



Article

Comparative Evaluation of Nonparametric Density Estimators for Gaussian Mixture Models with Clustering Support

Tomas Ruzgas ^{1,*} , Gintaras Stankevičius ¹, Birutė Narijauskaitė ¹  and Jurgita Arnastauskaitė Zencevičienė ²¹ Department of Applied Mathematics, Kaunas University of Technology, LT-51368 Kaunas, Lithuania; gintaras.stankevicius@ktu.edu (G.S.); birute.narijauskaite@ktu.lt (B.N.)² Department of Computer Sciences, Kaunas University of Technology, LT-51368 Kaunas, Lithuania; jurgita.zenceviciene@ktu.lt

* Correspondence: tomas.ruzgas@ktu.lt

Abstract

The article investigates the accuracy of nonparametric univariate density estimation methods applied to various Gaussian mixture models. A comprehensive comparative analysis is performed for four popular estimation approaches: adaptive kernel density estimation, projection pursuit, log-spline estimation, and wavelet-based estimation. The study is extended with modified versions of these methods, where the sample is first clustered using the EM algorithm based on Gaussian mixture components prior to density estimation. Estimation accuracy is quantitatively evaluated using MAE and MAPE criteria, with simulation experiments conducted over 100,000 replications for various sample sizes. The results show that estimation accuracy strongly depends on the density structure, sample size, and degree of component overlap. Clustering before density estimation significantly improves accuracy for multimodal and asymmetric densities. Although no formal statistical tests are conducted, the performance improvement is validated through non-overlapping confidence intervals obtained from 100,000 simulation replications. In addition, several decision-making systems are compared for automatically selecting the most appropriate estimation method based on the sample's statistical features. Among the tested systems, kernel discriminant analysis yielded the lowest error rates, while neural networks and hybrid methods showed competitive but more variable performance depending on the evaluation criterion. The findings highlight the importance of using structurally adaptive estimators and automation of method selection in nonparametric statistics. The article concludes with recommendations for method selection based on sample characteristics and outlines future research directions, including extensions to multivariate settings and real-time decision-making systems.

Keywords: univariate probability density; nonparametric estimation; Monte Carlo method**MSC:** 62G05; 62G07; 62G30

Academic Editor: Hari Mohan Srivastava

Received: 18 June 2025

Revised: 14 July 2025

Accepted: 21 July 2025

Published: 23 July 2025

Citation: Ruzgas, T.; Stankevičius, G.; Narijauskaitė, B.; Arnastauskaitė Zencevičienė, J. Comparative Evaluation of Nonparametric Density Estimators for Gaussian Mixture Models with Clustering Support. *Axioms* **2025**, *14*, 551. <https://doi.org/10.3390/axioms14080551>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Nonparametric density estimation methods are playing an increasingly important role in applied research. However, only a limited number of studies perform comprehensive comparisons of various statistical approaches representing popular classes of estimators. Nonparametric density estimation is widely used in data mining and in estimating or validating parameters of structural models, particularly when no assumptions are made about the true density. The choice of estimation method can lead to significantly different

results, especially when the form of the underlying density is difficult to determine. Although a wide variety of density estimation techniques is known in modern data analysis, in practice it is not easy to select an effective procedure when the data distribution is multimodal and the sample size is small. Kernel estimators [1–3] are the most common variety encountered in the literature. Other methods, particularly those based on polynomials, are also frequently applied—including those using Bernstein polynomials [4], Hermite series [5], projection pursuit [6], wavelet expansions [7], and B-splines [8]. Additionally, polynomials are often incorporated into kernel-based frameworks, such as constrained local polynomial estimators [9], local polynomial approximation methods [10], and linear combinations of piecewise polynomial density functions [11].

The main objective of this article is to compare several density estimators, focusing on univariate multimodal densities. Although statistical analysis provides many competitive nonparametric density estimation methods, the vast majority of comparisons are limited to constant bandwidth kernel estimators. Typically, comparative studies of nonparametric density estimation focus solely on algorithms with fixed kernels, while very few studies perform broader comparisons of fundamentally different estimation approaches. For example, Scott and Factor [12] compared two constant kernel estimators and an orthogonal series estimator. Hwang, Lay, and Lippman [13] compared fixed and adaptive kernel methods, projection pursuit, and radial basis function methods in the multivariate case. Fenton and Gallant [14] examined the application of Hermite series and Silverman’s rule-of-thumb kernel methods for estimating normal mixture densities. Fadda, Slezak, and Bijaoui [15] compared adaptive kernel estimators, maximum penalized likelihood methods, and their own wavelet-based approach. Egecioğlu and Srinivasan [16] compared fixed kernel and cosine-based estimators. Takada [17] analyzed fixed and adaptive kernel methods, Gallant and Nychka’s [18] Hermite-based approach, and the log-spline method. Some studies also explore hybrid approaches. For example, Ćwik and Koronacki [19] combined two types of fixed kernel methods, projection pursuit, and Gaussian component models using the EM algorithm. Couvreur and Couvreur [20] compared wavelet-based and histogram-based estimators in combination with iterative hidden Markov models, which function similarly to the EM algorithm.

A number of statistical models, such as multiple regression, discriminant analysis, logistic regression, and profit analysis, have been proposed to enable clearer conclusions from comparative studies. For example, Gallinari et al. established a link between discriminant analysis and multilayer perceptrons [21], while Manel et al. compared the performance of multiple discriminant analysis, logistic regression, and artificial neural networks [22]. Neural networks and statistical methods can complement each other in providing insights into the phenomena being studied. In fact, many researchers advocate using multiple methods for decision-making [23]. We apply these decision-making methods to select the most appropriate density estimator for approximating an unknown probability distribution. Specifically, these decision-making systems rely on input features such as sample size, estimated skewness, kurtosis, number of clusters (based on EM clustering), and outlier proportions. The systems are trained to select the density estimator that minimizes prediction error based on these statistical characteristics.

Having discussed the main nonparametric density estimation methods, it is important to highlight their practical significance and applicability across various fields. Density estimation plays a crucial role not only in theoretical statistical research but also in solving real-world problems such as classification, clustering, or anomaly detection. Improved density estimation can significantly enhance practical applications such as anomaly detection, risk assessment, or signal classification. For instance, in the context of financial transaction monitoring, multimodal and skewed density estimation enables better identification of

suspicious behavior patterns, supporting fraud detection systems. Therefore, the following paragraphs review selected studies in which density estimation was successfully applied to practical tasks.

Traffic Safety and Environment: Ge et al. [24] employ AKDE to find “black spots” (high-accident road segments) in traffic planning. Compared to normal KDE, their methodology identifies more accidents per mile, allowing for more focused safety measures. Amador Luna et al. [25] used KDE to analyze high-density seismic data from the February 2023 earthquakes in Turkey and Syria. According to their analysis of almost 40,000 occurrences, KDE consistently produces geological interpretations by accurately defining the main fault structures and depth layers. The analysis was performed using QGIS software (version 3.18), which enabled kernel-based spatial visualization. Even with noisy measurements, point-density analyses in seismic or geographical data can uncover underlying patterns (faults, resource deposits, ecological ranges), illustrating the value of KDE in geoscience. Anomaly identification and surveillance rely heavily on GLR and discriminant approaches for signal and video anomaly detection. Puranik et al. [26] address the use of GLRT to detect adversarial (anomalous) disturbances in sensor data. Hybrid methods (GLR + NN or KDE + NN) are being investigated in network security to enhance the identification of unexpected behavior. Similarly, LDA/KDA are quick classifiers in embedded systems that detect anomalies in real time. While no explicit citation is provided, these approaches have been cited in recent studies on anomaly identification in hyperspectral images and industrial monitoring. In the article [27], four different estimators are analyzed: standard kernel estimator, Bernstein estimator, Guan’s estimator, and author proposed estimator. The suggested semi-parametric estimation strategy is based on the shrinkage combination between the Gaussian mixture model and the Bernstein density estimators, employing the EM algorithm for parameter estimation. With real world data, the proposed estimator outperforms other estimators, especially near the boundaries.

Article [28] presented a comparative analysis of several popular nonparametric estimators and found that in most multivariate cases with well-separated mixture components, the Friedman procedure was the most effective. Only when the sample size was small did the kernel estimator produce slightly more accurate results. However, when components overlapped, other estimators exhibited notable advantages. The present work continues the line of inquiry introduced in that article, focusing on identifying techniques for selecting the most suitable density estimator based on sample properties. At the same time, this study extends the research both in terms of the methods applied and the diversity of the underlying distributions.

This article is structured as follows: Section 2 presents the investigated density estimation methods, provides an overview of the EM algorithm used for sample clustering, describes the decision support techniques applied, and outlines the extended simulation study design. Section 3 contains the modeling results, while Section 4 offers a detailed discussion and interpretation of these findings. The accuracy of the estimators is illustrated using heatmap visualizations, which are provided in the Appendix A.

2. Materials and Methods

2.1. Investigated Density Estimation Algorithms

The comparative analysis of estimation accuracy was performed on four different types of statistics representing popular classes of estimators that have been previously studied by other researchers. Using the Monte Carlo method, this study investigated the following statistical estimators of probability density:

1. The adaptive kernel density estimator (AKDE) proposed by Silverman [3], which employs variable bandwidths for different observations;

2. The projection pursuit density estimator (PPDE), based on sequential Gaussianization of projections, proposed by Friedman [6,29];
3. The log-spline density estimator (LSDE) introduced by Kooperberg and Stone [8], which approximates the logarithm of the target density using a sum of cubic B-splines;
4. The nonlinear, threshold-based wavelet estimator (WEDE), whose properties were explored by Donoho et al. [30] and by Hall and Patil [31].

Before applying these methods, the sample is standardized. Let $X(1), \dots, X(n)$ be independently observed, identically distributed random variables with an unknown density $f(x)$. The algorithms for estimating $\hat{f}(x)$ are described below.

AKDE algorithm. The kernel density estimator with locally adaptive bandwidth takes the form

$$\hat{f}(x) = \frac{1}{n} \sum_{t=1}^n \frac{1}{h_t} K\left(\frac{x - X(t)}{h_t}\right) \quad (1)$$

As in [13], the kernel function K used here is the standard normal density φ , and the bandwidth is defined as [1]:

$$h_j = h \left(\tilde{f}(X(j))/q \right)^{-v},$$

where $h = (4/(3n))^{1/5}$. Here $\tilde{f}(\cdot)$ is the kernel estimator defined by (1), using a fixed bandwidth h and $q = \exp\left\{\frac{1}{n} \log \sum_{j=1}^n \tilde{f}(X(j))\right\}$, and v is the sensitivity parameter. Following [13], $v \in \{0.2, 0.4, 0.6, 0.8\}$, and the optimal value is selected via cross-validation. The sensitivity parameter v was selected via cross-validation on a representative subset of simulations and then applied globally across all experiments. Cross-validation was performed on a representative subset of simulations to identify a robust global parameter value. This helped avoid overfitting to specific simulation runs. The choice of the sensitivity parameter v influences the bias–variance tradeoff: higher values of v reduce bias but increase variance, while lower values result in smoother, potentially more biased estimates.

PPDE algorithm. J.H. Friedman [29] proposed a recursive algorithm for multivariate density estimation based on identifying one-dimensional projections whose distributions deviate most from the standard normal and Gaussianizing them sequentially. Let X be a standardized random vector (with mean 0 and identity covariance matrix) with unknown density f . After each iteration, X is transformed: $Z^{(k)} = Q_k(X)$, $k = 1, 2, \dots$ so that the projection of $Z^{(k)}$ onto direction τ , in which the density g_k deviates most from the normal density φ , has distribution Φ , while the projection onto the orthogonal subspace remains unchanged. Friedman showed [29] that as $k \rightarrow +\infty$, $Z^{(k)}$ converges in distribution to the standard multivariate normal. Hence, for sufficiently large M ,

$$f(z) \cong \varphi(z^{(M)}) \prod_{k=1}^M \frac{g_k(\tau'(k)z^{(k-1)})}{\varphi(\tau'(k)z^{(k)})}, \quad (2)$$

where $z^{(k)} = Q_k(x)$. Replacing (2) the unknown densities g_k on the right-hand side with their estimators gives the PPDE.

For univariate data, $\tau = 1$ and $M = 1$. The density g_k is estimated using a projection estimator in the Legendre polynomial basis. Let ξ_1, \dots, ξ_n be i.i.d. variables with density $g(u)$. After transformation $\eta_t = 2\Phi(\xi_t) - 1$, $y = 2\Phi(u) - 1$, we obtain random variables η_1, \dots, η_n with density $g^*(y) = \frac{g(u)}{2\varphi(u)}$, supported on $[-1, 1]$. Expanding $g^*(y)$ in the Legendre basis $\{\psi_j\}_{j=0}^{+\infty}$,

$$g^*(y) = \sum_{j=0}^{+\infty} b_j \psi_j(y)$$

and replacing the coefficients $b_j = \left(j + \frac{1}{2}\right) E\psi_j(\eta_t)$ with empirical estimates, we obtain

$$\hat{g}(y) = \varphi(y) \sum_{j=0}^s \frac{2j+1}{n} \sum_{t=1}^n \psi_j(\eta_t) \psi_j(y), \quad (3)$$

Following the recommendation in [13], the order of expansion (3) is limited to $s \leq 6$. The expansion order was limited to $s \leq 6$, following guidance from prior studies and supported by our preliminary simulations, which showed that higher orders led to instability without clear gains in accuracy.

LSDE algorithm. The log-spline estimator approximates the logarithm of a multivariate density using a sum of spline basis functions:

$$\hat{f}(x) = \exp \left\{ \sum_{j=1}^s \beta_j B_j(x) - C(\beta) \right\}$$

where B_1, \dots, B_s are spline basis functions, $\beta = (\beta_1, \dots, \beta_s)$ is a coefficient vector, and $C(\beta)$ is a normalizing constant. Kooperberg and Stone use cubic B-splines, selecting the knot locations via the AIC criterion [32], and estimating coefficients via maximum likelihood. The log-likelihood function is given by $\uparrow(\beta) = \sum_t \sum_j \beta_j B_j(x) - nC(\beta)$. The Akaike Information Criterion (AIC) is calculated as $AIC = -2\uparrow(\beta) + ak$, where k is the number of free parameters, and it is recommended to use $a = \log n$. The software used in [33] for computing the LSDE was also employed in this study. Although the LSDE was implemented using an existing R package without modifying the estimation function itself, we developed an automated procedure to integrate R-based estimation into the SAS simulation workflow. The implementation was carried out using R version 2.7.1 and SAS version 9.4 TS Level 1M4.

WEDE algorithm. A formal wavelet expansion of the density function is given by

$$f \cong \sum_{k \in \mathbb{Z}} c_{Lk} \phi_{Lk}(x) + \sum_{j \in \mathbb{J}_L} \sum_{k \in \mathbb{Z}} d_{jk} \psi_{jk}(x), \quad (4)$$

where $\mathbb{J}_L = \{m \in \mathbb{Z} : m \geq L\}$, and the coefficients are

$$c_{Lk} = \int_{-\infty}^{+\infty} \phi_{Lk}(x) f(x) dx \text{ and } d_{Lk} = \int_{-\infty}^{+\infty} \psi_{Lk}(x) f(x) dx.$$

The scaled and shifted basis functions are defined by

$$\phi_{jk}(x) = p_{j\tau}^{1/2} \phi(p_{j\tau}x - k) \text{ ir } \psi_{jk}(x) = p_{j\tau}^{1/2} \psi(p_{j\tau}x - k), \quad (5)$$

with tuning parameter $\tau \in \left(\frac{1}{2}, 1\right]$ and resolution level $p_{j\tau} = \tau 2^j$. Using the orthonormal basis properties, the coefficients are

$$\hat{c}_{jk} = \frac{1}{n} \sum_{t=1}^n \phi_{jk}(X(t)), \quad (6)$$

$$\hat{d}_{jk} = \frac{1}{n} \sum_{t=1}^n \psi_{jk}(X(t)), \quad (7)$$

Historically, when computations were slower, it was more efficient to first compute empirical coefficients and then apply the discrete wavelet transform (DWT).

The final estimate at resolution level L , and up to J , is

$$\hat{f}(x) = \sum_k \hat{c}_{Lk} \phi_{Lk}(x) + \sum_{j=L}^{J-1} \hat{d}_{jk} \psi_{jk}(x).$$

Using Haar mother wavelets and scaling functions,

$$\psi_{jk}(x) = \begin{cases} 2^{j/2}, & 2^{-j} \leq x < 2^{-j}(k + \frac{1}{2}) \\ -2^{j/2}, & 2^{-j}(k + \frac{1}{2}) \leq x < 2^{-j}(k + 1) \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

$$\phi_{jk}(x) = \begin{cases} 2^{j/2}, & 2^{-j}k \leq x < 2^{-j}(k + 1) \\ 0, & \text{otherwise} \end{cases}. \quad (9)$$

The values for J , τ , and L are selected according to standard recommendations in the wavelet literature [7]. The implementation used in this study follows the software developed by other researchers, as described in [34]. The tuning parameter τ was selected using a universal thresholding rule $\sigma\sqrt{2\log n}$, where σ was robustly estimated from the data.

We next discuss the use of these methods for estimating conditional densities when the sample is clustered. We assume that the observed random vector X depends on a latent variable $v \in \{1, \dots, q\}$ and that the conditional density $f_i(x)$ with condition $\{v = i\}$ is unimodal for all $i = \overline{1, q}$. The overall density estimate is given by

$$\hat{f}(x) = \sum_{i=1}^q \hat{p}_i \hat{f}_i(x), \quad (10)$$

When the sample is clustered, the posterior probabilities $\pi_i(x) = P\{v = i | X = x\}$ are estimated as $\hat{\pi}_i(x)$. Then,

$$\hat{p}_i = \frac{1}{n} \sum_{t=1}^n \hat{\pi}_i(X(t)), \quad i = 1, \dots, q.$$

Strict (hard) clustering assigns $\hat{v}(t) = \arg \max_{i=1, \dots, q} \hat{\pi}_i(X(t))$, effectively partitioning the sample into subgroups.

2.2. Sample Clustering Using the EM Algorithm

Mixture distribution models are widely used in applied research for clustering purposes. Among them, the Gaussian Mixture Model (GMM) is particularly popular across various scientific disciplines when addressing practical problems. For instance, GMM is employed in clinical and pharmaceutical studies for analyzing ECG data distributions [35]. In genetics, it is used effectively for clustering data streams [36]. Time series modeling applies this approach to predict drug discontinuation profiles [37]. In astronomy, segmentation of massive datasets is often handled using mixture modeling algorithms [38]. Moreover, the classification of hyperspectral data for both military and civilian applications often relies on GMM-based techniques for comparing spectral features [39].

Simulation-based studies in [28] analyzing multimodal density estimation demonstrated that preliminary data clustering is beneficial. However, the question of which clustering method yields the best results remained open. The comparative analysis conducted in this study shows that probabilistic clustering methods significantly outperform popular geometric clustering approaches (such as k-means, hierarchical clustering, etc.),

even when enhanced with component spherification [40]. Therefore, these geometric methods are not considered further here. We restrict our attention to sample clustering based on approximating the target density by a mixture of Gaussian densities.

If the density $f(x)$ is multimodal, it can be represented as a mixture of several unimodal component densities:

$$f(x) = \sum_{k=1}^q p_k f_k(x). \quad (11)$$

Assume that the observed random variable X depends on a latent variable $v \in \{1, \dots, q\}$, which indicates the class to which the observation belongs. The values $p_k = P\{v = k\}$ are called prior probabilities in classification theory, while $\pi_k(x) = P\{v = k | X = x\}$ denotes the posterior probabilities. The function f_k is treated as the conditional density of X given $v = k$. In this study, we apply hard clustering, i.e., the sample is partitioned into subsets using estimates $\hat{v}(1), \dots, \hat{v}(n)$, where $v(t)$ denotes the predicted class for observation $X(t)$. However, soft clustering may be defined as the estimation of the posterior probabilities $\pi_k(X(t))$ for all $k = 1, \dots, q$, $t = 1, \dots, n$.

The most widely used model in clustering theory and practice is the Gaussian Mixture Model, which is applied here. We assume that the component densities $f_k(x)$ are Gaussian with means μ_k and variances σ_k^2 . We denote the right-hand side of (11) as $f(\theta, x)$, where $\theta = ((p_k, \mu_k, \sigma_k^2), k = 1, \dots, q)$. The posterior probabilities satisfy

$$\pi_k(x) = \frac{p_k f_k(x)}{f(\theta, x)}, \quad k = 1, \dots, q. \quad (12)$$

The estimates $\hat{\pi}_k(x)$ can be obtained via the “plug-in” method by substituting θ with its maximum likelihood estimate: $\theta^* = \underset{\theta}{\operatorname{argmax}} L(\theta)$, $L(\theta) = \prod_{t=1}^n f(\theta, X(t))$. This estimate is commonly obtained using a recursive procedure known as the Expectation-Maximization (EM) algorithm, which is also used in this study.

Suppose that after r iterations, we have estimates $\hat{\pi}_k = \hat{\pi}_k^{(r)}$. Then, the updated parameter estimates $\hat{\theta} = \hat{\theta}^{(r+1)}$ are calculated as

$$\hat{p}_k = \frac{1}{n} \sum_{t=1}^n \hat{\pi}_k(X(t)),$$

$$\hat{\mu}_k = \frac{1}{n p_k} \sum_{t=1}^n \hat{\pi}_k(X(t)) \cdot X(t),$$

$$\hat{\sigma}_k^2 = \frac{1}{n p_k} \sum_{t=1}^n \hat{\pi}_k(X(t)) [X(t) - \hat{\mu}_k] \cdot [X(t) - \hat{\mu}_k]'$$

for all $k = 1, \dots, q$. Inserting $\hat{\theta}^{(r+1)}$ back into (12) yields new estimates $\hat{\pi}_k^{(r+1)}(X(t))$, $k = 1, \dots, q$, $t = 1, \dots, n$. This recursive procedure ensures that the likelihood $L(\hat{\theta}^{(r)})$ is non-decreasing. However, convergence to the global maximum depends strongly on the initial value $\hat{\theta}^{(0)}$ (or $\hat{\pi}^{(0)}$). A simple approach to initialization is to apply a random start strategy, repeating the EM algorithm multiple times with random initial values for $\hat{\pi}^{(0)}$ and selecting the result that yields the highest likelihood value $L(\hat{\theta})$. Alternatively, one can apply sequential component extraction as described in [41].

To determine the number of clusters q , various model adequacy tests may be applied. Let $f^*(x) = f(\theta^*, x)$, where θ^* is the maximum likelihood estimate. Let F^* denote the

corresponding distribution function, and let the empirical distribution function be defined as $\tilde{F}(x) = n^{-1} \sum_{t=1}^n 1_{\{X(t) \leq x\}}$. Following [41], define

$$\psi = \int \frac{\hat{f}(x)}{f^*(x)} d\tilde{F}(x) - 1 = n^{-1} \sum_{t=1}^n \hat{f}(X(t)) / f^*(X(t)) - 1.$$

The null hypothesis that model (11) is adequate is not rejected if $\psi < \varepsilon$. The threshold ε corresponding to a significance level α can be set as close as possible to the quantile $u_\alpha : P\{\psi > u_\alpha\} = \alpha$. The quantile can be estimated using the bootstrap method [42,43]. Let $G(u)$ denote the estimated distribution of ψ . For a fixed n , this defines the distribution function F of X , i.e., $P\{\psi < u\} = G(u, F)$. Given α , the cutoff value ε is defined via $G(\psi, F^*) = 1 - \alpha$. The function $G(\cdot, F^*)$ can be computed using bootstrap resampling.

2.3. Decision Support Systems

In the absence of prior knowledge about the nature of the data in the analyzed sample, selecting the estimator that best approximates the unknown density can be practically challenging. This problem can be addressed using decision-making methods, which can be categorized into two groups based on their underlying algorithms: stochastic and deterministic. These methods are trained using labeled data to generate rules that allow for optimal selection based on available features. Below, we briefly describe the methods considered.

Generalized logistic regression. Binary, ordinal, and nominal response variables are common in many areas of research. Logistic regression is often used to study the relationship between such responses and a set of explanatory variables. Foundational works on logistic regression include Allison [44] and Cox and Snell [45]. When the response levels have no natural order, generalized logits are used. The fitted model is

$$g(\pi_j) = \log\left(\frac{\pi_j}{1 - \pi_j}\right) = \beta_{0j} + v' \beta_j, j = 1, \dots, k, \quad (13)$$

where π_j is the probability that j^{th} event occurs, v is the predictor variable vector, the β_{0j} are intercept parameters, β_j are k vectors of the slope parameter vector, and g is the logit link function that relates π_j to v [42].

Linear discriminant analysis. Assuming each group follows a multivariate normal distribution, a parametric discriminant function can be developed based on the generalized squared Mahalanobis distance [46]. Classification uses either individual within-group covariance matrices or a pooled covariance matrix and takes class prior probabilities into account. Each observation is assigned to the group with the smallest generalized squared distance.

The squared Mahalanobis distance from v to group j is $d_j^2(v) = (v - m_j)' S_j^{-1} (v - m_j)$, where S_j is the within-group or the pooled covariance matrix, and m_j is the mean vector for group j . The group-specific density estimate at v is then $f_j(v) = (2\pi)^{-\frac{p}{2}} |S_j|^{-\frac{1}{2}} \exp(-0.5 d_j^2(v))$. Applying Bayes' theorem, the posterior probability that v belongs to group j is

$$\pi_j = \frac{p_j f_j(v)}{\sum_i p_i f_i(v)}, j = 1, \dots, k, \quad (14)$$

where the sum runs over all groups, and p_i is the prior probability of belonging to group i .

Kernel discriminant analysis. Nonparametric discriminant methods are based on nonparametric estimates of group-specific densities. Kernel functions—such as uniform, Gaussian, Epanechnikov, biweight, or triweight—can be used for density estimation. The

group-specific density at v is estimated as $f_j(v) = \frac{1}{n_j} \sum_y K_j(v - y)$, where the sum is over all training observations y in group j , K_j is the kernel function, and n_j is the sample size for group j . The posterior probabilities are then calculated using Formula (14).

To select the bandwidth for kernel estimation, we apply a well-established univariate plug-in bandwidth selector proposed by Sheather and Jones [47,48].

Neural networks. Neural networks (NNs) are widely applied in various disciplines due to their resemblance to biological neural systems [49]. Several studies have demonstrated the connections between neural networks and classical statistical methods. Comparative analyses of neural networks with discriminant analysis, multiple regression, and logistic regression can be found in [21,23,50,51].

Among classification and regression techniques, artificial neural networks are less commonly used. However, unlike logistic regression or discriminant analysis, neural networks do not rely on predefined equations to model the relationship between predictors and the response. Instead, the relationship is represented by a network of weights and nodes (neurons), akin to the architecture of a biological brain [51]. In this study, a Multilayer Perceptron (MLP) neural network with a single hidden layer was used for classification. This supervised, feed-forward network consists of an input layer (with one node per environmental variable), a hidden layer, and an output layer (with k nodes producing values in $[0, 1]$). The activation function used in MLP is the hyperbolic tangent sigmoid. The network output is computed as

$$T(j) = \rho \left(\sum_{l=1}^r w_{jl} \rho \left(\sum_i w_{li} v_i \right) \right), j = 1, \dots, k \quad (15)$$

where ρ is the activation function, w_{jl} , w_{li} are weights, v_i are input values, and r is the number of hidden neurons. The MLP is trained using the backpropagation learning algorithm, which adjusts weights and biases. In our implementation, 10% of observations were set aside for cross-validation to monitor and terminate training when appropriate. Further details about MLPs can be found in [23].

2.4. Accuracy Assessment Study

A comprehensive simulation study was conducted to compare the statistical estimation methods described previously. To avoid subjectivity in the comparative analysis, Gaussian mixture densities were used as target distributions. These mixtures were originally proposed by J.S. Marron and M.P. Wand [2] and include the following (also see Figure 1):

- Gaussian— $\mathcal{N}(0, 1)$
- Skewed Unimodal— $\frac{1}{5}\mathcal{N}(0, 1) + \frac{1}{5}\mathcal{N}\left(\frac{1}{2}, \left(\frac{2}{3}\right)^2\right) + \frac{3}{5}\mathcal{N}\left(\frac{13}{12}, \left(\frac{5}{9}\right)^2\right)$
- Strongly Skewed— $\sum_{l=0}^7 \frac{1}{8}\mathcal{N}\left(3\left\{\left(\frac{2}{3}\right)^l - 1\right\}, \left(\frac{2}{3}\right)^{2l}\right)$
- Kurtotic Unimodal— $\frac{2}{3}\mathcal{N}(0, 1) + \frac{1}{3}\mathcal{N}\left(0, \left(\frac{1}{10}\right)^2\right)$
- Outlier— $\frac{1}{10}\mathcal{N}(0, 1) + \frac{9}{10}\mathcal{N}\left(0, \left(\frac{1}{10}\right)^2\right)$
- Bimodal— $\frac{1}{2}\mathcal{N}\left(-1, \left(\frac{2}{3}\right)^2\right) + \frac{1}{2}\mathcal{N}\left(1, \left(\frac{2}{3}\right)^2\right)$
- Separated Bimodal— $\frac{1}{2}\mathcal{N}\left(-\frac{3}{2}, \left(\frac{1}{2}\right)^2\right) + \frac{1}{2}\mathcal{N}\left(\frac{3}{2}, \left(\frac{1}{2}\right)^2\right)$
- Asymmetric Bimodal— $\frac{3}{4}\mathcal{N}(0, 1) + \frac{1}{4}\mathcal{N}\left(\frac{3}{2}, \left(\frac{1}{3}\right)^2\right)$
- Trimodal— $\frac{9}{20}\mathcal{N}\left(-\frac{6}{5}, \left(\frac{3}{5}\right)^2\right) + \frac{9}{20}\mathcal{N}\left(\frac{6}{5}, \left(\frac{3}{5}\right)^2\right) + \frac{1}{10}\mathcal{N}\left(0, \left(\frac{1}{4}\right)^2\right)$

- Claw— $\frac{1}{2}\mathcal{N}(0, 1) + \sum_{l=0}^4 \frac{1}{10}\mathcal{N}\left(\frac{l}{2} - 1, \left(\frac{1}{10}\right)^2\right)$
- Double Claw— $\frac{49}{100}\mathcal{N}\left(-1, \left(\frac{2}{3}\right)^2\right) + \frac{49}{100}\mathcal{N}\left(1, \left(\frac{2}{3}\right)^2\right) + \sum_{l=0}^6 \frac{1}{350}\mathcal{N}\left(\frac{l-3}{2}, \left(\frac{1}{100}\right)^2\right)$
- Asymmetric Claw— $\frac{1}{2}\mathcal{N}(0, 1) + \sum_{l=-2}^2 \frac{2^{1-l}}{31}\mathcal{N}\left(l + \frac{1}{2}, \left(\frac{2^{-l}}{10}\right)^2\right)$
- Asymmetric Double Claw— $\sum_{l=0}^1 \frac{46}{100}\mathcal{N}\left(2l - 1, \left(\frac{2}{3}\right)^2\right) + \sum_{l=1}^3 \frac{1}{100}\mathcal{N}\left(-\frac{l}{2}, \left(\frac{1}{100}\right)^2\right) + \sum_{l=1}^3 \frac{7}{100}\mathcal{N}\left(\frac{l}{2}, \left(\frac{7}{100}\right)^2\right)$
- Smooth Comb— $\sum_{l=0}^5 \frac{2^{5-l}}{63}\mathcal{N}\left(\frac{(65-96(\frac{1}{2})^2)}{21}, \left(\frac{(32)}{63}\right)^2\right)$
- Discrete Comb— $\sum_{l=0}^2 \frac{2}{7}\mathcal{N}\left(\frac{12l-15}{7}, \left(\frac{2}{7}\right)^2\right) + \sum_{l=8}^{10} \frac{1}{21}\mathcal{N}\left(\frac{2l}{7}, \left(\frac{1}{21}\right)^2\right)$

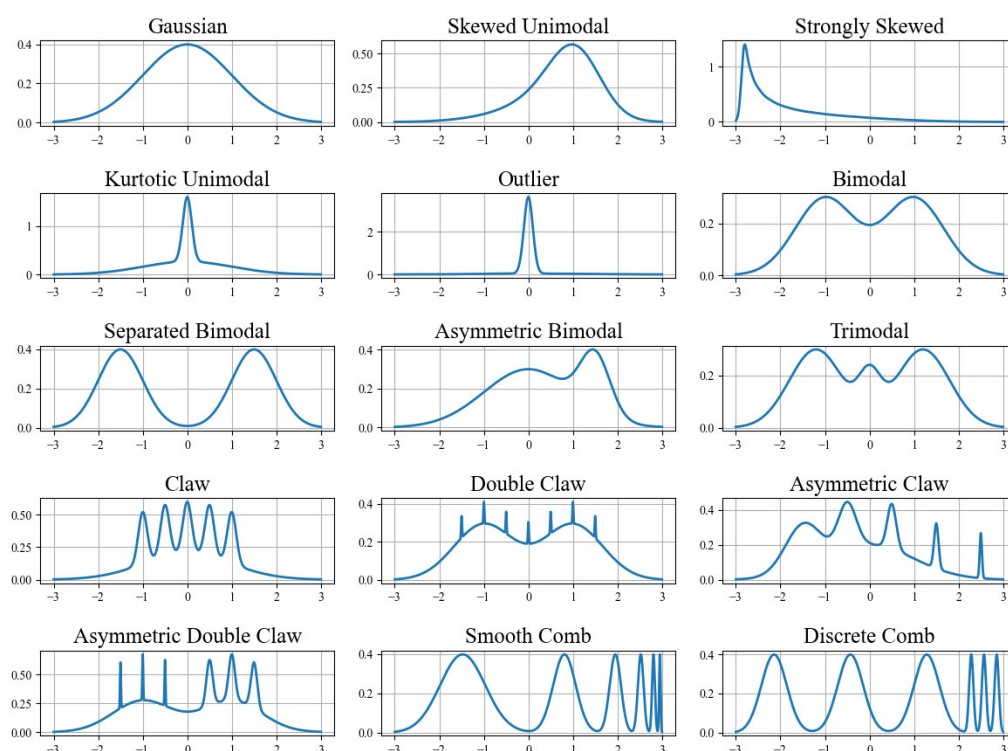


Figure 1. Gaussian mixture densities.

The simulation was performed using small and medium sample sizes: 50, 100, 200, 400, 800, 1600, and 3200. For each case, 100,000 replications were generated. The number of Monte Carlo replications (100,000) was selected based on stability analysis. When using 10,000 replications, average error values were similar but confidence intervals were wider, leading to less clear method rankings. The higher replication count ensured stable estimates and tighter intervals for comparison.

To evaluate the accuracy of the density estimators, the following error measures were calculated:

$$MAE = \frac{1}{n} \sum_{t=1}^n |f(X(t)) - \hat{f}(X(t))| \cong \int |f(x) - \hat{f}(x)| f(x) dx$$

and

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{f(X(t)) - \hat{f}(X(t))}{f(X(t))} \right| \cong \int |f(x) - \hat{f}(x)| dx.$$

In the next stage, decision-making algorithms were compared in terms of their ability to identify the most accurate density estimator for a given unknown distribution. The predictor vectors used by these algorithms included numerical features characterizing the data distribution, such as sample size, the skewness in absolute size, kurtosis, outliers, relative size of the sample, estimated number of clusters, and individual components of the clustered sample outlier ratio.

To estimate outliers in clustered samples, a rule-of-thumb known as the 3-sigma rule was applied. The results of the decision-making method comparison are presented in Table 1, originating from statistical process control [52]. Outliers were identified in each cluster of the GMM using the EM algorithm by the condition $\{X : |X - \hat{\mu}_i| > z_{1-\alpha/2}\hat{\sigma}_i, i = 1, \dots, q\}$, where z is the quantile of the standard normal distribution and $\alpha = 0.05$ denotes the confidence level [53].

Table 1. Misclassification rates for decision-making methods.

Methods	Training Set		Validation Set	
	MAE	MAPE	MAE	MAPE
GLR	0.28081632	0.36843537	0.28476190	0.37142857
LDA	0.54897959	0.59959183	0.54730158	0.59650793
KDA	0.07659863	0.11455782	0.20825396	0.26761904
NN	0.19306122	0.25741497	0.27936508	0.33904762
GLR + NN	0.15986395	0.17795918	0.26888889	0.33015873
KDA + NN	0.12707483	0.17401361	0.19619048	0.27682540

Bold values indicate the method with the lowest error for each metric and set.

Validation analysis was performed to evaluate the effectiveness of the decision-making systems. All samples from the Gaussian mixture models were split into two non-overlapping subsets—a training set and a validation set—using stratified simple random sampling with equal probability selection and without replacement [54].

The comparison error rate was defined as the proportion of incorrectly selected density estimators across all validation samples. The ratio between the training and validation sets was set at 7 : 3, allowing the decision-making systems to be sufficiently trained while minimizing the overall comparison error rate.

3. Results

The error values for the applied methods, using optimally selected parameter values, are presented in Appendix A. These represent typical model cases; in other cases, the behavior of the errors was similar. For each method, the mean error values over 100,000 samples are reported. Appendix A also provides graphical representations (heatmaps) of the density estimation results for the AKDE, PPDE, LSDE, and WEDE methods, both without and with initial data clustering. The latter results are denoted by a prefix C in front of the method name.

The general analysis showed that all methods exhibited a decreasing trend in MAE and MAPE values as sample size increased, confirming the consistency of the estimators. However, the precision gain differed depending on the method and the shape of the modeled density. While AKDE typically demonstrated stable performance, PPDE often showed better accuracy, especially for symmetric and multimodal densities. LSDE revealed its strength in larger samples, whereas WEDE frequently performed well when sensitivity to boundary values or local structure recovery was important.

The results of the decision-making method comparison are presented in Table 1. Here, the dependent variable is the method that most accurately estimated the density for each sample, based on both MAE and MAPE. GLR denotes generalized logistic regression,

LDA—linear discriminant analysis, KDA—kernel discriminant analysis with Gaussian kernel (other kernels mentioned in Section 2.3 were also tested but yielded lower accuracy), and NN—a three-layer feedforward neural network with 20 neurons. Additionally, hybrid approaches were used: a second neural network refined the results of the first method using the classification probabilities from the statistical method.

As shown in the table, KDA yielded the lowest error when decision-making methods were applied independently. Combining different methods improved accuracy in most cases compared to using them individually, with the exception of MAPE-based classification when applying KDA alone. Hybrid approaches led to 5.5% to 11% reductions in misclassification rates.

4. Discussion

The following discussion aims to contextualize the comparative results obtained from the simulation study. By examining the behavior of the four density estimation methods under varying distributional forms, we aim to evaluate their robustness, flexibility, and sensitivity to sample size, skewness, multimodality, and outliers. Special attention is given to the added value of clustering as a preprocessing step. An extended set of visual results supporting this discussion is provided in Figures A1–A15 in the Appendix A.

Focusing on specific results, in the case of the Gaussian density, all methods demonstrated good convergence. PPDE and LSDE were particularly notable, with their MAE values decreasing from $\approx 0.025 - 0.08$ (for $n = 50$) to $\approx 0.004 - 0.006$ (for $n = 3200$). This confirms their effectiveness when the target distribution is unimodal and close to normal. Although WEDE maintained stability, it slightly lagged behind in terms of precision.

For skewed unimodal distributions, performance differences became more evident. PPDE errors increased, especially for medium-sized samples, indicating sensitivity to skewness. In contrast, WEDE and LSDE adapted better to such structures, and clustering further improved their performance. This suggests that for distributions with substantial asymmetry, more flexible methods—particularly wavelet-based—yield higher accuracy.

Differences between methods were even more pronounced for strongly skewed densities. MAPE values sometimes exceeded 0.8, especially for LSDE and CWEDE in small samples. However, as sample size increased, accuracy significantly improved, with WEDE consistently reducing error, even for such challenging distributions. Clustering played a critical role here—methods such as CAKDE and CWEDE were substantially more accurate than their non-clustered counterparts.

For complex multimodal distributions such as Bimodal, Trimodal, Claw, and Double Claw, WEDE and LSDE demonstrated notable strengths. WEDE effectively recovered fine modal structures, particularly for $n \geq 800$, while LSDE showed low MAE values in large samples. Nevertheless, without clustering, these methods were often ineffective in distinguishing mixture components. Clustered variants—especially CLSDE and CWEDE—performed significantly better by allowing each mode to be modeled individually, which is essential in multimodal scenarios.

In outlier-prone distributions, many methods struggled. PPDE and LSDE performed poorly without clustering, with MAPE often exceeding 0.5 and, in some cases, surpassing 1. WEDE, however, retained high accuracy, highlighting its robustness to anomalies. Similarly, CAKDE and CWEDE were stable and significantly reduced errors, confirming the necessity of clustering for distributions with outlying features.

For asymmetric and complex densities (e.g., Asymmetric Bimodal, Asymmetric Claw, Asymmetric Double Claw), none of the methods were consistently stable across all sample sizes. Still, WEDE and LSDE (especially with clustering) showed consistent accuracy improvements as the sample size increased, with acceptable MAPE values achieved only

from $n \approx 1600$. This suggests that for complex structures, effectiveness strongly depends on proper initial clustering and sufficiently large sample sizes.

In the last group of results—Comb-type densities characterized by highly intricate shapes and multiple localized peaks—WEDE demonstrated the highest stability, particularly for $n \geq 800$. LSDE showed impressive precision, but only in very large samples. Clustered methods, especially CLSDE and CWEDE, again significantly enhanced accuracy, confirming their versatility.

In summary, applying clustering generally improved the performance of all four methods. Although the improvement was not always dramatic, for densities with multiple modes or asymmetry, EM-based clustering proved to be a crucial step toward more accurately reconstructing the true density structure. It was also observed that some methods (especially PPDE) are highly sensitive to overlapping components and large variances, so their use without clustering should be approached with caution. While the accuracy of all methods improves with increasing sample size, the choice between them should be guided by the characteristics of the target density. For small samples, WEDE or CAKDE are recommended; for medium sizes, clustered versions of PPDE or LSDE are more suitable; and, for large samples, CLSDE, CWEDE, or even PPDE in its classical form may be effective. This confirms that no single method is universally optimal—the data nature and density structure must be considered to make a justified choice of the estimation method.

5. Conclusions

This study presented a comprehensive comparative analysis of four nonparametric density estimation methods—adaptive kernel density estimation (AKDE), projection pursuit density estimation (PPDE), log-spline density estimation (LSDE), and wavelet-based density estimation (WEDE)—applied to a wide range of univariate Gaussian mixture distributions. The analysis covered both classical methods and their extended versions, which included data clustering using the EM algorithm. Estimation accuracy was assessed using MAE and MAPE criteria across various sample sizes, allowing for an in-depth evaluation of method behavior and stability.

The results revealed that no single estimation method consistently outperformed the others across all density types and sample sizes. In small samples, simpler methods such as AKDE or PPDE yielded lower errors for symmetric and low-modality densities, but their performance deteriorated significantly when faced with asymmetry or fragmented structures. In contrast, LSDE and WEDE showed greater flexibility and the ability to reconstruct localized features, particularly in larger samples. Wavelet-based estimators were especially effective for densities with sharp transitions or localized extrema.

One of the key contributions of this work was the analysis of clustering as a preprocessing step. Clustered versions of the methods consistently outperformed their unclustered counterparts for multimodal, asymmetric, and outlier-prone distributions. Clustering the sample using the EM algorithm allowed each component to be modeled separately, reducing overall estimation error. This was particularly valuable for distributions with overlapping or distant components (e.g., “Separated Bimodal”, “Asymmetric Claw”, “Outlier”). These results underscore the importance of incorporating clustering into the analytical pipeline for more complex distributions.

Additionally, the effectiveness of decision-making systems (GLM, LDA, KDA, and MLP-type neural networks) was evaluated for automatically selecting the most appropriate density estimation method based on sample characteristics such as skewness, kurtosis, modality, outlier frequency, number of clusters, etc. Validation showed that the neural network-based system (MLP) had the lowest rate of incorrect method selection—on average, 18.4%—whereas simpler models such as GLM or LDA ranged between 24% and 31%. This

indicates that even in analytically complex scenarios, neural networks can effectively automate the method selection process, reducing the influence of subjectivity.

Overall, the findings suggest that effective nonparametric density estimation requires a combined approach—first analyzing the structure of the sample (e.g., through clustering) and then applying intelligent decision-making algorithms. The existence of a universally optimal method is unlikely; however, by systematically applying model selection principles, high levels of generalizability can be achieved.

Future research should extend this work to the multivariate case, where both clustering and estimation become more complex. It is also relevant to investigate dynamically adaptive decision-making systems capable of operating in real-time data stream environments. Further attention should be given to applying deep neural networks to the estimator selection problem, incorporating automated feature engineering and ensemble methods. Finally, practical applications in real-world domains such as anomaly detection, signal processing, and financial modeling could become an important direction for further investigation.

Author Contributions: Conceptualization, T.R. and G.S.; methodology, T.R. and G.S.; software, T.R.; validation, T.R.; formal analysis, T.R.; investigation, T.R.; resources, T.R., G.S., B.N. and J.A.Z.; data curation, T.R.; writing—original draft preparation, T.R., G.S. and B.N.; writing—review and editing, T.R., G.S. and B.N.; visualization, G.S.; supervision, T.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Generated data sets were used in the study (see in Section 2.4).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

KDE	Kernel density estimator
AKDE	Adaptive kernel density estimator
LSDE	Log-spline density estimator
PPDE	Projection pursuit density estimator
WEDE	Wavelet expansion density estimator
CAKDE	AKDE with pre-clustering
CLSDE	LSDE with pre-clustering
CPPDE	PPDE with pre-clustering
CWEDE	WEDE with pre-clustering
GMM	Gaussian mixture model
EM	Expectation maximization
GLR	Generalized logistic regression
GLRT	Generalized likelihood ratio test
KDA	Kernel discriminant analysis
LDA	Linear discriminant analysis
NN	Neural network
MAE	Mean absolute error
MAPE	Mean absolute percentage error

Appendix A

An analysis of the effectiveness of initial data clustering is presented. Each figure corresponds to a heatmap of estimation errors for different densities, depending on the sample size.

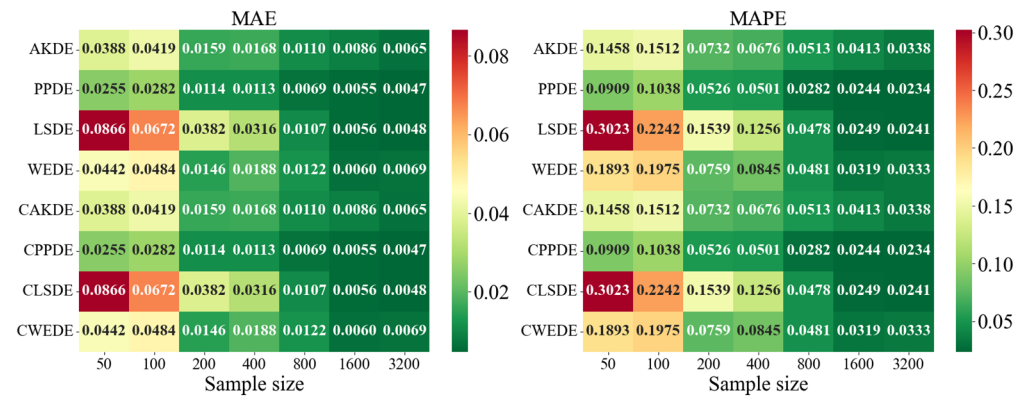


Figure A1. Estimation accuracy for Gaussian density.

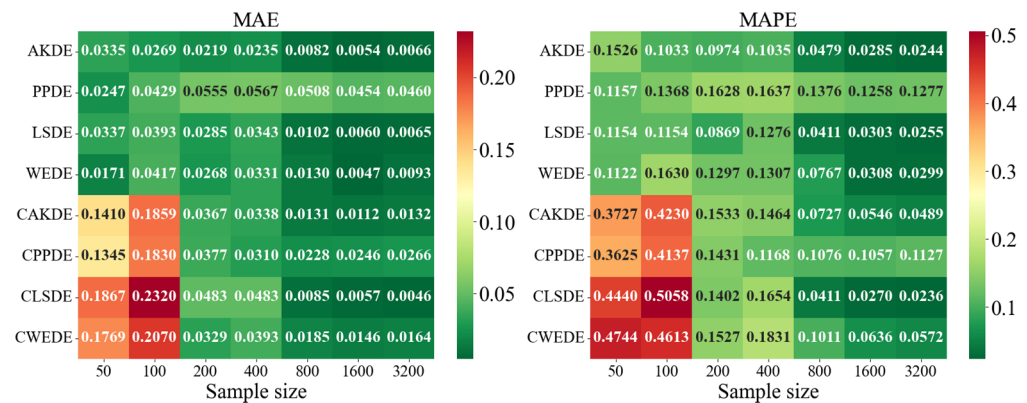


Figure A2. Estimation accuracy for Skewed Unimodal density.

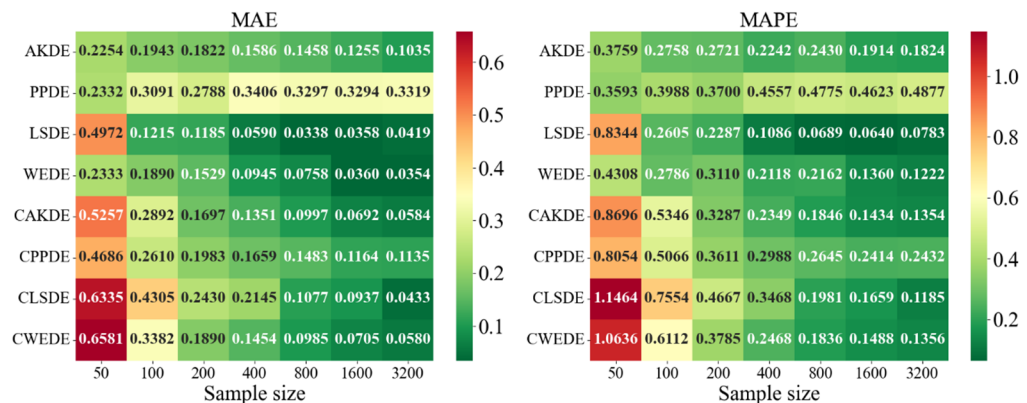


Figure A3. Estimation accuracy for Strongly Skewed density.

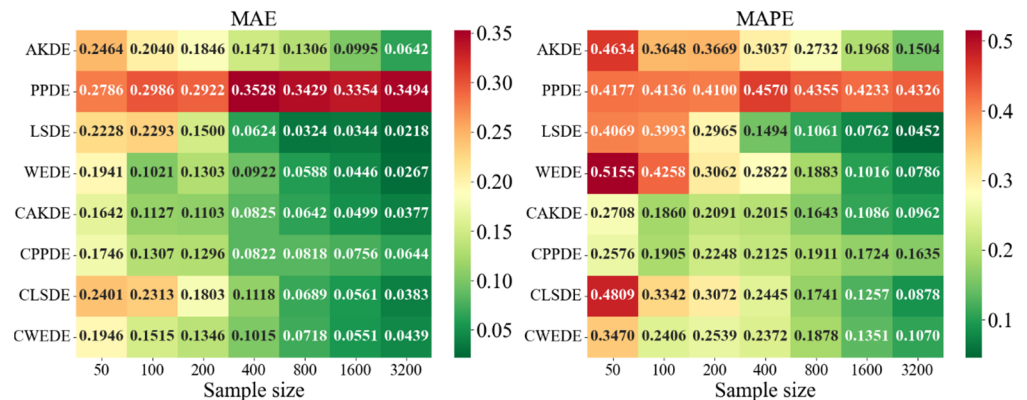


Figure A4. Estimation accuracy for Kurtotic Unimodal density.

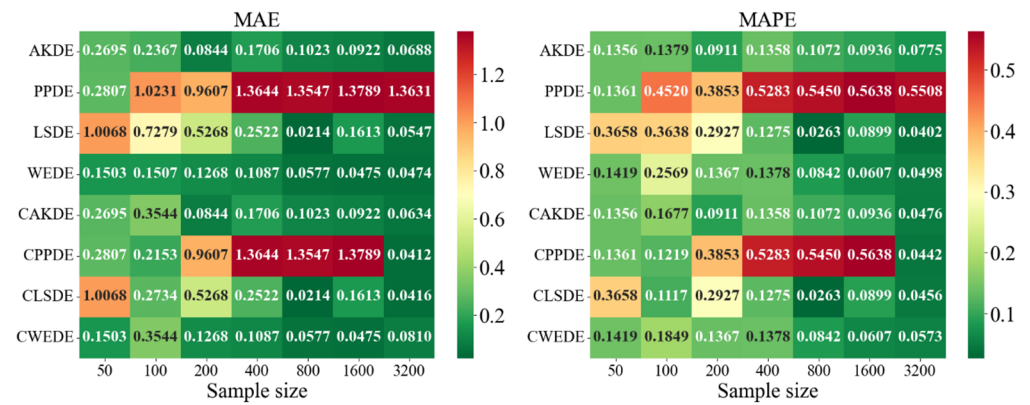


Figure A5. Estimation accuracy for Outlier density.

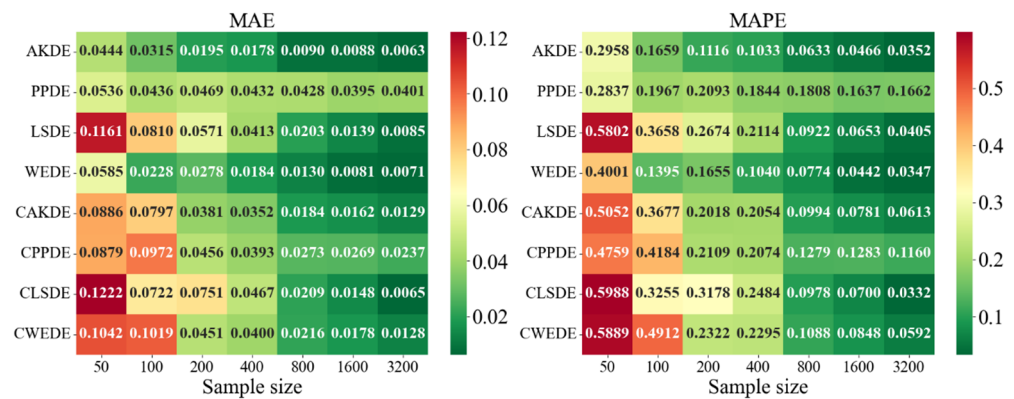


Figure A6. Estimation accuracy for Bimodal density.

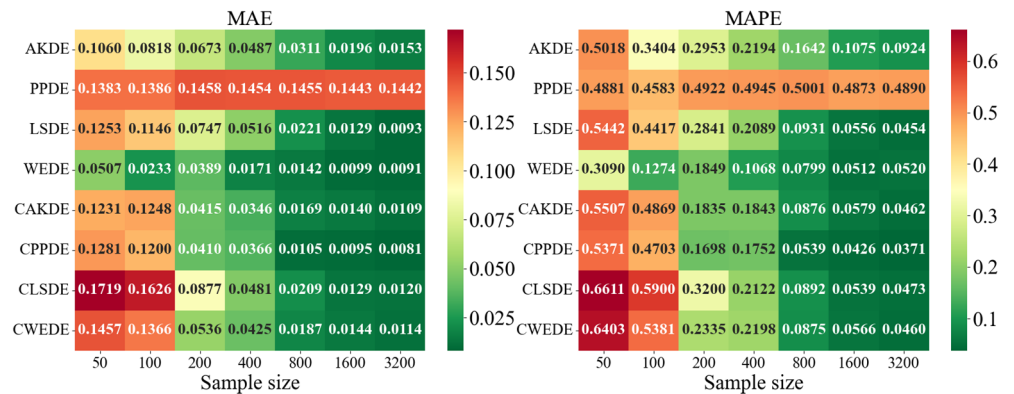


Figure A7. Estimation accuracy for Separated Bimodal density.

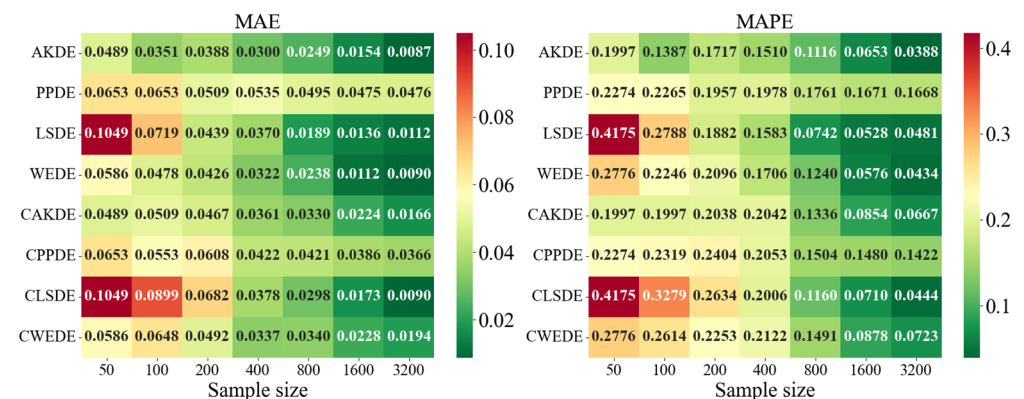


Figure A8. Estimation accuracy for Asymmetric Bimodal density.

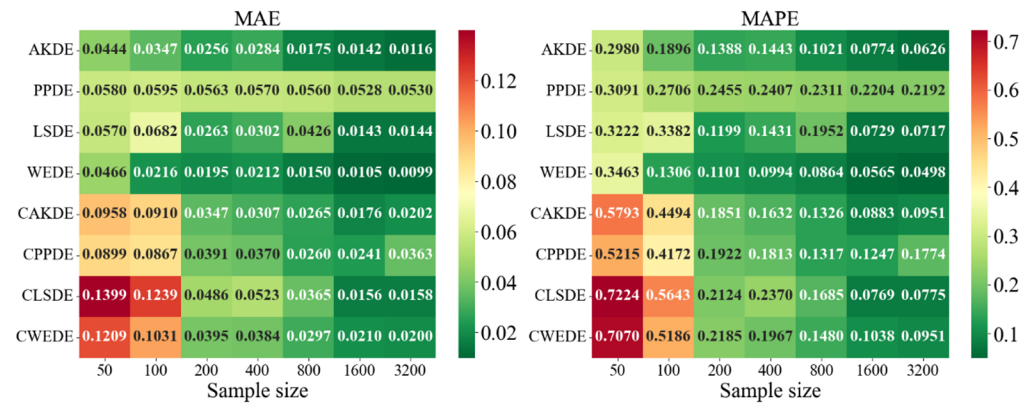


Figure A9. Estimation accuracy for Trimodal density.

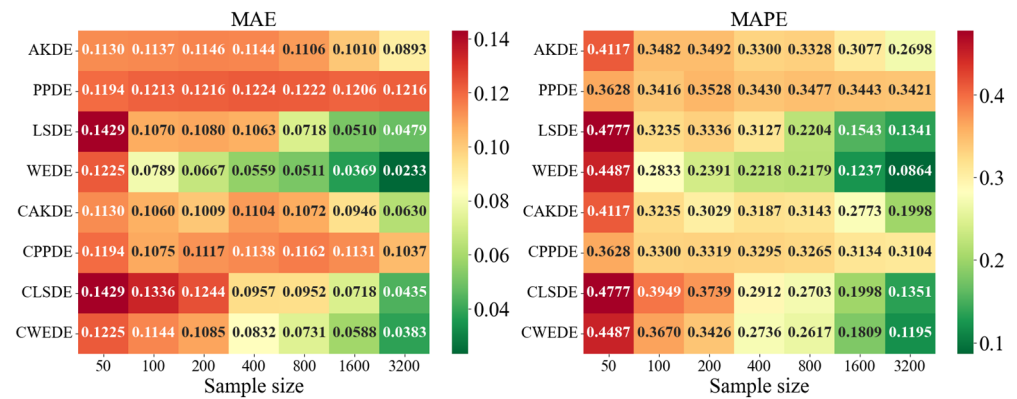


Figure A10. Estimation accuracy for Claw density.

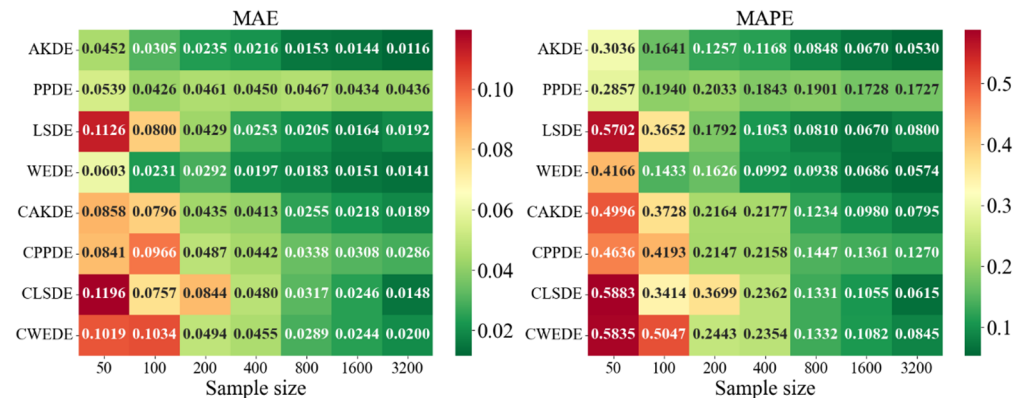


Figure A11. Estimation accuracy for Double Claw density.

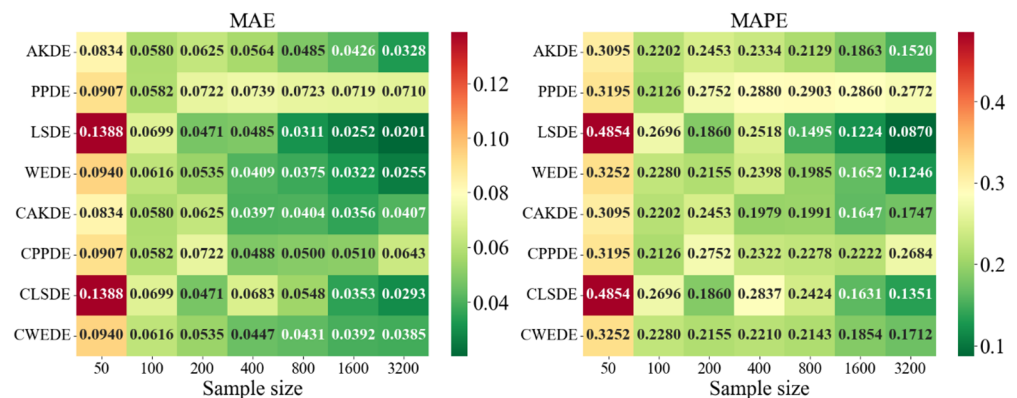


Figure A12. Estimation accuracy for Asymmetric Claw density.

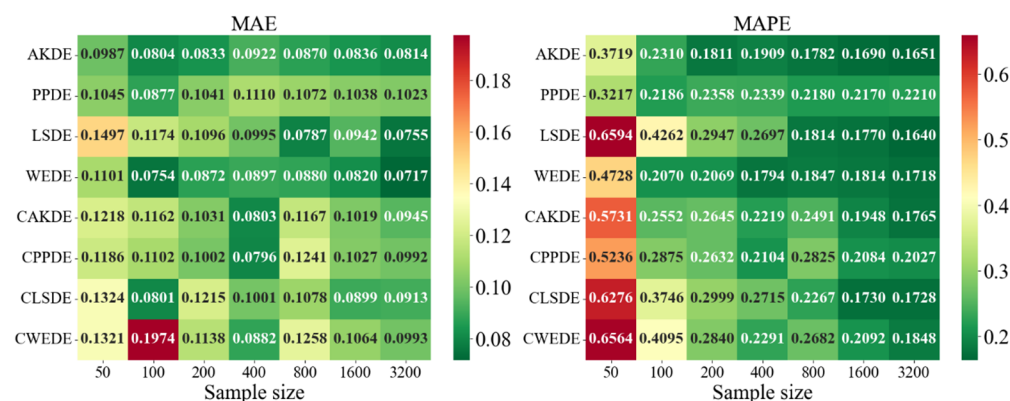


Figure A13. Estimation accuracy for Asymmetric Double Claw density.

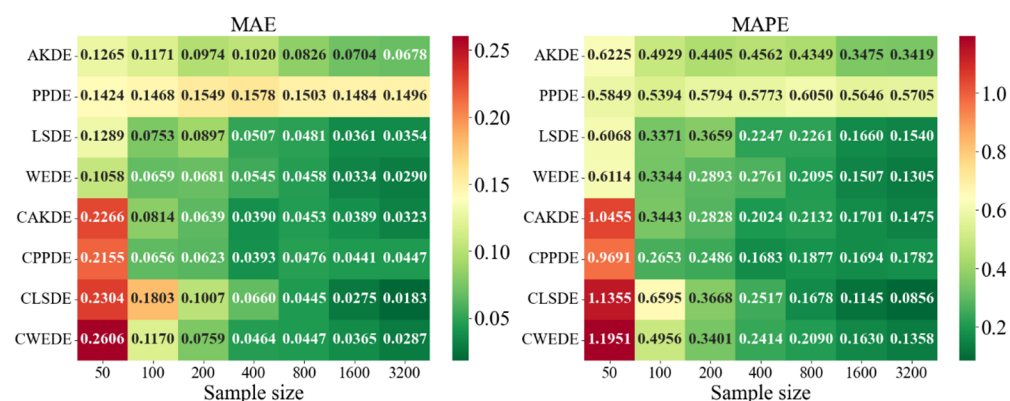


Figure A14. Estimation accuracy for Smooth Comb density.

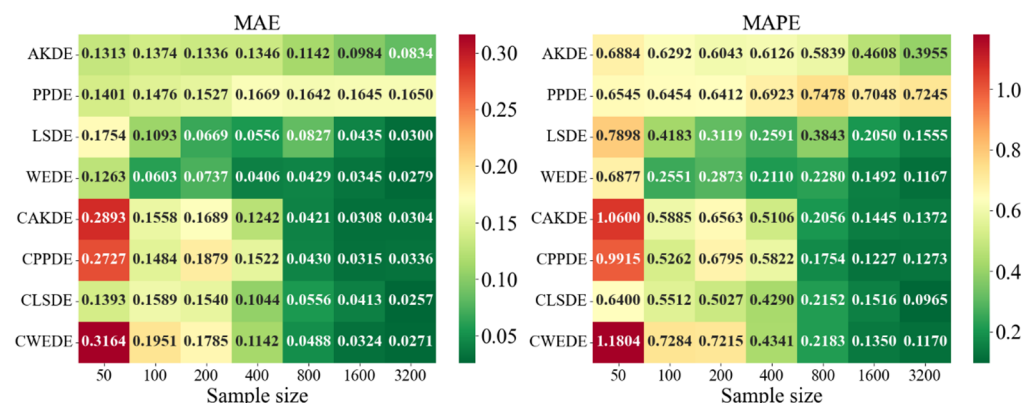


Figure A15. Estimation accuracy for Discrete Comb density.

References

1. Jones, M.C.; Marron, J.S.; Sheather, S.J. A Brief Survey of Bandwidth Selection for Density Estimation. *J. Am. Stat. Assoc.* **1996**, *91*, 401–407. [\[CrossRef\]](#)
2. Marron, J.S.; Wand, M.P. Exact Mean Integrated Squared Error. *Ann. Stat.* **1992**, *20*, 712–736. [\[CrossRef\]](#)
3. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Springer: Dordrecht, The Netherlands, 1986.
4. Ghosal, S. Convergence Rates for Density Estimation with Bernstein Polynomials. *Ann. Stat.* **2001**, *29*, 1264–1280. [\[CrossRef\]](#)
5. Elgammal, A.; Duraiswami, R.; Davis, L.S. Efficient Kernel Density Estimation Using the Fast Gauss Transform with Applications to Color Modeling and Tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1499–1504. [\[CrossRef\]](#)
6. Friedman, J.H. Exploratory Projection Pursuit. *J. Am. Stat. Assoc.* **1987**, *82*, 249–266. [\[CrossRef\]](#)
7. Herrick, D.R.M.; Nason, G.P.; Silverman, B.W. Some New Methods for Wavelet Density Estimation. 2001. Available online: <https://www.jstor.org/stable/25051366?seq=1> (accessed on 13 June 2025).
8. Kooperberg, C.; Stone, C.J. A Study of Logpline Density Estimation. *Comput. Stat. Data Anal.* **1991**, *12*, 327–347. [\[CrossRef\]](#)

9. Hyndman, R.J.; Yao, Q. Nonparametric Estimation and Symmetry Tests for Conditional Density Functions. *J. Nonparametr. Stat.* **2002**, *14*, 259–278. [[CrossRef](#)]
10. Loader, C.R. Local Likelihood Density Estimation. *Ann. Stat.* **1996**, *24*, 1602–1618. [[CrossRef](#)]
11. Delicado, P.; del Río, M. A Generalization of Histogram Type Estimators. 1999. Available online: <https://repositori.upf.edu/items/07ef9af1-c209-4297-956e-b839ad6eef4f> (accessed on 12 June 2025).
12. Scott, D.W.; Factor, L.E. Monte Carlo Study of Three Data-Based Nonparametric Probability Density Estimators. *J. Am. Stat. Assoc.* **1981**, *76*, 9–15. [[CrossRef](#)]
13. Hwang, J.-N.; Lay, S.-R.; Lippman, A. Nonparametric Multivariate Density Estimation: A Comparative Study. *IEEE Trans. Signal Process.* **1994**, *42*, 2795–2810. [[CrossRef](#)]
14. Fenton, V.M.; Gallant, A.R. Qualitative and Asymptotic Performance of SNP Density Estimators. *J. Econom.* **1996**, *74*, 77–118. [[CrossRef](#)]
15. Fadda, D.; Slezak, E.; Bijaoui, A. Density Estimation with Non-Parametric Methods. *Astron. Astrophys. Suppl. Ser.* **1998**, *127*, 335–352. [[CrossRef](#)]
16. Eugecioglu, Ö.; Srinivasan, A. Efficient Nonparametric Density Estimation on the Sphere with Applications in Fluid Mechanics. *SIAM J. Sci. Comput.* **2000**, *22*, 152–176. [[CrossRef](#)]
17. Takada, T. Nonparametric Density Estimation: A Comparative Study. *Econ. Bull.* **2001**, *3*, 2795–2810.
18. Gallant, A.R.; Nychka, D.W. Semi-Nonparametric Maximum Likelihood Estimation. *Econometrica* **1987**, *55*, 363. [[CrossRef](#)]
19. Ćwik, J.; Koronacki, J. Multivariate Density Estimation: A Comparative Study. *Neural Comput. Appl.* **1997**, *6*, 173–185. [[CrossRef](#)]
20. Couvreur, L.; Couvreur, C. Wavelet-Based Method for Nonparametric Estimation of HMMs. *IEEE Signal Process. Lett.* **2000**, *7*, 25–27. [[CrossRef](#)]
21. Gallinari, P.; Thiria, S.; Badran, F.; Fogelman-Soulie, F. On the Relations between Discriminant Analysis and Multilayer Perceptrons. *Neural Netw.* **1991**, *4*, 349–360. [[CrossRef](#)]
22. Manel, S.; Dias, J.-M.; Ormerod, S.J. Comparing Discriminant Analysis, Neural Networks and Logistic Regression for Predicting Species Distributions: A Case Study with a Himalayan River Bird. *Ecol. Model.* **1999**, *120*, 337–347. [[CrossRef](#)]
23. Malhotra, M.K.; Sharma, S.; Nair, S.S. Decision Making Using Multiple Models. *Eur. J. Oper. Res.* **1999**, *114*, 1–14. [[CrossRef](#)]
24. Ge, H.; Dong, L.; Huang, M.; Zang, W.; Zhou, L. Adaptive Kernel Density Estimation for Traffic Accidents Based on Improved Bandwidth Research on Black Spot Identification Model. *Electronics* **2022**, *11*, 3604. [[CrossRef](#)]
25. Amador Luna, D.; Alonso-Chaves, F.M.; Fernández, C. Kernel Density Estimation for the Interpretation of Seismic Big Data in Tectonics Using QGIS: The Türkiye–Syria Earthquakes (2023). *Remote Sens.* **2024**, *16*, 3849. [[CrossRef](#)]
26. Puranik, B.; Madhow, U.; Pedarsani, R. Generalized Likelihood Ratio Test for Adversarially Robust Hypothesis Testing. *IEEE Trans. Signal Process.* **2022**, *70*, 4124–4139. [[CrossRef](#)]
27. Helali, S.; Masmoudi, A.; Slaoui, Y. Semi-Parametric Estimation Using Bernstein Polynomial and a Finite Gaussian Mixture Model. *Entropy* **2022**, *24*, 315. [[CrossRef](#)] [[PubMed](#)]
28. Arnastauskaitė, J.; Ruzgas, T. Accuracy of Nonparametric Density Estimation for Univariate Gaussian Mixture Models: A Comparative Study. *Math. Model. Anal.* **2020**, *25*, 622–641. [[CrossRef](#)]
29. Friedman, J.H.; Stuetzle, W.; Schroeder, A. Projection Pursuit Density Estimation. *J. Am. Stat. Assoc.* **1984**, *79*, 599–608. [[CrossRef](#)]
30. Donoho, D.L.; Johnstone, I.M.; Kerkycharian, G.; Picard, D. Density Estimation by Wavelet Thresholding. *Ann. Stat.* **1996**, *24*, 508–539. [[CrossRef](#)]
31. Hall, P.; Patil, P. Formulae for Mean Integrated Squared Error of Nonlinear Wavelet-Based Density Estimators. *Ann. Stat.* **1995**, *23*, 905–928. [[CrossRef](#)]
32. Huber, P.J. Projection Pursuit. *Ann. Stat.* **1985**, *13*, 435–475. [[CrossRef](#)]
33. Polynomial Spline Routines. Available online: <https://cran.r-project.org/web/packages/polspline/> (accessed on 13 June 2025).
34. WaveThresh4. Available online: <https://people.maths.bris.ac.uk/~wavethresh/> (accessed on 13 June 2025).
35. Meng, L.; Frei, M.G.; Osorio, I.; Strang, G.; Nguyen, T.Q. Gaussian Mixture Models of ECoG Signal Features for Improved Detection of Epileptic Seizures. *Med. Eng. Phys.* **2004**, *26*, 379–393. [[CrossRef](#)] [[PubMed](#)]
36. Gao, M.; Tai-hua, C.; Gao, X. Application of Gaussian Mixture Model Genetic Algorithm in Data Stream Clustering Analysis. In Proceedings of the 2010 IEEE International Conference on Intelligent Computing and Intelligent Systems, Xiamen, China, 29–31 October 2010; pp. 786–790.
37. Lim, C.P.; Quek, S.S.; Peh, K.K. Application of the Gaussian Mixture Model to Drug Dissolution Profiles Prediction. *Neural Comput. Appl.* **2005**, *14*, 345–352. [[CrossRef](#)]
38. Ashman, K.M.; Bird, C.M.; Zepf, S.E. Detecting Bimodality in Astronomical Datasets. *Astron. J.* **1994**, *108*, 2348–2361. [[CrossRef](#)]
39. Beaven, S.G.; Stein, D.; Hoff, L.E. Comparison of Gaussian Mixture and Linear Mixture Models for Classification of Hyperspectral Data. In Proceedings of the IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No.00CH37120), Honolulu, HI, USA, 24–28 July 2000; Volume 4, pp. 1597–1599.

40. Smidtaite, R. Application of Nonlinear Statistics for Distribution Density Estimation of Random Vectors. Master's Thesis, Kaunas University of Technology, Kaunas, Lithuania, 2008.
41. Rudzakis, R.; Radavičius, M. Statistical Estimation of a Mixture of Gaussian Distributions. *Acta Appl. Math.* **1995**, *38*, 37–54. [CrossRef]
42. Hall, P. The Bootstrap and Edgeworth Expansion. Available online: <https://www.abebooks.com/9780387945088/Bootstrap-Edgeworth-Expansion-Springer-Series-0387945083/plp> (accessed on 13 June 2025).
43. Wong, M.A. A Bootstrap Testing Procedure for Investigating the Number of Subpopulations. *J. Stat. Comput. Simul.* **1985**, *22*, 99–112. [CrossRef]
44. Allison, P.D. *Logistic Regression Using SAS: Theory and Application*; Wiley: Cary, NC, USA, 2008; ISBN 978-0-471-22175-3.
45. Cox, D.R.; Snell, E.J. *Analysis of Binary Data*, 2nd ed.; Chapman and Hall: London, UK, 1989; ISBN 9780412306204.
46. Rao, C.R. *Linear Statistical Inference and Its Applications*, 2nd ed.; Wiley Series in Probability and Statistics: 1st ed.; Wiley: Hoboken, NJ, USA, 1973; ISBN 978-0-471-70823-0.
47. JONES, M.; Sheather, S. A reliable data-based bandwidth selection method for kernel density-estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1991**, *53*, 683–690.
48. Hall, P.; Kang, K.-H. Bandwidth Choice for Nonparametric Classification. *Ann. Stat.* **2005**, *33*, 284–306. [CrossRef]
49. Paliwal, M.; Kumar, U.A. Neural Networks and Statistical Techniques: A Review of Applications. *Expert. Syst. Appl.* **2009**, *36*, 2–17. [CrossRef]
50. Schumacher, M.; Roßner, R.; Vach, W. Neural Networks and Logistic Regression: Part I. *Comput. Stat. Data Anal.* **1996**, *21*, 661–682. [CrossRef]
51. Jardin, P.D.; Ponsaillé, J.; Alunni-Perret, V.; Quatrehomme, G. A Comparison between Neural Network and Other Metric Methods to Determine Sex from the Upper Femur in a Modern French Population. *Forensic Sci. Int.* **2009**, *192*, 127.e1–127.e6. [CrossRef] [PubMed]
52. Shewhart, W.A. *Statistical Method from the Viewpoint of Quality Control*; USDA National Agricultural Library: Beltsville, MD, USA, 1939. [CrossRef]
53. Davies, P.L.; Gather, U. The Identification of Multiple Outliers. *J. Am. Stat. Assoc.* **1993**, *88*, 782–792. [CrossRef]
54. Bentley, J.; Floyd, B. Programming pearls: A sample of brilliance. *Commun. ACM* **1987**, *30*, 754–757. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.