

Asymptotic normality of local linear regression estimator for mixtures with varying concentrations

Daniel Horbunov, Rostyslav Maiboroda*

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

danielhorbunov@knu.ua (D. Horbunov), rostmaiboroda@knu.ua (R. Maiboroda)

Received: 27 June 2025, Revised: 6 August 2025, Accepted: 7 August 2025,
Published online: 2 September 2025

Abstract Finite mixtures with different regression models for different mixture components naturally arise in statistical analysis of biological and sociological data. In this paper a model of mixtures with varying concentrations is considered in which the mixing probabilities are different for different observations. The modified local linear regression estimator (mLLRE) is considered for nonparametric estimation of the unknown regression function for the given component of mixture. The asymptotic normality of the mLLRE is proved in the case when the regressor's probability density function has jumps. Theoretically optimal bandwidth is derived. Simulations were made to estimate the accuracy of the normal approximation.

Keywords Nonparametric regression, mixture with varying concentrations, local linear regression, asymptotic normality, bandwidth selection

2020 MSC 62-04, 62G05, 62G08, 62G20

1 Introduction

In medical, biological and sociological studies the investigated population is frequently a mixture of subpopulations (components of the mixture) with different distributions of observed variables. If the subpopulation which a subject belongs to is not known exactly, the distribution of its variables is a mixture of subpopulations' distributions. In the classical finite mixture models (FMM) the concentrations of the components in the mixture (mixing probabilities) are the same for all observations. See [11, 9] and

*Corresponding author.

[14] for results on parametric estimation under FMM. In a more flexible mixture with varying concentrations model (MVC) the concentrations are different for different observations. See [7, 8] for the theory of nonparametric estimation in these models and their application to a DNA-microchip data, and [10] for the application of MVC in the analysis of a neurological data.

Regression models are applied usually to describe dependency between different numerical variables of one subject. In the case of homogeneous sample there exist many nonparametric estimators of the regression function, such as the Nadaraya–Watson estimator (NWE) and local linear regression estimator (LLRE) [4]. A modification of NWE (mNWE) for the estimation of regression function of some MVC component is presented in [2] which also contains the derivation of asymptotic normality for mNWE.

It is well known that for homogeneous samples NWE demonstrates an inappropriate bias in points where the regressor probability density function (PDF) has discontinuity (jump points). The bias of LLRE in this case is significantly smaller [3]. A modification of LLRE for MVC (mLLRE) was considered in [5]. The consistency of mLLRE was shown in [6] and the performance of mLLRE was compared to mNWE by simulations.

In this paper we continue the study of asymptotic mLLRE behavior in jump points and points of continuity of the regressor’s PDF. It is shown that under suitable assumptions mLLRE is asymptotically normal at jump points as well as in the continuity points of the regressor distribution. This result allows to calculate the theoretically optimal bandwidth for mLLRE which minimizes the asymptotic mean squared error.

Semiparametric models similar to the one considered in this paper were discussed in [15, 12] and [13]. In these papers some versions of EM-algorithm are used to estimate the regression functions of the mixture components. Since the EM-algorithm for mixtures is based on the iteratively reweighted likelihood maximization, to construct the estimators the authors need a parametric model for the error term in the regression model. In contrast to the EM technique the approach of this paper is nonparametric both by the regression function and the distribution of the errors.

The rest of the paper is organized as follows. In Section 2 the mixture of regression models is described, in terms of which the definition of the mLLRE is recalled. Section 3 contains the main result on asymptotic normality of the estimator. In Section 4 an optimal bandwidth parameter selection for the mLLRE is discussed. The proof of the main result is presented in Section 5. Simulations for the mLLRE are provided in Section 6. Conclusive remarks are placed in Section 7.

2 Mixture of regressions and the locally linear estimator

2.1 Mixture of regressions

Consider a sample with n subjects $\{O_j\}_{j=1}^n$. Each subject O_j belongs to one of the M subpopulations (components of the mixture). For each $j = \overline{1, n}$ the component which contains O_j is unknown. The numerical index of the containing component is denoted $\kappa_j = \kappa(O_j)$, $1 \leq \kappa_j \leq M$; it is a latent (unobserved) random variable, yet the distributions of κ_j are assumed to be known. The probabilities

$$p_{j:n}^{(k)} = \mathbf{P}(\kappa_j = k), \quad j = \overline{1, n}, \quad k = \overline{1, M} \quad (1)$$

are called *concentrations* of components of the mixture or *mixing probabilities*.

For each subject O_j one observes a bivariate vector of numerical variables $\xi_j = (X_j, Y_j)$ where $X_j = X(O_j)$ and $Y_j = Y(O_j)$ are the regressor and response respectively. The distribution of these variables is described by the regression model:

$$Y_j = g^{(\kappa_j)}(X_j) + \varepsilon_j, \quad j = \overline{1, n},$$

where $g^{(k)}$ is an unknown regression function for k -th component of mixture, $\varepsilon_j = \varepsilon(O_j)$ is a random error term. It is assumed that the vectors $\{(X_j, Y_j)\}_{j=1}^n$ are mutually independent for any fixed $n \geq 1$, and for all $j = \overline{1, n}$, X_j and ε_j are conditionally independent under the condition $\{\kappa_j = m\}$, $m = \overline{1, M}$.

For all $k = \overline{1, M}$ the conditional distribution of $X_j \mid \{\kappa_j = k\}$ has a Lebesgue density $f^{(k)}$, which does not depend of j . We assume that the distributions of errors ε_j satisfy the following conditions:

1. $\mathbf{E} [\varepsilon_j \mid \kappa_j = k] = 0$,
2. $\mathbf{Var} [\varepsilon_j \mid \kappa_j = k] = \sigma_{(k)}^2 < \infty$.

2.2 Minimax weights

In this paper we consider a modified locally linear estimator for $g^{(m)}$ at a fixed point $x_0 \in \mathbb{R}$ introduced in [5]. This estimator utilizes minimax weights for the estimation of component distributions (see [7]). Let us recall the construction of these weights.

In what follows the angle brackets mean averaging of a vector:

$$\langle \mathbf{v} \rangle_n = \frac{1}{n} \sum_{j=1}^n v_j, \quad \text{for any } \mathbf{v} = (v_1, \dots, v_n)^T \in \mathbb{R}^n.$$

Arithmetic operations with vectors in the angle brackets are performed entry-wise:

$$\langle \mathbf{v} \mathbf{u} \rangle_n = \frac{1}{n} \sum_{j=1}^n v_j u_j.$$

Consider a set of concentration vectors $\mathbf{p}^{(m)} = (p_{1:n}^{(m)}, \dots, p_{n:n}^{(m)})^T$, $m = \overline{1, M}$. Observe that $\langle \mathbf{p}^{(m)} \mathbf{p}^{(k)} \rangle_n$ can be considered as an inner product on \mathbb{R}^n . Assuming that the concentration vectors $\{\mathbf{p}^{(m)}\}_{m=1}^M$ are linearly independent, the Gram matrix $\Gamma_n = (\langle \mathbf{p}^{(k)} \mathbf{p}^{(l)} \rangle_n)_{k,l=1}^M$ is invertible. The weighting coefficients $a_{j:n}^{(m)}$ defined by the formula

$$a_{j:n}^{(m)} = \frac{1}{\det \Gamma_n} \sum_{m=1}^M (-1)^{m+k} \gamma_{km} p_{j:n}^{(m)}, \quad (2)$$

where γ_{km} is the (k, m) -th minor of Γ_n , are called *minimax weighting coefficients*. These weights can also be obtained by the formula

$$(a_{j:n}^{(1)}, \dots, a_{j:n}^{(M)}) = (p_{j:n}^{(1)}, \dots, p_{j:n}^{(M)}) \Gamma_n^{-1}.$$

The vector of minimax coefficients for the m -th component will be denoted by $\mathbf{a}^{(m)} = (a_{1:n}^{(m)}, \dots, a_{n:n}^{(m)})^T$. Observe that

$$\langle \mathbf{p}^{(k)} \mathbf{a}^{(m)} \rangle_n = \begin{cases} 1, & k = m, \\ 0, & k \neq m, \end{cases} \quad \text{for all } m = \overline{1, M}. \quad (3)$$

2.3 Construction of an estimator

The modified local linear regression estimator (mLLRE) for $g^{(m)}(x_0)$ was introduced in [5] as a generalization of local linear regression to the data described by the model of regression mixture (1). To define it one needs to choose a kernel function $K : \mathbb{R} \rightarrow \mathbb{R}_+$ and a bandwidth $h > 0$. For any $p, q \in \mathbb{Z}_+$ let

$$S_{p,q;n}^{(m)} = \frac{1}{nh} \sum_{j=1}^n a_{j:n}^{(m)} K\left(\frac{x_0 - X_j}{h}\right) \left(\frac{x_0 - X_j}{h}\right)^p Y_j^q.$$

Then the mLLRE can be defined as

$$\hat{g}_n^{(m)}(x_0) = \frac{S_{2,0;n}^{(m)} S_{0,1;n}^{(m)} - S_{1,1;n}^{(m)} S_{1,0;n}^{(m)}}{S_{2,0;n}^{(m)} S_{0,0;n}^{(m)} - (S_{1,0;n}^{(m)})^2}. \quad (4)$$

3 Asymptotic normality of an estimator

To formulate the result on the asymptotic normality of mLLRE we need some notations and definitions.

The symbol \xrightarrow{W} means weak convergence.

In what follows, the one-sided limits of a function $f(x)$ at a point x_0 are denoted by

$$f(x_0-) = \lim_{x \rightarrow x_0-0} f(x), \quad f(x_0+) = \lim_{x \rightarrow x_0+0} f(x),$$

assuming that these limits exist. With this notation, we define

$$I_d^{(k),-} = f^{(k)}(x_0+) \int_{-\infty}^0 z^d (K(z))^2 dz, \quad I_d^{(k),+} = f^{(k)}(x_0-) \int_0^{+\infty} z^d (K(z))^2 dz.$$

$$I_d^{(k)} = I_d^{(k),+} + I_d^{(k),-},$$

$$I_{d_x, d_y}^{(k)} = \begin{cases} I_{d_x}^{(k)}, & d_y = 0, \\ (I_{d_x}^{(k),+} g^{(k)}(x_0-) + I_{d_x}^{(k),-} g^{(k)}(x_0+)), & d_y = 1, \\ (I_{d_x}^{(k),+} ((g^{(k)}(x_0-))^2 + \sigma_{(k)}^2) + I_{d_x}^{(k),-} ((g^{(k)}(x_0+))^2 + \sigma_{(k)}^2)), & d_y = 2, \end{cases}$$

$$\Sigma_{d_x, d_y}^{(m)} = \sum_{k=1}^M \langle (\mathbf{a}^{(m)})^2 \mathbf{p}^{(k)} \rangle I_{d_x, d_y}^{(k)},$$

$$\Sigma^{(m)} = \begin{pmatrix} \Sigma_{0;0}^{(m)} & \Sigma_{0;1}^{(m)} & \Sigma_{1;0}^{(m)} & \Sigma_{1;1}^{(m)} & \Sigma_{2;0}^{(m)} \\ \Sigma_{0;1}^{(m)} & \Sigma_{0;2}^{(m)} & \Sigma_{1;1}^{(m)} & \Sigma_{1;2}^{(m)} & \Sigma_{2;1}^{(m)} \\ \Sigma_{1;0}^{(m)} & \Sigma_{1;1}^{(m)} & \Sigma_{2;0}^{(m)} & \Sigma_{2;1}^{(m)} & \Sigma_{3;0}^{(m)} \\ \Sigma_{1;1}^{(m)} & \Sigma_{1;2}^{(m)} & \Sigma_{2;1}^{(m)} & \Sigma_{2;2}^{(m)} & \Sigma_{3;1}^{(m)} \\ \Sigma_{2;0}^{(m)} & \Sigma_{2;1}^{(m)} & \Sigma_{3;0}^{(m)} & \Sigma_{3;1}^{(m)} & \Sigma_{4;0}^{(m)} \end{pmatrix},$$

$$\begin{aligned}
u_p^- &= \int_{-\infty}^0 z^p K(z) dz, & u_p^+ &= \int_0^{+\infty} z^p K(z) dz, \\
u_p &= u_p^- + u_p^+, \\
e_{p_x, p_y}^{(m)} &= (g^{(m)}(x_0))^{p_y} \cdot (f^{(m)}(x_0^-)u_{p_x}^+ + f^{(m)}(x_0^+)u_{p_x}^-), \quad p_y \in \{0, 1\}.
\end{aligned}$$

Now we are ready to formulate our main result on the asymptotic behavior of mLLRE.

Theorem 1. *Assume that the following conditions hold.*

1. For all $k = \overline{1, M}$, there exist $f^{(k)}(x_0^\pm)$, $g^{(k)}(x_0^\pm)$.
2. $g^{(m)}$ is twice continuously differentiable in some neighbourhood B of x_0 .
3. For all $k, k_1, k_2 = \overline{1, M}$ the limits

$$\begin{aligned}
\langle (\mathbf{a}^{(m)})^2 \mathbf{p}^{(k)} \rangle &:= \lim_{n \rightarrow +\infty} \langle (\mathbf{a}^{(m)})^2 \mathbf{p}^{(k)} \rangle_n, \\
\langle \mathbf{a}^{(m)} \mathbf{p}^{(k_1)} \mathbf{p}^{(k_2)} \rangle &:= \lim_{n \rightarrow \infty} \langle \mathbf{a}^{(m)} \mathbf{p}^{(k_1)} \mathbf{p}^{(k_2)} \rangle_n
\end{aligned}$$

exist and are finite.

4. There exists $\lim_{n \rightarrow \infty} \Gamma_n = \Gamma$, where $\Gamma_n = (\langle \mathbf{p}^{(k_1)} \mathbf{p}^{(k_2)} \rangle_n)_{k_1, k_2=1}^M$.
5. $h = h_n = Hn^{-1/5}$.
6. For some $A > 0$ and all z such that $|z| > A$, $K(z) = 0$.
7. Integrals $\int_{-\infty}^{\infty} |z|K(z)dz$ and $\int_{-\infty}^{\infty} z^4(K(z))^2 dz$ are finite.
8. $f^{(k)}(x)$, $g^{(k)}(x)$ are bounded for $x \in B$ for all $k = \overline{1, M}$.
9. $e_{2,0}^{(m)} e_{0,0}^{(m)} - (e_{1,0}^{(m)})^2 \neq 0$.
10. $\mathbf{E} \left[\varepsilon_j^4 \mid \kappa_j = k \right] < \infty$ for all $k = \overline{1, M}$.

Then

$$n^{2/5}(\hat{g}_n^{(m)}(x_0) - g^{(m)}(x_0)) \xrightarrow{W} N(\mu^{(m)}(x_0), S_{(m)}^2(x_0)), \quad (5)$$

where $\mu^{(m)}(x_0)$ and $S_{(m)}^2(x_0)$ are defined by

$$\begin{aligned}
\mu^{(m)}(x_0) &= H^2 \cdot \frac{\ddot{g}^{(m)}(x_0)}{2} \cdot \frac{(e_{2,0}^{(m)})^2 - e_{1,0}^{(m)} e_{3,0}^{(m)}}{e_{2,0}^{(m)} e_{0,0}^{(m)} - (e_{1,0}^{(m)})^2}, \\
S_{(m)}^2(x_0) &= \frac{1}{H} ((g^{(m)}(x_0))^2 \tilde{\Sigma}_0^{(m)} - 2(g^{(m)}(x_0)) \tilde{\Sigma}_1^{(m)} + \tilde{\Sigma}_2^{(m)}), \\
\tilde{\Sigma}_k^{(m)} &= \tilde{e}_{2,2}^{(m)} \Sigma_{0:k}^{(m)} - 2\tilde{e}_{1,2}^{(m)} \Sigma_{1:k}^{(m)} + \tilde{e}_{1,1}^{(m)} \Sigma_{2:k}^{(m)},
\end{aligned}$$

$$\tilde{e}_{p,q}^{(m)} = \frac{e_{p,0}^{(m)} e_{q,0}^{(m)}}{(e_{2,0}^{(m)} e_{0,0}^{(m)} - (e_{1,0}^{(m)})^2)^2},$$

where

$$\ddot{g}^{(m)}(x) = \frac{d^2 g^{(m)}(x)}{dx^2}.$$

4 Optimal bandwidth selection

The mLLRE $\hat{g}_n^{(m)}(x_0)$ defined by (4) depends on the bandwidth h , a tuning parameter that must be selected by the researcher to obtain an accurate estimator. The accuracy of $\hat{g}_n^{(m)}(x_0)$ usually is measured by the mean squared error

$$\text{MSE}(\hat{g}_n^{(m)}(x_0)) = \mathbf{E}[(\hat{g}_n^{(m)}(x_0) - g^{(m)}(x_0))^2] = \mathbf{Var}[\hat{g}_n^{(m)}(x_0)] + (\mathbf{bias}(\hat{g}_n^{(m)}(x_0)))^2,$$

where $\mathbf{bias}(\hat{g}_n^{(m)}(x_0)) = \mathbf{E}[\hat{g}_n^{(m)}(x_0)] - g^{(m)}(x_0)$ is the estimator's bias. In Theorem 1 we considered the choice $h = h_n = Hn^{-1/5}$, where H is some fixed constant. This rate of convergence for the bandwidth as $n \rightarrow \infty$ is optimal, since if h_n vanishes slower, the estimator has inappropriately high bias, while for more rapid h_n decay the variance of the estimator would be inappropriate. So, we need to choose the best constant H . By Theorem 1, $n^{2/5}(\hat{g}_n^{(m)}(x_0) - g^{(m)}(x_0))$ converges weakly to $\eta \sim N(\mu^{(m)}(x_0), S_{(m)}^2(x_0))$, so we will measure the asymptotic accuracy of $\hat{g}_n^{(m)}(x_0)$ by the asymptotic MSE (aMSE):

$$\text{aMSE}(H) = \mathbf{E}[\eta^2] = \left(\mu^{(m)}(x_0)\right)^2 + S_{(m)}^2(x_0) = H^4 \cdot E_{(m)}^2 + \frac{1}{H} \cdot V_{(m)},$$

where

$$E_{(m)} = \frac{\ddot{g}^{(m)}(x_0)}{2} \cdot \frac{(e_{2,0}^{(m)})^2 - e_{1,0}^{(m)} e_{3,0}^{(m)}}{e_{2,0}^{(m)} e_{0,0}^{(m)} - (e_{1,0}^{(m)})^2},$$

$$V_{(m)} = (g^{(m)}(x_0))^2 \tilde{\Sigma}_0^{(m)} - 2(g^{(m)}(x_0)) \tilde{\Sigma}_1^{(m)} + \tilde{\Sigma}_2^{(m)}.$$

An optimal bandwidth constant, which minimizes aMSE, is

$$H_*^{(m)} = \left(\frac{V_{(m)}}{4E_{(m)}^2} \right)^{1/5}.$$

Observe that $H_*^{(m)}$ cannot be calculated by the data, since it depends on unknown distributions of the mixture components. So it is an infeasible theoretically optimal bandwidth constant which can be used in comparisons to some empirical bandwidth selection rules.

5 Proofs

The proof of Theorem 1 is based on two lemmas.

Let

$$\begin{aligned} \mathbf{S}_n^{(m)} &= (S_{0,0}^{(m)}, S_{0,1}^{(m)}, S_{0,2}^{(m)}, S_{1,0}^{(m)}, S_{2,0}^{(m)})^T, \quad \mathbf{e}_n^m = \mathbf{E}[\mathbf{S}_n^{(m)}]. \\ \Delta_n^{(m)} &= \sqrt{nh}(\mathbf{S}_n^{(m)} - \mathbf{e}_n^{(m)}). \end{aligned} \quad (6)$$

Lemma 1. *Under Assumptions 1–4 and 6–10 of Theorem 1, if $h = h_n \rightarrow 0$ and $nh_n \rightarrow \infty$ as $n \rightarrow \infty$, then*

$$\Delta_n^{(m)} \xrightarrow{W} \Delta_\infty^{(m)} \sim N(\mathbf{0}, \Sigma^{(m)}). \quad (7)$$

For any $\mathbf{a} = (a_{0,0}, a_{0,1}, a_{1,0}, a_{1,1}, a_{2,0})^T \in \mathbb{R}^5$ let

$$\mathbf{U}(\mathbf{a}) = \frac{a_{2,0}a_{0,1} - a_{1,1}a_{1,0}}{a_{2,0}a_{0,0} - a_{1,0}^2}, \quad (8)$$

Lemma 2. *Under the assumptions of Theorem 1*

$$\mathbf{U}(\mathbf{e}_n^{(m)}) - g^{(m)}(x_0) = \frac{h^2}{2} \cdot \ddot{g}^{(m)}(x_0) \cdot \frac{(e_{2,0}^{(m)})^2 - e_{1,0}^{(m)} e_{3,0}^{(m)}}{e_{2,0}^{(m)} e_{0,0}^{(m)} - (e_{1,0}^{(m)})^2} + o(h^2), \quad n \rightarrow \infty,$$

Proof of Theorem 1. Consider

$$n^{2/5}(\hat{g}^{(m)}(x_0) - g^{(m)}(x_0)) = n^{2/5}(\mathbf{U}(\mathbf{S}_n^{(m)}) - \mathbf{U}(\mathbf{e}_n^{(m)})) + n^{2/5}(\mathbf{U}(\mathbf{e}_n^{(m)}) - g^{(m)}(x_0)). \quad (9)$$

Lemma 1 and the continuous mapping theorem (see Theorem 3.1 in [1]) yield

$$n^{2/5}(\mathbf{U}(\mathbf{S}_n^{(m)}) - \mathbf{U}(\mathbf{e}_n^{(m)})) \xrightarrow{W} \frac{1}{\sqrt{H}} \dot{\mathbf{U}}^T(\mathbf{e}_n^{(m)}) \Delta_\infty^{(m)},$$

where

$$\dot{\mathbf{U}}(\mathbf{a}) = \left(\frac{d}{da_{0,0}} \mathbf{U}(\mathbf{a}), \frac{d}{da_{0,1}} \mathbf{U}(\mathbf{a}), \frac{d}{da_{1,0}} \mathbf{U}(\mathbf{a}), \frac{d}{da_{1,1}} \mathbf{U}(\mathbf{a}), \frac{d}{da_{2,0}} \mathbf{U}(\mathbf{a}) \right)^T$$

is the gradient of \mathbf{U} . Tedious but straightforward algebra yields

$$\mathbf{Var} \left[\frac{1}{\sqrt{H}} \dot{\mathbf{U}}^T(\mathbf{e}_n^{(m)}) \Delta_\infty^{(m)} \right] = \frac{1}{H} \dot{\mathbf{U}}(\mathbf{e}_n^{(m)})^T \Sigma^{(m)} \dot{\mathbf{U}}(\mathbf{e}_n^{(m)}) = S_{(m)}^2(x_0). \quad (10)$$

By Lemma 2 for $h = Hn^{-1/5}$,

$$n^{2/5}(\mathbf{U}(\mathbf{e}_n^{(m)}) - g^{(m)}(x_0)) \rightarrow \frac{H^2}{2} \cdot \ddot{g}^{(m)}(x_0) \cdot \frac{(e_{2,0}^{(m)})^2 - e_{1,0}^{(m)} e_{3,0}^{(m)}}{e_{2,0}^{(m)} e_{0,0}^{(m)} - (e_{1,0}^{(m)})^2}. \quad (11)$$

Combining (9)–(11) one obtains the statement of Theorem 1. \square

To demonstrate Lemma 1 we need the Lindeberg–Feller central limit theorem.

Lemma 3. (Lindeberg’s CLT) *Let $\{\eta_{j:n}\}_{j=1}^n$, $n = 1, 2, \dots$, be a set of random vectors in \mathbb{R}^d , satisfying the following assumptions.*

1. *For any fixed $n \geq 1$, vectors $\{\eta_{j:n}\}_{j=1}^n$ are mutually independent.*
2. *For all $j = \overline{1, n}$, $n \geq 1$, one has $\mathbf{E}[\eta_{j:n}] = \mathbf{0}$.*
3. *If $\sigma_{j:n}^2 = \mathbf{Cov}(\eta_{j:n})$, then for $\sigma_n^2 = \sum_{j=1}^n \sigma_{j:n}^2$ there exists*

$$\sigma^2 = \lim_{n \rightarrow \infty} \sigma_n^2,$$

moreover σ^2 is a positive semidefinite matrix.

4. *For some $s > 2$ the following convergence holds:*

$$M_2(s) = \sum_{j=1}^n \mathbf{E}[\min(|\eta_{j:n}|^2, |\eta_{j:n}|^s)] \rightarrow 0.$$

Then

$$\sum_{j=1}^n \eta_{j:n} \xrightarrow{W} N(\mathbf{0}, \sigma^2).$$

For the proof, see [1], Theorem 8.4.1.

Proof of Lemma 1. To simplify notations, we introduce formally random vectors $(X_{(m)}, Y_{(m)}, \varepsilon_{(m)})$ with the distribution of $(X_j, Y_j, \varepsilon_j)$ given $\kappa_j = m$.

Conditions of Lemma 3 will be verified for $\{\eta_{j:n}^{(m)}\}$, where

$$\begin{aligned} \eta_{j:n}^{(m)} &= a_{j:n}^{(m)} \cdot \tilde{\eta}'_{j:n}, \quad \tilde{\eta}'_{j:n} = (\tilde{\eta}_{j:n} - \mathbf{E}[\tilde{\eta}_{j:n}]), \\ \tilde{\eta}_{j:n} &= \frac{1}{\sqrt{nh}} \cdot K\left(\frac{x_0 - X_j}{h}\right) \left(1, Y_j, \left(\frac{x_0 - X_j}{h}\right), \left(\frac{x_0 - X_j}{h}\right) Y_j, \left(\frac{x_0 - X_j}{h}\right)^2\right)^T. \end{aligned}$$

Similarly we define random variables $\tilde{\eta}_{(m):n}$ ($\tilde{\eta}'_{(m):n}$) that have a distribution of $\tilde{\eta}_{j:n}$ ($\tilde{\eta}'_{j:n}$) given $\kappa_j = m$. Obviously, $\Delta_n^{(m)} = \sum_{j=1}^n \eta_{j:n}^{(m)}$.

The first condition of Lemma 3 holds since (X_j, Y_j) are independent for different j . The second condition follows from the construction of $\eta_{j:n}^{(m)}$.

From now, we will proceed to the third condition of Lemma 3. For some p_x, p_y, q_x, q_y , consider

$$\mathbf{Cov}(S_{p_x, p_y}^{(m)}, S_{q_x, q_y}^{(m)}) = \Sigma_{p_x, p_y: q_x, q_y}^{(m)}(n) = Q_1^{(m)}(n, h) - Q_2^{(m)}(n, h),$$

$$Q_1^{(m)}(n, h) = \frac{1}{nh} \sum_{j=1}^n (a_{j:n}^{(m)})^2 \mathbf{E} \left[\left(K\left(\frac{x_0 - X_j}{h}\right) \right)^2 \left(\frac{x_0 - X_j}{h}\right)^{p_x + q_x} Y_j^{p_y + q_y} \right],$$

$$Q_2^{(m)}(n, h) = \frac{1}{nh} \sum_{j=1}^n (a_{j:n}^{(m)})^2 \mathbf{E} \left[K \left(\frac{x_0 - X_j}{h} \right) \left(\frac{x_0 - X_j}{h} \right)^{p_x} Y_j^{p_y} \right] \\ \times \mathbf{E} \left[K \left(\frac{x_0 - X_j}{h} \right) \left(\frac{x_0 - X_j}{h} \right)^{q_x} Y_j^{q_y} \right].$$

We will investigate $Q_1^{(m)}(n, h)$ and $Q_2^{(m)}(n, h)$ separately. First of all, note that

$$Q_1^{(m)}(n, h) = \frac{1}{h} \sum_{k=1}^M \langle (\mathbf{a}^{(m)})^2 \mathbf{p}^{(k)} \rangle_n \mathbf{E} \left[\left(K \left(\frac{x_0 - X_{(k)}}{h} \right) \right)^2 \left(\frac{x_0 - X_{(k)}}{h} \right)^{p_x + q_x} Y_{(k)}^{p_y + q_y} \right].$$

Consider the expectations in the sum and denote $d_x = p_x + q_x$ and $d_y = p_y + q_y \leq 4$. Then, for all $k = \overline{1, M}$,

$$\frac{1}{h} \mathbf{E} \left[\left(K \left(\frac{x_0 - X_{(k)}}{h} \right) \right)^2 \left(\frac{x_0 - X_{(k)}}{h} \right)^{p_x + q_x} Y_{(k)}^{p_y + q_y} \right] \\ = \sum_{l=0}^{d_y} \binom{d_y}{l} \cdot \mathbf{E} \left[\varepsilon_{(k)}^{d_y - l} \right] \cdot \int_{-\infty}^{+\infty} (K(z))^2 z^{d_x} (g^{(k)}(x_0 - hz))^l f^{(k)}(x_0 - hz) dz,$$

where $\binom{n}{k} = n!/(k!(n-k)!)$ is the binomial coefficient. By Assumptions 1, 6 and 7 we obtain

$$\int_{-\infty}^{+\infty} (K(z))^2 z^{d_x} (g^{(k)}(x_0 - hz))^l f^{(k)}(x_0 - hz) dz \\ \rightarrow (g^{(k)}(x_0 -)) ^l I_{d_x}^{(k),+} + (g^{(k)}(x_0 +)) ^l I_{d_x}^{(k),-}$$

as $n \rightarrow \infty$. So, for $d_x \in \{0, 1, 2, 3, 4\}$ and $d_y \in \{0, 1, 2\}$,

$$\frac{1}{h} \mathbf{E} \left[\left(K \left(\frac{x_0 - X_{(k)}}{h} \right) \right)^2 \left(\frac{x_0 - X_{(k)}}{h} \right)^{d_x} Y_{(k)}^{d_y} \right] \rightarrow I_{d_x, d_y}^{(k)}, \quad n \rightarrow \infty.$$

From the assumption $\mathbf{E}[\varepsilon_{(k)}] = 0$ and Assumption 4, we obtain

$$Q_1^{(m)}(n, h) \rightarrow \sum_{k=1}^M \langle (\mathbf{a}^{(m)})^2 \mathbf{p}^{(k)} \rangle I_{d_x, d_y}^{(k)}, \quad n \rightarrow \infty.$$

Now we will show that $Q_2^{(m)}(n, h) \rightarrow 0$ as $n \rightarrow \infty$. Note that

$$Q_2^{(m)}(n, h) = h \sum_{k_1, k_2=1}^M \langle (\mathbf{a}^{(m)})^2 p^{(k_1)} p^{(k_2)} \rangle_n Q_{p_x, p_y}^{(k_1)}(n, h) Q_{q_x, q_y}^{(k_2)}(n, h), \\ Q_{p_x, p_y}^{(k)}(n, h) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} K(z) z^{p_x} (g^{(k)}(x_0 - hz) + u)^{p_y} f^{(k)}(x_0 - hz) dz dF_{\varepsilon}^{(k)}(u),$$

where $F_{\varepsilon}^{(k)}(u) = \mathbf{P}(\varepsilon_{(k)} < u)$ is the cumulative distribution function of $\varepsilon_{(k)}$. The multiple integrals $Q_{p_x, p_y}^{(k)}(n, h)$ are bounded for $n \geq 1$, thus from Assumption 3 we obtain $Q_2^{(m)}(n, h) \rightarrow 0$ as $n \rightarrow \infty$.

Combining the asymptotics of $Q_1^{(m)}(n, h)$ and $Q_2^{(m)}(n, h)$ as $n \rightarrow \infty$, we obtain the asymptotics of covariances for $\Delta_n^{(m)}$:

$$\Sigma_{p_x, p_y; q_x, q_y}^{(m)}(n) \rightarrow \sum_{k=1}^M \langle (\mathbf{a}^{(m)})^2 \mathbf{p}^{(k)} \rangle I_{p_x+q_x, p_y+q_y}^{(k)} = \Sigma_{p_x+q_x; p_y+q_y}^{(m)}, \quad n \rightarrow \infty.$$

The third condition of Lemma 3 holds.

Finally we will show that the fourth condition of Lemma 3 holds. For some $s > 2$, note that

$$\begin{aligned} M_2(s) &= \sum_{j=1}^n \mathbf{E} [\min(|\eta_{j:n}|^2, |\eta_{j:n}|^s)] \\ &= \sum_{j=1}^n \sum_{k=1}^M p_{j:n}^{(k)} \mathbf{E} [\min(|a_{j:n}^{(m)}|^2 \cdot |\tilde{\eta}'_{(k):n}|^2, |a_{j:n}^{(m)}|^s \cdot |\tilde{\eta}'_{(k):n}|^s)] \\ &\leq C_\Gamma \cdot \sum_{j=1}^n \sum_{k=1}^M \mathbf{E} [\min(|\tilde{\eta}'_{(k):n}|^2, |\tilde{\eta}'_{(k):n}|^s)], \end{aligned}$$

since $p_{j:n}^{(m)} \leq 1$, $|a_{j:n}^{(m)}|^2 \leq \max(1, \sup_{j=1, \dots, n} |a_{j:n}^{(m)}|^s) = C_\Gamma < \infty$. By the inequality

$$|\mathbf{a} + \mathbf{b}|^s \leq 2^{s-1} (|\mathbf{a}|^s + |\mathbf{b}|^s), \quad \text{for any } \mathbf{a}, \mathbf{b} \in \mathbb{R}^d, \quad (12)$$

we obtain

$$\begin{aligned} &\sum_{j=1}^n \sum_{k=1}^M \mathbf{E} [\min(|\tilde{\eta}'_{(k):n}|^2, |\tilde{\eta}'_{(k):n}|^s)] \\ &\leq 2^{s-1} \cdot \sum_{k=1}^M n \cdot (\mathbf{E} [\min(|\tilde{\eta}_{(k):n}|^2, |\tilde{\eta}_{(k):n}|^s)] + \max(|\mathbf{E} [\tilde{\eta}_{(k):n}]|^2, |\mathbf{E} [\tilde{\eta}_{(k):n}]|^s)). \end{aligned}$$

We will show that, as $n \rightarrow \infty$,

$$n \cdot \mathbf{E} [\min(|\tilde{\eta}_{(k):n}|^2, |\tilde{\eta}_{(k):n}|^s)] \rightarrow 0, \quad (13)$$

$$n \cdot \max(|\mathbf{E} [\tilde{\eta}_{(k):n}]|^2, |\mathbf{E} [\tilde{\eta}_{(k):n}]|^s) \rightarrow 0. \quad (14)$$

Let us show (14). Observe that for any $p \geq 2$

$$n \cdot |\mathbf{E} [\tilde{\eta}_{(k):n}]|^p \rightarrow 0, \quad n \rightarrow \infty. \quad (15)$$

Really, the left-hand side of (15) can be expressed as follows:

$$n \cdot |\mathbf{E} [\tilde{\eta}_{(k):n}]|^p = n \cdot \left(E_{0,0}^{(k)} + E_{0,1}^{(k)} + E_{1,0}^{(k)} + E_{1,1}^{(k)} + E_{2,0}^{(k)} \right)^{p/2},$$

where $E_{p_x, p_y}^{(k)} = ((nh)^{-1/2} \cdot \mathbf{E}[K((x_0 - X_{(k)})/h)((x_0 - X_{(k)})/h)^{p_x} Y_{(k)}^{p_y}])^2$.

For instance,

$$E_{2,0}^{(k)} \leq \frac{h}{n} \cdot \left(\int_{-\infty}^{+\infty} z^2 K(z) f^{(k)}(x_0 - hz) dz \right)^2 \sim \frac{h}{n} \cdot C_{2,0}^{(m)},$$

as $n \rightarrow \infty$, where

$$C_{2,0}^{(m)} = \left(f^{(k)}(x_0+)u_2^- + f^{(k)}(x_0-)u_2^+ \right)^2.$$

By similar reasoning for the other terms we obtain

$$n \cdot |\mathbf{E}[\tilde{\eta}_{(k):n}]|^p \leq n \cdot \left(\frac{h}{n} \right)^{p/2} \cdot C^{(k)}, \quad (16)$$

for some $C^{(k)} < \infty$.

By Assumption 4,

$$n^{1-p/2} \cdot h^{p/2} \rightarrow 0, \quad n \rightarrow \infty, \quad (17)$$

since $p/2 \geq 1$. Then (16) and (17) yield (14).

To show (13), observe that for any $\tau > 0$

$$\mathbf{E} \left[\min(|\tilde{\eta}_{(k):n}|^2, |\tilde{\eta}_{(k):n}|^s) \right] \leq Z_n(\tau) + \tau^{s-2} Z_n(0), \quad (18)$$

where

$$Z_n(\tau) = \mathbf{E} \left[|\tilde{\eta}_{(k):n}|^2 \mathbf{1}_{\{|\tilde{\eta}_{(k):n}| \geq \tau\}} \right]. \quad (19)$$

We will show that $Z_n(0)$ is bounded and $Z_n(\tau) \rightarrow 0$ as $n \rightarrow \infty$ for any $\tau > 0$. So, taking τ small enough we can make the right-hand side of (18) as small as desired.

Let

$$V(z, x, u) = K^2(z)(1 + (g^{(k)}(x) + u)^2 + z^2 + z^2(g^{(k)}(x) + u)^2 + z^4).$$

Then

$$\begin{aligned} Z_n(\tau) &= \frac{1}{h} \mathbf{E} \left[V \left(\frac{x_0 - X_{(k)}}{h}, X_{(k)}, \varepsilon_{(k)} \right) \mathbf{1} \left\{ V \left(\frac{x_0 - X_{(k)}}{h}, X_{(k)}, \varepsilon_{(k)} \right) > \tau^2 nh \right\} \right] \\ &= \frac{1}{h} \int_{-A}^A \mathbf{E} \left[V \left(\frac{x_0 - x}{h}, x, \varepsilon_{(k)} \right) \mathbf{1} \left\{ V \left(\frac{x_0 - x}{h}, x, \varepsilon_{(k)} \right) > \tau^2 nh \right\} \right] f^{(k)}(x) dx \\ &= \int_{-A}^A \mathbf{E} \left[V(z, x_0 - hz, \varepsilon_{(k)}) \mathbf{1} \left\{ V(z, x_0 - hz, \varepsilon_{(k)}) > \tau^2 nh \right\} \right] f^{(k)}(x_0 - hz) dz. \end{aligned}$$

By Assumption 8, $g^{(k)}$ and $f^{(k)}$ are bounded in a neighborhood B of x_0 . For n large enough $[x_0 - hA, x_0 + hA] \in B$, so for $-A \leq z \leq A$,

$$V(z, x_0 - hz, u) f^{(k)}(x_0 - hz) \leq \bar{V}(z, u), \quad (20)$$

where

$$\bar{V}(z, u) = \bar{f} K^2(z)(1 + (\bar{g} + |u|)^2 + z^2 + z^2(\bar{g} + |u|)^2 + z^4),$$

$$\bar{f} = \sup_{x \in B} f^{(k)}(x), \quad \bar{g} = \sup_{x \in B} |g^{(k)}(x)|.$$

By Assumptions 7 and 10,

$$\int_{-A}^A \mathbf{E} [\bar{V}(z, \varepsilon(k))] dz = V^* < \infty. \quad (21)$$

So, by (20), $Z_n(0) < V^*$.

Observe that

$$\mathbf{1} \left\{ V \left(\frac{x_0 - x}{h}, x, u \right) > \tau^2 nh \right\} \rightarrow 0$$

as $n \rightarrow \infty$, since $nh \rightarrow \infty$ by Assumption 5. So, with (20) and (21) in mind, by the Lebesgue dominated convergence theorem we obtain $Z_n(\tau) \rightarrow 0$ as $n \rightarrow \infty$ for any $\tau > 0$.

Thus, for any $\delta > 0$ we can take $\tau > 0$ so small that $\tau^{s-2} Z_n(0) < \tau^{s-2} V^* < \delta/2$ and then n_0 so large that $Z_n(\tau) < \delta/2$ for $n > n_0$. By (18) this yields $M_2(s) \rightarrow 0$ as $n \rightarrow \infty$. So Assumption 4 of Lemma 3 holds.

Applying Lemma 3 to $\eta_{j:n}^{(m)}$ we obtain the statement of Lemma 1. \square

Proof of Lemma 2. Consider $c_n^{(m)} = e_{2,0:n}^{(m)} e_{0,1:n}^{(m)} - e_{1,1:n}^{(m)} e_{1,0:n}^{(m)}$ and $d_n^{(m)} = e_{2,0:n}^{(m)} e_{0,0:n}^{(m)} - (e_{1,0:n}^{(m)})^2$, where

$$e_{p_x, p_y:n}^{(m)} = \mathbf{E}[S_{p_x, p_y}^{(m)}] = \int_{-\infty}^{+\infty} K(z) z^{p_x} g^{(m)}(x_0 - hz) f^{(m)}(x_0 - hz) dz. \quad (22)$$

By continuity of \mathbf{U} and convergence $\mathbf{e}_n^{(m)} \rightarrow \mathbf{e}^{(m)}$, we get $\mathbf{U}(\mathbf{e}_n^{(m)}) \rightarrow \mathbf{U}(\mathbf{e}^{(m)}) = g^{(m)}(x_0)$.

We will examine the rate of convergence to zero for the difference

$$\mathbf{U}(\mathbf{e}_n^{(m)}) - g^{(m)}(x_0) = \frac{c_n^{(m)} - d_n^{(m)} g^{(m)}(x_0)}{d_n^{(m)}}. \quad (23)$$

From (22) one obtains

$$c_n^{(m)} = \int_{-\infty}^{+\infty} g^{(m)}(x_0 - hz) K(z) (e_{2,0:n}^{(m)} - z e_{1,0:n}^{(m)}) f^{(m)}(x_0 - hz) dz \quad (24)$$

and

$$d_n^{(m)} = \int_{-\infty}^{+\infty} K(z) (e_{2,0:n}^{(m)} - z e_{1,0:n}^{(m)}) f^{(m)}(x_0 - hz) dz. \quad (25)$$

By Taylor's expansion for $g^{(m)}$ in the neighborhood of x_0 , we obtain, as $n \rightarrow \infty$,

$$c_n^{(m)} - d_n^{(m)} g^{(m)}(x_0)$$

$$\begin{aligned}
&= \int_{-\infty}^{+\infty} (g^{(m)}(x_0 - hz) - g^{(m)}(x_0)) K(z) (e_{2,0:n}^{(m)} - ze_{1,0:n}^{(m)}) f^{(m)}(x_0 - hz) dz \\
&= \dot{g}^{(m)}(x_0)(-h) \int_{-A}^A z K(z) (e_{2,0:n}^{(m)} - ze_{1,0:n}^{(m)}) f^{(m)}(x_0 - hz) dz \\
&\quad + \frac{h^2}{2} \cdot \ddot{g}^{(m)}(x_0) \int_{-A}^A z^2 K(z) (e_{2,0:n}^{(m)} - ze_{1,0:n}^{(m)}) f^{(m)}(x_0 - hz) dz \\
&\quad + \int_{-A}^A R(hz) K(z) (e_{2,0:n}^{(m)} - ze_{1,0:n}^{(m)}) f^{(m)}(x_0 - hz) dz =: J_{1:n}^{(m)} + J_{2:n}^{(m)} + J_{3:n}^{(m)},
\end{aligned}$$

where $R(t)$ is some function, such that $|R(t)|/t^2 \rightarrow 0$ as $t \rightarrow 0$. By (22),

$$J_{1:n}^{(m)} = 0 \quad (26)$$

and

$$J_{2:n}^{(m)} = \frac{h^2}{2} \cdot \ddot{g}^{(m)}(x_0) ((e_{2,0:n}^{(m)})^2 - e_{1,0:n}^{(m)} e_{3,0:n}^{(m)}). \quad (27)$$

The asymptotics of

$$J_{3:n}^{(m)} = e_{2,0:n}^{(m)} J_{R,0:n}^{(m)} - e_{2,0:n}^{(m)} J_{R,1:n}^{(m)} \quad (28)$$

remains to be examined as $n \rightarrow \infty$, where

$$J_{R,p:n}^{(m)} = \int_{-A}^A R(hz) z^p K(z) f^{(m)}(x_0 - hz) dz.$$

where $p \in \{0, 1\}$. Since $e_{p,q:n}^{(m)} \rightarrow e_{p,q}^{(m)}$, it suffices to investigate the asymptotics of $J_{R,p:n}^{(m)}$.

Note that for $J_{R,p:n}^{(m),+} = \int_0^A R(hz) z^p K(z) f^{(m)}(x_0 - hz) dz$,

$$J_{R,p:n}^{(m),+} \sim f^{(m)}(x_0) \int_0^A R(hz) z^p K(z) dz, \quad n \rightarrow \infty,$$

For any $\varepsilon \in (0, 1)$, there exists such $N(\varepsilon)$ that $|R(t)| \leq \varepsilon t^2$, $n \geq N(\varepsilon)$. For $n \geq N(\varepsilon)$,

$$\left| \int_0^A R(hz) z^p K(z) dz \right| \leq \varepsilon h^2 \int_0^A z^{2+p} K(z) dz = o(h^2), \quad n \rightarrow \infty. \quad (29)$$

Similarly

$$J_{R,p:n}^{(m),-} = \int_{-A}^0 R(hz) z^p K(z) f^{(m)}(x_0 - hz) dz = o(h^2), \quad n \rightarrow \infty.$$

Thus,

$$J_{R,p:n}^{(m)} = J_{R,p:n}^{(m),+} + J_{R,p:n}^{(m),-} = o(h^2), \quad n \rightarrow \infty. \quad (30)$$

From (28) and (30), we get

$$J_{3;n}^{(m)} = o(h^2), \quad n \rightarrow \infty \quad (31)$$

From (26), (27) and (31), we obtain the statement of Lemma 2. \square

6 Simulations

6.1 Description of simulations

For simulations we considered a mixture of regressions with $M = 2$ components. The concentrations $\{p_{j:n}^{(m)}\}$ were defined by

$$p_{j:n}^{(1)} = \frac{j}{n}, \quad p_{j:n}^{(2)} = 1 - \frac{j}{n}, \quad j = \overline{1, n}.$$

The distribution of regressor X_j was the same for both components. Its PDF was

$$f(t) = \frac{3}{2} \cdot \mathbf{1}_{(0,1/2]}(t) + \frac{1}{2} \cdot \mathbf{1}_{(1/2,1)}(t), \quad t \in \mathbb{R}.$$

The distribution of ε_j was different for different experiments. Regression functions were defined as

$$g^{(m)}(t) = \begin{cases} \sin(2\pi t), & m = 1, \\ \cos(2\pi t), & m = 2. \end{cases}$$

Estimation was performed at $x_0 = 1/2$, which is a discontinuity point of $f(t)$. The simulation procedure was as follows:

1. For each sample size $n \in \{100, 500, 1000, 5000, 10000\}$, we generate $B = 1000$ copies of $\{(X_j, Y_j)\}_{j=\overline{1, n}}$ from the described model.
2. In each copy, the modified local linear regression estimator $\hat{g}_n^{(m)}(x_0)$ is computed at x_0 for each $m = \overline{1, M}$.
3. Having an array with B values of $\hat{g}_n^{(m)}(x_0)$, we compute sample bias and standard deviation

$$\begin{aligned} \text{Bias}_n^{(m)} &= n^{2/5} (\mathbf{E}_*[\hat{g}_n^{(m)}(x_0)] - g^{(m)}(x_0)), \\ \text{Std}_n^{(m)} &= n^{2/5} \cdot \sqrt{\mathbf{Var}_*[\hat{g}_n^{(m)}(x_0)]}. \end{aligned}$$

For $\hat{g}_n^{(m)}(x_0)$ we select an optimal bandwidth $h = H_*^{(m)} n^{-1/5}$ and the Epanechnikov kernel

$$K(t) = \mathbf{1}\{|t| < 1\} \cdot \frac{3}{4} \cdot (1 - t^2).$$

In this scenario, $H_*^{(1)}$ does not exist since $E_{(m)} = 0$. So, in the experiments, we let $H^{(1)} = H_*^{(2)}$.

Table 1. Computed values of sample bias and standard deviation for each n and m

n	$m = 1$		$m = 2$	
	$\text{Bias}_n^{(m)}$	$\text{Std}_n^{(m)}$	$\text{Bias}_n^{(m)}$	$\text{Std}_n^{(m)}$
100	0.7239	13.021	-3.5882	149.4207
500	0.2935	2.6544	1.0857	2.692
1000	0.0979	2.5753	1.3031	2.6607
5000	0.1256	2.6908	1.3599	2.6578
10000	-0.0103	2.5745	1.2657	2.5809
∞	0	2.6547	1.3274	2.6547

Table 2. Computed values of sample bias and standard deviation for each n and m

n	$m = 1$		$m = 2$	
	$\text{Bias}_n^{(m)}$	$\text{Std}_n^{(m)}$	$\text{Bias}_n^{(m)}$	$\text{Std}_n^{(m)}$
100	0.1082	4.4768	0.6359	6.724
500	0.1659	3.059	0.9511	3.0376
1000	0.1577	3.0195	1.1993	3.0792
5000	0.1386	2.9185	1.1616	3.012
10000	0.2131	2.8123	1.2613	2.9326
∞	0	2.9282	1.4641	2.9282

Table 3. Computed values of sample bias and standard deviation for each n and m

n	$m = 1$		$m = 2$	
	$\text{Bias}_n^{(m)}$	$\text{Std}_n^{(m)}$	$\text{Bias}_n^{(m)}$	$\text{Std}_n^{(m)}$
100	0.0616	4.5498	0.9148	5.3876
500	0.2661	2.8404	1.1379	2.9482
1000	0.2239	2.7614	1.1546	2.9566
5000	-0.0964	2.6225	1.4349	2.8296
10000	0.0435	2.626	1.3573	2.9192
∞	0	2.6938	1.4317	2.8635

6.2 Performance of mLLE

Experiment 1. In this experiment,, we consider $\varepsilon_j \sim N(0, 1.25)$. The results of Experiment 1 for mLLE are presented in Table 1.

Here $H_*^{(2)} \approx 0.6261$. The simulation results show the agreement with the asymptotic considerations for large n .

Experiment 2. In this experiment, we consider $\varepsilon_j \sim T_5$, where T_5 is the Student T distribution with 5 degrees of freedom. The results of Experiment 2 for mLLE are presented in Table 2.

Here $H_*^{(2)} \approx 0.6575$. These results are also in accordance with the asymptotic calculations.

Experiment 3. In this experiment, we consider

$$\varepsilon_j \mid \{\kappa_j = m\} \sim \begin{cases} N(0, 1.25), & m = 1, \\ T_5, & m = 2. \end{cases}$$

The results of Experiment 3 for mLLE are presented in Table 3.

Here $H_*^{(2)} \approx 0.6502$. The simulations results inherit the pattern similar to the observed in the previous experiments.

7 Conclusions

We examined the asymptotic behavior of the modified local linear regression estimator for a mixture of regressions model. We proved that the modified estimator is asymptotically normal. The obtained rate of convergence of this estimator to the unknown value of the regression function at a given point is the same, regardless of whether the density function of a regressor has a jump at this point or not.

Based on the proven asymptotic theory, the optimal bandwidth parameter was derived that minimizes the asymptotic standard deviation of the estimator. The derived asymptotic theory was tested using a simulation experiment. The results obtained in this experiment are consistent with the theoretical results.

The subject of a further research is the development of the theory of optimal choice of parameters for the modified local linear regression estimator, that is, the bandwidth parameter and the kernel function.

References

- [1] Borovkov, A.A.: Probability Theory. Springer, London (2013) [MR3086572](#). <https://doi.org/10.1007/978-1-4471-5201-9>
- [2] Dychko, H., Maiboroda, R.: A generalized Nadaraya-Watson estimator for observations obtained from a mixture. *Theory Probab. Math. Stat.* **100**, 61–76 (2020). [MR3992993](#). <https://doi.org/10.1090/tpms/1094>
- [3] Fan, J.: Local linear regression smoothers and their minimax efficiencies. *Ann. Stat.* **21**(1), 196–216 (1993). [MR1212173](#). <https://doi.org/10.1214/aos/1176349022>
- [4] Fan, J., Gijbels, I.: Local Polynomial Modelling and Its Applications. Chapman & Hall, London (1996) [MR1383587](#)
- [5] Horbunov, D., Maiboroda, R.: Cross-validation for local-linear regression by observations from mixture (in Ukrainian). *Bull. Taras Shevchenko Natl. Univ. Kyiv., Ser. Phys. Math.* **1**, 37–43 (2023). <https://doi.org/10.17721/1812-5409.2023/1.5>
- [6] Horbunov, D., Maiboroda, R.: Consistency of local linear regression estimator for mixtures with varying concentrations. *Mod. Stoch. Theory Appl.* **11**(3), 359–372 (2024). [MR4757725](#). <https://doi.org/10.15559/24-VMSTA250>
- [7] Maiboroda, R., Sugakova, O.: Estimation and Classification by Observations from Mixture (in Ukrainian). Kyiv University Publishers, Kyiv (2008)
- [8] Maiboroda, R., Sugakova, O.: Statistics of mixtures with varying concentrations with application to DNA microarray data analysis. *J. Nonparametr. Stat.* **24**(I), 201–205 (2012). [MR2885834](#). <https://doi.org/10.1080/10485252.2011.630076>
- [9] McLachlan, G.J., Peel, D.: Finite Mixture Models. Wiley-Interscience, New York (2000). [MR1789474](#). <https://doi.org/10.1002/0471721182>
- [10] Pidnebesna, A., Fajnerová, I., Horáček, J., Hlinka, J.: Mixture components inference for sparse regression: introduction and application for estimation of neuronal signal from fMRI BOLD. *Appl. Math. Model.* **116**, 735–748 (2023). [MR4522959](#). <https://doi.org/10.1016/j.apm.2022.11.034>

- [11] Titterton, D.M., Smith, A.F., Makov, U.E.: Analysis of Finite Mixture Distributions. Wiley, New York (1985) [MR0838090](#)
- [12] Yao, W., Xiang, S.: Semiparametric mixtures of nonparametric regressions. *Ann. Inst. Stat. Math.* **70**, 131–154 (2018). [MR3742821](#). <https://doi.org/10.1007/s10463-016-0584-7>
- [13] Yao, W., Xiang, S.: Semiparametric mixtures of regressions with single-index for model based clustering. *Adv. Data Anal. Classif.* **April** (2020). [MR4118951](#). <https://doi.org/10.1007/s11634-020-00392-w>
- [14] Yao, W., Xiang, S.: Mixture Models Parametric, Semiparametric, and New Directions. CRC Press, London (2024)
- [15] Young, D.S., Hunter, D.R.: Mixtures of regressions with predictor-dependent mixing proportions. *Comput. Stat. Data Anal.* **54**, 2253–2266 (2010). [MR2720486](#). <https://doi.org/10.1016/j.csda.2010.04.002>