MCMC-based Sensitivity Estimations

Pérez C, Martín J, and Rufo M. J.

Department of Mathematics, University of Extremadura, Avda. Universidad s/n, 10071 Cáceres, Spain
E-mail: carper@unex.es, jrmartin@unex.es, mrufo@unex.es

Abstract: The problem of calculating local parametric sensitivities is addressed. We propose a computationally low-cost method to estimate local sensitivities in Bayesian models. The proposed general method introduces a great flexibility because it can be applied to complex models that need to be solved by MCMC methods, and it allows to estimate the sensitivity measures and their errors with no additional random sampling. This sensitivity analysis method is easy to apply in practice as we show with an illustrative example.

Keywords: Sensitivity, Simulation, MCMC, Bayesian Inference, Bayesian Decision Theory.

1. INTRODUCTION

Many problems in statistics and operational research involve making decisions under uncertainty. Bayesian statistical methods provide a complete paradigm for both statistical inference and decision making under uncertainty. This methodology allows to combine information derived from observations with information elicited from experts. The range of its potential applicability is very wide. It is particularly useful for highly reliable components and systems where failures in test and field operations are very rare, requiring the use of all other engineering information. This methodology has become more popular due to the appearance of Markov Chain Monte Carlo (MCMC) methods (see Brooks [1] for a review). The application of these simulation techniques allows to obtain a numerical solution of problems based on really complex models. Sometimes, MCMC methods are the only computationally efficient alternative.

In addition to the solution, we need some description of its sensitivity with respect to reasonable changes and uncertainties in the specification of the inputs. Sensitivity analysis seeks to find out how the output of a model changes with variations in the inputs (see Saltelli et al. [2]). Such knowledge is important for (a) evaluating the applicability of the model, (b) determining parameters for which it is important to have more accurate values, and (c) understanding the behavior of the system being modeled. The output needs to be interpreted carefully whenever it changes significantly for input variations that are within the bounds of possible error. There are two kinds of sensitivity analysis: local and global. Local sensitivity studies parameter variations over neighborhoods around what are believed to be appropriate values, while global sensitivity considers parameter changes over the whole domain.

Further author information: (Corresponding author: C. Pérez)

Sensitivity analysis is required in many applications, for example, in those arising in engineering, medicine, archeology, or environment. It is particularly useful in reliability of hardware systems, space systems probabilistic risk analysis, nuclear power risk analysis or information security risk analysis. Sensitivity studies are demanded by several authors to be applied in models solved by MCMC methods (see, for example, Ríos and Ruggeri [3]). Some authors, like Hall et al. [4] and Halekoh and Vach [5], study parametric sensitivity by solving the model for some values of the parameters. The main disadvantage of this procedure is that they have to re-run the Markov chain, i.e, they have to generate new samples for the different parameter values. Therefore, it would be convenient to develop a general sensitivity method that can be applied to estimate local parametric sensitivities in Bayesian models solved by MCMC techniques. We address that issue in this paper.

The outline is as follows. In Section 2, a computationally low-cost method to estimate local parametric sensitivities is proposed. In order to show how the proposed method is easily applied in practice, an illustrative example is presented in Section 3. Finally, the conclusion is presented in Section 4.

2. LOCAL PARAMETRIC SENSITIVITY ESTIMATIONS

Suppose we are interested in the estimation of a quantity $\mathcal{I}$ that can be expressed as an integral of a function f over a multiple dimension domain with respect to a density g, i.e:

$$\mathcal{I} = \int_{\Theta} f(\theta)\, g(\theta)\, d\theta. \tag{1}$$

When g is the posterior distribution for θ, i.e, $g(\theta|x)$, this quantity could be, for example, the posterior mean. Note that $g(\theta)$ $(f(\theta))$ could depend on parameters, so a more convenient notation is $g_\lambda(\theta)$ $(f_\lambda(\theta))$ where λ represents a possibly multidimensional parameter in the space Λ. Firstly, we study the problem considering imprecision in g_λ, later we present a similar study for f_λ. In the former case, expression (1) becomes:

$$\mathcal{I}_\lambda = \int_{\Theta} f(\theta)\, g_\lambda(\theta)\, d\theta, \tag{2}$$

where Θ is independent of λ.

Suppose that sampling directly from $g_\lambda(\theta)$ is so complex that we need to use MCMC methods. Note that this is the case for most of the real problems. Let $\theta_1, \theta_2, \ldots, \theta_n$ be a sample generated from $g_{\lambda^0}(\theta)$ by MCMC methods, where λ^0 is a fixed quantity interior to Λ. Then, an estimate of $\mathcal{I}_{\lambda^0}$ is given by:

$$\widehat{\mathcal{I}_{\lambda^0}} = \frac{1}{n}\sum_{i=1}^{n} f(\theta_i). \tag{3}$$

Now, our interest is focused on evaluating the impact of changes on $\mathcal{I}_\lambda$ when λ varies in an infinitesimal neighborhood of λ^0, i.e, we want to make a local sensitivity analysis. The choice of a sensitivity analysis method depends on a great extent on (a) the sensitivity measures employed, (b) the accuracy in the estimates of the sensitivity measures, and (c) the computational cost involved. All these topics are studied in this section.

The first step is to define a local sensitivity measure. This measure must be easily interpretable and efficiently computed. Sometimes sensitivity is characterized through gradients or partial derivatives at the target point (see Turányi and Rabitz [6] and references therein). Suppose that all the partial derivatives exist. As a local sensitivity measure, we consider the gradient vector evaluated at λ^0 , i.e:

$$\nabla \mathcal{I}_{\lambda^0} = \left(\partial_{\lambda_1}\mathcal{I}_{\lambda^0}, \partial_{\lambda_2}\mathcal{I}_{\lambda^0}, \ldots, \partial_{\lambda_m}\mathcal{I}_{\lambda^0}\right). \tag{4}$$

Components in (4), i.e. the partial derivatives with respect to each λ_j evaluated at λ^0, indicate how rapidly $\mathcal{I}_\lambda$ is changing around an infinitesimal neighborhood of λ^0 along that axis. Therefore, they can be used as rates of change with respect to the parameter components. Then $\nabla\mathcal{I}_{\lambda^0}$ can be considered as a local sensitivity measure for the parameter λ at λ^0. The gradient vector represents the precise direction which has maximum increase of $\mathcal{I}_\lambda$ at λ^0. Furthermore, it indicates which component has the largest influence on the output.

In this context, the main problem is to calculate the gradient vector. We present a computationally low-cost method to estimate the components of (4). Under mild conditions, each component of $\nabla\mathcal{I}_{\lambda^0}$ can be expressed as:

$$\begin{aligned}\partial_{\lambda_j}\mathcal{I}_{\lambda^0} &= \int_\Theta \partial_{\lambda_j}\left(f(\theta)\, g_{\lambda^0}(\theta)\right) d\theta = \int_\Theta f(\theta)\partial_{\lambda_j} g_{\lambda^0}(\theta) d\theta = \\ &= \int_\Theta \frac{f(\theta)\partial_{\lambda_j} g_{\lambda^0}(\theta)}{g_{\lambda^0}(\theta)} g_{\lambda^0}(\theta)\, d\theta = E_{g_{\lambda^0}}\left(\frac{f(\theta)\,\partial_{\lambda_j} g_{\lambda^0}(\theta)}{g_{\lambda^0}(\theta)}\right)\end{aligned} \tag{5}$$

and estimated by:

$$\widehat{\partial_{\lambda_j}\mathcal{I}_{\lambda^0}} = \frac{1}{n}\sum_{i=1}^{n} \frac{f(\theta_i)\,\partial_{\lambda_j} g_{\lambda^0}(\theta_i)}{g_{\lambda^0}(\theta_i)}. \tag{6}$$

Also, we can estimate the error committed when estimating (5) by using (6). For each j, the estimate given in (6) is unbiased, so its error can be measured by its standard error (see e.g. Tanner [7]). The estimation of the error can be easily obtained from the generated sample that has been used to estimate $\mathcal{I}_{\lambda^0}$ and $\partial_{\lambda_j}\mathcal{I}_{\lambda^0}$.

The advantages of this local sensitivity analysis procedure are mainly two. First, it can be applied to complex models that need MCMC methods to sample from the objective densities. Second, the computations generally represent a very low additional cost because no further sampling is required. The same MCMC outputs obtained to estimate $\mathcal{I}_{\lambda^0}$ are used to estimate its sensitivity and the errors in the estimations. However, this approach can only be applied when we know a closed expression for g_{λ^0} and we can calculate its partial derivatives, what is not always possible. In fact, for complex models the explicit form for g_{λ^0} is usually analytically intractable. Nevertheless, we can obtain some results studying the practical implementation when g_{λ^0} is the posterior distribution. The following two cases are considered.

1. Prior sensitivity. Suppose that the prior distribution $\pi_\lambda(\theta)$ depends on a parameter λ, and let λ^0 be interior to Λ, then:

$$\mathcal{I}_{\lambda^0} = \int_\Theta f(\theta)\, p_{\lambda^0}(\theta|x)\, d\theta = \frac{\int_\Theta f(\theta)\, l(x|\theta)\, \pi_{\lambda^0}(\theta)\, d\theta}{\int_\Theta l(x|\theta)\, \pi_{\lambda^0}(\theta)\, d\theta}. \tag{7}$$

Under mild conditions that allow a derivative-integral interchange (see Spall [8]), we find that each component of $\nabla \mathcal{I}_{\lambda^0}$ can be expressed as:

$$\partial_{\lambda_j} \mathcal{I}_{\lambda^0} = \int_\Theta (f(\theta) - \mathcal{I}_{\lambda^0})\, \frac{\partial_{\lambda_j} \pi_{\lambda^0}(\theta)}{\pi_{\lambda^0}(\theta)}\, p_{\lambda^0}(\theta|x)\, d\theta.$$

The proof is mainly based on the derivative-integral interchange. The posterior steps are basic manipulations addressed to get the integral of a function with respect to the posterior distribution.

If $\theta_1, \theta_2, \dots, \theta_n$ is generated from the posterior distribution $p_{\lambda^0}(\theta|x)$ (mainly by MCMC methods), then the estimate of $\partial_{\lambda_j} \mathcal{I}_{\lambda^0}$ is given by:

$$\widehat{\partial_{\lambda_j} \mathcal{I}_{\lambda^0}} = \frac{1}{n} \sum_{i=1}^{n} \left(f(\theta_i) - \widehat{\mathcal{I}_{\lambda^0}} \right) \frac{\partial_{\lambda_j} \pi_{\lambda^0}(\theta_i)}{\pi_{\lambda^0}(\theta_i)}, \tag{8}$$

where $\widehat{\mathcal{I}_{\lambda^0}} = \frac{1}{n} \sum_{i=1}^{n} f(\theta_i)$ is the estimate of $\mathcal{I}_{\lambda^0}$. The Monte Carlo standard error estimate of (8) is given by:

$$\widehat{SE}(\widehat{\partial_{\lambda_j} \mathcal{I}_{\lambda^0}}) = \sqrt{\frac{1}{n\,(n-1)} \sum_{i=1}^{n} \left(\frac{(f(\theta_i) - \widehat{\mathcal{I}_{\lambda^0}}) \partial_{\lambda_j} \pi_{\lambda^0}(\theta_i)}{\pi_{\lambda^0}(\theta_i)} - \widehat{\partial_{\lambda_j} \mathcal{I}_{\lambda^0}} \right)^2} \tag{9}$$

Note that this case is more tractable because we need the partial derivatives for the prior distribution instead of the partial derivatives for the posterior distribution.

2. Function f sensitivity. If we consider that f belongs to a parametric class of functions, $\mathcal{F}_\lambda = \{f_\lambda, \lambda \in \Lambda\}$, then:

$$\mathcal{I}_{\lambda^0} = \int_\Theta f_{\lambda^0}(\theta)\, p(\theta|x)\, d\theta = \frac{\int_\Theta f_{\lambda^0}(\theta)\, l(x|\theta)\, \pi(\theta)\, d\theta}{\int_\Theta l(x|\theta)\, \pi(\theta)\, d\theta}.$$

Under the mild conditions analogous to the previous case, for each j we have:

$$\partial_{\lambda_j} \mathcal{I}_{\lambda^0} = \int_\Theta \partial_{\lambda_j} f_{\lambda^0}(\theta)\, p(\theta|x)\, d\theta,$$

and its estimate is given by:

$$\widehat{\partial_{\lambda_j} \mathcal{I}_{\lambda^0}} = \frac{1}{n} \sum_{i=1}^{n} \partial_{\lambda_j} f_{\lambda^0}(\theta_i), \tag{10}$$

where $\theta_1, \theta_2, \ldots, \theta_n \sim p(\theta|x)$. Now, the Monte Carlo standard error estimate of (10) is given by:

$$\widehat{SE}(\widehat{\partial_{\lambda_j} \mathcal{I}_{\lambda^0}}) = \sqrt{\frac{1}{n\,(n-1)} \sum_{i=1}^{n} \left(\partial_{\lambda_j} f_{\lambda^0}(\theta_i) - \widehat{\partial_{\lambda_j} \mathcal{I}_{\lambda^0}} \right)^2} \tag{11}$$

Note that if in any problem, the functions g and f depend on the same parameter λ, then the sensitivity measure proposed in this section can be estimated in the same sense. The estimate of each component can be expressed as the sum of analogous quantities to (8) and (10).

As a particular case, we can study the practical implementation of the proposed sensitivity measure in the context of Bayesian decision theory (among the many fine reviews are, for example, Berger [9] and French and Ríos [10]). Bayesian decision theory and inference describe a decision problem by a set of possible actions $a \in \Delta$, a set of states, or parameters, $\theta \in \Theta$, a prior distribution $\pi(\theta)$, a likelihood, $l(x|\theta)$ for the observed data x, and a loss (utility) function $l(a, \theta)$ $(u(a, \theta))$. The actions are ranked by the expected loss (utility). The optimal decision a^* is the action that minimizes (maximizes) the posterior expected loss (utility):

$$\begin{aligned} a^* &= \arg \min_{a \in \Delta} L(a), \\ L(a) &= \int l(a, \theta)\, p(\theta|x)\, d\theta = \frac{\int l(a,\theta)\, l(x|\theta)\, \pi(\theta)\, d\theta}{\int l(x|\theta)\, \pi(\theta)\, d\theta}. \end{aligned}$$

Practical implementation is hindered by the fact that $L(a)$ and hence the minimum a^* could be sensitive to the chosen prior $\pi(\cdot)$, likelihood $l(\cdot|\cdot)$ and/or loss function $l(\cdot)$. A skeptical decision maker will require, in addition to the optimal solution, some description of the sensitivity of a^* with respect to reasonable changes and uncertainties in the specification of the inputs. This type of sensitivity is known as functional sensitivity because the inputs are functions. Excellent summaries of Bayesian literature in this area are provided by Berger [11] and Ríos and Ruggeri [3].

In this context, we can investigate the local parametric sensitivity of $L_\lambda(a^*)$ where λ is a possibly multidimensional parameter that models the loss function and/or the prior distribution. Now, $f_\lambda(\theta) = l_\lambda(a^*, \theta)$ and the quantity of interest $\mathcal{I}_{\lambda^0}$ is $L_{\lambda^0}(a^*)$. Note that we refer to expected loss sensitivity instead of decision sensitivity (see Kadane and Srinivasan [12] for a distinction).

In the next section, we show how the proposed computationally low-cost sensitivity estimations and their errors can be easily calculated in practice.

3. APPLICATION

We consider an illustrative example relating to 10 power plant pumps. George et al. [13] provided a complete Bayesian hierarchical analysis of the pump failure data previously studied by Gaver and O'Muircheartaigh [14]. For the power plant pump i, the failure rate

i	1	2	3	4	5	6	7	8	9	10
t_i	94.32	15.72	62.88	12.76	5.24	31.44	1.05	1.05	2.09	10.48
x_i	5	1	5	14	3	19	1	1	4	22

Table 1. Pump failure data.

is denoted by θ_i and the length of operation time (in thousands of hours) is denoted by t_i. The data are given in Table 1.

Conditional on θ_i, the number of failures X_i is assumed to follow a Poisson distribution, $X_i|\theta_i \sim \text{Poisson}(\eta_i)$, $i = 1, \ldots, 10$, where $\eta_i = \theta_i\, t_i$ and X_i is independent of X_j for $i \neq j$. Conditional on α and β, independent gamma prior distributions are adopted for the failure rates, $\theta_i|\alpha, \beta \sim \text{Gamma}(\alpha, \beta)$. We assume the following prior specification for α and β :

$$\begin{aligned} \alpha &\sim \text{Exp}(\lambda_1), \\ \beta &\sim \text{Gamma}(\lambda_2, \lambda_3), \end{aligned}$$

where $\lambda_1 = 1$, $\lambda_2 = 0.1$, and $\lambda_3 = 1$. The model is graphically represented in Figure 1. This graph has been obtained by using DoodleBUGS that has been developed to specify graphical models in Bayesian context. This tool is included in WinBUGS PACKAGE (Spiegelhalter [15]).

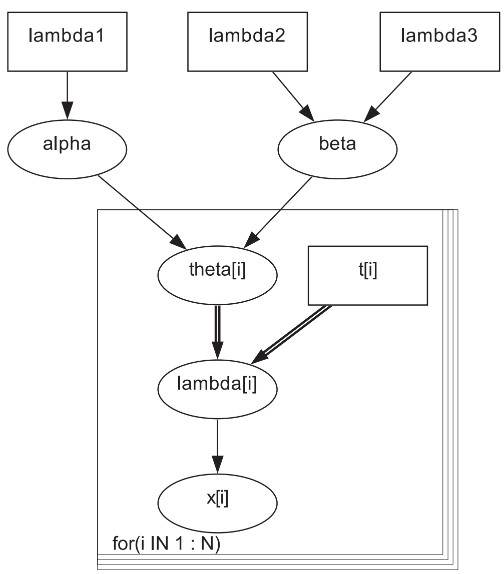

Figure 1. Graphical model.

We carry out a sensitivity analysis in the terms described in the previous section. We focus our interest on the posterior mean for the parameters θ_i, $i = 1, 2, \ldots, 10$. Those quantities represent the means of the failure rates after the Bayes update has been done. We study if the posterior means of the parameters are sensitive to the initial values of the prior specification, i.e. we study local sensitivity with respect to the parameter $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$ in the neighborhood of $\boldsymbol{\lambda}^0 = (1, 0.1, 1)$. In this case, the quantities of interest are $E_{\boldsymbol{\lambda}^0}[\theta_i|x]$, and, in order to simplify, they will be denoted by $\mathcal{E}_{\boldsymbol{\lambda}^0}(i)$, $i = 1, 2, \ldots, 10$.

By using WinBUBS, we can generate MCMC samples from the posterior distributions for all parameters. After we consider that the convergence has been achieved, we generate a sample of size $n = 10000$. The estimations of $\mathcal{E}_{\boldsymbol{\lambda}^0}(i)$, are given in Table 2.

Parameters	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
$\widehat{\mathcal{E}_{\boldsymbol{\lambda}^0}(i)}$	0.059	0.102	0.089	0.116	0.604	0.609	0.893	0.881	1.584	1.992

Table 2. Estimations of the posterior means.

Table 3 shows the estimations of the partial derivatives $\partial_{\lambda_j}\mathcal{E}_{\boldsymbol{\lambda}^0}(i)$, $j = 1, 2, 3$, $i = 1, 2, \ldots, 10$.

	λ_1	λ_2	λ_3
θ_1	$1.34 \cdot 10^{-4}$	$-7.11 \cdot 10^{-5}$	$7.25 \cdot 10^{-5}$
θ_2	$4.34 \cdot 10^{-4}$	$3.46 \cdot 10^{-4}$	$-4.37 \cdot 10^{-4}$
θ_3	$4.58 \cdot 10^{-5}$	$2.25 \cdot 10^{-4}$	$-2.09 \cdot 10^{-4}$
θ_4	$1.13 \cdot 10^{-5}$	$2.69 \cdot 10^{-4}$	$-2.28 \cdot 10^{-4}$
θ_5	$1.45 \cdot 10^{-4}$	$-1.41 \cdot 10^{-3}$	$7.66 \cdot 10^{-4}$
θ_6	$1.13 \cdot 10^{-4}$	$3.69 \cdot 10^{-4}$	$-3.68 \cdot 10^{-5}$
θ_7	$-2.27 \cdot 10^{-3}$	$-3.09 \cdot 10^{-3}$	$5.72 \cdot 10^{-3}$
θ_8	$2.41 \cdot 10^{-3}$	$7.69 \cdot 10^{-3}$	$-7.33 \cdot 10^{-3}$
θ_9	$3.83 \cdot 10^{-3}$	$-1.19 \cdot 10^{-3}$	$5.34 \cdot 10^{-4}$
θ_{10}	$7.58 \cdot 10^{-4}$	$-8.09 \cdot 10^{-2}$	$6.83 \cdot 10^{-3}$

Table 3. Estimations of the partial derivatives.

We consider that the rate of change for λ_1, λ_2 and λ_3 are within the reasonable limits with respect to the values of $\widehat{\mathcal{E}_{\boldsymbol{\lambda}^0}(i)}$, $i = 1, 2, \ldots, 10$. So the components of $\widehat{\nabla\mathcal{E}_{\boldsymbol{\lambda}^0}(i)}$ indicate that we can consider $\boldsymbol{\lambda}^{\mathbf{0}} = (1, 0.1, 1)$ as a robust value for the parameter $\boldsymbol{\lambda}$ in this model.

4. CONCLUSION

In Bayesian decision theory and inference the proposed local parametric sensitivity procedure can be very useful because it is a general technique applicable to complex models that need to be solved by MCMC methods. Besides, the MCMC simulations can be reused to estimate the sensitivity measures and their errors, avoiding the need of further sampling. This computationally low-cost method is easy to apply in practice and it is specially recommended to study sensitivities in reliability models.

ADKNOWLEDGEMENTS

Comments from James C. Spall are gratefully acknowledged. This research was partially supported by Junta de Extremadura, Spain (Project IPR00A075).

REFERENCES

1. S. P. Brooks. Markov chain Monte Carlo method and its application. *The Statistician*, 47:69–100, 1998.
2. A. Saltelli, K. Chan, and E.M. Scott. *Sensitivity Analysis*. Wiley, 2000.
3. D. Ríos Insua and F. Ruggeri. *Robust Bayesian Analysis*. Springer, 2000.
4. C. B. Hall, J. Y. Lynn Kuo, and R. B. Lipton. Bayesian and profile likelihood change point methods for modeling cognitive function over time. *Computational Statistics and Data Analysis*, 42(1-2):91–109, 2003.
5. U. Halekoh and W. Vach. A Bayesian approach to seriation problems in archaeology. *Computational Statistics and Data Analysis.*, 2003. In press, corrected proof, available online.
6. T. Turányi and H. Rabitz. *Sensitivity Analysis*, chapter Local Methods, pages 81–99. Wiley, 2003.
7. M.A. Tanner. *Tools for Statistical Inference*. Springer-Verlag, 1996.
8. J. C. Spall. *Introduction to Stochastic Search and Optimization. Estimation, Simulation and Control.* Wiley, 2003.
9. J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, 1985.
10. S. French and D. Ríos Insua. *Statistical Decision Theory*. Arnold, 2000.
11. J. O. Berger. An overview of robust Bayesian analysis. *Test*, 3:5–124, 1994.
12. J. Kadane and C. Srinivasan. Comment on “An Overview of Robust Bayesian Analysis” by Berger. *Test*, 3:93–95, 1994.
13. U. George, E. I. Makov and A. F. M. Smith. Conjugate likelihood distributions. *Scandinavian Journal of Statistics*, 20:147–156, 1993.
14. D. P. Gaver and I. G. O′Muircheartaigh. Robust empirical Bayes analysis of event rates. *Technometrics*, 29:1–15, 1987.
15. D. Spiegelhalter, A. Thomas, N. Best, and D. Lunn. *WinBUGS User Manual, Version 1.4*. MRC Biostatistics Unit, Cambridge, 2003.