Journal of Statistical Modelling: Theory and Applications
Vol. 6, No. 1, 2025, pp. 83-91
Yazd University Press 2025

JSMTA

Research Paper

Estimators of divergence criteria for two normal distributions with Bayesian approach

Samaneh Afshar Moghaddam, Parviz Nasiri*, Masoud Yarmohammadi
Department of Statistics, Payame Noor University, Tehran, Iran

Received: August 31, 2025/ Revised: October 27, 2025/ Accepted: November 09, 2025

Abstract: The use of statistical distributions for modeling, specifically evaluating the similarity between two probability distributions using various divergence measures, has recently attracted the attention of many of researchers to measure in context of machine learning. Given the importance of the topic, this article introduces several the divergence criteria, including Kullback-Leibler divergence, total variation divergence, alpha divergence, and power divergence, and computes the divergence parameters for two normal distributions. The parameters are estimated using both maximum likelihood and Bayesian methods. In the Bayesian approach, a conjugate distribution is used as the prior, taking into account the behavior of the parameters. Finally, the estimation methods for two normal distributions are evaluated bsased on the mean square error criterion.

Keywords: Bayesian estimation, Divergence criterion, Maximum likelihood estimation, Mean square error, Prior distribution.
Mathematics Subject Classification (2010):

1 Introduction

Comparison of different methods of estimators of the divergence criterion in a family of statistical distributions is an important issue in the comparison of statistical distributions for data analysis and selection of statistical models. It is possible to examine the similarity between distributions using the Kullback-Leibler (KL) divergence criterion, which was first introduced by Kullback and Leibler (1951). This criterion is known as one of the basic criteria for measuring the discrepancy between two probability distributions, which is widely used in various fields including metric learning,

*Corresponding author: pnasiri@pnu.ac.ir

machine learning, network security, biology, medicine, finance, and statistical data analysis Akaikei (1973). For more information, see the articles by Noh et al. (2018), Moulin and Veeravalli (2019), Pardo (2019), Ji et al. (2020), Koç et al. (2024), Ugwu et al. (2024), and Kaviani (2025). Ruding and Molapet Ding and Mullhaupt (2023) introduced an empirical estimator for the squared Hellinger distance measure between two continuous distributions and showed that the divergence estimation problem can be solved directly using the empirical cumulative distribution function and there is no need for an intermediate step of density estimation.

In the use of statistical distributions for data analysis and statistical modeling, there are various criteria for measuring the similarity between two probability distributions. For example, the KL divergence criteria, total variation divergence (TV-divergence), alpha divergence, and power divergence can be mentioned Hellinger (1909). In the application of these criteria, one of the important discussions is the estimation of these criteria Pérez-Cruz (2008). The KL divergence is another fundamental measure in information theory that indicates the degree of dissimilarity or relative distance between two probability distributions $P(x)$ and $Q(x)$. This measure determines how much information is lost or added. To introduce it, if a distribution such as $Q(x)$ is used for modeling instead of the true distribution $P(x)$, the KL divergence formula is written as:

$$D_{KL}(P||Q) = \int_R p(x) \log \frac{p(x)}{q(x)} dx,$$

where $p(x)$ and $q(x)$ are the probability density functions of the distributions $P(x)$ and $Q(x)$, respectively. The KL divergence criterion is always non-negative, and if the two distributions $P(x)$ and $Q(x)$ are identical, the value of KL becomes zero, and the larger this value, the greater the difference between the two distributions, indicating that using $Q(x)$ instead of $P(x)$ leads to more information loss.

Other methods in this field include TV-divergence, alpha divergence, and power divergence. Each of these measures has specific properties that can be useful in different applications. TV-divergence is a simple measure that measures the total difference between two distributions.

Alpha divergence and power are also measures that, based on different parameters, can measure different types of inconsistency between distributions. In this article, while introducing the KL divergence measures, TV-divergence, alpha divergence, and power divergence, and for assessing the similarity between two normal distributions, parameter estimation of the measures using maximum likelihood (ML) and Bayesian methods is presented. In Section 2, the divergence criteria are calculated for two normal distributions. In Section 3, the criteria estimators using ML and Bayesian methods are discussed. In Section 4, the estimators are compared using simulation with mean square error criterion.

2 Divergence criteria

In this section, the KL divergence criteria, TV-divergence, alpha divergence, and power divergence are introduced for two normal distributions with probability density functions $p(x)$ and $q(x)$ respectively,

$$p(x) = \frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2}, \quad -\infty < x < \infty \;, -\infty < \mu_1 < \infty, \quad \sigma_1 > 0, \tag{1}$$

$$q(x) = \frac{1}{\sigma_2\sqrt{2\pi}}e^{-\frac{1}{2\sigma_2^2}(x-\mu_2)^2}, \quad -\infty < x < \infty \;, -\infty < \mu_2 < \infty, \quad \sigma_2 > 0. \tag{2}$$

2.1 The KL divergence criterion

The KL divergence between two probability distributions $P(x)$ and $Q(X)$ is given by

$$D_{KL}(P||Q) = \int_R p(x)\log\frac{p(x)}{q(x)}dx.$$

Such that with considering the equations relations (1) and (2), can be written

$$\begin{aligned}
D(P||Q) &= \int_R q(x)\log\frac{p(x)}{q(x)}dx \\
&= \int_{-\infty}^{\infty}\frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2}\log\left(\frac{\sigma_2}{\sigma_1}e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2+\frac{1}{2\sigma_2^2}(x-\mu_2)^2}\right)dx \\
&= \log\left(\frac{\sigma_2}{\sigma_1}\right) - \frac{1}{2\sigma_1^2}\int_{-\infty}^{\infty}\frac{1}{\sigma_1\sqrt{2\pi}}(x-\mu_1)^2e^{-\frac{1}{2}(x-\mu_1)^2}dx \\
&\quad + \frac{1}{2\sigma_2^2}\int_{-\infty}^{\infty}\frac{1}{\sigma_1\sqrt{2\pi}}(x-\mu_2)^2e^{-\frac{1}{2}(x-\mu_1)^2}dx \\
&= \log\left(\frac{\sigma_2}{\sigma_1}\right) - \frac{1}{2} + \frac{1}{2\sigma_2^2}\int_{-\infty}^{\infty}\frac{1}{\sigma_1\sqrt{2\pi}}(x-\mu_2)^2e^{-\frac{1}{2}(x-\mu_1)^2}dx \\
&= \log\left(\frac{\sigma_2}{\sigma_1}\right) - \frac{1}{2} + \frac{1}{2\sigma_2^2}[E(X^2)+\mu_2^2-2\mu_1\mu_2] \\
&= \log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{1}{2}\frac{\sigma_1^2}{\sigma_2^2} + \frac{1}{2}\frac{1}{\sigma_2^2}(\mu_1-\mu_2)^2 - \frac{1}{2}.
\end{aligned} \tag{3}$$

In the special case for $\sigma_1^2 = \sigma_2^2 = 1$, $D(P||Q)$ is given by

$$D(P||Q) = \int_R p(x)\log\frac{p(x)}{q(x)} = \frac{1}{2}(\mu_1-\mu_2)^2.$$

2.2 total variation divergence criterion

The TV-divergence between two distributions $P(x)$ and $Q(x)$ is given by

$$D_{TV}(P||Q) = \frac{1}{2}\int_{-\infty}^{\infty}|p(x)-q(x)|dx.$$

By using of the (1) and (2), $D_{TV}(P||Q)$ can be written

$$D_{TV}(P||Q) = \frac{1}{2}\int_{-\infty}^{\infty}|p(x)-q(x)|dx$$

$$= \frac{1}{2}\int_{-\infty}^{\infty}\left|\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu_1)^2}-\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu_2)^2}\right|dx$$
$$= \frac{1}{2\sqrt{2\pi}}\int_{-\infty}^{\infty}\left|e^{-\frac{1}{2}(x-\mu_1)^2}-e^{-\frac{1}{2}(x-\mu_2)^2}\right|dx.$$

The following two cases
1- If $e^{-\frac{1}{2}(x-\mu_1)^2}-e^{-\frac{1}{2}(x-\mu_2)^2}>0$, then $x<\frac{\mu_1+\mu_2}{2}$,
2- If $e^{-\frac{1}{2}(x-\mu_1)^2}-e^{-\frac{1}{2}(x-\mu_2)^2}<0$, then $x>\frac{\mu_1+\mu_2}{2}$,
are considered to calculate the integral, it is easy to show that

$$D_{TV}(P||Q)=\begin{cases}\frac{1}{2}\int_{\frac{\mu_1+\mu_2}{2}}^{\infty}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu_2)^2}-\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu_1)^2}\right)dx,\\ \frac{1}{2}\int_{\infty}^{\frac{\mu_1+\mu_2}{2}}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu_1)^2}-\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu_2)^2}\right)dx.\end{cases}$$

But

$$\int_{\frac{\mu_1+\mu_2}{2}}^{\infty}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu_2)^2}-\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu_1)^2}\right)dx$$
$$= \int_{\frac{\mu_1-\mu_2}{2}}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz-\int_{-\frac{\mu_1-\mu_2}{2}}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz$$
$$= \phi\left(-\frac{\mu_1-\mu_2}{2}\right)-\phi\left(\frac{\mu_1-\mu_2}{2}\right),$$

such that, $\phi(z)=\int_{-\infty}^{t}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$. Therefore $D_{TV}(P||Q)$ is given by

$$D_{TV}(P||Q)=\frac{1}{2}\left(\phi\left(-\frac{\mu_1-\mu_2}{2}\right)-\phi\left(\frac{\mu_1-\mu_2}{2}\right)\right).$$

2.3 Alpha divergence criterion

The alpha divergence between two probability distributions $P(x)$ and $Q(x)$ is given by

$$D_\alpha(P||Q)=\frac{1}{\alpha(1-\alpha)}\left(1-\int_{\infty}^{-\infty}p(x)^\alpha q(x)^{1-\alpha}dx\right),\qquad \alpha\neq 0,1.$$

By according the equations (1) and (2), $D_\alpha(P||Q)$ can be written

$$D_\alpha(P||Q) = \frac{1}{\alpha(1-\alpha)}\left(1-\int_{\infty}^{-\infty}p(x)^\alpha q(x)^{1-\alpha}dx\right)$$
$$= \frac{1}{\alpha(1-\alpha)}-\frac{1}{\alpha(1-\alpha)}\int_{\infty}^{-\infty}p(x)^\alpha q(x)^{1-\alpha}dx$$
$$= \frac{1}{\alpha(1-\alpha)}-\frac{1}{\alpha(1-\alpha)}\int_{\infty}^{-\infty}\left(\frac{p(x)}{q(x)}\right)^\alpha q(x)\ dx.$$

But for $p(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu_1)^2}$ and $q(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu_2)^2}$, the value of $\left(\frac{p(x)}{q(x)}\right)^\alpha$ is given by

$$\left(\frac{p(x)}{q(x)}\right)^\alpha=\left(\frac{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu_1)^2}}{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu_2)^2}}\right)=e^{-\frac{1}{2}(\mu_1^2-\mu_2^2)\alpha+(\mu_1-\mu_2)\alpha x}.$$

Therefore, the alpha divergence criterion is equal to

$$\begin{aligned}
D_\alpha(P||Q) &= \frac{1}{\alpha(1-\alpha)} - \frac{1}{\alpha(1-\alpha)} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\mu_1^2-\mu_2^2)\alpha+(\mu_1-\mu_2)\alpha x} e^{-\frac{1}{2}(x-\mu_2)^2} dx \\
&= \frac{1}{\alpha(1-\alpha)} - \frac{1}{\alpha(1-\alpha)} e^{-\frac{1}{2}(\mu_1^2-\mu_2^2)\alpha-\frac{1}{2}\mu_2^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2+[\mu_2+(\mu_1-\mu_2)\alpha]x} dx \\
&= \frac{1}{\alpha(1-\alpha)} - \frac{1}{\alpha(1-\alpha)} e^{-\frac{1}{2}(\mu_1^2-\mu_2^2)\alpha-\frac{1}{2}\mu_2^2} e^{\frac{1}{2}[\mu_2+(\mu_1-\mu_2)\alpha]^2} \\
&= \frac{1}{\alpha(1-\alpha)} - \frac{1}{\alpha(1-\alpha)} e^{-\frac{1}{2}(\mu_1-\mu_2)^2\alpha+\frac{1}{2}(\mu_1^2-\mu_2^2)\alpha^2}.
\end{aligned}$$

2.4 Power divergence criterion

The power divergence between two probability distributions $P(x)$ and $Q(x)$ is given by

$$D_\lambda(P||Q) = \frac{2}{\lambda(\lambda+1)} \int_{-\infty}^{\infty} p(x) \left[\left(\frac{p(x)}{q(x)} \right)^\lambda - 1 \right] dx, \quad \lambda \neq 0.$$

By according the equations (1) and (2) $D_\lambda(P||Q)$ can be written

$$\begin{aligned}
D_\lambda(P||Q) &= \frac{2}{\lambda(\lambda+1)} \int_{-\infty}^{\infty} p(x) \left[\left(\frac{p(x)}{q(x)} \right)^\lambda - 1 \right] dx \\
&= \frac{2}{\lambda(\lambda+1)} \int_{-\infty}^{\infty} p(x) \left(\frac{p(x)}{q(x)} \right)^\lambda dx - \frac{2}{\lambda(\lambda+1)} \\
&= \frac{2}{\lambda(\lambda+1)} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu_1)^2} e^{-\frac{1}{2}(\mu_1^2-\mu_2^2)\lambda-(\mu_1-\mu_2)\lambda x} dx - \frac{2}{\lambda(\lambda+1)} \\
&= \frac{2}{\lambda(\lambda+1)} e^{-\frac{1}{2}(\mu_1^2-\mu_2^2)\lambda-\frac{1}{2}\mu_1^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2+[\mu_1+(\mu_1-\mu_2)\lambda]x} dx - \frac{2}{\lambda(\lambda+1)} \\
&= \frac{2}{\lambda(\lambda+1)} e^{-\frac{1}{2}(\mu_1^2-\mu_2^2)\lambda-\frac{1}{2}\mu_1^2} e^{\frac{1}{2}[\mu_1+(\mu_1-\mu_2)\lambda]^2} - \frac{2}{\lambda(\lambda+1)} \\
&= \frac{2}{\lambda(\lambda+1)} \left(e^{-\frac{1}{2}(\mu_1-\mu_2)^2\lambda+\frac{1}{2}(\mu_1^2-\mu_2^2)\lambda^2} - 1 \right).
\end{aligned}$$

Lemma 2.1. *If we consider $\alpha = \lambda = 3$, then the alpha divergence criterion is equal to the power divergence criterion*

Proof.

$$\begin{aligned}
\frac{D_\alpha(P||Q)}{D_\lambda(P||Q)} &= \frac{-\frac{1}{\alpha(1-\alpha)} \left(e^{-\frac{1}{2}(\mu_1-\mu_2)^2\alpha+\frac{1}{2}(\mu_1^2-\mu_2^2)\alpha^2} - 1 \right)}{\frac{2}{\alpha(\alpha+1)} \left(e^{-\frac{1}{2}(\mu_1-\mu_2)^2\alpha+\frac{1}{2}(\mu_1^2-\mu_2^2)\alpha^2} - 1 \right)} = -\frac{(\alpha+1)}{2(1-\alpha)}, \\
\frac{D_\alpha(P||Q)}{D_\lambda(P||Q)} &= 1, \rightarrow (\alpha+1) = 2(\alpha-1) \rightarrow \alpha = 3.
\end{aligned}$$

□

3 Estimation of divergence criteria

In this section, divergence criteria are estimated using ML and Bayesian methods. For this purpose, samples of $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ with of equal sizes are selected from normal $N(\mu_1, 1)$ and normal $N(\mu_2, 1)$ distributions, respectively.

3.1 Maximum likelihood estimation

According to the equation (3) for two normal distributions with means μ_1 and μ_2 for unit variances with random samples $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$, the KL divergence for ML estimators of the means of the two normal distributions is equal to

$$\hat{D}_{kl}(P||Q) = \frac{1}{2}\hat{\mu}_1^2 + \frac{1}{2}\hat{\mu}_2^2 - \hat{\mu}_1\hat{\mu}_2 = \frac{1}{2}\bar{X}^2 + \frac{1}{2}\bar{Y}^2 - \bar{X}\bar{Y}. \tag{4}$$

Similarly, the estimators of the TV-divergence criteria, alpha divergence, and power divergence are respectively equal to

$$\begin{aligned}
\hat{D}_{TV}(P||Q) &= \frac{1}{2}\left(\phi\left(-\frac{\hat{\mu}_1 - \hat{\mu}_2}{2}\right) - \phi\left(\frac{\hat{\mu}_1 - \hat{\mu}_2}{2}\right)\right) \\
&= \frac{1}{2}\left(\phi\left(-\frac{\bar{X} - \bar{Y}}{2}\right) - \phi\left(\frac{\bar{X} - \bar{Y}}{2}\right)\right), \qquad (5) \\
\hat{D}_{\alpha}(P||Q) &= \frac{1}{\alpha(1-\alpha)}\left(1 - e^{-\frac{1}{2}(\hat{\mu}_1 - \hat{\mu}_2)^2\alpha + \frac{1}{2}(\hat{\mu}_1^2 - \hat{\mu}_2^2)\alpha^2}\right) \\
&= \frac{1}{\alpha(1-\alpha)}\left(1 - e^{-\frac{1}{2}(\bar{X} - \bar{Y})^2\alpha + \frac{1}{2}(\bar{X}^2 - \bar{Y}^2)\alpha^2}\right), \qquad (6) \\
\hat{D}_{\lambda}(P||Q) &= \frac{2}{\lambda(\lambda+1)}\left(e^{-\frac{1}{2}(\hat{\mu}_1 - \hat{\mu}_2)^2\lambda + \frac{1}{2}(\hat{\mu}_1^2 - \hat{\mu}_2^2)\lambda^2} - 1\right) \\
&= \frac{2}{\lambda(\lambda+1)}\left(e^{-\frac{1}{2}(\bar{X} - \bar{Y})^2\lambda + \frac{1}{2}(\bar{X}^2 - \bar{Y}^2)\lambda^2} - 1\right). \qquad (7)
\end{aligned}$$

3.2 Bayesian estimation

In this section, we will use Bayesian estimates of divergence criteria with respect to a random sample $(X_1, \ldots, X_n)$ from $N(\mu_1, 1)$ for the Bayesian record, with the prior distribution $\pi(\mu_1) = \frac{1}{\sqrt{2\pi}}e^{-\frac{\mu_1^2}{2}}, -\infty < \mu_1 < \infty$, the posterior distribution is equal to

$$\pi(\mu_1|\underline{x}) \sim N\left(\frac{\sum_{i=1}^n x_i}{n+1}, \frac{1}{n+1}\right). \tag{8}$$

Similarly, for a random sample $(Y_1, \ldots, Y_n)$ from $N(\mu_2, 1)$ with prior distribution $\pi(\mu_2) = \frac{1}{\sqrt{2\pi}}e^{-\frac{\mu_2^2}{2}}, -\infty < \mu_2 < \infty$, the posterior distribution is equal to

$$\pi(\mu_2|\underline{y}) \sim N\left(\frac{\sum_{i=1}^n y_i}{n+1}, \frac{1}{n+1}\right). \tag{9}$$

According to relations (8) and (9), the Bayesian estimation of the parameters under the squared error loss function is equal to

$$\hat{\mu}_1 = \frac{n}{n+1}\bar{X} \qquad and \qquad \hat{\mu}_2 = \frac{n}{n+1}\bar{Y}.$$

By inserting $\hat{\mu}_1 = \frac{n}{n+1}\bar{X}$ and $\hat{\mu}_2 = \frac{n}{n+1}\bar{Y}$ in (4), (5), (6) and (7) of the Bayesian estimator, the KL divergence criteria, TV-divergence, alpha divergence and power divergence are respectively equal to

$$\begin{aligned}
\hat{D}_{KL}(P||Q) &= \frac{1}{2}\left(\frac{n}{n+1}\right)^2 \bar{X}^2 + \frac{1}{2}\left(\frac{n}{n+1}\right)^2 \bar{Y}^2 - \left(\frac{n}{n+1}\right)^2 \bar{X}\bar{Y}, \\
\hat{D}_{TV}(P||Q) &= \frac{1}{2}\left(\phi\left(-\frac{n}{2(n+1)}(\bar{X}-\bar{Y})\right) - \phi\left(\frac{n}{2(n+1)}(\bar{X}-\bar{Y})\right)\right), \\
\hat{D}_{\alpha}(P||Q) &= \frac{1}{\alpha(1-\alpha)}\left(1 - e^{-\frac{1}{2}\left(\frac{n}{n+1}\right)^2(\bar{X}-\bar{Y})^2\alpha + \frac{1}{2}\left(\frac{n}{n+1}\right)^2(\bar{X}^2-\bar{Y}^2)\alpha^2}\right), \\
\hat{D}_{\lambda}(P||Q) &= \frac{2}{\lambda(\lambda+1)}\left(e^{-\frac{1}{2}\left(\frac{n}{n+1}\right)^2(\bar{X}-\bar{Y})^2\lambda + \frac{1}{2}\left(\frac{n}{n+1}\right)^2(\bar{X}^2-\bar{Y}^2)\lambda^2} - 1\right).
\end{aligned}$$

Table 1: Bias and mean square error of the divergence estimator when $\alpha = 0.75$, $\lambda = 0.5$ and $\mu_1 = 1.5$.

n	μ_2		$D_{KL}(P\|\|Q)$ Bias	$D_{KL}(P\|\|Q)$ MSE	$D_{TV}(P\|\|Q)$ Bias	$D_{TV}(P\|\|Q)$ MSE	$D_{\alpha}(P\|\|Q)$ Bias	$D_{\alpha}(P\|\|Q)$ MSE	$D_{\lambda}(P\|\|Q)$ Bias	$D_{\lambda}(P\|\|Q)$ MSE
10	1.75	MLE	0.10085	0.03451	0.06860	0.00717	0.56553	0.47606	0.40351	0.23829
		Bayes	0.06373	0.02285	0.06310	0.00608	0.49251	0.35574	0.35618	0.18535
	2	MLE	0.15787	0.06018	0.06795	0.00706	0.66423	0.63582	0.40186	0.25420
		Bayes	0.02887	0.03938	0.06372	0.00625	0.61673	0.54093	0.38523	0.23079
20	1.75	MLE	0.06136	0.01128	0.04826	0.00367	0.42923	0.28194	0.30307	0.14126
		Bayes	0.03514	0.00907	0.04618	0.00337	0.39881	0.24137	0.28418	0.12429
	2	MLE	0.11452	0.02620	0.04746	0.00357	0.49113	0.36520	0.28768	0.13411
		Bayes	0.01580	0.02105	0.04590	0.00335	0.47146	0.33551	0.28231	0.12989
30	1.75	MLE	0.04792	0.00591	0.03920	0.00244	0.35565	0.19630	0.25035	0.09919
		Bayes	0.02360	0.00509	0.03808	0.00230	0.33829	0.17667	0.23977	0.09112
	2	MLE	0.09428	0.01605	0.03851	0.00236	0.40129	0.25049	0.23208	0.08925
		Bayes	0.00998	0.01386	0.03768	0.00226	0.39068	0.23697	0.22978	0.08803
40	1.75	MLE	0.04403	0.00491	0.03547	0.00196	0.32823	0.16601	0.22907	0.08105
		Bayes	0.02003	0.00438	0.03472	0.00188	0.31640	0.15316	0.22204	0.07609
	2	MLE	0.08757	0.01379	0.03479	0.00189	0.36823	0.20698	0.21027	0.07007
		Bayes	0.01006	0.01233	0.03425	0.00183	0.36085	0.19805	0.20877	0.06954
50	1.75	MLE	0.03917	0.00344	0.03217	0.00160	0.29807	0.13471	0.20839	0.06665
		Bayes	0.01583	0.00314	0.03162	0.00154	0.28920	0.12633	0.20313	0.06341
	2	MLE	0.07893	0.01051	0.03153	0.00154	0.33389	0.16924	0.19046	0.05743
		Bayes	0.00713	0.00962	0.03111	0.00150	0.32837	0.16360	0.18944	0.05721
60	1.75	MLE	0.03538	0.00270	0.02883	0.00128	0.26841	0.11088	0.18705	0.05418
		Bayes	0.01305	0.00250	0.02839	0.00125	0.26144	0.10506	0.18292	0.05199
	2	MLE	0.07128	0.00855	0.02822	0.00123	0.29908	0.13775	0.16966	0.04567
		Bayes	0.00612	0.00794	0.02788	0.00121	0.29458	0.13392	0.16872	0.04561

Table 2: Bias and mean square error of the divergence estimator when $\alpha = 0.5$, $\lambda = 0.5$ and $\mu_1 = 1$.

			$D_{KL}(P\|\|Q)$		$D_{TV}(P\|\|Q)$		$D_\alpha(P\|\|Q)$		$D_\lambda(P\|\|Q)$	
n	μ_2		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
10	1.2	MLE	0.09616	0.03153	0.06865	0.00718	0.41013	0.28314	0.27342	0.12584
		Bayes	0.06822	0.02105	0.06299	0.00605	0.35253	0.20376	0.23502	0.09056
	1.75	MLE	0.23804	0.10514	0.06667	0.00686	0.64894	0.60692	0.43263	0.26974
		Bayes	0.02665	0.07203	0.06442	0.00643	0.63744	0.57641	0.42496	0.25618
20	1.2	MLE	0.05443	0.00955	0.04835	0.00365	0.31394	0.15946	0.20929	0.07087
		Bayes	0.03765	0.00771	0.04621	0.00337	0.28976	0.13372	0.19318	0.05943
	1.75	MLE	0.17395	0.05193	0.04610	0.00339	0.47085	0.33590	0.31390	0.14929
		Bayes	0.01489	0.04282	0.04537	0.00329	0.46576	0.32828	0.31051	0.14590
30	1.2	MLE	0.04071	0.00475	0.03928	0.00245	0.26098	0.10864	0.17398	0.04829
		Bayes	0.02534	0.00410	0.03812	0.00231	0.24699	0.09636	0.16466	0.04283
	1.75	MLE	0.14310	0.03358	0.03731	0.00223	0.38043	0.22598	0.25362	0.10043
		Bayes	0.01144	0.02954	0.03692	0.00219	0.37889	0.22348	0.25259	0.09933
40	1.2	MLE	0.03657	0.00384	0.03555	0.00197	0.24253	0.09348	0.16169	0.04155
		Bayes	0.02131	0.00344	0.03476	0.00188	0.23279	0.08532	0.15519	0.03792
	1.75	MLE	0.13200	0.02868	0.03365	0.00177	0.34833	0.18495	0.23222	0.08220
		Bayes	0.00587	0.02596	0.03340	0.00174	0.34632	0.18279	0.23088	0.08124
50	1.2	MLE	0.03198	0.00261	0.03224	0.00161	0.21968	0.07435	0.14646	0.03304
		Bayes	0.01699	0.00238	0.03167	0.00155	0.21251	0.06909	0.14167	0.03071
	1.75	MLE	0.11923	0.02253	0.03047	0.00144	0.31524	0.15114	0.21016	0.06720
		Bayes	0.00638	0.02084	0.03024	0.00142	0.31406	0.15006	0.20937	0.06669
60	1.2	MLE	0.02851	0.00200	0.02890	0.00129	0.19839	0.06160	0.13226	0.02738
		Bayes	0.01395	0.00186	0.02845	0.00125	0.19274	0.05792	0.12849	0.02574
	1.75	MLE	0.10736	0.01840	0.02723	0.00115	0.28146	0.12236	0.18764	0.05438
		Bayes	0.00484	0.01722	0.02702	0.00114	0.28015	0.12172	0.18677	0.05410

4 Simulation

In this study, based on the simulation of data from normal distributions, the performance of ML and Bayesian estimators for the KL divergence criteria, TV-divergence, alpha divergence, and power divergence are examined and compared. To evaluate the estimator methods, while using bias and mean square error criteria,

$$\begin{aligned}
\hat{D}_{mle}(P||Q)_i - \bar{\hat{D}}_{mle}(P||Q) &= \hat{D}_{mle}(P||Q) - \frac{1}{1000}\sum_{i=1}^{1000}\hat{D}_{mle}(P||Q)_i, \\
MSE\left(\hat{D}_{mle}(P||Q)\right) &= \frac{1}{1000}\sum_{i=1}^{1000}\left(\hat{D}_{mle}(P||Q)_i - \bar{\hat{D}}_{mle}(P||Q)_i\right)^2.
\end{aligned}$$

An attempt has been made to measure the strengths and weaknesses of each method in different conditions. The simulation results using different methods are given in tables 1 and 2. It is worth noting that the simulation steps were performed using R software and the number of simulation iterations was increased to 1000 to obtain a more stable and accurate estimate. According to the simulation results, it can be observed that

i. As the sample size increases, the mean square error of all estimators of the divergence criterion decreases.

ii. Bayesian estimators of the KL divergence criterion, TV-divergence, alpha divergence, and power divergence perform better than ML estimators.

iii. Considering the mean square error criterion, the TV-divergence criterion performs better with respect to estimator methods compared to other divergence criteria.
iv. In conclusion, it can be said that the use of the Bayesian estimator causes a significant reduction in the mean square error.
v. The TV-divergence measure estimator perform better than other measures.
vi. For a fixed mean value of μ_1, with increasing the μ_2, the divergence measure of all estimators increases.

References

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

Ding, R. and Mullhaupt, A. (2023). Empirical squared Hellinger distance estimator and generalizations to a family of -divergence estimators. *Entropy,* **25**(4):612.

Hellinger, E. (1909). Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen. *Journal Für Die Reine Und Angewandte Mathematik,* **1909**(136):210–271.

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.

Kaviani, M. (2025). Forecasting the return of government exchange-traded funds based on linear and nonlinear models in machine learning algorithms. *Innovation Management and Operational Strategies,* **6**(1):59–69.

Koç, S., Erden, C., Ateş, Ç. and Ceviz, E. (2024). Evaluation of potential logistics village alternatives using bayesian best-worst method. *Optimality,* **1**(1):100–120.

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. (2019). *Statistical Inference for Engineers and Data Scientists.* Cambridge University Press.

Noh, Y.K., Sugiyama, M., Liu, S., du Plessis, M.C., Park, F.C. and Lee, D.D. (2018). Bias reduction and metric learning for nearest-neighbor estimation of Kullback-Leibler divergence. *Neural Computation,* **30**(7):1930–1960.

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,* pp. 1666-1670. IEEE.

Ugwu, D.N., Onyeagu, S.I. and Igbokwe, C.P. (2024). A new weighted T-X perks distribution: Characterization, simulation and applications. *Optimality,* **1**(1):66–81.