Bayes Estimation and Prediction of the Two-Parameter Gamma Distribution

Biswabrata Pradhan *& Debasis Kundu †

Abstract

In this article the Bayes estimates of two-parameter gamma distribution is considered. It is well known that the Bayes estimators of the two-parameter gamma distribution do not have compact form. In this paper, it is assumed that the scale parameter has a gamma prior and the shape parameter has any log-concave prior, and they are independently distributed. Under the above priors, we use Gibbs sampling technique to generate samples from the posterior density function. Based on the generated samples, we can compute the Bayes estimates of the unknown parameters and also can construct highest posterior density credible intervals. We also compute the approximate Bayes estimates using Lindley's approximation under the assumption of gamma priors of the shape parameter. Monte Carlo simulations are performed to compare the performances of the Bayes estimators with the classical estimators. One data analysis is performed for illustrative purposes. We further discuss about the Bayesian prediction of future observation based on the observed sample and it is observed that the Gibbs sampling technique can be used quite effectively, for estimating the posterior predictive density and also for constructing predictive interval of the order statistics from the future sample.

Keywords: Maximum likelihood estimators; Conjugate priors; Lindley's approximation; Gibbs sampling; Predictive density, Predictive distribution.

Subject Classifications: 62F15; 65C05

Corresponding Address: Debasis Kundu, Phone: 91-512-2597141; Fax: 91-512-2597500; e-mail: kundu@iitk.ac.in

*SQC & OR Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata, Pin 700108, India

†Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India

1 Introduction

The two-parameter gamma distribution has been used quite extensively in reliability and survival analysis particularly when the data are not censored. The two-parameter gamma distribution has one shape and one scale parameter. The random variable X follows gamma distribution with the shape and scale parameters as $\alpha > 0$ and $\lambda > 0$ respectively, if it has the following probability density function (PDF);

$$f(x|\alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}; \quad x > 0, \tag{1}$$

and it will be denoted by Gamma(α, λ). Here $\Gamma(\alpha)$ is the gamma function and it is expressed as

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx. \tag{2}$$

It is well known that the PDF of Gamma(α, λ) can take different shapes but it is always unimodal. The hazard function of Gamma(α, λ) can be increasing, decreasing or constant depending on $\alpha > 1$, $\alpha < 1$ or $\alpha = 1$ respectively. The moments of X can be obtained in explicit form, for example

$$E(X) = \frac{\alpha}{\lambda}, \quad \text{and} \quad V(X) = \frac{\alpha}{\lambda^2}. \tag{3}$$

A book length treatment on gamma distribution can be obtained in Bowman and Shenton [5], see also Johnson, Kotz and Balakrishnan [11] for extensive references till 1994.

Although, there is a vast literature available on estimation of the gamma parameters using the frequentest approach, but not much work has been done on the Bayesian inference of the gamma parameter(s). Damsleth [7] first showed theoretically, using the general idea of DeGroot that there exists conjugate priors for the gamma parameters. Miller [15] also used the same conjugate priors and showed that the Bayes estimates can be obtained only through numerical integration. Tsionas [18] considered the four-parameter gamma distribution of which (1) is a special case, and compute the Bayes estimates for a specific non-informative prior using Gibbs sampling procedure. Recently, Son and Oh [17] consider the model (1) and compute the Bayes estimates of the unknown parameters using Gibbs sampling procedure, under the vague priors, and compare their performances with the maximum likelihood estimators (MLEs) and the modified moment estimators. Very recently, Apolloni and Bassis [3] proposed an interesting method in estimating the parameters of a two-parameter gamma distribution, based on a completely different approach. The performances of the estimators proposed by Apolloni and Bassis [3] are very similar with the corresponding Bayes estimators proposed by Son and Oh [17].

The main aim of this paper is to use informative priors and compute the Bayes estimates of the unknown parameters. It is well known that in general if the proper prior information is available, it is better to use the informative prior(s) than the non-informative prior(s), see for example Berger [4] in this respect. In this paper it is assumed that the scale parameter has a gamma prior and the shape parameter has any log-concave prior and they are independently distributed. It may be mentioned that the assumption of independent priors for the shape and scale parameters is not very uncommon for the lifetime distributions, see for example Sinha [16] or Kundu [12].

Note that our priors are quite flexible, but in this general set up it is not possible to obtain the Bayes estimates in explicit form. First we propose Lindley's method to compute approximate Bayes estimates. It may be mentioned that Lindley's approximation plays an important role in the Bayesian analysis, see Berger [4]. The main reason might be, using Lindley's approximation it is possible to compute the Bayes estimate(s) quite accurately without performing any numerical integration. Therefore, if one is interested in computing the Bayes estimate(s) only, Lindley's approximation can be used quite effectively for this purpose.

Unfortunately, by using Lindley's method it is not possible to construct the highest posterior density (HPD) credible intervals. We propose to use Gibbs sampling procedure to construct the HPD credible intervals. To use Gibbs sampling procedure, it is assumed that the scale parameter has a gamma prior and the shape parameter has any independent log-concave prior. It can be easily seen that the prior proposed by Son and Oh [17] is a special case of the prior proposed by us. We provide an algorithm to generate samples directly from the posterior density function using the idea of Devroye [8]. The samples generated from the posterior distribution can be used to compute Bayes estimates and also to construct HPD credible intervals of the unknown parameters.

It should be mentioned that our method is significantly different than the methods proposed by Miller [15] or Son and Oh [17]. Miller [15] has worked with the conjugate priors and the corresponding Bayes estimates are obtained only through numerical integration. Moreover, Miller [15] did not report any credible intervals also. Son and Oh [17] obtained the Bayes estimates and the corresponding credible intervals by using the Gibbs sampling technique based on full conditional distributions. Whereas, in this paper we have suggested to generate Gibbs samples directly from the joint posterior distribution function.

We compare the estimators proposed by us with the classical moment estimators and also with the maximum likelihood estimators, by extensive simulations. As expected it is observed that when we have informative priors, the proposed Bayes estimators behave better than the classical maximum likelihood estimators but for non-informative priors their behavior are almost same. We provide a data analysis for illustrative purposes.

Bayesian prediction plays an important role in different areas of applied statistics. We further consider the Bayesian prediction of the unknown observable based on the present sample. It is observed that the proposed Gibbs sampling procedure can be used quite effectively for posterior predictive density of a future observation based on the present sample and also for constructing the associated predictive interval. We illustrate the procedure with an example.

The rest of the paper is organized as follows. In section 2, we provide the prior and posterior distributions. Approximate Bayes estimates using Lindley's approximation and using Gibbs sampling procedures are described in section 3. Numerical experiments are performed and their results are presented in section 4. One data analysis is performed in section 5 for illustrative purposes. In section 6 we discuss the Bayesian prediction problem, and finally we conclude the paper in section 7.

2 Prior and Posterior Distributions

In this section we explicitly provide the prior and posterior distributions. It is assumed that $\{x_1, \ldots, x_n\}$ is a random sample from $f(\cdot|\lambda, \alpha)$ as given in (1). We assume that λ has a prior $\pi_1(\cdot)$, and $\pi_1(\cdot)$ follows Gamma(a, b). At this moment we do not assume any specific prior on α. We simply assume that the prior on α is $\pi_2(\cdot)$ and the density function of $\pi_2(\cdot)$ is log-concave and it is independent of $\pi_1(\cdot)$.

The likelihood function of the observed data is

$$l(x_1, \ldots, x_n|\alpha, \lambda) = \frac{\lambda^{n\alpha}}{(\Gamma(\alpha))^n}\; e^{-\lambda T_1}\; T_2^{\alpha-1}, \tag{4}$$

where $T_1 = \sum_{i=1}^{n} x_i$ and $T_2 = \prod_{i=1}^{n} x_i$. Note that (T_1, T_2) are jointly sufficient for (α, λ). Therefore, the joint density function of the observed *data*, α and λ is

$$l(data, \alpha, \lambda) \propto \frac{1}{(\Gamma(\alpha))^n}\; \lambda^{b+n\alpha-1}\; e^{-\lambda(a+T_1)} T_2^{\alpha-1}\; \pi_2(\alpha). \tag{5}$$

The posterior density function of $\{\alpha, \lambda\}$ given the data is

$$l(\alpha, \lambda|data) = \frac{\frac{1}{(\Gamma(\alpha))^n}\lambda^{a+n\alpha-1}e^{-\lambda(b+T_1)}T_2^{\alpha-1}\pi_2(\alpha)}{\int_0^\infty\int_0^\infty \frac{1}{(\Gamma(\alpha))^n}\lambda^{a+n\alpha-1}e^{-\lambda(b+T_1)}T_2^{\alpha-1}\pi_2(\alpha)d\alpha d\lambda}. \tag{6}$$

From (6) it is clear that the Bayes estimate of $g(\alpha, \lambda)$, some function of α and λ under squared error loss function is the posterior mean, *i.e.*

$$\widehat{g}_B(\alpha, \lambda) = \frac{\int_0^\infty \int_0^\infty g(\alpha, \lambda) \frac{1}{(\Gamma(\alpha))^n} \lambda^{a+n\alpha-1} e^{-\lambda(b+T_1)} T_2^{\alpha-1} \pi_2(\alpha) d\alpha d\lambda}{\int_0^\infty \int_0^\infty \frac{1}{(\Gamma(\alpha))^n} \lambda^{a+n\alpha-1} e^{-\lambda(b+T_1)} T_2^{\alpha-1} \pi_2(\alpha) d\alpha d\lambda}. \tag{7}$$

Unfortunately, (7) cannot be computed for general $g(\alpha, \lambda)$. Because of that we provide two different approximations in the next section.

3 Bayes Estimation

In this section we provide the approximate Bayes estimates of the shape and scale parameters based on the prior assumptions mentioned in the previous section.

3.1 Lindley's Approximation

It is known that the (7) cannot be computed explicitly even if we take some specific priors on α. Because of that Lindley [14] proposed the an approximation to compute the ratio of two integrals such as (7). In this case we specify the priors on α and λ. It is assumed that λ follows Gamma(a, b) and α follows Gamma(c, d) and they are independent. Using the above priors, based on Lindley's approximation, the approximate Bayes estimates of α and λ under the squared error loss function are;

$$\widehat{\alpha}_B = \widehat{\alpha} + \frac{1}{2n(\widehat{\alpha}\psi'(\widehat{\alpha}) - 1)^2} \left[-\psi''(\widehat{\alpha})\widehat{\alpha}^2 + \psi'(\widehat{\alpha})\widehat{\alpha} - 2 \right] + \frac{a + c - 2 - d\widehat{\alpha} - b\widehat{\lambda}}{n(\widehat{\alpha}\psi'(\widehat{\alpha}) - 1)}, \tag{8}$$

$$\begin{aligned} \widehat{\lambda}_B &= \widehat{\lambda} + \frac{\widehat{\alpha}\widehat{\lambda}}{2n(\widehat{\alpha}\psi'(\widehat{\alpha}) - 1)^2} \left[-\psi''(\widehat{\alpha}) + 2(\psi'(\widehat{\alpha}))^2 - \frac{3\psi'(\widehat{\alpha})}{\widehat{\alpha}} \right] \\ &\quad + \frac{\widehat{\lambda}}{n(\widehat{\alpha}\psi'(\widehat{\alpha}) - 1)} \left(\frac{c-1}{\widehat{\alpha}} - d \right) + \frac{\widehat{\lambda}^2 \psi'(\widehat{\alpha})}{n(\widehat{\alpha}\psi'(\widehat{\alpha}) - 1)} \left(\frac{a-1}{\widehat{\lambda}} - b \right), \end{aligned} \tag{9}$$

respectively. Here $\widehat{\alpha}$ and $\widehat{\lambda}$ are the MLEs of α and λ respectively. Moreover, $\psi(x) = \frac{d}{dx} \ln \Gamma(x)$, $\psi'(x)$ and $\psi''(x)$ are its first and second derivatives respectively. The exact derivations of (8) and (9) can be obtained in Appendix A.

Although using Lindley's approximation we can obtain the Bayes estimates, but obtaining the HPD credible intervals are not possible. In the next subsection we propose Gibbs sampling procedure to generate samples from the posterior density function and in turn to compute Bayes estimates and HPD credible intervals.

3.2 Gibbs Sampling Procedure

In this subsection we propose Gibbs sampling procedure to generate samples from the posterior density function (6) under the assumption that λ follows Gamma(a, b) and α has any log-concave density function $\pi_2(\alpha)$ and they are independent. We need the following results for further development.

Theorem 1: The conditional distribution of λ given α and *data* is Gamma$(b + n\alpha, a + T_1)$.

Proof: Trivial and therefore it is omitted.

Theorem 2: The posterior density of α given the *data* is

$$l(\alpha|data) \propto \frac{\Gamma(a + n\alpha)}{(\Gamma(\alpha))^n} \times \frac{T_2^{\alpha-1}}{(b + T_1)^{a+n\alpha}} \times \pi_2(\alpha), \tag{10}$$

and $l(\alpha|data)$ is log-concave.

Proof: The first part is trivial so it is omitted and for the second part see Appendix B.

Now using Theorems 1 and 2 and following the idea of Geman and Geman [10], we propose the following scheme to generate (α, λ) from the posterior density function (10). Once we have the mechanism to generate samples from (10), we can use the samples to compute the approximate Bayes estimates and also to construct the HPD credible intervals.

Algorithm:

- Step 1: Generate α_1 from the log-concave density function (10) using the method proposed by Devroye [8].

- Step 2: Generate λ_1 from Gamma$(a + n\alpha_1, b + T_1)$.

- Step 3: Obtain the posterior samples $(\alpha_1, \lambda_1), \ldots, (\alpha_M, \lambda_M)$ by repeating the Steps 1 and 2, M times.

- Step 4: The Bayes estimates of α and λ with respect to the squared error loss function are

$$\widehat{E}(\alpha|data) = \frac{1}{M}\sum_{k=1}^{M} \alpha_k \quad \text{and} \quad \widehat{E}(\lambda|data) = \frac{1}{M}\sum_{k=1}^{M} \lambda_k$$

respectively. Then obtain the posterior variance of α and λ as

$$\widehat{V}(\alpha|data) = \frac{1}{M}\sum_{k=1}^{M} \left(\alpha_k - \widehat{E}(\alpha|data)\right)^2 \quad \text{and} \quad \widehat{V}(\lambda|data) = \frac{1}{M}\sum_{k=1}^{M} \left(\lambda_k - \widehat{E}(\lambda|data)\right)^2,$$

respectively

- Step 5: To compute the HPD credible interval of α order $\alpha_1, \ldots, \alpha_M$ as $\alpha_{(1)} < \ldots < \alpha_{(M)}$. Then construct all the 100(1-β)% credible intervals of α say

$$(\alpha_{(1)}, \alpha_{[M(1-\beta)]}), \ldots, (\alpha_{([M\beta])}, \alpha_{(M)}).$$

Here $[x]$ denotes the largest integer less than or equal to x. Then the HPD credible interval of α is that interval which has the shortest length. Similarly, the HPD credible interval of λ also can be constructed.

4 Simulation Study

In this section we investigate the performance of the proposed estimators through a simulation study. The simulation study is carried out for different sample size and with different hyper parameter values. In particular we take sample sizes n= 10, 15, 25 and 50. Both non-informative and informative priors are used for the shape and scale parameters. In case of non-informative prior we take $a = b = c = d = 0$. We call it as Prior 0. For the informative prior, we chose $a = b = 5$, $c = 2.25$ and d=1.5. We call it as Prior 1. In all these cases, we generate observation from a gamma distribution with $\alpha = 1.5$ and $\lambda = 1$. We compute the Bayes estimates using squared error loss function in all cases. For a particular sample, we compute Bayes estimate using Lindley's approximation, and Bayes estimate using 10000 MCMC samples. The 95% credible intevals are also computed using the MCMC samples. For comparison purpose we compute moment estimates (ME) and maximum likelihood estimates (MLE) and 95% confidence interval using the observed Fisher information matrix. We report average estimates obtained by all the methods along with mean squared error in parentheses in Table 1. The average 95% confidence intervals and HPD credible lengths are presented in Table 2. Since in all the cases the coverage percentages are very close to the nominal value, they are not reported here. All the results of Tables 1 and 2 are based on 1000 replications.

Table 1: The average values of the moment estimators (ME), maximum likelihood estimators (MLE) and Bayes estimator under Prior 0 and Prior 1 along with the MSE's in parentheses. In each cell, the first and second entry corresponds to α and λ, respectively.

n	ME	MLE	Bayes(Lindley)		Bayes(MCMC)	
			Prior 0	Prior 1	Prior 0	Prior 1
10	2.151(1.308)	1.905(0.746)	1.761(0.579)	1.218(0.119)	1.733(0.533)	1.544(0.089)
	1.488(0.750)	1.348(0.518)	1.246(0.391)	0.806(0.045)	1.216(0.381)	1.063(0.035)
15	1.902(0.649)	1.726(0.335)	1.642(0.271)	1.316(0.092)	1.649(0.314)	1.546(0.088)
	1.305(0.388)	1.186(0.222)	1.128(0.183)	0.854(0.037)	1.137(0.202)	1.046(0.037)
25	1.744(0.320)	1.637(0.162)	1.589(0.141)	1.447(0.059)	1.575(0.156)	1.532(0.065)
	1.169(0.173)	1.100(0.098)	1.068(0.087)	0.957(0.042)	1.079(0.111)	1.035(0.034)
50	1.627(0.129)	1.563(0.067)	1.541(0.062)	1.520(0.031)	1.540(0.060)	1.512(0.026)
	1.086(0.077)	1.049(0.043)	1.035(0.041)	1.015(0.017)	1.050(0.041)	1.020(0.019)

Some of the points are quite clear from Tables 1 and 2. As expected it is observed that as the sample size increases in all the cases the average biases and the mean squared

Table 2: The average 95% confidence intervals and HPD credible intervals.

Estimator	Parameter	n			
		10	15	25	50
MLE	α	(0.372, 3.690)	(0.607, 2.975)	(0.823, 2.533)	(1.009, 2.137)
	λ	(0.108, 2.789)	(0.300, 2.215)	(0.470, 1.836)	(0.616, 1.512)
MCMC (Prior 0)	α	(0.174, 3.373)	(0.304, 2.861)	(0.471, 2.410)	(0.715, 2.069)
	λ	(0.042, 2.561)	(0.113, 2.117)	(0.222, 1.749)	(0.392, 1.470)
MCMC (Prior 1)	α	(0.380, 2.373)	(0.464, 2.281)	(0.580, 2.146)	(0.782, 1.948)
	λ	(0.214, 1.665)	(0.254, 1.591)	(0.322, 1.504)	(0.445, 1.364)

errors decrease. It verifies the consistency properties of all the estimates. In Table 1 it is observed that the performance of Bayes estimates obtained using Lindley's approximation and Gibbs sampling procedure are quite similar in nature. That suggests that Lindley's approximation works quite well in this case. Moreover, the behavior (average biases and the mean squared errors) of the Bayes estimates under Prior 0 are very similar with the corresponding behavior of the MLEs, and they perform better than the moment estimates. The same phenomena were observed by Son and Oh [17]. But while using informative prior (Prior 1), the performance of the Bayes estimates are much better than the corresponding MLEs.

In Table 2 it is observed that the average confidence/ credible lengths decrease as the sample size increases. The asymptotic confidence intervals or the HPD credible intervals are slightly skewed for small sample sizes, but they became symmetric for large sample sizes. The performance of the Bayes estimates behave in a very similar manner with the corresponding MLEs (based on average confidence/ credible lengths and coverage percentages) when non-informative priors are used. But when we use the informative priors, the performance of the Bayes estimates are much better than the corresponding MLEs in terms of the shorter confidence/ credible lengths, although, the coverage percentages are properly maintained. Therefore, it is clear that if we have some prior information, the Bayes estimators and the corresponding credible intervals should be used rather than the MLEs and the associated asymptotic confidence intervals.

5 Data Analysis

In this section we analyze a data set from Lawless [13] to illustrate our methodology. The data on survival times in weeks for 20 male rat that were exposed to a high level of radiation are given below.

152	152	115	109	137	88	94	77	160	165
125	40	128	123	136	101	62	153	83	69

The sample mean and the sample variance are 113.45 and 1280.89 respectively. That gives the moment estimates of α and λ as $\widehat{\alpha}_{ME} = 10.051$ and $\widehat{\lambda}_{ME} = 0.089$. The maximum likelihood estimate of the parameters are $\hat{\alpha}_{MLE} = 8.799$ and $\hat{\lambda}_{MLE} = 0.078$ with the corresponding asymptotic variances as 7.46071 and 0.00061 respectively. Using these asymptotic variances, we obtain the 95% confidence intervals for α and λ as (3.4454, 14.1526) and (0.0296, 0.1264), respectively.

We further calculate the Bayes estimates of the unknown parameters, by using Lindley's approximation and Gibbs sampling procedure discussed before. Since we do not have any prior information we consider non-informative priors only for both the parameters. The Bayes estimates of α and λ based on Lindley's approximation are $\hat{\alpha}_{BL} = 8.391$ and $\hat{\lambda}_{BL} = 0.0740$. The Bayes estimates of the parameters by Gibbs sampling method based on 10000 MCMC samples are $\hat{\alpha}_{MC} = 8.397$ and $\hat{\lambda}_{MC} = 0.071$. The 95% HPD credible intervals for α and λ are (2.8252, 15.0854) and (0.0246, 0.1349), respectively.

As it has been observed in the simulation study, here also it is observed that the Bayes estimates and MLEs are very close to each other and they are different than the moment estimators. One of the natural question is whether gamma distribution fits this data set or not. There are several methods available to test the goodness of fit of a particular model to a given data set. For example Pearson's χ^2 test, and Kolomogorov-Smirnov test are extensively being used in practice. Since, for small sample sizes χ^2 test does not work well, we prefer to use the Kolomogorov-Smirnov test only.

We have computed the Kolomogorov-Smirnov distances between the empirical distribution function and the fitted distribution functions, and the associated p values (reported within brackets) for MEs, MLEs, Bayes (Lindley) and Bayes (GS) and they are 0.148 (0.741), 0.145 (0.760), 0.138 (0.811) and 0.128 (0.873). Therefore, based on the Kolomogorov-Smirnov distances we can say that all the methods work quite well but the Bayes estimates based on Gibbs sampling method performs slightly better than the rest.

6 Bayes Prediction

The Bayes prediction of unknown observable belongs to a future sample based on current available sample, known as informative sample, is an important feature in Bayes analysis, see for example Al-Jarallah and Al-Hussaini [2]. Al-Hussaini [1] provided a number of references on applications of Bayes predictions in different areas of applied statistics. In this section, we mainly consider the estimation of posterior predictive density of a future observation, based on the current *data*. The objective is to provide an estimate of the posterior predictive density function of the future observations of an experiment based on the results obtained from an informative experiment, see for example Dunsmore [9] for a nice discussion on this particular topic.

Let y be a future observation independent of the given *data* $x_1, \ldots, x_n$. Then the posterior predictive density of y given the observed *data* is defined as (see for example Chen Shao and

Ibrahim [6].

$$\pi(y|data) = \int_0^\infty \int_0^\infty f(y|\alpha,\lambda)\pi(\alpha,\lambda|data)d\alpha d\lambda. \tag{11}$$

Let us consider a future sample $\{y_1,\ldots,y_m\}$ of size m, independent of the informative sample $\{x_1,\ldots,x_n\}$ and let $y_{(1)} <,\ldots,< y_{(r)} <,\ldots, y_{(m)}$ be the sample order statistics. Suppose we are interested in the predictive density of the future order statistic $y_{(r)}$ given the informative set of data $\{x_1,\ldots,x_n\}$. If the probability density function of the r th order statistic in the future sample is denoted by $g_{(r)}(\cdot|\ \alpha,\lambda)$, then

$$g_{(r)}(y|\alpha,\lambda) = \frac{m!}{(r-1)!(m-r)!}\left[F(y|\alpha,\lambda)\right]^{r-1}\left[1-F(y|\alpha,\lambda)\right]^{m-r} f(y|\alpha,\lambda), \tag{12}$$

here $f(\cdot|\ \alpha,\lambda)$ is same as (1) and $F(\cdot|\ \alpha,\lambda)$ denotes the corresponding cumulative distribution function of $f(\cdot|\ \alpha,\lambda)$. If we denote the the predictive density of $y_{(r)}$ as $g^*_{(r)}(\cdot|\ \ data)$, then

$$g^*_{(r)}(y|data) = \int_0^\infty \int_0^\infty g_{(r)}(y|\alpha,\lambda)l(\alpha,\lambda|data)d\alpha d\lambda, \tag{13}$$

where $l(\alpha,\lambda|data)$ is the joint posterior density of α and λ as provided in (6). It is immediate that $g^*_{(r)}(y|data)$ cannot be expressed in closed form and hence it cannot be evaluated analytically.

Now we propose a simulation consistent estimator of $g^*_{(r)}(y|data)$, which can be obtained by using the Gibbs sampling procedure described in section 3. Suppose $\{(\alpha_i,\lambda_i), i = 1,\ldots,M\}$ is an MCMC sample obtained from $l(\alpha,\lambda|data)$ using the Gibbs sampling technique described in Section 3.2, then a simulation consistent estimator of $g^*_{(r)}(y|data)$ can be obtained as;

$$\widehat{g}^*_{(r)}(y|data) = \frac{1}{M}\sum_{i=1}^{M} g_{(r)}(y|\alpha_i,\lambda_i). \tag{14}$$

Along the same line if we want to estimate the predictive distribution of $y_{(r)}$, say $G^*_{(r)}(\cdot|data)$, then a simulation consistent estimator of $G^*_{(r)}(y|data)$ can be obtained as

$$\widehat{G}^*_{(r)}(y|data) = \frac{1}{M}\sum_{i=1}^{M} G_{(r)}(y|\alpha_i,\lambda_i), \tag{15}$$

here $G_{(r)}(y|\alpha,\lambda)$ denotes the distribution function of the density function $g_{(r)}(y|\alpha,\lambda)$, *i.e.*

$$\begin{aligned} G_{(r)}(y|\alpha,\lambda) &= \frac{m!}{(r-1)!(m-r)!}\int_0^y \left[F(z|\alpha,\lambda)\right]^{r-1}\left[1-F(z|\alpha,\lambda)\right]^{m-r} f(z|\alpha,\lambda)dz, \\ &= \frac{m!}{(r-1)!(m-r)!}\int_0^{F(y|\alpha,\lambda)} u^{r-1}(1-u)^{m-r}du. \end{aligned} \tag{16}$$

It should be noted that the same MCMC sample $\{(\alpha_i,\lambda_i), i=1,\ldots,M\}$ can be used to compute $\widehat{g}^*_{(r)}(y|data)$ or $\widehat{G}^*_{(r)}(y|data)$ for all y.

Another important problem is to construct a two sided predictive interval of the r-th order statistic $Y_{(r)}$ from a future sample $\{Y_1,\ldots,Y_m\}$ of size m, independent of the informative sample $\{x_1,\ldots,x_n\}$. Now we briefly discuss how to construct a $100\beta\%$ predictive interval for $Y_{(r)}$. Note that a symmetric $100\beta\%$ predictive interval for $Y_{(r)}$ can be obtained by solving the following two equations for the lower bound L and upper bound U, see for example Al-Jarallah and Al-Hussaini [2],

$$\frac{1+\beta}{2} = P\left[Y_{(r)} > L|data\right] = 1 - G^*_{(r)}(L|data), \quad \Rightarrow \quad G^*_{(r)}(L|data) = \frac{1}{2} - \frac{\beta}{2}, \tag{17}$$

$$\frac{1-\beta}{2} = P\left[Y_{(r)} > U|data\right] = 1 - G^*_{(r)}(U|data), \quad \Rightarrow \quad G^*_{(r)}(U|data) = \frac{1}{2} + \frac{\beta}{2}. \tag{18}$$

A one-sided predictive interval of the form (L,∞) with the coverage probability β can be obtained by solving

$$P\left[Y_{(r)} > L|data\right] = \beta \quad \Rightarrow \quad G^*_{(r)}(L|data) = 1-\beta \tag{19}$$

for L. Similarly, a one-sided predictive interval of the form $(0,U)$ with the coverage probability β can be obtained by solving

$$P\left[Y_{(r)} > U|data\right] = 1-\beta \quad \Rightarrow G^*_{(r)}(U|data) = \beta, \tag{20}$$

for U. It is not possible to obtain the solutions analytically. We need to apply suitable numerical techniques for solving these non-linear equations.

6.1 Example

For illustrative purposes, we would like to estimate the posterior predictive density of the first order statistic and also would like to construct a 95% symmetric predictive interval of the first order statistic of a future sample of size 20, based on the observation provided in the previous section.

Using the same 10000 Gibbs sample obtained before, we estimate the posterior predictive density function and also the posterior predictive distribution of the first order statistic as provided in (14) and (15) respectively. They are presented in Figure 1.

As it has been mentioned before that the construction of the predictive interval is possible only by solving non-linear equations (17) and (18). In this case we obtain the 95% symmetric predictive interval of the future first order statistic as (19.534, 80.912).

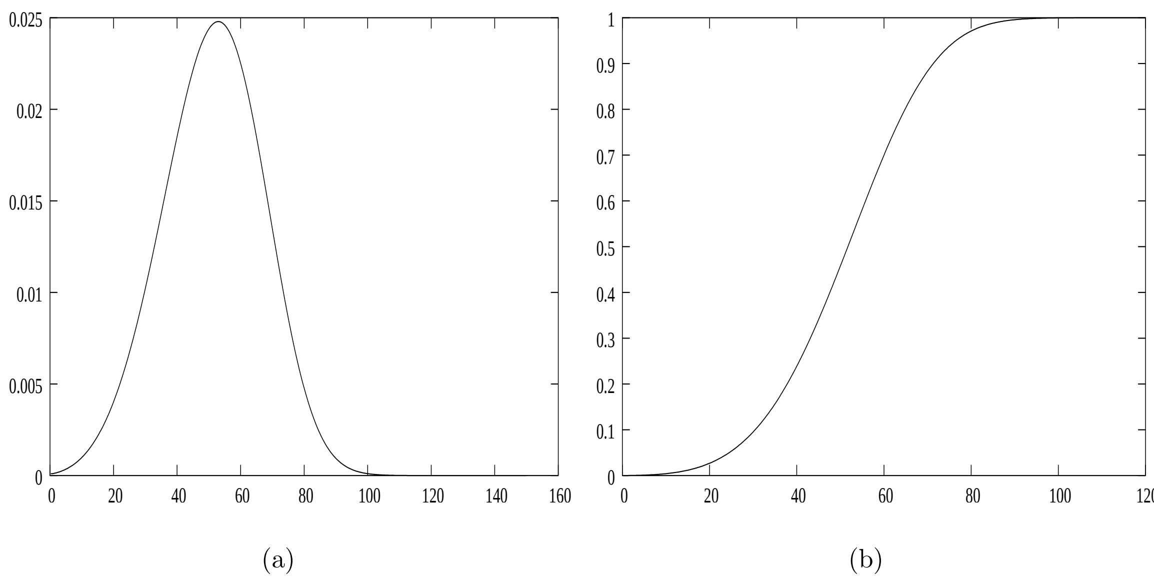

Figure 1: (a) Posterior predictive density function, (b) Posterior predictive distribution function of the first order statistic

7 Conclusions

In this paper we have considered the Bayesian inference of the unknown parameters of the two-parameter gamma distribution. It is an well known problem. It is assumed that the scale parameter has a gamma distribution, the shape parameter has any log-concave density function and they are independently distributed. The assumed priors are quite flexible in nature. We obtain the Bayes estimates and the corresponding credible intervals using Gibbs sampling procedure. Simulation results suggest that the Bayes estimates with non-informative priors behave like the maximum likelihood estimates, but for informative priors the Bayes estimates behave much better than the maximum likelihood estimates. As it had been mentioned before that Son and Oh [17] also considered the same problem and obtained the Bayes estimates and the associate credible intervals using Gibbs sampling technique under the assumption of vague priors. The priors proposed by Son and Oh [17] can be obtained as a special case of the priors proposed by us in this paper. Moreover, Son and Oh [17] generated the Gibbs samples from the full conditional distributions using adaptive rejection sampling technique. Whereas, we have generated the Gibbs samples, directly from the joint posterior density function. It is natural that generating Gibbs samples directly from the joint posterior density function, if possible, is preferable than generating from the full conditional distribution functions.

We have also considered the Bayesian prediction of the unknown observable based on the observed *data*. It is observed that in estimating the posterior predictive density function at any point, the Gibbs sampling procedure can be used quite effectively. Although, for constructing the predictive interval of a future observation, we need to solve two non-linear equations. Efficient numerical procedure is needed to solve these non-linear equations. More work is needed in this direction.

Acknowledgements

The authors would like to thank the referees and the associate editor for many helpful suggestions.

Appendix A

For the two-parameter case, using notation $(\lambda_1, \lambda_2) = (\alpha, \lambda)$, the Lindley's approximation can be written as follows;

$$\widehat{g} = g(\widehat{\lambda}_1, \widehat{\lambda}_2) + \frac{1}{2}\left(A + l_{30}B_{12} + l_{03}B_{21} + l_{21}C_{12} + l_{12}C_{21}\right) + p_1A_{12} + p_2A_{21}, \tag{21}$$

where

$$A = \sum_{i=1}^{2}\sum_{j=1}^{2} w_{ij}\tau_{ij}, \quad l_{ij} = \frac{\partial^{i+j}L(\lambda_1, \lambda_2)}{\partial\lambda_1^i\partial\lambda_2^j}, \quad i, j = 0, 1, 2, 3, \quad \& \quad i + j = 3,$$

$$p_i = \frac{\partial p}{\partial\lambda_i}, \quad w_i = \frac{\partial g}{\partial\lambda_i}, \quad w_{ij} = \frac{\partial^2 g}{\partial\lambda_i\partial\lambda_j}, \quad p = \ln\pi(\lambda_1, \lambda_2), \quad A_{ij} = w_i\tau_{ii} + w_j\tau_{ji},$$

$$B_{ij} = (w_i\tau_{ii} + w_j\tau_{ij})\tau_{ii}, \quad C_{ij} = 3w_i\tau_{ii}\tau_{ij} + w_j(\tau_{ii}\tau_{jj} + 2\tau_{ij}^2).$$

Now,

$$L(\alpha, \lambda) = n\alpha\ln\lambda - \lambda T_1 + (\alpha - 1)\ln T_2 - n\ln\Gamma(\alpha),$$

$$l_{30} = -n\psi''(\widehat{\alpha}), \quad l_{03} = \frac{2n\widehat{\alpha}}{\widehat{\lambda}^3}, \quad l_{21} = 0, \quad l_{12} = -\frac{n}{\widehat{\lambda}^2}.$$

The elements of the Fisher information matrix are;

$$\tau_{11} = \frac{\widehat{\alpha}}{n(\widehat{\alpha}\psi'(\widehat{\alpha}) - 1)}, \quad \tau_{12} = \tau_{21} = \frac{\widehat{\lambda}}{n(\widehat{\alpha}\psi'(\widehat{\alpha}) - 1)}, \quad \tau_{22} = \frac{\widehat{\lambda}^2\psi'(\widehat{\alpha})}{n(\widehat{\alpha}\psi'(\widehat{\alpha}) - 1)}$$

Now when $g(\alpha, \lambda) = \alpha$, then

$$w_1 = 1, \quad w_2 = 0, \quad w_{ij} = 0, \quad \text{for} \quad i, j = 1, 2.$$

Therefore,

$$A = 0, \quad B_{12} = \tau_{11}^2, \quad B_{21} = \tau_{21}\tau_{22}, \quad C_{12} = 3\tau_{11}\tau_{12}, \;\; C_{21} = (\tau_{22}\tau_{11}+2\tau_{21}^2), \;\; A_{12} = \tau_{11} \;\; A_{21} = \tau_{12}.$$

$$p = \ln \pi_2(\alpha) + \ln \pi_1(\lambda) = (a-1)\ln\lambda - b\lambda + (c-1)\ln\alpha - d\alpha \quad \text{and}$$

$$p_1 = \frac{c-1}{\widehat{\alpha}} - d, \quad p_2 = \frac{a-1}{\widehat{\lambda}} - b.$$

Now for the second part when $g(\alpha, \lambda) = \lambda$, then

$$w_1 = 0, \quad w_2 = 1, \quad w_{ij} = 0 \;\; \text{for} \;\; i, j = 1, 2, \quad \text{and}$$

$$A = 0, \quad B_{12} = \tau_{12}\tau_{11}, \quad B_{21} = \tau_{22}^2, \quad C_{12} = \tau_{11}\tau_{22} + 2\tau_{12}^2, \quad C_{21} = 3\tau_{22}\tau_{21}, \quad A_{12} = \tau_{21}, \quad A_{21} = \tau_{22}.$$

Appendix B

$$\ln l(\alpha|data) = k + \ln\Gamma(a+n\alpha) - n\ln\Gamma(\alpha) + \alpha\ln T_2 - (a+n\alpha)\ln(b+T_1) + \ln\pi_2(\alpha). \quad (22)$$

Note that to prove (22) is concave it is enough to show that

$$g(\alpha) = \ln\Gamma(a+n\alpha) - n\ln\Gamma(\alpha)$$

is concave, *i.e.*, $\frac{d^2}{d\alpha^2}g(\alpha) < 0$. Now

$$\frac{d}{d\alpha}g(\alpha) = n\psi(a+\alpha) - n\psi(\alpha)$$

and

$$\begin{aligned} \frac{1}{n} \times \frac{d^2}{d\alpha^2}g(\alpha) &= n\psi'(a+n\alpha) - \psi'(\alpha) \\ &= n(\psi'(a+n\alpha) - \psi'(n\alpha)) + n\psi'(n\alpha) - \psi'(\alpha). \end{aligned}$$

Since $\psi'(\cdot)$ is a decreasing function, therefore, $\psi'(a+n\alpha) - \psi'(n\alpha) \leq 0$. Now observe that

$$\frac{d}{d\alpha}(n\psi'(n\alpha) - \psi'(\alpha)) = n^2\psi''(n\alpha) - \psi''(\alpha) \geq 0,$$

as $\psi''(\cdot)$ is an increasing function. Therefore, $(n\psi'(n\alpha) - \psi'(\alpha))$ is an increasing function in α for all positive integers n. So we have for all $\alpha > 0$,

$$n\psi'(n\alpha) - \psi'(\alpha) \leq \lim_{\alpha \to \infty} (n\psi'(n\alpha) - \psi'(\alpha)).$$

Note that the proof will be complete if we can show that

$$\lim_{\alpha \to \infty} (n\psi'(n\alpha) - \psi'(\alpha)) = 0, \tag{23}$$

as, (23) implies, for all $\alpha > 0$,

$$n\psi'(n\alpha) - \psi'(\alpha) \leq 0.$$

Now (23) is obvious for fixed n, as $\psi'(x) \to 0$ as $x \to \infty$.

References

[1] Al-Hussaini, E. K. (1999), "Predicting observables from a general class of distributions", *Journal of Planning and Statistical Inference*, vol. 79, 79-91.

[2] Al-Jarallah, R. A. and Al-Hussaini E. K. (2007), "Bayes inference under a finite mixture of two-compound Gompertz components model", *Journal of Statistical Computation and Simulation*, vol. 77, 915-927.

[3] Apolloni, B. and Bassis, S. (2009), "Algorithmic inference of two-parameter gamma distribution", *Communications in Statistics - Simulation and Computation*, vol. 38, 1950 - 1968.

[4] Berger, J.O. (1985), *Statistical decision theory and Bayesian analysis*, Springer Verlag, New York.

[5] Bowman, K.O. and Shenton, L.R. (1988), *Properties of estimators for the gamma distribution*, Marcel Decker, New York.

[6] Chen, M.H., Shao, Q.M. and Ibrahim, J.G. (2000), *Monte Carlo Methods in Bayesian Computation*, Springer-Verlag, New York.

[7] Damsleth, E. (1975), "Conjugate classes for gamma distribution", *Scandinavian Journal of Statistics*, vol. 2, 80-84.

[8] Devroye, L. (1984), "A simple algorithm for generating random variates with a log-concave density", *Computing*, vol. 33, 247-257.

[9] Dunsmore, I. R. (1974), "The Bayesian predictive distribution in life testing models", *Technometrics* vol. 16, 455-460.

[10] Geman, S. and Geman, A. (1984), "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, 721 - 740.

[11] Johnson, R.L., Kotz, S. and Balakrishnan, N. (1995), *Continuous Univariate Distribution*, vol 1, 2-nd edition, Wiley, New York.

[12] Kundu, D. (2008), "Bayesian inference and reliability sampling plan for Weibull distribution", *Technometrics*, vol. 50, 144 - 154.

[13] Lawless, J. F. (2003), *Statistical Models and Methods for Lifetime Data*, Second edition, Wiley, New York, pp. 168.

[14] Lindley, D.V. (1980), "Approximate Bayes method", *Trabajos de Estadistica*, vol. 31, 223 237.

[15] Miller, R.B. (1980), "Bayesian analysis of the two-parameter gamma distribution", *Technometrics*, vol. 22, 65-69.

[16] Sinha, S.K. (1986), *Reliability and Life Testing*, John Wiley and Sons, New York.

[17] Son, Y.S. and Oh, M. (2006), “Bayes estimation of the two-parameter gamma distribution”, *Communication in Statistics - Simulation and Computation*, vol. 35, 285-293.

[18] Tsionas, E.G. (2001), “Exact inference in four-parameter generalized gamma distribution”, *Communications in Statistics - Theory and Methods*, vol. 30, 747-756.