

A Comparative Study of Maximum Likelihood Estimation and Bayesian Estimation for Erlang Distribution and Its Applications

Kaisar Ahmad and Sheikh Parvaiz Ahmad

Abstract

In this chapter, Erlang distribution is considered. For parameter estimation, maximum likelihood method of estimation, method of moments and Bayesian method of estimation are applied. In Bayesian methodology, different prior distributions are employed under various loss functions to estimate the rate parameter of Erlang distribution. At the end the simulation study is conducted in R-Software to compare these methods by using mean square error with varying sample sizes. Also the real life applications are examined in order to compare the behavior of the data sets in the parametric estimation. The comparison is also done among the different loss functions.

Keywords: Erlang distribution, prior distributions, loss functions, simulation study, applications

1. Introduction

Erlang distribution is a continuous probability distribution with wide applicability, primarily due to its relation to the exponential and gamma distributions. The Erlang distribution was developed by Erlang [1] to examine the number of telephone calls that could be made at the same time to switching station operators. This distribution can be expressed as waiting time and message length in telephone traffic. If the duration of individual calls are exponentially distributed then the duration of succession of calls is the Erlang distribution. The Erlang variate becomes gamma variate when its shape parameter is an integer (for details see Evans et al. [2]). Bhattacharyya and Singh [3] obtained Bayes estimator for the Erlangian queue under two prior densities. Haq and Dey [4] addressed the problem of Bayesian estimation of parameters for the Erlang distribution assuming different independent informative priors. Suri et al. [5] used Erlang distribution to design a simulator for time estimation of project management process. Damodaran et al. [6] obtained the expected time between failure measures. Further, they showed that the predicted failure times are closer to the actual failure times. Jodra [7] showed the

procedure of computing the asymptotic expansion of the median of Erlang distribution.

The probability density function of an Erlang variate is given by

$$f(x; \lambda, k) = \frac{\lambda^k}{(k-1)!} x^{k-1} e^{-\lambda x} \text{ for } x > 0, k \in N \text{ and } \lambda > 0. \quad (1)$$

where λ and k are the rate and the shape parameters, respectively, such that is k an integer number.

1.1 Graphical representation of pdf for Erlang distribution

In this chapter, Erlang distribution is considered. Some structural properties of Erlang distribution have been obtained. The parameter estimation of Erlang distribution is obtained by employing the maximum likelihood method of estimation, method of moments and Bayesian method of estimation in different sections of this chapter. In Bayesian approach, the parameters are estimated by using Jeffrey's and Quasi priors under different loss functions (**Figure 1**).

1.2 Relationship of Erlang distribution with other distributions

- i. The gamma distribution is a generalized form of the Erlang distribution.
- ii. If the shape parameter k is 1, then Erlang distribution reduces to exponential distribution.
- iii. If the scale parameter is 2, then Erlang distribution reduces to Chi-square distribution with 2 degrees of freedom.

Thus from the above descriptions, we can say that exponential distribution and Chi-square distribution are the sub-models of Erlang distribution.

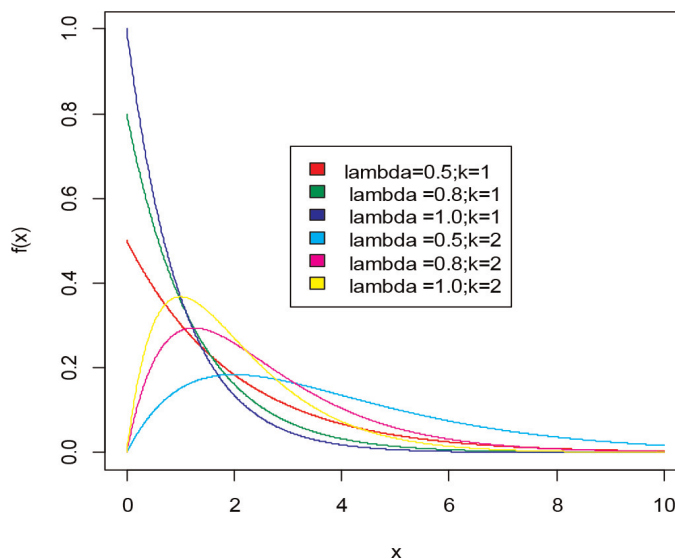


Figure 1.
Pdf's of Erlang distribution for different values of λ and k .

2. Methods used for parameter estimation

In this chapter, we have used different approaches for parameter estimation. The first two methods come under the classical approach which was founded by Fisher in a series of fundamental papers round about 1930.

The alternative approach is the Bayesian approach which was first discovered by Reverend Thomas Bayes. In this chapter, we have used two different priors for the parameter estimation. Also three loss functions are used which are discussed in their respective sections. A number of symmetric and asymmetric loss functions used by various researchers; see Zellner [8], Ahmad and Ahmad [9], Ahmad et al. [10], etc.

These methods of estimations are elaborated in their respective sections accordingly.

2.1 Maximum likelihood (MLH) estimation

The most general method of estimation is known as maximum likelihood (MLH) estimators, which was initially formulated by Gauss. Fisher in the early 1920 firstly introduced MLH as general method of estimation and later on developed by him in a series of papers. He revealed the advantages of this method by showing that it yields sufficient estimators, which are asymptotically MVUES. Thus the important feature of this method is that we look at the value of the random sample and then select our estimate of the unknown population parameter, the value of which the probability of getting the observed data is maximum.

Suppose the observed data sample values are (x_1, x_2, \dots, x_n) . When X is a discrete random variable, we can write $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = f(x_1, x_2, \dots, x_n)$, which is the value of joint probability distribution at the sample point (x_1, x_2, \dots, x_n) . Since the sample values has been observed and are therefore fixed numbers, we consider $f(x_1, x_2, \dots, x_n; \lambda)$ as the value of a function of the parameter λ , referred to as the likelihood function.

Similarly the definition applies when the random sample comes from a continuous population but in that case $f(x_1, x_2, \dots, x_n; \lambda)$ is the value of joint pdf at the sample point (x_1, x_2, \dots, x_n) . That is, the likelihood function at the sample value (x_1, x_2, \dots, x_n) which is given by

$$L(x_1, x_2, \dots, x_n | \lambda) = \prod_{i=1}^n f(x_i, \lambda).$$

Since the principle of maximum likelihood consists in finding an estimator of the parameter which maximizes the likelihood function for variation in the parameter. Thus if there exists a function $\hat{\lambda} = \hat{\lambda}(x_1, x_2, \dots, x_n)$ of the sample values which maximizes $L(x|\lambda)$ for variation in λ , then $\hat{\lambda}$ is to be taken as the estimator of λ . Usually we call $\hat{\lambda}$ as ML estimators. Thus $\hat{\lambda}$ is the solution

$$\text{iff } \frac{\partial L(x|\lambda)}{\partial \lambda} = 0 \text{ and } \frac{\partial^2 L(x|\lambda)}{\partial \lambda^2} < 0.$$

Since $L(x|\lambda) > 0$, so $\log L(x|\lambda)$ which shows that $L(x|\lambda)$ and $\log L(x|\lambda)$ attains their extreme values at $\hat{\lambda}$. Therefore, the equation becomes

$$\frac{1}{L(x|\lambda)} \frac{\partial L(x|\lambda)}{\partial \lambda} = 0 \Rightarrow \frac{\partial \log L(x|\lambda)}{\partial \lambda} = 0$$

a form which is more convenient from practical point of view.

The MLH estimation of the rate parameter of Erlang distribution is obtained in the following theorem:

Theorem 2.1: Let (x_1, x_2, \dots, x_n) be a random sample of size n from Erlang density function Eq. (1), then the maximum likelihood estimator of λ is given by

$$\hat{\lambda} = \frac{nk}{\sum_{i=1}^n x_i}.$$

Proof: The likelihood function of random sample of size n having Erlang density function Eq. (1) is given by

$$L(x; \lambda, k) = \left(\frac{(\lambda)^k}{(k-1)!} \right)^n \prod_{i=1}^n x_i^{k-1} e^{-\lambda \sum_{i=1}^n x_i}. \quad (2)$$

The log likelihood function is given by

$$\log L(x; \lambda, \beta) = nk \log \lambda + (k-1) \sum_{i=1}^n \log x_i - \lambda \sum_{i=1}^n x_i - n \log (k-1)!. \quad (3)$$

Differentiating Eq. (3) w.r.t. λ and equating to zero, we get

$$\begin{aligned} \frac{\partial}{\partial \lambda} \left(nk \log \lambda + (k-1) \sum_{i=1}^n \log x_i - \lambda \sum_{i=1}^n x_i - n \log (k-1)! \right) &= 0 \\ \hat{\lambda} &= \frac{nk}{\sum_{i=1}^n x_i}. \end{aligned} \quad (4)$$

2.2 Method of moments (MM)

One of the simplest and oldest methods of estimation is the method of moments. The method of moments was discovered by Karl Pearson in 1894. It is a method of estimation of population parameters such as mean, variance, etc. (which need not be moments), by equating sample moments with unobservable population moments and then solving those equations for the quantities to be estimated. The method of moments is special case when we need to estimate some known function of finite number of unknown moments.

Suppose $f(x; \lambda_1, \lambda_2, \dots, \lambda_p)$ be the density function of the parent population with p parameters $(\lambda_1, \lambda_2, \dots, \lambda_p)$. Let μ'_s be the s th moment of a random variable about origin and is given by

$$\mu'_s = \int_{-\infty}^{\infty} x^s f(x; \lambda_1, \lambda_2, \dots, \lambda_p) ; r = 1, 2, \dots, p.$$

In general $(\mu'_1, \mu'_2, \dots, \mu'_p)$ will be the functions of parameters $(\lambda_1, \lambda_2, \dots, \lambda_p)$. Let $(x_i; i = 1, 2, \dots, n)$ be a random sample of size n from the given population. The method of moments consists in solving the p -equations (i) for $(\lambda_1, \lambda_2, \dots, \lambda_p)$ in terms of $(\mu'_1, \mu'_2, \dots, \mu'_p)$. Then replacing these moments $(\mu'_s; s = 1, 2, 3, \dots, p)$ by the sample moments

$$\text{e.g., } \hat{\lambda}_i = \hat{\lambda}(\mu'_1, \mu'_2, \dots, \mu'_p) = \lambda_i(m'_1, m'_2, \dots, m'_p) ; i = 1, 2, \dots, p.$$

where m_i is the i th moment about origin in the sample.

Then by the method of moments $(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_p)$ are the estimators of respectively.

The MM estimation of the rate parameter of Erlang distribution is obtained in the following theorem:

Theorem 2.2: Let (x_1, x_2, \dots, x_n) be a random sample of size n from Erlang density function Eq. (1), then the moment estimator of λ is given by

$$\hat{\lambda} = \left(\frac{k}{\bar{x}} \right).$$

Proof: If the numbers (x_1, x_2, \dots, x_n) represents a set of data, then an unbiased estimator for the r th moment about origin is

$$\hat{m}_r = \frac{1}{n} \sum_{i=1}^n x_i^r \quad (5)$$

where \hat{m}_r stands for the estimate of m_r .

The r th moment of two parameter Erlang distribution about origin is given by

$$\mu'_r = \int_0^{\infty} x^r f(x; \lambda, k) dx \quad (6)$$

Using Eq. (5) in Eq. (6), we have

$$\mu'_r = \frac{\Gamma(r+k)}{\lambda^r (k-1)!}. \quad (7)$$

If $r = 1$ in Eq. (7), we get

$$\mu'_1 = \frac{k}{\lambda}.$$

If $r = 2$, then Eq. (7) becomes

$$\mu'_2 = \frac{k(k+1)}{\lambda^2}.$$

Thus the variance is given by $\sigma^2 = \frac{k}{\lambda^2}$.

When we divide σ^2 by $\mu_1'^2$, we get an expression which is a function of k only and is given by

$$\begin{aligned} \frac{\sigma^2}{\mu_1'^2} &= \frac{\left(\frac{k}{\lambda^2} \right)}{\left(\frac{k}{\lambda} \right)^2} \\ \frac{\sigma^2}{\mu_1'^2} &= \frac{1}{k}. \end{aligned} \quad (8)$$

On taking the square roots of Eq. (8), we have the coefficient of variation

$$\frac{\sigma}{\mu_1'} = \sqrt{\left(\frac{1}{k} \right)}. \quad (9)$$

The rate parameter λ can be then estimated using the following equation

$$m'_1 = \mu'_1.$$

If $r = 1$, in Eq. (5), then

$$m'_1 = \bar{x}.$$

Also if $r = 1$ in Eq. (7), then

$$\mu'_1 = \frac{k}{\lambda}.$$

Thus,

$$m'_1 = \mu'_1$$

$$\bar{x} = \frac{k}{\hat{\lambda}}, \text{ where } \bar{x} \text{ is the mean of the data and } \hat{\lambda} = \left(\frac{k}{\bar{x}} \right).$$

3. Bayesian method of estimation

Nowadays, the Bayesian school of thought is garnering more attention and at an increasing rate. This thought of statistics was given by Reverend Thomas Bayes. He first discovered the theorem that now bears his name. It was written up in a paper “An Essay Towards Solving a Problem in the Doctrine of Chances.” This paper was found after his death by his friend Richard Price, who had it published posthumously in the Philosophical Transactions of the Royal Society in 1763. Bayes showed how inverse probability could be used to calculate probability of antecedent events from the occurrence of the consequent event. His methods were adopted by Laplace and other scientists in the nineteenth century. By mid twentieth century interest in Bayesian methods was renewed by De Finetti, Jeffreys and Lindley, among others. They developed a complete method of statistical inference based on Bayes’ theorem. Bayesian analysis is to be used by practitioners for situations where scientists have a priori information about the values of the parameters to be estimated. In everyday life, uncertainty often permeates our choices, and when choices need to be made, past experience frequently proves a helpful aid. Bayesian theory provides a general and consistent framework for dealing with uncertainty.

Within Bayesian inference, there are also different interpretations of probability, and different approaches based on those interpretations. Early efforts to make Bayesian methods accessible for data analysis were made by Raiffa and Schlaifer [11], DeGroot [12], Zellner [13], and Box and Tiao [14]. The most popular interpretations and approaches are objective Bayesian inference and subjective Bayesian inference. Excellent expositions of these approaches are with Bayes and Price [15], Laplace [16], Jeffrey’s [17], Anscombe and Aumann [18], Berger [19, 20], Gelman et al. [21], Leonard and Hsu [22], De-Finetti [23]. Modern Bayesian data analysis and methods based on Markov chain Monte Carlo methods are presented in Bernardo and Smith [24], Robert [25], Gelman et al. [26], Marin and Robert [27], Carlin and Louis [28]. Good elementary introductions to the subject are Ibrahim et al. [29], Ghosh [30], Bansal [31], Koch [32], Hoff [33].

In Bayesian statistics probability is not defined as a frequency of occurrence but as the plausibility that a proposition is true, given the available information. The parameters are treated as random variables. The rules of probability are used

directly to make inferences about the parameters. Probability statements about parameters must be interpreted as “degree of belief.” We revise our beliefs about parameters after getting the data by using Bayes’ theorem. This gives our posterior distribution which gives the relative weights we give to each parameter value after analyzing the data. The posterior distribution comes from two sources: the prior distribution and the observed data. This means that the inference is based on the actual occurring data, not all possible data sets that might have occurred.

In this section, posterior distribution of Erlang distribution is obtained by using Jeffrey’s prior and Quasi prior. The rate parameter of Erlang distribution is estimated with the help of different loss functions. For parameter estimation we have used the approach as is used by Ahmad et al. [34], Ahmad et al. [35], etc. Some important prior distributions and loss functions which we have used in this article are given as below:

3.1 Prior distributions used

Prior distribution is the basic part of Bayesian analysis which represents all that is known or assumed about the parameter. Usually the prior information is subjective and is based on a person’s own experience and judgment, a statement of one’s degree of belief regarding the parameter.

Another important feature of the Bayesian analysis is the choice of the prior distribution. If the data have sufficient signal, even a bad prior will still not greatly influence the posterior. We can examine the impact of prior by observing the stability of posterior distribution related to different choices of priors. If the posterior distribution is highly dependent on the prior, then the data (the likelihood function) may not contain sufficient information. However, if the posterior is relatively stable over a choice of priors, then the data indeed contains significant information.

Prior distribution may be categorical in different ways. One common classification is a dichotomy that separated “proper” and “improper” priors.

A prior distribution is proper if it does not depend on the data and the value of integral $\int_{-\infty}^{\infty} g(\lambda)d\lambda$ or summation $\sum g(\lambda)$ is one. If the prior does not depend on the data and the distribution does not integrate or sum to one then we say that the prior is improper.

In this chapter, we have used two different priors Jeffrey’s prior and Quasi prior which are given below:

3.1.1 Jeffrey’s prior

An invariant form for the prior probability in estimation problems is given by Jeffery’s [36].

The general formula of the Jeffreys prior, which is defined by

$$g(\lambda) \propto \sqrt{I(\lambda)} \propto \left(-E \left[\frac{\partial^2 \log L(\lambda|x)}{\partial \lambda^2} \right] \right)^{\frac{1}{2}}.$$

where $I(\lambda)$ is the Fisher information for the parameter λ . When there are multiple parameters I is the Fisher information matrix, the matrix of the expected second partials

$$I(\lambda) = E\left(\frac{\partial^2 \log L(\lambda|x)}{\partial \lambda_i \partial \lambda_j}\right).$$

In this situation, the Jeffreys prior is given by

$$g(\lambda) \propto \sqrt{\det(I(\lambda))}.$$

Jeffrey suggested a thumb rule for specifying non-informative prior for parameter λ as

Rule 1: if $\lambda \in (-\infty, \infty)$ take $g(\lambda)$ to be constant, i.e., λ to be uniformly distributed.

Rule 2: if $\lambda \in (0, \infty)$ take $g(\lambda) \propto \frac{1}{\lambda}$, i.e., $\log \lambda$ to be uniformly distributed.

Under linear transformation, rule 1 is invariant and under any power transformation of λ , rule 2 is invariant.

3.1.2 Quasi prior

When there is no more information about the distribution parameter, one may use the Quasi density as given by $g(\lambda) = \frac{1}{\lambda^d}$; $\lambda > 0$ and $d > 0$.

3.2 Loss functions used

The concept of loss function is old as Laplace and was reintroduced in statistics by Abraham Wald [37]. In statistics, typically a loss function is used for parameter estimation, and the event in question is some function of the difference between estimated and true values for an occurrence of data. In the context of economics, loss function is usually economic cost. In optimal control, the loss is the penalty for failing to achieve a desired value.

The word “loss” is used in place of “error” and the loss function is used as a measure of the error or loss. Loss function is a measure of the error and presumably would be greater for large error than for small error. We would want the loss to be small or we want the estimate to be close to what it is estimating. Loss depends on sample and we cannot hope to make the loss small for every possible sample but can try to make the loss small on the average. Our objective is to select an estimator that makes this error or loss small, also which makes the average loss (risk) small and ideally select an estimator that has the small risk.

In this chapter, we have used three different Loss Functions which are as under:

3.2.1 Precautionary loss function (PLF)

The concept of precautionary loss function (PLF) was introduced by Norstrom [38]. He introduced an alternative asymmetric loss function and also presented a general class of precautionary loss functions as a special case. These loss functions approach infinitely near the origin to prevent underestimation, thus giving conservative estimators, especially when low failure rates are being estimated. These estimators are very useful when underestimation may lead to serious consequences. A very useful and simple asymmetric precautionary loss function (PLF) is

$$L(\hat{\lambda}, \lambda) = \frac{(\hat{\lambda} - \lambda)^2}{\lambda}.$$

3.2.2 Al-Bayyati's loss function (ALF)

The loss function proposed by Al-Bayyati [39] is an asymmetric loss function and is given by

$$l_A(\hat{\lambda}, \lambda) = \lambda^c (\hat{\lambda} - \lambda)^2 c \epsilon R.$$

where λ and $\hat{\lambda}$ represents the true and estimated values of the parameter. This loss function is frequently used because of its analytical tractability in Bayesian analysis.

3.2.3 LINEX loss function (LLF)

The idea of LINEX loss function (LLF) was founded by Klebanov [40] and used by Varian [41] in the context of real estate assessment. The formula of LLF is given by

$$L(\hat{\lambda}, \lambda) = (\exp(a(\hat{\lambda} - \lambda)) - a(\hat{\lambda} - \lambda) - 1)$$

where λ and $\hat{\lambda}$ represents the true and estimated values of the parameter and the constant c determines the shape of the loss function.

3.3 Posterior density under Jeffrey's prior

Let (x_1, x_2, \dots, x_n) be a random sample of size n having the Erlang density function Eq. (1) which is given by

$$f(x; \lambda, k) = \frac{\lambda^k}{(k-1)!} x^{k-1} e^{-\lambda x} \text{ for } x > 0, k \in N \text{ and } \lambda > 0$$

and the likelihood function Eq. (2) given as below

$$L(x; \lambda, k) = \left(\frac{(\lambda)^k}{(k-1)!} \right)^n \prod_{i=1}^n x_i^{k-1} e^{-\lambda \sum_{i=1}^n x_i}.$$

Jeffreys' non-informative prior for λ is given by

$$g(\lambda) \propto \sqrt{\det(I(\lambda))}.$$

where $I(\lambda) = -nE \left[\frac{\partial^2 \log f(x; \lambda, k)}{\partial \lambda^2} \right]$ is the Fisher's information matrix for the probability density function Eq. (1).

On solving the above expression, we have

$$g(\lambda) = \frac{1}{\lambda}. \quad (10)$$

By using the Bayes theorem, we have

$$\pi_1(\lambda | \underline{x}) \propto L(x | \lambda) g(\lambda). \quad (11)$$

Using Eqs. (2) and (10) in Eq. (11), we get

$$\begin{aligned}\pi_1(\lambda|\underline{x}) &\propto \frac{(\lambda)^{nk-1}}{(k-1)!} \prod_{i=1}^n x_i^{k-1} e^{-\lambda \sum_{i=1}^n x_i} \\ \pi_1(\lambda|\underline{x}) &= \rho \lambda^{nk-1} e^{-\lambda \sum_{i=1}^n x_i}\end{aligned}\quad (12)$$

where ρ is independent of λ and

$$\begin{aligned}\rho^{-1} &= \int_0^\infty \lambda^{nk-1} e^{-\lambda \sum_{i=1}^n x_i} d\lambda \\ \rho &= \frac{(\sum_{i=1}^n x_i)^{nk}}{\Gamma nk}.\end{aligned}$$

Using the value of ρ in Eq. (12)

$$\pi_1(\lambda|\underline{x}) = \left(\frac{\lambda^{nk-1} e^{-\lambda \sum_{i=1}^n x_i} (\sum_{i=1}^n x_i)^{nk}}{\Gamma nk} \right). \quad (13)$$

3.4 Posterior density under Quasi prior

Let (x_1, x_2, \dots, x_n) be a random sample of size n having the Erlang density function Eq. (2) and the likelihood function Eq. (2).

The Quasi for λ is given by

$$g(\lambda) = \frac{1}{\lambda^d}. \quad (14)$$

By using the Bayes theorem, we have

$$\pi_2(\lambda|\underline{x}) \propto L(x|\lambda)g(\lambda). \quad (15)$$

Using Eqs. (2) and (14) in Eq. (15), we have

$$\begin{aligned}\pi_2(\lambda|\underline{x}) &\propto \frac{(\lambda)^{nk-d}}{(k-1)!} \prod_{i=1}^n x_i^{k-1} e^{-\lambda \sum_{i=1}^n x_i} \\ \pi_2(\lambda|\underline{x}) &= \rho \lambda^{nk-d} e^{-\lambda \sum_{i=1}^n x_i}\end{aligned}\quad (16)$$

where ρ is independent of λ and

$$\begin{aligned}\rho^{-1} &= \int_0^\infty \lambda^{nk-2d} e^{-\lambda \sum_{i=1}^n x_i} d\lambda \\ \rho &= \frac{(\sum_{i=1}^n x_i)^{nk-d+1}}{\Gamma(nk-d+1)}.\end{aligned}$$

By using the value of ρ in Eq. (16), we have

$$\pi_2(\lambda|\underline{x}) = \left(\frac{\lambda^{nk-d} e^{-\lambda \sum_{i=1}^n x_i} (\sum_{i=1}^n x_i)^{nk-d+1}}{\Gamma(nk-d+1)} \right). \quad (17)$$

4. Estimation of parameters under Jeffrey's prior

In this section, parameter estimation of Erlang distribution is done by using Jeffreys' prior under different loss functions. The procedure of calculating the Bayesian estimate is already defined in Section 3. The estimates are obtained in the following theorems:

Theorem 4.1: Assuming the loss function $L_p(\hat{\lambda}, \lambda)$, the Bayesian estimator of the rate parameter λ , if the shape parameter k is known, is of the form

$$\hat{\lambda}_p = \frac{\sqrt{nk(nk+1)}}{(\sum_{i=1}^n x_i)}.$$

Proof: The risk function of the estimator λ under the precautionary loss function $L_p(\hat{\lambda}, \lambda)$ is given by the formula

$$R(\hat{\lambda}) = \int_0^{\infty} \frac{(\hat{\lambda} - \lambda)^2}{\hat{\lambda}} \pi_1(\lambda|\underline{x}) d\lambda. \quad (18)$$

Using Eq. (13) in Eq. (18), we get

$$\begin{aligned} R(\hat{\lambda}) &= \int_0^{\infty} \frac{(\hat{\lambda} - \lambda)^2}{\hat{\lambda}} \left(\frac{\lambda^{nk-1} (\sum_{i=1}^n x_i)^{nk} e^{-\lambda \sum_{i=1}^n x_i}}{\Gamma nk} \right) d\lambda \\ R(\hat{\lambda}) &= \frac{(\sum_{i=1}^n x_i)^{nk}}{\Gamma nk} \left[\hat{\lambda} \int_0^{\infty} \lambda^{nk-1} e^{-\lambda \sum_{i=1}^n x_i} d\lambda + \frac{1}{\hat{\lambda}} \int_0^{\infty} \lambda^{nk+1} e^{-\lambda \sum_{i=1}^n x_i} d\lambda - 2 \int_0^{\infty} \lambda^{nk} e^{-\lambda \sum_{i=1}^n x_i} d\lambda \right]. \end{aligned}$$

On solving the above expression, we get

$$R(\hat{\lambda}) = \hat{\lambda} + \frac{1}{\hat{\lambda}} \frac{nk(nk+1)}{(\sum_{i=1}^n x_i)^2} - \frac{2nk}{(\sum_{i=1}^n x_i)}.$$

Minimization of the risk with respect to $\hat{\lambda}$ gives us the optimal estimator i.e.,

$$\begin{aligned} \frac{\partial}{\partial \hat{\lambda}} [R(\hat{\lambda})] &= 0 \\ \hat{\lambda}_p &= \frac{\sqrt{nk(nk+1)}}{(\sum_{i=1}^n x_i)}. \end{aligned} \quad (19)$$

Theorem 4.2: Assuming the loss function $l_A(\hat{\lambda}, \lambda)$, the Bayesian estimator of the rate parameter λ , if the shape parameter k is known, is of the form

$$\hat{\lambda}_A = \frac{(nk+c)}{(\sum_{i=1}^n x_i)}.$$

Proof: The risk function of the estimator λ under the Al-Bayyati's loss function $L_A(\hat{\lambda}, \lambda)$ is given by the formula

$$R(\hat{\lambda}) = \int_0^{\infty} \lambda^c (\hat{\lambda} - \lambda)^2 \pi_1(\lambda|\underline{x}) d\lambda. \quad (20)$$

On substituting Eq. (13) in Eq. (20), we have.

$$R(\hat{\lambda}) = \int_0^{\infty} \lambda^c (\hat{\lambda} - \lambda)^2 \left(\frac{\lambda^{nk-1} (\sum_{i=1}^n x_i)^{nk} e^{-\lambda \sum_{i=1}^n x_i}}{\Gamma nk} \right) d\lambda$$

$$R(\hat{\lambda}) = \frac{(\sum_{i=1}^n x_i)^{nk}}{\Gamma nk} \left[\hat{\lambda}^2 \int_0^{\infty} \lambda^{nk+c-1} e^{-\lambda \sum_{i=1}^n x_i} d\lambda + \int_0^{\infty} \lambda^{nk+c+1} e^{-\lambda \sum_{i=1}^n x_i} d\lambda - 2\hat{\lambda} \int_0^{\infty} \lambda^{nk+c} e^{-\lambda \sum_{i=1}^n x_i} d\lambda \right].$$

On solving the above expression, we get

$$R(\hat{\lambda}) = \frac{\hat{\lambda}^2 \Gamma(nk + c)}{\Gamma nk (\sum_{i=1}^n x_i)^c} + \frac{\Gamma(nk + c + 2)}{\Gamma nk (\sum_{i=1}^n x_i)^{c+2}} - \frac{2\hat{\lambda} \Gamma(nk + c + 1)}{\Gamma nk (\sum_{i=1}^n x_i)^{c+1}}.$$

Minimization of the risk with respect to $\hat{\lambda}$ gives us the optimal estimator i.e.,

$$\frac{\partial}{\partial \hat{\lambda}} [R(\hat{\lambda})] = 0$$

$$\hat{\lambda}_A = \frac{(nk + c)}{(\sum_{i=1}^n x_i)}. \quad (21)$$

Theorem 4.3: Assuming the loss function $L_l(\hat{\lambda}, \lambda)$, the Bayesian estimator of the rate parameter λ , if the shape parameter k is known, is of the form

$$\hat{\lambda}_l = \frac{nk \log \left(1 + \frac{a}{(\sum_{i=1}^n y_i)} \right)}{a}.$$

Proof: The risk function of the estimator λ under the LINEX loss function $L_l(\hat{\lambda}, \lambda)$ is given by the formula

$$R(\hat{\lambda}) = \int_0^{\infty} (\exp(a(\hat{\lambda} - \lambda)) - a(\hat{\lambda} - \lambda) - 1) \pi_1(\lambda|\underline{x}) d\lambda. \quad (22)$$

Using Eq. (13) in Eq. (22), we have

$$R(\hat{\lambda}) = \int_0^{\infty} (\exp(a(\hat{\lambda} - \lambda)) - a(\hat{\lambda} - \lambda) - 1) \left(\frac{\lambda^{nk-1} (\sum_{i=1}^n x_i)^{nk} e^{-\lambda \sum_{i=1}^n x_i}}{\Gamma nk} \right) d\lambda$$

$$R(\hat{\lambda}) = \frac{(\sum_{i=1}^n x_i)^{nk}}{\Gamma nk} \left[\exp(a\hat{\lambda}) \int_0^{\infty} \left(\exp\left(-\left(a + \sum_{i=1}^n x_i\right)\lambda\right)\right) \lambda^{nk-1} d\lambda - a\hat{\lambda} \int_0^{\infty} \left(\exp\left(-\sum_{i=1}^n x_i\right)\lambda\right) \lambda^{nk-1} d\lambda \right. \\ \left. + a \int_0^{\infty} \left(\exp\left(-\sum_{i=1}^n x_i\right)\lambda\right) \lambda^{nk} d\lambda - \int_0^{\infty} \left(\exp\left(-\sum_{i=1}^n x_i\right)\lambda\right) \lambda^{nk-1} d\lambda \right].$$

On solving the above expression, we get

$$R(\hat{\lambda}) = \frac{(\sum_{i=1}^n x_i)^{nk} \exp(a\hat{\lambda})}{(a + \sum_{i=1}^n x_i)^{nk}} - a\hat{\lambda} + \frac{a\Gamma(nk+1)}{\Gamma nk (\sum_{i=1}^n x_i)} - 1.$$

Minimization of the risk with respect to $\hat{\lambda}$ gives us the optimal estimator i.e.,

$$\frac{\partial}{\partial \hat{\lambda}} [R(\hat{\lambda})] = 0$$

$$\hat{\lambda}_l = \frac{nk \log\left(1 + \frac{a}{(\sum_{i=1}^n x_i)}\right)}{a}. \quad (23)$$

5. Estimation of parameters under Quasi prior

In this section, parameter estimation of Erlang distribution is done by using QUASI prior under different loss functions. The procedure for obtaining the Bayesian estimate is available in Section 3. The estimates of parameter are obtained in the following theorems.

Theorem 5.1: Assuming the loss function $L_p(\hat{\lambda}, \lambda)$, the Bayesian estimator of the rate parameter λ , if the shape parameter k is known, is of the form

$$\hat{\lambda}_p = \frac{\sqrt{(nk-d+1)(nk-d+2)}}{(\sum_{i=1}^n x_i)}.$$

Proof: The risk function of the estimator λ under the precautionary loss function $L_p(\hat{\lambda}, \lambda)$ is given by the formula

$$R(\hat{\lambda}) = \int_0^{\infty} \frac{(\hat{\lambda} - \lambda)^2}{\hat{\lambda}} \pi_2(\lambda|\underline{x}) d\lambda \quad (24)$$

Using Eq. (17) in Eq. (24), we have

$$R(\hat{\lambda}) = \int_0^{\infty} \frac{(\hat{\lambda} - \lambda)^2}{\hat{\lambda}} \left(\frac{\lambda^{nk-d} (\sum_{i=1}^n x_i)^{nk-d+1} e^{-\lambda \sum_{i=1}^n x_i}}{\Gamma(nk-d+1)} \right) d\lambda$$

$$R(\hat{\lambda}) = \frac{(\sum_{i=1}^n x_i)^{nk-d+1}}{\Gamma(nk-d+1)} \left[\hat{\lambda} \int_0^{\infty} \lambda^{nk-d} e^{-\lambda \sum_{i=1}^n x_i} d\lambda + \frac{1}{\hat{\lambda}} \int_0^{\infty} \lambda^{nk-d+2} e^{-\lambda \sum_{i=1}^n x_i} d\lambda - 2 \int_0^{\infty} \lambda^{nk-d+1} e^{-\lambda \sum_{i=1}^n x_i} d\lambda \right].$$

On solving the above expression, we get

$$R(\hat{\lambda}) = \hat{\lambda} + \frac{1}{\hat{\lambda}} \frac{\Gamma(nk - d + 3)}{\Gamma(nk - d + 1)(\sum_{i=1}^n x_i)^2} - \frac{2\Gamma(nk - d + 2)}{\Gamma(nk - d + 1)(\sum_{i=1}^n x_i)}.$$

Minimization of the risk with respect to $\hat{\lambda}$ gives us the optimal estimator i.e.,

$$\begin{aligned} \frac{\partial}{\partial \hat{\lambda}} [R(\hat{\lambda})] &= 0 \\ \hat{\lambda}_p &= \frac{\sqrt{(nk - d + 1)(nk - d + 2)}}{(\sum_{i=1}^n x_i)}. \end{aligned} \quad (25)$$

Remark: Replacing $d = 1$ in Eq. (25), the same Bayes estimator is obtained as in Eq. (19).

Theorem 5.2: Assuming the loss function $l_A(\hat{\lambda}, \lambda)$, the Bayesian estimator of the rate parameter λ , if the shape parameter k is known, is of the form

$$\hat{\lambda}_A = \frac{(nk - d + c + 1)}{(\sum_{i=1}^n x_i)}.$$

Proof: The risk function of the estimator λ under the Al-Bayyati's loss function $L_A(\hat{\lambda}, \lambda)$ is given by the formula

$$R(\hat{\lambda}) = \int_0^{\infty} \lambda^c (\hat{\lambda} - \lambda)^2 \pi_2(\lambda | \underline{x}) d\lambda. \quad (26)$$

By using Eq. (17) in Eq. (26), we have.

$$\begin{aligned} R(\hat{\lambda}) &= \int_0^{\infty} \lambda^c (\hat{\lambda} - \lambda)^2 \left(\frac{\lambda^{nk-d} (\sum_{i=1}^n x_i)^{nk-d+1} e^{-\lambda \sum_{i=1}^n x_i}}{\Gamma(nk - d + 1)} \right) d\lambda \\ R(\hat{\lambda}) &= \frac{(\sum_{i=1}^n x_i)^{nk-d+1}}{\Gamma(nk - d + 1)} \left[\hat{\lambda}^2 \int_0^{\infty} \lambda^{nk-d+c} e^{-\lambda \sum_{i=1}^n x_i} d\lambda + \int_0^{\infty} \lambda^{nk-d+c+2} e^{-\lambda \sum_{i=1}^n x_i} d\lambda - 2\hat{\lambda} \int_0^{\infty} \lambda^{nk-d+c+1} e^{-\lambda \sum_{i=1}^n x_i} d\lambda \right]. \end{aligned}$$

On solving the above expression, we get

$$R(\hat{\lambda}) = \frac{\hat{\lambda}^2 \Gamma(nk - d + c + 1)}{\Gamma(nk - d + 1)(\sum_{i=1}^n x_i)^c} + \frac{\Gamma(nk - d + c + 3)}{\Gamma(nk - d + 1)(\sum_{i=1}^n x_i)^{c+2}} - \frac{2\hat{\lambda} \Gamma(nk - d + c + 2)}{\Gamma(nk - d + 1)(\sum_{i=1}^n x_i)^{c+1}}$$

Minimization of the risk with respect to $\hat{\lambda}$ gives us the optimal estimator, i.e.,

$$\frac{\partial}{\partial \hat{\lambda}} [R(\hat{\lambda})] = 0$$

$$\hat{\lambda}_A = \frac{(nk - d + c + 1)}{(\sum_{i=1}^n x_i)}. \quad (27)$$

Remark: Replacing $d = 1$ in Eq. (27), the same Bayes estimator is obtained as in Eq. (21).

Theorem 5.3: Assuming the loss function $L_l(\hat{\lambda}, \lambda)$, the Bayesian estimator of the rate parameter λ , if the shape parameter k is known, is of the form

$$\hat{\lambda}_l = \frac{(nk - d + 1) \log \left(1 + \frac{a}{(\sum_{i=1}^n x_i)} \right)}{a}.$$

Proof: The risk function of the estimator λ under the LINEX loss function $L_l(\hat{\lambda}, \lambda)$ is given by the formula

$$R(\hat{\lambda}) = \int_0^{\infty} (\exp(a(\hat{\lambda} - \lambda)) - a(\hat{\lambda} - \lambda) - 1) \pi_2(\lambda | \underline{x}) d\lambda. \quad (28)$$

Using Eq. (17) in Eq. (28), we have.

$$R(\hat{\lambda}) = \int_0^{\infty} (\exp(a(\hat{\lambda} - \lambda)) - a(\hat{\lambda} - \lambda) - 1) \left(\frac{\lambda^{nk-d} \left(\sum_{i=1}^n x_i \right)^{nk-d+1} e^{-\lambda \sum_{i=1}^n x_i}}{\Gamma(nk-d+1)} \right) d\lambda$$

$$R(\hat{\lambda}) = \frac{(\sum_{i=1}^n x_i)^{nk-d+1}}{\Gamma(nk-2c+1)} \left[\exp(a\hat{\lambda}) \int_0^{\infty} \left(\exp \left(- \left(a + \sum_{i=1}^n x_i \right) \lambda \right) \right) \lambda^{nk-d} d\lambda - a\hat{\lambda} \int_0^{\infty} \left(\exp \left(- \sum_{i=1}^n x_i \right) \lambda \right) \lambda^{nk-d} d\lambda \right. \\ \left. + a \int_0^{\infty} \left(\exp \left(- \sum_{i=1}^n x_i \right) \lambda \right) \lambda^{nk-d+1} d\lambda - \int_0^{\infty} \left(\exp \left(- \sum_{i=1}^n x_i \right) \lambda \right) \lambda^{nk-d} d\lambda \right].$$

On solving the above expression, we get

$$R(\hat{\lambda}) = \frac{\exp(a\hat{\lambda}) (\sum_{i=1}^n x_i)^{nk-d+1}}{(a + \sum_{i=1}^n x_i)^{nk-d+1}} - a\hat{\lambda} + \frac{a\Gamma(nk-d+2)}{\Gamma(nk-d+1) (\sum_{i=1}^n x_i)} - 1.$$

Minimization of the risk with respect to $\hat{\lambda}$ gives us the optimal estimator i.e.,

$$\frac{\partial}{\partial \hat{\lambda}} [R(\hat{\lambda})] = 0$$

$$\hat{\lambda}_l = \frac{(nk - d + 1) \log \left(1 + \frac{a}{(\sum_{i=1}^n x_i)} \right)}{a}. \quad (29)$$

Remark: Replacing $d = 1$ in Eq. (29), the same Bayes estimator is obtained as in Eq. (23).

6. Entropy estimation of Erlang distribution

The concept of entropy was introduced by Claude. Shannon [42] in the paper "A Mathematical theory of Communication." This concept of Shannon's entropy is the central role of information theory, sometimes referred as measure of uncertainty. Shannon entropy provides an absolute limit on the best possible lossless encoding or compression of any communication, assuming that the communication may be represented as a sequence of independent and identical distributed random variables. Entropy is typically measured in bits, when the log is to the base 2, and nats, when the log is to the base e .

Shannon's definition of entropy, when applied to an information source, can determine the minimum channel capacity required to reliably transmit the source as encoded binary digit. The entropy of a random variable is defined in terms of its probability distribution and can be shown to be a good measure of randomness or uncertainty. For deriving the entropy of probability distributions, we need the following two definitions that are more discussed in Cover et al. [43].

Definition (i): The entropy of the discrete random variable defined on the probability space is given by

$$H_P(f) = - \sum_{i=1}^n p(f = a) \log(p(f = a)).$$

It is obvious that $H_P(f) \geq 0$.

Definition (ii): The entropy of the continuous random variable defined on the real line is given by

$$H(f) = E(-\log(x)) = - \int_{-\infty}^{\infty} f(x) \log f(x) dx.$$

In this section, entropy estimation of two parameter Erlang distribution is discussed which given as below.

Theorem 6.1: Let (x_1, x_2, \dots, x_n) be n positive independent and identically distributed random samples drawn from a population having Erlang density function Eq. (2), then the Shannon's entropy of two parameter Erlang distribution is given by

$$H(f(x; \alpha, \beta)) = -\log\left(\frac{\lambda^k}{(k-1)!}\right) - (k-1)(\psi(k) - \log \lambda) + k.$$

Proof: Shannon's entropy for a continuous random variable is defined as

$$H(f(x; \lambda, k)) = E(-\log(f(x))) = - \int_{-\infty}^{\infty} f(x) \log f(x) dx \quad (30)$$

Using Eq. (1) in Eq. (30), we have

$$\begin{aligned} H(f(x; \alpha, \beta)) &= E\left\{-\log\left(\frac{\lambda^k}{(k-1)!} x^{k-1} e^{-\lambda x}\right)\right\} \\ H(f(x; \alpha, \beta)) &= -\log\left(\frac{\lambda^k}{(k-1)!}\right) - (k-1)E(\log(x)) + \lambda E(x). \end{aligned} \quad (31)$$

Now

$$\begin{aligned} E(\log(x)) &= \int_0^{\infty} \log(x) f(x) dx \\ E(\log(x)) &= \frac{\lambda^k}{(k-1)!} \int_0^{\infty} \log(x) x^{k-1} e^{-\lambda x} dx \end{aligned}$$

$$\text{Put } \lambda x = t \Rightarrow dx = \frac{dt}{\lambda}; \text{ as } x \rightarrow 0, t \rightarrow 0 \text{ and as } x \rightarrow \infty, t \rightarrow \infty$$

$$\begin{aligned}
 E(\log(x)) &= \frac{\lambda^{k-1}}{(k-1)!} \int_0^\infty \log\left(\frac{t}{\lambda}\right) \left(\frac{t}{\lambda}\right)^{k-1} e^{-t} dt \\
 E(\log(x)) &= \frac{1}{(k-1)!} \left[\int_0^\infty \log t (t)^{k-1} e^{-t} dt - \log \lambda \int_0^\infty (t)^{k-1} e^{-t} dt \right] \\
 E(\log(x)) &= \frac{\Gamma'(k)}{\Gamma k} - \log \lambda \\
 E(\log(x)) &= (\psi(k) - \log \lambda). \tag{32}
 \end{aligned}$$

Also

$$\begin{aligned}
 E(x) &= \int_0^\infty x f(x; \lambda, k) dx \\
 E(x) &= \frac{\lambda^k}{(k-1)!} \int_0^\infty x^k e^{-\lambda x} dx \\
 E(x) &= \frac{\lambda^k}{(k-1)!} \frac{\Gamma(k+1)}{\lambda^{k+1}} \\
 E(x) &= \frac{k}{\lambda}. \tag{33}
 \end{aligned}$$

Substitute the value of Eqs. (32) and (33) in Eq. (31), we have.

$$H(f(x; \alpha, \beta)) = -\log\left(\frac{\lambda^k}{(k-1)!}\right) - (k-1)(\psi(k) - \log \lambda) + k. \tag{34}$$

7. AIC and BIC criterion for Erlang distribution

For model selection the approach of Akaike information criterion (AIC) and Bayesian information criterion (BIC) based on entropy estimation are used. The Akaike information criterion (AIC) was introduced by Hirotugu Akaike [44] and proposed it as a measure of goodness of fit of an estimated statistical model. It is a measure of the relative quality of a statistical model for a given set of data. It has been found in information theory that it offers a relative estimate of the information lost when a given model is used to represent the process that generates the data.

The AIC is not a test of the model in the sense of hypothesis testing; rather it is a test between models—a tool for model selection. Given a data set, several competing models may be ranked according to their AIC, with the one having the lowest AIC being the best.

The formula for AIC is given by

$$AIC = 2K - 2\log(L(\hat{\lambda})).$$

where K is the number of parameters and $L(\hat{\lambda})$ is the maximized value of the likelihood function for the estimated model.

AICC was first introduced by Hurvich and Tsai [45] and its different derivations were proposed by Burnham and Anderson [46]. AICC is AIC with a correction for finite sample sizes and is given by

$$AICC = AIC + \frac{2K(K+1)}{(n-K-1)}.$$

Burnham and Anderson [47] strongly recommended that we should use AICC instead of AIC when the sample size is small or if K is large. Since AICC converges to AIC as the sample size is getting large. Using AIC, instead of AICC, when the sample size is not many times larger than K^2 , increases the probability of selecting models that have too many parameters, i.e., of over fitting. The probability of AIC over fitting can be substantial, in some cases.

The Bayesian information criterion (BIC) also known as Schwarz Criterion is used as a substitute for full calculation of the Bayes' factor since it can be calculated without specifying prior distribution. In BIC, the penalty for additional parameters is stronger than that of the AIC.

The formula for the BIC is given by

$$BIC = K \log n - 2 \log (L(\hat{\lambda})).$$

where K is the number of parameters, n is the sample size and $L(\hat{\lambda})$ is the maximized value of the likelihood function.

The AIC and BIC of two parameter Erlang distribution are obtained in this section, which are given below.

The Shannon's entropy of two parameter Erlang distribution is given by

$$\begin{aligned} SH(ED) &= \log(k-1)! - \log \lambda^k - (k-1)E(\log x) + \lambda E(x) \\ \hat{H}(ED) &= \log(k-1)! - \log \lambda^k - (k-1) \log \bar{x} + \lambda \bar{x}. \end{aligned} \quad (35)$$

Also

$$\begin{aligned} ll(x; \lambda, k) &= n \log \lambda^k - n \log(k-1)! + (k-1) \sum_{i=1}^n \log x_i - \lambda \sum_{i=1}^n x_i - ll(x; \hat{\lambda}, \hat{k}) \\ &= n [\log(k-1)! - \log \lambda^k - (k-1) \log \bar{x} + \lambda \bar{x}]. \end{aligned} \quad (36)$$

Comparing Eqs. (35) and (36), we have

$$ll(x; \hat{\lambda}, \hat{k}) = -n\hat{H}(ED).$$

The AIC and BIC methodology attempts to find the model that best explains the data with a minimum of their values, we have.

$ll(x; \hat{\lambda}, \hat{k}) = -n\hat{H}(ED)$, then for Erlang family we have

$$AIC = 2K + 2n\hat{H}(ED) \quad (37)$$

or

$$AIC = 2K - 2ll$$

and

$$BIC = K \log n + 2n\hat{H}(ED) \quad (38)$$

or

$$BIC = K \log n - 2ll.$$

8. Simulation study of Erlang distribution

We have generated the data for Erlang distribution of different sample sizes (15, 30 and 60) in R Software for each pairs of (λ, k) , where $(k = 1, 2)$ and $(\lambda = 0.5, 1.0)$. The value for the loss parameter ($C_1 = -1, 1$) and $(a = 0.5, 1.0)$. The values of extension are $(C = 0.5, 1.0)$. The estimates of rate parameter for each method are calculated. The results are presented in the following tables.

N	k	λ	$\hat{\lambda}_{ML}$	$\hat{\lambda}_p$	$\hat{\lambda}_A$		$\hat{\lambda}_l$	
					c = -1	c = 1	a = 0.5	a = 1.0
15	1	0.5	0.15350	0.16179	0.10530	0.18428	0.13350	0.12597
	2	1.0	0.51869	0.54673	0.43100	0.58896	0.47443	0.44430
30	1	0.5	0.03480	0.03238	0.03238	0.03266	0.04191	0.03259
	2	1.0	0.06850	0.04258	0.07049	0.02085	0.03077	0.02171
60	1	0.5	0.03910	0.03320	0.02856	0.03488	0.03110	0.03060
	2	1.0	0.10349	0.09918	0.08981	0.10244	0.09398	0.09200

ML = maximum likelihood estimate, p = precautionary LF, A = Al-Bayyati's LF, l = LINEX LF.

Table 1.
Mean squared error for $\hat{\lambda}$ under Jeffrey's prior.

N	k	λ	d	$\hat{\lambda}_{ML}$	$\hat{\lambda}_p$	$\hat{\lambda}_A$		$\hat{\lambda}_l$	
						c = -1	c = 1	a = 0.5	a = 1.0
15	1	0.5	0.5	0.15350	0.14710	0.11081	0.18979	0.13901	0.13148
			1.0	0.15350	0.16179	0.10530	0.18428	0.13350	0.12597
	2	1.0	0.5	0.51869	0.51528	0.43951	0.59747	0.48294	0.45281
			1.0	0.51869	0.54673	0.43100	0.58896	0.47443	0.44430
30	1	0.5	0.5	0.03480	0.03286	0.02179	0.02206	0.21630	0.02200
			1.0	0.03480	0.03238	0.03238	0.03266	0.04191	0.03259
	2	1.0	0.5	0.06850	0.05828	0.06361	0.01397	0.02389	0.01483
			1.0	0.06850	0.04258	0.07049	0.02085	0.03077	0.02171
60	1	0.5	0.5	0.03910	0.03206	0.03206	0.03534	0.03156	0.03106
			1.0	0.03910	0.03320	0.02856	0.03488	0.03110	0.03060
	2	1.0	0.5	0.10349	0.09641	0.09021	0.10284	0.09439	0.09241
			1.0	0.10349	0.09918	0.08981	0.10244	0.09398	0.09200

ML = maximum likelihood estimate, p = precautionary LF, A = Al-Bayyati's LF, l = LINEX LF.

Table 2.
Mean squared error for $\hat{\lambda}$ under Quasi prior.

9. Comparison of Erlang distribution (ED) with its sub-models

The flexibility and potentiality of the Erlang distribution is compared with its sub models, which is examined by using different criteria like AIC, BIC and AICC with the help of the following illustration.

Illustration I:

We provide the compatibility of the Erlang distribution (ED) with their sub-models; Chi square and exponential distributions. For this purpose, we generated the data set for Erlang distribution of large sample size (i.e., 200) in R Software for each pairs of (λ, k) , where $(k = 2)$ and $(\lambda = 2.5)$. The data analysis is given in the following table:

Model	Parameter estimate	Standard error	Measures			
			$-\log l$	AIC	BIC	AICC
Exponential	$\hat{\lambda} = 0.5945$	0.04203	315.6853	633.3705	636.6689	581.9245
Chi-square	$k = 1.090$	0.0586	312.3096	625.6346	628.9329	633.3902
Erlang	$k = 2$		289.9327	581.8654	585.1637	625.6543
	$\hat{\lambda} = 1.1890$	0.0594				

Table 3.
AIC, BIC and AICC criterion for different sub-models of ED.

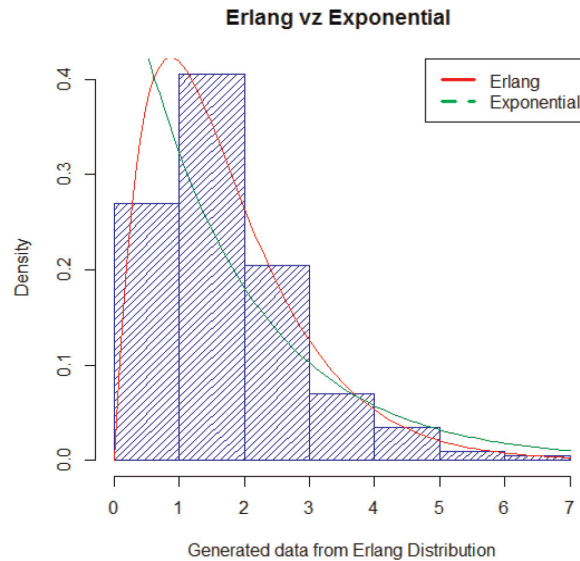


Illustration II:

The data set is taken from Lawless [48]. The observations involves the number of million revolutions between failures for each of 23 ball bearings, the individual bearings were inspected periodically to determine whether “failure” had occurred. Treating the failure times as continuous, the 23 failure times are:

(17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.40, 51.84, 51.96, 54.02, 55.56, 67.80, 68.64, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04, 173.40)

Model	Parameter estimate	Standard error	Measures			
			$-\log l$	AIC	BIC	AICC
Exponential	$\hat{\lambda} = 0.01387$	0.002877	121.4324	244.8648	246.0003	245.0248
Chi-square	$k = 2.9420$	1.1745	177.5835	357.1670	358.3025	357.327
Erlang	$k = 1$		113.0298	230.0597	232.3307	230.5212
	$\hat{\lambda} = 0.05577$	0.01683				

Table 4.
AIC, BIC and AICC criterion for different sub-models of ED.

10. Results and discussion

We primarily studied the maximum likelihood (MLH) estimation and Bayesian estimation to estimate the rate parameter of Erlang distribution. In Bayesian method, we use Jeffreys' prior and Quasi prior under three different loss functions. These methods are compared through simulation technique and the results are presented in the **Tables 1** and **2** respectively.

From the results obtained in **Tables 1** and **2**, we observe that in most of the cases, Bayesian estimator under Al-Bayyati's loss function has the smallest mean squared error (MSE) values for Jeffrey's prior and Quasi prior as compared to other loss functions and the maximum likelihood estimator. Thus we can conclude that Bayes estimator under Al-Bayyati's loss function is efficient when the loss parameter C is -1 .

Also we estimated the unknown parameters of sub-models of Erlang distribution. The Akaike information criterion (AIC), Bayesian information criterion (BIC), and the corrected Akaike information criterion (AICC) are used to compare the candidate distributions. The best distribution corresponds to lower $-\log L$, AIC, BIC, AICC statistics value.

From the results obtained in **Tables 3** and **4**, we observe that the Erlang distribution is a competitive distribution as compared to its sub-models (i.e., exponential distribution and Chi-square distribution). In fact, based on the values of the AIC, BIC and AICC criteria, it shows clear picture that the Erlang distribution provides the best fit for these data among all the models considered.

11. Conclusions

In this paper we have generated three types of data sets with varying sample sizes for Erlang distribution. These data sets were simulated and behavior of the data was checked in case of parameter estimation for Erlang distribution in R Software. By the virtue of the data analysis we are able predict the estimate of rate parameter for Erlang distribution under three different functions by using two different prior distributions. With the help of these results we can also do comparison between loss functions and the priors.

Also the comparison of Erlang distribution with its sub-models was carried out. The results acquired in **Tables 3** and **4**, it shows the clear picture that Erlang distribution performs better as compared to its sub-models. Thus we can say that Erlang distribution is efficient as compared to its sub-models (i.e., exponential distribution and Chi-square distribution) on the basis of the above procedures.

Acknowledgements

The authors acknowledge editor of this journal for encouragement to finalize the chapter. Further, the authors acknowledge profound thanks to anonymous referee for giving critical comments which have immensely improved the presentation of the chapter. Also I extend my sincere thanks to all the authors whose papers I have consulted for this work.

Conflict of interest

The authors declare that there is no conflict of interests regarding the publication of this chapter.

Author details


Kaisar Ahmad^{1*} and Sheikh Parvaiz Ahmad²

1 Department of Statistics, Sheikh-ul-Alam Memorial Degree College, Budgam, J&K, India

2 Department of Statistics, University of Kashmir, Srinagar, J&K, India

*Address all correspondence to: ahmadkaisar31@gmail.com

IntechOpen

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Erlang AK. The theory of probabilities and telephone conversations. *Nyt Tidsskrift for Matematik B*. 1909;20(6):87-98
- [2] Evans M, Hastings N, Peacock B. *Statistical Distributions*. 3rd ed. New York: John Wiley and Sons, Inc.; 2000
- [3] Bhattacharyya SK, Singh NK. Bayesian estimation of the traffic intensity in M/Ek/1 queue. *Far. East. Journal of Mathematical Sciences*. 1994; 2:57-62
- [4] Haq A, Dey S. Bayesian estimation of Erlang distribution under different prior distributions. *Journal of Reliability and Statistical Studies*. 2001;4(1):1-30
- [5] Suri PK, Bhushan B, Jolly A. Time estimation for project management life cycles: A simulation approach. *International Journal of Computer Science and Network Security*. 2009; 9(5):211-215
- [6] Damodaran D, Gopal G, Kapur PK. A Bayesian Erlang software reliability model. *Communication in Dependability and Quality Management*. 2010;13(4):82-90
- [7] Jodra P. Computing the asymptotic expansion of the median of the Erlang distribution. *Mathematical Modelling and Analysis*. 2012;17(2):281-292
- [8] Zellner A. Bayesian estimation and prediction using asymmetric loss function [PhD thesis]. *Journal of American Statistical Association* exponential distribution using simulation. Iraq: Baghdad University; 1986. vol. 81, pp. 446-451
- [9] Ahmad SP, Ahmad K. Bayesian analysis of Weibull distribution using R Software. *Australian Journal of Basic and Applied Sciences*. 2013;7(9): 156-164. ISSN: 1991-8178
- [10] Ahmad K, Ahmad SP, Ahmed A, Reshi JA. Bayesian analysis of generalized gamma distribution using R Software. *Journal of Statistics Applications & Probability*. 2015;2(4): 323-335
- [11] Raffia H, Schlaifer R. *Applied Statistical Decision Theory*. Division of Research, Graduate School of Business Administration, Harvard University; 1961
- [12] DeGroot MH. *Optimal Statistical Decisions*. New York: McGraw-Hill; 1970
- [13] Zellner A. Bayesian and non-Bayesian analysis of the log-normal distribution and log normal regression. *Journal of the American Statistical Association*. 1971;66:327-330
- [14] Box GEP, Tiao GC. *Bayesian Inference in Statistical Analysis*. New York: John Wiley and Sons; 1973
- [15] Bayes TR. An Essay Towards Solving Problems In The Doctrine Of Chances. *Philosophical Transactions of the Royal Society A*. 1763;53:370-418 Reprinted in *Biometrika*, 48, 296-315
- [16] Laplace PS. *Essai Philosophique sur Les Probabilités*, Paris. This Book Went Through Five Editions (The Fifth was in 1825) Revised By Laplace. The Sixth Edition Appeared In English Translation By Dover Publications, New York, In 1951. While This Philosophical Essay Appeared Separately In 1814, It Also Appeared As A Preface To His Earlier Work. *Theorie Analytique Des Probabilités*; 1774. pp. 621-656
- [17] Jeffrey H. *Theory of Probability*. III edition (I edition 1939, II edition 1994) ed. Oxford University Press: Clarendon Press; 1961
- [18] Anscombe F, Alhmann R. A definition of subjective Probability. *The*

- Annals of Mathematical Statistics. 1963; **34**(1):199-205
- [19] Berger JO. Statistical Decision Theory and Bayesian Analysis. 2nd ed. New York: Springer-Verlag; 1985
- [20] Berger JO, Bernardo JM. Estimating a product of means: Bayesian analysis with reference priors. American Journal of Mathematical Statistical Association. 1989;**84**:200-207
- [21] Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis. London: Chapman and Hall; 1995
- [22] Leonardo T, Hsu JSJ. Bayesian Methods. Cambridge: Cambridge University Press; 1999
- [23] De Finetti B. A critical essay on the theory of probability and on the value of science. "probabilismo". Erkenntnis. 1931;**31**:169-223 English translation as "Probabilism"
- [24] Bernardo JM, Smith AFM. Bayesian Theory. Chichester, West Sussex: John Wiley and Sons; 2000
- [25] Robert CP. The Bayesian Choice. 2nd ed. New York: Springer-Verlag; 2001
- [26] Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis. 2nd ed. Boca Raton, FL: Chapman and Hall/CRC; 2004
- [27] Marin JM, Robert CP. Bayesian Core: A Practical Approach to Computational Bayesian Statistics. Springer-Verlag; 2007
- [28] Carlin BP, Louis TA. Bayesian Methods for Data Analysis. IIrd ed. Boca Raton, FL: Chapman and Hall/CRC; 2008
- [29] Ibrahim JG, Chen M-H, Sinha D. Bayesian Survival Analysis. New York: Springer Verlag; 2001
- [30] Ghosh JK, Delampady M, Samanta T. An Introduction to Bayesian Analysis Theory and Methods. Springer-Verlag; 2006
- [31] Bansal AK. Bayesian Parametric Inference. New Dehli: Narosa Publishing House; 2007
- [32] Koch KR. Introduction to Bayesian Statistics. New York: Springer-Verlag; 2007
- [33] Hoff PD. A First Course in Bayesian Statistical Methods. Springer-Verlag; 2009
- [34] Ahmad K, Ahmad SP, Ahmed A. Classical and Bayesian approach in estimation of scale parameter of inverse Weibull distribution. Mathematical Theory and Modeling. 2015;**5**. ISSN 2224-5804
- [35] Ahmad K, Ahmad SP, Ahmed A. Classical and Bayesian approach in estimation of scale parameter of Nakagami distribution. Journal of Probability and Statistics. 2016;**2016** Article ID 7581918, 1-8
- [36] Jefferys H. An invariant form for the prior probability in estimation problems. Proceedings of The Royal Society Of London, Series. A. 1946;**186**: 453-461
- [37] Wald A. Statistical Decision Functions. Wiley; 1950
- [38] Norstrom JG. The use of precautionary loss functions in risk analysis. IEEE Transactions on Reliability. 1996;**3**:400-403
- [39] Al-Bayyati. Comparing methods of estimating Weibull failure models using simulation [PhD thesis]. Iraq: College of Administration and Economics, Baghdad University; 2002
- [40] Klebnov LB. Universal loss function and unbiased estimation. Dokl. Akad.

Nank SSSR Soviet-Maths. Dokl. T. 1972;
203:1249-1251

[41] Varian HR. A Bayesian approach to real estate assessment. In: Fienberg SE, Zellner A, editors. *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage*. Amsterdam: North-Holland; 1975. pp. 195-208

[42] Shannon CE. A mathematical theory of communication. *Bell System Technical Journal*. 1948;**27**:623-659

[43] Cover TM, Thomas JA. *Elements of Information Theory*. New York: Wiley; 1991

[44] Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, editors. *2nd International Symposium on Information Theory*. Budapest: Akademia Kiado; 1973. pp. 267-281

[45] Hurvich CM, Tsai CL. Regression and time series model selection in small samples. *Biometrika*. 1989;**76**:297-307

[46] Burnham KP, Anderson DR. Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*. 2004; **33**:261-304

[47] Burnham KP, Anderson DR. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd ed. Springer-Verlag; 2002

[48] Lawless JF. *Statistical Models and Methods for Lifetime Data*. 2nd ed. Wiley-InterScience; 2002