldsymbol{\omega}}(KL(\pi\|\pi_k)), \tag{2}$$

with the constrains $\sum_{j=1}^k \delta_j^* = 1$ and $\delta_j^* \geqslant 0$.

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:

$$E_{\boldsymbol{\omega}}(E_{\mu|\boldsymbol{\omega}}(\log \pi_h(\mu))) - E_{\boldsymbol{\omega}}(E_{\mu|\boldsymbol{\omega}}(\log \pi_1(\mu))) = 0, \quad \text{for } h = 2, 3, \ldots, k, \tag{3}$$

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

$$E_{\boldsymbol{\omega}}(E_{\mu|\boldsymbol{\omega}}(\log \pi_l(\mu))) = \int_\chi \int_\Omega \pi(\mu|\boldsymbol{\omega}) \log \pi_l(\mu) d\mu \pi(\boldsymbol{\omega}) d\boldsymbol{\omega} = \sum_{j=1}^k \delta_j^* E_{\pi_j}(\log \pi_l(\mu)),$$

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

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

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