

Application of Bayesian Computational Techniques in Estimation of Posterior Distributional Properties of Lognormal Distribution

Munni Begum¹⁾ · M. Masoom Ali²⁾

Abstract

In this paper we presented a Bayesian inference approach for estimating the location and scale parameters of the lognormal distribution using iterative Gibbs sampling algorithm. We also presented estimation of location parameter by two non iterative methods, importance sampling and weighted bootstrap assuming scale parameter as known. The estimates by non iterative techniques do not depend on the specification of hyper parameters which is optimal from the Bayesian point of view. The estimates obtained by more sophisticated Gibbs sampler vary slightly with the choices of hyper parameters. The objective of this paper is to illustrate these tools in a simpler setup which may be essential in more complicated situations.

KeyWords : Gibbs Sampling Algorithm, Importance sampling, Lognormal distribution, Weighted bootstrap

1. Introduction

Bayesian inference approach has now encompassed almost all branches of statistical science. In most Bayesian inference problems, the joint posterior distribution of the parameters of interest does not have a closed analytic form. It is the normalizing constant of the posterior distribution that usually does not have a closed form. In such cases, computation for the posterior summarizations becomes tedious. One of the simplest solutions to such problem is to use a large sample approximation to the posterior distribution obtained by expanding it in a

1) First Author : Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420

2) Department of Mathematical Sciences, Ball State University , Muncie, IN 47306-0490
E-mail : mali@bsu.edu

Taylor series. However, this approach requires that the 'unnormalized' posterior has a unique mode and also it gives only first order approximation. Tierney and Kadane (1986) and Tierney, Kass and Kadane (1989) presented a more accurate second order approximation for posterior expectations using Laplace's method.

Noniterative Monte Carlo methods such as, importance sampling, rejection sampling and weighted bootstrap depend on an appropriate importance or envelop function. In particular, for many high dimensional problems, it is quite difficult to find an acceptably accurate importance or envelop function from which samples could still be drawn easily. Importance sampling approach was outlined by Hammersley and Handscombe (1964) and later championed for Bayesian analysis by Geweke (1989). Excellent summaries on rejection sampling can be found in the books by Riply (1987) and Devroye (1986). In subsequent sections, we presented applications of the importance sampling, weighted bootstrap and the Gibbs sampling techniques. The iterative Gibbs sampling technique originally introduced by Geman and Geman (1984), was exposed to statisticians by Gelfand and Smith (1990) in the context of Bayesian inference. They illustrated the use of the Gibbs sampler as a method of calculating Bayesian marginal posterior and predictive densities for normal data. Our objective is to explore the usefulness of this powerful tool to other frequently used distributions. One such distribution is the lognormal distribution which belongs to log-location-scale family. These are also known as lifetime distribution for their wide scale application in survival data analysis. In sections 2 and 3 an overview of the Gibbs sampling technique and its application to lognormal distribution in estimating posterior characteristics are presented. Section 4 presents a numerical illustration of these techniques.

2. Gibbs Sampling

Under standard notations, we denote densities as $[.]$ so that $[Y_1, Y_2]$, $[Y_1|Y_2]$ and $[Y_1]$ stand for the joint, conditional and marginal densities. We assume that the full conditional distributions $\{p(y_i|y_j, j \neq i), i = 1, \dots, k\}$ are available for sampling, meaning that samples may be generated by some method. Under mild conditions (Besag, 1974) the one dimensional conditional distributions uniquely determine the joint distribution $[Y_1, Y_2, \dots, Y_k]$ and all the marginal distributions $[Y_i], i = 1, \dots, k$. The Gibbs sampling algorithm proceeds as follows:

Step 0 We consider a set of arbitrary starting values $(Y_1^{(0)}, \dots, Y_k^{(0)})$

Step 1 Then draw $Y_1^{(1)}$ from $[Y_1|Y_2^{(0)}, \dots, Y_k^{(0)}]$.

Step 2 Draw $Y_2^{(1)}$ from $[Y_2|Y_1^{(1)}, Y_3^{(0)}, \dots, Y_k^{(0)}]$.

Step k Draw $Y_k^{(1)}$ from $[Y_k | Y_1^{(1)}, Y_2^{(1)}, \dots, Y_{k-1}^{(1)}]$.

After 1 iteration, we have $(Y_1^{(1)}, \dots, Y_k^{(1)})$ and after t such iterations, we obtain $(Y_1^{(t)}, \dots, Y_k^{(t)})$ which is our t -th Gibbs sample. Geman and Geman (1984) showed that $(Y_1^{(t)}, \dots, Y_k^{(t)}) \rightarrow [Y_1, \dots, Y_k]$ in distribution as $t \rightarrow \infty$ and the convergence is exponential in t . In the context of Bayesian analysis, we are required to sample from the joint posterior distribution $[Y_1, \dots, Y_k | x]$. But with the use of Gibbs sampler it suffices to draw from each of the univariate conditional distributions only. Then an estimate of marginal density of Y_i can be obtained from each of the univariate conditional distributions $p(y_i | y_j, i \neq j)$. Subsequently the distributional properties can easily be estimated. For example, a marginal density estimate of the marginal density of Y_i is given by,

$$\hat{p}(y_i | x) = \frac{1}{m} \sum_{j=1}^m P(y_i | y_{k,j}^{(t)}, k \neq i; x). \quad (2.1)$$

The marginal density estimated by this method is more reliable than that estimated by other methods such as kernel density estimation.

3. The Lognormal Distribution

The lognormal distribution with two parameters may be defined as the distribution of a random variable whose logarithm is normally distributed. Such a variable is necessarily positive and in real life from the sizes of organisms to the number of species in Biology, incomes in Economics and survival/failure times in clinical trials, variates are inherently positive. Also it is very common to use a logarithmic transformation of a variable in data analysis when it deviates considerably from normality. The parameter estimates from the inverse transformation is biased, which makes the use of normal distribution less attractive for inference purposes. An elaborate presentation of the theory and application on lognormal distribution can be found in Crow and Shimizu (1988). In this paper we presented a Bayesian inference approach for the location and scale parameters of the lognormal distribution using iterative Gibbs sampling algorithm. We also presented estimation of location parameter only by two non iterative methods, importance sampling and weighted bootstrap. Scale parameter is assumed to be known to keep the computation simple.

3.1 Posterior Computations for Lognormal Distribution

We consider a positive random variable Y to be lognormally distributed with parameters μ and σ^2 if $X = \ln Y$ is normally distributed with mean μ and variance σ^2 . Then the likelihood function for μ and σ^2 can be written as,

$$L(\mu, \tau) = (2\pi)^{-\frac{n}{2}} \left(\prod_{i=1}^n Y_i \right)^{-1} (\tau)^{\frac{n}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{i=1}^n (\ln Y_i - \mu)^2 \right\}, \quad (3.1)$$

where $\tau = \frac{1}{\sigma^2}$. To carry out Bayesian inference, we can elicit the joint prior distribution for the parameters μ and τ , $\pi(\mu, \tau) = \pi(\mu|\tau)\pi(\tau)$ as follows:

$$\begin{aligned} \pi(\mu|\tau) &\sim N(\mu_0, \tau^{-1}\sigma_0^2), \\ \pi(\tau) &\sim G\left(\frac{\delta_0}{2}, \frac{\gamma_0}{2}\right), \end{aligned}$$

where μ_0 , σ_0^2 , δ_0 and γ_0 are hyper parameters, N and G stand for the Gaussian and Gamma distributions, respectively. For such a prior elicitation, the joint posterior distribution can be written up to the proportionality constant as:

$$\pi(\mu, \tau|y) \propto \tau^{\frac{n}{2} + \frac{\delta_0}{2} - \frac{1}{2}} \exp \left\{ -\frac{\tau}{2} \left[\sum_{i=1}^n (\ln Y_i - \mu)^2 + \frac{1}{\sigma_0^2} (\mu - \mu_0)^2 + \gamma_0 \right] \right\}.$$

In this case, the proportionality constant is straightforward to evaluate which may not be the case in high dimensional models. The joint posterior distribution of μ and τ can be evaluated as,

$$\pi(\mu, \tau|y) \propto \frac{1}{k} \tau^{\frac{n}{2} + \frac{\delta_0}{2} - \frac{1}{2}} \exp \left\{ -\frac{\tau}{2} \left[\sum_{i=1}^n (\ln Y_i - \mu)^2 + \frac{1}{\sigma_0^2} (\mu - \mu_0)^2 + \gamma_0 \right] \right\},$$

$$\text{where } k = \frac{(2\pi)^{\frac{1}{2}} (n + \sigma_0^{-2})^{-\frac{1}{2}} \Gamma\left(\frac{n + \delta_0}{2}\right)}{\frac{1}{2} \left[\gamma_0 + \frac{\mu_0^2}{\sigma_0^2} + \sum (\ln Y_i)^2 - \frac{(\mu_0 \sigma_0^{-2} + \sum \ln Y_i)^2}{n + \sigma_0^{-2}} \right]}.$$

The marginal posterior densities of μ and τ are straightforward to obtained as given below:

$$P[\mu|y] \propto \left[1 + \frac{c(\mu - a)^2}{b - ca^2} \right]^{-\frac{n + \delta_0 + 1}{2}},$$

which is the kernel of t distribution with location and dispersion parameters a and $\left[\frac{(n + \delta_0)c}{b - ca^2} \right]^{-1}$, respectively and degrees of freedom $n + \delta_0$. Also, a , b , and c in the above expression is defined as,

$$\begin{aligned} a &= \frac{n\sigma_0^{-2} + \sum_{i=1}^n \ln Y_i}{n + \sigma_0^{-2}}, \\ b &= \gamma_0 + \frac{\mu_0^2}{\sigma_0^2} + \sum_{i=1}^n \ln Y_i, \\ c &= n + \sigma_0^{-2}. \end{aligned}$$

Similarly, the marginal distribution of τ can be obtained as,

$$P[\tau|y] \propto \tau^{\frac{n + \delta_0}{2} - 1} \exp \left\{ -\frac{\tau}{2} \left[\gamma_0 + \frac{\mu_0}{\sigma_0^2} + \sum_{i=1}^n (\ln Y_i)^2 - (n + \sigma_0^{-2})a^2 \right] \right\}.$$

That is,

$$\tau|y \sim G \left[\frac{n + \delta_0}{2}, \frac{1}{2} \left(\gamma_0 + \frac{\mu_0}{\sigma_0^2} + \sum_{i=1}^n (\ln Y_i)^2 - (n + \sigma_0^{-2})a^2 \right) \right].$$

Therefore, inference about the parameters μ and σ^2 can be made from these marginal distributions. In the next two sections we discuss two methods for making inference about μ and σ^2 without requiring to evaluate their joint distributions.

3.2 Non Iterative Techniques

The importance sampling technique starts with selecting an importance density which approximates the posterior density as well as is easy to sample from. We choose t -distribution with 5 d.f. with certain location and scale as the importance density $g(\mu)$, which could be taken as an approximation of the likelihood times

the prior. We define the weight function, $w(\mu)$, as $w(\mu) = \frac{L(\mu)\pi(\mu)}{g(\mu)}$. Next draw μ_1, \dots, μ_N from $g(\mu)$. Then get $\tilde{E}(\mu|y) = \frac{\sum_{i=1}^N \mu_i w(\mu_i)}{w(\mu_i)}$, $\tilde{E}(\mu^2|y) = \frac{\sum_{i=1}^N \mu_i^2 w(\mu_i)}{w(\mu_i)}$ and $\widetilde{Var}(\mu|y) = \tilde{E}[\mu^2|y] - \tilde{E}^2[\mu|y]$.

Like importance sampling, under weighted bootstrap method we draw μ_1, \dots, μ_N from a density $g(\mu)$ that approximates the likelihood times the prior. We consider, $g(\mu)$ as t_5 as well. Now we define $w(\mu_i) = \frac{L(\mu_i)\pi(\mu_i)}{g(\mu_i)}$ and $q_i = \frac{w_i}{\sum_{i=1}^N w_i}$. Finally we draw μ^* from the discrete distribution over μ_1, \dots, μ_N which places mass q_i at μ_i . Then $\tilde{E}(\mu|y) = \frac{1}{N} \sum_{i=1}^N \mu_i^*$ and $\widetilde{Var}(\mu|y) = \frac{1}{N-1} \sum_{i=1}^N (\mu_i^* - \tilde{E}(\mu|y))^2$.

A numerical illustration of importance sampling and weighted bootstrap for estimating the location parameter μ assuming the scale parameter τ known is given in section 4.

3.3 The Gibbs Sampling Algorithm

In order to apply the Gibbs algorithm, which we outlined in section 2, for estimating the parameters μ and σ^2 , we consider the two stage hierarchy for the prior information of the parameters μ and τ . We consider $\tau = \frac{1}{\sigma^2}$ for which prior elicitation is easier. We elicit the prior distributions for μ and τ as,

$$\begin{aligned}\pi(\mu) &\sim N(\mu_0, \sigma_0^2), \\ \pi(\tau) &\sim G\left(\frac{\delta_0}{2}, \frac{\gamma_0}{2}\right).\end{aligned}$$

The full conditional distributions for μ and τ can be derived as under;

$$\begin{aligned}\pi(\mu|\tau, y) &\sim N\left[\frac{\tau \ln Y_i + \mu_0 \sigma_0^{-2}}{\tau + \sigma_0^{-2}}, (\tau + \sigma_0^{-2})^{-1}\right], \\ \pi(\tau|\mu, y) &\sim G\left(\frac{n}{2} + \frac{\delta_0}{2}, \frac{1}{2} \left[\sum_{i=1}^n (\ln Y_i - \mu)^2 + \gamma_0 \right]\right).\end{aligned}$$

Now the algorithm proceeds as follows: Consider an arbitrary starting point

$\{\mu^{(0)}, \tau^{(0)}\}$. Draw $\mu^{(1)} \sim \pi(\mu | \tau^{(0)}, y)$ and $\tau^{(1)} \sim \pi(\tau | \mu^{(1)}, y)$. We repeat these steps until desired samples are drawn. After t iteration we have $(\mu^{(t)}, \tau^{(t)})$. According to Geman and Geman (1984), $(\mu^{(t)}, \tau^{(t)}) \rightarrow [\mu, \tau]$ in distribution as $t \rightarrow \infty$.

In following section we consider a numerical illustration of the techniques presented in subsection 3.1 for estimating the posterior location and scale parameters of lognormal distribution.

4. Numerical Illustration

For importance sampling we simulated samples from lognormal distribution with location 15 and scale 4.5. That is, we generated samples from the distribution,

$$f(y | \mu, \tau) = (2\pi)^{-\frac{1}{2}} (\tau Y_i)^{-1} (4.5)^{\frac{1}{2}} \exp\left\{-\frac{4.5}{2} (\ln Y_i - 15)^2\right\}.$$

We elicit a normal prior for μ assuming τ as known and setting its value to 1, as

$$\pi(\mu | \mu_0) \propto \exp\left\{-\frac{1}{2} (\mu - \mu_0)^2\right\}.$$

So the likelihood times the prior becomes,

$$L(\mu) * \pi(\mu) \propto \exp\left\{-\frac{1}{2} \sum_{i=1}^n (\ln Y_i - \mu)^2\right\} \exp\left\{-\frac{1}{2} (\mu - \mu_0)^2\right\}.$$

The estimation of μ is then carried out through importance sampling technique given in subsection 3.2 assuming a t distribution with 5 d.f as the importance density. In any Bayesian analysis, one should carry out the 'sensitivity analysis' by varying the choice of the hyper parameters in the prior distribution. It enables one to see if there is any impact of these choices on inference. We conducted the same analysis for three sets of hyper parameter, μ_0 in the prior distribution of μ . For the following choices of hyper parameter, μ_0 the estimated posterior means and variances of μ by importance sampling are listed below:

Table 1 Posterior mean and variances for μ by importance sampling

$\pi(\mu \mu_0)$	$\tilde{E}(\mu y)$	$\tilde{V}(\mu y)$
$N(12, 1)$	15.259	$1.1 \times e^{-12}$
$N(10, 1)$	15.009	$5.654 \times e^{-13}$
$N(8, 1)$	14.961	$6.605 \times e^{-7}$

Under same assumptions, the estimated posterior means and variances of μ by weighted bootstrap method are listed below:

Table 2 Posterior mean and variances for μ by weighted bootstrap

$\pi(\mu \mu_0)$	$\tilde{E}(\mu y)$	$\tilde{V}(\mu y)$
$N(12, 1)$	14.958	$1.125 \times e^{-10}$
$N(10, 1)$	14.958	$1.654 \times e^{-11}$
$N(8, 1)$	14.904	$2.605 \times e^{-11}$

We assumed a t distribution with 5 degrees of freedom, location 8.35 and scale 2.25 as the importance density to approximate the likelihood time the prior in both cases. It is to be noted that we do not require to estimate τ for given μ using these non iterative techniques since the likelihood times the prior for τ assuming μ known has a Gamma kernel. So the posterior mean and variance of τ can be obtained easily from the properties of Gamma distribution. Finally to implement the Gibbs algorithm we generated samples from lognormal distribution with location 0.735 and scale 0.389. The prior specifications for μ and τ and the full conditionals are given in the subsection 3.3. We considered five arbitrary sets of starting values for μ and τ as $\{1, 1\}$, $[0.5, 0.5]$, $\{3, 4\}$, $\{2.5, 1.5\}$ and $\{5, 3\}$. Then we ran the Gibbs sampler for 1000 iteration to sample from the full conditionals for each starting values. From each of these five runs we let burn-in phase of 750 samples and retained one fourth of the Gibbs samples (250) giving a total sample size of 1250 for subsequent computations. Then for three sets of hyper parameters for μ and τ the posterior estimates of means and variances of μ and τ are obtained as follows:

Table 3 Posterior mean and variances for μ by Gibbs sampling

μ_0	τ_0	$\tilde{E}(\mu y)$	$\tilde{V}(\mu y)$
2	1	1.987	0.979
3	1.5	3.012	1.548
4.5	2.5	4.509	2.477

Table 4 Posterior mean and variances for τ by Gibbs sampling

δ_0	γ_0	$\tilde{E}(\tau y)$	$\tilde{V}(\tau y)$
2.1	1	0.980	616.868
2.5	1.2	0.141	7.015
1.5	0.5	0.002	0.001

The posterior mean and variances for μ and τ are highly dependent on the choice of the hyper parameters as well as the simulated data. The choices of the hyper parameters were made arbitrarily after running the Gibbs algorithm for several other choices [data not shown].

5. Discussion

In this paper a simple illustration of some heavily involved Bayesian computation techniques has been presented for a distribution that arises frequently in lifetime data analysis. In subsection 3.1 we presented a straight forward computation of the marginal posterior densities of μ and τ after computing their joint posterior distributions. Hence, any inference regarding the parameter of interest could be carried out based on these marginal densities. However, we illustrate the Bayesian computational tools such as, importance sampling, weighted bootstrap and Gibbs sampling for estimating the properties of the Lognormal distribution. In particular, we implement these tools to carry out the posterior estimation of the location and scale parameters of the lognormal distribution without requiring to evaluate the joint posterior distribution of μ and τ . This fact is highly important for estimating parameters in high dimensional models in practical problems. Also application of Gibbs algorithm which suited best in high dimensional cases is very straight forward to implement.

It is to be noted that the estimated posterior mean obtained by non iterative techniques, such as, importance sampling and w weighted bootstrap, do not vary for different choices of hyper parameters. This situation is sometimes referred to as that 'the likelihood is dominated by the data, not by prior information'. In this case, Bayesian and frequentist inference are fairly similar. That is, estimation does

not heavily depend on the prior information. However, we observe reverse scenario in the estimation by iterative Gibbs sampler where the posterior estimates vary with the different choices of hyper parameters, which is also difficult to interpret since the more sophisticated Gibbs algorithm should provide more plausible results. One explanation could be that we specified prior distribution for one parameter assuming the other as known in the non - iterative schemes. This may reduce the variability of estimates for different choices of hyper parameters in the end. And it was not the case with iterative Gibbs sampling where we specified priors for both location and scale parameters. Also convergence issue comes with any iterative procedure. One has to check at least graphically if the algorithm had converged or not. However, the algorithm converged quickly in our simple case[graphs are not provided]. Another limitation to be mentioned here is that the posterior estimates obtained in section 4 depend on the simulation of the sample observations since we did not apply the methods to any real data set. Having reported the limitations we need to mention possible merits of this work as well. In many social and behavioral studies the outcomes of interest are usually skewed and the common practice is to model these using regression procedure under normality assumption on the log transformed outcome variables. Back transformation of the estimates and their standard errors are possible but is difficult to interpret. Bayesian techniques illustrated here for the Lognormal distribution may be an alternative way to estimate the model parameters and the results may be compared to those obtained using regression analysis.

Reference

1. Besag, J., Spatial Interaction and the Statistical Analysis of Lattice Systems (with discussion). *J. Roy. Statist. Soc., Ser. B*, **36**,192-236.
2. Carlin, B.P. and Louis, T.A., *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall(2000).
3. Crow, E.L. and Shimizu, K., *Lognormal Distributions Theory and Applications*. Dekker(1988).
4. Devroye, L., *Non Uniform Random Variate Generation*. Springer-Verlag(1986).
5. Gelfand, A. and Smith, A.F.M., Sampling Based Approaches to Calculating Marginal Densities. *J.Amer.Statist. Assoc.* **85**(1990) 398-409.
6. Gelfand, A. and Smith, A.F.M. et.al., Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling. *J.Amer.Statist. Assoc.* **85**(1990) 972-985.
7. Geman, S. and Geman, D., Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6** (1984) 721-740.

8. Geweke, J., Bayesian Inference in Econometric Models Using Monte Carlo Integration. *Econometrica*. **57**(1989), 1317-1339.
9. Hammersley, J.M. and Handscombe, D.C., *Monte Carlo Methods*. Chapman and Hall (1964).
10. Riply, B.D., *Stochastic Simulation*. John Wiley and Sons (1987).
11. Tierney, L. and Kadane, J.B., Accurate Approximations for Posterior Moments and Marginal Densities. *J.Amer.Statist. Assoc.* **81**(1986) 82-86.
12. Tierney, L., Kass, R.E., and Kadane, J.B., Fully Exponential Laplace Approximations to Expectations and Variances of Non-positive Functions. *J.Amer.Statist. Assoc.* **84**(1989) 710-716.

[received date : Jan. 2004 , accepted date : Feb. 2004]