Journal of Statistical Modelling: Theory and Applications
Vol. 7, No. 1, 2026, pp. 237-247
Yazd University Press 2026

JSMTA

Research Paper

Statistical inference for Kullback–Leibler and total variation divergences for two normal populations with a common unknown variance

Shohreh Enami* and Ehsan Golzade Gervi
Department of Statistics, Payame Noor University, Tehran, Iran

Received: June 20, 2026/ Revised: October 17, 2026/ Accepted: November 08, 2026

Abstract: This article focuses on the estimation of two divergence measures, the Kullback–Leibler divergence and the total variation divergence, between two normal distributions with a common unknown variance. We provide the maximum likelihood estimator and a Bayesian estimator for these divergences. In the Bayesian framework, improper prior distributions are considered for the population parameters, and both Bayesian point estimators and interval estimators, including credible intervals and highest posterior density intervals, are constructed. In the classical approach, we also present the asymptotic normal confidence intervals. The performance of the point estimators is evaluated using the root mean squared error criterion, while the interval estimators are assessed based on coverage probability and average length through a simulation study. Finally, we use a real dataset to illustrate the proposed methods.

Keywords: Bayesian credible interval; Bayesian estimation; Highest posterior density interval; Kullback-Leibler divergence; Total variation divergence.
Mathematics Subject Classification (2010): 62F10, 62F15.

1 Introduction

Estimating the divergence criteria within a family of statistical distributions is an important issue in statistical data analysis and model selection. According to Akaike (1998), divergence measures allow researchers to quantify the similarity or dissimilarity between probability distributions, which is fundamental in many areas such as metric

*Corresponding author: enami@pnu.ac.ir

learning, machine learning, network security, biology, medicine, finance, and statistical data analysis.

One of the most widely used divergence measures is the Kullback–Leibler (KL) divergence, first introduced by Kullback and Leibler (1951). This criterion measures the degree of dissimilarity between two probability distributions and has become a cornerstone in information theory and statistical inference. Another fundamental measure is the total variation (TV) divergence, which quantifies the maximum difference between the probabilities assigned to any event by two distributions. Both criteria are non-negative and equal zero if and only if the two distributions are identical. For further applications and theoretical developments, see Noh et al. (2014), Moulin and Veeravalli (2018), Pardo (2019), and Ji et al. (2020).

Based on Pérez-Cruz (2008), a key challenge in applying these divergence measures is their estimation from practical data. Recently, Afshar Moghaddam et al. (2025) proposed the point estimators for certain divergence measures between two normal populations with unit variances. Specifically, they introduced the maximum likelihood estimator (MLE) as well as a plug-in estimator obtained by substituting the Bayes estimators of the population means into the expressions of various divergence measures, particularly the KL and TV divergences.

Let $F_1(\cdot)$ and $F_2(\cdot)$ be two continuous distribution functions with corresponding density functions $f_1(\cdot)$ and $f_2(\cdot)$, respectively. The KL divergence, denoted $D_{\mathrm{KL}}(F_1 \parallel F_2)$, is a measure of the dissimilarity between the two distribution functions $F_1(\cdot)$ and $F_2(\cdot)$. It is defined as follows

$$D_{\mathrm{KL}}(F_1 \parallel F_2) = \int_{\mathbb{R}} f_1(x) \log \frac{f_1(x)}{f_2(x)} \mathrm{d}x.$$

Another statistical criterion for measuring the dissimilarity between $F_1(\cdot)$ and $F_2(\cdot)$ is the TV divergence, denoted $D_{\mathrm{TV}}(F_1 \parallel F_2)$, which is defined as follows:

$$D_{\mathrm{TV}}(F_1 \parallel F_2) = \frac{1}{2} \int_{\mathbb{R}} |f_1(x) - f_2(x)| \mathrm{d}x.$$

Both the KL divergence and the TV divergence are non-negative, and they equal zero if and only if $F_1(\cdot)$ and $F_2(\cdot)$ are identical. Moreover, larger values of these divergence measures indicate greater dissimilarity between $F_1(\cdot)$ and $F_2(\cdot)$.

Now, suppose that F_1 and F_2 correspond to two normally distributed populations with a common variance σ^2 and respective means θ_1 and θ_2. Then, according to Equation (3) of Afshar Moghaddam et al. (2025), we have

$$D_{\mathrm{KL}}(F_1 \parallel F_2) = \frac{1}{2\sigma^2}(\theta_2 - \theta_1)^2 = \frac{\gamma^2}{2\sigma^2}, \tag{1}$$

where $\gamma = \theta_2 - \theta_1$. To obtain $D_{\mathrm{TV}}(F_1 \parallel F_2)$, we can write

$$D_{\mathrm{TV}}(F_1 \parallel F_2) = \frac{1}{2} \int_{\mathbb{R}} (2\pi\sigma^2)^{-1/2} \left| e^{-\frac{1}{2\sigma^2}(x-\theta_1)^2} - e^{-\frac{1}{2\sigma^2}(x-\theta_2)^2} \right| \mathrm{d}x. \tag{2}$$

Note that

$$e^{-\frac{1}{2\sigma^2}(x-\theta_1)^2} - e^{-\frac{1}{2\sigma^2}(x-\theta_2)^2} > 0, \text{ if and only if } \begin{cases} x > \frac{\theta_1+\theta_2}{2}, & \text{if } \theta_1 > \theta_2, \\ x < \frac{\theta_1+\theta_2}{2}, & \text{if } \theta_1 < \theta_2, \end{cases}$$

$$e^{-\frac{1}{2\sigma^2}(x-\theta_1)^2} - e^{-\frac{1}{2\sigma^2}(x-\theta_2)^2} < 0, \text{ if and only if } \begin{cases} x < \frac{\theta_1+\theta_2}{2}, & \text{if } \theta_1 > \theta_2, \\ x > \frac{\theta_1+\theta_2}{2}, & \text{if } \theta_1 < \theta_2. \end{cases}$$

Therefore by (2), if $\theta_1 > \theta_2$, we have

$$\begin{aligned} D_{\mathrm{TV}}(F_1 \parallel F_2) &= \frac{1}{2}\int_{\frac{\theta_1+\theta_2}{2}}^{\infty} (2\pi\sigma^2)^{-1/2}\left[e^{-\frac{1}{2\sigma^2}(x-\theta_1)^2} - e^{-\frac{1}{2\sigma^2}(x-\theta_2)^2}\right]\mathrm{d}x \\ &\quad + \frac{1}{2}\int_{-\infty}^{\frac{\theta_1+\theta_2}{2}} (2\pi\sigma^2)^{-1/2}\left[e^{-\frac{1}{2\sigma^2}(x-\theta_2)^2} - e^{-\frac{1}{2\sigma^2}(x-\theta_1)^2}\right]\mathrm{d}x \\ &= \Phi\left(\frac{\theta_1-\theta_2}{2\sigma}\right) - \Phi\left(\frac{\theta_2-\theta_1}{2\sigma}\right), \end{aligned}$$

where $\Phi(\cdot)$ denotes the distribution function of the standard normal distribution. Similarly, if $\theta_1 < \theta_2$, we have

$$\begin{aligned} D_{\mathrm{TV}}(F_1 \parallel F_2) &= \frac{1}{2}\int_{-\infty}^{\frac{\theta_1+\theta_2}{2}} (2\pi\sigma^2)^{-1/2}\left[e^{-\frac{1}{2\sigma^2}(x-\theta_1)^2} - e^{-\frac{1}{2\sigma^2}(x-\theta_2)^2}\right]\mathrm{d}x \\ &\quad + \frac{1}{2}\int_{\frac{\theta_1+\theta_2}{2}}^{\infty} (2\pi\sigma^2)^{-1/2}\left[e^{-\frac{1}{2\sigma^2}(x-\theta_2)^2} - e^{-\frac{1}{2\sigma^2}(x-\theta_1)^2}\right]\mathrm{d}x \\ &= \Phi\left(\frac{\theta_2-\theta_1}{2\sigma}\right) - \Phi\left(\frac{\theta_1-\theta_2}{2\sigma}\right). \end{aligned}$$

Hence, we obtain $D_{\mathrm{TV}}(F_1 \parallel F_2)$ for each θ_1, θ_2, and σ of the following closed form

$$D_{\mathrm{TV}}(F_1 \parallel F_2) = 2\Phi\left(\frac{|\gamma|}{2\sigma}\right) - 1. \tag{3}$$

In this paper, we focus on the KL divergence and TV divergence for two normal distributions with a common but unknown variance. In Section 2, we present the classical inference, including the MLEs and asymptotic normal (AN) confidence intervals for these divergence measures. Section 3 is devoted to Bayesian inference, where we adopt improper prior distributions for the population parameters. Within this framework, we propose a Bayes estimator and construct Bayesian credible intervals as well as highest posterior density (HPD) intervals for the divergence measures. In Section 4, the performance of the proposed methods is evaluated through simulation studies. The point estimators (MLE and Bayes) are compared using the root mean squared error (RMSE) criterion, while the interval estimators are assessed based on coverage probability (CP) and average length (AL). Section 5 is devoted to illustrating our methods using a real dataset. Finally, in Section 6, we make a discussion about the paper.

2 Classical inference

Suppose that X_1 and X_2 are two independent random variables with $X_1 \sim N(\theta_1, \sigma^2)$ and $X_2 \sim N(\theta_2, \sigma^2)$. Let $X_{11}, \ldots, X_{1n_1}$ be a random sample drawn from X_1, and

let $X_{21}, \ldots, X_{2n_2}$ be another random sample drawn from X_2. Then, the likelihood function can be expressed as

$$\begin{aligned} L(\theta_1, \theta_2, \sigma^2) &= (2\pi\sigma^2)^{-\frac{n_1+n_2}{2}} e^{-\frac{1}{2\sigma^2}\left[\sum_{j=1}^{n_1}(X_{1j}-\theta_1)^2+\sum_{j=1}^{n_2}(X_{2j}-\theta_2)^2\right]}, \\ &\quad -\infty < \theta_1, \theta_2 < \infty,\ \sigma > 0. \end{aligned}$$

Hence, the MLEs of parameters θ_1, θ_2 and σ^2 are obtained as

$$\hat{\theta}_1 = \bar{X}_1, \qquad \hat{\theta}_2 = \bar{X}_2, \qquad \hat{\sigma}^2 = \frac{1}{n_1+n_2}\left[(n_1-1)S_1^2 + (n_2-1)S_2^2\right],$$

where $S_i^2 = \sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2/(n_i-1)$ $i = 1, 2$ are the sample variances. Therefore, the MLEs of $D_{\mathrm{KL}}(F_1 \parallel F_2)$ and $D_{\mathrm{TV}}(F_1 \parallel F_2)$ are given by

$$\widehat{D}_{\mathrm{KL}}(F_1 \parallel F_2) = \frac{\hat{\gamma}^2}{2\hat{\sigma}^2}, \quad \widehat{D}_{\mathrm{TV}}(F_1 \parallel F_2) = 2\Phi\left(\frac{|\hat{\gamma}|}{2\hat{\sigma}}\right) - 1, \tag{4}$$

where $\hat{\gamma} = \hat{\theta}_2 - \hat{\theta}_1$ is the MLE of γ.

To construct the AN confidence interval, we use the fact that, by the central limit theorem, for large values of n_1 and n_2, we have

$$\begin{pmatrix} \bar{X}_i \\ S_i^2 \end{pmatrix} - \begin{pmatrix} \theta_i \\ \sigma^2 \end{pmatrix} \approx N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{n_i}\begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix} \right), \qquad i = 1, 2.$$

Thus, by simple algebraic manipulation, we can write

$$\begin{pmatrix} \hat{\gamma} \\ \hat{\sigma}^2 \end{pmatrix} - \begin{pmatrix} \gamma \\ \sigma^2 \end{pmatrix} \approx N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix} \right),$$

where $\sigma_{11} = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\sigma^2$ and $\sigma_{22} = 2\frac{(n_1-1)^2+(n_2-1)^2}{(n_1+n_2)^2}\sigma^4$. Using the multivariate delta method, after simplification, we get

$$\begin{aligned} \widehat{D}_{\mathrm{KL}}(F_1 \parallel F_2) - D_{\mathrm{KL}}(F_1 \parallel F_2) &\approx N(0, \eta^2), \\ \widehat{D}_{\mathrm{TV}}(F_1 \parallel F_2) - D_{\mathrm{TV}}(F_1 \parallel F_2) &\approx N(0, \tau^2), \end{aligned}$$

where

$$\begin{aligned} \eta^2 &= 2\left(\frac{1}{n_1} + \frac{1}{n_2}\right) D_{\mathrm{KL}}(F_1 \parallel F_2) + 2\frac{(n_1-1)^2 + (n_2-1)^2}{(n_1+n_2)^2} D_{\mathrm{KL}}^2(F_1 \parallel F_2), \\ \tau^2 &= \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\phi^2\left(\frac{|\gamma|}{2\sigma}\right) + \frac{(n_1-1)^2 + (n_2-1)^2}{2(n_1+n_2)^2\sigma^2}\gamma^2\phi^2\left(\frac{|\gamma|}{2\sigma}\right), \quad \gamma \neq 0, \end{aligned}$$

and $\phi(\cdot)$ denotes the density function of the standard normal distribution. Therefore, by the fact that $D_{\mathrm{KL}}(F_1 \parallel F_2) \geq 0$, a $100(1-\alpha)\%$ AN confidence interval for $D_{\mathrm{KL}}(F_1 \parallel F_2)$ is

$$\left(\max\left(\widehat{D}_{\mathrm{KL}}(F_1 \parallel F_2) - \hat{\eta} z_{1-\alpha/2}, 0\right), \widehat{D}_{\mathrm{KL}}(F_1 \parallel F_2) + \hat{\eta} z_{1-\alpha/2} \right),$$

and by the fact that $0 \leq D_{\mathrm{TV}}(F_1 \parallel F_2) \leq 1$, that for $D_{\mathrm{TV}}(F_1 \parallel F_2)$ is given by

$$\left(\max\left(\widehat{D}_{\mathrm{TV}}(F_1 \parallel F_2) - \hat{\tau} z_{1-\alpha/2}, 0\right), \min\left(1, \widehat{D}_{\mathrm{TV}}(F_1 \parallel F_2) + \hat{\tau} z_{1-\alpha/2}\right) \right),$$

where $z_{1-\alpha/2}$ denotes the $(1-\alpha/2)$th quantile of the standard normal distribution, and $\hat{\eta}$ and $\hat{\tau}$ are the MLEs of η and τ, respectively.

3 Bayesian inference

Suppose that θ_1, θ_2, and σ are independent with the following improper priors:

$$\pi(\theta_1) = \pi(\theta_2) = 1, \qquad -\infty < \theta_1, \theta_2 < \infty, \qquad \pi(\sigma) \propto \frac{1}{\sigma}, \quad \sigma > 0. \tag{5}$$

From Section 2.5.1 in Box and Tiao (1973), we conclude that

$$\frac{(n_1+n_2)\hat{\sigma}^2}{\sigma^2}\Big|\hat{\sigma}^2 \sim \chi^2_{n_1+n_2-2}, \quad \gamma\big|(\sigma^2, \hat{\gamma}) \sim N\left(\hat{\gamma}, \sigma^2(\frac{1}{n_1}+\frac{1}{n_2})\right). \tag{6}$$

Let $\widetilde{D}_{\mathrm{KL}}(F_1 \parallel F_2)$ denote the Bayes estimator of $D_{\mathrm{KL}}(F_1 \parallel F_2)$ under the squared error loss function. Then, using (1) and (6), we obtain a closed-form expression for $\widetilde{D}_{\mathrm{KL}}(F_1 \parallel F_2)$ as follows

$$\begin{aligned}
\widetilde{D}_{\mathrm{KL}}(F_1 \parallel F_2) &= E\left[D_{\mathrm{KL}}(F_1 \parallel F_2)\big|\hat{\sigma}^2, \hat{\gamma}\right] = \frac{1}{2}E\left[\frac{\gamma^2}{\sigma^2}\Big|\hat{\sigma}^2, \hat{\gamma}\right] \\
&= \frac{1}{2}E\left[E\left[\frac{\gamma^2}{\sigma^2}\Big|\sigma^2, \hat{\sigma}^2, \hat{\gamma}\right]\right] = \frac{1}{2}E\left[\frac{\hat{\gamma}^2}{\sigma^2} + \frac{1}{n_1} + \frac{1}{n_2}\Big|\hat{\sigma}^2, \hat{\gamma}\right] \\
&= \frac{1}{2}\left[\frac{(n_1+n_2-2)\hat{\gamma}^2}{(n_1+n_2)\hat{\sigma}^2} + \frac{1}{n_1} + \frac{1}{n_2}\right] \\
&= \frac{n_1+n_2-2}{n_1+n_2}\widehat{D}_{\mathrm{KL}}(F_1 \parallel F_2) + \frac{1}{2n_1} + \frac{1}{2n_2}.
\end{aligned} \tag{7}$$

However, a closed-form expression for the Bayes estimator of $D_{\mathrm{TV}}(F_1 \parallel F_2)$, denoted by $\widetilde{D}_{\mathrm{TV}}(F_1 \parallel F_2)$, cannot be obtained. To address this, as well as to construct Bayesian credible intervals and HPD intervals for $D_{\mathrm{KL}}(F_1 \parallel F_2)$ and $D_{\mathrm{TV}}(F_1 \parallel F_2)$, we employ the method proposed by Chen and Shao (1999). Let $\hat{\gamma}$ and $\hat{\sigma}^2$ be the observed values of the MLEs of the parameters γ and σ, respectively. We then propose the following algorithm to obtain the Bayes estimators for $D_{\mathrm{TV}}(F_1 \parallel F_2)$, as well as the Bayesian credible intervals for both $D_{\mathrm{KL}}(F_1 \parallel F_2)$ and $D_{\mathrm{TV}}(F_1 \parallel F_2)$.

Algorithm 3.1. *Given n_1, n_2, $\hat{\gamma}$ and $\hat{\sigma}^2$,*
(i) *Generate W from the chi-squared distribution with n_1+n_2-2 degrees of freedom and set $\sigma^2 = (n_1+n_2)\hat{\sigma}^2/W$ based on the first equation in* (6)*;*
(ii) *Generate T from the normal distribution with the mean $\hat{\gamma}$ and the variance $\sigma^2(1/n_1 + 1/n_2)$, and set $\gamma = T$ based on the second equation in* (6)*;*
(iii) *Compute $D^1_{\mathrm{KL}}(F_1 \parallel F_2)$ and $D^1_{\mathrm{TV}}(F_1 \parallel F_2)$ by substituting the simulated values of*

σ^2 and γ in Equations (1) *and* (3)*, respectively;*
(iv) *Repeat the steps* (i)-(iii) *for M times to get*

$$D^1_{\mathrm{KL}}(F_1 \parallel F_2), \ldots, D^M_{\mathrm{KL}}(F_1 \parallel F_2),$$
$$D^1_{\mathrm{TV}}(F_1 \parallel F_2), \ldots, D^M_{\mathrm{TV}}(F_1 \parallel F_2);$$

(v) *Average of the values $D^1_{\mathrm{TV}}(F_1 \parallel F_2), \ldots, D^M_{\mathrm{TV}}(F_1 \parallel F_2)$ is a Bayes estimator for parameter $D_{\mathrm{TV}}(F_1 \parallel F_2)$ under the squared error loss function;*
(vi) *Arrange the obtained values increasingly, as*

$$D^{(1)}_{\mathrm{KL}}(F_1 \parallel F_2) < \cdots < D^{(M)}_{\mathrm{KL}}(F_1 \parallel F_2),$$
$$D^{(1)}_{\mathrm{TV}}(F_1 \parallel F_2) < \cdots < D^{(M)}_{\mathrm{TV}}(F_1 \parallel F_2);$$

(vii) *A $100(1-\alpha)\%$ Bayesian credible interval for $D_{\mathrm{KL}}(F_1 \parallel F_2)$ is*

$$\left(D^{([M\frac{\alpha}{2}])}_{\mathrm{KL}}(F_1 \parallel F_2), D^{([M(1-\frac{\alpha}{2})])}_{\mathrm{KL}}(F_1 \parallel F_2)\right),$$

and that for $D_{\mathrm{TV}}(F_1 \parallel F_2)$ is given by

$$\left(D^{([M\frac{\alpha}{2}])}_{\mathrm{TV}}(F_1 \parallel F_2), D^{([M(1-\frac{\alpha}{2})])}_{\mathrm{TV}}(F_1 \parallel F_2)\right),$$

where $[k]$ denotes the floor of k;
(viii) *The $100(1-\alpha)\%$ HPD interval for $D_{\mathrm{KL}}(F_1 \parallel F_2)$ is the shortest interval of the form*

$$\left(D^{(j)}_{\mathrm{KL}}(F_1 \parallel F_2), D^{(j+[M(1-\alpha)])}_{\mathrm{KL}}(F_1 \parallel F_2)\right),$$

and that for $D_{\mathrm{TV}}(F_1 \parallel F_2)$ is the shortest interval of the form

$$\left(D^{(j)}_{\mathrm{TV}}(F_1 \parallel F_2), D^{(j+[M(1-\alpha)])}_{\mathrm{TV}}(F_1 \parallel F_2)\right),$$

such that $j = 0, 1, \ldots, M - [M(1-\alpha)]$, and $D^{(0)}_{\mathrm{KL}}(F_1 \parallel F_2) = D^{(0)}_{\mathrm{TV}}(F_1 \parallel F_2) = 0$.

In recent years, many statistical inferences have been made using the concept of the generalized pivotal variable. This concept, along with generalized pivotal confidence intervals, was introduced by Weerahandi (1993). To make inferences about the parameters $D_{\mathrm{KL}}(F_1 \parallel F_2)$ and $D_{\mathrm{TV}}(F_1 \parallel F_2)$, it is necessary to construct generalized pivotal variables for γ and σ^2. Suppose that $\boldsymbol{X} = (X_{11}, \ldots, X_{1n_1}, X_{21}, \ldots, X_{2n_2})$ and $\boldsymbol{x}$ is its realization. It is clear that

$$Z = \frac{\hat{\gamma} - \gamma}{\sigma} \sim N\left(0, \frac{1}{n_1} + \frac{1}{n_2}\right), \quad W = \frac{(n_1+n_2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2},$$

where Z and W are independent. Suppose $\hat{\gamma}_{\mathrm{obs}}$ and $\hat{\sigma}^2_{\mathrm{obs}}$ are the observed MLEs of the parameters γ and σ^2 based on $\boldsymbol{x}$, respectively. Define

$$G_1(\boldsymbol{X}; \boldsymbol{x}, \gamma, \sigma^2) = \frac{(n_1+n_2)\hat{\sigma}^2_{\mathrm{obs}}}{W}, \qquad G_2(\boldsymbol{X}; \boldsymbol{x}, \gamma, \sigma^2) = \hat{\gamma}_{\mathrm{obs}} - \sqrt{G_1(\boldsymbol{X}; \boldsymbol{x}, \gamma, \sigma^2)}Z.$$

Is is clear that $G_1(\boldsymbol{X};\boldsymbol{x},\gamma,\sigma^2)=\sigma^2$ and $G_2(\boldsymbol{X};\boldsymbol{x},\gamma,\sigma^2)=\gamma$. In addition, for the fixed observed data $\boldsymbol{x}$, the probabilistic behaviours of $G_1(\boldsymbol{X};\boldsymbol{x},\gamma,\sigma^2)$ and $G_2(\boldsymbol{X};\boldsymbol{x},\gamma,\sigma^2)$ as functions of the random $\boldsymbol{X}$, do not depend on any unknown parameter. Therefore, $G_1(\boldsymbol{X};\boldsymbol{x},\gamma,\sigma^2)$ and $G_2(\boldsymbol{X};\boldsymbol{x},\gamma,\sigma^2)$ are the generalized pivotal variables for σ^2 and γ, respectively.

It can be readily seen that these generalized pivotal variables are exactly the same as σ^2 and γ introduced in Steps (i)–(ii) of Algorithm 3.1. Consequently, the generalized pivotal inference for the parameters of interest coincides with the Bayesian inference based on the priors given in (5). For further details on generalized pivotal inference, readers may refer to Weerahandi (1995).

Table 1: The RMSE of the MLE and the Bayes estimator.

				$D_{\mathrm{KL}}(F_1 \parallel F_2)$		$D_{\mathrm{TV}}(F_1 \parallel F_2)$	
$\lvert\gamma\rvert$	σ	n_1	n_2	MLE	Bayes	MLE	Bayes
0.5	1	10	10	0.3949	0.3981	0.1603	0.1251
		10	30	0.2575	0.2710	0.1311	0.1054
		50	30	0.1345	0.1388	0.0896	0.0795
		50	50	0.1148	0.1176	0.0793	0.0731
0.5	1.5	10	10	0.3161	0.3367	0.1508	0.1310
		10	30	0.2011	0.2253	0.1220	0.1051
		50	30	0.0794	0.1059	0.0833	0.0699
		50	50	0.0805	0.0864	0.0746	0.0637
1	1	10	10	0.7186	0.6654	0.1741	0.1449
		10	30	0.4699	0.4595	0.1348	0.1228
		50	30	0.2702	0.2666	0.0869	0.0850
		50	50	0.2316	0.2291	0.0746	0.0741
1	1.5	10	10	0.5005	0.4837	0.1686	0.1300
		10	30	0.3240	0.3293	0.1377	0.1143
		50	30	0.1791	0.1805	0.0910	0.0856
		50	50	0.1552	0.1500	0.0791	0.0764
2	1	10	10	1.6771	1.4757	0.1323	0.1298
		10	30	1.0583	0.9970	0.1045	0.1045
		50	30	0.6214	0.6013	0.0680	0.0676
		50	50	0.5232	0.5094	0.0588	0.0587
2	1.5	10	10	1.0071	0.9102	0.1051	0.1483
		10	30	0.6331	0.6076	0.1267	0.1212
		50	30	0.3754	0.3669	0.0825	0.0813
		50	50	0.3228	0.3168	0.0721	0.0710

4 Simulation study

In this section, we conduct a simulation study using the `R` software to evaluate the performance of the proposed point and interval estimation methods. The RMSE is used as the criterion for evaluating the point estimators, while the CP and AL are employed to assess the credible intervals across different values of $|\gamma|$, σ^2, n_1, and n_2.

For given parameter values and after generating $\hat{\gamma}$ and $\hat{\sigma}^2$, we first computed $\widehat{D}_{\mathrm{KL}}(F_1 \parallel F_2)$ and $\widehat{D}_{\mathrm{TV}}(F_1 \parallel F_2)$ using Equation (4), and obtained the the 95% AN confidence intervals for $D_{\mathrm{KL}}(F_1 \parallel F_2)$ and $D_{\mathrm{TV}}(F_1 \parallel F_2)$. Next, $\widetilde{D}_{\mathrm{KL}}(F_1 \parallel F_2)$ was obtained from Equation (7). Then, we obtained $\widetilde{D}_{\mathrm{TV}}(F_1 \parallel F_2)$ along with the

Table 2: The CP (first rows) and AL (second rows) of the AN, BC and HPD intervals.

				$D_{\mathrm{KL}}(F_1\|\|F_2)$			$D_{\mathrm{TV}}(F_1\|\|F_2)$		
$\|\gamma\|$	σ	n_1	n_2	AN	BC	HPD	AN	BC	HPD
0.5	1	10	10	0.8849	0.9751	0.9940	0.9945	0.9746	0.9636
				1.0315	1.1829	0.9882	0.6194	0.5026	0.4742
		10	30	0.8746	0.9461	0.9934	0.9997	0.9758	0.9646
				0.8337	0.8499	0.7199	0.5750	0.4348	0.4128
		50	30	0.9001	0.9766	0.9463	0.9835	0.9778	0.9291
				0.5536	0.4749	0.4203	0.4657	0.3177	0.3087
		50	50	0.9152	0.9543	0.9377	0.9808	0.9539	0.9222
				0.5025	0.4070	0.3667	0.4421	0.2866	0.2809
0.5	1.5	10	10	0.9219	0.9742	0.9957	0.9937	0.9745	0.9695
				0.7727	1.0048	0.8247	0.5721	0.4803	0.4485
		10	30	0.8993	0.9730	0.9919	1.0000	0.9729	0.9665
				0.5664	0.6970	0.5773	0.5031	0.4085	0.3821
		50	30	0.8686	0.9762	0.9905	1.0000	0.9798	0.9741
				0.3263	0.3447	0.2932	0.3724	0.2895	0.2747
		50	50	0.8756	0.9742	0.9897	1.0000	0.9742	0.9724
				0.2895	0.2846	0.2451	0.3470	0.2611	0.2496
1	1	10	10	0.9121	0.9717	0.9453	0.9792	0.9720	0.9136
				2.4195	2.0737	1.8323	0.7767	0.5649	0.5483
		10	30	0.9289	0.9485	0.9388	0.9854	0.9482	0.9170
				2.1677	1.5619	1.4188	0.7544	0.4869	0.4782
		50	30	0.9808	0.9524	0.9361	0.9961	0.9523	0.9452
				1.7415	0.9739	0.9360	0.7500	0.3324	0.3307
		50	50	0.9875	0.9500	0.9336	0.9980	0.9495	0.9468
				1.6582	0.8399	0.8159	0.7261	0.2903	0.2891
1	1.5	10	10	0.8713	0.9746	0.9936	0.9929	0.9745	0.9626
				1.4099	1.4310	1.2189	0.6763	0.5279	0.5039
		10	30	0.8762	0.9757	0.9921	0.9999	0.9758	0.9669
				1.1784	1.0569	0.9172	0.6521	0.4607	0.4439
		50	30	0.9409	0.9516	0.9367	0.9842	0.9516	0.9199
				0.8698	0.6299	0.5779	0.5682	0.3338	0.3293
		50	50	0.9543	0.9479	0.9334	0.9872	0.9479	0.9181
				0.8081	0.5429	0.5076	0.5467	0.2962	0.2938
2	1	10	10	0.9807	0.9520	0.9415	0.9740	0.9509	0.9291
				7.6437	4.7832	4.5356	0.8225	0.5093	0.4898
		10	30	0.9948	0.9549	0.9436	0.9988	0.9545	0.9419
				7.4021	3.5132	3.4111	0.8155	0.4017	0.3919
		50	30	1.0000	0.9511	0.9466	1.0000	0.9505	0.9431
				6.4417	2.2480	2.2178	0.7909	0.2639	0.2607
		50	50	1.0000	0.9500	0.9457	1.0000	0.9503	0.9475
				6.2323	1.9498	1.9280	0.7788	0.2317	0.2294
2	1.5	10	10	0.9452	0.9468	0.9390	0.9798	0.9465	0.9059
				3.7536	2.8358	2.5824	0.8900	0.5722	0.5586
		10	30	0.9668	0.9542	0.9417	0.9932	0.9560	0.9213
				3.4944	2.1380	2.0105	0.8893	0.4470	0.4226
		50	30	0.9952	0.9467	0.9392	0.9995	0.9473	0.9434
				2.9656	1.3498	1.3192	0.8593	0.3148	0.3128
		50	50	0.9976	0.9474	0.9403	0.9998	0.9478	0.9439
				2.8553	1.1754	1.1549	0.8385	0.2746	0.2732

95% Bayesian credible intervals for $D_{\mathrm{KL}}(F_1 \parallel F_2)$ and $D_{\mathrm{TV}}(F_1 \parallel F_2)$ by implementing Algorithm 3.1 with $M = 10{,}000$ repetitions. The RMSEs of the point estimators, as well as the CPs and ALs of the AN confidence intervals, the Bayesian credible intervals

and the HPD intervals, were computed based on 10,000 replications. These results are reported in Tables 1 and 2, respectively.

The summary of results in Tables 1-2 are listed below:

i. We observe that for every value of $|\gamma|$ and σ, the RMSE of all estimators decreases as either n_1 or n_2 increases, provided that the other is held fixed.

ii. For estimating $D_{\mathrm{KL}}(F_1 \parallel F_2)$, the RMSE of the MLE is smaller than that of the Bayes estimator in all cases. In contrast, for estimating $D_{\mathrm{TV}}(F_1 \parallel F_2)$, the Bayes estimator outperforms the MLE across all cases.

iii. For interval inference, we observe that the AN method does not attain the nominal confidence level of 0.95 in the cases considered. In addition, when constructing Bayesian intervals for both criteria $D_{\mathrm{KL}}(F_1 \parallel F_2)$ and $D_{\mathrm{TV}}(F_1 \parallel F_2)$, it is observed that in most cases, the CP of the Bayesian credible interval is closer to the nominal level of 0.95 than that of the HPD intervals.

iv. It is observed that for every value of $|\gamma|$ and σ, the AL of all estimators decreases as either n_1 or n_2 increases, provided that the other is held fixed.

v. The simulation results generally confirm that the AL of the HPD interval is shorter than that of the equal-tailed Bayesian credible interval.

Overall, we recommend using the MLE for estimating $D_{\mathrm{KL}}(F_1 \parallel F_2)$ and the Bayesian method for estimating $D_{\mathrm{TV}}(F_1 \parallel F_2)$. Additionally, the Bayesian credible interval is recommended for constructing interval estimates for both divergence criteria.

5 Real data

In this section, we present the analysis of a real data for illustrative purposes. The dataset show the breaking strengths of jute fibers at two different gauge lengths. These same datasets were previously used by Xia et al. (2009). Let X_1 and X_2 denote the breaking strengths of jute fibers at gauge lengths of 10 mm and 20 mm, respectively. We used the F test, to check the equality of variances between the two groups, obtaining a p-value of 0.74. We obtained the estimates $\hat{\mu}_1 = 365.7297$, $\hat{\mu}_2 = 340.74$, and $\hat{\sigma}^2 = 55654.64$. Using the Kolmogorov-Smirnov test, we verified that $X_1 \sim N(\hat{\mu}_1, \hat{\sigma}^2)$ with a p-value of 0.7906, and $X_2 \sim N(\hat{\mu}_2, \hat{\sigma}^2)$ with a p-value of 0.1964. Consequently, we obtain the following estimates: $\widehat{D}_{\mathrm{KL}}(F_1 \parallel F_2) = 0.0056$, $\widetilde{D}_{\mathrm{KL}}(F_1 \parallel F_2) = 0.0388$, $\widehat{D}_{\mathrm{TV}}(F_1 \parallel F_2) = 0.0422$, and $\widetilde{D}_{\mathrm{TV}}(F_1 \parallel F_2) = 0.0882$. Furthermore, for $D_{\mathrm{KL}}(F_1 \parallel F_2)$, the AN confidence interval, the Bayesian credible interval and the HPD interval are $(0, 0.0603)$, $(0.0004, 0.1940)$ and $(0, 0.1486)$, respectively. For $D_{\mathrm{TV}}(F_1 \parallel F_2)$, the corresponding intervals are $(0, 0.2478)$, $(0.0035, 0.2443)$ and $(0, 0.2144)$. In section Simulation Study 4, we recommended the Bayesian credible interval to construct interval estimates for both KL and TV divergence criteria. Therefore, we consider the 95% credible intervals of $(0.0004, 0.1940)$ and $(0.0035, 0.2443)$ for KL and TV, respectively.

6 Discussion

In this paper, we considered the estimation of the and TV divergence measures between two normal distributions with a common unknown variance. We first provided a

classical inference, including the MLE and the AN confidence interval, for these divergence measures. Then, we made point and interval inferences based on the Bayesian approach. In the Bayesian framework, improper prior distributions were considered for the population parameters, and both Bayesian point estimators and interval estimators, including credible intervals and HPD intervals, were constructed. The performance of the point estimators was evaluated using the RMSE criterion, while the interval estimators are assessed based on coverage probability and average length via simulation study. Our simulation results showed that for estimating $D_{\mathrm{KL}}(F_1 \parallel F_2)$, the RMSE of the MLE is smaller than that of the Bayes estimator in all cases. In contrast, for estimating $D_{\mathrm{TV}}(F_1 \parallel F_2)$, the Bayes estimator outperforms the MLE across all cases. For interval inference, we observed that the AN method does not attain the nominal confidence level of 0.95 in the cases considered. In addition, when constructing Bayesian intervals for both criteria $D_{\mathrm{KL}}(F_1 \parallel F_2)$ and $D_{\mathrm{TV}}(F_1 \parallel F_2)$, we observed that in most cases, the CP of the Bayesian credible interval is closer to the nominal level of 0.95 than that of the HPD intervals. The simulation results generally confirmed that the AL of the HPD interval is shorter than that of the equal-tailed Bayesian credible interval. Overall, we recommend using the MLE for estimating $D_{\mathrm{KL}}(F_1 \parallel F_2)$ and the Bayesian method for estimating $D_{\mathrm{TV}}(F_1 \parallel F_2)$. Additionally, the Bayesian credible interval is recommended for constructing interval estimates for both divergence criteria. Finally, we used the breaking strengths of jute fibers at two different gauge lengths dataset to illustrate the proposed methods.

Acknowledgments

The authors would like to thank the anonymous reviewers for their helpful comments and insightful suggestions, which improved the paper significantly.

References

Afshar Moghaddam, S., Nasiri, P. and Yarmohammadi, M. (2025). Estimators of divergence criteria for two normal distributions with Bayesian approach. *Journal of Statistical Modelling: Theory and Applications*, **6**(1):83–91.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, pp. 199-213, New York, NY: Springer New York.

Box, G.E.P. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley Publishing Company.

Chen, M.H. and Shao, Q.M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, **8**(1):69–92.

Ji, S., Zhang, Z., Ying, S., Wang, L., Zhao, X. and Gao, Y. (2020). Kullback-Leibler divergence metric learning. *IEEE Transactions on Cybernetics*, **52**(4):2047–2058.

Kullback, S. and Leibler, R.A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, **22**(1):79–86.

Moulin, P. and Veeravalli, V.V. (2018). *Statistical Inference for Engineers and Data Scientists*. Cambridge University Press.

Noh, Y.K., Sugiyama, M., Liu, S., Plessis, M.C., Park, F.C., and Lee, D.D. (2014). Bias reduction and metric learning for nearest-neighbor estimation of Kullback-Leibler divergence. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, pp. 669–677. PMLR, Iceland.

Pardo, L. (2019). New developments in statistical information theory based on entropy and divergence measures. *Entropy*, **21**(4):391.

Pérez-Cruz, F. (2008). Kullback-Leibler divergence estimation of continuous distributions. In *2008 IEEE International Symposium on Information Theory*, 1666–1670.

Weerahandi, S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association*, **88**(423):899–905.

Weerahandi, S. (1995). *Exact Statistical Methods for Data Analysis*. New York: Springer Verlag.

Xia, Z.P., Yu. J.Y., Cheng, L.D., Liu. L.F. and Wang, W.M. (2009). Study on the breaking strength of jute fibers using modified Weibull distribution. *Journal of Composites Part A: Applied Science and Manufacturing*, **40**(1):54–59.