ISSN 1684-8403
Journal of Statistics
Volume 19, 2012. pp. 83-90

Numerical Methods for Bayesian Analysis

Nasir Abbas[1], Syed Mohsin Ali Kazmi[2] and Muhammad Aslam[3]

Abstract

Bayesian Inference is a technique of statistical inference that, in addition to using the sample information, utilizes prior information about Parameter(s) to draw results about the Parameters. But the beauty is subdued by huge and cumbersome algebraic calculations necessary to find Posterior estimates. This article suggests numerical methods to derive Posterior distributions under all types of priors – uninformative and informative – and to find Bayes Estimates. We use both numerical differentiation and numerical integration to serve the purpose. The entire estimation procedure is illustrated using real as well as simulated datasets.

Keywords

Numerical differentiation, Numerical integration, Fisher's information, Jeffreys prior, Squared error loss function (SELF), Bayes estimator, Exponential distribution, Normal distribution

1. Introduction

The main difference between the Bayesian and the Frequentistic schools of thoughts is that the former associate randomness with population Parameters and formally incorporate in their analysis any prior information pertaining to Parameters. Prior information about Parameters is updated with current information (data) to yield Posterior Distribution, which is a work-bench for the Bayesians. Adams (2005) throws light on the advantages of using Bayesian approach. But the major problem, which Bayesians face, is the calculation of Posterior Estimates via the Posterior Distribution.

[1] Department of Statistics, Government Postgraduate College, Jhang, Pakistan.
Email: nabbasgcj@yahoo.com
[2] Sustainable Development Policy Institute, Islamabad, Pakistan.
[3] Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan.

The problem is even aggravated when we use the Jeffreys Prior. It renders the Posterior Distribution and hence the Posterior Inference even more complicated. Different numerical techniques, like Gibbs sampler, numerical integration etc. are used to address these problems. WinBUGS is recently-developed software that is being extensively used to get the Posterior summaries of Parameters.

In this study, an effort is made to suggest numerical technique that efficiently deals with all the problems of Posterior Estimation. It not only gives us the Posterior Estimates of Parameters but also accommodates the Jeffreys Prior if it is suggested to be used. The use of other type of priors – uninformative uniform prior, conjugate prior and other informative priors – is comparatively easy than that of the Jeffreys Prior.

The breakup of the study is as follows: In Section 2, the Fisher's information and the Jeffreys Prior are explained along with formulae for numerical differentiation. Section 3 is concerned with the Quadrature method of numerical integration. In Section 4, the estimation methodology is explained. Section 5 presents illustrative examples of the entire estimation procedure. We have considered the cases of one as well as two Parameters by taking into account the Exponential and the Normal Distributions. Multivariate Distributions may similarly be accounted for. Section 6 concludes the entire study and discusses the results.

2. The Jeffreys Prior

The situations where we do not have much information about the Parameters of a model, we use an uninformative prior proposed by Jeffreys (1946, 1961) and is defined as the density of Parameters proportional to the square root of the determinant of the Fisher's information matrix. Let the dataset X be drawn from a certain Distribution $f(X|\theta)$ that depends upon the vector of Parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_t)$. The Likelihood Function is denoted by $L(X;\theta)$ and its Fisher's Information is given by $I(\theta) = -E_{\boldsymbol{X}|\boldsymbol{\theta}}\left\{\frac{\partial^2 L(X;\theta)}{\partial\theta^2}\right\}$. The Fisher's Information measures the sensitivity of an estimator in the neighborhood of the Maximum Likelihood Estimate (MLE), as it is proportional to the expected curvature of the Likelihood at the MLE. Jeffreys, more generally, suggests invariance prior (Berger, 1985) which takes the form

$$p(\theta) \propto \sqrt{\det\{I(\theta)\}}, \qquad \text{where } \theta \in \Omega \tag{2.1}$$

If the nature of expressions involved in the determinant is complicated, we may use the numerical methods for finding the second partial derivatives to calculate the Jeffreys Prior using the relations

$$f''_{uu} = \frac{\partial^2 f(u,v)}{\partial u^2}\bigg|_{(u_0,v_0)} = \frac{f(u_0-h,v_0)-2f(u_0,v_0)+f(u_0+h,v_0)}{h^2}, \tag{2.2}$$

and

$$f''_{uv} = \frac{\partial^2 f(u,v)}{\partial u \partial v}\bigg|_{(u_0,v_0)} = \frac{f(u_0+h,u_0+h)+f(u_0-h,u_0-h)-h^2(f''_{uu}+f''_{vv})-2f(u_0,v_0)}{h^2}, \tag{2.3}$$

where $f(u,v)$ is a bi-variate function and f''_{uu} denotes the partial double derivative with regard to the random variable u. For more details, one can see Bernardo (1979), Berger and Bernardo (1989, 1992a, 1992b), Datta and Ghosh (1995), Jeffreys (1961).

3. The Quadrature Method

We usually need to evaluate multiple integrals to find Bayes Estimates, for example, Posterior means, predictive probabilities, Posterior probabilities for Hypotheses testing etc. based on complicated nature of the Posterior Distribution, particularly when there is a Vector of Parameters and the expressions involve complicated algebraic functions. Considering a one-dimensional case, the Quadrature refers to any method for numerically approximating the value of the definite integral $\int_a^b p(\theta)d\theta$, where $p(\theta)$ may be any proper density. The procedure is to calculate it at a number of points in the range 'a' through 'b' and find the result as a weighted average as

$$\int_a^b p(\theta)d\theta = \sum_{i=0}^{n} \varepsilon_i\, p(\theta_i) \tag{3.1}$$

where $a = \theta_0 < \theta_1 < \theta_2, \dots, \theta_n = b,\ \theta_{i+1} = \theta_1 + \varepsilon_i$, for all $i = 0,1,2,\cdots,n$ and ε_i stands for the size of increment used to approach 'b' from 'a'. Here it is important to note that the accuracy and size of the increment are inversely related to each other. Two-dimensional integrals may be evaluated using the relation

$$\int_a^b \int_c^d p(\theta_i,\theta_j)d\theta_i d\theta_j \cong \sum_{i=0}^{n_i} \sum_{i=0}^{n_i} \varepsilon_i\, \varepsilon_j\, p(\theta_i,\theta_j) \tag{3.2}$$

where $\min(\theta_i) = \theta_0 < \theta_1 < \theta_2, \dots, \theta_{n_i} = \max(\theta_i)$, for all θ_i, $\min(\theta_j) = \theta_0 < \theta_1 < \theta_2, \dots, \theta_{n_j} = \max(\theta_j)$, for all θ_j; ε_i and ε_j respectively denote the size of increments in the Parametric values θ_i and θ_j, and $p(\theta_i,\theta_j)$ symbolizes any Bivariate Density.

4. Estimation Methodology

To find the Bayes Estimates of Posterior Distributions, we proceed as follows:

- Use one of the formulae (3.1) or (3.2) for Quadrature, and repeat the calculation process $(b-a)/\varepsilon$ times if there is only one Parameter to be estimated. For more Parameters, we use Nested Loops to evaluate Quadrature. The Loop control variable(s) must be initialized, incremented and checked for the desired number of iterations to reach the terminal value(s). If the ranges involve infinities, the reasonable numbers may be used to represent infinite ranges. For this, we may continuously and gradually expand the limits till the convergence of the Posterior Estimates. The increment of Loops should be accordingly set to attain the precision required. The smaller be the value of increment, the greater the precision would be. The Loop control variables are actually the Parameters to be estimated.
- For an observed sample of size n with values $X_1, X_2, \dots, X_n$ taken from a certain Distribution $f(X|\theta)$, define the Likelihood Function $L(X;\theta)$ of the Distribution of data set (current information) of random variable(s) as $L(X;\theta) = \prod_{i=1}^{n} f(X_i|\theta)$ or logarithm of the Likelihood Function as $l(X;\theta) = \sum_{i=1}^{n} \ln f(X_i|\theta)$.
- Using (2.1), derive the Jeffreys prior as $p(\theta) \propto \sqrt{\det\{I(\theta)\}}$ based on the Fisher's information matrix $I(\theta) = -E_{X|\theta}\left\{\frac{\partial^2 L(X;\theta)}{\partial\theta^2}\right\}$. The numerical differentiation may be carried out using relations (2.2) and/or (2.3). For simplicity, we may use Kernel Density – a density without normalizing constant – to derive the Jeffreys Prior, because the Jeffreys Prior is only the function of Parameters and the normalizing constant too is always independent of the Parameters of the Distribution. Remember that the Posterior Distributions are always the Distributions of the population Parameters considered as random variables.
- Obtain the Posterior Distribution of the Parameters of interest by multiplying the Likelihood Function with the Jeffreys Prior obtained as $f(\theta|X) = k^{-1} p(\theta).L(X;\theta)$, where k is the normalizing constant defined by $k = \int_{-\infty}^{\infty} p(\theta).L(X;\theta)d\theta$. In case of using the Kernel Density, the normalizing constant is obtained by integrating out the Parameter(s) on the entire range(s). The Kernel Density is then divided by the normalizing

constant to get a proper density. In this scenario, the Posterior Distribution is automatically used to yield the desired Posterior Bayes Estimates.

- The entire estimation algorithm may be understood by Figure 1.

5. Illustrations

For the purpose of illustration of the entire estimation procedure, we take two examples: one for single-Parameter Density and the other for two-Parameter Density and consider the Exponential and the Normal Distributions respectively.

5.1 One-Parameter Distribution: For a one-Parameter case, for instance, we consider the Exponential Distribution with density $f(x|\theta) = \theta \exp(-\theta x)$, $\theta > 0$, for $x \geq 0$; and zero elsewhere. Let an observed sample of size 'n' with values $X_1, X_2, \ldots, X_n$ be taken from the Exponential Distribution. The Likelihood Function is $L(\boldsymbol{X};\theta) = \theta^n \exp(-\theta \sum_{i=1}^{n} x_i)$ and the logarithm of the Likelihood Function is $l(\boldsymbol{X};\theta) = n \ln \theta - \theta \sum_{i=1}^{n} x_i$, which implies $\frac{\partial^2 l(\boldsymbol{X};\theta)}{\partial \theta^2} = -\frac{n}{\theta^2}$. Since it does not depend upon **x**, so we get the Fisher's information as $I(\theta) = -E\left\{\frac{\partial^2 l(X;\theta)}{\partial \theta^2}\right\} = \frac{n}{\theta^2}$ and hence the Jeffreys Prior takes the form $J(\theta) \propto \theta^{-1}$. The Posterior Distribution follows the Gamma Distribution as $f(\theta|X) \propto \theta^{n-1} \exp(-\theta \sum_{i=1}^{n} x_i)$, i.e., $\theta|X \sim g(n, \sum_{i=1}^{n} x_i)$ with Posterior mean $E_{\theta|x}(\theta) = \frac{n}{\sum_{i=1}^{n} x_i} = (\bar{x})^{-1}$.

For illustration, let the time in minutes required to serve a customer at certain facility have an Exponential Distribution with unknown Parameter θ. If the average time required to serving a random sample of 20 customers is observed to be 3.8 minutes. Obviously, the Posterior Distribution for the Parameter θ under the Jeffreys Prior, as derived in Section 5.1, is the Gamma with Parameters 20 and 20 × 3.8 = 76, i.e., $\theta|X \sim g(20, 76)$ and the Posterior Bayes estimator under the Squared-Error Loss Function is $(\bar{x})^{-1}$, i.e., 0.263158.
Using the numerical estimation criteria explained in Section 4, we run a set of *C* codes to get $E_{\theta|x}(\theta) \equiv (\bar{x})^{-1} =$ 0.263158. Even for complex Posterior Distributions, the numerical estimation criteria give good results.

5.2 Two-Parameter Distribution: Similarly, if we consider the Normal Distribution with Parameters mean μ and variance δ^2, both unknown, with density $f(x|\mu, \delta^2) = 1/\sqrt{2\pi\delta^2} . \exp\{-(x-\mu)^2/(2\delta^2)\}$, $-\infty \leq \mu \leq \infty$, $\delta > 0$, for $-\infty \leq x \leq \infty$; and zero elsewhere. For a sample of size n, the Likelihood Function is $L(X;\mu,\delta^2) = (2\pi\delta^2)^{-\frac{n}{2}}\exp\{-\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\delta^2}\}$ and the logarithm of the Likelihood Function is $l(X;\mu,\delta^2) = -\frac{n}{2}\ln(2\pi\delta^2) - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\delta^2}$, which implies $\frac{\partial^2 l(\boldsymbol{X};\theta)}{\partial\mu^2} = -\frac{n}{\delta^2}$, $\frac{\partial^2 l(\boldsymbol{X};\theta)}{\partial\mu\partial\delta^2} = -(n\mu - \sum_{i=0}^{n} x_i)/(\delta^2)^2$, and $\frac{\partial^2 l(\boldsymbol{X};\theta)}{\partial(\delta^2)^2} = -\frac{n}{2(\delta^2)^2} + (\delta^2)^{-3}\sum_{i=1}^{n}(x_i-\mu)^2$.

The partial double-derivative vanishes under expectation and Fisher's information, therefore, takes the form $I(\mu,\delta^2) = n^2(\delta^2)^{-3}$. The Jeffreys Prior is derived to be $J(\mu,\delta^2) \propto (\delta^2)^{-3/2}$ for $-\infty \leq \mu \leq \infty, \sigma > 0$. The Joint Posterior Distribution, $f(\mu,\delta^2|X)$, takes the form

$$f(\mu,\delta^2|X), \propto (\delta^2)^{-(n+3)/2}\{-\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\delta^2}\}, \quad \text{for } -\infty \leq \mu \leq \infty,\ \sigma > 0. \quad (5.1)$$

After some algebra, it can be shown that the Posterior Marginal Distribution of precision δ^2 follows the Gamma Distribution and the Posterior Conditional Distribution of $\mu|\delta^{-2}$ follows the Normal Distribution.

For illustration, let an observed sample with 7 values, 20.87, 18.83, 21.36, 17.77, 18.97, 26.66 and 24.24, be taken from the Normal Distribution $N(\mu,\sigma^2)$ with both the Parameters unknown. The observed mean μ and variance σ^2 are found to be 21.24286 and 10.25492 respectively. We now utilize the estimation criteria of Section 4 and find the Posterior Estimates of mean μ and variance σ^2 by running C codes. The desired Estimates of mean μ and variance σ^2 are found to be 21.2429 and 8.65912 respectively. The difference in variance may be due to the short size of dataset. This difference becomes negligible for datasets of large sizes.

For the simulated dataset of size 100 from the Normal Distribution with mean 0 and variance 1, we get $\sum x_i = -10.266$ and $\sum x_i^2 = 124.7118$ with mean $\mu = -0.10266$ and variance $\sigma^2 = 1.24907$. The estimated mean and variance through the suggested criteria are found to be -0.10266 and 1.23617 respectively which are fairly close to the theoretical results.

6. Concluding Remarks

In this article, an effort is made to elaborate on the numerical methods to find the Jeffreys Prior and the Posterior Bayes Estimates. Numerical differentiation and Quadrature are considered to find the Jeffreys Prior and the Posterior Bayes Estimates. For instance of a one-Parameter case, we used the Exponential Distribution to derive the Jeffreys Prior and the Posterior Estimates, whereas for the two-Parameter case, we used the Normal Distribution. The observed and simulated datasets are studied. It is seen that the theoretical and numerical results fairly agree.

The same procedure can easily be employed for the case of uninformative uniform, informative and conjugate priors too. The method works equally well when priors and datasets are assumed to follow non regular density functions. The complicated Posterior Distributions can also be handled with equal ease and accuracy.

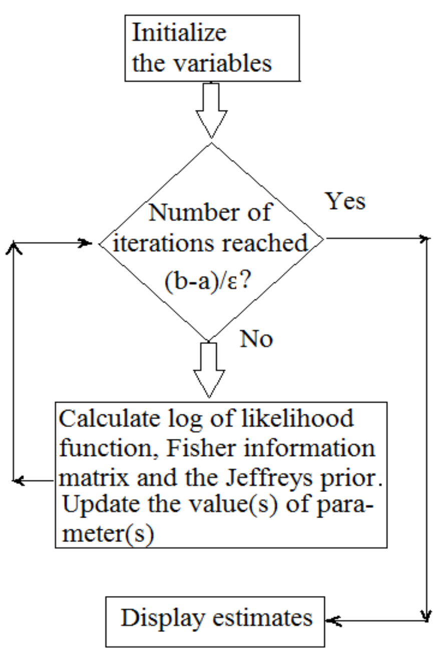

Figure 1: The numerical-estimation procedure

References

1. Adams, E. S. (2005). Bayesian Analysis of linear dominance hierarchies. *Animal Behaviour*, **69**, 1191-1201.
2. Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. 2nd Edition. Springer-Verlag, Inc. New York.
3. Berger, J. O. and Bernardo, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *Journal of American Statistical Association*, **84**, 200-207.
4. Berger, J. O. and Bernardo, J. M. (1992a). *On the Development of Reference Priors, Bayesian Statistics (4th edition)*. Bernardo, J. M., Berger, J. O. Dawid, A. P. and Smith, A. M. F., Valencia, Spain: University Press 35-60.
5. Berger, J. O. and Bernardo, J. M. (1992b). Ordered group reference priors with application to multibinomial problem. *Biometrika*, **70**, 25-38.
6. Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference (with discussion). *Journal of the Royal Statistical Society, Serial B*, **41**, 113-147.
7. Datta, G. S. and Ghosh, M. (1995). Some remarks on noninformative priors. *Journal of the American Statistical Association*, **90(432)**, 1357-1363.
8. Jeffreys, H. (1946). An invariant form for the prior probability estimation problems. *Proceeding of the Royal Society of London, Series A*, **186**, 453-461.
9. Jeffreys, H. (1961). *Theory of Probability*. Claredon Press, Oxford UK.