Austrian Journal of Statistics
2024, Volume 53, 53–69.
http://www.ajs.or.at/
doi:10.17713/ajs.v53i3.1694

A Relationship between Classical and Bayesian Estimation Procedures through Fisher Information

Shreya Bhunia
Aliah University
Department of Mathematics and Statistics

Babulal Seal
Aliah University
Department of Mathematics and Statistics

Abstract

The aim of this article is to investigate the association between the classical and Bayesian approaches through Fisher information. For any particular distribution, the computation of Fisher information is quite significant, as it provides the amount of information about the unknown parameter inferred from the observed data and is related to classical methods of estimation. Also, in the light of some prior knowledge, we may estimate the unknown parameter through Bayesian approach. Specifically, we want to see a relationship between information and Bayes estimation. In this article, the scale parameter of the one-parameter exponential distribution is estimated under the weighted squared error and Kullback-Leibler distance loss functions. The information acquired from both the classical and Bayesian methodologies have been connected through the risk intensity and error intensity which have been introduced in this article. The results of extensive simulation studies using these intensity measures show that the Bayes estimator performs more intensely as the amount of Fisher information increases. It is seen that the Fisher information, which is pivotal to many classical estimation methods, has a relationship with the Bayesian method depending on prior distribution, at least in this case, as the intensity measures of the Bayes estimator decrease with the increase in information. Further, to comprehend the theoretical notion of association, two real-life datasets have been included to show usefulness in practical field.

Keywords: Fisher information, risk intensity, error intensity, Kullback-Leibler divergence.

1. Introduction

In statistics, Fisher information plays an important role both in the classical and Bayesian inference. In frequentist context, Fisher information provides knowledge about the parameter gained from the observed data. It is used to compute the asymptotic variance of the maximum likelihood estimator. Also, the Cramér-Rao inequality states that the inverse of Fisher information is a lower bound for the variance of any unbiased estimator. Meanwhile, in Bayesian estimation problem, the Fisher information is used to define the Jeffreys prior which is commonly used as a non-informative prior. In addition to these fundamental characteristics, the use of Fisher information serves a variety of other applications also. Goel and Degroot (1981) defined a Bayesian analog of the Fisher information, which provides the information that an observed sample has about the posterior distribution. Fisher information can also be expressed in terms of hazard function which is investigated by Efron and Johnstone (1990). Also, some thought-provoking insights on the relationship between Bayesian and classical estimation using the continuous uniform distribution and exponential distribution have been explained by Rossman, Short, and Parks (1998) and Elfessi and Reineke (2001) respectively. They explored how different parameter combinations may be used to obtain the classical estimators from the Bayes estimators under different prior selections.

Despite the fact that the connections between the classical and Bayesian approaches are well-known in literature. However, the purpose of this article is to establish a relationship between them depending on how much knowledge the Fisher information and the prior information provide about the parameter. When the model parameter is unknown and we do not have any prior knowledge, the traditional Fisher information can provide us some insight about the parameter. Whereas, in Bayesian context, prior knowledge of the model parameter plays an important role to form the posterior distribution based on the additional information (Banerjee and Seal 2022). This article attempts to provide a relationship between the classical and Bayesian techniques by defining the risk and error intensities, which quantify the amount of contribution, a Bayes estimator makes per unit of information. In this article, we see the above-discussed relationship in exponential distribution which has the following probability density function (pdf)

$$f(x;\theta) = \theta e^{-\theta x}, \quad x > 0, \theta > 0, \tag{1}$$

and cumulative distribution function (cdf)

$$F(x;\theta) = 1 - e^{-\theta x}, \quad x > 0, \theta > 0. \tag{2}$$

In statistics and probability, the exponential distribution has been extensively used in analyzing lifetime data, particularly where the associated hazard rate is constant. The exponential distribution is the continuous analog of the geometric distribution and it is also a special case of some well-known statistical models, including the Weibull, Rayleigh, gamma, and Erlang distributions (Dhungana and Kumar 2022). Due to the mathematical simplicity and its lack of memory property, this distribution has received considerable attention among the researchers. Jaheen (2004) derived the empirical Bayes estimator for the parameter of the exponential distribution based on the record values. The upper record range statistic has been considered by Nasiri, Hosseini, and Yarmohammadi (2012) to draw inference from the parameter of the exponential distribution. The Bayes estimates of the parameter of the exponential distribution under Kullback-Leibler distance loss and squared error loss function (SELF) are derived and compared by Abufoudeh, Awwad, and Bdair (2019). Abu Awwad, Abufoudeh, and Bdair (2019) estimated the parameter of the exponential distribution using both classical and Bayesian approaches and also predicted the future record values depending on a sequence of past records.

In this paper, relationship between the Fisher information and the Bayes estimate is described for the scale parameter θ of the exponential distribution. The Bayes estimate of θ is obtained by assuming gamma distribution as the prior information, where both the prior parameters being known. In classical way, the Fisher information is used to measure the available information in the sample. Both the weighted squared error loss and the loss derived by utilizing the densities of exponential model in Kullback-Leibler distance measure are used to derive the Bayes estimates. Risk intensity and error intensity are the two measures introduced here to compare the performance of the Bayes estimators. These measures are actually defined as contribution of the Bayes estimator for per unit information. For comparison purpose, an extensive simulation study is performed for different combinations of prior parameters and varying sample sizes. Two datasets namely, vinyl chloride data and the survival times (in years) of a group of gastric cancer patients are also considered to illustrate the performance of the methodology in real-life situations. For these two datasets, exponential distribution is fitted well and the performance of Bayes estimator becomes better when the amount of information increases.

The content of this article is organized as follows. The Bayes estimators under both KL distance loss and WSELF as well as the Fisher information of the scale parameter θ of exponential distribution are derived in Section 2. To compare the performance of the estimators as well as to examine the relationship between the Fisher information and the Bayes estimators, numerical results obtained through simulation for different parameter combinations are presented in Section 3. Two real-life data have been used in Section 4 to illustrate these in real situations. Finally, the conclusion of the work is discussed in Section 5.

2. Bayes estimation and Fisher information of the scale parameter

Specification of the prior knowledge for the target parameter θ, is the fundamental step to obtain the Bayes estimate. Here, we consider the conjugate prior due to its appealing analytical properties. However, the conjugate prior requires the specification of a few hyper-parameters.

The prior distribution of θ is taken as, $\theta \sim Gamma(\alpha, \beta)$,

$$i.e.\ \ \pi(\theta) = \frac{\beta^{\alpha}}{\Gamma\alpha}\theta^{\alpha-1}e^{-\beta\,\theta};\ \ \theta > 0,\ \alpha > 0,\ \beta > 0,$$

where, α and β are the respective shape and scale parameters.

Once the prior is selected, the posterior distribution is defined as the proportional to the likelihood function for the data given parameter multiplied by the prior distribution (Banerjee and Bhunia 2022). Then, the posterior distribution of θ becomes,

$$\begin{aligned} \Pi(\theta \mid \mathbf{x}) &= \frac{\theta^{(n+\alpha)-1}e^{-(\beta+\sum_{i=1}^{n}x_i)}}{\int_0^\infty \theta^{(n+\alpha)-1}e^{-(\beta+\sum_{i=1}^{n}x_i)}\ d\theta} \\ &= \frac{(\beta+\sum_{i=1}^{n}x_i)^{n+\alpha}}{\Gamma(n+\alpha)}\theta^{(n+\alpha)-1}e^{-(\beta+\sum_{i=1}^{n}x_i)\theta}. \end{aligned} \tag{3}$$

Therefore, the posterior distribution of $\theta \mid \mathbf{x}$ belongs to the gamma family with parameters $(n+\alpha)$ and $(\beta + \sum_{i=1}^{n} x_i)$.

$$i.e.\ \ \theta \mid \mathbf{x} \sim Gamma(n+\alpha,\ \beta + \sum_{i=1}^{n} x_i).$$

Now, the Bayes estimates of θ will be derived under a distance loss function using Kullback-Leibler divergence measure and the traditionally used weighted squared error loss function respectively.

2.1. Kullback-Leibler distance loss function

In statistics, the Kullback-Leibler divergence measure, also known as the relative entropy measure, is successfully utilized to calculate the distance loss function (Kullback and Leibler 1951). This type of loss functions can be obtained directly from the measure of discrepancies between two probability density functions. In literature, the term 'intrinsic loss' is used alternatively to indicate this kind of loss functions (Rong, Tang, and Zhou 2017; Ni and Sun 2021).

For a continuous random variable X, suppose $f(x)$ and $g(x)$ are two probability density functions, then the KL divergence measure is given as,

$$D_{KL}(f(x)||g(x)) = E_f\left[log\left(\frac{f(x)}{g(x)}\right)\right] = \int_{-\infty}^{\infty} f(x)\,log\left(\frac{f(x)}{g(x)}\right)dx, \tag{4}$$

where the expectation is taken with respect to the density function $f(x)$.

Now, in this study, we consider two density functions as $f(x \mid \theta_1)$ and $f(x \mid \theta_2)$, from the exponential distribution to derive the KL distance loss function using (4).

$$\begin{aligned} D(f(. \mid \theta_1)||f(. \mid \theta_2)) &= E_{\theta_1}\left[log\left(\frac{f(x \mid \theta_1)}{f(x \mid \theta_2)}\right)\right] \\ &= E_{\theta_1}\left[log\left(\frac{\theta_1 e^{-\theta_1 x}}{\theta_2 e^{-\theta_2 x}}\right)\right] \\ &= log\left(\frac{\theta_1}{\theta_2}\right) + (\theta_2 - \theta_1)\frac{1}{\theta_1} \\ &= log\left(\frac{\theta_1}{\theta_2}\right) + \frac{\theta_2}{\theta_1} - 1. \end{aligned}$$

Therefore, the loss function for the scale parameter θ can be expressed as,

$$L(\delta_1(x), \theta) = log\left(\frac{\delta_1(x)}{\theta}\right) + \frac{\theta}{\delta_1(x)} - 1. \tag{5}$$

The posterior risk of the obtained loss function is expressed here as,

$$\begin{aligned} & E_{\theta|\mathbf{x}}\left[L(\delta_1(x), \theta)\right] \\ = \; & E_{\theta|\mathbf{x}}\left[log\delta_1(x) - log\theta + \frac{\theta}{\delta_1(x)} - 1\right] \\ = \; & log\delta_1(x) - E[log\theta \mid \mathbf{x}] + \frac{1}{\delta_1(x)}E[\theta \mid \mathbf{x}] - 1. \end{aligned}$$

Now, after minimizing the above posterior risk, we obtain the Bayes estimate as $\delta_1(x) = E(\theta \mid \mathbf{x})$, which is nothing but the posterior mean. Therefore, the final expression for the Bayes estimate under KL distance loss function becomes,

$$\delta_1(x) = \frac{n + \alpha}{\beta + \sum_{i=1}^{n} x_i}. \tag{6}$$

2.2. Weighted squared error loss function

In Bayesian inference, the most commonly used loss function is the squared error loss function. This is a symmetric loss which assumes that the under-estimation and over-estimation of the same magnitude are equally serious. But, for estimating the scale parameter, Ferguson (2014) modified this loss function and defined it as,

$$L(\theta, \delta(x)) = \left(\frac{\delta(x)}{\theta} - 1\right)^2,$$

where, $\delta(x)$ is the estimate of θ. Generalization on the squared error loss is termed as the weighted squared error loss function (WSELF). Hence, for the scale parameter θ, the considered WSELF becomes,

$$L(\theta, \delta_2(x)) = (\delta_2(x) - \theta)^2 \, \frac{1}{\theta^2}; \quad with \quad w(\theta) = \frac{1}{\theta^2}. \tag{7}$$

The expression for the Bayes estimate under WSELF can be written as,

$$\delta_2(x) = \frac{E(\theta^{-1} \mid \mathbf{x})}{E(\theta^{-2} \mid \mathbf{x})}.$$

So, the Bayes estimator is derived as,

$$\begin{aligned} \delta_2(x) &= \frac{\int_0^\infty \theta^{(n+\alpha-1)-1} e^{-(\beta+\sum_{i=1}^n x_i)\theta} \, d\theta}{\int_0^\infty \theta^{(n+\alpha-2)-1} e^{-(\beta+\sum_{i=1}^n x_i)\theta} \, d\theta} \\ &= \frac{(n+\alpha-2)\Gamma(n+\alpha-2)}{\Gamma(n+\alpha-2)} (\beta + \sum_{i=1}^n x_i)^{-1} \\ &= \frac{n+\alpha-2}{\beta+\sum_{i=1}^n x_i}. \end{aligned} \tag{8}$$

2.3. Fisher information

The amount of information, we gain about an unknown parameter from a sample is described by the Fisher information. Suppose, X is the continuous random variable having density function (1) with the scale parameter θ, then the information from X about θ is expressed as the expected amount of the second derivation of the log-likelihood function under the standard regularity condition (Lehmann and Casella 2006).

Let, $x_1, x_2, \cdots, x_n$ be a random sample from the exponential density (1), then the log-likelihood function is

$$\begin{aligned} log\, f(\mathbf{x} \mid \theta) &= \sum_{i=1}^n log\, f(x_i \mid \theta) \\ &= nlog\theta - \theta \sum_{i=1}^n x_i. \end{aligned} \tag{9}$$

Now, the Fisher information quantity, denoted as $I(\theta)$ is

$$\begin{aligned} I(\theta) &= -E_\theta \left[\frac{-\partial^2 log f(\mathbf{x} \mid \theta)}{\partial \theta^2} \right] \\ &= -E_\theta \left[-\frac{n}{\theta^2} \right] \\ &= \frac{n}{\theta^2}. \end{aligned} \tag{10}$$

3. Relationship between classical and Bayesian approach through simulation study

In this section, an attempt has been made to formulate a relationship between the classical and Bayesian methods by gathering information from both estimation methods. According to the classical idea, we can gain some knowledge regarding the parameter of interest through the Fisher information quantity. Whereas, under Bayesian paradigm, the prior knowledge about the concerned parameter plays an important role to derive the Bayes estimator. So, Fisher information and the Bayes estimator are taken as the fundamental pillars in order to investigate this relationship. Now, being an estimator of the scale parameter θ, $\delta(x)$ is associated with an amount of risk $R(\theta, \delta(x))$, which quantifies the behaviour of the estimator. Also, the goodness of the estimator has been studied by calculating the absolute deviation of the estimator $\delta(x)$ from the true parameter value θ. It actually represents the error that is occurred while taking the decision $\delta(x)$. Both the absolute deviation and risk values have been obtained to determine the performance of the Bayes estimator. Thus, we investigate the relationship between the classical and Bayesian method through Fisher information by introducing the following measures on the basis of the risk and error quantities.

- **Risk Intensity:** It is defined as the risk per unit of information and is expressed as the ratio between the risk value and the Fisher information.

$$R_I = \frac{R(\theta, \delta(\cdot))}{I(\theta)}. \tag{11}$$

- **Error Intensity:** It measures the error per unit of information and is expressed as the ratio between the average absolute deviation and the Fisher information.

$$E_I = \frac{E \mid \theta - \delta(\cdot) \mid}{I(\theta)}. \tag{12}$$

It measures the intensity of a Bayes estimator for per unit information. More specifically, it quantifies how much contribution has been made by a Bayes estimator for per unit of information. As, the analytical derivations of the intensity measures seem to be difficult, we proceed further by doing an extensive simulation study with the help of following algorithm.

Step 1: Fix some known values of the hyper-parameters α and β.

Step 2: Generate $\theta_1, \theta_2, \cdots \theta_m$ from $Gamma(\alpha, \beta)$ and consider the first realization as θ.

Step 3: Generate random sample **X** of size n from $exp(\theta)$.

Step 4: Calculate the Fisher information $I(\theta)$ from (10).

Step 5: Repeat **Step 3** for K times to generate K simulated samples $X_1, X_2, \cdots, X_k$.

Step 6: Obtain the Bayes estimators $\delta_1(x)$ and $\delta_2(x)$ by using (6) and (8) respectively.

Step 7: Calculate the absolute deviation $|\theta - \delta_1(x)|$ and $|\theta - \delta_2(x)|$ and take an average over K replicates.

Step 8: Obtain loss values K times using $L(\delta_1(x), \theta)$ and $L(\theta, \delta_2(x))$ mentioned in (5) and (7) respectively.

Step 9: Derive the risk values $R(\delta_1(x), \theta)$ and $R(\theta, \delta_2(x))$ by taking average over the respective loss values.

Step 10: Calculate the relative measures by using (11) and (12) for both the loss functions.

Step 11: Repeat **Step 3-10** for the remaining generated $\theta_i's$; $i = 2, 3, \cdots, m$.

Initially, the hyper-parameters of the prior distribution are chosen as (α, β) = (3.5, 2), (2, 0.5), (5, 3), (4,0.75) and for each pair, we generate m = 10 times θ from Gamma(α, β). Now, for each θ_i, random samples of size n = 25, 50, 75, 100 are simulated from the baseline distribution. The study is conducted over $K = 5000$ simulated samples by using R Core Team (2021). The average absolute deviations, risk values of the Bayes estimator $\hat{\theta}_i$ are obtained over the K replications and then we divide them by the corresponding Fisher information value $I(\theta_i)$ to determine the risk and error intensity measures. The association between $I(\theta)$ and the Bayes estimators $\hat{\theta}$ has been evaluated based on the simulation results which are presented in Tables 1 - 4. It may also be noted that, while displaying the simulation results graphically, we have employed certain multipliers with the Fisher information values to plot them with the intensity measures into a fixed range.

From the simulation study, it has been observed that as the sample size increases the information also increases. Both the risk and error intensities of the estimator decrease gradually when the sample size increases. The risk and error intensities of the Bayes estimator increase with the decrease in the Fisher information amount. The risk intensity under the KL distance

Table 1: Fisher information with the error intensity and risk intensity of θ when α = 3.5, β = 2

Sample sizes	Generated θ	$I(\theta)$	Under WSELF E_I	Under WSELF R_I	Under KL E_I	Under KL R_I
n = 25	0.2168	531.8895	0.0000736	0.0001072	0.0000931	0.0000519
	0.5085	96.6848	0.0008857	0.0005005	0.0011464	0.0002674
	0.7995	39.1114	0.0032473	0.0010700	0.0042571	0.0006201
	0.9477	27.8355	0.0052741	0.0014088	0.0069302	0.0008440
	1.1052	20.4672	0.0081757	0.0018018	0.0107328	0.0011104
	1.3656	13.4058	0.0149909	0.0025341	0.0194803	0.0016073
	1.8369	7.4092	0.0358121	0.0042303	0.0443493	0.0026575
	2.1016	5.6603	0.0539102	0.0054998	0.0641004	0.0033211
	2.8611	3.0540	0.1463043	0.0112971	0.1474976	0.0055083
	3.8818	1.6591	0.4349131	0.0276051	0.3354107	0.0093149
n = 50	0.2168	1063.7791	0.0000243	0.0000223	0.0000278	0.0000111
	0.5085	193.3696	0.0003027	0.0001129	0.0003490	0.0000590
	0.7995	78.2227	0.0011429	0.0002597	0.0013222	0.0001405
	0.9477	55.6709	0.0018807	0.0003535	0.0021745	0.0001939
	1.1052	40.9344	0.0029499	0.0004666	0.0034040	0.0002588
	1.3656	26.8116	0.0054871	0.0006843	0.0062881	0.0003836
	1.8369	14.8183	0.0132106	0.0011874	0.0147650	0.0006611
	2.1016	11.3206	0.0198303	0.0015452	0.0216983	0.0008445
	2.8611	6.1081	0.0520811	0.0029998	0.0521883	0.0014790
	3.8818	3.3182	0.1447990	0.0065333	0.1242736	0.0026119
n = 75	0.2168	1595.6686	0.0000129	0.0000094	0.0000142	0.0000047
	0.5085	290.0543	0.0001622	0.0000487	0.0001799	0.0000251
	0.7995	117.3341	0.0006181	0.0001145	0.0006861	0.0000605
	0.9477	83.5064	0.0010208	0.0001574	0.0011324	0.0000840
	1.1052	61.4016	0.0016065	0.0002097	0.0017790	0.0001127
	1.3656	40.2174	0.0030011	0.0003113	0.0033057	0.0001684
	1.8369	22.2275	0.0072510	0.0005461	0.0078420	0.0002944
	2.1016	16.9809	0.0108736	0.0007107	0.0115878	0.0003788
	2.8611	9.1621	0.0281175	0.0013518	0.0282824	0.0006744
	3.8818	4.9773	0.0758182	0.0027944	0.0683296	0.0012043
n = 100	0.2168	2127.5582	0.0000083	0.0000050	0.0000089	0.0000025
	0.5085	386.7391	0.0001046	0.0000265	0.0001129	0.0000137
	0.7995	156.4455	0.0004008	0.0000632	0.0004329	0.0000331
	0.9477	111.3418	0.0006637	0.0000874	0.0007160	0.0000461
	1.1052	81.8688	0.0010470	0.0001172	0.0011276	0.0000621
	1.3656	53.6232	0.0019628	0.0001757	0.0021033	0.0000934
	1.8369	29.6366	0.0047576	0.0003119	0.0050236	0.0001649
	2.1016	22.6412	0.0071373	0.0004074	0.0074502	0.0002133
	2.8611	12.2161	0.0183702	0.0007742	0.0183414	0.0003847
	3.8818	6.6364	0.0487318	0.0015685	0.0447212	0.0006952

Table 2: Fisher information with the error intensity and risk intensity of θ when α = 2.0, β = 0.5

Sample sizes	Generated θ	$I(\theta)$	Under WSELF E_I	Under WSELF R_I	Under KL E_I	Under KL R_I
n = 25	1.1467	19.0126	0.0098170	0.0022854	0.0103711	0.0010197
	1.5028	11.0697	0.0218097	0.0037864	0.0224047	0.0016731
	1.5118	10.9383	0.0221970	0.0038285	0.0227871	0.0016915
	1.9187	6.7909	0.0447808	0.0059391	0.0446935	0.0026158
	2.6236	3.6320	0.1124083	0.0105017	0.1078482	0.0047039
	3.9727	1.5840	0.3833096	0.0224108	0.3560793	0.0112158
	4.7607	1.1031	0.6587123	0.0315349	0.6168807	0.0174801
	7.1049	0.4952	2.2604123	0.0720139	2.3631928	0.0568726
	7.9422	0.3963	3.2313417	0.0930883	3.5618769	0.0827155
	10.3878	0.2317	7.9077413	0.1841034	10.0503240	0.2162744
n = 50	1.1467	38.0251	0.0034132	0.0005426	0.0035094	0.0002554
	1.5028	22.1395	0.0076357	0.0009164	0.0077378	0.0004289
	1.5118	21.8767	0.0077726	0.0009271	0.0078738	0.0004339
	1.9187	13.5818	0.0157898	0.0014674	0.0157647	0.0006853
	2.6236	7.2640	0.0400200	0.0026732	0.0391817	0.0012585
	3.9727	3.1681	0.1376500	0.0059237	0.1324692	0.0029504
	4.7607	2.2061	0.2366788	0.0084196	0.2285479	0.0044302
	7.1049	0.9905	0.8006407	0.0189293	0.8240474	0.0124197
	7.9422	0.7927	1.1332720	0.0240463	1.2061488	0.0172068
	10.3878	0.4634	2.6778158	0.0445449	3.1708695	0.0405125
n = 75	1.1467	57.0377	0.0018371	0.0002376	0.0018747	0.0001138
	1.5028	33.2092	0.0041179	0.0004034	0.0041604	0.0001924
	1.5118	32.8150	0.0041919	0.0004082	0.0042342	0.0001946
	1.9187	20.3726	0.0085327	0.0006496	0.0085342	0.0003091
	2.6236	10.8960	0.0216859	0.0011930	0.0213900	0.0005701
	3.9727	4.7521	0.0748337	0.0026697	0.0729127	0.0013225
	4.7607	3.3092	0.1287096	0.0038033	0.1256437	0.0019546
	7.1049	1.4857	0.4323255	0.0084933	0.4402062	0.0051186
	7.9422	1.1890	0.6089175	0.0107188	0.6363288	0.0069114
	10.3878	0.6950	1.4141172	0.0193275	1.6119308	0.0152568
n = 100	1.1467	76.0502	0.0011898	0.0001310	0.0012067	0.0000641
	1.5028	44.2790	0.0026706	0.0002232	0.0026877	0.0001089
	1.5118	43.7534	0.0027187	0.0002258	0.0027356	0.0001102
	1.9187	27.1635	0.0055424	0.0003609	0.0055347	0.0001759
	2.6236	14.5280	0.0141175	0.0006668	0.0139554	0.0003265
	3.9727	6.3362	0.0488421	0.0015062	0.0478978	0.0007595
	4.7607	4.4122	0.0840685	0.0021540	0.0825987	0.0011179
	7.1049	1.9810	0.2820841	0.0048305	0.2861611	0.0028415
	7.9422	1.5853	0.3966343	0.0060914	0.4106815	0.0037833
	10.3878	0.9267	0.9133839	0.0108931	1.0191121	0.0080104

Table 3: Fisher information with the error intensity and risk intensity of θ when α = 5.0, β = 3.0

Sample sizes	Generated θ	$I(\theta)$	Under WSELF E_I	Under WSELF R_I	Under KL E_I	Under KL R_I
n = 25	0.3660	186.6284	0.0003709	0.0003363	0.0005194	0.0001796
	0.6513	58.9356	0.0018632	0.0008283	0.0027937	0.0005325
	0.9117	30.0771	0.0046842	0.0013227	0.0073451	0.0009825
	1.0396	23.1317	0.0067062	0.0015756	0.0106673	0.0012404
	1.1732	18.1633	0.0093442	0.0018522	0.0150043	0.0015319
	1.3898	12.9430	0.0149821	0.0023506	0.0240935	0.0020458
	1.7725	7.9573	0.0305816	0.0035358	0.0470500	0.0030522
	1.9836	6.3538	0.0434242	0.0044724	0.0638141	0.0036476
	2.5787	3.7596	0.1064150	0.0089748	0.1281432	0.0054282
	3.3615	2.2125	0.2932022	0.0216659	0.2543766	0.0079321
n = 50	0.3660	373.2569	0.0001202	0.0000678	0.0001475	0.0000363
	0.6513	117.8712	0.0006374	0.0001875	0.0008075	0.0001103
	0.9117	60.1542	0.0016722	0.0003304	0.0021585	0.0002079
	1.0396	46.2634	0.0024358	0.0004109	0.0031607	0.0002654
	1.1732	36.3267	0.0034474	0.0005026	0.0044843	0.0003316
	1.3898	25.8860	0.0056262	0.0006700	0.0073022	0.0004513
	1.7725	15.9147	0.0115452	0.0010422	0.0146185	0.0006967
	1.9836	12.7075	0.0162820	0.0013057	0.0201030	0.0008487
	2.5787	7.5191	0.0380274	0.0023975	0.0419781	0.0013330
	3.3615	4.4249	0.0967489	0.0050994	0.0875711	0.0020849
n = 75	0.3660	559.8853	0.0000633	0.0000280	0.0000737	0.0000149
	0.6513	176.8069	0.0003416	0.0000808	0.0004066	0.0000456
	0.9117	90.2313	0.0009088	0.0001470	0.0010944	0.0000868
	1.0396	69.3951	0.0013310	0.0001853	0.0016077	0.0001114
	1.1732	54.4900	0.0018922	0.0002295	0.0022887	0.0001398
	1.3898	38.8290	0.0031049	0.0003106	0.0037474	0.0001918
	1.7725	23.8720	0.0063918	0.0004881	0.0075747	0.0003003
	1.9836	19.0613	0.0089894	0.0006097	0.0104695	0.0003686
	2.5787	11.2787	0.0205320	0.0010832	0.0221645	0.0005910
	3.3615	6.6374	0.0503569	0.0021621	0.0470221	0.0009466
n = 100	0.3660	746.5138	0.0000403	0.0000148	0.0000453	0.0000078
	0.6513	235.7425	0.0002202	0.0000437	0.0002510	0.0000242
	0.9117	120.3084	0.0005905	0.0000813	0.0006784	0.0000464
	1.0396	92.5268	0.0008678	0.0001035	0.0009988	0.0000597
	1.1732	72.6534	0.0012377	0.0001292	0.0014250	0.0000752
	1.3898	51.7721	0.0020399	0.0001770	0.0023418	0.0001037
	1.7725	31.8293	0.0042144	0.0002818	0.0047639	0.0001639
	1.9836	25.4151	0.0059266	0.0003531	0.0066076	0.0002022
	2.5787	15.0383	0.0134221	0.0006234	0.0141204	0.0003288
	3.3615	8.8498	0.0322612	0.0012091	0.0302929	0.0005356

Table 4: Fisher information with the error intensity and risk intensity of θ when $\alpha = 4.0$, $\beta = 0.75$

Sample sizes	Generated θ	$I(\theta)$	Under WSELF E_I	Under WSELF R_I	Under KL E_I	Under KL R_I
n = 25	0.8476	34.7983	0.0044733	0.0016986	0.0053203	0.0007406
	1.7550	8.1168	0.0364404	0.0059919	0.0411020	0.0026128
	2.6257	3.6262	0.1140000	0.0113575	0.1220186	0.0049742
	3.0623	2.6659	0.1757418	0.0143606	0.1833578	0.0063272
	3.5232	2.0140	0.2608447	0.0177578	0.2653803	0.0079062
	4.2790	1.3654	0.4529358	0.0239666	0.4443494	0.0109801
	5.6333	0.7878	1.0150302	0.0383224	0.9509682	0.0191999
	6.3885	0.6126	1.4917137	0.0492130	1.3869649	0.0263416
	8.5405	0.3427	3.8829793	0.0998682	3.6990883	0.0652249
	11.4083	0.1921	11.1942079	0.2426187	11.5149843	0.1947541
n = 50	0.8476	69.5966	0.0014658	0.0003488	0.0016223	0.0001640
	1.7550	16.2336	0.0124542	0.0013514	0.0133221	0.0006289
	2.6257	7.2524	0.0403072	0.0027819	0.0418120	0.0012903
	3.0623	5.3318	0.0630543	0.0036496	0.0644676	0.0016954
	3.5232	4.0281	0.0948257	0.0046712	0.0956295	0.0021800
	4.2790	2.7308	0.1672480	0.0065929	0.1654864	0.0031219
	5.6333	1.5756	0.3774582	0.0109445	0.3649164	0.0054410
	6.3885	1.2251	0.5524950	0.0140204	0.5310470	0.0072412
	8.5405	0.6855	1.3840074	0.0265717	1.3460508	0.0156580
	11.4083	0.3842	3.7124851	0.0572368	3.8048596	0.0400564
n = 75	0.8476	104.3949	0.0007749	0.0001457	0.0008359	0.0000700
	1.7550	24.3505	0.0066742	0.0005822	0.0070119	0.0002770
	2.6257	10.8786	0.0218359	0.0012302	0.0224182	0.0005827
	3.0623	7.9977	0.0343010	0.0016322	0.0348649	0.0007735
	3.5232	6.0421	0.0517774	0.0021107	0.0521399	0.0010028
	4.2790	4.0962	0.0917507	0.0030184	0.0911767	0.0014473
	5.6333	2.3634	0.2077933	0.0050652	0.2030563	0.0025086
	6.3885	1.8377	0.3036748	0.0064823	0.2955978	0.0032958
	8.5405	1.0282	0.7471919	0.0119859	0.7325354	0.0067145
	11.4083	0.5763	1.9396278	0.0243972	1.9794414	0.0157780
n = 100	0.8476	139.1932	0.0004957	0.0000776	0.0005244	0.0000380
	1.7550	32.4673	0.0043035	0.0003161	0.0044624	0.0001537
	2.6257	14.5047	0.0141678	0.0006792	0.0144438	0.0003293
	3.0623	10.6636	0.0223211	0.0009079	0.0225817	0.0004405
	3.5232	8.0561	0.0337858	0.0011828	0.0339241	0.0005752
	4.2790	5.4615	0.0601002	0.0017090	0.0597243	0.0008373
	5.6333	3.1512	0.1365676	0.0029038	0.1341071	0.0014594
	6.3885	2.4502	0.1996371	0.0037293	0.1955025	0.0019127
	8.5405	1.3710	0.4882509	0.0068769	0.4809000	0.0038048
	11.4083	0.7683	1.2453126	0.0136778	1.2678072	0.0085181

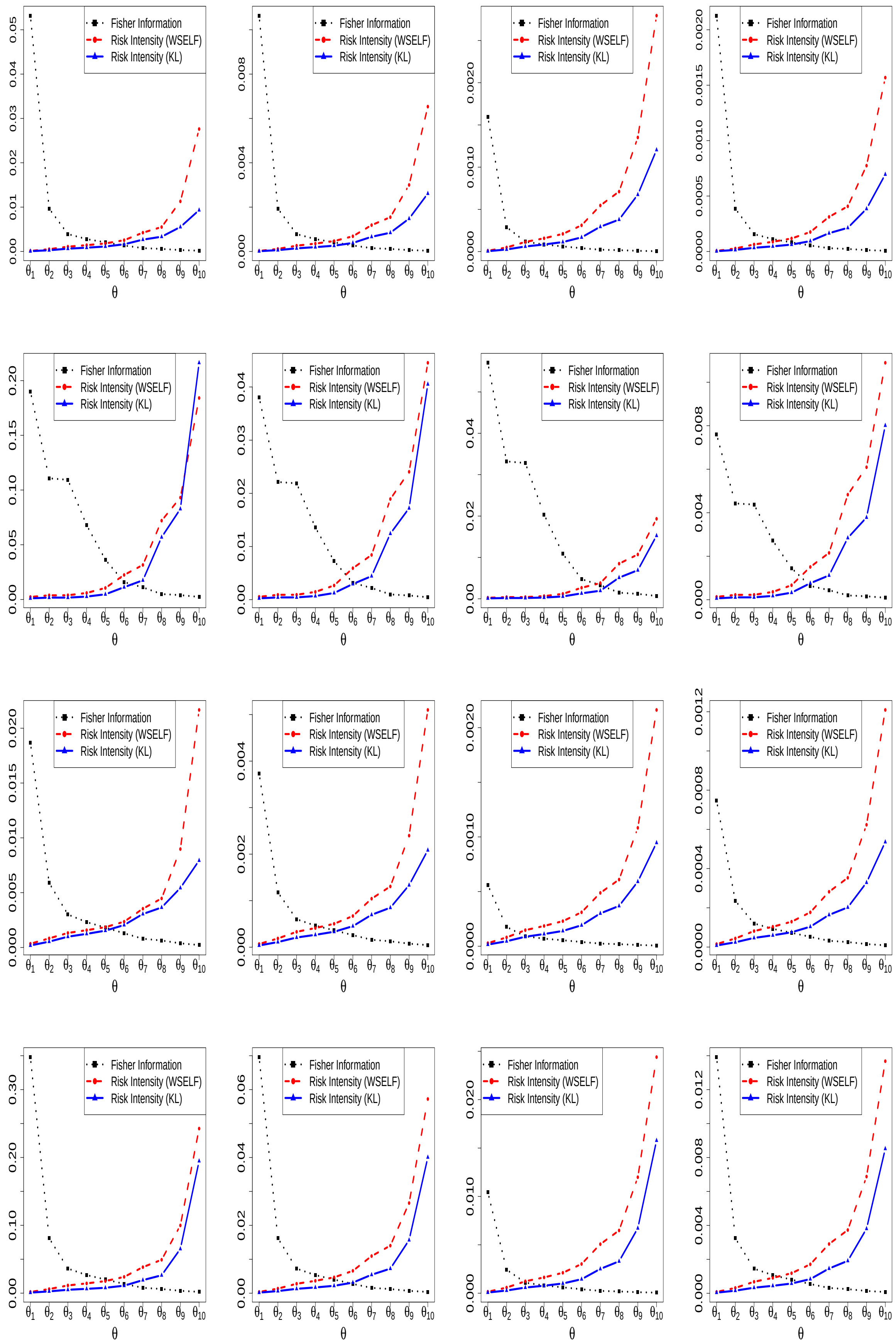

Figure 1: Risk intensity plots for varying sample sizes (left to right) and when prior parameters are $\alpha = 3.5$, $\beta = 2.0$ (first row), $\alpha = 2.0$, $\beta = 0.5$ (second row), $\alpha = 5.0$, $\beta = 3.0$ (third row) and $\alpha = 4.0$, $\beta = 0.75$ (fourth row)

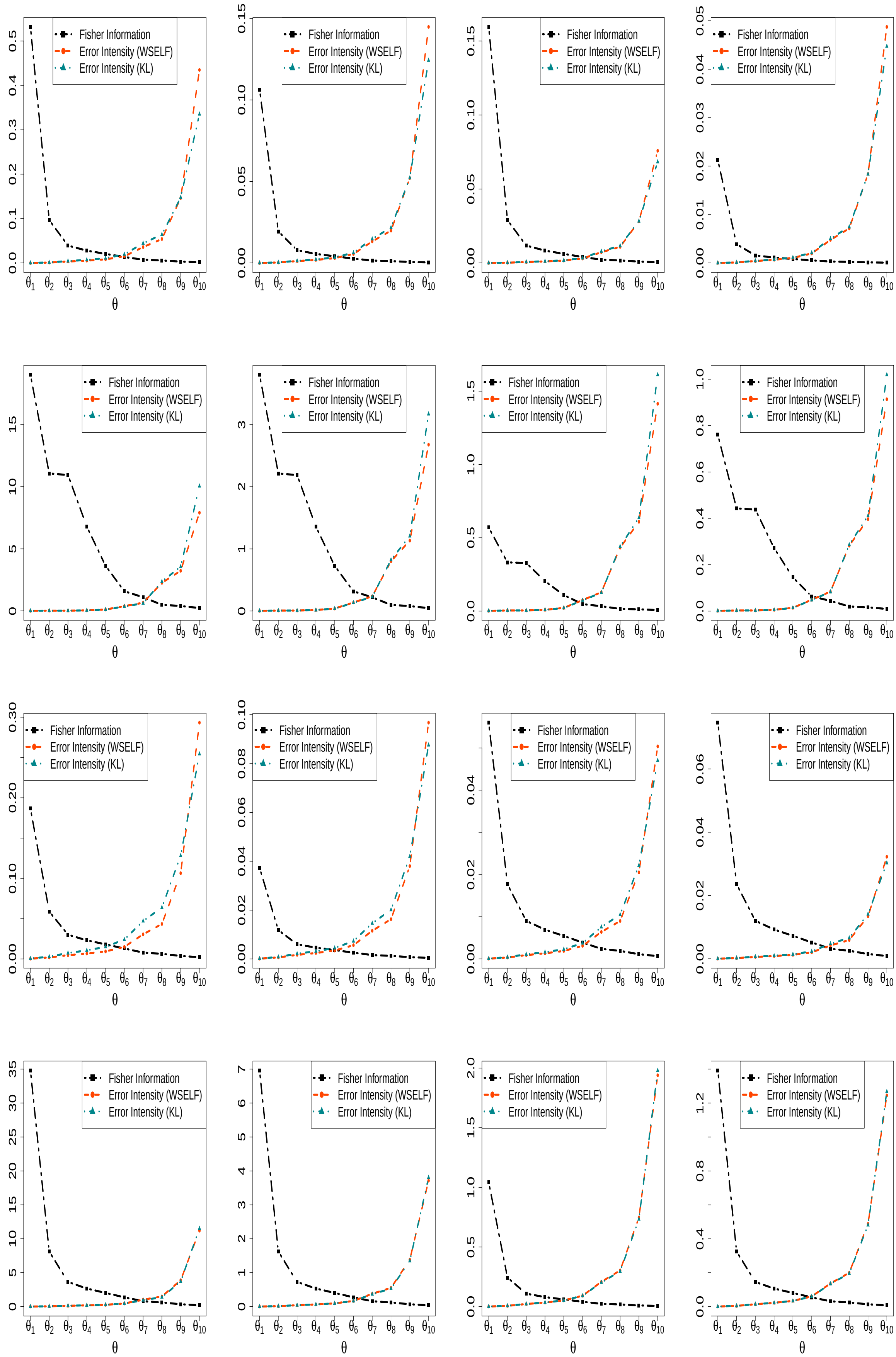

Figure 2: Error intensity plots for varying sample sizes (left to right) and when prior parameters are $\alpha = 3.5$, $\beta = 2.0$ (first row), $\alpha = 2.0$, $\beta = 0.5$ (second row), $\alpha = 5.0$, $\beta = 3.0$ (third row) and $\alpha = 4.0$, $\beta = 0.75$ (fourth row)

loss function is smaller than that under the WSELF loss function. For large sample, the error intensities of the Bayes estimators under the KL and WSELF loss are almost same.

4. Real data applications

In this section, for illustration of the methodology discussed above, two applications are given using real-life data. The first data is about vinyl chloride (mg/l), obtained from clean-up gradient monitoring wells and the second data is regarding the survival times (in years) of a group of gastric cancer patients. Before proceeding with further analysis, we validate the fits of the considered datasets to the exponential distribution. For each data set, we estimate the model parameter using maximum likelihood method and use these in determining the goodness-of-fit test statistic values. The Kolmogorov-Smirnov (K-S), Cramér-Von-Mises (CVM) and Anderson-Darling (AD) test statistics with their corresponding p-values have been examined in order to verify the fitting (Bhunia and Banerjee 2022). Table 5 displays the MLE values with their standard errors and the goodness-of-fit measures along with their associated P-values for both the datasets. Also, some diagnostic plots such as histogram with the fitted pdf and empirical cdf have been presented in Figures 3 and 4 respectively. Based on the Table 5 and Figures 3 and 4, it is clear that the exponential distribution provides a reasonable fit and a potential model for analyzing considered datasets. The R Core Team (2021) (version 3.6.1) is used to do the necessary computation.

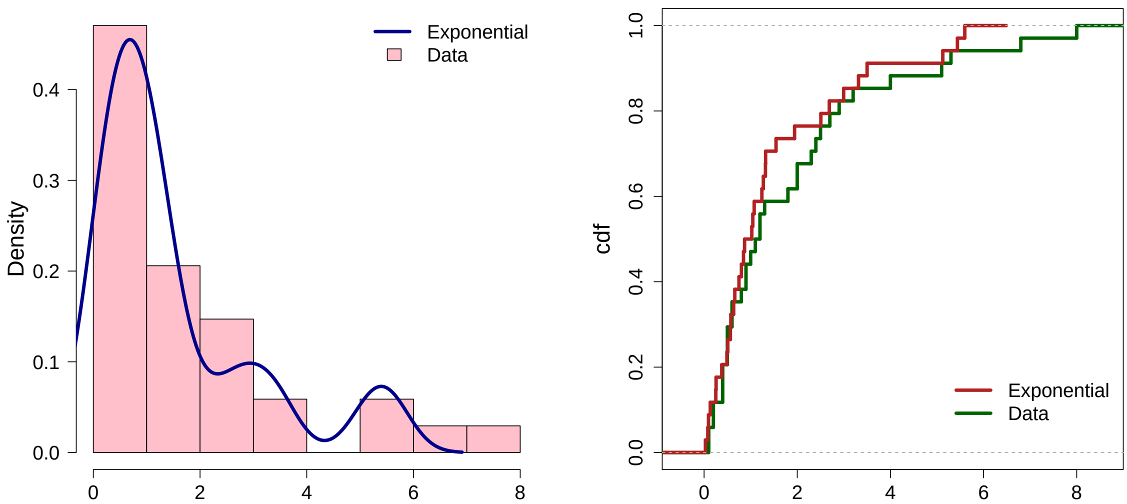

Figure 3: Histogram with fitted pdf, empirical with fitted cdf of vinyl chloride data

Table 5: Parameter estimates with their standard errors (in parentheses) and goodness-of-fit measures for real datasets

Data	MLE (S.E.)	K-S (p-value)	CVM (p-value)	AD (p-value)
Vinyl chloride data	$\hat{\theta} = 0.53208$ (0.09125)	0.08896 (0.95071)	0.04051 (0.93312)	0.27196 (0.95738)
Gastric cancer data	$\hat{\theta} = 0.79042$ (0.11783)	0.08997 (0.82779)	0.05717 (0.83430)	0.41881 (0.82912)

We have used simulated random samples that are replicas of the actual datasets. These replicated samples are drawn from the considered distribution by using the MLE of the

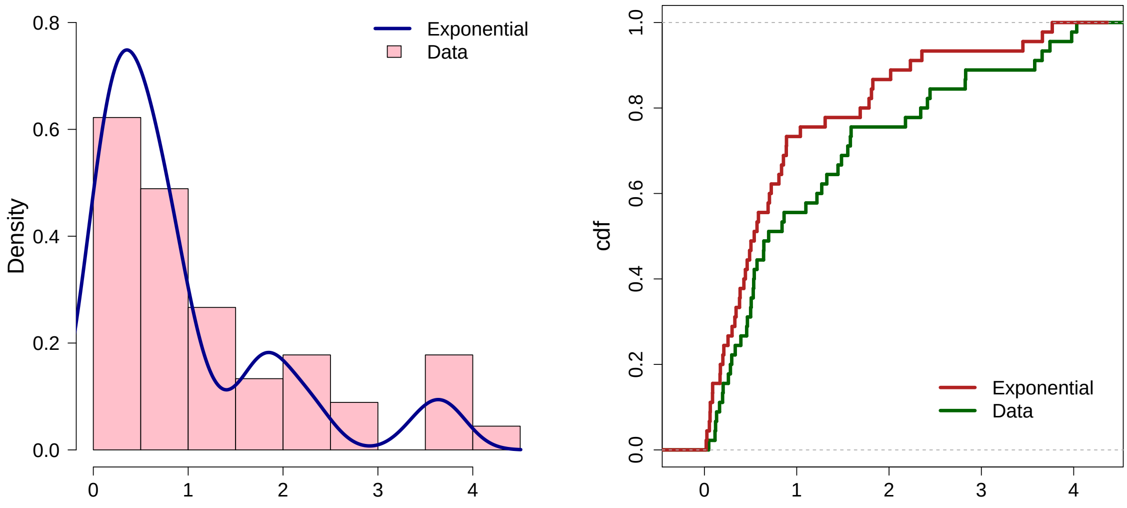

Figure 4: Histogram with fitted pdf, empirical with fitted cdf of gastric cancer data

unknown parameter for each of the data set. Further, the Bayes estimates of the target parameter have been obtained under WSELF and KL loss functions. Now, to investigate the relationship between the classical and Bayesian approach, we compute the information and the risk intensity values for both the datasets.

4.1. Data set I

The first data set is taken from Bhaumik, Kapur, and Gibbons (2009) and represents 34 observation of the vinyl chloride data obtained from clean upgradient monitoring wells in mg/L. It is a volatile organic compound. This constituent is of particular interest in environmental investigations because it is both anthropogenic and carcinogenic. The data are given as: 5.1, 1.2, 1.3, 0.6, 0.5, 2.4, 0.5, 1.1, 8, 0.8, 0.4, 0.6, 0.9, 0.4, 2, 0.5, 5.3, 3.2, 2.7, 2.9, 2.5, 2.3, 1, 0.2, 0.1, 0.1, 1.8, 0.9, 2, 4, 6.8, 1.2, 0.4, 0.2.

Table 6: Fisher information and risk intensity values for different choices of hyper-parameters using vinyl chloride data.

θ	Fisher Information	$\alpha = 2.0,\ \beta = 1.0$		$\alpha = 3.0,\ \beta = 0.5$		$\alpha = 5.0,\ \beta = 2.0$	
		R_I (WSELF)	R_I (KL)	R_I (WSELF)	R_I (KL)	R_I (WSELF)	R_I (KL)
0.5196997	125.88501	0.0002286	0.0001237	0.0002677	0.0001347	0.0003071	0.0001713
0.5461529	113.98571	0.0002516	0.0001362	0.0002948	0.0001481	0.0003349	0.0001883
0.5481330	113.16365	0.0002533	0.0001371	0.0002969	0.0001492	0.0003370	0.0001896
0.5539309	110.80715	0.0002585	0.0001400	0.0003030	0.0001522	0.0003433	0.0001934
0.5849944	99.35176	0.0002871	0.0001555	0.0003369	0.0001689	0.0003773	0.0002144

Table 7: Fisher information and risk intensity values for different choices of hyper-parameters using gastric cancer data

θ	Fisher Information	$\alpha = 2.0,\ \beta = 1.0$		$\alpha = 3.0,\ \beta = 0.5$		$\alpha = 5.0,\ \beta = 2.0$	
		R_I (WSELF)	R_I (KL)	R_I (WSELF)	R_I (KL)	R_I (WSELF)	R_I (KL)
0.7402762	82.11547	0.0002280	0.0001158	0.0002633	0.0001264	0.0002797	0.0001616
0.8069154	69.11249	0.0002690	0.0001365	0.0003110	0.0001488	0.0003240	0.0001897
0.8101432	68.56285	0.0002711	0.0001376	0.0003134	0.0001499	0.0003262	0.0001911
0.8206677	66.81559	0.0002778	0.0001410	0.0003214	0.0001536	0.0003334	0.0001957
0.9142761	53.83414	0.0003416	0.0001732	0.0003957	0.0001880	0.0003334	0.0001957

4.2. Data set II

The second data set represents the survival times (in years) of a group of gastric cancer patients given chemotherapy and radiation treatment. It is a subset of data reported by Stablein, Carter Jr, and Novak (1981). The data consists of 45 observations and given as: 0.047, 0.115, 0.121, 0.132, 0.164, 0.197, 0.203, 0.260, 0.282, 0.296, 0.334, 0.395, 0.458, 0.466, 0.501, 0.507, 0.529, 0.534, 0.540, 0.570, 0.641, 0.644, 0.696, 0.841, 0.863, 1.099, 1.219, 1.271, 1.326, 1.447, 1.485, 1.553, 1.581, 1.589, 2.178, 2.343, 2.416, 2.444, 2.825, 2.830, 3.578, 3.658, 3.743, 3.978, 4.033.

It has been observed from Tables 6 and 7 that, for every choices of the hyper-parameters, Fisher information and the risk intensity have the inverse proportional relationship with each other. More specifically, as the overall data information increases, the risk intensity is gradually decreasing. Therefore, the efficiency of the Bayes estimator is improved with increased information. It implies that the risk intensity depends on the overall data information, establishing a connection between the classical and Bayesian approaches.

5. Conclusion

In statistics, both the classical and Bayesian approaches can provide us with the information about the model parameter. In classical way, Fisher information quantifies the information regarding the parameter from the observed sample. Additionally, before the data is considered, a certain amount of information about the parameter can be guessed from the prior information. In this article, an attempt has been made to establish a connection between the classical and Bayesian approaches through Fisher information. Here, we derive the Fisher information and the Bayes estimate of the scale parameter by considering the exponential distribution. We have taken the weighted squared error loss function which is traditionally used to estimate the scale parameter. Also, the Kullback-Leibler distance loss function has been derived directly by utilizing density of exponential distribution to obtain the Bayes estimators. Two measures namely, risk intensity and error intensity have been introduced to compare the performance of the estimators as well as to examine the relationship between the Fisher information and the Bayes estimators.

From the numerical results, it has been observed that the Bayes estimator becomes more precise as the amount of information increases. The risk and error intensity of the estimators continuously decrease as Fisher information increases. In comparison to the WSELF loss function, the risk intensity under the KL distance loss function is minimal. However, the error intensity of the Bayes estimators under the KL and WSELF losses are almost equal for large samples. The numerical outcomes have been supported by the real-life application by demonstrating the applicability of the theoretical notion in real-world circumstances. The explanation above leads us to the conclusion that the relationship between the classical and Bayesian techniques through Fisher information may contribute to the literature as a novel insight.

References

Abu Awwad RR, Abufoudeh GK, Bdair OM (2019). "Statistical Inference of Exponential Record Data under Kullback-Leibler Divergence Measure." *Statistics in Transition new series*, **20**(2), 1–14. doi:10.21307/stattrans-2019-011.

Abufoudeh GK, Awwad RRA, Bdair OM (2019). "Bayesian Estimation under Kullback-Leibler Divergence Measure Based on Exponential Data." *Investigacion Operacional*, **40**(1), 61–72.

Banerjee P, Bhunia S (2022). "Minimax Estimation of the Scale Parameter of Inverse Rayleigh Distribution under Symmetric and Asymmetric Loss Functions." *Reliability: Theory & Applications*, **17**(4 (71)), 218–231. doi:10.24412/1932-2321-2022-471-218-231.

Banerjee P, Seal B (2022). "Partial Bayes Estimation of Two Parameter Gamma Distribution under Non-Informative Prior." *Statistics, Optimization & Information Computing*, **10**(4), 1110–1125. doi:10.19139/soic-2310-5070-1110.

Bhaumik DK, Kapur K, Gibbons RD (2009). "Testing Parameters of a Gamma Distribution for Small Samples." *Technometrics*, **51**(3), 326–334. doi:10.1198/tech.2009.07038.

Bhunia S, Banerjee P (2022). "Some Properties and Different Estimation Methods for Inverse A (α) Distribution with an Application to Tongue Cancer Data." *Reliability: Theory & Applications*, **17**(1 (67)), 251–266. doi:10.24412/1932-2321-2022-167-251-266.

Dhungana GP, Kumar V (2022). "Exponentiated Odd Lomax Exponential Distribution with Application to COVID-19 Death Cases of Nepal." *PloS One*, **17**(6), e0269450. doi:10.1371/journal.pone.0269450.

Efron B, Johnstone IM (1990). "Fisher's Information in Terms of the Hazard Rate." *The Annals of Statistics*, pp. 38–62. doi:10.1214/aos/1176347492.

Elfessi A, Reineke DM (2001). "A Bayesian Look at Classical Estimation: The Exponential Distribution." *Journal of Statistics Education*, **9**(1). doi:10.1080/10691898.2001.11910648.

Ferguson TS (2014). *Mathematical Statistics: A Decision Theoretic Approach*, volume 1. Academic press.

Goel PK, Degroot MH (1981). "Information about Hyperparameters in Hierarchical Models." *Journal of the American Statistical Association*, **76**(373), 140–147. doi:10.2307/2287059.

Jaheen Z (2004). "Empirical Bayes Analysis of Record Statistics Based on LINEX and Quadratic Loss Functions." *Computers & Mathematics with Applications*, **47**(6-7), 947–954. doi:10.1016/S0898-1221(04)90078-8.

Kullback S, Leibler RA (1951). "On Information and Sufficiency." *Annals of Mathematical Statistics*, **22**(1), 79–86. doi:10.1214/aoms/1177729694.

Lehmann EL, Casella G (2006). *Theory of Point Estimation*. Springer Science & Business Media.

Nasiri P, Hosseini S, Yarmohammadi M (2012). "A New Approach to Statistical Inference for Exponential Distribution Based on Record Values." *Canadian Journal of Pure and Applied Sciences*, **6**, 2033–2038.

Ni S, Sun D (2021). "Intrinsic Bayesian Estimation of Linear Time Series Models." *Statistical Theory and Related Fields*, **5**(4), 275–287. doi:10.1080/24754269.2020.1744073.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rong Y, Tang M, Zhou J (2017). "Intrinsic Losses Based on Information Geometry and Their Applications." *Entropy*, **19**(8), 405. doi:10.3390/e19080405.

Rossman AJ, Short TH, Parks MT (1998). "Bayes Estimators for the Continuous Uniform Distribution." *Journal of Statistics Education*, **6**(3). doi:10.1080/10691898.1998.11910622.

Stablein DM, Carter Jr WH, Novak JW (1981). "Analysis of Survival Data with Nonproportional Hazard Functions." *Controlled Clinical Trials*, **2**(2), 149–159. doi:10.1016/0197-2456(81)90005-2.

Affiliation:

Shreya Bhunia
Department of Mathematics and Statistics
Aliah University
Action Area II, Plot No.- IIA/27,
Newtown, Kolkata, West Bengal, India-700160
E-mail: shreyabhunia.stat@gmail.com

Dr. Babulal Seal
Professor in Statistics
Department of Mathematics and Statistics
Aliah University
Action Area II, Plot No.- IIA/27,
Newtown, Kolkata, West Bengal, India-700160
E-mail: babulal_seal@yahoo.com

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 53 *Submitted:* 2023-03-18
2024 *Accepted:* 2023-09-26