Minimum Hellinger Distance Estimation for a Three-component location shifted Mixture Model

Minimum Hellinger Distance (MHD) is based on Hellinger Distance, a measure of divergence between two probability distributions which is defined 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 [6].

(MHDE) estimates parameters by minimizing the Hellinger distance between a parametric model and a nonparametric estimate of the distribution. MHD is the smallest Hellinger distance between the empirical and theoretical distributions within a specified family, offering a robust measure of model fit [2][7].

Minimum Hellinger Distance Estimation (MHDE) is valued for its efficiency and robustness, making it a powerful tool in statistical inference, especially when dealing with model misspecification or outliers. Its versatility is evident in various applications, particularly in mixture models, where it provides more reliable and stable parameter estimates compared to other methods.

A semi-parametric location-shifted mixture model is used to analyze complex data with unknown distributions or significant heterogeneity. It combines parametric and non-parametric methods, assuming the data comes from a mixture of distributions, with each component having a location parameter that shifts according to a specific pattern [3]

.

In a standard mixture model, data comes from multiple sub-populations, each with its own probability distribution. A location-shifted mixture model introduces a parameter to shift each component's location, typically affecting the mean or median, allowing for better capture of the data's structure. The semi-parametric aspect provides flexibility: location shifts are parameterized, while component distributions can be modeled non-parametrically, enabling the model to handle complex distributional shapes for more robust analysis [4][1].

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 [8].

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 [5], 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

$$\tilde{f}^{(k+1)}(x) = \begin{cases} \frac{1}{1-\hat{\pi}^{(k)}} \sum_{i\geq 0} \left(\frac{-\hat{\pi}^{(k)}}{1-\hat{\pi}^{(k)}}\right)^i \hat{h}\left(x + \hat{\mu}_2^{(k)} + i(\hat{\mu}_2^{(k)} - \hat{\mu}_1^{(k)})\right), & \text{if } \hat{\pi}^{(k)} < 0.5, \\ \frac{1}{\hat{\pi}^{(k)}} \sum_{i\geq 0} \left(\frac{\hat{\pi}^{(k)}-1}{\hat{\pi}^{(k)}}\right)^i \hat{h}\left(x + \hat{\mu}_1^{(k)} + i(\hat{\mu}_1^{(k)} - \hat{\mu}_2^{(k)})\right), & \text{if } 0.5 < \hat{\pi}^{(k)} < 1, \end{cases}$$

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

and then the updated 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 . Once we obtain the converged MHDE $\hat{\theta}$, we can plug it to get a more accurate estimate of the nonparametric nuisance parameter f.

References

[1] Tatiana Benaglia, Didier Chauveau, and David R Hunter. An em-like algorithm for semi-and nonparametric estimation in multivariate mixtures. *Journal of Computational and Graphical Statistics*, 18(2):505–526, 2009.

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

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

[4] 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.

[5] RJ Karunamuni and J Wu. Minimum hellinger distance estimation in a nonparametric mixture model. *Journal of Statistical Planning and Inference*, 139(3):1118–1133, 2009.

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

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

[8] 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.