Scholars Journal of Physics, Mathematics and Statistics OPEN ACCESS

Abbreviated Key Title: Sch J Phys Math Stat
ISSN 2393-8056 (Print) | ISSN 2393-8064 (Online)
Journal homepage: http://saspjournals.com/sjpms

Intrinsic Variance Estimation for Gaussian Mixture Distribution

Junhao Guo*

College of Mathematics, Sichuan University, Chengdu, 610064, China

DOI: 10.21276/sjpms.2019.6.7.2 | **Received:** 02.05.2019 | **Accepted:** 09.05.2019 | **Published:** 22.07.2019
***Corresponding author:** Junhao Guo

Abstract **Review Article**

The intrinsic estimation for parameters are extensively studied in the literatures, which does not depend on the coordinate systems or model parametrization. In this paper, the intrinsic variance estimation for Gaussian mixture distribution is provided. The optimal estimators are proposed by minimizing the mean squared Rao distance and the mixture of symmetrized Kullback–Leibler divergence between normal distributions.

Keywords: Intrinsic estimation, Gaussian mixture distribution, Variance estimation, Rao distance, Kullback–Leibler divergence.

INTRODUCTION

Mean square error (MSE), as a classical metric, is extensively applied to parameter estimation. However, it is heavily dependent on the coordinate systems. For example, given two parameters of some specific distribution according to a one-to-one nonlinear reparametrization, the corresponding minimum mean square error estimators usually do not satisfy this mapping [1]. Therefore, the intrinsic version of MSE should be considered.

Although the loss function is an effective instrument for modeling, the choice of a suitable loss function still presents a concern. The selected loss function should enable us to measure the discrepancy between probability density functions. Entropy loss or Kullback–Leibler (KL) divergence, Hellinger distance and symmetrized KL divergence [2-5] were successively used to develop the parameter intrinsic estimation. As for intrinsic Bayesian estimation, the Rao distance was considered as the most proper loss [6]. The intrinsic estimators of Bernoulli distribution, multivariate normal distribution with known covariance matrix, normal distribution with known mean (a multiple of sample variance) [6] and multivariate normal distribution with zero mean (the same form as [6]) [7] have been provided.

In this paper, we obtain the intrinsic variance estimation for Gaussian mixture distribution. Like [6] and [7], the provided estimators are different scale numbers times sample variance. Firstly, we propose the definition of the intrinsic MSE. Then, the intrinsic estimators are provided by minimizing the mean squared Rao distance and the mixture of symmetrized KL divergence.

Intrinsic Variance Estimation

Set $\mathcal{X}$ is the sample space of the random variable X and $f\left(x \mid \theta\right)$ is the probability density function, where $\theta' = [\theta_1, \ldots, \theta_n]$ is continuous parameters in the parameter space Θ. Suppose that the probability density function satisfies the certain regularity conditions 8. Then, the information matrix is defined as

$$g_{ij}(\theta) = \mathbb{E}\left[\frac{\partial}{\theta_i}\ln f(x \mid \theta)\frac{\partial}{\theta_j}\ln f(x \mid \theta)\right], \quad i, j = 1, \ldots, n,$$

where $\mathbb{E}$ is the mathematical integration with regard to $f(x \mid \theta)$. Then, the Riemannian metric in the parameter space Θ is

$$ds = \sqrt{\sum_{i,j=1}^{n} g_{ij} d\theta_i d\theta_j}\,.$$

For two probability density functions $f\left(x \mid \theta_1\right)$ and $f\left(x \mid \theta_2\right)$, the Rao distance $\rho\left(\theta_1, \theta_2\right)$ is the geodesic distance with respect to the Riemannian metric.

As for two probability density functions $f\left(x \mid \theta_1\right)$ and $f\left(x \mid \theta_2\right)$, the symmetrized KL divergence is

$$J\left(f\left(x \mid \theta_1\right), f\left(x \mid \theta_2\right)\right) = \frac{1}{2}\left[\int_x f\left(x \mid \theta_1\right)\ln\frac{f\left(x \mid \theta_1\right)}{f\left(x \mid \theta_2\right)}dx + \int_x f\left(x \mid \theta_2\right)\ln\frac{f\left(x \mid \theta_2\right)}{f\left(x \mid \theta_1\right)}dx\right].$$

Next, we adopt the squared Rao distance or mixture of KL divergence as the loss of intrinsic MSE instead of square error.

Suppose that the probability density function of the random variable $X_1, \ldots, X_N$ is

$$\left(1 - \sum_{j=1}^{M}\varepsilon_j\right)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{x^2}{2\sigma^2}\right\} + \sum_{j=1}^{M}\varepsilon_j\frac{1}{\sqrt{2\pi c_j\sigma^2}}\exp\left\{-\frac{x^2}{2c_j\sigma^2}\right\},$$

where $\varepsilon_j, j = 1, \ldots, M$ are mixing coefficients and $c_j, j = 1, \ldots, M$ are positive numbers. Let

$$\sigma_X^2 = a\sigma^2 = \left(1 - \sum_{j=1}^{M}\varepsilon_j + \sum_{j=1}^{M}\varepsilon_j c_j\right)\sigma^2,$$

where σ_X^2 is the variance of random variable $X_1, \ldots, X_N$.

Then, we propose the intrinsic variance estimator $\alpha S = \alpha\sum_{i=1}^{N} X_i^2$, where α is a scalar.

Theorem 1. By minimizing the mean squared Rao distance $\rho\left(\sigma_X^2, \alpha S\right)$, we obtain the intrinsic variance estimator

$$\bar{\alpha} S = \exp\left\{-\mathbb{E}\left[\ln \frac{S}{\sigma_X^2}\right]\right\} S.$$

Proof. Obviously, the random variable $X_1, \dots, X_N$ follow elliptical distribution. From 9, we have

$$\rho^2\left(\sigma_X^2, \alpha S\right) = \left(\sigma_X^2\right)^2\left(3b_h - \frac{1}{4}\right)\left(\ln\alpha + \ln\frac{S}{\sigma_X^2}\right)^2,$$

where $b_h = \dfrac{\mathbb{E}(u^2 w^2)}{3}$, $u = \dfrac{X^2}{\sigma^2}$, $w = \dfrac{\partial \ln \sigma f}{\partial u}$.

We can exchange the expectation and differentiation. Thus,

$$\frac{d}{dc}\mathbb{E}\left[\rho^2\left(\sigma_X^2, \alpha S\right)\right] = \frac{d}{dc}\mathbb{E}\left[\left(\sigma_X^2\right)^2\left(3b_h - \frac{1}{4}\right)\left(c + \ln\frac{S}{\sigma_X^2}\right)^2\right]$$

$$= \mathbb{E}\left[\frac{d}{dc}\left(\sigma_X^2\right)^2\left(3b_h - \frac{1}{4}\right)\left(c + \ln\frac{S}{\sigma_X^2}\right)^2\right]$$

$$= 2\left(\sigma_X^2\right)^2\left(3b_h - \frac{1}{4}\right)\left(c + \mathbb{E}\left[\ln\frac{S}{\sigma_X^2}\right]\right),$$

where $c = \ln\alpha$.

$\bar{c} = -\mathbb{E}\left[\ln\dfrac{S}{\sigma_X^2}\right]$ attains minimum value of mean squared Rao distance if $3b_h - \dfrac{1}{4} \neq 0$. When $3b_h - \dfrac{1}{4} = 0$, $\mathbb{E}\left[\rho^2\left(\sigma_X^2, \alpha S\right)\right]$ is constant for any c. Therefore, $\bar{c} = -\mathbb{E}\left[\ln\dfrac{S}{\sigma_X^2}\right]$ minimizes the mean squared Rao distance.

Thus, the scalar $\bar{\alpha}$ of intrinsic variance estimator is

$$\bar{\alpha} = \exp\left(\bar{c}\right) = \exp\left\{-\mathbb{E}\left[\ln\frac{S}{\sigma_X^2}\right]\right\}.$$

There is no analytical expression for KL divergence about Gaussian mixture distribution. Thus, the discrepancy between Gaussian mixture distributions is measured in the sense of the mixture of symmetrized KL divergence between normal distributions in this paper.

$$\left(1 - \sum_{j=1}^{M}\epsilon_j\right) J\left(N\left(0, \sigma_1^2\right), N\left(0, \sigma_2^2\right)\right) + \sum_{j=1}^{M}\epsilon_j J\left(N\left(0, c_j\sigma_1^2\right), N\left(0, c_j\sigma_2^2\right)\right), \quad (1)$$

where $N\left(0,\sigma_1^2\right)$ is normal distribution with mean 0 and variance σ_1^2.

Lemma 1. By minimizing the expectation of metric (1) between variance and estimator, we obtain the intrinsic variance estimator

$$\bar{\alpha}S = \sqrt{\frac{\mathbb{E}\left[\frac{\sigma_X^2}{S}\right]}{\mathbb{E}\left[\frac{S}{\sigma_X^2}\right]}}S.$$

Proof. From **Error! Reference source not found.** and metric (1), the mixture of symmetrized KL divergence between variance σ_X^2 and estimator αS is

$$\left(1-\sum_{j=1}^{M}\epsilon_j\right)\frac{1}{2}\left[\frac{\frac{\alpha S}{a}}{\sigma^2}+\frac{\sigma^2}{\frac{\alpha S}{a}}-2\right]+\sum_{j=1}^{M}\epsilon_j\frac{1}{2}\left[\frac{\frac{c_j\alpha S}{a}}{c_j\sigma^2}+\frac{c_j\sigma^2}{\frac{c_j\alpha S}{a}}-2\right]=\frac{1}{2}\left[\frac{\alpha S}{a\sigma^2}+\frac{a\sigma^2}{\alpha S}-2\right].$$

Then, let

$$\frac{d}{d\alpha}\mathbb{E}\left[\frac{1}{2}\frac{\alpha S}{a\sigma^2}+\frac{1}{2}\frac{a\sigma^2}{\alpha S}-1\right]=0.$$

We can exchange the expectation and differentiation. Then,

$$\bar{\alpha} = \sqrt{\frac{\mathbb{E}\left[\frac{a\sigma^2}{S}\right]}{\mathbb{E}\left[\frac{S}{a\sigma^2}\right]}}.$$

Lemma 2. If the joint probability density function of the random variable $X_1,\ldots,X_N$ is

$$\left(1-\sum_{j=1}^{M}\varepsilon_j\right)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\sum_{i=1}^{N}\frac{x_i^2}{2\sigma^2}\right\}+\sum_{j=1}^{M}\varepsilon_j\frac{1}{\sqrt{2\pi c_j\sigma^2}}\exp\left\{-\sum_{i=1}^{N}\frac{x_i^2}{2c_j\sigma^2}\right\}.$$

Then, the intrinsic variance estimator in the sense of minimizing mean squared $\rho\left(\sigma_X^2,\alpha S\right)$ is

$$\bar{\alpha}S = \frac{\exp\left(-\psi\left(\frac{N}{2}\right)\right)a}{2\prod_{j=1}^{M}c_j^{\epsilon_j}}S,$$

where $\psi\left(x\right)$ is digamma function.

Proof. From lemma 2.1 in 9, the probability density function of the random variable S is

$$\left(1-\sum_{j=1}^{M}\epsilon_j\right)\frac{1}{2^{\frac{N}{2}}\Gamma\left(\frac{N}{2}\right)}\left(\sigma^2\right)^{-\frac{N}{2}}s^{\frac{N-2}{2}}\exp\left\{-\frac{s}{2\sigma^2}\right\}$$
$$+\sum_{j=1}^{M}\epsilon_j\frac{1}{2^{\frac{N}{2}}\Gamma\left(\frac{N}{2}\right)}\left(c_j\sigma^2\right)^{-\frac{N}{2}}s^{\frac{N-2}{2}}\exp\left\{-\frac{s}{2c_j\sigma^2}\right\}.$$

Then, the probability density function of the random variable $S_1=\dfrac{S}{a\sigma^2}$ is

$$\left(1-\sum_{j=1}^{M}\epsilon_j\right)\frac{1}{2^{\frac{N}{2}}\Gamma\left(\frac{N}{2}\right)}\left(\frac{1}{a}\right)^{-\frac{N}{2}}s_1^{\frac{N-2}{2}}\exp\left\{-\frac{as_1}{2}\right\}$$
$$+\sum_{j=1}^{M}\epsilon_j\frac{1}{2^{\frac{N}{2}}\Gamma\left(\frac{N}{2}\right)}\left(\frac{c_j}{a}\right)^{-\frac{N}{2}}s_1^{\frac{N-2}{2}}\exp\left\{-\frac{as_1}{2c_j}\right\}.$$

Thus, the expectation of the random variable $\ln\dfrac{S}{a\sigma^2}$ is

$$\left(1-\sum_{j=1}^{M}\epsilon_j\right)\left[\psi\left(\frac{N}{2}\right)+\ln 2+\ln\frac{1}{a}\right]+\sum_{j=1}^{M}\epsilon_j\left[\psi\left(\frac{N}{2}\right)+\ln 2+\ln\frac{c_j}{a}\right]$$
$$=\psi\left(\frac{N}{2}\right)+\ln 2-\ln a+\sum_{j=1}^{M}\epsilon_j\ln c_j.$$

Then, the scale factor of intrinsic variance estimator is

$$\bar{\alpha}=\frac{\exp\left(-\psi\left(\frac{N}{2}\right)\right)a}{2\prod_{j=1}^{M}c_j^{\epsilon_j}}.$$

CONCLUSION

For Gaussian mixture distribution with zero mean, we obtain intrinsic variance estimation by minimizing the mean squared Rao distance and the mixture of symmetrized KL divergence, which is a scalar times sample variance. Finally, the intrinsic estimator is proposed when the joint probability density function of the N random variable is mixture of multivariate normal distribution with a diagonal covariance matrix.

REFERENCES

1. Oller, J. M. (1993). On an intrinsic analysis of statistical estimation. In *Multivariate Analysis: Future Directions 2* (pp. 421-437). North-Holland.

2. Bernardo, J. M., & Juárez, M. A. (2003). Intrinsic estimation. In Bernardo, J. M., Bayarri, M. J. Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., & West, M. editors, Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting, pages 456–476. Oxford University Press, New York, NY, USA.

3. Bernardo, J. M. (2007). Intrinsic point estimation of the normal variance. *Bayesian statistics and its applications*, 110-121.

4. Bernardo, J. M. (2007). Objective Bayesian point and region estimation in location-scale models. *Statistics and Operations Research Transactions*, 31(1): 3-44

5. Bernardo, J. M. (2011). Integrated objective Bayesian estimation and hypothesis testing. *Bayesian statistics*, *9*, 1-68.

6. García, G., & Oller, J. M. (2007). What does intrinsic mean in statistical estimation? [J]. SORT: Statistics and *Operations Research Transactions*, 30(2):125–146.

7. Rong, Y., Tang, M., & Zhou, J. (2017). Intrinsic Losses Based on Information Geometry and Their Applications. *Entropy*, *19*(8), 405.

8. Amari, S., & Nagaoka, H. (2000). Methods of Information Geometry. American Mathematical Society: Providence, RI, USA.

9. Berkane, M., Oden, K., & Bentler, P. M. (1997). Geodesic estimation in elliptical distributions. *Journal of Multivariate Analysis*, 63(1):35-46.

10. Sutradhar, B. C., & Ali, M. M. (1989). A generalization of the Wishart distribution for the elliptical model and its moments for the multivariate t model. *Journal of Multivariate Analysis*, 29(1):155–162.