Weighted Means of Likelihood Ratio as Indices Measuring Similarity Between Densities

Hideyoshi Ko (✉ ko@suzuka-u.ac.jp)

Suzuka University of Medical Science: Suzuka Iryo Kagaku Daigaku

Research Article

Keywords: Density estimation, Parametric Inference, Nonparametric Inference, Mixture Distribution

Posted Date: November 30th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1081144/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Weighted Means of Likelihood Ratio as Indices Measuring Similarity Between Densities

Hideyoshi Ko^{1*}

^{1*}Department of Clinical Engineering, Suzuka University of Medical Science, Kishioka, Suzuka-city, 5100293, Mie, Japan.

Corresponding author(s). E-mail(s): ko@suzuka-u.ac.jp;

Abstract

Criteria for similarity between probability density functions are important in the field of statistics such as density estimation. In this short paper, a set of indices measuring similarity between probability densities is proposed using the weighted means of the likelihood ratio function. Numerical simulations demonstrate that the estimates of these indices are easily obtained from observations and could be useful for both parametric and nonparametric density estimation with numerical optimization.

Keywords: Density estimation, Parametric Inference, Nonparametric Inference, Mixture Distribution

1 Introduction

Criteria for similarity between probability density functions are important in the field of statistics such as density estimation, goodness-of-fit test, etc. Today, the supremum norm, the integrated absolute error (IAE, L^1 -norm in other words), the integrated squared error (ISE, equivalent to the squared L^2 -norm), the likelihood ratio function, the Kullback-Leibler distance, the Hellinger distance, and other tools are ready for practical applications (Scott, 2015). Although these tools are useful enough, there could be another one since the choice of optimality criteria may be somewhat subjective. In this paper, a set of indices measuring similarity between probability densities is proposed, and numerical examples demonstrate that the estimates of these indices are easily obtained from observations and useful for both parametric and nonparametric density estimation with numerical optimization.

2 Methods

We can deduce from the high utility of the likelihood ratio that the set of weighted means $I_1(g; f), \dots, I_n(g; f)$ of the form

$$I_k(g; f) := \int_0^1 b_k(y) \left[1 - \frac{g \circ F^{-1}(y)}{f \circ F^{-1}(y)} \right] dy \quad (1)$$

can be used to evaluate the similarity between two density functions f and g , where n is a positive integer, F is the cumulative distribution function corresponding to the density f , and b_k is the density function of the beta distribution with shape parameters k and $n-k+1$. Indeed, $I_k(g; f) = 0$ for all $k = 1, \dots, n$ when $g = f$ almost everywhere. Moreover, if we assume that the approximation

$$1 - \frac{g \circ F^{-1}(y)}{f \circ F^{-1}(y)} = \sum_{k=1}^n \alpha_k b_k(y) + \varepsilon(y)$$

with the small error term $\varepsilon(y)$ is available with real coefficients α_k by virtue of sharp localized shape of b_k for large n , the similarity between f and g would be inferred from the values of $I_k(g; f)$. That is to say, if there exist small positive real numbers ε_0 and δ_0 such that $\sup_y |\varepsilon(y)| \leq \varepsilon_0$ and $|I_k(g; f)| \leq \delta_0$ for all k , then g is similar to f , since

$$\begin{aligned} & \int_0^1 \left[1 - \frac{g \circ F^{-1}(y)}{f \circ F^{-1}(y)} \right]^2 dy \\ &= \left| \int_0^1 \left(\sum_{k=1}^n \alpha_k b_k(y) + \varepsilon(y) \right) \left[1 - \frac{g \circ F^{-1}(y)}{f \circ F^{-1}(y)} \right] dy \right| \\ &\leq \left| \sum_{k=1}^n \alpha_k I_k(g; f) \right| + \left| \int_0^1 \varepsilon(y) \left[1 - \frac{g \circ F^{-1}(y)}{f \circ F^{-1}(y)} \right] dy \right| \\ &\leq \delta_0 \left(\sum_{k=1}^n |\alpha_k| \right) + \varepsilon_0 \int_0^1 \left| 1 - \frac{g \circ F^{-1}(y)}{f \circ F^{-1}(y)} \right| dy \end{aligned}$$

A little calculation shows that $I_k(g; f)$ is also given by

$$\begin{aligned} I_k(g; f) &= \int_{\mathbb{R}} \left(1 - \frac{F(x)^{k-1} [1 - F(x)]^{n-k}}{B(k, n-k+1)} \right) g(x) dx \quad (2) \\ &= 1 - \int_{\mathbb{R}} \frac{f_k(x)}{f(x)} g(x) dx \quad (3) \end{aligned}$$

where f_k is the density function of the order statistic $X_{(k)}$ of a random sample of size n from the population with the density f , and B is the beta function. An important property $\sum_{k=1}^n I_k(g; f) = 0$ is derived from equation (3) and $\sum_{k=1}^n f_k(x) = nf(x)$, and the relation $I_k(\sum_{l=1}^m \lambda_l g_l; f) = \sum_{l=1}^m \lambda_l I_k(g_l; f)$ for densities $g_1, \dots, g_m$ with coefficients $\lambda_1, \dots, \lambda_m$ satisfying $\sum_{l=1}^m \lambda_l = 1$ is obvious by the definition.

It is convenient to calculate the estimate of $I_k(g; f)$ from n observations $x_1, x_2, \dots, x_n$ by equation (2) in the following two ways:

(a) The empirical density function $\hat{g}(x) = \frac{1}{n} \chi_{\{x_1, x_2, \dots, x_n\}}(x)$ is employed, where χ is an indicator function, and the estimate $I_k(\hat{g}; f)$ for given

f is obtained by

$$I_k(\hat{g}; f) = 1 - \binom{n-1}{k-1} \sum_{j=1}^n F(x_j)^{k-1} [1 - F(x_j)]^{n-k} \quad (4)$$

(b) The empirical distribution function $\hat{F}(x) = \frac{1}{n} \sum_{j=1}^n \chi_{\{y: y \geq x_j\}}(x)$ is employed, and the estimate $I_k(g; \hat{f})$ for given g is obtained by

$$I_k(g; \hat{f}) = \sum_{j=1}^{n-1} w_{k,j} [G(x_{j+1}) - G(x_j)] \quad (5)$$

where

$$w_{k,j} = 1 - B(k, n-k+1)^{-1} j^{k-1} (n-j)^{n-k} n^{1-n} \quad (6)$$

and G is the distribution function corresponding to g .

Numerical examples in the next section present applications of these estimates to probability density estimation.

3 Numerical Examples

In this section, numerical simulations of parametric and nonparametric density estimation are demonstrated as the applications of the proposed indices $I_k(g; f)$. Each density estimation was performed with a random sample of size $n = 51$ and repeated 10000 times in each setting. All numerical experiments were implemented in Mathematica 12.3.1 (Wolfram Research Inc., Champaign, IL). `FindMinimum`, `NIntegrate` and `SmoothKernelDistribution` of Mathematica functions by default settings were used for searching local minima, calculating L^1 and L^2 -norms, and obtaining estimates by the conventional nonparametric density estimator, respectively.

Parametric Density Estimation

All observations were generated from the Cauchy distribution with location parameter $\mu_0 = 2$ and scale parameter $\sigma_0 = 2$. The values of $I_k(g; f)$ were estimated by equation (4), where F was set to be the distribution function of the Cauchy distribution with unknown location parameter μ and scale parameter σ . Each density estimation was

performed as solving the unconstrained optimization problem

$$\min_{\mu, \sigma} \sum_{k=1}^n I_k(\hat{g}; f)^2 \quad (7)$$

to determine unknown parameters. Figure 1 shows the histograms and density histogram of the estimates $\hat{\mu}$ and $\hat{\sigma}$, and the descriptive statistics of these estimates are given in Table 1. The estimates may be unbiased and the distribution of $\hat{\mu}$ might be normal, but the distribution of $\hat{\sigma}$ is skewed.

Nonparametric Density Estimation

Observations were generated from the mixture distribution whose density was $f(x) = 0.3\phi_{2,1}(x) + 0.4\phi_{6,0.8}(x) + 0.3\phi_{10,2}(x)$, where $\phi_{\mu,\sigma}(x)$ is the density function of the normal distribution with mean μ and standard deviation σ . In this numerical example, the idea of kernel density estimation was employed, and the Gaussian kernels g_l , $l = 1, \dots, 2n/3 + 1$ whose location and scale parameters were calculated from the sets of ordered observations $x_{(l)}, \dots, x_{(l+n/3-1)}$ were applied. The value of $I_k(g_l; f)$ for each density g_l was estimated by equation (5) with Stirling's formula for calculating factorials in the beta function. Each density estimation was performed as solving the constrained optimization problem

$$\begin{aligned} \min_{\lambda_1, \lambda_2, \dots, \lambda_{2n/3+1} \geq 0} & \sum_{k=1}^n \left[\sum_{l=1}^{2n/3+1} \lambda_l I_k(g_l; \hat{f}) \right]^2 \\ \text{subject to} & \sum_{l=1}^{2n/3+1} \lambda_l = 1 \end{aligned} \quad (8)$$

to determine the optimal mixture weights λ_l for g_l , and the estimated density $\hat{g}$ was given by $\hat{g} = \sum_{l=1}^{2n/3+1} \lambda_l g_l$. Figure 2 shows the histograms of $\|\hat{g} - f\|_{L^1}$ and $\|\hat{g} - f\|_{L^2}$, and the descriptive statistics of them are presented in Table 2 – the integrals over $\mathbb{R}$ required to obtain these norms were approximated by the numerical integrals over the interval $[-5, 20]$. The L^1 -norm best-fit and worst-fit estimates are depicted in Figure 3. One should notice that the density estimation with the proposed method is more accurate than that with conventional one but less precise, and the values of L^1 and L^2 -norms are not normally distributed.

In addition, it must be mentioned that the proposed estimator would not always be superior to the conventional one in accuracy, since the problem of bandwidth selection heavily affects kernel density estimates.

4 Conclusion

The set of indices to measure similarity between probability densities is proposed in this paper, and its applications to parametric and nonparametric density estimation have been presented. Numerical examples have shown that the density estimation with the proposed indices could have higher accuracy and less precision than that with the conventional one. Although the proposed indices are simply explained as the weighted means of the likelihood ratio function, a more detailed theoretical analysis must be carried out for further applications. In addition, similar indices for multivariate distributions should be explored, since there is an increasing interest in multivariate density estimation, and the definition of the proposed set of indices is easily extended to the multivariate case with the density function of the multivariate beta distribution.

Declarations

- **Funding:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
- **Competing interests:** None.

References

Scott DW (2015) Multivariate Density Estimation: Theory, Practice, and Visualization, Second Edition. John Wiley & Sons

Table 1 Descriptive statistics of estimates in parametric density estimation for the Cauchy distribution.

	Mean	Median (IQR)	Min - Max
Estimates of location parameter $\mu_0 = 2$	1.998	1.996 (1.728 - 2.265)	0.492 - 3.739
Estimates of scale parameter $\sigma_0 = 2$	2.085	2.041 (1.782 - 2.340)	0.845 - 5.388

Fig. 1 Histograms of estimates in parametric density estimation by the proposed method for Cauchy distribution with location parameter $\mu_0 = 2$ and scale parameter $\sigma_0 = 2$.**Table 2** Descriptive statistics of L^1 and L^2 -norms in nonparametric density estimation.

		Proposed method	Conventional method
L^1 -norm	Mean	0.322	0.399
	Median (IQR)	0.315 (0.258 - 0.379)	0.397 (0.370 - 0.427)
	Min - Max	0.088 - 0.742	0.192 - 0.662
L^2 -norm	Mean	0.116	0.132
	Median (IQR)	0.111 (0.089 - 0.137)	0.132 (0.123 - 0.141)
	Min - Max	0.027 - 0.359	0.066 - 0.237

Fig. 2 Histograms of L^1 and L^2 -norms in nonparametric density estimation by the proposed method for the mixture distribution of the density $0.3\phi_{2,1}(x) + 0.4\phi_{6,0.8}(x) + 0.3\phi_{10,2}(x)$.

Fig. 3 The L^1 -norm best-fit (left panel) and worst-fit (right panel) estimated densities by the proposed method in nonparametric inference for the mixture distribution of the density $0.3\phi_{2,1}(x) + 0.4\phi_{6,0.8}(x) + 0.3\phi_{10,2}(x)$. Solid curve is the true density. Dashed and dotted curves represent the estimated densities by the proposed and conventional method, respectively.