Consistency 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)

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. A modified local linear estimation (mLLE) technique is developed to estimate the regression functions of the mixture component nonparametrically. Consistency of the mLLE is demonstrated. Performance of mLLE and a modified Nadaraya–Watson estimator (mNWE) is assessed via simulations. The results confirm that the mLLE technique overcomes the boundary effect typical to the NWE.

Keywords   Nonparametric regression, mixture with varying concentrations, boundary effect, consistency

1   Introduction

Models of mixtures with varying concentrations (MVC) naturally arise in statistical analysis of sociological and biomedical data [9, 11]. The MVC model is a generalization of a classical finite mixture model FMM [10, 12] in which the concentrations of components in the mixture are different for different observations. Statistics of data described by MVC models is considered in [8].

Regression models provide powerful technology of dependencies analysis in multivariate data. For applications of parametric regression mixture models in behavioral

---

*Corresponding author.

science, see [5]. In context of the MVC model, linear [6] and nonlinear [7] regression models were considered. A modification of the Nadaraya–Watson estimator for MVC is considered in [1]. It is well known that for homogeneous data the Nadaraya–Watson estimator suffers from so-called boundary effect: it has particularly severe bias at the boundary points of the regressor support [3]. As a remedy to cure this sickness a local-linear regression (LLR) technique (or, more general local-polynomial) is used [2]. A modification of local-linear estimator for MVC is proposed in [4].

This paper focuses on the investigation of the modified LLR estimator properties. In Section 2 we describe the MVC model and particularly the model of regression mixture with varying concentrations. In Section 3 modifications of the Nadaraya–Watson (mNWE) and local-linear estimators (mLLE) for MVC are presented. Section 4 is devoted to the conditions for consistency of mLLE. Results of simulations are presented in Section 5. Concluding remarks are placed in Section 6.

2   Regression mixture

Consider a sample with n subjects $O_1, \ldots, O_n$, where each subject belongs to one of M subpopulations (mixture components). The number of components $\kappa_j = \kappa(O_j)$, which the j-th subject belongs to, is unknown. But the probability $p_{j;n}^m = \mathbf{P}(\kappa_j = m)$ that O_j belongs to m-th component is known for all $1 \le j \le n$ and $1 \le m \le M$. The probabilities $\{p_{j;n}^m\}$ are called the mixing probabilities or concentrations of the components in the mixture.

For each subject one observes a set of observed variables $\xi_j = \xi(O_j) \in \mathbb{R}^d$. Let $F^m(A) := \mathbf{P}(\xi(O) \in A | \kappa(O) = m)$ be the (unknown) distribution of $\xi(O)$ if O belongs to the m-th component. Then

$$\mathbf{P}(\xi_j \in A) = \sum_{m=1}^{M} p_{j;n}^m F^m(A), \quad A \in \mathcal{B}(\mathbb{R}^d), \tag{1}$$

where $\mathcal{B}(\mathbb{R}^d)$ is the Borel σ-algebra on $\mathbb{R}^d$.

In what follows we assume that ξ_j, $j = 1, \ldots, n$, are independent random vectors. Formula (1) is called the MVC model. A partial case of MVC is the mixture of regressions, at which a regression model is assumed for each mixture component distribution $F^m(A)$, $m = 1, \ldots, M$.

In this paper we restrict ourselves to bivariate vectors of observed features. For each subject O_j in the sample a vector $\xi_j = (X_j, Y_j)$ is observed, where $X_j = X(O_j)$ and $Y_j = Y(O_j)$ are the regressor and response connected by the following regression model:

$$Y_j = g^{(\kappa_j)}(X_j) + \varepsilon_j, \quad 1 \le j \le n, \tag{2}$$

where $g^{(m)}$ is an unknown regression function, which corresponds to the m-th component of mixture, $\varepsilon_j = \varepsilon(O_j)$ is a random error term. The error terms have zero mean and finite variance for each mixture component:

$$\mathbf{E}\big[\varepsilon(O_j) \mid \kappa(O_j) = m\big] = 0, \ \mathbf{Var}\big[\varepsilon(O_j) \mid \kappa(O_j) = m\big] = \sigma_{(m)}^2 < \infty, \quad 1 \le j \le n.$$

It is assumed that X_j and ε_j are independent for fixed κ_j.

Let $F_X^m(A) = \mathbf{P}(X(O) \in A \mid \kappa(O) = m)$ be the distribution of the regressor for m-th component. In what follows we assume that $F_X^m(A)$ is dominated by the Lebesgue measure for all $1 \le m \le M$. The corresponding probability densities are denoted by $f^{(m)}(x)$. These density functions are unknown.

Our aim is to estimate the unknown regression function $g^{(m)}(x)$ for m-th component of mixture in (2).

3 Estimators

3.1 Nonparametric regression in homogeneous samples

Let us recall how nonparametric regression estimators are defined in the case of homogeneous data. That is, we will assume here that (2) holds with $M = 1$ and $g(x) = g^{(1)}(x)$ is to be estimated.

The classical Nadaraya–Watson estimator (NWE) $\hat{g}_n^{NW}(x_0)$ for $g(x_0)$ is defined by

$$\hat{g}_n^{NW}(x_0) = \frac{\sum_{j=1}^n Y_j K(\frac{x_0 - X_j}{h})}{\sum_{j=1}^n K(\frac{x_0 - X_j}{h})},$$

where $K : \mathbb{R} \to \mathbb{R}$ is a kernel function, $h > 0$ is a bandwidth. The kernel K defines the weight of an observation (X_j, Y_j) in the estimator. A typical example of K used in NWE is the Epanechnikov kernel $K^{Ep}(t) = \frac{3}{4}(1 - t^2)\mathbf{1}\{|t| < 1\}$, where $\mathbf{1}\{A\}$ is an indicator function of an event A. The bandwidth can be interpreted as the width of the neighborhood around x_0 to which we localize the estimator of $g(x_0)$.

To construct the local linear estimator (LLE) one considers the localized least squares functional of the form

$$J(x_0; a, b) = \sum_{j=1}^n K\left(\frac{x_0 - X_j}{h}\right)\left(Y_j - \left(a + b(x_0 - X_j)\right)\right)^2. \qquad (3)$$

Let $(\hat{a}(x_0), \hat{b}(x_0))$ be the point of minimum of $J(x_0; a, b)$ over $(a, b) \in \mathbb{R}^2$. Then LLE for $g(x_0)$ is

$$\hat{g}_n^{LL}(x_0) = \hat{a}(x_0).$$

The estimator $\hat{g}^{LL}(x_0)$ can be calculated by (8)–(9) below with the weights $w_{j;n} = K((x_0 - X_j)/h)$.

3.2 The minimax weighting coefficients

To modify the estimators for MVC data we will need the minimax coefficients for components distribution estimation defined in [9].

Let $\mathbf{p}^m = (p_{1;n}^m, \ldots, p_{n;n}^m)^T$ be the vector of concentrations for the m-th component of mixture, $m = 1, \ldots, M$.

The averaging operation will be denoted by angle brackets:

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

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

$$\langle \mathbf{p}^m \mathbf{p}^k \rangle_n = \frac{1}{n} \sum_{j=1}^{n} p_{j;n}^m p_{j;n}^k.$$

Note that $\langle \mathbf{p}^m \mathbf{p}^k \rangle_n$ can be considered as an inner product on $\mathbb{R}^n$.

In what follows we assume that $\{\mathbf{p}^m\}_{m=1}^M$ are linearly independent. Denote by $\mathbf{\Gamma}_n = (\langle \mathbf{p}^k \mathbf{p}^l \rangle_n)_{k,l=1}^M$ the Gram matrix of $\{\mathbf{p}^m\}_{m=1}^M$. The weighting coefficients $a_{j;n}^k$, defined by

$$a_{j;n}^k = \frac{1}{\det \mathbf{\Gamma}_n} \sum_{m=1}^{M} (-1)^{m+k} \gamma_{km} p_{j;n}^m, \tag{4}$$

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

$$(a_{j;n}^1, \dots, a_{j;n}^M) = (p_{j;n}^1, \dots, p_{j;n}^M) \mathbf{\Gamma}_n^{-1}.$$

The vector of minimax coefficients for the m-th component will be denoted by $\mathbf{a}^m = (a_{1;n}^m, \dots, a_{n;n}^m)^T$. These coefficients can be used in weighted empirical distributions

$$\hat{F}_n^m(A) = \frac{1}{n} \sum_{j=1}^{n} a_{j;n}^m \mathbf{1}\{\xi_j \in A\}.$$

Observe that

$$\langle \mathbf{p}^k \mathbf{a}^m \rangle_n = \begin{cases} 1 & \text{if } k = m, \\ 0 & \text{if } k \neq m, \end{cases} \quad \text{for all } 1 \leq m \leq M. \tag{5}$$

By (5) and (1)

$$\mathbf{E}\big[\hat{F}_n^m(A)\big] = \sum_{k=1}^{M} \langle \mathbf{a}^m \mathbf{p}^k \rangle_n F^k(A) = F^m(A),$$

so $\hat{F}_n^m(A)$ is an unbiased estimator for $F^m(A)$. It is shown in [9] that under the MVC model (1) $\hat{F}_n^m(A)$ is a minimax estimator for $F_m(A)$ in the class of all unbiased estimators.

3.3 Nonparametric estimators for MVC

Here and below we assume that ξ_j, $j = 1, \dots, n$, are described by the regression MVC model (2).

In [1] a modified NWE (mNWE) for $g^{(m)}(x_0)$ is proposed of the form

$$\hat{g}_n^{NW(m)}(x_0) = \frac{\sum_{j=1}^{n} a_{j;n}^m Y_j K(\frac{x_0 - X_j}{h})}{\sum_{j=1}^{n} a_{j;n}^m K(\frac{x_0 - X_j}{h})}. \tag{6}$$

It is shown in [1] that, under suitable assumptions, $\hat{g}_n^{NW(m)}(x_0)$ is a consistent and asymptotically normal estimator of $g^{(m)}(x_0)$.

To derive a modified local linear estimator (mLLE) we start with the weighted least squares functional

$$J^{(m)}(x_0; a, b) = \sum_{j=1}^{n} a_{j;n}^m K\left(\frac{x_0 - X_j}{h}\right)(Y_j - (a + b(x_0 - X_j)))^2. \qquad (7)$$

Observe that some weights $a_{j;n}^m$ can be negative, so this functional can attain negative values and be unbounded from below. Despite this we consider the stationary point of $J^{(m)}(x_0; a, b)$, i.e. the solution $(\hat{a}(x_0), \hat{b}(x_0))$ to the equation system

$$\begin{cases} \frac{\partial}{\partial a} J^{(m)}(x_0; a, b) = 0, \\ \frac{\partial}{\partial b} J^{(m)}(x_0; a, b) = 0. \end{cases}$$

The mLLE for $g^{(m)}(x_0)$ is defined as $\hat{g}_n^{LL(m)}(x_0) := \hat{a}(x_0)$. By a simple algebra it can be calculated as

$$\hat{g}_n^{LL(m)}(x_0) = \frac{S_Y S_{XX} - S_X S_{XY}}{S_1 S_{XX} - (S_X)^2}, \qquad (8)$$

where

$$S_1 = \sum_{j=1}^{n} w_{j;n}^m, \ S_X = \sum_{j=1}^{n} w_{j;n}^m (x_0 - X_j), \ S_Y = \sum_{j=1}^{n} w_{j;n}^m Y_j,$$

$$S_{XX} = \sum_{j=1}^{n} w_{j;n}^m (x_0 - X_j)^2, \ S_{XY} = \sum_{j=1}^{n} w_{j;n}^m Y_j (x_0 - X_j), \qquad (9)$$

$$w_{j;n}^m = a_{j;n}^m K\big((x_0 - X_j)/h\big).$$

Since $\hat{g}_n^{NW(m)}(x_0) = S_Y/S_1$, we have

$$\hat{g}_n^{LL(m)}(x_0) = \hat{g}_n^{NW(m)}(x_0) - \frac{S_1 S_{XY} - S_X S_Y}{S_1 S_{XX} - (S_X)^2} \cdot \frac{S_X}{S_1}. \qquad (10)$$

So, mLLE can be considered as the mNWE with an additional correction term. This term takes a value close to zero if $S_X/S_1 \approx 0$. As we will see below, this is usually the case if x_0 is an interior point of the regressor support and K is an even function. Therefore if x_0 lies in the interior of the regressor support, mLLE behavior is nearly the same as of mNWE. But at the boundary points the correction makes the bias of mLLE smaller than of mNWE.

4 Consistency of the modified local-linear regression estimator

The consistency conditions of the weightened local-linear estimator are given in the following theorem.

Theorem 1. *Let* $1 \leq m \leq M$ *be a fixed number of components. Assume that:*

1. x_0 *is a continuity point of* $f^{(m)}(x)$ *and* $g^{(m)}(x)$;

2. $h = h_n$, such that $h_n \to 0$ and $nh_n \to +\infty$ as $n \to \infty$;

3. the kernel function K is bounded on $\mathbb{R}$ and satisfies $\int_{-\infty}^{\infty} |K(z)|z^2 dz < \infty$ and $\int_{-\infty}^{\infty} K^2(z)z^4 dz < \infty$;

4. $g^{(k)}(x)$, $f^{(k)}(x)$ are bounded on $x \in \mathbb{R}$ for all $1 \leq k \leq M$;

5. $f^{(m)}(x_0) > 0$;

6. $\Delta(K) := \int_{-\infty}^{+\infty} K(z)dz \int_{-\infty}^{+\infty} z^2 K(z)dz - (\int_{-\infty}^{+\infty} zK(z)dz)^2 \neq 0$;

7. there exists $c_0 > 0$, such that $\det \Gamma_n \geq c_0$ for all $n \geq 1$.

Then $\hat{g}_n^{LL(m)}(x_0)$ is a consistent estimator of $g^{(m)}(x_0)$, i.e.

$$\hat{g}_n^{LL(m)}(x_0) \xrightarrow{P} g^{(m)}(x_0), \quad n \to \infty.$$

Remarks.

1. Assumption 2 of the theorem is necessary. This is also required for the LLR estimator in case of homogeneous sample, see Theorem 1 of [2].

2. If the kernel K has a bounded support then Assumption 4 can be relaxed to a local version: $g^{(k)}$ and $f^{(k)}$ are bounded in some open neighborhood of x_0.

3. If the kernel $K(z)$ is nonnegative the assumption $\Delta(K) \neq 0$ is equivalent to $K(\cdot) \neq 0$ on a set of positive Lebesgue measure.

4. The requirement that $f^{(m)}(x_0) > 0$ is crucial. If $f^{(m)}(x) = 0$ for all x in some neighborhood of x_0, then $X^{(m)}$ will not attain values at this neighborhood, hence $g^{(m)}(x_0)$ cannot be estimated nonparametrically. On the other hand, when x_0 is a boundary point of the support of $X^{(m)}$, consistent estimation is possible.

5. Condition $\det \Gamma_n > 0$, is equivalent to linear independence of concentrations vectors $\{\mathbf{p}^k\}$, $k = 1, \dots, M$. Assumption 7 can be considered as an asymptotic version of the linear independence condition. It is worth noting that this assumption outrules the classical FMMs at which $p_{j;n}^m = p^m$ does not depend on m.

Proof. Multiplying both the numerator and denominator in (8) by $(nh_n^2)^{-2}$, one obtains

$$\hat{g}_n^{LL(m)}(x_0) = \frac{\frac{S_Y}{nh_n}\frac{S_{XX}}{nh_n^3} - \frac{S_X}{nh_n^2}\frac{S_{XY}}{nh_n^2}}{\frac{S_1}{nh_n}\frac{S_{XX}}{nh_n^3} - (\frac{S_X}{nh_n^2})^2}. \tag{11}$$

Let us analyze each term separately.

We will start from $S_{XX}/(nh_n^3)$.

$$\mathbf{E}\left[\frac{S_{XX}}{nh_n^3}\right] = \frac{1}{nh_n^3}\sum_{j=1}^{n} a_{j;n}^m \mathbf{E}\left[K\left((x_0 - X_j)/h_n\right)(x_0 - X_j)^2\right]$$

$$= \frac{1}{nh_n^3} \sum_{j=1}^{n} a_{j;n}^m \sum_{k=1}^{M} p_{j;n}^k \mathbf{E}\left[K\big((x_0 - X(O))/h_n\big)(x_0 - X(O))^2 \mid \kappa(O) = k \right]$$

$$= \frac{1}{h_n^3} \sum_{k=1}^{M} \langle \mathbf{a}^m \mathbf{p}^k \rangle_n \mathbf{E}\left[K\big((x_0 - X(O))/h_n\big)(x_0 - X(O))^2 \mid \kappa(O) = k \right]$$

$$= \frac{1}{h_n^3} \mathbf{E}\left[K\big((x_0 - X(O))/h_n\big)(x_0 - X(O))^2 \mid \kappa(O) = m \right],$$

due to (5). So

$$\mathbf{E}\left[\frac{S_{XX}}{nh_n^3} \right] = \frac{1}{h_n^3} \int\limits_{-\infty}^{+\infty} K\big((x_0 - x)/h_n\big)(x_0 - x)^2 f^{(m)}(x)\,dx.$$

The substitution

$$z = (x_0 - x)/h_n, \quad dz = -dx/h_n, \tag{12}$$

yields

$$\mathbf{E}\left[\frac{S_{XX}}{nh_n^3} \right] = \int\limits_{-\infty}^{+\infty} K(z) z^2 f^{(m)}(x_0 - h_n z)\,dz.$$

By Assumptions 1 and 2, $K(z)z^2 f^{(m)}(x_0 - h_n z) \to K(z)z^2 f^{(m)}(x_0)$ for all $z \in \mathbb{R}$. By Assumption 4, there exists $C_f < \infty$, such that $f^{(m)}(x) < C_f$ for all $x \in \mathbb{R}$. By Assumption 3, $\int_{-\infty}^{\infty} |K(z)|z^2 dz < \infty$, so $C_f |K(z)|z^2$ is an integrable majorant of $K(z)z^2 f^{(m)}(x_0 - h_n z)$ and by the Lebesgue dominated convergence theorem

$$\mathbf{E}\left[\frac{S_{XX}}{nh_n^3} \right] \to f(x_0) \int_{-\infty}^{\infty} K(z)z^2 dz, \quad \text{as } n \to \infty. \tag{13}$$

Let us bound the variance of $S_{XX}/(nh_n^3)$:

$$\mathbf{Var}\left[\frac{S_{XX}}{nh_n^3} \right] = \frac{1}{(nh_n^3)^2} \sum_{j=1}^{n} (a_{j;n}^m)^2 \, \mathbf{Var}\left[K\big((x_0 - X_j)/h_n\big)(x_0 - X_j)^2 \right]$$

$$\leq \frac{1}{n^2 h_n^6} \sum_{j=1}^{n} (a_{j;n}^m)^2 \mathbf{E}\left[\big(K\big((x_0 - X_j)/h_n\big)(x_0 - X_j)^2 \big)^2 \right]$$

$$= \frac{1}{n^2 h_n^6} \sum_{j=1}^{n} (a_{j;n}^m)^2 \sum_{k=1}^{M} p_{j;n}^k \mathbf{E}\left[\big(K\big((x_0 - X(O))/h_n\big) \big)^2 (x_0 - X(O))^4 \mid \kappa(O) = k \right]$$

$$= \frac{1}{nh_n^6} \sum_{k=1}^{M} \langle (\mathbf{a}^m)^2 \mathbf{p}^k \rangle_n \int\limits_{-\infty}^{+\infty} \big(K\big((x_0 - x)/h_n\big) \big)^2 (x_0 - x)^4 f^{(k)}(x)\,dx.$$

Applying once more the substitution (12) we obtain

$$\mathbf{Var}\left[\frac{S_{XX}}{nh_n^3} \right] \leq \frac{1}{nh_n} \sum_{k=1}^{M} \langle (\mathbf{a}^m)^2 \mathbf{p}^k \rangle_n \int\limits_{-\infty}^{+\infty} \big(K(z) \big)^2 z^4 f^{(k)}(x_0 - h_n z)\,dz. \tag{14}$$

Observe that

$$\left\langle (\mathbf{a}^m)^2 \mathbf{p}^k \right\rangle_n \leq \left(\max_{1 \leq j \leq n} |a_{j;n}^m| \right)^2$$

since $0 \leq p_{j;n}^k \leq 1$.

By (4) and Assumption 7,

$$|a_{j;n}^k| = \left| \frac{1}{\det \Gamma_n} \sum_{m=1}^{M} (-1)^{m+k} \gamma_{km} p_{j;n}^m \right| \leq c_0^{-1} \sum_{m=1}^{M} |\gamma_{km}| \leq c_0^{-1} M! =: C_0 < \infty,$$

since $|\gamma_{km}| \leq (M-1)!$.

So

$$\mathbf{Var}\left[\frac{S_{XX}}{nh_n^3} \right] \leq \frac{1}{nh_n} M C_0^2 C_f \int_{-\infty}^{\infty} K^2(z) z^4 dz \to 0, \quad n \to \infty,$$

by Assumption 2.

Combining this with (13) we obtain

$$\frac{S_{XX}}{nh_n^3} \xrightarrow{\mathrm{P}} f^{(m)}(x_0) \int_{-\infty}^{+\infty} K(z) z^2 dz, \quad \text{as } n \to \infty. \tag{15}$$

Consider now asymptotics of $S_{XY}/(nh^2)$. With (2) in mind by the same way as above we obtain

$$\mathbf{E}\left[\frac{S_{XY}}{nh_n^2} \right] = \frac{1}{h_n^2} \sum_{k=1}^{M} \langle \mathbf{a}^m \mathbf{p}^k \rangle_n \mathbf{E}\left[K\big((x_0 - X(O))/h_n\big)(x_0 - X(O)) Y(O) \mid \kappa(O) = k \right]$$

$$= \frac{1}{h_n^2} \sum_{k=1}^{M} \langle \mathbf{a}^m \mathbf{p}^k \rangle_n \mathbf{E}\left[K\big((x_0 - X(O))/h_n\big)(x_0 - X(O)) g^{(\kappa(O))}(X(O)) \mid \kappa(O) = k \right]$$

$$= \frac{1}{h_n^2} \int_{-\infty}^{+\infty} K\big((x_0 - x)/h_n\big)(x_0 - x) g^{(m)}(x) f^{(m)}(x) dx.$$

(Here the assumption $\mathbf{E}[\varepsilon(O) \mid \kappa(O) = m] = 0$ and conditional independence of $X(O)$ and $\varepsilon(O)$ were used.)

The substitution (12) yields

$$\mathbf{E}\left[\frac{S_{XY}}{nh_n^2} \right] = \int_{-\infty}^{\infty} K(z) z g^{(m)}(x_0 - h_n z) f^{(m)}(x - h_n z) dz.$$

Applying the Lebesgue dominated convergence theorem, in view of Assumptions 1 and 6, we obtain

$$\mathbf{E}\left[\frac{S_{XY}}{nh_n^2} \right] \to g^{(m)}(x_0) f^{(m)}(x_0) \int_{-\infty}^{\infty} K(z) z dz, \quad \text{as } n \to \infty. \tag{16}$$

Then

$$\mathbf{Var}\left[\frac{S_{XY}}{nh_n^2}\right] \le \frac{1}{(nh_n^2)^2} \sum_{j=1}^{n} (a_{j;n}^m)^2 \mathbf{E}\big[\big(K\big((x_0 - X_j)/h_n\big)(x_0 - X_j)Y_j\big)^2\big]$$

$$= \frac{1}{nh_n^4} \sum_{k=1}^{M} \langle(\mathbf{a}^m)^2\mathbf{p}^k\rangle_n \mathbf{E}\big[\big(g^{(k)}\big(X(O)\big)K\big((x_0 - X(O))/h_n\big)(x_0 - X(O))\big)^2 \mid \kappa(O) = k\big]$$

$$+ \frac{1}{nh_n^4} \sum_{k=1}^{M} \sigma_{(k)}^2 \langle(\mathbf{a}^m)^2\mathbf{p}^k\rangle_n \mathbf{E}\big[\big(K\big((x_0 - X(O))/h_n\big)(x_0 - X(O))\big)^2 \mid \kappa(O) = k\big]$$

$$= \frac{1}{nh_n} \sum_{k=1}^{M} \langle(\mathbf{a}^m)^2\mathbf{p}^k\rangle_n \int_{-\infty}^{\infty} \big(K(z)z\big)^2 \big(g^{(k)}(x_0 - h_n z)\big)^2 f^{(k)}(x_0 - h_n z)dz$$

$$+ \frac{1}{nh_n} \sum_{k=1}^{M} \sigma_{(k)}^2 \langle(\mathbf{a}^m)^2\mathbf{p}^k\rangle_n \int_{-\infty}^{\infty} \big(K(z)z\big)^2 f^{(k)}(x_0 - h_n z)dz.$$

So, by Assumptions 2, 3 and 4 for some $C < \infty$, we obtain

$$\mathbf{Var}\left[\frac{S_{XY}}{nh_n^2}\right] \le \frac{C}{nh_n} \to 0, \quad \text{as } n \to \infty.$$

This with (16) yields

$$\frac{S_{XY}}{nh_n^2} \xrightarrow{\ \mathrm{P}\ } f^{(m)}(x_0)g^{(m)}(x_0) \int_{-\infty}^{+\infty} K(z)zdz, \quad \text{as } n \to \infty. \tag{17}$$

By the same way we obtain

$$\frac{S_X}{nh_n^2} \xrightarrow{\ \mathrm{P}\ } f^{(m)}(x_0) \int_{-\infty}^{+\infty} zK(z)dz,$$

$$\frac{S_Y}{nh_n} \xrightarrow{\ \mathrm{P}\ } f^{(m)}(x_0)g^{(m)}(x_0) \int_{-\infty}^{+\infty} K(z)dz,$$

$$\frac{S_1}{nh_n} \xrightarrow{\ \mathrm{P}\ } f^{(m)}(x_0) \int_{-\infty}^{+\infty} K(z)dz,$$

as $n \to \infty$. Combining this with (11), (15) and (17) we obtain

$$\hat{g}_n^{LL(m)}(x_0) = \frac{\frac{S_Y}{nh_n}\frac{S_{XX}}{nh_n^3} - \frac{S_X}{nh_n^2}\frac{S_{XY}}{nh_n^2}}{\frac{S_1}{nh_n}\frac{S_{XX}}{nh_n^3} - (\frac{S_X}{nh_n^2})^2} \xrightarrow{\ \mathrm{P}\ }$$

$$\xrightarrow{\ \mathrm{P}\ } \frac{g^{(m)}(x_0)(f(x_0))^2\Delta(K)}{(f(x_0))^2\Delta(K)} = g^{(m)}(x_0), \quad \text{as } n \to \infty.$$

(Here Assumptions 5 and 6 where used.) This is just the statement of the theorem.
□

5   Simulation results

5.1   Description of simulations

To assess quality of mLLE and compare it to mNWE we performed a small simulation study. In all the simulation experiments we used a two-component MVC model with concentrations defined as follows:

$$p^1_{j;n} = \frac{j}{n}, \ p^2_{j;n} = 1 - \frac{j}{n}, \quad 1 \le j \le n.$$

The regression functions of the components were defined as

$$g^m(x) = (-1)^m x(1-x), \quad m = 1, 2.$$

The distribution of regressor X is uniform on $[0, 1]$ for both the components. The error terms had different distributions in different experiments.

Performance of the estimators was examined at two points $x_0 = 0.5$ (an interior point of the regressor support) and $x_0 = 0$ (a boundary point).

For sample sizes n from 100 through 10000 we generated $B = 1000$ independent samples from the model. Assuming that x_0 is fixed, the mLLE $\hat{g}_n^{LL(m)}(x_0)$ and the mNWE estimator $\hat{g}_n^{NW(m)}(x_0)$ were calculated by each generated sample. By the obtained samples of the estimations we calculated the observed bias

$$\text{Bias}\big(\hat{g}_n^{LL(m)}(x_0)\big) = \mathbf{E}_*\big[\hat{g}_n^{LL(m)}(x_0) - g^m(x_0)\big]$$

and variance

$$\text{Var}\big(\hat{g}_n^{LL(m)}(x_0)\big) = \mathbf{E}_*\big[\big(\hat{g}_n^{LL(m)}(x_0) - \mathbf{E}_*\big[\hat{g}_n^{LL(m)}(x_0)\big]\big)^2\big].$$

(Similarly for $\hat{g}_n^{NW(m)}$.) Here $m = 1, 2$ and $\mathbf{E}_*[\cdot]$ is a sample mean. For all of the experiments we used the Epanechnikov kernel and the bandwidth parameter was $h_n = Hn^{-1/5}$, where $H = H_{opt}^{NW} = 1.816$ is a theoretically optimal value for mNWE, see [1].

Three experiments were performed.

- *Experiment 1.* In this experiment the regression errors' distribution was Gaussian, $\varepsilon_j \sim N(0, 1.25)$ for both components, so this is the case of light-tailed errors.

- *Experiment 2.* The regression errors were generated from the Student-T distribution with 10 degrees of freedom, i.e. $\varepsilon_j \sim T(10)$ for both components. In this case the the variance of errors is the same as in the previous case, but distribution is heavy-tailed.

- *Experiment 3.* In this experiment the errors' distributions were different in the first and second components: $\varepsilon_j|\{\kappa_j = 1\} \sim N(0, 1.25)$ and $\varepsilon_j|\{\kappa_j = 2\} \sim T(10)$.

For all of experiments, the numerical results are shown only for the first component of mixture, since the biases and variances of the estimators for the second component are nearly of the same magnitude, so they are not of significant interest.

5.2 Performance of mLLE

The results of Experiment 1 for mLLE are presented in Table 1.

Table 1. Experiment 1 results on the mLLE

	$x_0 = 0$		$x_0 = 0.5$	
n	Bias	Var	Bias	Var
100	−0.14934	9.78188	0.0783	0.05874
250	−0.04822	0.15793	0.0709	0.02123
500	−0.03528	0.09786	0.05721	0.01149
1000	−0.01801	0.04796	0.04349	0.00697
2500	−0.01372	0.0254	0.02951	0.00329
5000	−0.01285	0.01394	0.02405	0.00193
10000	−0.00887	0.00828	0.01492	0.0011

In this experiment the bias of mLLE at the boundary point $x = 0$ was even smaller in magnitude then at the inner point $x = 0.5$ for large samples ($n \geq 250$). The variance of mLLE was smaller at the point $x = 0.5$. For the sample size $n = 100$ the mLLE had extremely large variance at $x = 0$. Since only nearly 50 observations in such samples belong to the first component, this is a very small sample size for a nonparametric estimation.

The results of Experiment 2 that are shown in Table 2.

Table 2. Experiment 2 results on the mLLE

	$x_0 = 0$		$x_0 = 0.5$	
n	Bias	Var	Bias	Var
100	−0.07368	0.50335	0.06813	0.05452
250	−0.0472	0.19456	0.06015	0.02402
500	−0.02049	0.09438	0.0523	0.01202
1000	−0.02602	0.05402	0.03873	0.00683
2500	−0.01697	0.02364	0.03016	0.00322
5000	−0.00879	0.01374	0.02307	0.00171
10000	−0.01294	0.00728	0.01571	0.00112

By these results we observe that the heavy-tailed regression errors did not significantly affect the mLLE performance. The bias at the boundary point is still of the same magnitude as in the inner point of the regressor support.

The results of Experiment 3 are shown in Table 3.

Table 3. Experiment 3 results on the mLLE

	$x_0 = 0$		$x_0 = 0.5$	
n	Bias	Var	Bias	Var
100	−0.0646	0.47884	0.06129	0.05877
250	−0.06522	0.18535	0.06101	0.02189
500	−0.0309	0.09735	0.05424	0.0113
1000	−0.03353	0.05129	0.04028	0.00645
2500	−0.02261	0.02417	0.02939	0.00298
5000	−0.01109	0.01384	0.02022	0.00187
10000	−0.01169	0.00796	0.01528	0.00099

In this experiment we observe the same pattern of decrease of mLLE biases and variances as in Experiments 1 and 2. So difference of the errors distributions caused no significant effect on the estimator.

5.3   Results on comparison of mLLE and mNWE

To compare performance of mLLE and mNWE we calculated the ratios of biases for these two estimators $\mathrm{Bias}(\hat{g}_n^{LL(1)}(x_0))/\mathrm{Bias}(\hat{g}_n^{NW(1)}(x_0))$ and the ratios of their variances $\mathrm{Var}(\hat{g}_n^{LL(1)}(x_0))/\mathrm{Var}(\hat{g}_n^{NW(1)}(x_0))$.

The results of the experiments for $x_0 = 0.5$ are presented in Figure 1. Both the biases and variances ratios are very close to 1 even for small sample sizes. They apparently tend to 1 when n increases.
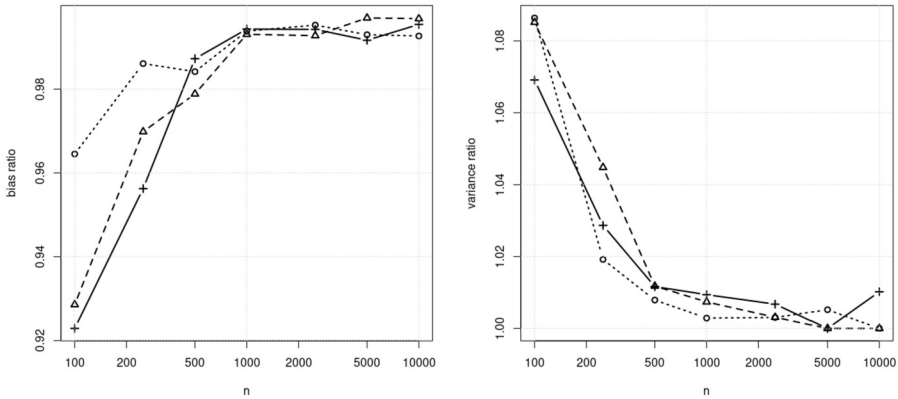

Fig. 1. Ratios of biases (left panel) and variances (right panel) of mLLE and mNWE at $x_0 = 0.5$ in Experiment 1 ($\circ$), Experiment 2 ($\triangle$) and Experiment 3 ($+$)

The same ratios for $x_0 = 0$ are presented in Figure 2. (Extremely high ratio $\mathrm{Var}(\hat{g}_n^{LL(1)}(x_0))/\mathrm{Var}(g_n^{NW(1)}(x_0)) = 109.92$ for $n = 100$ in Experiment 1 is not presented to avoid collapsing of all other points.)

The left panel of Figure 2 shows that the bias of mLLE is significantly smaller then of mNWE at the boundary point $x_0 = 0$. The biases' ratio for $n = 10000$ vary from 0.097 in Experiment 1 to 0.14 in Experiment 2. On the other hand, the variances of mLLE are larger then of mNWE (Three to four times larger for $n = 10000$). Recall that the bandwidth h in the experiments was taken to be asymptotically optimal for mNWE, not for mLLE. Increasing h one can achieve a reduction in variance due to a moderate increase of the bias.

As an unknown referee noted, it can be of interest to compare the mean squared errors of the estimators, i.e.

$$\mathbf{MSE}\big[\hat{g}_n^{(m)}(x)\big] = \mathbf{E}_*\big(\hat{g}_n^{(m)}(x) - g^{(m)}\big)^2 = \mathrm{Var}\big(\hat{g}_n^{(m)}(x)\big) + \mathrm{Bias}^2\big(\hat{g}_n^{(m)}(x)\big).$$

In all the experiments the ratio

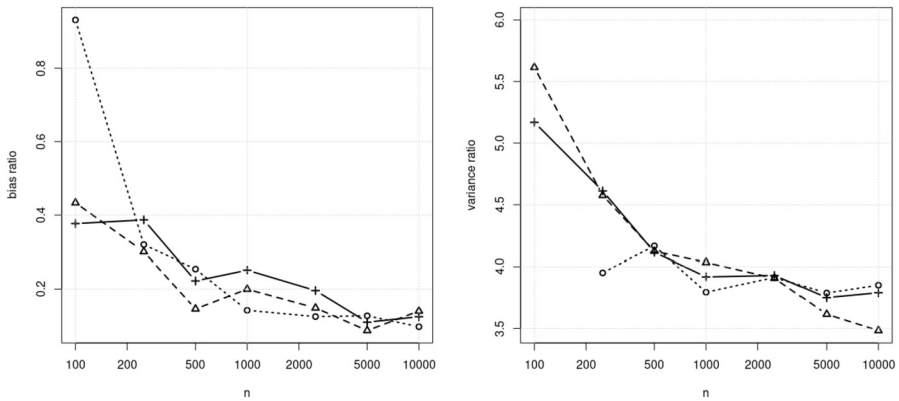$$R = \mathbf{MSE}\big[\hat{g}_n^{LL(m)}(x_0)\big]/\mathbf{MSE}\big[\hat{g}_n^{NW(m)}(x_0)\big]$$

Fig. 2. Ratios of biases (left panel) and variances (right panel) of mLLE and mNWE at $x_0 = 0$ in Experiment 1 ($\circ$), Experiment 2 ($\triangle$) and Experiment 3 ($+$)

was near to 1 at $x_0 = 0.5$ for all sample sizes. At $x_0 = 0$, R was larger then 1 for smaller n (from 2.280 in Experiment 2 to 2.31 in Experiment 1 for $n = 250$) and smaller then 1 for $n = 10000$ (from 0.8037 in Experiment 1, to 0.7461 in Experiment 3). So, the mLLE estimators outperformed the mNWE ones for large sample sizes.

6 Conclusion

We have discussed a modification of local linear regression technique for the nonparametric estimation in the model of regression mixture. Consistency of the obtained estimator is demonstrated. Results of simulations confirm significant reduction of boundary effect by the use of mLLE. Much efforts are needed to develop a practical algorithm for bandwidth selection, especially at boundary points of the regressor distribution support.

Acknowledgement

References

[1] Dychko, H., Maiboroda, R.: A generalized Nadaraya-Watson estimator for observations obtained from a mixture. Theory Probab. Math. Stat. **100**, 61–76 (2020). https://doi.org/10.1090/tpms/1094

[2] Fan, J.: Local Linear Regression Smoothers and their minimax efficiencies. Ann. Stat. **21**(1), 196–216 (1993). https://doi.org/10.1214/aos/1176349022

[3] Hastie, T., Loader, C.: Local regression: automatic kernel carpentry. Stat. Sci. **8**(2), 120–129 (1993). https://doi.org/10.1214/ss/1177011002

[4] Horbunov, D., Maiboroda, R.: Cross-validation for local-linear regression by observations from mixture. Bull. Taras Shevchenko Natl. Univ. Kyiv., Ser. Phys. Math. **1**, 37–43 (2023) (in Ukrainian). https://doi.org/10.17721/1812-5409.2023/1.5

[5] Lamont, A.E., Vermunt, J.K., Van Horn, Regression, M.L.: Mixture models: does modeling the covariance between independent variables and latent classes improve the results? Multivar. Behav. Res. **51**(1), 35–52 (2016). https://doi.org/10.1080/00273171.2015.1095063

[6] Liubashenko, D., Maiboroda, R.: Linear regression by observations from mixture with varying concentrations. Mod. Stoch. Theory Appl. **2**, 343–353 (2015). MR3456142. https://doi.org/10.15559/15-VMSTA41

[7] Maiboroda, R., Miroshnichenko, V., Sugakova, O.: Jackknife for nonlinear estimating equations. Mod. Stoch. Theory Appl. **9**(4), 377–399 (2022). MR4510379. https://doi.org/10.15559/22-vmsta208

[8] Maiboroda, R., Sugakova, O.: Estimation and Classification by Observations from Mixture. Kyiv University Publishers, Kyiv (2008) (in Ukrainian)

[9] Maiboroda, R., Sugakova, O.: Statistics of mixtures with varying concentrations with application to DNA microarray data analysis. J. Nonparametr. Stat. **24**(1), 201–205 (2012). MR2885834. https://doi.org/10.1080/10485252.2011.630076

[10] McLachlan, G.J., Peel, D.: Finite Mixture Models. Wiley–Interscience, New York (2000). MR1789474. https://doi.org/10.1002/0471721182

[11] 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

[12] Titterington, D.M., Smith, A.F., Makov, U.E.: Analysis of Finite Mixture Distributions. Wiley, New York (1985). MR0838090