Proceedings of the 1993 Winter Simulation Conference
G.W. Evans, M. Mollaghasemi, E.C. Russell, W.E. Biles (eds.)

MONTE CARLO ESTIMATION OF BAYESIAN ROBUSTNESS

Jin Wang
Bruce W. Schmeiser

School of Industrial Engineering
Purdue University
West Lafayette, IN 47907, U.S.A.

ABSTRACT

Bayesian estimation procedures often require Monte Carlo integration with respect to the posterior distribution. We propose a Monte Carlo estimator of an arbitrary posterior-distribution property, as well as its gradient with respect to prior-distribution hyperparameters and to the observed data. Unlike most Monte Carlo samplers for Bayesian problems, we sample from the prior distribution, which is usually more tractable than the posterior distribution. We discuss sufficient conditions for interchanging expected value and differentiation, so that the gradient can be estimated by averaging observations of the stochastic gradient. In addition to the gradient estimator, we suggest asymptotically valid standard error and confidence-interval estimators. We give two numerical examples.

1 INTRODUCTION

The Bayesian calculation problem is to evaluate the posterior integral $\delta^\pi(x,\lambda) = E^{\pi(\theta|x,\lambda)}[g(\theta)]$, where $\pi(\theta|x,\lambda)$ is the posterior distribution, x is the vector of observed data, λ is the vector of hyperparameters of the prior distribution, and g is any measurable function.

We are interested in estimating the *local sensitivity* of $\delta^\pi(x,\lambda)$ to the hyperparameters λ and *local resistance* to the data x. (See, for example, Polasek 1982.) In particular, we estimate the gradients $\nabla_\lambda \delta^\pi(x,\lambda)$ and $\nabla_x \delta^\pi(x,\lambda)$ by simulating only at a single value of the hyperparameters λ and observed data x. We obtain i.i.d. realizations from the prior distribution, compute various stochastic gradients, and combine them to estimate the gradients.

Given a particular application, the two key issues are (1) whether the gradient estimator converges and (2) the derivation of the various stochastic gradients.

2 LOCAL SENSITIVITY

Here we consider local sensitivity of the general Bayesian posterior integral $\delta^\pi(x,\lambda) = E^{\pi(\theta|x,\lambda)}[g(\theta)]$ to λ. Define $n(x,\lambda)$ and $d(x,\lambda)$:

$$\delta^\pi(x,\lambda) = \frac{\int_\Theta g(\theta) f(x|\theta)\pi(\theta|\lambda)d\theta}{\int_\Theta f(x|\theta)\pi(\theta|\lambda)d\theta} \equiv \frac{n(x,\lambda)}{d(x,\lambda)},$$

where $f(x|\theta)$ is the likelihood of data x. The local sensitivity, the derivative of $\delta^\pi(x,\lambda)$ with respect to the parameter λ, is

$$\frac{\partial}{\partial\lambda}\delta^\pi(x,\lambda) = \frac{1}{d(x,\lambda)}\left[\frac{\partial}{\partial\lambda}n(x,\lambda) - \delta^\pi(x,\lambda)\frac{\partial}{\partial\lambda}d(x,\lambda)\right].$$

Let $\theta_i(\lambda)$ be i.i.d realizations from the prior density $\pi(\theta|\lambda)$. Define the component estimators

$$\begin{aligned}
\overline{n(x,\lambda)} &= \frac{1}{n}\sum_{i=1}^{n} g(\theta_i(\lambda))\; f(x|\theta_i(\lambda)),\\
\overline{d(x,\lambda)} &= \frac{1}{n}\sum_{i=1}^{n} f(x|\theta_i(\lambda)),\\
\overline{\frac{\partial}{\partial\lambda}n(x,\lambda)} &= \frac{1}{n}\sum_{i=1}^{n}\left\{\left[\frac{\partial}{\partial\theta}g(\theta_i(\lambda))\right] f(x|\theta_i(\lambda))\right.\\
&\quad + \left. g(\theta_i(\lambda))\left[\frac{\partial}{\partial\theta}f(x|\theta_i(\lambda))\right]\right\}\frac{\partial}{\partial\lambda}\theta_i(\lambda),
\end{aligned}$$

and

$$\overline{\frac{\partial}{\partial\lambda}d(x,\lambda)} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}f(x|\theta_i(\lambda))\frac{\partial}{\partial\lambda}\theta_i(\lambda).$$

The combined estimator of the local sensitivity is

$$\widehat{\frac{\partial}{\partial\lambda}\delta^\pi(x,\lambda)} = \frac{1}{\overline{d(x,\lambda)}}\left[\overline{\frac{\partial}{\partial\lambda}n(x,\lambda)} - \frac{\overline{n(x,\lambda)}}{\overline{d(x,\lambda)}}\;\overline{\frac{\partial}{\partial\lambda}d(x,\lambda)}\right].$$

Each component is the infinitesimal perturbation analysis estimator (Glasserman 1991). The parameter λ could be a vector.

3 LARGE-SAMPLE BEHAVIOR

Using results from Billingsley (1986) and Reiman & Weiss (1989), and assuming some regularity conditions, we have shown that

(a) the component estimators are unbiased,
(b) the combined estimator is strongly consistent and asymptotically normal if all expectations exist and
(c) asymptotically valid standard error and confidence intervals are available.

4 NUMERICAL EXAMPLES

The first example is a reliability problem to which we apply the combined estimator method directly. The second is a simple textbook example for which importance sampling is needed to obtain numerical stability. The method works well in both examples.

4.1 Non-Homogeneous Poisson Process.

Crow (1974) considers the problem of estimating a Non-Homogenous Poisson process rate function $\lambda(t) = \beta(1-\alpha)t^{1-\alpha}$ using observed failure-time data and independent prior distributions on α and β. We estimate the gradient of the posterior means of α and β with respect to parameters of the prior distributions. Let N_t represent the number of failures experienced by time t. Let $P(N_t = k) = [\Lambda^k(t)\exp\{-\Lambda(t)\}]/k!$ for $k = 0, 1, \ldots$ and $EN_t = \Lambda(t)$, where $\Lambda(t) = \int_0^t \lambda(u)du$ and $\lambda(u)$ is the rate of occurrence of failures.

If failures are observed at time $t_0 = 0 < t_1 < t_2 < \cdots < t_{n_0}$, the likelihood is $L = \{\beta(1-\alpha)\}^{n_0}\{\prod_{i=1}^{n_0} t_i\}^{-\alpha}\exp\{-\beta t_{n_0}^{1-\alpha}\}$. Assume that α and β have independent prior distributions with densities $\phi(\alpha)$ and $\psi(\beta)$, respectively. Suppose $\phi(\alpha)$ and $\psi(\beta)$ are U(a,b) and Gamma(m,μ) densities, respectively, with a, b and m known. Then μ is the hyperparameter of our Bayesian model and there is no closed form for the posterior density.

We are interested in the Bayesian point estimators for α and β: the posterior means of α and β. We consider sensitivity of the Bayesian estimators with respect to the hyperparameter μ. We need to calculate all stochastic derivatives in order to obtain the combined estimators. We use the data of Sander and Badoux (1991) to run our Monte Carlo experiment.

The Monte Carlo simulation results shows that $\alpha^{\pi}(t, m, \mu)$ is less dependent than $\beta^{\pi}(t, m, \mu)$ on the hyperparameter μ. The reason is that the prior of α is U(0,1), which is independent of μ, and the prior of β is Gamma(m,μ), which is not independent of μ.

4.2 Multivariate Eatimation

Following Berger (1985, page 247), we discuss the Bayesian estimation of a vector $\theta = (\theta_1, \ldots, \theta_m)^t$ with a hyperparameter vector $\lambda = (\lambda_1, \ldots, \lambda_m)^t$. We assume that $X = (X_1, \ldots, X_m)^t \sim \mathcal{N}_m(\theta, I)$ and $\theta \sim \mathcal{N}_m(0, \Sigma)$, where Σ =diagonal$(\lambda_1, \ldots, \lambda_m)$.

We consider sensitivity of the posterior means and posterior covariances with respect to λ. Since we have selected a specific model, we can find the analytical expressions so that we can compare the Monte Carlo results to the true values. Numerical problems arose when we coded the combined estimator. Importance sampling solved this problem.

The Monte Carlo experimental results are very good in the sense of comparing the combined estimators with the true values.

ACKNOWLEDGMENTS

This work was supported in part by NSF grants DMS-87-17799 and DDM-93-00058. This paper is a summary of Wang and Schmeiser (1993).

REFERENCES

Billingsley, P. 1986. *Probability and Measure*. New York: John Wiley & Sons.

Berger, J.O. 1985. *Statistical Decision Theory and Bayesian Analysis*. 2nd ed. New York: Springer-Verlag.

Crow, L.H. 1974. Reliability analysis for complex repairable systems. In *Reliability and Biometry*, ed. F. Proschan and R.J. Serfling, 379–410. Philadelphia: SIAM.

Glasserman, P. 1991. *Gradient Estimation via Perturbation Analysis*. Boston, Mass: Kluwer Academic Publishers.

Polasek, W. 1982. Local Sensitivity Analysis and Matrix Derivative. In *Operations Research in Progress*, ed. G. Feichtinger and P. Kall, 425–443. Dordrecht, Holland: Reidel.

Reiman, M.I., and A. Weiss, 1989. Sensitivity analysis for simulations via likelihood ratios. *Operations Research* 37:830–844.

Sander, P., and R. Badoux, 1991. *Bayesian Methods in Reliability*. Boston, Mass: Kluwer Academic Publishers.

Wang, J., and B. Schmeiser, 1993. Monte Carlo estimation of Bayesian robustness. Technical Report SMS 93-11, School of Industrial Engineering, Purdue University.