Written Report

August 27, 2024

Most papers focus on Minimum Hellinger Distance Estimation (MHDE), where the estimation is achieved by minimizing the Hellinger distance between a parametric model and a nonparametric estimation of the distribution. Although these papers emphasize the estimation process rather than isolating the concept of Minimum Hellinger Distance (MHD) itself, they do provide the necessary theoretical foundation for understanding MHD. Specifically, MHD is defined as the smallest possible Hellinger distance between the empirical distribution and the theoretical distribution within a specified family of distributions. In essence, MHD captures the minimum divergence between these distributions (Beran, 1977; Lindsay, 1994).

Hellinger Distance which measures the distance between two probability distributions and is given by:

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

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

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)

[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.