UNIVERSITATIS IAGELLONICAE ACTA MATHEMATICA, FASCICULUS XLIII
2005

ON SOME TWO-STEP DENSITY ESTIMATION METHOD

BY JOLANTA JARNICKA

Abstract. We introduce a new two-step kernel density estimation method, based on the EM algorithm and the generalized kernel density estimator. The accuracy obtained is better in particular, in the case of multimodal or skewed densities.

1. Introduction. Density estimation has been investigated in many papers. In particular, the kernel estimation method, introduced in [**11**], has received much attention.

Let us recall some problems connected with the kernel density estimation.

Let $X_i : \Omega \longrightarrow \mathbb{R}$, $i = 1, \ldots, n$, be a sequence of random variables defined on a probability space $(\Omega, \mathfrak{F}, P)$. Suppose that they are identically and independently distributed with an unknown density f, and that $\{x_i\}$ is a sequence of corresponding observations.

The kernel density estimator is characterized by two components: the bandwidth $h(n)$ and the kernel K.

DEFINITION 1.1. Let $\{h(n)\}_{n=1}^{\infty} \subset (0, +\infty)$ with $h(n) \longrightarrow 0$. Let $K : \mathbb{R} \longrightarrow \mathbb{R}$ be a measurable nonnegative function. The *kernel density estimator* is given by the formula

$$\widehat{f}_h(x) = \frac{1}{nh(n)} \sum_{i=1}^{n} K\Big(\frac{x - x_i}{h(n)}\Big). \tag{1.1}$$

We call h the *bandwidth* (window width or smoothing parameter). We call K the *kernel*.

Statistical properties of this estimator, like biasedness, asymptotical unbiasedness and efficiency, were presented in [**11**], [**10**], and [**13**].

The kernel determines the regularity, symmetry about zero and shape of the estimator, while the bandwidth – the amount of smoothing. In particular, $\widehat{f_h}$ is a density, provided that $K \geq 0$ and $\int_{\mathbb{R}} K(t)\,dt = 1$.

Considerable research has been carried out on the question of how one should select K in order to optimize properties of the kernel density estimator $\widehat{f_h}$ (see e.g. **[5]**, **[14]**, and **[3]**) and numerous examples have been discussed. Suggested choice is a density, symmetric about zero, with finite variance, like e.g. the Gaussian kernel $K(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, the Epanechnikov kernel $K(t) = \frac{3}{4\sqrt{5}}(1 - \frac{1}{5}x^2)$ for $|x| < \sqrt{5}$ or the rectangular kernel $K(t) = \frac{1}{2}$ for $|x| < 1$ (see e.g. **[13]** or **[4]** for more examples). But in some situations nonsymmetric kernels (e.g. the exponential kernel $K(t) = e^{-x}$ for $x > 0$) guarantee better results. A comparative study of this problem was carried out in **[7]**.

The main problem of the kernel density estimation is still the choice of the bandwidth $h(n)$. Various measures of accuracy of the estimator $\widehat{f_h}$ have been used to obtain the optimal bandwidth (see e.g. **[3]**, **[13]**). We will here focus on the *mean integrated squared error* given by

$$\mathrm{MISE}(\widehat{f_h}) = E \int_{\mathbb{R}} (\widehat{f_h}(x) - f(x))^2\,dx,$$

and its asymptotic version for $h(n) \to 0$ and $nh(n) \to \infty$ as $n \to \infty$, often abbreviated by $\mathrm{AMISE}(\widehat{f_h})$ (see **[5]**, **[8]**). The mean integrated squared error can be written as a sum of integrated squared bias and integrated variance of $\widehat{f_h}$:

$$\mathrm{MISE}(\widehat{f_h}) = \int_{\mathbb{R}} (E(\widehat{f_h}(x)) - f(x))^2\,dx + \int_{\mathbb{R}} V(\widehat{f_h}(x))\,dx.$$

We want to keep both the bias and the variance possibly small, so the optimal bandwidth $h(n)$ should be chosen so that MISE is minimal.

Under additional assumptions on the density f (we assume that f is twice differentiable and that $\int_{\mathbb{R}} \Big(f''(x)\Big)^2\,dx < \infty$) and for a kernel being a symmetric density with a finite variance, we can provide asymptotic approximations for the bias and the variance of $\widehat{f_h}$ (see **[13]**). Having obtained the formula for the asymptotic mean integrated squared error, we get the optimal bandwidth $h(n)$ in the following form:

$$h_{\mathrm{opt}}(n) = \left(\frac{\int_{\mathbb{R}} K^2(t)\,dt}{\Big(\int_{\mathbb{R}} t^2 K(t)\,dt\Big)^2 \int_{\mathbb{R}} \Big(f''(x)\Big)^2\,dx} \right)^{\frac{1}{5}} n^{-\frac{1}{5}}. \tag{1.2}$$

Although the integrals $\int_{\mathbb{R}} K^2(t)\,dt$ and $\int_{\mathbb{R}} t^2K(t)\,dt$ can be calculated if K is known, formula (1.2) depends on an unknown density f and hence, in practice, the ability to calculate $h_{\text{opt}}(n)$ depends on obtaining an estimate for $\int_{\mathbb{R}} \Big(f''(x)\Big)^2\,dx$.

Several ways to solve this problem have appeared in the literature. One of the simplest methods, suggested in [**13**] (often called the *Silverman's rule of thumb*), is to assume that f is known. For example assuming that f is the density of a normal distribution with parameters μ and σ^2, and using the Gaussian kernel, we obtain $h_{\text{opt}}(n) = 1.06\sigma n^{-\frac{1}{5}}$, where σ can be estimated by the standard deviation. Other best known methods (beside those giving automatic procedures for selecting the bandwidth, like the cross-validation (see e.g. [**13**])) are based on the idea of constructing a pilot function and then using it in actual estimation (see e.g. [**13**], [**9**], [**1**], [**12**], [**15**], and [**6**]). The problem is that none of these methods guarantees good results for a wide class of estimated density functions.

In this paper we propose a new two-step kernel density estimation method, which generalizes methods considered in [**13**] and [**6**]. Its practical application is in fact based on a solution of some optimization problem.

In Section 2 we present the idea of the two-step method and prove some basic properties of the proposed generalized kernel density estimator. Section 3 is dedicated to the problem of the choice of optimal bandwidths $\{h_j(n)\}_{j=1}^m$. In Section 4 we propose an algorithm which enables us to construct a pilot. The last section contains a discussion of an example, which presents possibilities of our method.

We emphasize that we have carried a lot of experiments and computer simulations in order to check the efficiency of the new method in comparison with the methods mentioned. According to the statistical analysis, we can state that in some cases (e.g. when estimating bimodal or skewed densities) our method gives better results than other known methods and in numerous cases its efficiency is comparable to the one of the methods mentioned (see [**7**]).

2. The idea of the two-step method. Let $(\Omega, \mathfrak{F}, P)$ be a probability space, $X_i : \Omega \longrightarrow \mathbb{R}$ $(i = 1, 2, \ldots, n)$ – a sequence of random variables and suppose they are independently and identically distributed. We assume that f is an unknown density function of X_i and $\{x_i\}_{i=1}^n$ is the sequence of observations on X_i. Consider a family of functions $\{\phi_j(x)\}_{j=1}^m$, such that

$$0 \le \phi_j(x) \le 1, \quad \sum_{j=1}^{m} \phi_j(x) = 1, \quad x \in \mathbb{R},$$

and a sequence of corresponding bandwidths $\{h_j(n)\}_{j=1}^m$, such that $h_j(n) > 0$ and $h_j(n) \to 0$, $j = 1, \ldots, m$.

Assume also that the kernel $K : \mathbb{R} \to \mathbb{R}$ is a nonnegative measurable function.

DEFINITION 2.1. We define the generalized kernel density estimator as a function given by

$$\widehat{f}_\phi(x) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{\phi_j(x_i)}{h_j(n)} K\Big(\frac{x - x_i}{h_j(n)}\Big). \tag{2.1}$$

The proposed method consists of two steps:

(1) We choose a pilot function f_0, by parametric estimation, using the EM algorithm (see Section 4.1), under the assumption that f_0 is given as follows

$$f_0(x) = \sum_j \alpha_j f_j(x), \quad \text{where} \quad f_j(x) = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{1}{2}\left(\frac{x-\mu_j}{\sigma_j}\right)^2}, \ j = 1, \ldots, m.$$

(2) Considering the generalized kernel density estimator (nonparametric estimation) we choose a sequence of bandwidths $h_j(n) > 0$, $h_j(n) \to 0$, for $n \to \infty$, $j = 1, \ldots, m$, using the pilot calculated in step (1), by taking

$$\phi_j(x) = \frac{\alpha_j f_j(x)}{f_0(x)} = \frac{\alpha_j f_j(x)}{\sum_{k=1}^m \alpha_k f_k(x)}. \tag{2.2}$$

REMARK 2.2. The first modification with respect to the traditional kernel estimation method is the fact that we choose a sequence of bandwidths $h_j(n)$ for $j = 1, \ldots, m$ instead of the single value $h(n)$. This may cause some complication in calculations but guarantees a better accuracy of the estimator.

REMARK 2.3. Formula (2.1) corresponds to the methods described in **[1]**, **[13]** and in **[15]**. The idea of the construction of the pilot function by parametric estimation comes from **[6]**.

REMARK 2.4. Note that for $m = 1$, taking $\phi_1(x) \equiv 1$, we get the kernel density estimator (1.1), with the bandwidth $h_1(n)$.

REMARK 2.5. As in the case of the kernel density estimator (1.1), the kernel K has an essential effect on the properties of $\widehat{f}_\phi$. In particular, if K is a density, and so $K(x) \geq 0$ and $\int_{\mathbb{R}} K(x)\,dx = 1$, then also $\widehat{f}_\phi(x) \geq 0$ and

$$\int_{\mathbb{R}} \widehat{f}_\phi(x)\,dx = \int_{\mathbb{R}} \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} \frac{\phi_j(y)}{h_j(n)} K\Big(\frac{x-y}{h_j(n)}\Big)\,dx$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} \phi_j(y) \int_{\mathbb{R}} \frac{1}{h_j(n)} K\Big(\frac{x-y}{h_j(n)}\Big)\,dx.$$

Putting $\frac{x-y}{h_j(n)} = t$, for a fixed j, gives $\int_{\mathbb{R}} \widehat{f}_\phi(x)\,dx = \int_{\mathbb{R}} K(t)\,dt = 1$.

2.1. *Properties of the generalized kernel density estimator.* We will now present some statistical properties of $\widehat{f}_\phi$ and then use its asymptotic behaviour in a criterion for the choice of $h_j(n)$, $j = 1, \dots, m$.

Under the above assumptions and notation, the following lemma is true.

LEMMA 2.6. *Assume that* $E\Big(\phi_j(x_i)K\Big(\frac{x-x_i}{h_j(n)}\Big)\Big) < \infty$ *and*

$$E\Big(\phi_j(x_i)\phi_k(x_i)K\Big(\frac{x-x_i}{h_j(n)}\Big)K\Big(\frac{x-x_i}{h_k(n)}\Big)\Big) < \infty, \quad j,k = 1,\dots,m.$$

Then

(2.3)
$$E(\widehat{f}_\phi(x)) = \sum_{j=1}^{m} \int_{\mathbb{R}} K(y)\phi_j(x + h_j(n)y) f(x + h_j(n)y)\,dy,$$

(2.4)
$$V(\widehat{f}_\phi(x)) = \frac{1}{n}\sum_{j,k=1}^{n} \frac{1}{h_j(n)h_k(n)} \int_{\mathbb{R}} K\Big(\frac{x-y}{h_j(n)}\Big)K\Big(\frac{x-y}{h_k(n)}\Big)\phi_j(y)\phi_k(y)f(y)\,dy$$
$$-\frac{1}{n}\Big(\sum_{j=1}^{m} \int_{\mathbb{R}} K(y)\phi_j(x + h_j(n)y) f(x + h_j(n)y)\,dy\Big)^2.$$

PROOF. Fix an $n \in \mathbb{N}$. Since K is a measurable nonnegative function and $\{X_i\}_{i=1}^n$ is a sequence of identically and independently distributed random variables, $\Big\{\frac{\phi_j(x_i)}{h_j(n)}K\Big(\frac{x-x_i}{h_j(n)}\Big)\Big\}_{i=1}^n$ forms a sequence of identically and independently distributed random variables. Thus

$$E(\widehat{f}_\phi(x)) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} \frac{1}{h_j(n)} E\Big(\phi_j(x_i)K\Big(\frac{x-x_i}{h_j(n)}\Big)\Big)$$
$$= \sum_{j=1}^{m} \frac{1}{h_j(n)} E\Big(\phi_j(x_i)K\Big(\frac{x-x_i}{h_j(n)}\Big)\Big).$$

Observe that, for every $j = 1, \ldots, m$

$$\frac{1}{h_j(n)} E\Big(\phi_j(x_i) K\Big(\frac{x - x_i}{h_j(n)}\Big)\Big) = (h_j(n))^{-1} \int_{\mathbb{R}} K\Big(\frac{x - t}{h_j(n)}\Big) \phi_j(t) f(t)\, dt.$$

Letting $y = \frac{x-t}{h_j(n)}$, we have $\left|\frac{dy}{dt}\right| = \frac{1}{h_j(n)}$ and

$$\frac{1}{h_j(n)} \int_{\mathbb{R}} K(y) \phi_j(x + h_j(n)y) f(x + h_j(n)y) h_j(n)\, dy$$
$$= \int_{\mathbb{R}} K(y) \phi_j(x + h_j(n)y) f(x + h_j(n)y)\, dy.$$

Hence, summing both sides over $j = 1, \ldots, m$, we get the expected value of the generalized kernel estimator

$$E(\widehat{f}_\phi(x)) = \sum_{j=1}^{m} \int_{\mathbb{R}} K(y) \phi_j(x + h_j(n)y) f(x + h_j(n)y)\, dy.$$

To prove (2.4), we use the following equation

$$V(\widehat{f}_\phi(x)) = E(\widehat{f}_\phi^2(x)) - (E(\widehat{f}_\phi(x))^2.$$

There is

$$\begin{aligned} V(\widehat{f}_\phi(x)) = \frac{1}{n^2} \sum_{i=1}^{n} \Big(& E\Big(\sum_{j=1}^{m} \frac{\phi_j(x_i)}{h_j(n)} K\Big(\frac{x - x_i}{h_j(n)}\Big)\Big)^2 \\ & - \Big[E\Big(\sum_{j=1}^{m} \frac{\phi_j(x_i)}{h_j(n)} K\Big(\frac{x - x_i}{h_j(n)}\Big)\Big)\Big]^2\Big) \\ = \frac{1}{n^2} \sum_{i=1}^{n} \Big(& E\Big(\sum_{j=1}^{m} \frac{\phi_j(x_i)}{h_j(n)} K\Big(\frac{x - x_i}{h_j(n)}\Big)\Big)\Big(\sum_{k=1}^{m} \frac{\phi_k(x_i)}{h_k(n)} K\Big(\frac{x - x_i}{h_k(n)}\Big)\Big) \\ & - \Big[\sum_{j=1}^{m} \frac{1}{h_j(n)} E\Big(\phi_j(x_i) K\Big(\frac{x - x_i}{h_j(n)}\Big)\Big)\Big]^2\Big). \end{aligned}$$

Thanks to the independence of x_i, $i = 1, \dots, n$, the first term can be written as:

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i=1}^{n}\Big(E\Big(\sum_{j=1}^{m}\frac{\phi_j(x_i)}{h_j(n)}K\Big(\frac{x-x_i}{h_j(n)}\Big)\Big)\Big(\sum_{k=1}^{m}\frac{\phi_k(x_i)}{h_k(n)}K\Big(\frac{x-x_i}{h_k(n)}\Big)\Big)\\
&=\frac{1}{n}\int_{\mathbb{R}}\Big(\sum_{j=1}^{m}\frac{\phi_j(y)}{h_j(n)}K\Big(\frac{x-y}{h_j(n)}\Big)\Big)\Big(\sum_{k=1}^{m}\frac{\phi_k(y)}{h_k(n)}K\Big(\frac{x-y}{h_k(n)}\Big)\Big)f(y)\,dy\\
&=\frac{1}{n}\int_{\mathbb{R}}\sum_{j=1}^{m}\sum_{k=1}^{m}\frac{\phi_j(y)\phi_k(y)}{h_j(n)h_k(n)}K\Big(\frac{x-y}{h_j(n)}\Big)K\Big(\frac{x-y}{h_k(n)}\Big)f(y)\,dy\\
&=\frac{1}{n}\sum_{j=1}^{m}\sum_{k=1}^{m}\frac{1}{h_j(n)h_k(n)}\int_{\mathbb{R}}K\Big(\frac{x-y}{h_j(n)}\Big)K\Big(\frac{x-y}{h_k(n)}\Big)\phi_j(y)\phi_k(y)f(y)\,dy.
\end{aligned}$$

Hence the variance of the generalized kernel estimator is given by

$$\begin{aligned}
V(\widehat{f}_\phi(x)) = \frac{1}{n}\sum_{j,k=1}^{n}\frac{1}{h_j(n)h_k(n)}\int_{\mathbb{R}}K\Big(\frac{x-y}{h_j(n)}\Big)K\Big(\frac{x-y}{h_k(n)}\Big)\phi_j(y)\phi_k(y)f(y)\,dy\\
-\frac{1}{n}\Big(\sum_{j=1}^{m}\int_{\mathbb{R}}K(y)\phi_j(x+h_j(n)y)f(x+h_j(n)y)\,dy\Big)^2,
\end{aligned}$$

which completes the proof. □

REMARK 2.7. Taking $m = 1$ and $\phi_1(x) \equiv 1$ in Theorem 2.6, we obtain the formulae for the expected value and variance of the traditional kernel estimator $\widehat{f}_h$. They can be found for example in **[13]**.

In order to show some statistical properties of $\widehat{f}_\phi$, we use the following lemma (see **[10]** for a similar result for the traditional kernel estimator).

LEMMA 2.8. *Under the above assumptions, let K be a bounded and integrable function such that $\lim_{|t|\to\infty} tK(t) = 0$, and let $h_j(n) > 0$, $h_j(n) \to 0$, for $n \to \infty$ and $j = 1, \dots, m$. Then for every point x of continuity of f there is*

$$\sum_{j=1}^{m}\frac{1}{h_j(n)}E\Big(\phi_j(x_i)K\Big(\frac{x-x_i}{h_j(n)}\Big)\Big) \longrightarrow f(x)\int_{\mathbb{R}}K(t)\,dt, \quad n\to\infty.$$

PROOF. Our goal is to show that for every $j = 1, \dots, m$

$$\lim_{n\to\infty}\int_{\mathbb{R}}(f(x+h_j(n)z)\phi_j(x+h_j(n)z) - f(x)\phi_j(x+h_j(n)z))K(z)\,dz = 0.$$

For a fixed j, there is

$$\Big|\int_{\mathbb{R}}(f(x+h_j(n)z)\phi_j(x+h_j(n)z)-f(x)\phi_j(x+h_j(n)z))K(z)\,dz\Big|$$
$$\leq\int_{\mathbb{R}}|(f(x+h_j(n)z)\phi_j(x+h_j(n)z)-f(x)\phi_j(x+h_j(n)z))K(z)|\,dz.$$

Letting $-\zeta=h_j(n)z$, we obtain $\left|\frac{dz}{d\zeta}\right|=\frac{1}{h_j(n)}$, and

$$\int_{\mathbb{R}}\Big|(f(x-\zeta)\phi_j(x-\zeta)-f(x)\phi_j(x-\zeta))K\Big(-\frac{\zeta}{h_j(n)}\Big)\Big|\frac{1}{h_j(n)}\,d\zeta$$
$$=\frac{1}{h_j(n)}\int_{\mathbb{R}}\Big|(f(x-\zeta)\phi_j(x-\zeta)-f(x)\phi_j(x-\zeta))K\Big(-\frac{\zeta}{h_j(n)}\Big)\Big|\,d\zeta,$$

where

$$\Big|(f(x-\zeta)\phi_j(x-\zeta)-f(x)\phi_j(x-\zeta))K\Big(-\frac{\zeta}{h_j(n)}\Big)\Big|$$
$$=\Big|f(x-\zeta)-f(x)\Big|\Big|\phi_j(x-\zeta)\Big|K\Big(-\frac{\zeta}{h_j(n)}\Big),$$

for K is nonnegative. Observe that $|\phi_j(x-\zeta)|\leq 1$, whenever $x-\zeta\in\mathbb{R}$. Therefore,

$$\frac{1}{h_j(n)}\int_{\mathbb{R}}\Big|f(x-\zeta)-f(x)\Big|\Big|\phi_j(x-\zeta)\Big|K\Big(-\frac{\zeta}{h_j(n)}\Big)\,d\zeta$$
$$\leq\frac{1}{h_j(n)}\int_{\mathbb{R}}\Big|f(x-\zeta)-f(x)\Big|K\Big(-\frac{\zeta}{h_j(n)}\Big)\,d\zeta.$$

Take an arbitrary $\delta>0$. Splitting the integration interval into two parts: $\{\zeta:\ |\zeta|\geq\delta\}$ and $\{\zeta:\ |\zeta|<\delta\}$ and applying the triangle inequality to the first term, we get

$$\frac{1}{h_j(n)}\int_{\{\zeta:\,|\zeta|\geq\delta\}}\Big|f(x-\zeta)-f(x)\Big|K\Big(-\frac{\zeta}{h_j(n)}\Big)\,d\zeta$$
$$\leq\frac{1}{h_j(n)}\int_{\{\zeta:\,|\zeta|\geq\delta\}}\Big|f(x-\zeta)\Big|K\Big(-\frac{\zeta}{h_j(n)}\Big)\,d\zeta$$
$$+\frac{1}{h_j(n)}\int_{\{\zeta:\,|\zeta|\geq\delta\}}\Big|f(x)\Big|K\Big(-\frac{\zeta}{h_j(n)}\Big)\,d\zeta.$$

Since K is bounded and f is a density, we get

$$\frac{1}{h_j(n)}\int_{\{\zeta:\,|\zeta|\geq\delta\}}\frac{f(x-\zeta)}{\zeta}\zeta K\Big(-\frac{\zeta}{h_j(n)}\Big)\,d\zeta \leq \sup_{|y|\geq\frac{\delta}{h_j(n)}} yK(y)\int_{\{\zeta:\,|\zeta|\geq\delta\}} f(x-\zeta)\,d\zeta$$

and

$$\frac{1}{h_j(n)}\int_{\{\zeta:\,|\zeta|\geq\delta\}} f(x)K\Big(-\frac{\zeta}{h_j(n)}\Big)\,d\zeta = \frac{1}{h_j(n)}f(x)\int_{\{\zeta:\,|\zeta|\geq\delta\}} K\Big(-\frac{\zeta}{h_j(n)}\Big)\,d\zeta = f(x)\int_{|z|\geq\frac{\delta}{h_j(n)}} K(z)\,dz.$$

Hence, for $n\to\infty$, the first term tend to 0:

$$\frac{1}{h_j(n)}\int_{\{\zeta:\,|\zeta|\geq\delta\}}\Big|f(x-\zeta)-f(x)\Big|K\Big(-\frac{\zeta}{h_j(n)}\Big)\,d\zeta \leq \sup_{|y|\geq\frac{\delta}{h_j(n)}} yK(y)\int_{\{\zeta:\,|\zeta|\geq\delta\}} f(x-\zeta)\,d\zeta + f(x)\int_{|z|\geq\frac{\delta}{h_j(n)}} K(z)\,dz \longrightarrow 0.$$

From the continuity of f at x, the second integral may be estimated as follows

$$\frac{1}{h_j(n)}\int_{\{\zeta:\,|\zeta|<\delta\}}\Big|f(x-\zeta)-f(x)\Big|K\Big(-\frac{\zeta}{h_j(n)}\Big)\,d\zeta \leq \frac{\varepsilon_\delta}{h_j(n)}\int_{\{\zeta:\,|\zeta|<\delta\}} K\Big(-\frac{\zeta}{h_j(n)}\Big)\,d\zeta = \varepsilon_\delta\int_{\{y:|y|<\frac{\delta}{h_j(n)}\}} K(y)\,dy$$

and converges to zero as $n\to\infty$.

Thus, for every $j=1,\dots,m$,

$$\frac{1}{h_j(n)}E\Big(\phi_j(x_i)K\Big(\frac{x-x_i}{h_j(n)}\Big)\Big)\longrightarrow f(x)\phi_j(x)\int_{\mathbb{R}}K(t)\,dt,\ \text{for}\quad n\to\infty.$$

After summing both sides over j the proof is completed. □

Remark 2.9. By the additional assumption that $\int_{\mathbb{R}}K(t)\,dt=1$,

$$E(\widehat{f}_\phi(x))-f(x)\to 0,\qquad n\to\infty,$$

and hence the generalized kernel estimator is asymptotically unbiased. Provided that for every $j=1,\dots,m$, $nh_j(n)\to\infty$ for $n\to\infty$, we obtain $V(\widehat{f}_\phi)\to 0$ for $n\to\infty$.

3. Criterion for the bandwidth choice. Like in the case of the traditional kernel estimator (1.1), we will use asymptotic mean integrated squared error AMISE($\widehat{f}_\phi$) as a criterion for the choice of $h_j(n) > 0$, $h_j(n) \to 0$, for $n \to \infty$, $j = 1, \dots, m$. We obtain the formula for AMISE($\widehat{f}_\phi$) by asymptotic approximations of $E(\widehat{f}_\phi(x)) - f(x)$ and $V(\widehat{f}_\phi)$. We will use a well known corollary of Taylor's formula:

COROLLARY 3.1. *Suppose that $g(x)$ is an arbitrary function, $(n-1)$-times differentiable in a neighbourhood of x_0 and that $g^{(n)}(x_0)$ at x_0 exists. Then*

$$\forall_{\varepsilon>0}\ \exists_{\delta>0,\,0<\theta<\delta}:\ \Big|g(x_0+\theta) - g(x_0) - \theta g'(x_0) - \dots - \frac{\theta^n}{n!}g^{(n)}(x_0)\Big| < \frac{\theta^n}{n!}\varepsilon.$$

Under the assumption from the last section we will formulate the criterion for the choice of the sequence of bandwidths. From now on, we assume that K is a measurable nonnegative function such that

$$\int_{\mathbb{R}} t^2 K(t)\,dt < \infty \qquad \text{and} \qquad \int_{\mathbb{R}} K(t)\,dt = 1.$$

We will consider two cases: for a symmetric and nonsymmetric kernel, starting with the symmetric case.

THEOREM 3.2. *Under the assumptions from the last section, suppose that functions $\phi_j f$, $j = 1, \dots, m$ are differentiable in a neighbourhood of x and there exist derivatives $(\phi_j(x)f(x))''$, $j = 1, \dots, m$ at x, and a kernel K satisfies $\int_{\mathbb{R}} tK(t)\,dt = 0$. Then*

$$E(\widehat{f}_\phi(x)) = f(x) + \frac{1}{2}\sum_{j=1}^{m} h_j(n)^2(\phi_j(x)f(x))'' \int_{\mathbb{R}} t^2 K(t)\,dt + o\Big(\sum_j h_j(n)^2\Big),$$

$$V(\widehat{f}_\phi(x)) = \frac{1}{n}\sum_{j,k=1}^{m} \frac{1}{h_j(n)h_k(n)} \int_{\mathbb{R}} K\Big(\frac{x-y}{h_j(n)}\Big)K\Big(\frac{x-y}{h_k(n)}\Big)\phi_j(y)\phi_k(y)f(y)\,dy + o\Big(\frac{1}{n\sum_j h_j(n)}\Big).$$

PROOF. Let $\varepsilon > 0$. By Corollary 3.1, for every $j = 1, \dots, m$, there exists a $\delta > 0$ such that $0 < h_j(n)t < \delta$ and

$$|\phi_j(x + h_j(n)t)f(x + h_j(n)t) - \phi_j(x)f(x) - h_j(n)t(\phi_j(x)f(x))' - \frac{1}{2}h_j(n)^2t^2(\phi_j(x)f(x))''| < \frac{h_j(n)^2t^2}{2!}\varepsilon.$$

Multiplying inequalities

$$-\frac{h_j(n)^2t^2}{2!}\varepsilon < \phi_j(x+h_j(n)t)f(x+h_j(n)t) - \phi_j(x)f(x) - h_j(n)t(\phi_j(x)f(x))' - \frac{1}{2}h_j(n)^2t^2(\phi_j(x)f(x))'' < \frac{h_j(n)^2t^2}{2!}\varepsilon$$

by $K(t)$, and integrating with respect to t, we obtain

$$-\frac{h_j(n)^2}{2!}\varepsilon\int_{\mathbb{R}} t^2K(t)\,dt < \int_{\mathbb{R}} \phi_j(x+h_j(n)t)f(x+h_j(n)t)K(t)\,dt - \phi_j(x)f(x) - \frac{1}{2}h_j(n)^2(\phi_j(x)f(x))''\int_{\mathbb{R}} t^2K(t)\,dt < \frac{h_j(n)^2}{2!}\varepsilon\int_{\mathbb{R}} t^2K(t)\,dt,$$

provided that K is a symmetric density. Summing over $j = 1, \dots, m$, by (2.3), we get

$$-\frac{1}{4}\sum_{j=1}^{m} h_j(n)^2\varepsilon\int_{\mathbb{R}} t^2K(t)\,dt < E(\hat{f}_\phi(x)) - f(x) - \frac{1}{2}\sum_{j=1}^{m} h_j(n)^2(\phi_j(x)f(x))''\int_{\mathbb{R}} t^2K(t)\,dt < \frac{1}{4}\sum_{j=1}^{m} h_j(n)^2\varepsilon\int_{\mathbb{R}} t^2K(t)\,dt.$$

Therefore,

$$\Big|E(\widehat{f}_\phi(x)) - f(x) - \frac{1}{2}\sum_{j=1}^{m} h_j(n)^2(\phi_j(x)f(x))''\int_{\mathbb{R}} t^2K(t)\,dt\Big| < \frac{1}{4}\sum_{j=1}^{m} h_j(n)^2\varepsilon\int_{\mathbb{R}} t^2K(t)\,dt.$$

Since $h_j(n) > 0$, $j = 1, \dots, m$, denoting $\mu_2 = \int_{\mathbb{R}} t^2K(t)\,dt$, we obtain

$$\left|\frac{E(\widehat{f}_\phi(x)) - f(x)}{\sum_{j=1}^{m} h_j(n)^2} - \frac{1}{2}(\phi_j(x)f(x))''\mu_2\right| < \frac{1}{4}\varepsilon\mu_2.$$

Observe that $\forall_k\ h_k(n)^2 \le \sum_{j=1}^{m} h_j(n)^2$ and $h_j(n) \to 0, \quad n \to \infty$ for $j = 1, \dots, m$. Since $\varepsilon > 0$ is arbitrary, we get

$$E(\widehat{f}_\phi(x)) - f(x) - \frac{1}{2}(\phi_j(x)f(x))''\mu_2 = o\Big(\sum_{j=1}^{m} h_j(n)^2\Big).$$

Now, since

$$V(\widehat{f}_\phi(x)) = \frac{1}{n^2}\sum_{i=1}^{n}\Big(E\Big(\sum_{j=1}^{m}\frac{\phi_j(x_i)}{h_j(n)}K\Big(\frac{x-x_i}{h_j(n)}\Big)\Big)\Big(\sum_{k=1}^{m}\frac{\phi_k(x_i)}{h_k(n)}K\Big(\frac{x-x_i}{h_k(n)}\Big)\Big) - \Big[\sum_{j=1}^{m}\frac{1}{h_j(n)}E\Big(\phi_j(x_i)K\Big(\frac{x-x_i}{h_j(n)}\Big)\Big)\Big]^2\Big),$$

applying Lemma 2.8 and Remark 2.9, we conclude that the second term is $o(\frac{1}{n\sum_j h_j(n)})$. By (2.4), the first term equals

$$\frac{1}{n}\sum_{j,k=1}^{m}\frac{1}{h_j(n)h_k(n)}\int_{\mathbb{R}}K\Big(\frac{x-y}{h_j(n)}\Big)K\Big(\frac{x-y}{h_k(n)}\Big)\phi_j(y)\phi_k(y)f(y)\,dy,$$

which completes the proof. □

In the nonsymmetric case, we proceed in the following way.

THEOREM 3.3. *Under the assumptions from the last section, if the functions $f\phi_j$ are differentiable at x, for $j = 1,\dots,m$ and*

$$\int_{\mathbb{R}} tK(t)\,dt \neq 0, \text{ and } \Big|\int_{\mathbb{R}} tK(t)\,dt\Big| < \infty,$$

then

$$E(\widehat{f}_\phi(x)) = f(x) + \sum_{j=1}^{m} h_j(n)(\phi_j(x)f(x))'\int_{\mathbb{R}} tK(t)\,dt + o\Big(\sum_j h_j(n)\Big),$$

$$V(\widehat{f}_\phi(x)) = \frac{1}{n}\sum_{j,k=1}^{m}\frac{1}{h_j(n)h_k(n)}\int_{\mathbb{R}}K\Big(\frac{x-y}{h_j(n)}\Big)K\Big(\frac{x-y}{h_k(n)}\Big)\phi_j(y)\phi_k(y)f(y)\,dy + o\Big(\frac{1}{n\sum_j h_j(n)}\Big).$$

PROOF. Let $\varepsilon > 0$ be arbitrary. By the Peano formula, for every $j = 1,\dots,m$, there exists a $\delta > 0$, $0 < h_j(n)t < \delta$ such that:

$$|\phi_j(x + h_j(n)t)f(x + h_j(n)t) - \phi_j(x)f(x) - h_j(n)t(\phi_j(x)f(x))'| < h_j(n)t\varepsilon.$$

Multiplying by $K(t)$, and integrating with respect to t, we get by assumptions on K,

$$\Big|\int_{\mathbb{R}}\phi_j(x + h_j(n)t)f(x + h_j(n)t)K(t)\,dt - \phi_j(x)f(x) - h_j(n)(\phi_j(x)f(x))'\mu_1\Big| < h_j(n)\varepsilon\mu_1,$$

where $\mu_1 = \int_{\mathbb{R}} tK(t)\,dt$.

Summing over $j = 1, \dots, m$, by (2.3), we obtain

$$\Big|E(\widehat{f}_\phi(x)) - f(x) - \sum_{j=1}^{m} h_j(n)(\phi_j(x)f(x))'\mu_1\Big| < \sum_{j=1}^{m} h_j(n)\varepsilon\mu_1.$$

Therefore,

$$\left|\frac{E(\widehat{f}_\phi(x)) - f(x)}{\sum_{j=1}^{m} h_j(n)} - (\phi_j(x)f(x))'\mu_1\right| < \varepsilon\mu_1,$$

since for every j, $h_j(n) \to 0$, $n \to \infty$ and hence $\sum_j h_j(n)$ is bounded, so

$$E(\widehat{f}_\phi(x)) - f(x) = \sum_{j=1}^{m} h_j(n)\mu_1(\phi_j(x)f(x))' = o\Big(\sum_j h_j(n)\Big).$$

In the second part of the proof we can proceed as in the symmetric case (see the proof of Theorem 3.2). □

Under the above assumptions, the following corollaries are true.

COROLLARY 3.4. *By Theorem 3.2, if K is symmetric and*

$$\int_{\mathbb{R}} (\phi_j(x)f(x))''(\phi_k(x)f(x))'' \, dx < \infty,$$

for every $j, k = 1, \dots, m$, then the asymptotic mean integrated squared error is given by:

$$\text{(3.1)}\quad \text{AMISE}(\widehat{f}_\phi) = \sum_{j=1}^{m}\sum_{k=1}^{m} \frac{1}{4} h_j^2(n)h_k^2(n)\mu_2^2 \int_{\mathbb{R}} (\phi_j(x)f(x))''(\phi_k(x)f(x))'' \, dx$$
$$+ \frac{1}{n}\sum_{j=1}^{m}\sum_{k=1}^{m} \frac{1}{h_j(n)h_k(n)}\Big(\int_{\mathbb{R}} K\Big(\frac{t}{h_j(n)}\Big)K\Big(\frac{t}{h_k(n)}\Big)\, dt\Big)\int_{\mathbb{R}} \phi_j(y)\phi_k(y)f(y)\, dy,$$

where $\mu_2 = \int_{\mathbb{R}} t^2 K(t)\, dt$.

COROLLARY 3.5. *By Theorem 3.3, if K is nonsymmetric and*

$$\int_{\mathbb{R}} (\phi_j(x)f(x))'(\phi_k(x)f(x))' \, dx < \infty,$$

for every $j,k=1,\dots,m$, then the asymptotic mean integrated squared error is given by

$$\text{AMISE}(\widehat{f}_\phi)=\sum_{j=1}^m\sum_{k=1}^m h_j(n)h_k(n)\mu_1^2\int_{\mathbb{R}}(\phi_j(x)f(x))'(\phi_k(x)f(x))'\,dx \tag{3.2}$$
$$+\frac{1}{n}\sum_{j=1}^m\sum_{k=1}^m\frac{1}{h_j(n)h_k(n)}\Big(\int_{\mathbb{R}}K\Big(\frac{t}{h_j(n)}\Big)K\Big(\frac{t}{h_k(n)}\Big)\,dt\Big)\int_{\mathbb{R}}\phi_j(y)\phi_k(y)f(y)\,dy,$$

where $\mu_1=\int_{\mathbb{R}}tK(t)\,dt$.

Unfortunately the complexity of formulae (3.1) and (3.2) prevent the derivation of the explicit formulae for $h_j(n),\ j=1,\dots,m$, hence the choice of the sequence of bandwidths should be done numerically, by minimization of AMISE, for a symmetric and nonsymmetric kernel, respectively.

Obviously the above formulae get simpler if we know the kernel. For example, in the case of the Gaussian kernel and for fixed j,k, we get $\mu_2=\int_{\mathbb{R}}t^2K_g(t)\,dt=1$ and

$$\begin{aligned}\int_{\mathbb{R}}K_g\Big(\frac{t}{h_j(n)}\Big)K_g\Big(\frac{t}{h_k(n)}\Big)\,dt&=\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}e^{-\frac{t^2}{2}\left(\frac{1}{h_j(n)^2}+\frac{1}{h_k(n)^2}\right)}\,dt\\&=\frac{1}{\sqrt{2\pi}}\frac{h_j(n)h_k(n)}{\sqrt{h_j(n)^2+h_k(n)^2}}.\end{aligned}$$

Therefore,

$$\begin{aligned}\text{AMISE}(\widehat{f}_\phi)=\sum_{j=1}^m\sum_{k=1}^m\Big(&\frac{1}{4}h_j^2(n)h_k^2(n)\int_{\mathbb{R}}(\phi_j(x)f(x))''(\phi_k(x)f(x))''\,dx\\&+\frac{1}{\sqrt{2\pi}n}\frac{1}{\sqrt{h_j(n)^2+h_k(n)^2}}\int_{\mathbb{R}}\phi_j(x)\phi_k(x)f(x)\,dx\Big).\end{aligned}$$

In case of the rectangular kernel, for fixed j,k, $\mu_2=\int_{\mathbb{R}}t^2K_p(t)\,dt=\frac{1}{3}$ and

$$\int_{\mathbb{R}}K_p\Big(\frac{t}{h_j(n)}\Big)K_p\Big(\frac{t}{h_k(n)}\Big)\,dt=\frac{1}{2}\min\{h_j(n),h_k(n)\},$$

and hence

$$\begin{aligned}\text{AMISE}(\widehat{f}_\phi)=\sum_{j=1}^m\sum_{k=1}^m\Big(&\frac{1}{36}h_j^2(n)h_k^2(n)\int_{\mathbb{R}}(\phi_j(x)f(x))''(\phi_k(x)f(x))''\,dx\\&+\frac{\min\{h_j(n),h_k(n)\}}{2nh_j(n)h_k(n)}\int_{\mathbb{R}}\phi_j(x)\phi_k(x)f(x)\,dx\Big).\end{aligned}$$

Using the nonsymmetric exponential kernel, we obtain

$$\operatorname{AMISE}(\widehat{f}_\phi) = \sum_{j=1}^{m}\sum_{k=1}^{m}\Big(h_j(n)h_k(n)\int_{\mathbb{R}}(\phi_j(x)f(x))'(\phi_k(x)f(x))'\,dx + \frac{h_j(n)h_k(n)}{n(h_j(n)+h_k(n))}\int_{\mathbb{R}}\phi_j(x)\phi_k(x)f(x)\,dx\Big),$$

since for fixed j,k, there is $\mu_1 = \int_{\mathbb{R}} tK_w(t)\,dt = 1$

$$\int_{\mathbb{R}} K_w\Big(\frac{t}{h_j(n)}\Big)K_w\Big(\frac{t}{h_k(n)}\Big)\,dt = \int_0^{+\infty} e^{-t\left(\frac{1}{h_j(n)^2}+\frac{1}{h_k(n)^2}\right)}\,dt = \frac{h_j(n)h_k(n)}{h_j(n)+h_k(n)}.$$

Remark 3.6. If K is symmetric, taking $m=1$ and $\phi_1(x)\equiv 1$ (see Remark 2.4), by (3.1), we obtain:

$$\begin{aligned}\operatorname{AMISE}(\widehat{f}_\phi) &= \frac{h_1(n)^4}{4}\left(\int_{\mathbb{R}} t^2K(t)\,dt\right)^2\int_{\mathbb{R}}((\phi_1(x)f(x))'')^2\,dx \\ &\quad + \frac{1}{nh_1(n)}\int_{\mathbb{R}}K^2(t)\,dt\int_{\mathbb{R}}\phi_1(x)^2f(x)\,dx \\ &= \frac{h_1(n)^4}{4}\left(\int_{\mathbb{R}} t^2K(t)\,dt\right)^2\int_{\mathbb{R}}((f(x))'')^2\,dx + \frac{1}{nh_1(n)}\int_{\mathbb{R}}K^2(t)\,dt,\end{aligned}$$

which gives the formula for AMISE for the traditional kernel estimator (1.1).

Remark 3.7. Using nonsymmetric kernel, for $m=1$, and $\phi_1(x)\equiv 1$, by (3.2), we get the formula for AMISE for estimator (1.1):

$$\begin{aligned}\operatorname{AMISE}(\widehat{f}_\phi) &= h_1(n)^2\left(\int_{\mathbb{R}} tK(t)\,dt\right)^2\int_{\mathbb{R}}((\phi_1(x)f(x))')^2\,dx \\ &\quad + \frac{1}{nh_1(n)}\int_{\mathbb{R}}K^2(t)\,dt\int_{\mathbb{R}}\phi_1(x)^2f(x)\,dx \\ &= h_1(n)^2\left(\int_{\mathbb{R}} tK(t)\,dt\right)^2\int_{\mathbb{R}}((f(x))')^2\,dx + \frac{1}{nh_1(n)}\int_{\mathbb{R}}K^2(t)\,dt\end{aligned}$$

(see **[7]** for results on the nonsymmetric case in the traditional kernel estimation).

In Section 4 we will construct the pilot function by parametric estimation. The main problem is the choice of number of components in the model

$$\alpha_1 f_1(x|\theta_1) + \alpha_2 f_2(x|\theta_2) + \cdots + \alpha_m f_m(x|\theta_m).$$

This is a known and much investigated problem in many methods of parametric estimation. However, in our case, it is not so important, since the pilot is just

a tool for further estimation. So we suggest the choice of $m = 2$, because it does not decrease the efficiency of the method considered.

REMARK 3.8. In the case $m = 2$, we choose two bandwidths $h_1(n)$ and $h_2(n)$, minimizing:

- for the Gaussian kernel:

(3.3)

$$\begin{aligned}\mathrm{AMISE}(\widehat{f}_\phi) &= \frac{1}{2}h_1^2(n)h_2^2(n)\int_{\mathbb{R}}(\phi_1(x)f(x))''(\phi_2(x)f(x))''\,dx\\ &\quad+\frac{1}{4}h_1^4(n)\int_{\mathbb{R}}((\phi_1(x)f(x))'')^2\,dx+\frac{1}{4}h_2^4(n)\int_{\mathbb{R}}((\phi_2(x)f(x))'')^2\,dx\\ &\quad+\frac{1}{2nh_1(n)\sqrt{\pi}}\int_{\mathbb{R}}\phi_1^2(x)f(x)\,dx+\frac{1}{2nh_2(n)\sqrt{\pi}}\int_{\mathbb{R}}\phi_2^2(x)f(x)\,dx\\ &\quad+\frac{\sqrt{2}}{n\sqrt{\pi(h_1^2(n)+h_2^2(n))}}\int_{\mathbb{R}}\phi_1(x)\phi_2(x)f(x)\,dx,\end{aligned}$$

- for the rectangular kernel:

$$\begin{aligned}\mathrm{AMISE}(\widehat{f}_\phi) &= \frac{1}{18}h_1^2(n)h_2^2(n)\int_{\mathbb{R}}(\phi_1(x)f(x))''(\phi_2(x)f(x))''\,dx\\ &\quad+\frac{1}{36}h_1^4(n)\int_{\mathbb{R}}((\phi_1(x)f(x))'')^2\,dx+\frac{1}{36}h_2^4(n)\int_{\mathbb{R}}((\phi_2(x)f(x))'')^2\,dx\\ &\quad+\frac{1}{2nh_1(n)}\int_{\mathbb{R}}\phi_1^2(x)f(x)\,dx+\frac{1}{2nh_2(n)}\int_{\mathbb{R}}\phi_2^2(x)f(x)\,dx\\ &\quad+\frac{\min\{h_1(n),h_2(n)\}}{nh_1(n)h_2(n)}\int_{\mathbb{R}}\phi_1(x)\phi_2(x)f(x)\,dx,\end{aligned}$$

- for the exponential kernel:

$$\begin{aligned}\mathrm{AMISE}(\widehat{f}_\phi) &= h_1^2(n)\int_{\mathbb{R}}((\phi_1(x)f(x))')^2\,dx+h_2^2(n)\int_{\mathbb{R}}((\phi_2(x)f(x))')^2\,dx\\ &\quad+\frac{h_1(n)}{2n}\int_{\mathbb{R}}\phi_1^2(x)f(x)\,dx+\frac{h_2(n)}{2n}\int_{\mathbb{R}}\phi_2^2(x)f(x)\,dx\\ &\quad+2h_1(n)h_2(n)\int_{\mathbb{R}}(\phi_1(x)f(x))'(\phi_2(x)f(x))'\,dx\\ &\quad+2\frac{h_1(n)h_2(n)}{n(h_1(n)+h_2(n))}\int_{\mathbb{R}}\phi_1(x)\phi_2(x)f(x)\,dx.\end{aligned}$$

4. Construction of the pilot function – parametric estimation. In this section we present some way of construction of the pilot function, based on a modification of the well-known maximum likelihood method. We apply the Expectation Maximization Algorithm, introduced in **[2]** and used to estimate

the parameters in stochastic models, to hidden Markov models, to estimate a hazard function, as well as to estimate the parameters in mixture models. The last of these applications is the matter of our interest and will be used in the normal mixture model

$$\alpha_1 N(\mu_1, \sigma_1^2) + \alpha_2 N(\mu_2, \sigma_2^2) + \cdots + \alpha_m N(\mu_m, \sigma_m^2).$$

4.1. *The EM Algorithm.* Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ be a sample of independent observations from an absolutely continuous distribution with a density function $f_X(x|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes the parameters of the distribution.

Definition 4.1. We define the *likelihood function* and the *log-likelihood function*

$$L_{\boldsymbol{x}}(\boldsymbol{\theta}|\boldsymbol{x}) := \prod_{i=1}^{n} f_X(x_i|\theta),$$

$$l_{\boldsymbol{x}}(\boldsymbol{\theta}|\boldsymbol{x}) := \ln L_{\boldsymbol{x}}(\boldsymbol{\theta}|\boldsymbol{x}) = \sum_{i=1}^{n} \ln f_X(x_i|\theta).$$

The *maximum likelihood estimator* of $\boldsymbol{\theta}$ is given by

$$\widehat{\boldsymbol{\theta}} = \operatorname*{argmax}_{\boldsymbol{\theta}\in\Theta} L_{\boldsymbol{x}}(\boldsymbol{\theta}|\boldsymbol{x}) = \operatorname*{argmax}_{\boldsymbol{\theta}\in\Theta} l_{\boldsymbol{x}}(\boldsymbol{\theta}|\boldsymbol{x}).$$

To present the EM algorithm we consider data model

$$\{(x_i, y_i),\ i = 1, \ldots, n\}, \tag{4.1}$$

from the distribution of the random variable (X, Y), where $\boldsymbol{x} = (x_1, \ldots, x_n)$ is a sample of independent observations from the absolutely continuous distribution of X, with the density f_X (treated as a marginal density of (X, Y)), and $\boldsymbol{y} = (y_1, \ldots, y_n)$ denotes the latent or missing data. Sample $\boldsymbol{x}$, together with $\boldsymbol{y}$ is called the *complete data*.

Definition 4.2. For data model (4.1), the *likelihood* and *log-likelihood functions* are defined as follows:

$$L(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{y}) = \prod_{i=1}^{n} f(x_i, y_i|\theta), \tag{4.2}$$

$$l(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{y}) = \ln L(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} \ln f(x_i, y_i|\theta). \tag{4.3}$$

From now on we will use the following notation

$$f_X(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{i=1}^{n} f_X(x_i|\theta), \tag{4.4}$$

$$f(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i,y_i|\theta), \tag{4.5}$$

$$f_{Y|X}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}) = \prod_{i=1}^{n} f_{Y|X}(y_i|x_i,\theta). \tag{4.6}$$

COROLLARY 4.3. *For the considered data model, the maximum likelihood estimator of θ is given by*

$$\widehat{\boldsymbol{\theta}} = \operatorname*{argmax}_{\boldsymbol{\theta}\in\Theta} \left(l_{\boldsymbol{x}}(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{y}) - \ln f_{Y|X}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})\right),$$

where $f_{Y|X}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})$ denotes the conditional density of Y, given $X=\boldsymbol{x}$ and $\boldsymbol{\theta}$.

PROOF. By the definition of conditional density,

$$f_X(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{f(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta})}{f_{Y|X}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})},$$

and hence, by Definition 4.1,

$$\begin{aligned}\widehat{\boldsymbol{\theta}} &= \operatorname*{argmax}_{\boldsymbol{\theta}\in\Theta} l_{\boldsymbol{x}}(\boldsymbol{\theta}|\boldsymbol{x}) = \operatorname*{argmax}_{\boldsymbol{\theta}\in\Theta}(\ln f(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta}) - \ln f_{Y|X}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})) \\ &= \operatorname*{argmax}_{\boldsymbol{\theta}\in\Theta}(l(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{y}) - \ln f_{Y|X}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})).\end{aligned}$$

□

Now, let

$$Q(\boldsymbol{\theta},\boldsymbol{\theta}_t) := E\left(\ln f(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta})\middle|\,\boldsymbol{x},\boldsymbol{\theta}_t\right), \tag{4.7}$$

$$H(\boldsymbol{\theta},\boldsymbol{\theta}_t) := E\left(\ln f_{Y|X}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})\middle|\,\boldsymbol{x},\boldsymbol{\theta}_t\right), \tag{4.8}$$

where $E(\cdot|\boldsymbol{x},\boldsymbol{\theta})$ denotes the conditional expected value, given $X=\boldsymbol{x}$ and fixed $\boldsymbol{\theta}=\boldsymbol{\theta}_t$, where $\boldsymbol{\theta}_t$ is the value of parameter $\boldsymbol{\theta}$ in the iteration t.

The EM algorithm gives an iterative procedure that maximizes $Q(\boldsymbol{\theta},\boldsymbol{\theta}_t)$, for every $t=0,1,2,\ldots$ giving $\boldsymbol{\theta}_{t+1} = \operatorname{argmax}_{\boldsymbol{\theta}} Q(\boldsymbol{\theta},\boldsymbol{\theta}_t)$. After sufficient number of iterations, $\operatorname{argmax}_{\boldsymbol{\theta}} l_{\boldsymbol{x}}(\boldsymbol{\theta}|\boldsymbol{x})$ will be found with a high probability (see **[2]**, **[16]**).

THEOREM 4.4. *Under the above assumptions the procedure of*

$$\boldsymbol{\theta}_{t+1} = \operatorname*{argmax}_{\boldsymbol{\theta}\in\Theta} Q(\boldsymbol{\theta},\boldsymbol{\theta}_t),$$

guarantees that

$$l_{\boldsymbol{x}}(\boldsymbol{\theta}_{t+1}|\boldsymbol{x}) \geq l_{\boldsymbol{x}}(\boldsymbol{\theta}_t|\boldsymbol{x}),$$

with equality iff

$$Q(\boldsymbol{\theta}_{t+1}|\boldsymbol{x},\boldsymbol{\theta}_t) = Q(\boldsymbol{\theta}_t|\boldsymbol{x},\boldsymbol{\theta}_t),$$

for $t = 0, 1, 2, \ldots$

PROOF. Consider the expression

$$Q(\boldsymbol{\theta}_{t+1}|\boldsymbol{x},\boldsymbol{\theta}_t) - H(\boldsymbol{\theta}_{t+1}|\boldsymbol{x},\boldsymbol{\theta}_t).$$

From formulae (4.7) and (4.8), there follows

$$\begin{aligned} Q(\boldsymbol{\theta}_{t+1}|\boldsymbol{x},\boldsymbol{\theta}_t) &- H(\boldsymbol{\theta}_{t+1}|\boldsymbol{x},\boldsymbol{\theta}_t) \\ &= E\left(\ln f(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta}_{t+1})|\boldsymbol{x},\boldsymbol{\theta}_t\right) - E\left(\ln f_{Y|X}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}_{t+1})|\boldsymbol{x},\boldsymbol{\theta}_t\right) \\ &= E\left(\ln \frac{f(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta}_{t+1})}{f_{Y|X}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}_{t+1})}\Big|\boldsymbol{x},\boldsymbol{\theta}_t\right), \end{aligned}$$

and by the definition of conditional density and (4.4)

$$l_{\boldsymbol{x}}(\boldsymbol{\theta}_{t+1}|\boldsymbol{x}) = \ln \frac{f(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta}_{t+1})}{f_{Y|X}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}_{t+1})}.$$

Thus

$$Q(\boldsymbol{\theta}_{t+1}|\boldsymbol{x},\boldsymbol{\theta}_t) - H(\boldsymbol{\theta}_{t+1}|\boldsymbol{x},\boldsymbol{\theta}_t) = E(l_{\boldsymbol{x}}(\boldsymbol{\theta}_{t+1}|\boldsymbol{x})|\boldsymbol{x},\boldsymbol{\theta}_t).$$

By the monotonicity of the conditional expected value, $l_{\boldsymbol{x}}(\boldsymbol{\theta}_{t+1}|\boldsymbol{x}) - l_{\boldsymbol{x}}(\boldsymbol{\theta}_t|\boldsymbol{x}) \geq 0$, provided that

$$(Q(\boldsymbol{\theta}_{t+1}|\boldsymbol{x},\boldsymbol{\theta}_t) - Q(\boldsymbol{\theta}_t|\boldsymbol{x},\boldsymbol{\theta}_t)) + (H(\boldsymbol{\theta}_t|\boldsymbol{x},\boldsymbol{\theta}_t) - H(\boldsymbol{\theta}_{t+1}|\boldsymbol{x},\boldsymbol{\theta}_t)) \geq 0.$$

Note that the first term $Q(\boldsymbol{\theta}_{t+1}|\boldsymbol{x},\boldsymbol{\theta}_t) - Q(\boldsymbol{\theta}_t|\boldsymbol{x},\boldsymbol{\theta}_t) \geq 0$, by (4.4). The second one, by (4.8) and the definition of the conditional expected value, can be written as

$$\begin{aligned} H(\boldsymbol{\theta}_t|\boldsymbol{x},\boldsymbol{\theta}_t) &- H(\boldsymbol{\theta}_{t+1}|\boldsymbol{x},\boldsymbol{\theta}_t) \\ &= E\left(\ln f_{Y|X}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}_t)|\boldsymbol{x},\boldsymbol{\theta}_t\right) - E\left(\ln f_{Y|X}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}_{t+1})|\boldsymbol{x},\boldsymbol{\theta}_t\right) \\ &= E\left(\ln \frac{f_{Y|X}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}_t)}{f_{Y|X}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}_{t+1})}\Big|\boldsymbol{x},\boldsymbol{\theta}_t\right) = \int \ln \frac{f_{Y|X}(y|\boldsymbol{x},\boldsymbol{\theta}_t)}{f_{Y|X}(y|\boldsymbol{x},\boldsymbol{\theta}_{t+1})} f_{Y|X}(y|\boldsymbol{x},\boldsymbol{\theta}_t)\, dy. \end{aligned}$$

As

$$\ln \frac{f_{Y|X}(y|\boldsymbol{x},\boldsymbol{\theta}_t)}{f_{Y|X}(y|\boldsymbol{x},\boldsymbol{\theta}_{t+1})} = -\ln\left(1 + \frac{f_{Y|X}(y|\boldsymbol{x},\boldsymbol{\theta}_{t+1}) - f_{Y|X}(y|\boldsymbol{x},\boldsymbol{\theta}_t)}{f_{Y|X}(y|\boldsymbol{x},\boldsymbol{\theta}_t)}\right)$$

and $\ln(1+t) < t$, for every $t > -1,\ t \neq 0$, the proof is completed. □

The EM algorithm can be stated as follows:

Start: Set the initial value $\boldsymbol{\theta}_0$.

Each step consists of two substeps. For every $t = 0, 1, 2, \ldots$ we proceed as follows:

Expectation step:

Let $\boldsymbol{\theta}_t$ be the current estimates of the unknown parameters $\boldsymbol{\theta}$. By (4.7), we calculate

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}_t) =& E\left(\ln f(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta})|\boldsymbol{x}, \boldsymbol{\theta}_t\right) = \int \ln f(\boldsymbol{x}, y|\boldsymbol{\theta}) f_{Y|X}(y|\boldsymbol{x}, \boldsymbol{\theta}_t)\, dy \\ &= \int \ln f(\boldsymbol{x}, y|\boldsymbol{\theta}) \frac{f(\boldsymbol{x}, y|\boldsymbol{\theta}_t)}{f_X(\boldsymbol{x}|\boldsymbol{\theta}_t)}\, dy. \end{aligned}$$

Maximization step:

Since $f_X(\boldsymbol{x}|\boldsymbol{\theta}_t)$ is independent of $\boldsymbol{\theta}$, we get

$$\boldsymbol{\theta}_{t+1} = \operatorname*{argmax}_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}_t) = \operatorname*{argmax}_{\boldsymbol{\theta}} \int \ln f(\boldsymbol{x}, y|\boldsymbol{\theta}) f(\boldsymbol{x}, y|\boldsymbol{\theta}_t)\, dy.$$

By iterating EM steps, the algorithm converges to the parameters that should be the maximum likelihood parameters, but the convergence speed is rather low (see e.g. **[16]**). There is some research to accelerate the convergence speed of the EM algorithm, but the procedure is usually not easy and difficult to carry out. In this paper we will focus on the traditional form of the EM algorithm and apply it to the estimation of a mixture of the normal densities.

4.2. *Fitting the mixture of the normal densities.* As previously, consider $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ – a sample of independent observations from a density $f_X(x|\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ denotes the parameters of the distribution and consider the following model:

Suppose that f_X is given by the mixture density

$$f_X(x|\boldsymbol{\theta}) = \sum_{j=1}^{m} \alpha_j f_j(x|\theta_j),$$

where

$$\alpha_j > 0, \quad j = 1, \ldots, m, \quad \sum_{j=1}^{m} \alpha_j = 1,$$

$$f_j(x|\theta_j) = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{1}{2}\left(\frac{x-\mu_j}{\sigma_j}\right)^2}, \quad \theta_j = (\mu_j, \sigma_j^2).$$

We call α_j the prior probability of the j-th mixture component, for $j = 1, \ldots, m$.

Suppose also, as in (4.1), that $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ denotes latent data from the distribution of random variable $Y = (Y_1, Y_2, \ldots, Y_n)$, such that $P(Y_i = j) = \alpha_j,\ j = 1, \ldots, m,\ i = 1, \ldots, n$.

Calculate $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_t)$ for the above model. The conditional density of Y, given $X = \boldsymbol{x}$ and $\boldsymbol{\theta}_t$, is given by

$$f_{Y|X}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}_t) = \prod_{i=1}^{n} f_{Y|X}(y_i|x_i, \theta_t), \tag{4.9}$$

and the joint density of X and Y

$$f(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i, y_i|\theta_{y_i}), \tag{4.10}$$

which by the definition of the conditional density gives

$$f(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n} \alpha_{y_i} f_{y_i}(x_i|\theta_{y_i}). \tag{4.11}$$

From formulae (4.9), (4.11) and (4.7), we obtain

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}_t) =& E(\ln f(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta})|\boldsymbol{x}, \boldsymbol{\theta}_t) = E(\ln (\prod_{i=1}^{n} \alpha_{y_i} f_{y_i}(x_i|\theta_{y_i}))|\boldsymbol{x}, \boldsymbol{\theta}_t) \\ =& \sum_{j=1}^{m} \sum_{i=1}^{n} \ln (\alpha_{y_i} f_{y_i}(x_i|\theta_{y_i})) P(Y_i = j|\boldsymbol{x}, \boldsymbol{\theta}_t) \\ =& \sum_{j=1}^{m} \sum_{i=1}^{n} \ln (\alpha_j f_j(x_i|\theta_j)) P(j|x_i, \boldsymbol{\theta}_t). \end{aligned} \tag{4.12}$$

Thus, in each iteration t of the EM algorithm, the expectation step, calculates $q_{ij}^t = P(j|x_i, \boldsymbol{\theta}_t)$. From now on, we will denote by $\widehat{\alpha}_j^t$, $\widehat{\mu}_j^t$ and $(\widehat{\sigma}_j^t)^2$ estimators of the parameters $\alpha_j,\ \mu_j,\ \sigma_j^2$ in iteration t.

LEMMA 4.5. *Under the above assumptions,*

$$\widehat{\alpha}_j^t = \frac{1}{n} \sum_{i=1}^{n} q_{ij}^t, \tag{4.13}$$

$$\widehat{\mu}_j^t = \frac{\sum_{i=1}^{n} q_{ij}^t x_i}{\sum_{i=1}^{n} q_{ij}^t}, \tag{4.14}$$

$$(\widehat{\sigma}_j^t)^2 = \frac{\sum_{i=1}^{n} (x_i - \widehat{\mu}_j^t)^2 q_{ij}^t}{\sum_{i=1}^{n} q_{ij}^t}. \tag{4.15}$$

PROOF. To find α_j, we use the Lagrange multiplier λ, with the constraint of $\sum_{j=1}^m \alpha_j = 1$. For $j = 1, \dots, m$, we calculate

$$\frac{\partial}{\partial \alpha_j}\Big(Q(\boldsymbol{\theta}, \boldsymbol{\theta}_t) + \lambda\Big(\sum_{j=1}^m \alpha_j - 1\Big)\Big) = 0.$$

By formula (4.12),

$$\frac{\partial}{\partial \alpha_j}\Big(\sum_{j=1}^m \sum_{i=1}^n \ln\left(\alpha_j f_j(x_i|\theta_j)\right) q_{ij}^t + \lambda\Big(\sum_{j=1}^m \alpha_j - 1\Big)\Big) = 0, \;\; j = 1, \dots, m,$$

and hence

$$\sum_{i=1}^n q_{ij}^t = -\lambda \alpha_j, \; j = 1, \dots, m.$$

Summing both sides over j, we get $\lambda = -n$, and thus (4.13).

To prove (4.14) and (4.15), we maximize the function

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}_t) = \sum_{j=1}^m \sum_{i=1}^n \ln\left(\alpha_j f_j(x_i|\theta_j)\right) q_{ij}^t.$$

By the assumptions of the model,

$$f_j(x|\theta_j) = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{1}{2}\left(\frac{x-\mu_j}{\sigma_j}\right)^2}, \quad \theta_j = (\mu_j, \sigma_j^2).$$

We get the function given by

$$\widetilde{Q}(\mu_j, \sigma_j) = \sum_{j=1}^m \sum_{i=1}^n \Big(\ln(\alpha_j) - \frac{1}{2}\ln(2\pi) - \ln(\sigma_j) - \frac{1}{2}\Big(\frac{x_i - \mu_j}{\sigma}\Big)^2\Big) q_{ij}^t. \tag{4.16}$$

By differentiating (4.16) with respect to μ and σ, for $j = 1, \dots, m$, we obtain

$$\begin{aligned}
\frac{\partial \widetilde{Q}}{\partial \mu_j} &= \sum_{i=1}^n (x_i - \mu_j) q_{ij}^t, \\
\frac{\partial \widetilde{Q}}{\partial \sigma_j} &= \sum_{i=1}^n \Big(\frac{1}{\sigma_j} + \frac{(x_i - \mu_j)^2}{\sigma_j^3}\Big) q_{ij}^t.
\end{aligned}$$

Solving the equations $\sum_{i=1}^n (x_i - \mu_j) q_{ij}^t = 0$ and $\sum_{i=1}^n \Big(\frac{1}{\sigma_j} + \frac{(x_i-\mu_j)^2}{\sigma_j^3}\Big) q_{ij}^t = 0$, we get (4.14) and (4.15). □

The scheme of the EM algorithm, for the mixture of normal densities can be stated as follows:

Start: We set α_j^0, μ_j^0 i $(\sigma_j^0)^2$.
For $j = 1, 2, \ldots, m$, we iterate $(t = 1, 2, \ldots)$:
E-step: We calculate q_{ij}^t,
M-step: We calculate

$$\widehat{\alpha}_j^t = \frac{1}{n}\sum_{i=1}^{n} q_{ij}^t,$$
$$\widehat{\mu}_j^t = \frac{\sum_{i=1}^n q_{ij}^t x_i}{\sum_{i=1}^n q_{ij}^t},$$
$$(\widehat{\sigma}_j^t)^2 = \frac{\sum_{i=1}^n (x_i - \widehat{\mu}_j^t)^2 q_{ij}^t}{\sum_{i=1}^n q_{ij}^t}.$$

The above scheme with the exact formulae can easily be used to construct the pilot function.

5. Example. In this section we present an example of how to apply the two-step method in practice. Suppose that we are given a sample of 200 independent observations from an absolutely continuous distribution with an unknown density f.

In this example we consider the mixture of two Weibull densities, with parameters: $\alpha_1 = 0.25$, $a_1 = 2.9$, $b_1 = 3.5$ and $\alpha_2 = 0.75$, $a_2 = 1.6$, $b_2 = 0.6$, which is nonsymmetric and bimodal.

At first we construct a pilot function as a mixture of $m = 2$ normal densities

$$f_0(x) = \alpha_1 f_1(x|\mu_1, \sigma_1^2) + \alpha_2 f_2(x|\mu_2, \sigma_2^2),$$

where

$$f_j(x|\mu_1, \sigma_1^2) = \frac{1}{\sqrt{2\pi}\sigma_j}\, e^{-\frac{1}{2}\frac{(x-\mu_j)^2}{\sigma_j^2}}, \quad j = 1, 2,$$

using the EM algorithm. We set the initial values of the estimates of the parameters:

$$\widehat{\alpha}_1^0 = 0.5, \quad \widehat{\alpha}_2^0 = 0.5, \quad \widehat{\mu}_1^0 = -0.5, \quad \widehat{\mu}_2^0 = 0.5, \quad (\widehat{\sigma}_1^0)^2 = 0.5, \quad (\widehat{\sigma}_2^0)^2 = 0.5,$$

and after 30 iterations of the EM algorithm we get

$$\widehat{\alpha}_1 = 0.315, \quad \widehat{\mu}_1 = 2.698, \quad (\widehat{\sigma}_1)^2 = 1.560,$$
$$\widehat{\alpha}_2 = 0.685, \quad \widehat{\mu}_2 = 0.489, \quad (\widehat{\sigma}_2)^2 = 0.079.$$

Therefore, the pilot is given by the formula

$$f_0(x) = \frac{\widehat{\alpha}_1}{\sqrt{2\pi(\widehat{\sigma}_1)^2}}\, e^{-\frac{1}{2}\frac{(x-\widehat{\mu}_1)^2}{(\widehat{\sigma}_1)^2}} + \frac{\widehat{\alpha}_2}{\sqrt{2\pi(\widehat{\sigma}_2)^2}}\, e^{-\frac{1}{2}\frac{(x-\widehat{\mu}_2)^2}{(\widehat{\sigma}_2)^2}}.$$

In the second step we use the generalized kernel density estimator (2.1) with the Gaussian kernel. In this case

$$\phi_1(x) = \frac{\frac{\widehat{\alpha}_1}{\sqrt{2\pi(\widehat{\sigma}_1)^2}}\, e^{-\frac{1}{2}\frac{(x-\widehat{\mu}_1)^2}{(\widehat{\sigma}_1)^2}}}{f_0(x)}, \qquad \phi_2(x) = \frac{\frac{\widehat{\alpha}_2}{\sqrt{2\pi(\widehat{\sigma}_2)^2}}\, e^{-\frac{1}{2}\frac{(x-\widehat{\mu}_2)^2}{(\widehat{\sigma}_2)^2}}}{f_0(x)}.$$

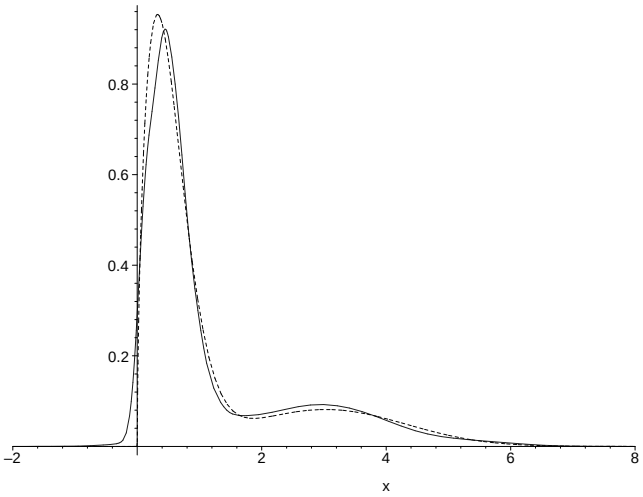
We obtain the bandwidths h_1 and h_2 minimizing the AMISE$(\widehat{f}_\phi)$ given by formula (3.3). By numerical minimization, we get $h_1 = 0.543$ and $h_2 = 0.110$.

Hence, we have the density estimator given by the formula

$$\widehat{f}_\phi(x) = \frac{1}{200}\sum_{i=1}^{200}\Big(\frac{\phi_1(x)}{h_1}K\Big(\frac{x-x_i}{h_1}\Big) + \frac{\phi_2(x)}{h_2}K\Big(\frac{x-x_i}{h_2}\Big)\Big),$$

where $x_1, \ldots, x_{200}$ denote the sample given, and $\phi_1(x)$, $\phi_2(x)$, h_1, h_2 are calculated above.

To show the results of the estimation, we present the graphs of the estimator $\widehat{f}_\phi$ and the estimated density in the picture below.



One can see that the estimator $\widehat{f}_\phi$ (solid line) is very close to the estimated density f (dotted line), which gives evidence to a high efficiency of the method proposed.

References

1. Abramson I.S., *Arbitrariness of the Pilot Estimator in Adapative Kernel Methods*, J. Multivariate Anal., **12** (1982), 562–567.
2. Dempster A.P., Laird N.M., Rubin D.B., *Maximum-likelihood from Incomplete Data via EM Algorithm*, J. Roy. Statist. Soc. Ser. B (1977), 39.
3. Devroye L., Gjörfi L., *Nonparametric Density Estimation: the L^1 view*, Willey, New York, 1985.
4. Domański Cz., Pruska K., *Nieklasyczne metody statystyczne*, PWE, Warszawa, 2000.
5. Hall P., Marron J.S., *Choice of Kernel Order in Density Estimation*, Ann. Statist., **16**, **1** (1987), 161–173.
6. Hjort N.L., Glad I.K., *Nonparametric Density Estimation with a Parametric Start*, Ann. Statis., **23, 3** (1995), 882–904.
7. Jarnicka J., *Dwukrokowa metoda estymacji gęstości*, preprint.
8. Marron J.S., Wand M.P., *Exact Mean Integrated Squared Error*, Ann. Statist., **20**, **2** (1992), 712–736.
9. Park B.U., Marron J.S., *Comparison of Data-Driven Bandwidth Selectors*, J. Amer. Statist. Assoc., **85** (1990), 66–72.
10. Parzen E., *On Estimation of a Probability Density Function and Mode*, Ann. Math. Statist., **33** (1962), 1065–1076.
11. Rosenblatt M., *Remarks on some Nonparametrics Estimates of a Density Function*, Ann. Math. Statist., **27** (1956), 827–837.
12. Sheater S.J., Jones M.C., *A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation*, J. Roy. Statist. Soc. Ser. B, **53**, **3** (1991), 683–690.
13. Silverman B.W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, New York, 1986.
14. Terrel G.R., Scott D.W., *Variable Kernel Density Estimation*, Ann. Statist., **20**, **3** (1992), 1236–1265.
15. Wand M.P., Marron J.S., Ruppert D., *Transformations in Density Estimation*, J. Amer. Statist. Assoc., **86**, **414** (1991), 343–353.
16. Wu C.F.J., *On the Convergence Properties of the EM Algorithm*, Ann. Statist., **11**, **1** (1983), 95–103.

Received November 2, 2005

e-mail: jj@gamma.im.uj.edu.pl