Written Report

August 27, 2024

The concept of Minimum Hellinger Distance (MHD) is grounded in the broader context of Hellinger Distance, a measure of the divergence between two probability distributions. The Hellinger Distance is mathematically defined as:

$$H(p,q) = \frac{1}{\sqrt{2}} \left(\int_X \left(\sqrt{p(x)} - \sqrt{q(x)} \right)^2 dx \right)^{\frac{1}{2}}$$

where $p(x)$ and $q(x)$ are the two probability distributions over the domain X(Lucien,2012).

Building on this concept, Minimum Hellinger Distance Estimation (MHDE) focuses on estimating parameters by minimizing the Hellinger distance between a parametric model and a nonparametric estimation of the distribution. Although many studies emphasize the estimation process itself, they provide the necessary theoretical foundation for understanding MHD as a concept. Specifically, MHD is defined as the smallest possible Hellinger distance between the empirical distribution and the theoretical distribution within a specified family of distributions (Beran, 1977; Lindsay, 1994). In essence, MHD captures the minimum divergence between these distributions, providing a robust measure of model fit.

The context of Minimum Hellinger Distance Estimation (MHDE) is particularly intriguing due to its combination of efficiency and robustness, as highlighted by Beran (1977). These properties make MHDE a powerful tool in statistical inference, especially in scenarios where traditional methods may be vulnerable to model misspecification or outliers. The versatility of MHDE is further demonstrated by its successful application across various statistical models, notably in the analysis of mixture models, where it has been shown to yield more reliable and stable parameter estimates compared to other methods.

A semi-parametric location-shifted mixture model is a versatile statistical tool used to analyze complex data structures, particularly when the underlying distribution is not fully known or when the data exhibits significant heterogeneity. This model blends parametric and non-parametric approaches by assuming that the data is generated from a mixture of distributions, where each component can have a location parameter that shifts according to a specified pattern

(Hall & Zhou, 2003).

In a standard mixture model, data is assumed to come from a mixture of several sub-populations, each described by its own probability distribution. In contrast, a location-shifted mixture model introduces a parameter that allows each component to shift its location, typically affecting the mean or median of the distribution. This location shift enables the model to more accurately capture the underlying structure of the data, especially in cases where different sub-populations exhibit different central tendencies (Jiang & Tanner, 1999).

The semi-parametric aspect of the model refers to its flexibility: while the location shifts are parameterized, the component distributions can be modeled non-parametrically. This feature allows the model to handle complex distributional shapes that may not be well-represented by standard parametric forms, leading to a more robust and accurate analysis (Benaglia, Chauveau, & Hunter, 2009).
The location shifted mixture model can be expressed as:

$$g(x) = \sum_{i=1}^{k} \pi_i f(x - \mu_i)$$

where:

- k is the number of mixture components.
- π_i are the mixing proportions, with $\sum_{i=1}^{k} \pi_i = 1$.
- $f(x)$ is the base distribution, which might be parametric or non-parametric.
- μ_i are the location parameters for each component, shifting the base distribution $f(x)$.

In this paper, we investigate the Minimum Hellinger Distance Estimator for a semi-parametric location shifted mixture model for three components(k=3).

$$g(x) = \sum_{i=1}^{k} \pi_i f(x - \mu_i)$$

where f is an even function(Xiaofan et al,2018)

MHDE of parameter θ

Given a random sample $X_1, \ldots, X_n$ from the mixture , we can estimate $h_{\theta,f}$ nonparametrically by

$$\hat{h}(x) = \frac{1}{nb_n} \sum_{j=1}^{n} K\left(\frac{x - X_j}{b_n}\right),$$

where K is a kernel density function and b_n is a sequence of bandwidths satisfying $b_n \to 0$ and $nb_n \to \infty$ as $n \to \infty$. With this choice of $\hat{h}$ and a consistent estimate $\hat{\theta}^{(0)}$ for $\hat{f}$ in (10), we propose an MHDE of θ given by

$$\hat{\theta} = \arg\min_{\theta \in \Theta_0} \left\| h_{\theta,\hat{f}}^{1/2} - \hat{h}^{1/2} \right\|,$$

where Θ_0.

By Wu and Karunamuni [**?**], the MHDE $\hat{\theta}$ is consistent and asymptotically normally distributed. To relax the dependence of $\hat{f}$ on the initial estimate $\hat{\theta}^{(0)}$ and increase the accuracy of $\hat{f}$, we calculate $\hat{\theta}$ iteratively. In each iteration, we update $\hat{\theta}^{(0)}$ with the current MHDE to get an updated $\hat{f}$ and consequently an updated $\hat{\theta}$ until convergence.

Following this idea, we propose the following iterative algorithm to calculate the MHDE $\hat{\theta}$:

Suppose the k-th estimates of θ and f are respectively $\hat{\theta}^{(k)} = (\hat{\pi}^{(k)}, \hat{\mu}_1^{(k)}, \hat{\mu}_2^{(k)})^\top$ and $\hat{f}^{(k)}$, and $\hat{h}$ is the kernel density estimator of $h_{\theta,f}$ given in (11).

Step 1. With current estimates $\hat{\pi}^{(k)}, \hat{\mu}_1^{(k)}$, and $\hat{\mu}_2^{(k)}$, calculate

$$\tilde{f}_s^{(k+1)}(x) = \frac{\tilde{f}^{(k+1)}(-x) + \tilde{f}^{(k+1)}(x)}{2},$$

and then update the estimate of f by

$$\hat{f}^{(k+1)}(x) = \frac{\tilde{f}_s^{(k+1)}(x) I\{\tilde{f}_s^{(k+1)}(x) \geq 0\}}{\int \tilde{f}_s^{(k+1)}(x) I\{\tilde{f}_s^{(k+1)}(x) \geq 0\} dx}.$$

Step 2. With the updated $\hat{f}^{(k+1)}$, find the updated $\hat{\pi}^{(k+1)}, \hat{\mu}_1^{(k+1)}$, and $\hat{\mu}_2^{(k+1)}$ by minimizing

$$\left\| \left(\hat{\pi}^{(k+1)} \hat{f}^{(k+1)}(\cdot - \hat{\mu}_1^{(k+1)}) + (1 - \hat{\pi}^{(k+1)}) \hat{f}^{(k+1)}(\cdot - \hat{\mu}_2^{(k+1)}) \right)^{1/2} - \hat{h}^{1/2}(\cdot) \right\|.$$

Then repeat Steps 1 and 2 until $\hat{\theta}$ converges.

We use Matlab software for our numerical studies. In Step 2, we use the function 'fminsearch' in Matlab to find the optimizer $\hat{\theta}$. The proposed MHD method not only provides a good estimation of the parameter θ but also improves the estimation of the nonparametric nuisance parameter f. With an initial estimate of θ, we can only give a crude estimate of f as given in (10). Once we obtain the converged MHDE $\hat{\theta}$, we can plug it into (13) to get a more accurate estimate of the nonparametric nuisance parameter f.

[6]. [7]. [1] [5] [4] [2] [3]

References

[1] Rudolf Beran. Minimum hellinger distance estimates for parametric models. *The annals of Statistics*, pages 445–463, 1977.

[2] Peter Hall and Xiao-Hua Zhou. Nonparametric estimation of component distributions in a multivariate mixture. *The annals of statistics*, 31(1):201–224, 2003.

[3] Wenxin Jiang and Martin A Tanner. Hierarchical mixtures-of-experts for exponential family regression models: approximation and maximum likelihood estimation. *The Annals of Statistics*, 27(3):987–1011, 1999.

[4] Lucien Le Cam. *Asymptotic methods in statistical decision theory*. Springer Science & Business Media, 2012.

[5] Bruce G Lindsay. Efficiency versus robustness: the case for minimum hellinger distance and related methods. *The annals of statistics*, 22(2):1081–1114, 1994.

[6] JH Westcott. The parameter estimation problem. *IFAC Proceedings Volumes*, 1(1):789–797, 1960.

[7] Jingjing Wu and Xiaofan Zhou. Minimum hellinger distance estimation for a semiparametric location-shifted mixture model. *Journal of Statistical Computation and Simulation*, 88(13):2507–2527, 2018.