

Markov Chain Monte Carlo to Study the Estimation of the Coefficient of Variation

M. A. W. Mahmoud
Mathematics Department,
Faculty of Science, Al-Azhar University,
Nasr City (11884), Cairo, Egypt.

A. H. Abd Ellah
Mathematics Department,
Faculty of Science, Sohag University,
Sohag 82524, Egypt

A. A. Soliman
Mathematics Department,
Faculty of Science,
Islamic University, Madinah, Saudi Arabia.

R. M. El-Sagheer
Mathematics Department,
Faculty of Science, Al-Azhar University,
Nasr City (11884), Cairo, Egypt.

ABSTRACT

The coefficient of variation (CV) of a population is defined as the ratio of the population standard deviation to the population mean. It is regarded as a measure of stability or uncertainty, and can indicate the relative dispersion of data in the population to the population mean. In this article, based on the upper record values, we study the behavior of the CV of a random variable that follows a Lomax distribution. Specifically, we compute the maximum likelihood estimations (MLEs) and the confidence intervals of CV based on the observed Fisher information matrix using asymptotic distribution of the maximum likelihood estimator and also by using the bootstrapping technique. In addition, we propose to apply Markov Chain Monte Carlo (MCMC) techniques to tackle this problem, which allows us to construct the credible intervals. A numerical example based on a real data is presented to illustrate the implementation of the proposed procedure. Finally, Monte Carlo simulations are performed to observe the behavior of the proposed methods.

General Terms:

Computer Science Algorithms

Keywords:

Lomax distribution, Coefficient of variation, Markov chain Monte Carlo, Upper record value, Bootstrap.

1. INTRODUCTION

The coefficient of variation can be used to compare distributions obtained with different units, such as, for example, the variability of the weights of newborns (measured in grams) with the size of adults (measured in centimeters). This approach has been used by several authors to obtain the CV estimates see Pang et al. [26] and Pang et al. [27]. The CV has long been widely used as a descriptive and inferential quantity in several fields such as chemistry, engineering, finance, medical sciences, physics, and telecommunications. In chemical experiments, it is often used as a yardstick of precision

of measurements; two measurement methods may be compared on the basis of their respective CV . In finance, the CV can be used as a measure of relative risks (Miller and Karson [24]): In clinical and diagnostic areas, the CV is also often used as a yardstick of the precision of measurements (Reh and Scheffler [29]). In physiological science, the CV can be applied to assess the homogeneity of bone samples (Hamer et al. [17]). It has been used as a tool in uncertainty analysis of fault trees (Ahn [3]) and in assessing the strength of ceramics Gong and Li [13]. Many statistical procedures concerning CV are based on the normal distribution. However, several phenomena do not agree with the normality assumption due to asymmetry or to the presence of heavy-andlight tails in the distribution of the data. Thus, the statistical inference under normal populations cannot be adequate in the mentioned cases.

The inverse of the coefficient of variation ($1/CV$), called Sharpe's ratio or index, is very popular in finance as a measure of portfolio performance see Knight and Satchell [19]. The wide use of this index stems from the fact that it is an individual measure of performance and not an equilibrium one. Since this is a ratio, Sharpe's index measures the slope of the indifference curve in the mean-standard deviation space so that a higher value of this index implies a higher mean-variance expected utility. For applications of Sharpe's ratio in the analysis of portfolio selection models and in a market risk, see Rachev et al. [28] and Marrison [22]. For a recent illustration of its use in insurance see, Dacorogna and Rüttener [9]. 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. For example, 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 information. A summary of the Bayesian activity is presented by Berger [6].

The Lomax distribution belongs to the class of decreasing failure rate distributions see also Chahkandi and Ganjali, [8]. Sometimes it is called Pareto distribution of the second kind or Pareto Type-II distribution. It was introduced by Lomax [20] as a model for

business failure data. For its applications as lifetime distribution and extensions, we refer to Marshall and Olkin [21]. Bryson [7] has argued that Lomax distributions provide a very good alternative to common lifetime distributions like exponential, Weibull, or gamma distributions when the experimenter presumes that the population distribution may be heavy-tailed. Details on Pareto distributions as well as areas of application can be found in Arnold [4].

Lomax distribution has been shown to be useful for modeling and analyzing the life time data in medical and biological sciences, engineering, etc. So, it has been received the greatest attention from theoretical and applied statisticians primarily due to its use in reliability and lifetesting studies. Many statistical methods have been developed for this distribution, for a review of Lomax distribution see Habibullh and Ahsanullah [15], Upadhyay and Peshwani [32] and Abd Ellah [1,2]. A great deal of research has been done on estimating the parameters of a Lomax using both classical and Bayesian techniques.

Therefore, the purpose of this paper is to develop the Bayes estimates and Markov Chain Monte Carlo (MCMC) techniques to compute the credible intervals and bootstrap confidence intervals of the coefficient of variation CV based on upper record values from the Lomax distribution.

Let $X_{U(1)}, X_{U(2)}, X_{U(3)}, \dots, X_{U(n)}$ be the first upper record value of size n arising from a sequence $\{X_i\}$ of i.i.d Lomax variables with the probability density function pdf

$$f(x) = \alpha\beta^\alpha(x + \beta)^{-(\alpha+1)}, \quad (1)$$

$$x \geq 0, \alpha, \beta > 0.$$

and cumulative distribution function cdf

$$F(x) = 1 - \beta^\alpha(x + \beta)^{-\alpha}, \quad (2)$$

$$x \geq 0, \alpha, \beta > 0,$$

where β is the scale parameter and α is the shape parameter. The Lomax distribution has the following properties

(1) The expected value of X

$$E(X) = \int_0^\infty xf(x) dx$$

$$= \alpha\beta^\alpha \int_0^\infty x(x + \beta)^{-(\alpha+1)} dx \quad (3)$$

$$= \frac{\beta}{(\alpha - 1)}, \quad \alpha > 1$$

(2) The expected value of X^2

$$E(X^2) = \int_0^\infty x^2 f(x) dx$$

$$= \alpha\beta^\alpha \int_0^\infty x^2(x + \beta)^{-(\alpha+1)} dx \quad (4)$$

$$= \frac{2\beta^2}{(\alpha - 1)(\alpha - 2)}, \quad \alpha > 2$$

The theoretical coefficient of variation under the Lomax distribution is thus given by

$$CV = \frac{\sqrt{E(X^2) - (E(X))^2}}{E(X)} \quad (5)$$

$$= \sqrt{\frac{\alpha}{(\alpha - 2)}}, \quad \alpha > 2.$$

The rest of the paper is organized as follows. In Section 2, we discuss the maximum likelihood estimations (MLEs) and the confidence intervals of CV . Parametric bootstrap confidence intervals are discussed in Section 3. Section 4 describes MCMC for estimating the empirical posterior distribution of CV and its interval estimation. Section 5 contains the analysis of a real life data set to illustrate our proposed method. Simulation studies are reported in order to give an assessment of the performance of the different estimation methods in Section 6. Finally we conclude with some comments in Section 7.

2. MLE OF CV

This section describes the ML method for estimating CV based on upper record values from the Lomax distribution. Let $\underline{x} = x_{U(1)}, x_{U(2)}, \dots, x_{U(n)}$ be the first upper record values of size n from Lomax (α, β) . The likelihood function for observed record \underline{x} given by see Arnold et al. [5]

$$\ell(\alpha, \beta|\underline{x}) = f(x_{u(n)}) \prod_{i=1}^{n-1} \frac{f(x_{u(i)})}{1 - F(x_{u(i)})}. \quad (6)$$

From Equations (1), (2), and (6), the likelihood function is given by

$$\ell(\alpha, \beta|\underline{x}) = \alpha^n \beta^\alpha (x_{u(n)} + \beta)^{-\alpha} \prod_{i=1}^n (x_{u(i)} + \beta)^{-1}. \quad (7)$$

The log-likelihood function may then be written as

$$L(\alpha, \beta|\underline{x}) = \log \ell(\alpha, \beta|\underline{x})$$

$$= n \log \alpha + \alpha \log \beta -$$

$$\alpha \log (x_{u(n)} + \beta) - \sum_{i=1}^n \log (x_{u(i)} + \beta), \quad (8)$$

by taking derivative in Equation (8) with respect to α , and β , and equating each result to zero, obtained

$$\frac{\partial L(\alpha, \beta|\underline{x})}{\partial \alpha} = \frac{n}{\alpha} + \log \beta - \log (x_{u(n)} + \beta) = 0 \quad (9)$$

and

$$\frac{\partial L(\alpha, \beta|\underline{x})}{\partial \beta} = \frac{\alpha}{\beta} - \frac{\alpha}{(x_{u(n)} + \beta)} - \sum_{i=1}^n \frac{1}{(x_{u(i)} + \beta)} = 0. \quad (10)$$

From (9), the ML estimate of α denoted by $\hat{\alpha}$ is

$$\hat{\alpha} = \frac{n}{\log (x_{u(n)} + \beta) - \log \beta} \quad (11)$$

By substituting Equation (11) in Equation (10), we get

$$\frac{n}{\hat{\beta} [\log (x_{u(n)} + \hat{\beta}) - \log \hat{\beta}]} -$$

$$\frac{n}{(x_{u(n)} + \hat{\beta}) [\log (x_{u(n)} + \hat{\beta}) - \log \hat{\beta}]} -$$

$$\sum_{i=1}^n \frac{1}{(x_{u(i)} + \hat{\beta})} = 0. \quad (12)$$

Since (12) cannot be solved analytically, thus the MLE estimate $\hat{\beta}$ of the parameter β can be obtained by solving the nonlinear likeli-

hood Equation (12) using, for example, the Newton–Raphson iteration scheme, and hence the corresponding ML estimate $\hat{\alpha}$ of the parameter α is computed from Equation (11) as

$$\hat{\alpha} = \frac{n}{\log(x_{u(n)} + \hat{\beta}) - \log \hat{\beta}}, \quad (13)$$

then, using the invariance property of the ML estimators, the ML estimate of CV , denoted by \hat{CV} , is given by

$$\hat{CV} = \sqrt{\frac{\hat{\alpha}}{(\hat{\alpha} - 2)}}, \hat{\alpha} > 2. \quad (14)$$

2.1 Asymptotic likelihood method

To find an asymptotic variance of the ML estimate \hat{CV} , we shall first write the observed Fisher information matrix $V = (v_{ij})$, where the partial derivatives v_{ij} 's are given by

$$\begin{aligned} v_{11} &= \frac{\partial L^2(\alpha, \beta | \underline{x})}{\partial \alpha^2} \\ &= -\frac{n}{\alpha^2} \end{aligned} \quad (15)$$

$$\begin{aligned} v_{22} &= \frac{\partial L^2(\alpha, \beta | \underline{x})}{\partial \beta^2} \\ &= -\frac{\alpha}{\beta^2} + \frac{\alpha}{(x_{u(n)} + \beta)^2} \\ &\quad + \sum_{i=1}^n \frac{1}{(x_{u(i)} + \beta)^2} \end{aligned} \quad (16)$$

$$\begin{aligned} v_{12} &= v_{21} \\ &= \frac{\partial L^2(\alpha, \beta | \underline{x})}{\partial \alpha \partial \beta} \\ &= \frac{1}{\beta} - \frac{1}{(x_{u(n)} + \beta)}. \end{aligned} \quad (17)$$

An approximate estimate of the variance-covariance matrix of $\hat{\alpha}$ and $\hat{\beta}$ is $\mathbf{V}_{ij}^{-1}|_{\hat{\alpha}, \hat{\beta}}$. In order to find an approximate estimate of the variance of \hat{CV} using the Delta method, see Greene [14], let

$$G' = \left(\frac{\partial CV}{\partial \alpha}, \frac{\partial CV}{\partial \beta} \right), \quad (18)$$

where $\frac{\partial CV}{\partial \alpha}$ and $\frac{\partial CV}{\partial \beta}$ are the first derivatives of the CV with respect to the parameters α and β . The approximate asymptotic variance of \hat{CV} is given by

$$\widehat{Var}(\hat{CV}) \simeq [G' I^{-1} G']|_{\hat{\alpha}, \hat{\beta}}.$$

The asymptotic distribution of the MLE \hat{CV} of CV satisfies:

$$(\hat{CV} - CV) / \sqrt{\widehat{Var}(\hat{CV})} \sim N(0, 1).$$

This yields that the asymptotic $100(1 - \alpha)\%$ confidence interval for CV is given by

$$\hat{CV} \pm Z_{1-\gamma/2} \sqrt{\widehat{Var}(\hat{CV})}. \quad (19)$$

3. PARAMETRIC BOOTSTRAP CONFIDENCE INTERVALS OF CV

This section discuss two confidence intervals based on the parametric bootstrap methods: (i) percentile bootstrap method (Boot-p) based on the idea of Efron [10], (ii) bootstrap-t method (Boot-t) based on the idea of Hall [16]. We illustrate briefly how to estimate confidence intervals of CV using both methods.

1. From the original sample $\{x_{U(1)}, x_{U(2)}, \dots, x_{U(n)}\}$, compute MLEs $(\hat{\alpha}, \hat{\beta})$ of (α, β) and \hat{CV} of CV using Equation (14).
2. Using $\hat{\alpha}$ and $\hat{\beta}$ generate a bootstrap sample $\{x_{U(1)}^*, x_{U(2)}^*, \dots, x_{U(n)}^*\}$. Use this sample to compute the MLE $(\hat{\alpha}^*, \hat{\beta}^*)$ of (α, β) and \hat{CV}^* of CV .
3. Repeat step 2, B boot times, to get parametric bootstrap estimates $\hat{CV}_1^*, \dots, \hat{CV}_B^*$ of CV .

In the following, we propose to use two types of parametric bootstrap confidence intervals for the CV .

3.1 Boot-p

Suppose that $G(z) = P(\hat{CV}^* \leq z)$ be the cumulative distribution function of \hat{CV}^* . Define $\hat{CV}_{Boot-p}^* = G^{-1}(z)$ for given z . The approximate bootstrap $100(1 - \gamma)\%$ confidence interval of CV is given by

$$\left(\hat{CV}_{Boot-p}^* \left(\frac{\gamma}{2} \right), \hat{CV}_{Boot-p}^* \left(1 - \frac{\gamma}{2} \right) \right). \quad (20)$$

3.2 Boot-t

Compute the following statistic $T^* = \frac{\sqrt{B}(\hat{CV}^* - \hat{CV})}{\sqrt{Var(\hat{CV}^*)}}$. For the

T^* values, determine the upper and lower bounds of the $100(1 - \gamma)\%$ confidence interval of CV as follows: let $H(z) = P(T^* \leq z)$ be the cumulative distribution function of T^* . For a given z , define

$$\hat{CV}_{Boot-t}^*(z) = \hat{CV}^* + B^{-1/2} \sqrt{Var(\hat{CV}^*)} H^{-1}(z).$$

The approximate $100(1 - \gamma)\%$ confidence interval of CV is given by

$$\left(\hat{CV}_{Boot-t}^* \left(\frac{\gamma}{2} \right), \hat{CV}_{Boot-t}^* \left(1 - \frac{\gamma}{2} \right) \right). \quad (21)$$

4. BAYES ESTIMATOR OF CV USING MCMC

This section describes Bayesian MCMC methods that have been used to estimate the CV of Lomax distribution. The Bayesian approach is introduced and its computational implementation with MCMC algorithms is described. The two most popular MCMC algorithms are the Gibbs sampler, named by Geman and Geman [11], and the M-H algorithm based on the papers by Metropolis et al. [23] and Hastings [18]. Details of the MCMC methods can be found in Gilks et al. [12]. Using these algorithms it is possible to implement posterior simulation in essentially any problem which allow pointwise evaluation of the prior distribution and likelihood function. Gibbs sampling procedure and the M-H algorithm are used to generate samples from the posterior density function and in turn compute the Bayes point estimates and also construct the corresponding credible intervals based on the generated posterior samples.

By considering model (1), assume the following gamma prior densities for α and β as

$$h_1(\alpha|a, b) = \begin{cases} \frac{b^a}{\Gamma(a)} \alpha^{a-1} \exp(-b\alpha) & \text{if } \alpha > 0 \\ 0 & \text{if } \alpha \leq 0. \end{cases} \quad (22)$$

and

$$h_2(\beta|c, d) = \begin{cases} \frac{d^c}{\Gamma(c)} \beta^{c-1} \exp(-d\beta) & \text{if } \beta > 0 \\ 0 & \text{if } \beta \leq 0. \end{cases} \quad (23)$$

Multiplying $h_1(\alpha|a, b)$ by $h_2(\beta|c, d)$ we obtain the joint prior density of α and β ; given by

$$h(\alpha, \beta) = \frac{b^a d^c}{\Gamma(a)\Gamma(c)} \alpha^{a-1} \beta^{c-1} \exp(-b\alpha - d\beta) \quad (24)$$

Based on the likelihood function of the observed sample is same as (8) and the joint prior in (24), the joint posterior density of α and β given the data is

$$h^*(\alpha, \beta|\underline{x}) = \frac{L(\text{data}|\alpha, \beta) \times h(\alpha, \beta)}{\int_0^\infty \int_0^\infty L(\text{data}|\alpha, \beta) \times h(\alpha, \beta) d\alpha d\beta}, \quad (25)$$

therefore, the Bayes estimate of any function of α and β say $g(\alpha, \beta)$, under squared error loss function is

$$\begin{aligned} \tilde{g}(\alpha, \beta) &= E_{\alpha, \beta|\text{data}}(g(\alpha, \beta)) \\ &= \frac{\int_0^\infty \int_0^\infty g(\alpha, \beta) L(\text{data}|\alpha, \beta) h(\alpha, \beta) d\alpha d\beta}{\int_0^\infty \int_0^\infty L(\text{data}|\alpha, \beta) h(\alpha, \beta) d\alpha d\beta}. \end{aligned} \quad (26)$$

Generally, the ratio of two integrals given by (26) can not be obtained in a closed form. In this case, we use the MCMC method to generate samples from the posterior distributions and then compute the Bayes estimator of $g(\alpha, \beta)$ under the squared errors loss (SEL) function. Therefore, we propose the approaches of MCMC technique to approximate (26). See, for example, Robert and Casella [31] and Recently, Rezaei et al. [30].

4.1 MCMC technique

In this subsection we consider the MCMC method to generate samples from the posterior distributions and then compute the Bayes estimates of CV of the Lomax distribution under the squared errors loss (SEL) function. A wide variety of MCMC schemes are available, and it can be difficult to choose among them. An important sub-class of MCMC methods are Gibbs sampling and more general Metropolis-within-Gibbs samplers. The advantage of using the MCMC method over the MLE method is that we can always obtain a reasonable interval estimate of the parameters by constructing the probability intervals based on the empirical posterior distribution. This is often unavailable in maximum likelihood estimation. Indeed, the MCMC samples may be used to completely summarize the posterior uncertainty about the parameters and, through a kernel estimate of the posterior distribution. This is also true of any function of the parameters, CV in particular. Suppose we wish to give point and interval estimates for CV .

The expression for the joint posterior can be obtained up to proportionality by multiplying the likelihood with the joint prior and this

can be written as

$$h^*(\alpha, \beta) \propto \alpha^{n+a-1} \beta^{\alpha+c-1} \times \exp[-(b\alpha + d\beta + \alpha \log(x_{u(n)} + \beta))] \times \prod_{i=1}^n (x_{u(i)} + \beta)^{-1}, \quad (27)$$

from (27) it is clear that the posterior density function of α given β is proportional to

$$h_1^*(\alpha|\beta) \propto \alpha^{n+a-1} \times \exp -\alpha [b + \log(x_{u(n)} + \beta) - \log \beta]. \quad (28)$$

Therefore, the posterior density function of α given β , is gamma with parameters $(n + a)$ and $(b + \log(x_{u(n)} + \beta) - \log \beta)$. The posterior density function of β given α can be written as

$$h_2^*(\beta|\alpha) \propto \beta^{\alpha+c-1} \times \exp \left[-d\beta - \alpha \log(x_{u(n)} + \beta) - \sum_{i=1}^n \log(x_{u(i)} + \beta) \right]. \quad (29)$$

The posterior distribution of β given α Equation (29) cannot be reduced analytically to well known distributions and therefore it is not possible to sample directly by standard methods. We, employ the Metropolis-within-Gibbs method instead to sample. The choice of the hyperparameters (a, b, c, d) are which make (29) close to the proposal distribution and obviously more convergence of the MCMC iteration. We propose the following MCMC algorithm to draw samples from the posterior density functions; and in turn compute the Bayes estimates and also, construct the corresponding credible intervals of CV .

Algorithm 1.

1. $\beta_0 = \hat{\beta}$, $M = n\text{burn}$.
2. Generate α_1 from gamma distribution $h_1^*(\alpha|\beta)$.
3. Generate β_1 from $h_2^*(\beta|\alpha)$ using (M-H) algorithm see Metropolis et al. [23].
4. Compute CV_1 from (14).
5. Repeat steps 2-4 N times we obtain CV_1, CV_2, \dots, CV_N .
6. Obtain the Bayes estimate of CV with respect to the SEL function as

$$\hat{E}(CV|\text{data}) = \frac{1}{N - M} \sum_{i=M+1}^N CV_i. \quad (30)$$

7. To compute the credible intervals of CV , order $CV_{M+1}, CV_{M+2}, \dots, CV_N$ as

$CV_{(1)}, CV_{(2)}, \dots, CV_{(N-M)}$ Then the $100(1 - \gamma)\%$ symmetric credible intervals

$$(CV_{((N-M)\gamma/2)}, CV_{((N-M)(1-\gamma/2))}). \quad (31)$$

5. ILLUSTRATIVE EXAMPLE

To illustrate the inferential procedures developed in the preceding sections, we choose the real data set which was also used in Lawless (1982-pp 185), these data are from Nelson [25] concerning the data on time to breakdown of an insulating fluid between electrodes at a voltage of 34 K. V. (minutes). The 19 times to breakdown are:

0.96	4.15	0.19	0.78	8.01	31.75	7.35
6.50	8.27	33.91	32.52	3.16	4.85	2.78
4.67	1.31	12.06	36.71	72.89		

Therefore, we observe the upper record values from the observed data as follows: 0.96, 4.15, 8.01, 31.75, 33.91, 36.71, 72.89. A model suggested by engineering considerations is that, for a fixed voltage level, time to breakdown has a Lomax distribution. Based on these seven upper record values, we run the chain for 10 000 times and discard the first 1000 values as 'burn-in'. When the non-informative prior distribution is used, the joint posterior distribution of the parameters is proportional to the likelihood function. The Bayes point estimates and 95% credible intervals for CV are computed. Also performed a simple bootstrap procedure to generate the sampling distribution of CV based on the observed seven upper record values. The point estimates of CV using the maximum likelihood method and bootstrap methods as well as 95% bootstrap confidence interval are presented in Table 1. If we adopt the Bayesian approach, we have the results in Table 2. The descriptive statistics as well as the 95% credible interval for CV based on the MCMC generated sample are also given in Table 2. As we can see from the histograms of the posterior distributions of CV in Figure 1.

Table 1. MLE and Bootstrap results of CV

Method	\hat{CV}	95% C.I.	Length
MLE	1.7071	[-0.7929, 4.2071]	5.0000
Boot-p	1.9469	[1.2078, 5.1171]	3.9093
Boot-t	1.8439	[0.0896, 3.9770]	3.8874

Table 2. MCMC descriptive statistics of CV

Mean	Median	Mode	SD
1.9667	1.7969	1.7573	1.0379
Skewness	95% C. I.	Length	
4.5488	[1.1282, 5.4117]	4.2835	

where 95% C. I. means that 95% Credible Interval.

6. MONTE CARLO SIMULATIONS

In this section, we report some numerical experiments performed to evaluate the behavior of the MCMC methods for different upper record samples from a Lomax distribution, different parameter values, and different priors. Then we use MCMC method to estimate the CV based on 10 000 samples and discard the first 1000 values as 'burn-in'. Thus we can construct the 95% confidence intervals. The simulation runs then repeat $N = 1000$ times. We compute the average Bayes estimates, mean squared errors (MSEs), coverage percentages (C.P), and the average lengths (A.L) based on N replications.

In our present study, we set the different sample sizes n , we also, used two sets of parameter values $\alpha = 3, \beta = 3.1$ and $\alpha = 4.2, \beta = 3.1$. We used different hyperparameters (a, b, c, d), mainly to explore their effects on the estimates of CV . First, we used the noninformative gamma priors for both shape parameters, ($a = b = c = d = 0$). Note that as the hyperparameters go to 0, the prior density becomes inversely proportional to its argument and also becomes improper. This density is commonly used

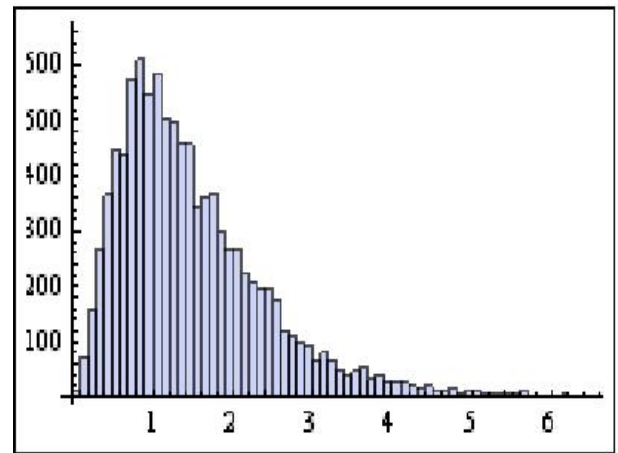


Fig. 1. Histogram of CV

as an improper prior for parameters in the range of 0 to infinity, and this prior is not specifically related to the gamma density. Other than noninformative prior, we also used an informative prior ($a = 5, b = c = d = 0.5$) for the two sets of parameter values. The simulation results are presented in Tables 3 and 4.

Table 3. MCMC estimates of the CV with MSE, average length of 95% C.I and C.P when $\alpha = 3$ and $\beta = 3.1$.

n	Mean	MSE	A.L	C.P
Noninformative prior				
4	1.7369	0.1399	3.3707	0.993
5	1.7451	0.1398	3.4572	0.966
6	1.7544	0.1395	3.4641	0.980
7	1.7820	0.1382	3.4889	0.969
8	1.7933	0.1343	3.5198	0.978
9	1.7964	0.1316	3.5675	0.985
10	1.8168	0.1299	3.5866	0.995
12	1.7960	0.1238	3.6658	0.975
15	1.7970	0.1189	3.6797	0.960
17	1.7572	0.1014	3.7353	0.955
20	1.7418	0.0939	3.8963	0.950
22	1.7289	0.0754	3.9179	0.965
23	1.7319	0.0678	3.9421	0.970
25	1.7422	0.0564	3.9581	0.965
Informative prior				
4	1.6427	0.1234	2.3413	0.994
5	1.6993	0.1224	2.3580	0.963
6	1.6741	0.1221	2.3974	0.977
7	1.7051	0.1219	2.4182	0.990
8	1.7081	0.1162	2.4231	0.995
9	1.7076	0.1124	2.4505	0.975
10	1.6813	0.0931	2.5001	0.970
12	1.7584	0.0882	2.5456	0.950
15	1.8169	0.0874	2.5645	0.940
17	1.7849	0.0823	2.5712	0.930
20	1.8106	0.0734	2.6260	0.935
22	1.8183	0.0675	2.6796	0.949
23	1.7712	0.0612	2.7506	0.960
25	1.7482	0.0458	2.7770	0.950

Table 4. MCMC estimates of the *CV* with MSE, average length of 95% C.I and C.P when $\alpha = 4.2$ and $\beta = 3.1$.

n	Mean	MSE	A.L	C.P
Noninformative prior				
4	1.5712	0.1304	2.3893	0.967
5	1.5728	0.1289	2.4152	0.970
6	1.5765	0.0910	2.5231	0.966
7	1.5559	0.0883	2.5743	0.999
8	1.5622	0.0841	2.6306	0.987
9	1.5777	0.0812	2.6951	0.971
10	1.5118	0.0809	2.7107	0.980
12	1.5003	0.0799	2.7327	0.970
15	1.5121	0.0791	2.7807	0.955
17	1.5008	0.0722	2.8484	0.970
20	1.4779	0.0705	2.8883	0.960
22	1.4824	0.0686	2.9101	0.961
23	1.4928	0.0627	2.9463	0.945
25	1.4630	0.0573	2.9660	0.955
Informative prior				
4	1.4445	0.0320	1.5745	0.970
5	1.4856	0.0317	1.6893	0.968
6	1.4896	0.0316	1.7103	0.956
7	1.4565	0.0313	1.7576	0.998
8	1.4533	0.0312	1.7896	0.990
9	1.4644	0.0303	1.7939	0.952
10	1.4159	0.0301	1.8167	0.995
12	1.4454	0.0277	1.8917	0.985
15	1.4169	0.0268	1.9346	0.965
17	1.4068	0.0253	1.9388	0.945
20	1.4075	0.0226	1.9750	0.965
22	1.4119	0.0220	1.9835	0.970
23	1.4118	0.0217	2.0358	0.955
25	1.4039	0.0202	2.0226	0.960

7. CONCLUSIONS

This article study the estimation of the *CV* of a random variable that follows a Lomax distribution, when the available data are upper record values. Maximum likelihood, parametric bootstrap and MCMC Bayes methods are proposed, to obtain point estimates, as well as confidence intervals of the *CV*. We introduced illustrated example using the observed sample of real data. We found that the Bayes estimates cannot be obtained in explicit form. We used the MCMC technique to compute the approximate Bayes estimates and the corresponding credible intervals.

From results in Tables 1 and 2, we observed that most of the methods work well in general. The MCMC method provides an alternative method for parameter estimation of the Lomax distribution. Indeed, the MCMC sample may be used to completely summaries posterior uncertainty about the parameters, through a kernel estimate of the posterior distribution. Bayes estimates of the unknown parameters and then the *CV* can be obtained using Gibbs sampling and the Metropolis-Hastings procedures.

A simulation study was conducted to examine the performance of the MCMC Bayes estimators for different parameters values, different sample size (*n*) and different priors. We observe that, comparing the two Bayes estimators based on noninformative and informative priors clearly shows that the Bayes estimators based on informative prior perform better than the Bayes estimators based on noninformative prior, in terms of both MSEs and the length of the credible intervals.

8. REFERENCES

- [1] A. H. Abd Ellah, *Bayesian one sample prediction bounds for the Lomax distribution*, Indian J. Pure Appl. Math. **42** (2011), no. 5, 335–356.
- [2] A. H. Abd Ellah, *Comparison of estimates using record statistics from Lomax model : Bayesian and non Bayesian approaches*, J. Stat. Res. Train. Center **3** (2006), 139-158.
- [3] K. Ahn, *Use of coefficient of variation for uncertainty analysis in fault tree analysis*, Reliab. Engin. System Safety **47** (1995), no 3, 229–230.
- [4] B. C. Arnold, *Pareto Distributions. In: Statistical Distributions in Scientific Work*, International Co-operative Publishing House, Burtonsville, MD, 1983.
- [5] B. C. Arnold, N. Balakrishnan and H. N. Nagaraja, *Records*, Wiley, New York, 1998.
- [6] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. New York, Springer Verlag, 1985.
- [7] M. C. Bryson, *Heavy-tailed distributions: properties and tests*, Technometrics **16** (1974), no. 1, 61–68.
- [8] M. Chahkandi and M. Ganjali, *On some lifetime distributions with decreasing failure rate*, Comput. Statist. Data Anal. **53** (2009), no. 12, 4433–4440.
- [9] M. M. Dacorogna and E. Rüttener, *Why time-diversified equalisation reserves are something worth having*, Converium Research Report, Switzerland, 2006.
- [10] B. Efron, *The bootstrap and other resampling plans*, In: CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, PA, 1982.
- [11] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Trans. Pattern Anal. Mach. Intell. **6** (1984), 721–741.
- [12] W. R. Gilks, S. Richardson and D. J. Spiegelhalter, *Markov chain Monte Carlo in Practices*, Chapman and Hall, London, 1996.
- [13] J. Gong and Y. Li, *Relationship between the estimated Weibull modulus and the coefficient of variation of the measured strength for ceramics*, J. Ame. Ceramics Society, **82** (1999), no. 2, 449–452.
- [14] W. H. Greene, *Econometric Analysis, 5th Ed. New York University*, 2002.
- [15] M. Habibullah, and M. Ahsanullah, *Estimation of parameters of a Pareto distribution by generalized order statistics*, Comm. Statist. Theory Methods **29** (2000), 1597-1609.
- [16] P. Hall, *Theoretical comparison of bootstrap confidence intervals*, Ann. Stat. **16** (1988), 927-953.
- [17] A. J. Hamer, J. R. Strachan, M. M. Black, C. Ibbotson and R. A. Elson, *A new method of comparative bone strength measurement*, J. Med. Engin. Tech. **19** (1995), no. 1, 1–5.
- [18] W. K. Hastings, *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika **57** (1970), 97–109.
- [19] J. Knight and S. Satchell, *Are-examination of Sharpe's ratio for log-normal prices*, Appl. Math. Finance, **12** (2005), no. 1, 87–100.
- [20] K. S. Lomax, *Business failure: Another example of the analysis of the failure data*, JASA **49** (1954), 847-852.
- [21] A. W. Marshall and I. Olkin, *Life Distributions. Structure of Nonparametric, Semiparametric, and Parametric Families*, Springer, New York, NY, 2007.

- [22] C. Marrison, *The Fundamentals of Risk Measurement*, McGraw-Hill, New York, 2002.
- [23] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *Equations of state calculations by fast computing machines*, J. Chem. Phys. **21** (1953), 1087–1091.
- [24] E. G. Miller and M. J. Karson, *Testing Equality of Two Coefficients of Variation*, Amer. statist. Assoc. Proceedings. Business and economics **1** (1977), 278-283.
- [25] W. B. Nelson, *Applied Life Data Analysis*, Wiley, New York, 1982.
- [26] W. K. Pang, W. T. Y. Bosco, M. D. Troutt and H. H. Shui, *A simulation-based approach to the study of coefficient of variation of dividend yields*, European J. Oper. Res. **189** (2008), 559–569.
- [27] W. K. Pang, P. K. Leung, W. K. Huang and W. Liu, *On interval estimation of the coefficient of variation for the three-parameter Weibull, lognormal and gamma distribution: A simulation based approach*, European J. Oper. Res. **164** (2005), 367–377.
- [28] S. T. Rachev, I. Huber and S. Ortobelli, *Portfolio choice with heavy tailed distributions*. Tech. Rep., University of Karlsruhe, Germany, 2004. .
- [29] W. Reh, and B. Scheffler, *Significance tests and confidence intervals for coefficient of variation*, Comput. Stat. Data Anal. **22** (1996), no. 4, 449–453.
- [30] S. Rezaei, R. Tahmasbi and M. Mahmoudi, *Estimation of $P[Y_i|X]$ for generalized Pareto distribution*, J. Statist. Plann. Inference, **140** (2010), 480-494.
- [31] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Second edition, Springer, New York, 2004.
- [32] S. K. Upadhyay and M. Peshwani, *Choice between Weibull and lognormal models: a simulation based Bayesian study*, Comm. Statist. Theory Methods **32** (2003), 381-405.