A semiparametric mixture model with application in genetics*

Magali SAN CRISTOBAL-GAUDY[1,2]

and

Irène GIJBELS[2]

[1] Laboratoire de Génétique Cellulaire, Institut National de la Recherche Agronomique, BP 28, 31326 CASTANET TOLOSAN Cédex, FRANCE

[2] Institut de Statistique, Université Catholique de Louvain, 34 Voie du Roman Pays, 1348 LOUVAIN-LA-NEUVE, BELGIQUE

Abbreviated title: Semiparametric mixture model

Acknowledgements

This research was conducted while the first author was a fellow of the Institut de Statistique, on postdoctoral leave from INRA. Thanks to Christian RITTER, Berwin TURLACH (UCL), Claude CHEVALET (INRA) for helpfull discussions on this topic.

*This research was supported partially by 'Programme d'Action de Recherche Concertée', N° 93-98 – 164.

A semiparametric mixture model with application in genetics*

Abstract

We consider a semiparametric approach to a mixture model of unknown densities. These probability density functions are assumed to be shape invariant. This study was motivated by a problem encountered in genetics. Estimators for the parametric and nonparametric components in the model are considered. The nonparametric component is estimated using a kernel type estimation procedure. It is shown that the estimator for the parametric component is uniformly consistent and is asymptotically normally distributed. Tests procedures, relying on the obtained results are proposed.

Abbreviated title: Semiparametric mixture model

MSC 1991 subject classification: Primary 62F03, 62G05, 62G10; secondary 62P10.

Keywords: asymptotic normality, consistency, kernel density estimators, mixture model, semiparametric model, shape invariance, testing hypothesis.

1 Introduction

The motivation for the developments discussed in this paper comes from a problem in genetics. The statistical theorems presented here can however be applied in several domains. We first give a brief overview of the biological problem motivating the present study. More details about the problem can be found in San Cristobal-Gaudy and Chevalet (1993).

A few words in genetics need to be introduced. Different versions of a *gene* are referred to as *alleles*. Each gene is located at a specific position on the *genome*. This position of the gene on the genome is called a *locus*. Hence, the genome can be roughly represented by a straight line on which genes are located. There are two main approaches to the genetical analysis of quantitative traits, such as growth rate of a chicken, technological yield of pig meat, dairy production of cows. In the first approach, a *major gene* is assumed to influence highly the trait of interest. If there exists two alleles Q and q of this gene, individuals carrying the Q allele are quite different from those carrying the q allele. It is then possible to see two modes on the histogram: a mixture of two populations is observed, corresponding to the Q and q carriers. An other approach is taken if no major gene could be identified. Then one uses the so-called *infinitesimal model* which assumes that an infinite number of genes have a very small influence on the quantitative trait. In this case, a Gaussian distribution is observed.

Models between the above two extreme models are worthwhile investigating. One can consider a finite number of genes having a non-negligible effect on the measured trait. The aim is to locate each such genes, at some QTL (*Quantitative Trait Locus*). But here, the histogram seems unimodal, although it should be a mixture, depending on the genotypes at the QTL's. *Genetic maps* are then a useful tool to bring these mixtures into the light (Elsen, 1993), if there are enough markers (in practice 10 to 20 for a chromosome). If an individual carries the A allele (resp. a) at the marker locus and the Q allele (resp. q) at the QTL, then the two alleles A and Q (resp. a and q) are mostly transmitted together to the progeny. On the contrary, if the two loci are independent, the A carriers are Q carriers or q carriers with, say, equal probability.

So, in practice one is faced with observations which can approximatively be viewed as a mixture of the alleles at the marker loci. The accuracy of the approximation is high when the QTL and the marker locus are very close on the chromosome. On the contrary, if these loci are genetically independent (i.e. far away from each other on the genome) no mixture is observed.

The observation of marker alleles depends highly on the studied species, and is very different for plants, animals and humans. Results obtained for one group can not straightforwardly be applied to another group, due to the differences in biology and genetics between the groups. Here we consider the PiGMaP design, which consists of two generations families of pigs, giving full and half sibs. For a full description of this particular design we refer the reader to San Cristobal-Gaudy and Chevalet (1993).

It is of common interest to test whether a mixture of two Gaussian random variables is underlying the observations, or only one Gaussian random variable. The maximum likelihood ratio test is usually applied, with known (Aitkin and Rubin, 1985; Goffinet *et al*, 1992) or unknown proportions (Ghosh and Sen, 1985; Pons, 1993) using nonstandard likelihood asymptotics. But this test is not robust against departures from normality. This brought up the idea of introducing a nonparametric component in the model. A whole nonparametric treatment is not desirable, because some parameters have a biological meaning. In the model we discuss here, only the Gaussian aspect is eliminated, leading to a mixture of two unknown densities with the same shape, differing only in location and scale. A similar development was given by Härdle and Marron (1990) in a regression setup.

The paper is organized as follows. In Section 2 we introduce the model and define the estimates for the parametric and nonparametric components. The consistency of the parametric estimate is established in Section 3, and its asymptotic normality is investigated in Section 4. This asymptotic normality result is then used to develop tests concerning respectively the parametric and the nonparametric component. Section 5 contains a discussion on this issue of hypothesis testing. The study presented here gives by no means a complete treatment of the motivating problem in genetics, but is only a first attempt to introduce such a semiparametric approach in this area. Further interesting open questions and directions for research are addressed in the last section.

2 Parametric comparison of nonparametric probability density functions

Let us consider $Y_i = (X_i, A_i)$, for $i = 1, \ldots, n$, independent identically distributed random variables from the population random variable $Y = (X, A)$. The random variable X is continuous with conditional density $f_{X|A=a}$:

$$f_{X|A=a}(x) = (1-a)g_1(x) + ag_2(x), \tag{2.1}$$

where $g_1(\cdot)$ and $g_2(\cdot)$ are unknown probability density functions. One can think of the random variable X as one of the quantitative traits (such as growth rate, dairy production etc.) mentioned in Section 1. The conditional event $\{A = a\}$ refers to the specific version a of the gene under consideration.

Let $\{v_0, \ldots, v_k\}$ be the finite set of values of A, with $0 = v_0 < v_1 < \ldots < v_k = 1$. The v_k's are probabilities determined by marker observation. The distribution of $\underline{A} = [I\{A = v_0\}, \ldots, I\{A = v_k\}]$ is multinomial:

$$\underline{A} \sim M(1; \{\alpha_0, \ldots, \alpha_k\}) \tag{2.2}$$

with parameters $\alpha_0, \ldots, \alpha_k$, which are assumed to be known.

To make the following developments coherent, suppose that α_0 and α_k are close to 0.5, i.e. that we know in most of the cases if a data point belongs to the $A = 0$ group or to the $A = 1$ group. This restriction can be justified by the biological knowledge of highly polymorphic markers. So, the marker genotypes and their exact transmission of genes are mainly known (see Section 1). In Section 6 we comment further on this assumption.

We assume further that the densities g_1 and g_2 have the same shape:

$$g_1(x) = \theta_{3,0} + \theta_{4,0}\, g_2(\theta_{1,0} + \theta_{2,0}x) \tag{2.3}$$

where $\theta_0 = (\theta_{1,0}, \theta_{2,0}, \theta_{3,0}, \theta_{4,0})$ is the true unknown value of the parameter $\theta = (\theta_1, \theta_2, \theta_3, \theta_4) \in \Theta \subseteq \mathbb{R}^4$. One of the tasks consists of estimating θ_0 from the observations $(X_1, A_1), (X_2, A_2), \cdots, (X_n, A_n)$. We will assume that $\theta_2 \neq 0$. The parameter space Θ is subject to some natural restrictions, which are as follows. Since we are dealing with probability density functions, we find that

- $\theta_3 = 0$, since $\lim_{x\to\pm\infty} g_j(x) = 0,\ j = 1,2$
- Suppose that the density functions g_1 and g_2 have support $I\!R$. Then

$$\int_{I\!R} g_1(x)\ dx = (\theta_4/|\theta_2|) \int_{I\!R} g_2(y)\ dy = 1.$$

which implies that

$$\theta_4 = |\theta_2|.$$

The same restriction holds through when g_1 and g_2 are compactly supported. For the sake of simplicity we restrict to those two cases.

As a consequence the parametric link between g_1 and g_2 is given by

$$g_1(x) = |\theta_{2,0}|\, g_2(\theta_{1,0} + \theta_{2,0}x) = G(x, \theta_0), \tag{2.4}$$

for $x \in \mathrm{supp}(g_1)$ and $\theta_{1,0} + \theta_{2,0}x \in \mathrm{supp}(g_2)$. Here we introduced the notation

$$G(x,\theta) = |\theta_2|\, g_2(\theta_1 + \theta_2 x)$$

for $\theta \in \Theta$. The parameters θ_1 and θ_2 are related to the means and variances of random variables, say B_1 and B_2, having density functions g_1 and g_2 respectively. Indeed, denoting by μ_j and σ_j^2 the means and variances of B_j, $j = 1, 2$, it is easy to see that

$$\begin{cases} |\theta_2| = \dfrac{\sigma_2}{\sigma_1} \\ \theta_1 = \mu_2 - \mu_1\, \theta_2. \end{cases} \tag{2.5}$$

This provides an easy interpretation of the parameters θ_1 and θ_2: if θ_2 is positive (respectively negative), it is a scale parameter which is simply the ratio of the standard deviations (respectively its opposite), while θ_1 is a shift parameter equal to the differences (respectively the sum) of the means on a common scale between the two random variables B_1 and B_2.

The observations Y_i for which $A_i = 0$ (respectively $A_i = 1$) will be used to estimate the density function g_1 (respectively g_2) with the following kernel type estimator:

$$\hat{g}_1(x) = \frac{1}{n_0 h(n_0)} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h(n_0)}\right) I\{A_i = 0\} = \frac{\sum_{i=1}^{n} K_h(x - X_i) I\{A_i = 0\}}{\sum_{i=1}^{n} I\{A_i = 0\}} \tag{2.6}$$

$$\hat{g}_2(x) = \frac{1}{n_k h(n_k)} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h(n_k)}\right) I\{A_i = 1\} = \frac{\sum_{i=1}^{n} K_h(x - X_i) I\{A_i = 1\}}{\sum_{i=1}^{n} I\{A_i = 1\}} \tag{2.7}$$

where

$$n_0 = \sum_{i=1}^{n} I\{A_i = 0\}, \qquad n_k = \sum_{i=1}^{n} I\{A_i = 1\}, \qquad K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$$

and where $h(n_0)$ and $h(n_k)$ denote the bandwidths and K is a nonnegative weight function, the so-called kernel.

In view of (2.4) a natural loss function is

$$L(\theta) = \int [g_1(x) - |\theta_2|\; g_2(\theta_1 + \theta_2 x)]^2 w(x)\; dx, \tag{2.8}$$

where $w(.)$ is a nonnegative weight function. This loss function can be estimated straightforwardly by replacing the density functions g_1 and g_2 by their kernel estimators $\hat{g}_1$ and $\hat{g}_2$ defined in (2.6) and (2.7), leading to

$$\hat{L}(\theta) = \int [\hat{g}_1(x) - |\theta_2|\; \hat{g}_2(\theta_1 + \theta_2 x)]^2 w(x)\; dx$$

and an estimator for θ is provided by

$$\hat{\theta} = \underset{\theta \in \Theta}{\text{Argmin}} \quad \hat{L}(\theta). \tag{2.9}$$

This paper has as objective to establish the asymptotic properties of the estimator $\hat{\theta}$. Those will then be used to develop statistical tests. The developments are similar to those established for the case of regression curves in Härdle and Marron (1990).

3 Consistency of the parameter estimate

We first study the consistency of the estimator $\hat{\theta}$. The uniform consistency of $\hat{\theta}$ is ensured under some conditions which read as follows:

- The loss function $L(\theta)$ is locally convex near θ_0:

$$\forall \varepsilon > 0, \exists D(\varepsilon) > 0, |\theta - \theta_0| > \varepsilon \;\Rightarrow\; L(\theta) - L(\theta_0) > D(\varepsilon)$$

- The bandwidths h and h', used to estimate $g_1(\cdot)$, respectively $g_2(\cdot)$, depend only on n:

$$h(n_0) = h(n) = h \quad \text{and} \quad h'(n_k) = h'(n) = h'$$

  and belong moreover to the interval $B_n = [n^{-1+\delta}, n^{-\delta}]$, where $0 < \delta < 1$.

- K is a bounded, compactly supported probability density function, having bounded variation. Further we assume that K is Hölder continuous:

$$\exists \alpha, \beta > 0, \forall u, v, |K(u) - K(v)| \leq \alpha |u - v|^{\beta}, \quad 0 < \beta \leq 1.$$

- The density functions g_1 and g_2 are uniformly continuous and satisfy

$$\int g_1^2(x)\; w(x)\; dx < \infty \quad \text{and} \quad \sup_{\theta \in \Theta} \int G^2(x, \theta)\; w(x)\; dx < \infty.$$

THEOREM 1. *Under the above assumptions, $\hat\theta$ is consistent for θ_0, uniformly over $h, h' \in B_n$, in the sense that*

$$\sup_{h,h'\in B_n} |\hat\theta - \theta_0| \stackrel{a.s.}{\longrightarrow} 0\,.$$

Proof of Theorem 1.
Denote $\sup\limits_h = \sup\limits_{h,h'\in B_n}$ and $\sup\limits_\theta = \sup\limits_{\theta\in\Theta}$. Recall that

$$X_n \stackrel{\text{a.s.}}{\longrightarrow} X \qquad \Longleftrightarrow \qquad \forall\varepsilon > 0 : \lim_{n\to\infty} P\left[\bigcup_{m=n}^{\infty}\{|X_m - X| > \varepsilon\}\right] = 0.$$

Note that, for any $\varepsilon > 0$ given,

$$\begin{aligned}
\left\{\sup_h |\hat\theta - \theta_0| > \varepsilon\right\} &\subseteq \left\{\sup_h (L(\hat\theta) - L(\theta_0)) > D(\varepsilon)\right\}\\
&\subseteq \left\{\sup_h (L(\hat\theta) - \hat L(\hat\theta) + \hat L(\theta_0) - L(\theta_0)) > D(\epsilon)\right\}\\
&\subseteq \left\{\sup_\theta \sup_h |\hat L(\theta) - L(\theta)| > \frac{D(\varepsilon)}{2}\right\},
\end{aligned}$$

since $L(\cdot)$ is locally convex near θ_0 (first inclusion) and $\hat\theta$ minimises $\hat L(\cdot)$ (second inclusion). We obtain that

$$0 \le \lim_{n\to\infty} P\left\{\bigcup_{m=n}^{\infty}\{\sup_h |\hat\theta - \theta_0| > \varepsilon\}\right\} \le \lim_{n\to\infty} P\left\{\bigcup_{m=n}^{\infty}\{\sup_\theta\sup_h |\hat L(\theta) - L(\theta)| > \frac{D(\varepsilon)}{2}\}\right\}$$

and hence it suffices to show that

$$\sup_\theta\sup_h |\hat L(\theta) - L(\theta)| \stackrel{\text{a.s.}}{\longrightarrow} 0 \tag{3.1}$$

Define $\hat G(x;\theta) = |\theta_2|\hat g_2(\theta_1 + \theta_2 x)$. Then, by rearranging terms, adding and subtracting $2\,\hat g_1(x)\; G(x,\theta)$ and by applying the triangle inequality, we find that

$$\begin{aligned}
|\hat L(\theta) - L(\theta)| &= \left|\int (\hat g_1(x) - \hat G(x;\theta))^2 w(x)\; dx - \int (g_1(x) - G(x;\theta))^2 w(x)\; dx\right|\\
&\le \int \Big[|(\hat g_1(x) - g_1(x))(\hat g_1(x) + g_1(x))| + 2|\hat g_1(x)(G(x,\theta) - \hat G(x;\theta))|\\
&\quad +2|G(x,\theta)(g_1(x) - \hat g_1(x))| + |(\hat G(x,\theta) - G(x;\theta))(\hat G(x,\theta) + G(x,\theta))|\Big]\; w(x)\; dx.
\end{aligned}$$

Applying the Cauchy-Schwartz inequality, it is clear that it suffices to show that

$$\int g_1^2(x)\; w(x)\; dx < \infty, \tag{3.2}$$

$$\sup_h \int \hat{g}_1^2(x)\ w(x)\ dx < \infty \qquad \text{a.e.}, \tag{3.3}$$

$$\sup_\theta \int G^2(x,\theta)\ w(x)\ dx < \infty, \tag{3.4}$$

$$\sup_\theta \sup_h \int \hat{G}^2(x,\theta)\ w(x)\ dx < \infty \qquad \text{a.e.}, \tag{3.5}$$

$$\sup_h \int (\hat{g}_1(x) - g_1(x))^2\ w(x)\ dx \to 0 \quad \text{a.s.}, \tag{3.6}$$

$$\sup_\theta \sup_h \int (\hat{G}(x,\theta) - G(x,\theta))^2\ w(x)\ dx \to 0 \quad \text{a.s.} \tag{3.7}$$

In the sequel we comment on each of these. The statements (3.2) and (3.4) are part of the conditions of the theorem. Statement (3.3) is an immediate consequence of the uniform almost sure convergence of $\hat{g}_1$ to g_1 which holds under assumption (3.2). Similarly, under assumption (3.4), the uniform almost sure convergence of $\hat{g}_2$ to g_2 leads to (3.5).

In order to show (3.6) we first of all remark that, for $B_n' \subseteq B_n$,

$$\begin{aligned}
&\left\{\sup_h \int (\hat{g}_1(x) - g_1(x))^2 w(x)\ dx > \varepsilon\right\} \\
&\subseteq \left\{\sup_{h_1 \in B_n'} \int (\hat{g}_{1,h_1}(x) - g_1(x))^2 w(x)\ dx > \frac{\varepsilon}{2}\right\} \\
&\quad \bigcup \left\{\sup_{h \in B_n} \inf_{h_1 \in B_n'} \left| \int (\hat{g}_{1,h}(x) - g_1(x))^2 w(x)\ dx - \int (\hat{g}_{1,h_1}(x) - g_1(x))^2 w(x)\ dx \right| > \frac{\varepsilon}{2}\right\},
\end{aligned}$$

where we used the extra subscripts to avoid confusion. For the second term on the right-hand side of the above inequality we apply a factorisation argument followed by the Cauchy-Schwartz inequality, and find that we basically should look at

$$\hat{g}_{1,h}(x) - \hat{g}_{1,h_1}(x) = \frac{1}{n_0} \sum_{i=1}^n \left[\frac{1}{h} K\left(\frac{x - X_i}{h}\right) - \frac{1}{h_1} K\left(\frac{x - X_i}{h_1}\right)\right] I\{A_i = 0\},$$

which indicates already that Hölder continuity of K has to be assumed. Choosing B_n' appropriately, and so that $\#(B_n') \leq n^\zeta$ for some $\zeta > 0$, we can deal with this second term. So, in order to obtain (3.6) it remains to show that

$$\sup_{h \in B_n'} \mathrm{ISE}(h) \xrightarrow{\text{a.s.}} 0,$$

with

$$\mathrm{ISE}(h) = \int (\hat{g}_1(x) - g_1(x))^2 w(x)\ dx.$$

The estimator $\hat{g}_1$ can be considered as a fractional delta sequence estimator as defined in Marron and Härdle (1986). Remark that by the strong law of large numbers

$$\frac{1}{n} \sum_{i=1}^n I\{A_i = 0\} \xrightarrow{\text{a.s.}} \alpha_0,$$

and that this convergence does not depend on x or h. Therefore, we can apply the corollary of Theorem 3 in Marron and Härdle (1986, p. 100), which tells us that

$$\lim_{n\to\infty} \sup_{h\in B_n'} \left| \frac{\mathrm{ISE}(h) - \mathrm{MISE}^*(h)}{\mathrm{MISE}^*(h)} \right| = 0 \quad \text{a.e.} \tag{3.8}$$

with

$$\mathrm{MISE}^*(h) = E \int \left[\frac{1}{n} \sum_{i=1}^{n} g_1^*(x, Y_i) \right]^2 w(x)\, dx$$

where

$$g_1^*(x, Y_i) = \frac{1}{\alpha_0} \left[K_h(x - X_i) - g_1(x) \right] I\{A_i = 0\}.$$

Now,

$$0 \le \sup_{h\in B_n'} \mathrm{ISE}(h) \le \sup_{h\in B_n'} \left| \frac{\mathrm{ISE}(h) - \mathrm{MISE}^*(h)}{\mathrm{MISE}^*(h)} \right| . \sup_{h\in B_n'} \mathrm{MISE}^*(h) + \sup_{h\in B_n'} \mathrm{MISE}^*(h)$$

and hence it remains to show that

$$\sup_{h\in B_n'} \mathrm{MISE}^*(h) \longrightarrow 0. \tag{3.9}$$

We have that

$$\mathrm{MISE}^*(h) = \int \frac{1}{n^2 \alpha_0^2}\, E(M_h^*(x))\, w(x)\, dx,$$

with

$$\begin{aligned} E(M_h^*(x)) \quad &= \quad \int\!\!\int \sum_{i=1}^{n} \sum_{j=1}^{n} \left[K_h(x - u_i) - g_1(x) \right] \left[K_h(x - u_j) - g_1(x) \right] \\ &\qquad \times I\{t_i = 0\}\, I\{t_j = 0\}\, dF_i dF_j \end{aligned}$$

where we used the notation

$$dF_i = dF_{X_i, A_i}(u_i, t_i) = \sum_a P\{A_i = a\}\, f_{X_i | A_i = a_i}(u_i)\, du_i\, d\delta_a(t_i)$$

and denoted by δ_a the Dirac mass in $a \in \{v_0, \ldots, v_k\}$, $P\{A_i = a\} = \alpha_a$, $\sum_a \alpha_a = 1$. Since the random variables $Y_1, \ldots, Y_n$ are independent and identically distributed we obtain that

$$\begin{aligned} E(M_h^*(x)) \quad &= \quad n\alpha_0 \int [K_h(x-u) - g_1(x)]^2 g_1(u)\, du \\ &\qquad + n(n-1)\alpha_0^2 \left(\int [K_h(x-u) - g_1(x)] g_1(u)\, du \right)^2 \\ &= \quad nh\alpha_0 \int \left[\frac{1}{h} K(v) - g_1(x) \right]^2 g_1(x - hv)\, dv \\ &\qquad + n(n-1)\alpha_0^2 \left(\int K(v)\, g_1(x - hv)\, dv - g_1(x) \right)^2 \end{aligned}$$

Applying the assumptions on h, K, and g_1 we find (3.9), which finishes the proof of statement (3.6).

Using similar arguments we show that (3.7) holds. This completes the proof of Theorem 1.

□

4 Asymptotic normality of the parametric estimate

In this section we establish the asymptotic normality result for the estimator $\hat{\theta}$. We will need the following assumptions.

- The bandwidths are such that $h' = |\theta_2| h$.
- The Hessian matrix
$$H(\theta_0) = \left[\frac{\partial^2}{\partial\theta_j \partial\theta_{j'}} L(\theta) \Big|_{\theta=\theta_0} \right]_{j,j'=1,2}$$
is positive definite.
- The function $g_1(x)$ is three times uniformly continuous differentiable, with
$$\int g_1^2(x) w(x)\, dx < \infty, \qquad \int g_1'^2(x) w(x)\, dx < \infty, \qquad \int x^2 g_1'^2(x) w(x)\, dx < \infty,$$
$$\int g_1''^2(x) w(x)\, dx < \infty, \qquad \int x^4 g_1''^2(x) w(x)\, dx < \infty, \qquad \text{and} \qquad \int g_1'''^2(x) w(x)\, dx < \infty.$$
- K is a symmetric probability density function with compact support, which is zero at the endpoints of its support,
$$\int K'^2(u)\, du < \infty,$$
and K has a Hölder continuous second derivative.
- The weight function w is two times continuously differentiable.
- The bandwidth h may be selected in a data-driven way resulting into $\hat{h}$. However, the selection procedure has to satisfy the following condition: $\hat{h} = h_0 + O_p(n^{-3/10})$, with h_0 proportional to $n^{-1/5}$. This ensures $P\{\hat{h} \in B_n^*\} \to 1$, with $B_n^* = [h_0 - n^{\alpha-3/10}, h_0 + n^{\alpha-3/10}]$ and $\alpha \in (0, 1/10)$.

The asymptotic normality result is stated in the next theorem. The following additional notations are needed:
$$G_j(x,\theta) = \frac{\partial}{\partial\theta_j} G(x,\theta), \qquad \hat{G}_j(x,\theta) = \frac{\partial}{\partial\theta_j} \hat{G}(x,\theta),$$
$$G_{jj'}(x,\theta) = \frac{\partial^2}{\partial\theta_j \partial\theta_{j'}} G(x,\theta), \qquad \hat{G}_{jj'}(x,\theta) = \frac{\partial^2}{\partial\theta_j \partial\theta_{j'}} \hat{G}(x,\theta), \qquad j, j' = 1, 2,$$
$$\hat{L}_j(\theta) = \frac{\partial}{\partial\theta_j} \hat{L}(\theta) \quad j = 1, 2, \qquad \nabla \hat{L}(\theta) = \begin{pmatrix} \hat{L}_1(\theta) \\ \hat{L}_2(\theta) \end{pmatrix}, \qquad \text{and} \qquad \hat{H}(\theta) = \left[\frac{\partial^2}{\partial\theta_j \partial\theta_{j'}} \hat{L}(\theta) \right]_{j,j'=1,2}.$$

THEOREM 2. *Under the above assumptions,*
$$n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{D} N\left(0, H^{-1}(\theta_0)\, \Sigma\, H^{-1}(\theta_0)\right),$$

where

$$\Sigma = 4 \begin{pmatrix} \left(\frac{1}{\alpha_0}+\frac{1}{\alpha_k}\right)\beta_{11} - 2\beta_1^2 & \left(\frac{1}{\alpha_0}+\frac{1}{\alpha_k}\right)\beta_{12} - 2\beta_1\beta_2 \\ \left(\frac{1}{\alpha_0}+\frac{1}{\alpha_k}\right)\beta_{12} - 2\beta_1\beta_2 & \left(\frac{1}{\alpha_0}+\frac{1}{\alpha_k}\right)\beta_{22} - 2\beta_2^2 \end{pmatrix}, \tag{4.1}$$

with

$$\beta_{\ell\ell'} = \int g_1(x) G_\ell(x,\theta_0) G_{\ell'}(x,\theta_0) w^2(x)\, dx \qquad \ell,\ell' = 1,2. \tag{4.2}$$

and

$$\beta_\ell = \int g_1(x) G_\ell(x,\theta_0) w(x)\, dx \qquad \ell = 1,2\,. \tag{4.3}$$

Proof of Theorem 2.
Recall the definition of the estimator $\hat\theta$ given in (2.9). We have that

$$0 = \nabla \hat L(\hat\theta) = \nabla \hat L(\theta_0) + \hat H(\hat\xi_n)(\hat\theta - \theta_0),$$

with $\hat\xi_n$ between $\hat\theta$ and θ_0, which leads to

$$n^{1/2}(\hat\theta - \theta_0) = -\hat H^{-1}(\hat\xi_n) n^{1/2} \nabla \hat L(\theta_0),$$

and hence the result in Theorem 2 is an immediate consequence of the next two lemmas.

LEMMA 1.

$$n^{1/2}\nabla \hat L(\theta_0) \ \xrightarrow{\text{D}}\ \text{N}(0,\Sigma).$$

LEMMA 2.

$$\hat H(\hat\xi_n) \ \xrightarrow{\text{P}}\ H(\theta_0).$$

For proving Lemma 1 we will use the decomposition

$$\begin{aligned} \hat L_j(\theta_0) &= -2\int [\hat g_1(x) - \hat G(x,\theta_0)]\, \hat G_j(x,\theta_0)\, w(x)\, dx \\ &= -2Z_j(h_0) - 2[Z_j(h) - Z_j(h_0)] \\ &\quad -2\int [\hat g_1(x) - \hat G(x,\theta_0)]\, [G_j(x,\theta_0) - \hat G_j(x,\theta_0)]\, w(x)\, dx \\ &\quad -2\int [E\hat g_1(x) - E\hat G(x,\theta_0)]\, G_j(x,\theta_0)\, w(x)\, dx, \end{aligned}$$

with

$$Z_j(h) = \int [\hat g_{1,h}(x) - E\hat g_{1,h}(x) + E\hat G_h(x,\theta_0) - \hat G_h(x,\theta_0)]\, G_j(x,\theta_0)\, w(x)\, dx,$$

where again we introduce the subscript h to avoid confusion. From the above decomposition it is clear that Lemma 1 will follow from the results stated in the lemmas below.

LEMMA 1.1.

$$\sup_h \left| \int [\hat g_1(x) - \hat G(x,\theta_0)]\, [G_j(x,\theta_0) - \hat G_j(x,\theta_0)]\, w(x)\, dx \right| = o_P(n^{-1/2}).$$

LEMMA 1.2.
$$\sup_h \left| \int [E\hat{g}_1(x) - E\hat{G}(x,\theta_0)]\, G_j(x,\theta_0)\; w(x)\; dx \right| = o(n^{-1/2}).$$

LEMMA 1.3.
$$\sup_h |Z_j(h) - Z_j(h_0)| = o_P(n^{-1/2}) \qquad j = 1,2.$$
and

LEMMA 1.4.
$$2n^{1/2} Z(h_0) \xrightarrow{\text{D}} \text{N}(0,\Sigma), \quad \text{with } Z(h_0) = \begin{pmatrix} Z_1(h_0) \\ Z_2(h_0) \end{pmatrix}.$$

We will now give the proofs of Lemmas 1.1 — 1.4 and Lemma 2.

Proof of Lemma 1.1
Since $g_1(x) = G(x,\theta_0)$ (see (2.4)) we get, by applying the Cauchy-Schwarz inequality, that
$$\begin{aligned}
&\left| \int [\hat{g}_1(x) - \hat{G}(x,\theta_0)][G_j(x,\theta_0) - \hat{G}_j(x,\theta_0)] w(x)\; dx \right| \\
&\le \left[\int (\hat{g}_1(x) - g_1(x))^2 w(x)\; dx \right]^{1/2} \left[\int (G_j(x,\theta_0) - \hat{G}_j(x,\theta_0))^2 w(x)\; dx \right]^{1/2} \\
&+ \left[\int (G(x,\theta_0) - \hat{G}(x,\theta_0))^2 w(x)\; dx \right]^{1/2} \left[\int (G_j(x,\theta_0) - \hat{G}_j(x,\theta_0))^2 w(x)\; dx \right]^{1/2}.
\end{aligned}$$
Therefore it suffices to show that
$$\sup_h \int (\hat{g}_1(x) - g_1(x))^2 w(x)\; dx = o_P(n^{-7/10}) \tag{4.4}$$
$$\sup_h \int (\hat{G}(x,\theta_0) - G(x,\theta_0))^2 w(x)\; dx = o_P(n^{-7/10}) \tag{4.5}$$
$$\sup_h \int (\hat{G}_j(x,\theta_0) - G_j(x,\theta_0))^2 w(x)\; dx = o_P(n^{-3/10}). \tag{4.6}$$
Similarly as in the proof of Theorem 1 we will apply the corollary of Theorem 3 in Marron and Härdle (1986). Since
$$\text{MISE}^*(h) = O\left(\frac{1}{nh}\right) + O\left(\frac{1}{n}\right) + O\left(\frac{h}{n}\right) + O(h^4)$$
and h is proportional to $n^{-1/5}$, one obtains that $\text{MISE}^* = O(n^{-4/5})$, which shows (4.4). In a similar way we obtain the result in (4.5).

The same kind of reasoning applies to $\hat{G}_j - G_j$ leading now to
$$\text{MISE}^*(h) = O\left(\frac{1}{nh^3}\right) + O\left(\frac{1}{nh}\right) + O\left(\frac{h}{n}\right) + O(h^4)$$

which implies that $\text{MISE}^* = O(n^{-2/5})$, and this proves (4.6). $\square$

Proof of Lemma 1.2
Note first of all that it is easy to verify that

$$E_{X,A}\,\hat{g}_1(x) = c_0(n,\alpha_0)\int K(u)\,g_1(x-hu)\,du \tag{4.7}$$

and

$$E_{X,A}\,\hat{G}(x,\theta_0) = c_0(n,\alpha_k)\int K(u)\,g_1\left(x - \text{sign}(\theta_{2,0})hu\right)\,du, \tag{4.8}$$

where

$$c_0(n,\alpha) = 1-(1-\alpha)^n.$$

Note further that

$$G_1(x,\theta_0) = \frac{1}{\theta_{2,0}}g_1'(x),$$

and

$$G_2(x,\theta_0) = \frac{1}{\theta_{2,0}}[g_1(x) + x\,g_1'(x)].$$

The above expressions enable us to write

$$\begin{aligned}
&\sup_h\left|\int[E\hat{g}_1(x) - E\hat{G}(x,\theta_0)]\,G_j(x,\theta_0)\,w(x)\,dx\right| \\
&\quad= |(1-\alpha_k)^n - (1-\alpha_0)^n| \\
&\quad\quad\times\sup_h\left|\int\int G_j(x,\theta_0)\,K(v)\,g_1(x-hv)\,w(x)\,dx\,dv\right|\{1+o(1)\}.
\end{aligned}$$

The first factor appearing on the right-hand side of the above equality is of order $o(n^{-1/2})$, while the second factor is finite. This concludes the proof of the lemma. $\square$

Proof of Lemma 1.3
Rewriting $Z_j(h)$ as

$$\begin{aligned}
Z_j(h) &= \int(\hat{g}_{1,h}(x) - \hat{G}_h(x,\theta_0))\,G_j(x,\theta_0)\,w(x)\,dx \\
&\quad - \int(E\hat{g}_{1,h}(x) - E\hat{G}_h(x,\theta_0))\,G_j(x,\theta_0)\,w(x)\,dx,
\end{aligned}$$

and by using expressions (4.7) and (4.8) we obtain

$$\begin{aligned}
Z_j(h) - Z_j(h_0) &= \frac{n}{n_0}\frac{1}{n}\sum_{i=1}^n \widetilde{X}_i(j) - \frac{n}{n_k}\frac{1}{n}\sum_{i=1}^n \widetilde{\widetilde{X}}_i(j) \\
&\quad -[c_0(n,\alpha_0) - c_0(n,\alpha_k)] \\
&\quad\quad \times\int\int K(v)\,G_j(x,\theta_0)\,w(x)\,[g_1(x-hv) - g_1(x-h_0v)]\,dx\,dv\{1+o(1)\}
\end{aligned} \tag{4.9}$$

with

$$
\begin{aligned}
\widetilde{X}_i(j) &= \int [K_h(x-X_i) - K_{h_0}(x-X_i)]\ G_j(x,\theta_0)\ I\{A_i=0\} w(x)\ dx, \\
\widetilde{\widetilde{X}}_i(j) &= \int |\theta_{2,0}| \left[K_{|\theta_{2,0}|h}(\theta_{1,0}+\theta_{2,0}x - X_i) - K_{|\theta_{2,0}|h_0}(\theta_{1,0}+\theta_{2,0}x - X_i)\right] \\
&\qquad \times G_j(x,\theta_0)\ I\{A_i=1\}\ w(x)\ dx.
\end{aligned}
$$

It is important to remark that the random variables $\widetilde{X}_1(j), \widetilde{X}_2(j), \cdots, \widetilde{X}_n(j)$ are independent and identically distributed. A similar remark can be made for the random variables $\widetilde{\widetilde{X}}_1(j), \widetilde{\widetilde{X}}_2(j), \cdots,$ $\widetilde{\widetilde{X}}_n(j)$. Moreover the two sets of random variables are mutually independent. The third term in expression (4.9) is negligible since it can be bounded by

$$
|(1-\alpha_0)^n - (1-\alpha_k)^n|\ |h-h_0| \left| \int\int K(v) G_j(x,\theta_0) w(x) g_1'(x)\ dv\ dx + O(h_0) \right| = o(n^{-1/2}),
$$

using a Taylor expansion. So, the main contribution to (4.9) comes from the sums. We will now investigate the mean and variance of the random variables $\widetilde{X}_i(j)$ and $\widetilde{\widetilde{X}}_i(j)$ appearing in those sums. The mean of $\widetilde{X}_i(j)$ is

$$
E\widetilde{X}_1(j) = \alpha_0 \int\int [K_h(x-u) - K_{h_0}(x-u)]\, G_j(x,\theta_0) w(x)\ dx\ g_1(u) du.
$$

This can be evaluated further by considering a Taylor expansion up to remainder term of order 2 of the function $K_h(x-u)$ near $h = h_0$, as well as a Taylor expansion of the function $g_1(x - h_0 v)$ near $h_0 = 0$ up to a remainder of order 3. This then reveals that

$$
E\widetilde{X}_1(j) = \alpha_0 h_0 (h-h_0) \int v^2 K(v) dv \int G_j(x,\theta_0) g_1''(x) w(x) dx + o\left(h_0(h-h_0)\right).
$$

Furthermore, it is easy to see that

$$
\frac{1}{\alpha_0} E\widetilde{X}_1(j) = \frac{1}{\alpha_k} E\widetilde{\widetilde{X}}_1(j).
$$

Similar considerations yield that

$$
E\widetilde{X}_1^2(j) = \alpha_0 h_0^2 (h-h_0)^2 \left(\int v^2 K(v) dv\right)^2 \int (G_j w)''(u) g_1(u) du + o\left(h_0^2(h-h_0)^2\right),
$$

and a similar kind of expression for $E\widetilde{\widetilde{X}}_1^2(j)$. Note further that the weak law of large number provides that

$$
\frac{n_0}{n} \overset{\mathrm{P}}{\longrightarrow} \alpha_0, \qquad \text{and} \qquad \frac{n_k}{n} \overset{\mathrm{P}}{\longrightarrow} \alpha_k. \tag{4.10}
$$

These facts together with an application of the central limit theorem leads to

$$
\frac{n}{n_0} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{X}_i(j) - \frac{n}{n_k} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{\widetilde{X}}_i(j) = O_P\left(h_0^2 (h-h_0)^2\right) = o_P(1),
$$

if $h_0^2(h-h_0)^2 \to 0$, which is indeed the case. $\square$

Proof of Lemma 1.4
We start with writing

$$Z_j(h_0) = Z_{j,1}(h_0) - Z_{j,2}(h_0) \qquad j = 1,2$$

where

$$Z_{j,1}(h) = \int [\hat{g}_{1,h}(x) - E\hat{g}_{1,h}(x)]\, G_j(x,\theta_0)\ w(x)\ dx$$

and

$$Z_{j,2}(h) = \int [\hat{G}_h(x,\theta_0) - E\hat{G}_h(x,\theta_0)]\, G_j(x,\theta_0)\ w(x)\ dx.$$

Recalling expressions (4.7) and (4.8) it is easy to see that

$$\begin{aligned}
Z_j(h) &= Z_{j,1}(h) - Z_{j,2}(h) \\
&= \frac{1}{n_0}\sum_{i=1}^{n} X_i^{(0)}(j) - \frac{1}{n_k}\sum_{i=1}^{n} X_i^{(1)}(j) \\
&\quad -c_0(n,\alpha_0)\int\int K(u)g_1(x-hu)G_j(x,\theta_0)w(x)\ dxdu \\
&\quad +c_0(n,\alpha_k)\int\int K(u)g_1\left(x-\text{sign}(\theta_{2,0})hu\right)G_j(x,\theta_0)w(x)\ dxdu,
\end{aligned} \tag{4.11}$$

where we introduced the random variables

$$X_i^{(0)}(j) = \int K_h(x-X_i)\ G_j(x,\theta_0)I\{A_i=0\}\ w(x)\ dx$$

and

$$X_i^{(1)}(j) = |\theta_{2,0}|\int K_{|\theta_{2,0}|h}(\theta_{1,0}+\theta_{2,0}x-X_i)\ G_j(x,\theta_0)\ I\{A_i=1\}w(x)\ dx,$$

$i = 1,\ldots,n$. Note that the random variables $X_i^{(0)}(j)$ and $X_i^{(1)}(j)$ are independent. The sums in (4.11) involving these random variables are the main terms in that expression and hence determine its behavior. We now investigate the characteristics of those random variables, denoted by

$$\mu_j^{(\ell)} = E\left(X_1^{(\ell)}(j)\right) \qquad \sigma_j^{(\ell)2} = \text{Var}\left(X_1^{(\ell)}(j)\right), \qquad \ell = 0,1$$

and also

$$\sigma_{12}^{(\ell)} = \text{Cov}\left(X_1^{(\ell)}(1), X_1^{(\ell)}(2)\right) \qquad \ell = 0,1.$$

Straightforward calculations yield

$$\begin{aligned}
\mu_j^{(0)} &= \alpha_0\int\int K(u)g_1(x-hu)G_j(x,\theta_0)w(x)dxdu \\
&= \alpha_0\int g_1(x)G_j(x,\theta_0)w(x)dx + o(1) \\
&= \alpha_0\beta_j + o(1),
\end{aligned}$$

and similarly

$$\mu_j^{(1)} = \alpha_k\beta_j + o(1),$$

recalling the definitions given in (4.2) and (4.3). Furthermore,

$$\sigma_j^{(0)2} = \alpha_0 \beta_{jj} - \alpha_0^2 \beta_j^2 + o(1),$$

and also

$$\sigma_j^{(1)2} = \alpha_k \beta_{jj} - \alpha_k^2 \beta_j^2 + o(1).$$

For the covariance structures we find that

$$\sigma_{12}^{(0)} = \alpha_0 \beta_{12} - {\alpha_0}^2 \beta_1 \beta_2 + o(1)$$

and

$$\sigma_{12}^{(1)} = \alpha_k \beta_{12} - {\alpha_k}^2 \beta_1 \beta_2 + o(1).$$

Applying the weak convergence results in (4.10) we find from expression (4.11) that

$$2n^{1/2} Z_j(h_0) \quad = \quad 2\frac{n}{n_0}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(X_i^{(0)}(j) - \mu_j^{(0)}\right) - 2\frac{n}{n_k}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(X_i^{(1)}(j) - \mu_j^{(1)}\right) + o_P(1).$$

Concerning the main terms in the above expression we have that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} X_i^{(0)}(1) - \mu_1^{(0)} \\ X_i^{(0)}(2) - \mu_2^{(0)} \end{pmatrix} \xrightarrow{\mathrm{D}} \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \alpha_0 \begin{pmatrix} \beta_{11} - \alpha_0 \beta_1^2 & \beta_{12} - \alpha_0 \beta_1 \beta_2 \\ \beta_{12} - \alpha_0 \beta_1 \beta_2 & \beta_{22} - \alpha_0 \beta_2^2 \end{pmatrix}\right)$$

and similarly

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} X_i^{(1)}(1) - \mu_1^{(1)} \\ X_i^{(1)}(2) - \mu_2^{(1)} \end{pmatrix} \xrightarrow{\mathrm{D}} \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \alpha_k \begin{pmatrix} \beta_{11} - \alpha_k \beta_1^2 & \beta_{12} - \alpha_k \beta_1 \beta_2 \\ \beta_{12} - \alpha_k \beta_1 \beta_2 & \beta_{22} - \alpha_k \beta_2^2 \end{pmatrix}\right)$$

where we applied Theorem B p.28 in Serfling (1980). The result now follows from (4.10) and the above asymptotic normality results. □

Proof of Lemma 2.
The matrix $\hat{H}(\hat{\xi}_n)$ contains the elements

$$\hat{L}_{jj'}(\hat{\xi}_n) = 2\int \left[\hat{G}_j(x,\hat{\xi}_n)\,\hat{G}_{j'}(x,\hat{\xi}_n) - \left(\hat{g}_1(x) - \hat{G}(x,\hat{\xi}_n)\right)\,\hat{G}_{jj'}(x,\hat{\xi}_n)\right]\,w(x)\,dx$$

whereas the matrix $H(\theta_0)$ consists of the quantities

$$L_{jj'}(\theta_0) = 2\int \left[G_j(x,\theta_0)\,G_{j'}(x,\theta_0) - (g_1(x) - G(x,\theta_0))\,G_{jj'}(x,\theta_0)\right]\,w(x)\,dx.$$

In order to show that

$$\hat{L}_{jj'}(\hat{\xi}_n) - L_{jj'}(\theta_0) \xrightarrow{\mathrm{P}} 0$$

we consider the following decompositions

$$\begin{aligned}
&\int \left[\hat{G}_j(x,\hat{\xi}_n)\,\hat{G}_{j'}(x,\hat{\xi}_n) - G_j(x,\theta_0)\,G_{j'}(x,\theta_0)\right] w(x)dx \\
&= \int \left[\hat{G}_j(x,\hat{\xi}_n) - G_j(x,\hat{\xi}_n)\right] \left[\hat{G}_{j'}(x,\hat{\xi}_n) - G_{j'}(x,\hat{\xi}_n)\right] w(x)dx \\
&\quad + \int \left[\hat{G}_j(x,\hat{\xi}_n) - G_j(x,\hat{\xi}_n)\right] \left[G_{j'}(x,\hat{\xi}_n) - G_{j'}(x,\theta_0)\right] w(x)dx \\
&\quad + \int \left[\hat{G}_j(x,\hat{\xi}_n) - G_j(x,\hat{\xi}_n)\right] G_{j'}(x,\theta_0) w(x)dx \\
&\quad + \int \left[G_j(x,\hat{\xi}_n) - G_j(x,\theta_0)\right] \left[\hat{G}_{j'}(x,\hat{\xi}_n) - G_{j'}(x,\hat{\xi}_n)\right] w(x)dx \\
&\quad + \int \left[G_j(x,\hat{\xi}_n) - G_j(x,\theta_0)\right] \left[G_{j'}(x,\hat{\xi}_n) - G_{j'}(x,\theta_0)\right] w(x)dx \\
&\quad + \int \left[G_j(x,\hat{\xi}_n) - G_j(x,\theta_0)\right] G_{j'}(x,\theta_0) w(x)dx \\
&\quad + \int G_j(x,\theta_0) \left[\hat{G}_{j'}(x,\hat{\xi}_n) - G_{j'}(x,\hat{\xi}_n)\right] w(x)dx \\
&\quad + \int G_j(x,\theta_0) \left[G_{j'}(x,\hat{\xi}_n) - G_{j'}(x,\theta_0)\right] w(x)dx,
\end{aligned} \tag{4.12}$$

and

$$\begin{aligned}
&\int \left\{\left[\hat{g}_1(x) - \hat{G}(x,\hat{\xi}_n)\right]\ \hat{G}_{jj'}(x,\hat{\xi}_n) - \left[g_1(x) - G(x,\theta_0)\right]\ G_{jj'}(x,\theta_0)\right\} w(x)dx \\
&= \int \left[\hat{g}_1(x) - g_1(x)\right] \left[\hat{G}_{jj'}(x,\hat{\xi}_n) - G_{jj'}(x,\hat{\xi}_n)\right] w(x)dx \\
&\quad + \int \left[\hat{g}_1(x) - g_1(x)\right] \left[G_{jj'}(x,\hat{\xi}_n) - G_{jj'}(x,\theta_0)\right] w(x)dx \\
&\quad + \int \left[\hat{g}_1(x) - g_1(x)\right] G_{jj'}(\theta_0) w(x)dx \\
&\quad - \int \left[\hat{G}(x,\hat{\xi}_n) - G(x,\hat{\xi}_n)\right] \left[\hat{G}_{jj'}(x,\hat{\xi}_n) - G_{jj'}(x,\hat{\xi}_n)\right] w(x)dx \\
&\quad - \int \left[\hat{G}(x,\hat{\xi}_n) - G(x,\hat{\xi}_n)\right] \left[G_{jj'}(x,\hat{\xi}_n) - G_{jj'}(x,\theta_0)\right] w(x)dx \\
&\quad - \int \left[\hat{G}(x,\hat{\xi}_n) - G(x,\hat{\xi}_n)\right] G_{jj'}(x,\theta_0) w(x)dx \\
&\quad - \int \left[G(x,\hat{\xi}_n) - G(x,\theta_0)\right] \left[\hat{G}_{jj'}(x,\hat{\xi}_n) - G_{jj'}(x,\hat{\xi}_n)\right] w(x)dx \\
&\quad - \int \left[G(x,\hat{\xi}_n) - G(x,\theta_0)\right] \left[G_{jj'}(x,\hat{\xi}_n) - G_{jj'}(x,\theta_0)\right] w(x)dx \\
&\quad - \int \left[G(x,\hat{\xi}_n) - G(x,\theta_0)\right] G_{jj'}(x,\theta_0) w(x)dx.
\end{aligned} \tag{4.13}$$

We apply the Cauchy-Schwartz inequality to (4.12) and see that we need to show that, for all j,

$$\int \left[G_j(x,\hat{\xi}_n) - G_j(x,\theta_0)\right]^2 w(x)dx \xrightarrow{\text{P}} 0, \tag{4.14}$$

$$\sup_h \int \left[\hat{G}_j(x,\hat{\xi}_n) - G_j(x,\hat{\xi}_n)\right]^2 w(x)dx \xrightarrow{\text{P}} 0, \tag{4.15}$$

and

$$\int G_j^2(x,\theta_0)w(x)dx < \infty. \tag{4.16}$$

The last statement follows immediately from the integral assumptions on g_1 and its derivatives given before the formulation of Theorem 2. Relation (4.14) is a direct consequence of the uniform continuity of the functions involved. Statement (4.15) is proved using similar arguments as those used in the proof of Lemma 1.1. We omit the details here.

In order to deal with (4.13) we end up with proving that, for all j and j',

$$\sup_h \int \left[\hat{g}_1(x) - g_1(x)\right]^2 w(x)dx \xrightarrow{\mathrm{P}} 0, \tag{4.17}$$

$$\sup_h \int \left[\hat{G}(x,\hat{\xi}_n) - G(x,\hat{\xi}_n)\right]^2 w(x)dx \xrightarrow{\mathrm{P}} 0, \tag{4.18}$$

$$\sup_h \int \left[\hat{G}_{jj'}(x,\hat{\xi}_n) - G_{jj'}(x,\hat{\xi}_n)\right]^2 w(x)dx < \infty \qquad \text{a.e.}, \tag{4.19}$$

$$\int \left[G_{jj'}(x,\hat{\xi}_n) - G_{jj'}(x,\theta_0)\right]^2 w(x)dx < \infty \qquad \text{a.e.}, \tag{4.20}$$

and

$$\int \left[G_{jj'}(x,\theta_0)\right]^2 w(x)dx < \infty. \tag{4.21}$$

Statements (4.17) and (4.18) are immediate consequences of statements (4.4) and (4.5) which were established in the proof of Lemma 1.1. Using similar arguments as in that proof one can show (4.19). Relation (4.20) follows from the uniform continuity of the functions involved. Finally, (4.21) is obtained immediately from the assumptions on the integrals in terms of g_1 and its derivatives. □

5 Testing hypothesis

Two types of hypothesis tests are of our interest: one concerns the parametric component of the model whereas the other one focuses on the nonparametric component.

We first look at a null hypothesis of the form

$$H_0 : M(\theta_0 - \theta^*) = 0,$$

where M is a given $r \times 2$ matrix of rank $r \in \{1,2\}$. Such a kind of hypothesis is important when investigating the availability of QTL detection. In view of Theorem 2 it can be suggested to reject H_0 when $(\hat{\theta}-\theta^*)M^T\hat{\Sigma}^{*-1}M(\hat{\theta}-\theta^*)$ is larger than the 95th percentile of the χ_r^2 distribution, where $\hat{\Sigma}^*$ is a consistent estimate of $\Sigma^* = MH(\theta_0)^{-1}\Sigma H(\theta_0)^{-1}M^T$.

An important issue in QTL detection is to study the particular null hypothesis

$$H_0 : \theta_{1,0} = 0 \quad \text{and} \quad \theta_{2,0} = 1,$$

where M was taken to be the 2×2 identity matrix and $\theta_1^* = 0$ and $\theta_2^* = 1$. Here the null hypothesis states that the densities g_1 and g_2 are identical and hence that there is no mixture.

In other words, the QTL and the marker locus are independent (see Section 1). Hence, no QTL is located in a neighborhood of the chosen marker.

The second type of hypothesis is more concerned with the choice of the shape invariant model: do g_1 and g_2 have the same shape? A reasonable approach is to take a test statistic which is based on the minimisation criterion $\hat{L}(\hat{\theta})$. The asymptotic distribution of this quantity is established in the following theorem.

THEOREM 3. *Under the assumptions of Theorem 2 we have that*

$$n^{1/2}\left(\hat{L}(\hat{\theta}) - n^{-1}h_0^{-1}C_\mu\right) \xrightarrow{D} N(0, 4C_\sigma^2)$$

where

$$C_\mu = \left(\frac{1}{\alpha_0} + \frac{1}{\alpha_k}\right) \int g_1(x)\, w(x)\, dx \int K^2(v)\, dv \tag{5.1}$$

and

$$C_\sigma^2 = 2\left(\frac{1}{\alpha_0} + \frac{1}{\alpha_k}\right) \int g_1^3(x)\, w^2(x)\, dx - 6\left(\int g_1^2(x)\, w(x)\, dx\right)^2. \tag{5.2}$$

provided that the values α_0 and α_k are such that they satisfy $\alpha_0, \alpha_k < 1$ and $4\alpha_0\alpha_k < \alpha_0 + \alpha_k$.

Based on the result established in the above theorem we can propose the following test statistic for testing the correctness of the shape invariant model. Reject the null hypothesis if

$$\hat{L}(\hat{\theta}) > \frac{1}{n\hat{h}}\hat{C}_\mu + \frac{1}{\sqrt{n}} 2\hat{C}_\sigma z_{1-\alpha},$$

where $z_{1-\alpha}$ denotes the $(1-\alpha)^{th}$ quantile of the standard normal distribution function, and where $\hat{C}_\mu$ and $\hat{C}_\sigma$ are consistent estimators of C_μ and C_σ respectively.

We now give the proof of the asymptotic normality result in Theorem 3.

Proof of Theorem 3.
First of all note that by Taylor expansion

$$\hat{L}(\theta_0) = \hat{L}(\hat{\theta}) + (\theta_0 - \hat{\theta})\nabla\hat{L}(\hat{\theta}) + \frac{1}{2}(\theta_0 - \hat{\theta})^T \hat{H}(\hat{\xi}^n)(\theta_0 - \hat{\theta}),$$

where $\hat{\xi}^n$ is between θ_0 and $\hat{\theta}$. Since $\nabla\hat{L}(\hat{\theta}) = 0$ this leads to

$$n^{1/2}\left(\hat{L}(\hat{\theta}) - n^{-1}h_0^{-1}C_\mu\right) = n^{1/2}\left(\hat{L}(\theta_0) - n^{-1}h_0^{-1}C_\mu\right) - n^{1/2}\frac{1}{2}(\theta_0 - \hat{\theta})^T \hat{H}(\hat{\xi}^n)(\theta_0 - \hat{\theta})$$

and therefore it is enough to show that

$$(\theta_0 - \hat{\theta})^T \hat{H}(\hat{\xi}^n)(\theta_0 - \hat{\theta}) = o_P(n^{-1/2}) \tag{5.3}$$

and

$$n^{1/2}\left(\hat{L}(\theta_0) - n^{-1}h_0^{-1}C_\mu\right) \xrightarrow{\mathrm{D}} \mathrm{N}(0, 4C_\sigma^2). \tag{5.4}$$

The result in (5.3) follows immediately from applying Theorem 2, since this theorem provides that

$$\hat{H}(\hat{\xi}^n) \xrightarrow{\mathrm{P}} H(\theta_0) > 0$$

and

$$n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{\mathrm{D}} \mathrm{N}(0, H^{-1}(\theta_0)\Sigma H^{-1}(\theta_0)).$$

In order to prove (5.4) we start with decomposing $\hat{L}(\theta_0)$ as follows:

$$\begin{aligned} \hat{L}(\theta_0) &= \int \left[\hat{g}_1(x) - \hat{G}(x,\theta_0)\right]^2 w(x)dx \\ &= \int \hat{g}_1^2(x)w(x)dx + \int \hat{G}^2(x,\theta_0)w(x)dx - 2\int \hat{g}_1(x)\hat{G}(x,\theta_0)w(x)dx, \end{aligned} \tag{5.5}$$

and we will deal with each of these terms separately.

It is easy to see that the first integral in (5.5) equals

$$\int \hat{g}_1^2(x)w(x)dx = \frac{1}{n}\left(\frac{n}{n_0}\right)^2 \frac{1}{n}\sum_{i=1}^n X_i^{(0)} + \left(\frac{n}{n_0}\right)^2 \frac{n(n-1)}{n^2} U_n^{(0)}, \tag{5.6}$$

where we introduced the random variables

$$X_i^{(0)} = \int K_h^2(x - X_i)\, I\{A_i = 0\} w(x)dx,$$

and the notation

$$U_n^{(0)} = \frac{2}{n(n-1)} \sum_{1\le i<j\le n} h^{(0)}(Y_i, Y_j),$$

with

$$h^{(0)}(y_1,y_2) = \int K_h(x-x_1)I\{a_1=0\}K_h(x-x_2)I\{a_2=0\}w(x)dx,$$

$$y_j = (x_j, a_j) \qquad j = 1,2.$$

We now study the terms appearing on the right-hand side of (5.6). The random variables $X_1^{(0)}, X_2^{(0)}, \cdots, X_n^{(0)}$ are independently identically distributed with mean

$$\mu^{(0)} = \alpha_0 \frac{1}{h}\int\int K^2(v)g_1(x-hv)w(x)dxdv \equiv \alpha_0\tilde{\mu}(h),$$

and variance

$$\sigma^{(0)2} = \alpha_0 \int \left[\frac{1}{h}\int K^2(x)w(u+hx)dx\right]^2 g_1(u)\,du - \left(\mu^{(0)}\right)^2.$$

Moreover the central limit theorem provides an asymptotic normality result for the properly normalized sum of random variables $X_1^{(0)}, X_2^{(0)}, \cdots, X_n^{(0)}$, namely

$$\sqrt{n}\left(\frac{\frac{1}{n}\sum_i X_i^{(0)} - \mu^{(0)}}{\sigma^{(0)}}\right) \xrightarrow{\mathrm{D}} \mathrm{N}(0,1). \tag{5.7}$$

What can we say about the behavior of the second term on the right-hand side of (5.6)? The statistic $U_n^{(0)}$ appearing in (5.6) is a U-statistic of degree 2 with kernel $h^{(0)}(\cdot,\cdot)$. We partially adopt the notations of Serfling (1980). The U-statistic $U_n^{(0)}$ is an unbiased estimator of the parameter

$$\theta^{(0)} = \alpha_0^2 \widetilde{\theta}(h,h),$$

where

$$\widetilde{\theta}(h_1,h_2) = \int \left\{ \int \int K(u)K(v) g_1(x-h_1u) g_1(x-h_2v) du dv \right\} w(x) dx. \tag{5.8}$$

The kernel $h^{(0)}(\cdot,\cdot)$ satisfies the condition $Eh^{(0)2}(Y_1,Y_2) < \infty$. Indeed, this is easily seen using the assumptions on the functions K, g_1, and g_2. Moreover we have that

$$h_1^{(0)}(y_1) = Eh^{(0)}(y_1,Y_2) = \alpha_0 \int \int K(x) I\{a_1=0\} K(v) g_1(x_1+hx-hv) w(x_1+hx) dx dv.$$

Remark also that, effectively, $Eh_1^{(0)}(Y_1) = \theta^{(0)}$.

Furthermore,

$$Eh_1^{(0)2}(Y_1) = \alpha_0^3 \widetilde{M}(h,h),$$

where

$$\begin{aligned}\widetilde{M}(h_1,h_2) & \\ &= \int \left\{ \int \int K(v)K(u) g_1(z+uh_1-vh_2) w(z+uh_1) du dv \right\}^2 g_1(z) dz. \end{aligned} \tag{5.9}$$

We denote

$$\zeta_1^{(0)} = Eh_1^{(0)2}(Y_1) - \theta^{(0)2}. \tag{5.10}$$

By applying the Cauchy-Schwarz inequality it is easy to see, under the assumption on α_0 stated in the theorem, that $\zeta_1^{(0)} > 0$. Then application of Theorem A p. 192 in Serfling (1980) provides the following asymptotic normality result for the U-statistics $U_n^{(0)}$

$$\sqrt{n}\left(\frac{U_n^{(0)} - \theta^{(0)}}{2\sqrt{\zeta_1^{(0)}}} \right) \xrightarrow{\text{D}} \text{N}(0,1). \tag{5.11}$$

This ends the discussion on the first integral in (5.5).

The second integral in (5.5) can be dealt with in a similar way. This integral can be rewritten as follows:

$$\int \hat{G}_1^2(x,\theta_0) w(x) dx = \frac{1}{n}\left(\frac{n}{n_k}\right)^2 \frac{1}{n}\sum_{i=1}^n X_i^{(1)} + \left(\frac{n}{n_k}\right)^2 \frac{n(n-1)}{n^2} U_n^{(1)}, \tag{5.12}$$

where now

$$X_i^{(1)} = \theta_{2,0}^2 \int K^2_{|\theta_{2,0}|h}(\theta_{1,0} + \theta_{2,0}x - X_i)\, I\{A_i = 1\} w(x) dx,$$

and

$$U_n^{(1)} = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} h^{(1)}(Y_i,Y_j).$$

Here $U_n^{(1)}$ denotes a U-statistic of degree 2 with kernel

$$
\begin{aligned}
&h^{(1)}(y_1, y_2) \\
&= \int |\theta_{2,0}| K_{|\theta_{2,0}|h}(\theta_{1,0} + \theta_{2,0}x - x_1) I\{a_1 = 1\} |\theta_{2,0}| K_{|\theta_{2,0}|h}(\theta_{1,0} + \theta_{2,0}x - x_2) I\{a_2 = 1\} w(x) dx.
\end{aligned}
$$

Using similar arguments as before it is shown that

$$
\sqrt{n} \left(\frac{\frac{1}{n}\sum_i X_i^{(1)} - \mu^{(1)}}{\sigma^{(1)}} \right) \xrightarrow{\text{D}} \text{N}(0,1), \tag{5.13}
$$

where

$$
\mu^{(1)} = \alpha_k \widetilde{\mu}(\gamma_0 h), \qquad \gamma_0 \equiv \text{sign}(\theta_{2,0}),
$$

$$
\sigma^{(1)2} = \alpha_k \int \left[\frac{1}{h} \int K^2(x) w(u + \gamma_0 h x) dx \right]^2 g_1(u)\, du - \left(\mu^{(1)}\right)^2.
$$

For the U-statistic $U_n^{(1)}$ we obtain

$$
\sqrt{n} \left(\frac{U_n^{(1)} - \theta^{(1)}}{2\sqrt{\zeta_1^{(1)}}} \right) \xrightarrow{\text{D}} \text{N}(0,1), \tag{5.14}
$$

where we used the notations

$$
\theta^{(1)} = \alpha_k^2 \widetilde{\theta}(\gamma_0 h, \gamma_0 h),
$$

$$
\zeta_1^{(1)} = E h_1^{(1)2}(Y_1) - \theta^{(1)2},
$$

and

$$
E h_1^{(1)2}(Y_1) = \alpha_k^3 \widetilde{M}(\gamma_0 h, \gamma_0 h).
$$

The third integral in (5.5) equals

$$
\int \hat{g}_1(x) \hat{G}(x, \theta_0) w(x) dx = \frac{n(n-1)}{2 n_0 n_k} U_n, \tag{5.15}
$$

with

$$
U_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} h(Y_i, Y_j),
$$

a U-statistic of degree 2 with kernel

$$
\begin{aligned}
h(y_1, y_2) \quad = \quad & \int \Big[K_h(x - x_1) I\{a_1 = 0\} |\theta_{2,0}| K_{|\theta_{2,0}|h}(\theta_{1,0} + \theta_{2,0}x - x_2) I\{a_2 = 1\} \\
& + K_h(x - x_2) I\{a_2 = 0\} |\theta_{2,0}| K_{|\theta_{2,0}|h}(\theta_{1,0} + \theta_{2,0}x - x_1) I\{a_1 = 1\} \Big] \, w(x) dx.
\end{aligned}
$$

For this U-statistic we have the following asymptotic normality result

$$
\sqrt{n} \left(\frac{U_n - \theta^{(U)}}{2\sqrt{\zeta_1}} \right) \xrightarrow{\text{D}} \text{N}(0,1), \tag{5.16}
$$

where

$$\theta^{(U)} = 2\alpha_0\alpha_k\widetilde{\theta}(h, \gamma_0 h),$$
$$\zeta_1 = Eh_1(Y_1) - \theta^{(U)2},$$

and

$$Eh_1^2(Y_1) = (\alpha_0\alpha_k^2 + \alpha_0^2\alpha_k)\widetilde{M}(h, \gamma_0 h).$$

This result holds through under the conditions of the theorem, which enable us to apply the appropriate theorems available for U-statistics.

Combining decompositions (5.5), (5.6), (5.12) and (5.15) it is easy to see that

$$\begin{aligned}
n^{1/2}\left(\hat{L}(\theta_0) - n^{-1}h_0^{-1}C_\mu\right) \quad = \quad & \sqrt{n}\left\{\frac{1}{n}\left(\frac{n}{n_0}\right)^2\frac{1}{n}\sum_{i=1}^{n}X_i^{(0)} - \frac{1}{n}\frac{1}{\alpha_0^2}\mu^{(0)}\right\} \\
& +\sqrt{n}\left\{\frac{1}{n}\left(\frac{n}{n_k}\right)^2\frac{1}{n}\sum_{i=1}^{n}X_i^{(1)} - \frac{1}{n}\frac{1}{\alpha_k^2}\mu^{(1)}\right\} \\
& +\sqrt{n}\left\{\left(\frac{n}{n_0}\right)^2\frac{n(n-1)}{n^2}U_n^{(0)} - \frac{1}{\alpha_0^2}\theta^{(0)}\right\} \\
& +\sqrt{n}\left\{\left(\frac{n}{n_k}\right)^2\frac{n(n-1)}{n^2}U_n^{(1)} - \frac{1}{\alpha_k^2}\theta^{(1)}\right\} \\
& -\sqrt{n}\left\{\frac{n(n-1)}{n_0 n_k}U_n - \frac{1}{\alpha_0\alpha_k}\theta^{(U)}\right\} + o_P\left(\frac{1}{\sqrt{n}h_0}\right).
\end{aligned}$$

Finally it can be checked that

$$\mathrm{Cov}(U_n^{(0)}, U_n^{(1)}) = 0, \qquad \mathrm{Cov}(U_n^{(\ell)}, U_n) = 0 \qquad \ell = 0, 1.$$

and that

$$\frac{4}{\alpha_0^4}\zeta_1^{(0)} + \frac{4}{\alpha_1^4}\zeta_1^{(1)} + \frac{4}{\alpha_0^2\alpha_k^2}\zeta_1 = 4C_\sigma^2 + O(h_0).$$

which completes the proof. □

6 Discussion

The presented semiparametric approach to the problem of mixtures seems to be quite appropriate from the biometrical point of view. Theorem 2 means that the rate of convergence of the estimator $\hat{\theta}$ for the true parameter θ_0 of interest is the same here as for the full parametric approach. But the gain is the robustness against departure from normality.

Keeping up with the point of view that the semiparametric approach is satisfactory, one could think of replacing the kernel estimator applied here by another type of density estimator (see for example Silverman (1986)).

In our estimation procedure, described in Section 2, we only use the data points of which we know that they are belonging to the groups $A = 0$ or $A = 1$. The remaining data belong to a mixture with known proportions:

$$(1-a)\,g_1(x) + a\,g_2(x), \quad a \in \{v_1, \ldots, v_{k-1}\}.$$

In the pig breeding scheme considered here this type of data represents approximately 10% of the population. An improvement of the estimation method using all the available data is one of the subjects for further research.

Some straightforward extensions can be envisaged. The mixture of two densities is easily extended to a mixture of M densities:

$$f_{X|A_1=a_1,\ldots,A_M=a_M}(x) = a_1\, g_1(x) + \ldots + a_M\, g_M(x)$$

with $\sum a_i = 1$ and where $g_1, \ldots, g_M$ are M unknown densities with similar shape. For reasons of estimability, a reference cell type constraint (similar to the one in this paper) is considered:

$$g_1(x) = |\theta_{i2}|\, g_i(\theta_{i1} + \theta_{i2}x), \quad i = 2, \ldots, M.$$

Note that this corresponds to $\theta_{11} = 0$ and $\theta_{12} = 1$. The estimation procedure in this situation of a mixture of M densities, and its corresponding theorems are similar. A null mean like constraint (see Kneip and Engel (1992)) is equivalent to:

$$\sum_i \theta_{i1} = 0 \qquad \text{and} \qquad \frac{1}{M}\sum_i \theta_{i2} = 1.$$

At a first glance the following parameterisation may be preferred (the two component mixture is considered for sake of simplicity):

$$\delta = \mu_2 - \mu_1 \text{ and } \theta = \sigma_2/\sigma_1$$

so that

$$g_1(x) = \theta\, g_2(\theta x + \delta - \mu_1(\theta - 1))$$

since the null hypothesis $H_0 : \mu_1 = \mu_2$, without any restriction on the scale parameter, is directly written as $H_0 : \delta = 0$. This is in contrast with $H_0 : \theta_1 + \mu_1(\theta_2 - 1) = 0$ with the parameterisation chosen in this article. Nevertheless, the nuisance parameter μ_1 has to be estimated in both cases, so that it does not matter in practice which parameterisation is used.

We should mention that in reality, for the genetical problem, the data are not independent. In fact, they are correlated, due to full sib or half sib relationships. But the covariance matrix of $\{X_i|A_i = a_i\}$ has a known form. Further work is needed to incorporate this information. A similar remark can be made concerning the independence assumption of $A_1, \ldots, A_n$.

For pig breeders, the aim is to detect QTL, i.e. to decide whether $\delta = 0$ if the QTL has no mean effect, or $\theta_2 = 1$ if the QTL has no scale effect. The results presented here concern a particular locus l. The question also arises how to detect the position l_0 of a QTL. In case of several QTL's the test statistics T_l and $T_{l'}$ will be correlated, with a higher correlation if the loci l and l' are close on the genome (i.e. genetically linked). One method to deal with this problem is to adopt the same kind of graphical procedure as in Lander and Botstein (1989): locate a QTL at the maximum of the $\{T_l\}_{l\in L}$ line, and give a standard error of this estimation by choosing an appropriate threshold based on stochastic processes.

In this paper we established some asymptotic results for the estimators. Simulation studies are needed to shed light on the finite sample behaviour of the estimators and the test statistics. The proposed semiparametric approach should be compared with the parametric approach based on the Gaussian hypothesis, using maximum likelihood, maximum penalized likelihood (Leroux, 1992), Bayesian inference (Robert, 1994), and other semiparametric models (see e.g. Roeder, 1992).

It is well-known that the choice of a bandwidth is crucial in kernel estimation. An interesting overview of the practical performance of popular data-driven bandwidth selectors is provided in Park and Turlach (1992).

REFERENCES

Aitkin M. and Rubin D.B.(1985). Estimation and hypothesis testing in finite mixture models. *J. Roy. Statist. Assoc.* B **47**, 67-75.

Elsen (1993). *Detection and use of marker genes in farm animals.* International ERASMUS Course, Liège, Belgique, 12-16 July 1993.

Ghosh J.K. and Sen P.K. (1985). On the asymtotic performance of the log likelihood ratio statistic for the mixture model and related results. *In* Proceedings of the Berkeley conference in honor of J. Neyman and J. Kiefer, vol 2, Le Cam and Olshen eds.

Goffinet B., Loisel P. and Laurent B. (1992). Testing mixture models when the proportions are known. *Biometrika* **79**, 842-846.

Härdle W. and Marron J.S. (1990). Semiparametric comparison of regression curves. *Ann. Statist.* **18**, 63-89.

Kneip A. and Engel J. (1992). Model estimation in nonlinear regression under shape invariance. *Sonderforschungsbereich 303, Information und die Koordination wirtschatlisher Aktivitäten, Discussion Paper A-367, Rheinische Friedrich Wilhelm Universität Bonn, Germany.*

Lander E.S. and Botstein D. (1989). Mapping Mendelian factors undelying quantitative traits using RFLP linkage maps. *Genetics* **121**, 743-756.

Leroux B.G. (1992). Consistent estimation of a mixing distribution. *Ann. Statist.* **20**, 1350-1360.

Marron J.S. and Härdle W. (1986). Random approximations to some measures of accuracy in nonparametric curve estimation. *J. Multivariate Anal.* **20**, 91-113.

Park B.U. and Turlach B.A. (1992). Practical performance of several data driven bandwith selectors. *Comput. Statist.* **7**, 251-270.

Pons O. (1993). Homogeneity test in distribution mixtures and application to major genes detection. *Statistics* **24**, 149-160.

Robert C.P. (1994). Mixtures of distributions: inference and estimation. *In* Practical MCMC.

Roeder K. (1992). Semiparametric estimation of normal mixture densities. *Ann. Statist.* **20**, 929-943.

Rosenblatt M. (1971). Curve estimates. *Ann. Math. Statist.* **42**, 1815-1842.

San Cristobal-Gaudy M. and Chevalet C. (1993). Some statistical problems involved by the detection of quantitative trait loci in animal genetics. *Discussion paper 9316, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgique.*

Serfling, R.J. (1980). *Approximation theorems of mathematical statistics.* Wiley, New York.

Silverman B.W. (1986). *Density estimation for statistics and data analysis.* Chapman and Hall.