Local sensitivity analysis for Bayesian mixture models

Department of Mathematics

M. J. Rufo, C. J. Pérez and J. Martín

University of Extremadura, Spain

E-mails: mrufo@unex.es, carper@unex.es, jrmartin@unex.es

1. Introduction

- Mixture models are receiving considerable significance in the last years (see, for example McLachlan and Peel [2] and Böhning and Seidel [1]). They help in dealing with practical situations in reliability and survival analysis.
- Rufo et al. [5] provided a general approach to address Bayesian analysis of finite mixture models of distributions from natural exponential families with quadratic variance function (NEF-QVF). These families contain distributions such as normal, gamma, Poisson, binomial distributions,... Exponential and Bernoulli distributions are special cases.
- We propose a general method to estimate local prior sensitivities in these mixture models. An advantage of this method is that it allows a direct implementation of the estimates of the sensitivity measures and their errors. Besides, the generated samples to estimate the parameters in the mixture model are re-used to perform estimates of the sensitivity measures and their errors, as in Pérez et al. [4].

2. Background

2.1. *Natural exponential families*. The parametric family of probability measures $\{P_\theta\}$ is a natural exponential family (NEF) if its density with respect to η satisfies:

$$dP_\theta(x) = \exp\{x\theta - M(\theta)\}\, d\eta(x),\ \theta \in \Theta,$$

where $M(\theta) = \log \int e^{x\theta} d\eta(x)$ and $\Theta = \{\theta \in \mathbb{R} : M(\theta) < \infty\} \neq \emptyset$.

The mapping $\mu = \mu(\theta) = M'(\theta)$ is differentiable, with inverse $\theta = \theta(\mu)$. It provides an alternative parameterization for $\{P_\theta\}$ called mean parameterization. If this parameterization is used then the density with respect to the measure η is given by:

$$p(x|\mu) = \exp\{x\theta(\mu) - M(\theta(\mu))\},\ \mu \in \Omega,$$

where $\Omega = \mu(\Theta)$ is the mean parameter space.

A NEF has quadratic variance function (QVF) when $V(\mu) = v_0 + v_1\mu + v_2\mu^2$ where v_0, v_1 and v_2 are real constants.

2.2. *Conjugate prior distribution*. If the mean parameterization is used, then the conjugate prior distribution for μ is (see Morris [3]):

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

where $\mu_0 \in \Omega$, $m > 0$ and $K(m, \mu_0)$ is chosen to make $\int_\Omega \pi(\mu)\, d\mu = 1$.

3. Estimating local parametric sensitivity

- Suppose the interest is focused on the estimation of the posterior expectation for $f(\theta)$. If local parametric changes in the prior distribution, $\pi_\lambda(\theta)$, are considered, then it is defined:

$$\mathcal{I}_\lambda = \int_\Theta f(\theta)\, p_\lambda(\theta|x)\, d\theta = \frac{\int_\Theta f(\theta) l(\theta|x)\pi_\lambda(\theta) d\theta}{\int_\Theta l(\theta|x)\pi_\lambda(\theta) d\theta}$$

where λ is a multidimensional parameter. Let $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(N)}$ be a sample generated from $p_{\lambda^0}(\theta|x)$, where λ^0 is a fixed quantity. Then, an estimate of $\mathcal{I}_{\lambda^0}$ is given by: $\widehat{\mathcal{I}_{\lambda^0}} = \frac{1}{N}\sum_{t=1}^N f\left(\theta^{(t)}\right)$.

- A local parametric sensitivity analysis is performed. As a local sensitivity measure, the gradient vector evaluated at λ^0 is considered, that is:

$$\nabla \mathcal{I}_{\lambda^0} = \left(\partial_{\lambda_1}\mathcal{I}_{\lambda^0}, \partial_{\lambda_2}\mathcal{I}_{\lambda^0}, \dots, \partial_{\lambda_m}\mathcal{I}_{\lambda^0}\right).$$

- It is found that (see Pérez et al. [4]): $\partial_{\lambda_j}\mathcal{I}_{\lambda^0} = E_{p_{\lambda^0}}\left[(f(\theta) - \mathcal{I}_{\lambda^0})\, \partial_{\lambda_j} \log \pi_\lambda(\theta)\Big|_{\lambda=\lambda^0}\right]$.
- Then, the estimate of $\partial_{\lambda_j}\mathcal{I}_{\lambda^0}$ is given by:

$$\widehat{\partial_{\lambda_j}\mathcal{I}_{\lambda^0}} = \frac{1}{N}\sum_{t=1}^N (f(\theta^{(t)}) - \widehat{\mathcal{I}_{\lambda^0}})\partial_{\lambda_j} \log \pi_\lambda(\theta^{(t)})\Big|_{\lambda=\lambda^0}.$$

- The Monte Carlo standard error estimate is given by:

$$\widehat{SE}\left(\widehat{\partial_{\lambda_j}\mathcal{I}_{\lambda^0}}\right) = \sqrt{\frac{1}{N(N-1)}\sum_{t=1}^N ((f(\theta^{(t)}) - \widehat{\mathcal{I}_{\lambda^0}})\partial_{\lambda_j} \log \pi_\lambda(\theta^{(t)})\Big|_{\lambda=\lambda^0} - \widehat{\partial_{\lambda_j}\mathcal{I}_{\lambda^0}})^2}.$$

References

[1] Böhning, D. and Seidel, (2003) Editorial: recent developments in mixture models, *Computational Statistics and Data Analysis*, **41**, 349-357.

[2] McLachlan, G. and Peel, D. (2000) *Finite mixture models*, Wiley: New York.

[3] Morris C. N. (1983) Natural exponential families with quadratic variance fuctions: Statistical theory, *The Annals of Statistics*, **11**, 515-529.

[4] Pérez, C. J., Martín, J. and Rufo, M. J. (2006) Sensitivity estimations for Bayesian inference models solved by MCMC methods, *Reliability Engineering and System Safety*, **91**, 1310-1314.

[5] Rufo, M. J., Martín, J. and Pérez, C. J. (2006) Bayesian analysis of finite mixture models of distributions from exponential families, *Computational Statistics*, **21**, 621-637.

This research has been supported by *Ministerio de Educación y Ciencia*, Spain (Project TSI2004-06801-C04-03).

4. Application to NEF-QVF mixtures

4.1. *NEF-QVF mixtures*. Models in which data $x_1, x_2, \dots, x_n$ are assumed to be independent observations from a mixture density with k components are considered:

$$p(x|\boldsymbol{\omega}, \boldsymbol{\mu}) = \sum_{j=1}^k \omega_j p(x|\mu_j),$$

where ω_j are the mixture weights (which are restricted to be non-negative and sum to unity) and $p(x|\mu_j)$ are distributions from a NEF-QVF, with μ_j the mean for component j. The mean parameterization is used since general expressions for the estimates of the components in gradient vector will be obtained. This allows a direct implementation of the method proposed by Pérez et al. [4] for all the families of NEF-QVF.

The conjugate prior distribution for each μ_j, $\pi(\mu_j)$, depends on the parameters $m > 0$ and $\mu_0 \in \Omega$. Let $\lambda^0 = (m^0, \mu_0^0)$ be a fixed quantity. Local sensitivity with respect to the parameter $\lambda = (m, \mu_0)$ in the neighborhood of $\lambda^0 = (m^0, \mu_0^0)$ is considered.

4.2. *Estimate of the gradient vector*. Firstly, a sample from the posterior distribution of interest is generated. A useful method is Gibbs sampling by using latent allocations variables. The additional problem of the symmetry in the posterior distribution is solved by using the general algorithm proposed in Rufo et al. [5]. This approach is based on permutated sample points.

Next, the partial derivatives are calculated:

$$\partial_{\mu_0}\pi_\lambda(\mu_j) = (\partial_{\mu_0}K(m, \mu_0)) \exp\{m\mu_0\theta_j(\mu_j) - mM(\theta_j(\mu_j))\} V^{-1}(\mu_j) + \\ + K(m, \mu_0)\exp\{m\mu_0\theta_j(\mu_j) - mM(\theta_j(\mu_j))\} V^{-1}(\mu_j)(m\theta_j(\mu_j)),$$

$$\partial_m\pi_\lambda(\mu_j) = (\partial_m K(m, \mu_0)) \exp\{m\mu_0\theta_j(\mu_j) - mM(\theta_j(\mu_j))\} V^{-1}(\mu_j) + \\ + K(m, \mu_0)\exp\{m\mu_0\theta_j(\mu_j) - mM(\theta_j(\mu_j))\} V^{-1}(\mu_j)(\mu_0\theta_j(\mu_j) - M(\theta_j(\mu_j)))$$

where $\partial_{\mu_0}K(m, \mu_0)$ and $\partial_m K(m, \mu_0)$ are straightforward to calculate.

Therefore, the estimates of the partial derivatives are:

$$\widehat{\partial_{\mu_0}\mathcal{I}_{\lambda^0}} = \frac{1}{N}\sum_{t=1}^N (f(\mu_j^{(t)}) - \widehat{\mathcal{I}_{\lambda^0}})\left(\frac{\partial_{\mu_0}K(m, \mu_0)}{K(m, \mu_0)} + m\theta_j(\mu_j^{(t)})\right)\Bigg|_{\lambda=\lambda^0}$$

$$\widehat{\partial_m\mathcal{I}_{\lambda^0}} = \frac{1}{N}\sum_{t=1}^N (f(\mu_j^{(t)}) - \widehat{\mathcal{I}_{\lambda^0}})\left(\frac{\partial_m K(m, \mu_0)}{K(m, \mu_0)} + \mu_0\theta_j(\mu_j^{(t)}) - M(\theta_j(\mu_j^{(t)}))\right)\Bigg|_{\lambda=\lambda^0}$$

where $\mu_j^{(1)}, \mu_j^{(2)}, \dots, \mu_j^{(N)}$, $j = 1, 2, \dots, k$ is the permuted sample and $\widehat{\mathcal{I}_{\lambda^0}}$ is the estimate of the posterior quantity interest with this sample.

5. Application to reliability

The application of the proposed method is illustrated by using a real data set. The data set shows the times (in minutes) to break down of an insulating fluid under two level of voltage stress. The observations come from two populations with different voltage stress, but it is not known which observation is coming from what populations. Then, a mixture of two exponential distributions is fitted to the data:

$$p(x|\boldsymbol{\omega}, \boldsymbol{\mu}) = \omega_1 \cdot \frac{\exp(-x/\mu_1)}{\mu_1} + \omega_2 \cdot \frac{\exp(-x/\mu_2)}{\mu_2},$$

where $\omega_1 + \omega_2 = 1$, and $\mu_1, \mu_2 > 0$.

The prior distribution for $\boldsymbol{\omega} = (\omega_1, \omega_2) \sim$ Diritchlet(δ, δ) with $\delta = 1$. By other side, a conjugate prior distribution is chosen for μ_j, $j = 1, 2$, i.e., $\mu_j \sim$ InvertedGamma$(m + 1, m\mu_0)$, with $\mu_0 = 1$ and $m = 0.05$. The interest is focused on studying how the posterior means of the parameters μ_j, vary, when changes on the parameters μ_0 and m are realized in a neighborhood of $\mu_0 = 1$ and $m = 0.05$.

A random sample of size 150,000 for the posterior distributions of the weight and mean vectors is generated by using the Gibbs sampling method. A post-processing study on the generated samples has been performed to solve the symmetry in the posterior distribution. The permuted samples have remained invariant. The reason is that two well-differentiated populations have been detected.

The posterior expectation for μ_j is represented by $\mathcal{E}_{\mu_0,m}(\mu_j) = E[\mu_j|\boldsymbol{x}, \mu_0, m]$ for $j = 1, 2$.

j	$\widehat{\mathcal{E}_{\mu_0,m}(\mu_j)}$	$\widehat{\partial_{\mu_0}\mathcal{E}_{\mu_0,m}(\mu_j)}$	$\widehat{\partial_m\mathcal{E}_{\mu_0,m}(\mu_i)}$	$\widehat{SE}(\widehat{\partial_{\mu_0}\mathcal{E}_{\mu_0,m}(\mu_i)})$	$\widehat{SE}(\widehat{\partial_m\mathcal{E}_{\mu_0,m}(\mu_i)})$
1	0.581545	0.002949	0.025551	0.000354	0.006316
2	12.107760	0.010073	-2.263961	0.001667	0.244742

Estimated posterior means, partial derivatives and standard errors

The partial derivatives indicates how rapidly the failure posterior mean is changing around an infinitesimal neighborhood of the parameter values of the prior distribution.

For the first component, the partial derivatives with respect to μ_0 and m can be considered small enough relative to the value of the posterior mean. Therefore, the rate of change in both directions can be considered small, leading to a locally robust estimation.

With respect to the second component, note the increasing in the magnitude of the estimation of the posterior mean The partial derivative for μ_0 is very small relative to the posterior mean. The partial derivative with respect to m represents less than 18.5% of the value of the posterior mean. Although, it represents a larger rate of change than the corresponding to μ_0, less than 20% of relative percentage can be assumed. Therefore, the estimation of this posterior mean can also be considered as locally robust for the parameter values of the prior distribution.