Philipps Universität Marburg

Identification and Estimation in semiparametric two-component mixtures

Diploma thesis

by

Daniel Hohmann

Großseelheimer Straße 61
35039 Marburg

Supervisor

Prof. Dr. Hajo Holzmann

Marburg, August 23, 2010

Acknowledgments

I am very much indebted to Prof. Dr. Hajo Holzmann for mentoring this thesis. He brought basic ideas to my work and always made time for support and discussions. Also, I am grateful for suggesting this topic, leading me to a highly interesting field of research, what actually makes me looking forward to future work.

Further on I would like to thank the entire work group for the pleasant atmosphere, the enjoyable afternoon breaks, and the considerable help provided. Particularly, I want to mention Florian and Johannes, whose supportive comments helped me to improve this work a lot. Thanks also go to Thorsten for sharing both the office and bad jokes all day long. Kristin, thank you for being by my side.

Without any doubts my family deserves the biggest thanks, though, especially my parents Elisabeth and Paul. Thank you for supporting me in all respects. My studies would not have been possible without you.

Finally, thanks go to all my fellow students and friends.

Contents

1. Introduction **1**

2. Identifiability **3**

2.1. Definitions 3
2.2. Identifiability of the semiparametric two-component mixture 5
2.3. Almost periodic functions 9
2.4. A location parameter extension 12
2.4.1. Identifiability results in case of degeneracy 15
2.4.2. Global results 17
2.5. A scale parameter extension 20

3. Estimation **24**

3.1. Estimating the model parameter 24
3.2. Empirical processes 27
3.2.1. Glivenko-Cantelli theorem 28
3.2.2. Donsker's theorem and the functional Delta-method 30
3.2.3. Smooth Estimates 33
3.3. Consistency 36
3.4. Asymptotic normality 40

4. Simulation study **56**

5. Outlook **59**

Bibliography **60**

A. Notation **63**

A.1. Total and partial derivatives 63
A.2. Stochastic Landau symbols 64

B. Declaration **65**

1. Introduction

During the last century mixture models have become of great interest in the field of statistic analysis. All the more when group structure within the given data is known or at least supposed, modeling heterogeneity by means of a mixture of distributions is highly reasonable and forms the basic tool for advanced cluster analysis, latent class analysis, and discriminant analysis. But applications are more wide-spread. The use of mixture models also provides a useful framework for the general modeling and fitting of unknown distributional shapes, e. g. density estimation, even if the single components do not have a physical interpretation. The main reason, of course, is the higher flexibility one gains through working with mixtures of varied distributions, without, however, losing control of the shape and the number of components as it is in general kernel density estimation, for instance.

First approaches to statistic analysis using mixture models are due to Newcomb [19] and Pearson [20]. In his now classic paper Pearson fits a mixture of two normal distributions to a sample of crab measurements, wherewith he attempts to substantiate the presence of two subspecies within the observed population. The unknown parameters of his model are determined by means of the so called *method of moments*. In the following decades finite mixture models, especially parametric mixtures, were studied exhaustively. With the advent of high speed computers the methodology of estimation turned to *maximum likelihood* techniques, attributable to Rao [21], and later also to the application of the *EM algorithm*, first formalized by Dempster, Laird, and Rubin [8].[1]

In the seventies, however, first questions about the predominant usage of parametric mixture models arose. In fact, given observed data, stipulating the underlying parametric family of distributions is hard but essential. This is why the field of interest also turned to the study of so called semiparametric mixture models. In the early eighties Hall [10] and Titterington [24], for example, introduced estimators for the mixing proportions of a finite mixture without imposing any parametric assumptions to the component distributions. The proposed methodology, however, needs the number of components to be fixed and known. Further, training data has to be on hand, not only for the mixture itself but rather for each single component. About twenty years later Hettmansperger and Thomas [13] and Cruz-Medina, Hettmansperger, and Thomas [7] suggested a generalized method, the so called *cut-point approach*, an estimation algorithm for both the mixing proportions and the unknown number of components. Essential for this methodology is the presence of independent and identically distributed repeated measurements for each subject.

[1] For a brief sketch of the history of finite mixture models see e. g. McLachlan and Peel [18].

Training data for the single mixture components is no longer required. Hall and Zhou [11] considered a related model, a semiparametric two-component mixture for d-variate data, where the component distributions are unspecified except of the assumption that each observation, conditioned on the component which it comes from, has independent marginals. They established identifiability of their model for $d \geq 3$ and suggested a strongly consistent estimation method, including estimates for the mixing proportions and the two d-variate component distributions. Hunter, Wang, and Hettmansperger [14] and Bordes, Mottelet, and Vandekerkhove [5], independently of each other, considered a k-component mixture for univariate data with coinciding mixture components up to individual location parameters. Assuming the component distributions to be symmetric, they proved identifiability of their mixture model for $k = 2$ and $k = 3$ and introduced an estimator, which, in case of identifiability, is strongly consistent for the mixing proportions and the location parameters under mild assumptions. At the same time Bordes, Delmas, and Vandekerkhove [4] treated a semiparametric two-component mixture having symmetric but potentially different component distributions. In order to obtain identifiability in this generalized setup they assumed one of the mixture components to be known. Applying this model is reasonable whenever one component is actually known, of course, but also when at least additional training data is available for one of the mixture components. Thus, it can be ranged in between the models of Hall [10] and Hunter et al. [14]. A consistent estimator for the mixing proportions and the location parameter of the unknown component was provided.

In this work we will exclusively address the type of mixture model considered by Bordes et al. [4] and also Bordes and Vandekerkhove [3], i. e. modeling univariate data by means of a semiparametric two-component mixture with symmetric component distributions, where adequate information about one of the components is assumed to on hand. The thesis is organized as follows.

Section 2 gives basic definitions of identifiability and a mathematical constitution of the considered mixture model first. A brief summary of the identifiability results established by Bordes et al. [4] follows. Thereafter, we consider two extensions of this mixture model. We introduce an additional location parameter on the one hand and an additional scale parameter on the other hand and suggest conditions providing identifiability in these generalized setups.

In section 3 an adapted semiparametric estimator for the location parameter extended mixture model is introduced. After recapitulating some basic theory of empirical processes, this estimator is proved to be strongly consistent for its Euclidean part and asymptotically normal as a whole.

Finally, we give a numerical validation of the established estimator in section 4 and a brief outlook to future work in section 5.

2. Identifiability

2.1. Definitions

Given a family of probability distributions $\mathcal{F}$, commonly represented by the corresponding cumulative distribution functions (cdf) or probability density functions (pdf), a finite mixture model is determined by its number of components $k \in \mathbb{N}$, an underlying parameter space Θ, and a mixture m, stipulating the mixing regulation. Frequently, for some $k, d \in \mathbb{N}$,

$$\Theta = \left\{ (p_1, \ldots, p_k) \in \mathbb{R}^k \mid 0 \leq p_1, \ldots, p_k \leq 1, \sum_{i=1}^k p_i = 1 \right\} \times \mathcal{F}^k$$

and

$$m(x; \theta) = \sum_{i=1}^{k} p_i f_i(x) \quad , \quad x \in \mathbb{R}^d \, , \tag{2.1.1}$$

where $\theta = (p_1, \ldots, p_k, f_1, \ldots, f_k) \in \Theta$ is the model parameter. (2.1.1) is referred to as a k-*component mixture* for d-variate data. The p_i's are called the *mixing weights* or *mixing proportions*, the f_i's are the *component distributions* or *component densities*. If the family $\mathcal{F}$ has a parametric shape, e. g.

$$\{x \mapsto \sigma^{-1} f((x - \mu)/\sigma) \mid \mu \in \mathbb{R}^d, \sigma > 0\}$$

for some fixed pdf f, the mixture (model) is said to be parametric. Otherwise, if $\mathcal{F}$ cannot be specified parametrically, it is referred to as being *semiparametric*. As in (2.1.1) the number of components k is often assumed to be fixed. In many semiparametric setups this is even essential to obtain any identifiability results, see e. g. Hunter et al. [14], Bordes et al. [5].

Loosely spoken, *identifiability* describes the fact that a particular parameter configuration yields a unique mixture, or the other way around, that a given mixture density completely determines the corresponding model parameter. Let us denote by $\mathcal{M}_m(\Theta)$ the set of mixture densities induced by mixture m and the parameter space Θ, i. e.

$$\mathcal{M}_m(\Theta) = \{x \mapsto m(x; \theta) \mid \theta \in \Theta\} \, .$$

Correspondingly, we define $\mathcal{M}_m(\Theta_0)$ for all subsets $\Theta_0 \subset \Theta$. Further, let us introduce the map $\pi_m : \Theta \to \mathcal{M}_m(\Theta)$, canonically defined by

$$\pi_m : \theta \mapsto m(x;\theta) \ .$$

As motivated by the latter description, the definition of identifiability can be put from two directions. First, we formally introduce the identifiability of a considered parameter space.

Definition 2.1 (Identifiability parameter space)**.** *A parameter space $\Theta_0 \subset \Theta$ is said to be identifiable if the restriction of π_m to Θ_0 is one to one, i. e. each pair of parameters $\theta_1, \theta_2 \in \Theta_0$, $\theta_1 \neq \theta_2$ induces two different mixtures.*

To assess the difference of two parameters in this semiparametric setup one e. g. applies the product norm

$$\|\theta\| = \sum_{i=1}^{k} |p_i| + \sum_{i=1}^{k} \int_{\mathbb{R}} |f_i(x)| \, dx \ ,$$

provided that the f_i's are Lebesgue densities. With this, $\theta^{(1)}, \theta^{(2)} \in \Theta$ are considered as being equal if there exists a permutation σ of $\{1, \ldots, k\}$ such that

$$\sum_{i=1}^{k} |p_i^{(1)} - p_{\sigma(i)}^{(2)}| + \sum_{i=1}^{k} \int_{\mathbb{R}} |f_i^{(1)}(x) - f_{\sigma(i)}^{(2)}(x)| \, dx = 0 \ .$$

Note that this approach also handles nonidentifiability due to label switching as two parameters which only differ in the order of their components are considered as being equal. This is reasonable since such parameters in the general setup (2.1.1) always induce the same mixture pdf.

Second, we define the identifiability of a single mixture.

Definition 2.2 (Identifiability mixture)**.** *Let $\Theta_0 \subset \Theta$ and $\theta_0 \in \Theta_0$. The mixture $\pi_m(\theta_0)$ is said to be identifiable over the parameter space Θ_0 if it has a unique representative in it, i. e.*

$$\#\big(\pi_m^{-1}(\pi_m(\theta_0)) \cap \Theta_0\big) = 1 \ .$$

Remark. (i) There is an obvious link between the definitions 2.1 and 2.2. In fact, a parameter space Θ_0 is identifiable if and only if all mixtures in $\mathcal{M}_m(\Theta_0)$ are identifiable over Θ_0.

(ii) The concept of identifiability is highly important when it comes to the estimation of the model parameter based on training data for the mixture. Within a nonidentifiable setup an estimate $\hat{\theta}$ has no meaningful interpretation. Also, whenever a mixture is nonidentifiable, consistency for the estimator can clearly not be obtained.

2.2. Identifiability of the semiparametric two-component mixture

Let $\mathcal{E}$ denote the set of zero-symmetric (even) Lebesgue densities on $\mathbb{R}$, i. e.

$$\mathcal{E} = \left\{ f : \mathbb{R} \to \mathbb{R} \mid f \geq 0, \int f(x)\, dx = 1, f(x) = f(-x), x \in \mathbb{R} \right\} .$$

Then, the mixture model studied by Bordes et al. [4],[3] is defined by the parameter space

$$\Theta = [0,1] \times \mathbb{R} \times \mathcal{E}$$

and the overlying mixture

$$g(x;\theta) = (1-p) f_0(x) + p f(x-\mu) \quad , \quad x \in \mathbb{R} , \tag{2.2.1}$$

where $\theta = (p, \mu, f) \in \Theta$ is the semiparametric model parameter, including the mixing proportion p, the unknown component density f, and the corresponding location parameter μ. The density $f_0 \in \mathcal{E}$ is fixed and assumed to be known throughout the following.

Even though we assume both f_0 and f to be even it is easy to see that identifiability of the whole parameter space Θ fails. If the mixing proportion p is equal to zero, for instance, then any $f_1, f_2 \in \mathcal{E}$ and $\mu_1, \mu_2 \in \mathbb{R}$ yield

$$g(x; p, \mu_1, f_1) = g(x; p, \mu_2, f_2) \quad , \quad x \in \mathbb{R}$$

and hence $\pi_g(p, \mu_1, f_1) = \pi_g(p, \mu_2, f_2)$. Also, if $\mu = 0$, then the pdf $x \mapsto g(x; p, \mu, f)$ is even itself, yielding

$$\pi_g(p, \mu, f) = \pi_g(1, 0, g(x; p, \mu, f)) .$$

If $\mu = 0$ as well as $f = f_0$, then

$$\pi_g(p_1, \mu, f) = \pi_g(p_2, \mu, f)$$

for any $p_1, p_2 \in [0,1]$.

In order to avoid these problematic cases in advance we restrict further studies to the parameter space

$$\Theta_1 = (0,1] \times \mathcal{R}_1 \times \mathcal{E} , \tag{2.2.2}$$

where for convenience $\mathcal{R}_1 := \mathbb{R} \backslash \{0\}$.

Unfortunately, identifiability of this restricted parameter space cannot be obtained, either. Simple counterexamples can be given.

Example 2.3. Let us consider a mixture of two *uniform distributions* with distinct but connected supports and mixing proportion $p_1 \in (0,1]$, e. g.

$$(1-p_1)(2a)^{-1}\mathbf{1}_{(-a,a)}(x) + p_1(2b)^{-1}\mathbf{1}_{(-b,b)}(x-(a+b)) \quad , \quad x \in \mathbb{R} \tag{2.2.3}$$

for some $a, b > 0$. With $f_0(x) = (2a)^{-1}\mathbf{1}_{(-a,a)}(x)$, $f_1(x) = (2b)^{-1}\mathbf{1}_{(-b,b)}(x)$, and $\theta_1 = (p_1, a+b, f)$, mixture (2.2.3) can be written as

$$\pi_g(\theta_1) \ .$$

Whenever p_1 satisfies

$$p_1\big(1 + a/b\big) \le 1$$

the pdf $\pi_g(\theta_1)$ is not identifiable over Θ_1, though. With

$$f_2(x) = (2(a+b))^{-1}\mathbf{1}_{(-a-b,a+b)}(x-b) \ ,$$

which covers the whole support $(-a, a+2b)$ of $\pi_g(\theta_1)$, for all $x \in (a, a+2b)$,

$$g(x;\theta_1) = p_1/(2b) \quad , \quad f_2(x) = (2(a+b))^{-1} \ .$$

Therefore, defining

$$p_2 := p_1 \, \frac{2(a+b)}{2b} = p_1\Big(1 + \frac{a}{b}\Big) \in (0,1]$$

and $\theta_2 := (p_2, b, f_2)$,

$$g(x;\theta_2) = g(x;\theta_1) \quad , \quad x \in (a, a+2b) \ .$$

Moreover, for all $x \in (-a, a)$,

$$\begin{aligned} g(x;\theta_2) &= (1-p_2)(2a)^{-1} + p_2(2(a+b))^{-1} \\ &= \frac{2b - 2p(a+b)}{4ab} + \frac{p}{2b} = \frac{1-p}{2a} \\ &= g(x;\theta_1) \ , \end{aligned}$$

too. Hence, $\pi_g(\theta_2) = \pi_g(\theta_1)$, whereas $\theta_1 \ne \theta_2$.

Necessary nonidentifiability conditions

In order to work for a parameter space which is identifiable with respect to mixture g, as a first step we derive a relation, which holds whenever two parameters reveal a situation of nonidentifiability.

Let us denote by $\mathcal{E}_k \subset \mathcal{E}$, $k \in \mathbb{N}$ the set of even pdfs having a k^{th} order moment,

i. e.

$$\mathcal{E}_k = \{f \in \mathcal{E} \mid \int |x|^k f(x)\, dx < \infty\} .$$

Denoting by $g^{(k)}$ the k^{th} order moment of g, for $(p, \mu, f) \in \Theta_1$ we obtain

$$\begin{aligned} g^{(1)}(p, \mu, f) &= p\mu , \\ g^{(2)}(p, \mu, f) &= (1-p)\xi_0 + p(\xi + \mu^2) , \text{ and} \\ g^{(3)}(p, \mu, f) &= p(3\xi\mu + \mu^3) , \end{aligned}$$

referring to the second order moments of f_0 and f as ξ_0 and ξ, respectively. Note that, being even, f_0 and f have first and third order moments equal to zero.

Lemma 2.4. *Let $\theta_1, \theta_2 \in \Theta_1$, cf. (2.2.2), $\theta_i = (p_i, \mu_i, f_i)$, $i = 1, 2$. If $\pi_g(\theta_1) = \pi_g(\theta_2)$, then either $\theta_1 = \theta_2$ or θ_2 can be expressed in terms of θ_1 according to*

$$\begin{aligned} p_2 &= \frac{2\mu_1^2 p_1}{\mu_1^2 + 3(\xi_1 - \xi_0)} , \\ \mu_2 &= \frac{\mu_1^2 + 3(\xi_1 - \xi_0)}{2\mu_1} , \text{ and} \\ \xi_2 &= \xi_1 + \frac{(\mu_1^2 + \xi_1 - \xi_0)(\mu_1^2 - 3(\xi_1 - \xi_0))}{4\mu_1^2} , \end{aligned} \tag{2.2.4}$$

where ξ_0, ξ_1, and ξ_2 respectively denote the second order moments of f_0, f_1, and f_2.

The detailed proof of lemma 2.4 can be found in Bordes et al. [4], p. 735, proposition 1. Basically, the system of moment equations

$$g^{(k)}(\theta_1) = g^{(k)}(\theta_2) \quad , \quad k = 1, 2, 3 ,$$

induced by the equality $\pi_g(\theta_1) = \pi_g(\theta_2)$, leads to

$$\left(2p_1\mu_1\mu_2 - p_1(\mu_1^2 + 3(\xi_1 - \xi_0))\right)(\mu_1 - \mu_2) = 0 ,$$

what is obviously fulfilled if and only if μ_2 is equal to μ_1 or if μ_2 is as in (2.2.4). Either way, the corresponding parameters p_2 and f_2 can be determined easily thereafter.

Sufficient identifiability conditions

So as to obtain satisfactory identifiability results for mixture g we now in addition take account of the mixture's Fourier transform. Recall that the Fourier transform of a Lebesgue integrable function f, say, is given by

$$\hat{f} : \mathbb{R} \to \mathbb{C} \quad , \quad t \mapsto \hat{f}(t) = \int_{\mathbb{R}} e^{ity} f(y)\, dy .$$

Applying some basic rules for computation[1] we obtain

$$\hat{g}(t;\theta) = (1-p)\hat{f}_0(t) + pe^{i\mu t}\hat{f}(t) \quad , \quad t \in \mathbb{R} ,$$

where $\hat{f}_0$ and $\hat{f}$ denote the Fourier transforms of f_0 and f, respectively. Since f_0 and f are even functions, the corresponding Fourier transforms are real, $\hat{f}_0, \hat{f} : \mathbb{R} \to \mathbb{R}$. As a result, splitting up $\hat{g}$ into its real and imaginary part yields

$$\begin{aligned} &\operatorname{Re} \hat{g}(t; p, \mu, f) = (1-p)\hat{f}_0(t) + p\cos(\mu t)\hat{f}(t) \quad \text{and} \\ &\operatorname{Im} \hat{g}(t; p, \mu, f) = p\sin(\mu t)\hat{f}(t) \end{aligned} \tag{2.2.5}$$

by means of *Euler's formula.*

There are many ways of using the Fourier transform $\hat{g}$ as for the derivation of adequate conditions to the model parameter (p, μ, f) ensuring identifiability. Frequently, regularity conditions with respect to the component's Fourier transforms are claimed, e. g.

$$\lim_{t\to\infty} \hat{f}(t)/\hat{f}_0(t) = 0 \ .$$

Adapting this approach to mixture g actually leads to suitable identifiability results as we will see later, cf. corollary 2.18. In this section, though, we bring up the technique used by Bordes et al. [4], at what, strongly based on the symmetry of f_0 and f, the Fourier transform $\hat{g}$ supplies satisfactory information while imposing none but a rather weak condition on $\hat{f}_0$.

Theorem 2.5. *If $\hat{f}_0 > 0$, then the parameter space*

$$\Theta^* = \big\{(0,1] \times \mathcal{R}_1 \times \mathcal{E}_3\big\} \setminus \bigcup_{z\in\mathbb{Z}} \big\{(p,\mu,f) \mid \xi = \xi_0 + \frac{z-2}{3z}\,\mu^2\big\} ,$$

where ξ_0 and ξ respectively denote the second order moments of f_0 and f, is identifiable, i. e. each mixture in $\mathcal{M}_g(\Theta^*)$ has a unique representative in Θ^*.

The proof of theorem 2.5 is stated in Bordes et al. [4], p. 736, proposition 2. The basic idea is as follows. Provided that $(p_1, \mu_1, f_1), (p_2, \mu_2, f_2) \in (0,1] \times \mathcal{R}_1 \times \mathcal{E}_3$ give $\pi_g(p_1, \mu_1, f_1) = \pi_g(p_2, \mu_2, f_2)$, the system of equations induced by (2.2.5) yields

$$(p_2 - p_1)\sin(\mu_2 t)\hat{f}_0(t) = p_1 \sin((\mu_1 - \mu_2)t)\hat{f}_1(t) \quad , \quad t \in \mathbb{R} . \tag{2.2.6}$$

With $p_1 \neq p_2$, the particular argument $t^* = \pi/(\mu_1 - \mu_2)$ gives

$$\sin\Big(\frac{\mu_2}{\mu_1 - \mu_2}\,\pi\Big) = 0 ,$$

[1]Cf. Bauer [1], p. 191.

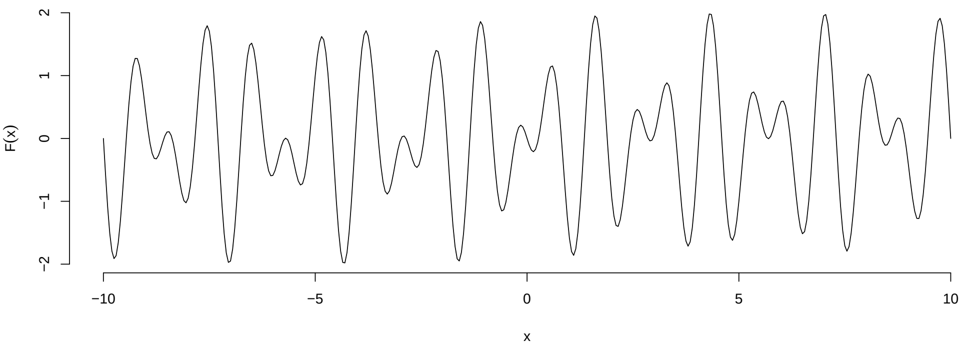

Figure 2.1.: The almost periodic function $x \mapsto F(x) = \sin(1.5\pi x) - \sin(2.2\pi x)$.

which holds true if and only if $z^* := \mu_2/(\mu_1 - \mu_2)$ is integer, so that

$$\mu_2 = \mu_1 \cdot z^*/(z^* + 1)$$

for some $z^* \in \mathbb{Z}$. Regarding lemma 2.4 we obtain

$$\xi_1 = \xi_0 + \frac{z^* - 1}{3(z^* + 1)}\, \mu_1^2 \,.$$

If $p_1 = p_2$, then we straightforward conclude $(p_1, \mu_1, f_1) = (p_2, \mu_2, f_2)$. Hence, the parameter space as stated in fact excludes all the crucial parameter configurations.

2.3. Almost periodic functions

For the further analysis of identifiability of semiparametric two-component mixtures we will often deal with functions resulting from the intermixture of different sines, e. g.

$$x \mapsto \alpha_1 \sin(\beta_1 x) + \alpha_2 \sin(\beta_2 x) \quad , \quad x \in \mathbb{R} \tag{2.3.1}$$

with $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{R}$. Figure 2.1 gives an example of (2.3.1). In general, these functions are known not to be *periodic*. Nevertheless, several useful properties can be derived.

Definition 2.6 (Almost periodic function)**.** *Let $f : \mathbb{R} \to \mathbb{R}$.*

(i) The function f is said to be periodic if there exists $\tau \in \mathbb{R}$ such that

$$f(x + \tau) = f(x) \quad , \quad x \in \mathbb{R} \,.$$

The number τ is called a period or translation number of f.

(ii) We say the function f is almost periodic if for each arbitrary small value $\varepsilon > 0$ there exists a real number $L_\varepsilon > 0$ such that every interval $I \subset \mathbb{R}$ of length

greater or equal to L_ε contains at least one real number $\tau_{\varepsilon,I}$ giving

$$|f(x+\tau_{\varepsilon,I}) - f(x)| < \varepsilon \quad , \quad x \in \mathbb{R} .$$

The number $\tau_{\varepsilon,I}$ is called an ε-translation number of f.

(iii) The function f is said to be normal if for any sequence $(y_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ one can find a subsequence $(y'_n)_{n\in\mathbb{N}}$ such that the function series $(f(x+y'_n))_{n\in\mathbb{N}}$ converges uniformly in x.

The simplest periodic functions one may think of is the trigonometric function $x \mapsto \sin(x)$. A corresponding period is 2π, for instance. Specifying an almost periodic or a normal function seems to be much harder for now. Therefore, we observe the following coherences at the first.

Lemma 2.7. *(i) Each periodic function is almost periodic.*

(ii) A continuous function is almost periodic if and only if it is normal.

Proof. (i) Let $f : \mathbb{R} \to \mathbb{R}$ be a periodic function with translation number τ and let us for some $a \in \mathbb{R}$ consider the arbitrary real interval $I = [a, a+|\tau|]$ of length $|\tau|$. Choosing $z \in \mathbb{Z}$ as the smallest integer such that $z\tau \geq a$, then $z\tau \in I$ and, if $z \geq 0$,

$$|f(x+z\tau) - f(x)| \leq \sum_{k=0}^{z-1} |f(x+(k+1)\tau) - f(x+k\tau)| = 0 \quad , \quad x \in \mathbb{R} .$$

Likewise, if $z < 0$,

$$|f(x+z\tau) - f(x)| \leq \sum_{k=0}^{|z|-1} |f(x-(k+1)\tau) - f(x-k\tau)| = 0 \quad , \quad x \in \mathbb{R} .$$

Therefore, for any arbitrary small value $\varepsilon > 0$ we (independently from ε) choose $L_\varepsilon := |\tau|$ and, according to definition 2.6 (ii), conclude that f is almost periodic.

(ii) The equivalence of almost periodicity and normality for continuous functions on $\mathbb{R}$ is stated in Corduneanu [6], p. 140, theorem 6.6, for instance. □

Theorem 2.8. *(i) The set of continuous almost periodic functions forms a vector space over the field $\mathbb{R}$.*

(ii) If $f : \mathbb{R} \to \mathbb{R}$ is almost periodic and

$$\lim_{x\to\infty} f(x) = 0 ,$$

then $f(x) = 0$ for all $x \in \mathbb{R}$.

Proof. (i) Let $f, g : \mathbb{R} \to \mathbb{R}$ be continuous almost periodic functions. Let $\alpha \in \mathbb{R}$ be any scalar. If $\alpha = 0$, then $\alpha f(x) \equiv 0$ and hence αf is almost periodic since constant functions clearly are. So, let $\alpha \neq 0$ from now on.

Consider $\varepsilon > 0$ and an arbitrary interval I of length greater or equal to L_δ, at what $\delta > 0$ is defined by $\delta := \varepsilon/|\alpha|$. Then, per definition, there exists a δ-translation number $\tau_{\delta,I} \in I$ of f such that

$$|\alpha f(x+\tau_{\delta,I}) - \alpha f(x)| = |\alpha| \cdot |f(x+\tau_{\delta,I}) - f(x)| < |\alpha| \cdot \delta = \varepsilon \quad , \quad x \in \mathbb{R} .$$

Thus, the continuous function αf is almost periodic.

Let us consider an arbitrary sequence $(y_n)_{n\in\mathbb{N}} \subset \mathbb{R}$. According to lemma 2.7 (ii) we choose a subsequence $(y'_n)_{n\in\mathbb{N}}$ such that $x \mapsto f(x+y'_n)$ converges uniformly as n tends to infinity. Further, we choose a subsequence $(y''_n)_{n\in\mathbb{N}}$ of $(y'_n)_{n\in\mathbb{N}}$ such that $x \mapsto g(x+y''_n)$ converges uniformly, too. Therewith, $x \mapsto (f+g)(x)$ is continuous and the function series $x \mapsto (f+g)(x+y''_n)$ converges uniformly. We conclude that $f+g$ is almost periodic, see again lemma 2.7 (ii).

(ii) The second part we prove by contradiction. Suppose that there exists $x_0 \in \mathbb{R}$ such that $|f(x_0)| \geq \varepsilon$ for some $\varepsilon > 0$. Since $f(x) \to 0$ as $x \to \infty$, there exists $T \in \mathbb{R}$ such that

$$|f(x)| < \varepsilon/2 \quad , \quad x \geq T .$$

As f is almost periodic, we choose $L_{\varepsilon/2} > 0$ according to definition 2.6 (ii) and an appropriate $\varepsilon/2$-translation number $\tau_{\varepsilon/2,[T+|x_0|,T+|x_0|+L_{\varepsilon/2}]} =: \tau^*$. By means of triangle inequality we obtain

$$|f(x_0)| \leq |f(x_0) - f(x_0+\tau^*)| + |f(x_0+\tau^*)| < \varepsilon ,$$

what contradicts the assumption $|f(x_0)| \geq \varepsilon$ and thus concludes the proof. ☐

Corollary 2.9. *Let $\alpha, \beta \in \mathbb{R}$ be such that*

$$\lim_{x\to\infty} \alpha \sin(\beta x) = 0 .$$

Then $\alpha = 0$ or $\beta = 0$ holds.

Proof. The function $x \mapsto F(x) := \alpha \sin(\beta x)$ is periodic with period $2\pi/\beta$, for instance. By lemma 2.7 (i) F is almost periodic and therefore from theorem 2.8 (ii) it follows

$$F(x) = 0 \quad , \quad x \in \mathbb{R} .$$

If $\beta \neq 0$, then the particular argument $x = \pi/(2\beta)$ yields $F(x) = \alpha$ and thus $\alpha = 0$. If vice versa $\alpha \neq 0$, then $\beta = 0$ follows. ☐

Corollary 2.10. *Let $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{R}\backslash\{0\}$, $\alpha_1, \alpha_2 > 0$, and let*

$$F(x) = \alpha_1 \sin(\beta_1 x) - \alpha_2 \sin(\beta_2 x) . \tag{2.3.2}$$

If $F(x) \to 0$ as $x \to \infty$, then $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$.

Proof. The function F is almost periodic by lemma 2.7 (i) and theorem 2.8 (i). Therefore, as $F(x) \to 0$, $x \to \infty$, from theorem 2.8 (ii) it follows $F(x) \equiv 0$, yielding

$$\alpha_1 \sin(\beta_1 x) = \alpha_2 \sin(\beta_2 x) \quad , \quad x \in \mathbb{R} . \tag{2.3.3}$$

Since both hand sides of (2.3.3) are continuous functions respectively taking on all values in $[-\alpha_1, \alpha_1]$ and $[-\alpha_2, \alpha_2]$ we conclude $\alpha_1 = \alpha_2$. Canceling α_1, α_2 and taking derivatives of (2.3.3) leads to

$$\beta_1 \cos(\beta_1 x) = \beta_2 \cos(\beta_2 x) \quad , \quad x \in \mathbb{R} ,$$

where the particular argument $x = 0$ gives $\beta_1 = \beta_2$. The proof is concluded. □

2.4. A location parameter extension

So as to derive more general identifiability results, in this section we extend the previously considered mixture model in as much as we introduce an additional location parameter ν to the mixture component that was assumed to be completely known up to now. Accordingly, in the following we consider the parameter space

$$\Theta = [0, 1] \times \mathbb{R}^2 \times \mathcal{E}$$

and the mixture

$$h(x; \theta) = (1 - p) f_0(x - \nu) + p f(x - \mu) \tag{2.4.1}$$

with $\theta = (p, \nu, \mu, f) \in \Theta$. Further on, the pdf $f_0 \in \mathcal{E}$ is fixed and assumed to be known.

Definition 2.11 (Degenerate mixture)**.** *If $p \in \{0, 1\}$, then the mixture $\pi_h(\theta)$ is said to be degenerate. Correspondingly, we characterize the mixture as being 0-degenerate or 1-degenerate. Otherwise, $\pi_h(\theta)$ is said to be nondegenerate.*

Obviously, identifiability cannot be obtained in case of degeneracy. For example,

$$\pi_h(1, \nu_1, \mu, f) = \pi_h(1, \nu_2, \mu, f)$$

for all $\nu_1, \nu_2 \in \mathbb{R}$. We handle this problem by giving the following enhanced definition, where

$$\iota_k : \Theta \to \mathbb{R} \cup \mathcal{E} \quad , \quad k = 1, \ldots, 4$$

denotes the projection onto the k^{th} component.

Definition 2.12 (Identifiability of a degenerate mixture)**.** *Let $\Theta_0 \subset \Theta$ and $\theta_0 \in \Theta_0$ such that $\pi_h(\theta_0)$ is degenerate.*

(i) If $\pi_h(\theta_0)$ *is 0-degenerate, then it is said to be identifiable over* Θ_0 *if for all* $\theta \in \pi_h^{-1}(\pi_h(\theta_0)) \cap \Theta_0$ *it holds*

$$\iota_1(\theta) = 0 \quad , \quad \iota_2(\theta) = \iota_2(\theta_0) \; .$$

(ii) If $\pi_h(\theta_0)$ *is 1-degenerate, then it is said to be identifiable over* Θ_0 *if for all* $\theta \in \pi_h^{-1}(\pi_h(\theta_0)) \cap \Theta_0$ *it holds*

$$\iota_1(\theta) = 1 \quad , \quad \iota_3(\theta) = \iota_3(\theta_0) \quad , \quad \iota_4(\theta) = \iota_4(\theta_0) \; .$$

Motivated by the fact that coinciding component densities form a source of non-identifiability,

$$h(x; p, \nu, \mu, f_0) = h(x; 1-p, \mu, \nu, f_0) \quad , \quad x \in \mathbb{R} \; ,$$

we consider the following regularity conditions with respect to the unknown component pdf f, where $\hat{f}_0$ and $\hat{f}$ again denote the Fourier transforms of f_0 and f, respectively.

Condition C1. For large $t \in \mathbb{R}$ it holds $\hat{f}_0(t) \neq 0$ and

$$\lim_{t \to \infty} \frac{\hat{f}(t)}{\hat{f}_0(t)} = 0 \; .$$

Condition C2. For large $t \in \mathbb{R}$ it holds $\hat{f}(t) \neq 0$ and

$$\lim_{t \to \infty} \frac{\hat{f}_0(t)}{\hat{f}(t)} = 0 \; .$$

The conditions C1 and C2 guarantee that the component pdfs f_0 and f do not coincide, but their Fourier transforms even differ significantly, what makes them "distinguishable". Let us define the disjoint sets

$$\mathcal{E}^{f_0,i} := \{ f \in \mathcal{E} \mid f \text{ meets condition Ci} \} \quad , \quad i = 1, 2$$

and the union

$$\mathcal{E}^{f_0} := \mathcal{E}^{f_0,1} \cup \mathcal{E}^{f_0,2} \; .$$

As we will further require the existence of some moments of pdf h let

$$\mathcal{E}_k^{f_0} := \mathcal{E}_k \cap \mathcal{E}^{f_0} \quad , \quad k \in \mathbb{N} \; . \tag{2.4.2}$$

With $\theta = (p, \nu, \mu, f)$, provided that $f_0, f \in \mathcal{E}_3$,

$$\begin{aligned} h^{(1)}(\theta) &= (1-p)\nu + p\mu \ , \\ h^{(2)}(\theta) &= (1-p)(\xi_0 + \nu^2) + p(\xi + \mu^2) \ , \text{ and} \\ h^{(3)}(\theta) &= (1-p)(3\xi_0\nu + \nu^3) + p(3\xi\mu + \mu^3) \ , \end{aligned} \tag{2.4.3}$$

where $h^{(k)}$ denotes the k^{th} order moment of h and where ξ_0 and ξ denote the second order moments of f_0 and f, respectively.

Moreover, let us conformably denote by $\hat{h}$ the Fourier transform of mixture pdf h, that is

$$\hat{h}(t;\theta) = (1-p)\exp(i\nu t)\hat{f}_0(t) + p\exp(i\mu t)\hat{f}(t) \quad , \quad t \in \mathbb{R} \ .$$

Splitting up $\hat{h}$ into its real and imaginary part, minding again that $\hat{f}_0$ and $\hat{f}$ are real functions, we have

$$\begin{aligned} \operatorname{Re}\hat{h}(t;\theta) &= (1-p)\cos(\nu t)\hat{f}_0(t) + p\cos(\mu t)\hat{f}(t) \ , \\ \operatorname{Im}\hat{h}(t;\theta) &= (1-p)\sin(\nu t)\hat{f}_0(t) + p\sin(\mu t)\hat{f}(t) \ . \end{aligned} \tag{2.4.4}$$

Finally, we give a technical result, which helps for later calculations.

Lemma 2.13. *If $\theta_1, \theta_2 \in \Theta$, $\theta_i = (p_i, \nu_i, \mu_i, f_i)$, $i = 1, 2$ give $\pi_h(\theta_1) = \pi_h(\theta_2)$, then*

$$\hat{f}_1(t)p_1\sin((\mu_1-\mu_2)t) = \hat{f}_0(t)[(1-p_1)\sin((\mu_2-\nu_1)t) - (1-p_2)\sin((\mu_2-\nu_2)t)]$$

holds true for all $t \in \mathbb{R}$.

Proof. By the almost sure identity of $h(x;\theta_1)$ and $h(x;\theta_2)$, regarding (2.4.4), we obtain

$$\begin{aligned} (1-p_1)\cos(\nu_1 t)\hat{f}_0(t) + p_1\cos(\mu_1 t)\hat{f}_1(t) &= (1-p_2)\cos(\nu_2 t)\hat{f}_0(t) + p_2\cos(\mu_2 t)\hat{f}_2(t) \ , \\ (1-p_1)\sin(\nu_1 t)\hat{f}_0(t) + p_1\sin(\mu_1 t)\hat{f}_1(t) &= (1-p_2)\sin(\nu_2 t)\hat{f}_0(t) + p_2\sin(\mu_2 t)\hat{f}_2(t) \end{aligned}$$

for all $t \in \mathbb{R}$. Multiplying the first line by $\sin(\mu_2 t)$, we plug in the second equation into the first one, yielding

$$\begin{aligned} &\hat{f}_1(t)p_1[\sin(\mu_1 t)\cos(\mu_2 t) - \sin(\mu_2 t)\cos(\mu_1 t)] \\ &= \hat{f}_0(t)[(1-p_1)(\sin(\mu_2 t)\cos(\nu_1 t) - \sin(\nu_1 t)\cos(\mu_2 t)) \\ &\qquad - (1-p_2)(\sin(\mu_2 t)\cos(\nu_2 t) - \sin(\nu_2 t)\cos(\mu_2 t))] \end{aligned}$$

and hence

$$\begin{aligned} &\hat{f}_1(t)p_1[\sin(\mu_1 t)\cos(-\mu_2 t) + \sin(-\mu_2 t)\cos(\mu_1 t)] \\ &= \hat{f}_0(t)[(1-p_1)(\sin(\mu_2 t)\cos(-\nu_1 t) + \sin(-\nu_1 t)\cos(\mu_2 t)) \\ &\qquad - (1-p_2)(\sin(\mu_2 t)\cos(-\nu_2 t) + \sin(-\nu_2 t)\cos(\mu_2 t))] \ . \end{aligned}$$

By means of an addition theorem for the sine[2],

$$\begin{aligned}&(1-p_1)(\sin(\mu_2 t)\cos(-\nu_1 t)+\sin(-\nu_1 t)\cos(\mu_2 t))\\&-(1-p_2)(\sin(\mu_2 t)\cos(-\nu_2 t)+\sin(-\nu_2 t)\cos(\mu_2 t))\\=\ &(1-p_1)\sin((\mu_2-\nu_1)t)-(1-p_2)\sin((\mu_2-\nu_2)t)\ ,\end{aligned}$$

what leads to the stated assertion. □

2.4.1. Identifiability results in case of degeneracy

For convenience, let

$$\mathcal{R}_2 = \mathbb{R}^2 \backslash \{(x,x) \mid x \in \mathbb{R}\}$$

in what follows.

Theorem 2.14. *Let* $\theta_1 = (p_1, \nu_1, \mu_1, f_1) \in \{0,1\} \times \mathbb{R}^2 \times \mathcal{E}_1^{f_0}$*, cf. (2.4.2).*

(i) If $p_1 = 0$*, then* $\pi_h(\theta_1)$ *is identifiable over* $[0,1] \times \mathbb{R}^2 \times \mathcal{E}_1^{f_0}$*.*

(ii) If $p_1 = 1$*, then* $\pi_h(\theta_1)$ *is identifiable over* $[0,1] \times \mathcal{R}_2 \times \mathcal{E}_1^{f_0}$*.*

(iii) If in particular $p_1 = 1$ *and* f_1 *meets condition C1, then* $\pi_h(\theta_1)$ *is identifiable over* $[0,1] \times \mathbb{R}^2 \times \mathcal{E}_1^{f_0}$*.*

Proof. (i) Let $\theta_2 = (p_2, \nu_2, \mu_2, f_2) \in [0,1] \times \mathbb{R}^2 \times \mathcal{E}_1^{f_0}$ be such that $\pi_h(\theta_2) = \pi_h(\theta_1)$. Since $p_1 = 0$, for almost all $x \in \mathbb{R}$,

$$f_0(x-\nu_1) = (1-p_2)f_0(x-\nu_2) + p_2 f_2(x-\mu_2) \tag{2.4.5}$$

and

$$\nu_1 = (1-p_2)\nu_2 + p_2\mu_2 \tag{2.4.6}$$

by (2.4.3). If it was $p_2 = 1$, then (2.4.6) would give $\nu_1 = \mu_2$ and thus $f_0 = f_2$ by (2.4.5), which is a conflict with $f_2 \in \mathcal{E}_1^{f_0}$. Therefore, $p_2 \in [0,1)$. Choosing some $\mu_3 \in \mathbb{R}\backslash\{\nu_1, \nu_2, \mu_2\}$,

$$\pi_h(p_1, \nu_1, \mu_3, f_1) = \pi_h(p_1, \nu_1, \mu_1, f_1) = \pi_h(p_2, \nu_2, \mu_2, f_2)\ ,$$

yielding

$$\hat{f}_2(t)p_2\sin((\mu_2-\mu_3)t) = \hat{f}_0(t)[(1-p_2)\sin((\mu_3-\nu_2)t) - \sin((\mu_3-\nu_1)t)] \quad , \quad t \in \mathbb{R} \tag{2.4.7}$$

[2] For all $x, y \in \mathbb{R}$ it holds $\sin(x+y) = \sin(x)\cos(y) + \sin(y)\cos(x)$, cf. Beardon [2], p. 38, theorem 4.2.1.

by lemma 2.13. If $f_2 \in \mathcal{E}^{f_0,2}$, then (2.4.7) gives

$$\lim_{t\to\infty} p_2 \sin((\mu_2 - \mu_3)t) = 0 \ .$$

As $\mu_2 \neq \mu_3$ it follows $p_2 = 0$ by corollary 2.9 and then $\nu_1 = \nu_2$ by (2.4.6). If $f_2 \in \mathcal{E}^{f_0,1}$, then (2.4.7) leads to

$$\lim_{t\to\infty} (1-p_2)\sin((\mu_3 - \nu_2)t) - \sin((\mu_3 - \nu_1)t) = 0 \ .$$

As $p_2 < 1$ corollary 2.10 yields $p_2 = 0$ and $\nu_2 = \nu_1$, too.

(ii) Let $p_1 = 1$ and $\mu_2 \neq \nu_2$, for almost all $x \in \mathbb{R}$ yielding

$$f_1(x-\mu_1) = (1-p_2)f_0(x-\nu_2) + p_2 f_2(x-\mu_2) \tag{2.4.8}$$

as well as

$$\mu_1 = (1-p_2)\nu_2 + p_2\mu_2 \ . \tag{2.4.9}$$

By lemma 2.13 we have

$$\hat{f}_1(t)\sin((\mu_1-\mu_2)t) = -\hat{f}_0(t)(1-p_2)\sin((\mu_2-\nu_2)t) \quad , \quad t \in \mathbb{R} \ , \tag{2.4.10}$$

Again, we handle (2.4.10) by differentiating different cases. If $f_1 \in \mathcal{E}^{f_0,1}$, then it follows

$$\lim_{t\to\infty} (1-p_2)\sin((\mu_2 - \nu_2)t) = 0 \ ,$$

yielding $p_2 = 1$ as $\mu_2 \neq \nu_2$ and therewith $\mu_2 = \mu_1$ by (2.4.9). If $f_1 \in \mathcal{E}^{f_0,2}$, then (2.4.10) gives

$$\lim_{t\to\infty} \sin((\mu_1 - \mu_2)t) = 0$$

and hence $\mu_1 = \mu_2$, also leading to $p_2 = 1$ by (2.4.9) as $\nu_2 \neq \mu_2$. In both cases we additionally have $f_1 = f_2$ by (2.4.8).

(iii) Let $p_1 = 1$ and particularly $f_1 \in \mathcal{E}^{f_0,1}$. It remains to consider the case $\mu_2 = \nu_2$. From (2.4.9) it follows $\mu_1 = \mu_2 = \nu_2$ and thus

$$\hat{f}_1(t) = (1-p_2)\hat{f}_0(t) + p_2\hat{f}_2(t) \quad , \quad t \in \mathbb{R} \tag{2.4.11}$$

by (2.4.8). At first this gives $p_2 \neq 0$ as $\hat{f}_1 \neq \hat{f}_0$. Since $\hat{f}_1(t)/\hat{f}_0(t) \to 0$ we obtain

$$\lim_{t\to\infty} \hat{f}_2(t)/\hat{f}_0(t) = (p_2-1)/p_2 \ .$$

As $f_2 \in \mathcal{E}^{f_0}$, this holds true if and only if $p_2 = 1$. The remaining can be concluded as in (ii). □

Remark. Claiming condition C2, identifiability cannot be obtained in the case of

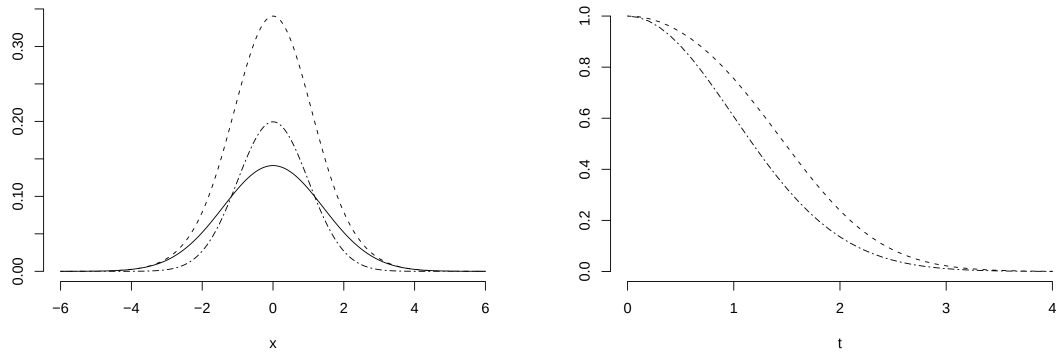

Figure 2.2.: Nonidentifiability in case of degeneracy: The left-hand side shows the pdf f_1 (dashed) resulting from the intermixture of two centered Gaussian pdfs f_0 (solid) and f (two-dashed) with variances $\sigma_0 = 2$ and $\sigma = 1$, respectively, at mixing proportion $p = 0.5$. The right-hand side shows the corresponding Fourier transform ratios $\hat{f}_0/\hat{f}_1$ (dashed) and $\hat{f}_0/\hat{f}$ (two-dashed).

coinciding component location parameters in general, cf. theorem 2.14 (ii) and (iii). For example, with $p \in (0,1)$, $\mu \in \mathbb{R}$, and $f_1 \in \mathcal{E}^{f_0,2}$, let us consider the mixture $\pi_h(p,\mu,\mu,f_1)$, i. e.

$$x \mapsto (1-p)f_0(x-\mu) + pf_1(x-\mu) \ .$$

As the pdf $f_2 := (1-p)f_0 + pf_1$ is even itself it follows

$$\pi_h(1,\mu,\mu,f_2) = \pi_h(p,\mu,\mu,f_1) \ ,$$

where f_2 meets condition C2, too. In fact,

$$\lim_{t\to\infty} \frac{\hat{f}_0(t)}{\hat{f}_2(t)} = \lim_{t\to\infty} \frac{\hat{f}_0(t)}{(1-p)\hat{f}_0(t) + p\hat{f}(t)} = \lim_{t\to\infty} \frac{\hat{f}_0(t)/\hat{f}(t)}{(1-p)\hat{f}_0(t)/\hat{f}(t) + p} = 0 \ .$$

Figure 2.2 illustrates this situation based on the mixture of two centered Gaussian pdfs with variances $\sigma_0 = 2$ and $\sigma = 1$.

2.4.2. Global results

Lemma 2.15. *Let* $\theta_i = (p_i,\nu_i,\mu_i,f_i) \in [0,1] \times \mathbb{R}^2 \times \mathcal{E}_3$, $i = 1,2$. *If in particular* $(p_1,\nu_1,\mu_1) \in (0,1) \times \mathcal{R}_2$ *as well as* $\mu_1 = \mu_2$ *and* $\pi_h(\theta_1) = \pi_h(\theta_2)$, *then* $\theta_1 = \theta_2$.

Proof. With (2.4.3) and $\mu_1 = \mu_2$, the almost sure identity of $h(x;\theta_1)$ and $h(x;\theta_2)$ gives

$$\begin{aligned} (1-p_1)\nu_1 + p_1\mu_1 &= (1-p_2)\nu_2 + p_2\mu_1 \ , \\ (1-p_1)(\xi_0+\nu_1^2) + p_1(\xi_1+\mu_1^2) &= (1-p_2)(\xi_0+\nu_2^2) + p_2(\xi_2+\mu_1^2) \ , \text{ and} \\ (1-p_1)(3\xi_0\nu_1+\nu_1^3) + p_1(3\xi_1\mu_1+\mu_1^3) &= (1-p_2)(3\xi_0\nu_2+\nu_2^3) + p_2(3\xi_2\mu_1+\mu_1^3) \ , \end{aligned}$$

where ξ_0, ξ_1, and ξ_2 label the second order moments of f_0, f_1, and f_2, respectively. The first moment equation can be written as

$$(1-p_2)\nu_2 = (1-p_1)\nu_1 + (p_1 - p_2)\mu_1 \ . \tag{2.4.12}$$

Multiplying the second moment equation by $(1-p_2)$ we obtain

$$(1-p_2)^2\nu_2^2 = (1-p_2)\left((1-p_1)(\xi_0+\nu_1^2) + p_1(\xi_1+\mu_1^2) - (1-p_2)\xi_0 - p_2(\xi_2+\mu_1^2)\right) . \tag{2.4.13}$$

Applying (2.4.12) and (2.4.13) to the $(1-p_2)^2$ times of the third moment equation yields

$$(\mu_1-\nu_1)^3(1-p_1)(p_2-p_1)(p_2+p_1-2) = 0 \ . \tag{2.4.14}$$

Since $\mu_1 \neq \nu_1$ and $p_1 < 1$ we conclude $p_2 = p_1$. Applying this to the first moment equation gives

$$(1-p_1)\nu_1 + p_1\mu_1 = (1-p_1)\nu_2 + p_1\mu_1 \ ,$$

which leads to $\nu_2 = \nu_1$ as $p_1 < 1$. Finally, the representation

$$f(x) = (h(x+\mu; p, \nu, \mu, f) + (1-p)f_0(x+\mu-\nu))/p \tag{2.4.15}$$

additionally provides the almost sure identity of f_1 and f_2. □

Theorem 2.16. *The parameter space*

$$\Theta^* = \left([0,1]\times\mathbb{R}^2\times\mathcal{E}_3^{f_0}\right)\setminus\left\{(p,\nu,\mu,f) \mid \mu=\nu\ ,\ f\in\mathcal{E}^{f_0,2}\right\}$$

is identifiable, i. e. each mixture in $\mathcal{M}_h(\Theta^*)$ has a unique representative in Θ^*.

Proof. Let $\theta_i = (p_i,\nu_i,\mu_i,f_i) \in \Theta^*$, $i = 1,2$ such that $\pi_h(\theta_1) = \pi_h(\theta_2)$. If $p_1 = 0$ or $p_1 = 1$ and $\mu_2 \neq \nu_2$, then identifiability follows from theorem 2.14 (i) and (ii), respectively. If $p_1 = 1$ and $\mu_2 = \nu_2$, though, then

$$\hat{f}_1(t) = (1-p_2)\hat{f}_0(t) + p_2\hat{f}_2(t) \quad , \quad t\in\mathbb{R}\ , \tag{2.4.16}$$

cf. (2.4.11), and particularly $f_2 \in \mathcal{E}^{f_0,1}$ due to the definition of Θ^*, leading to

$$\lim_{t\to\infty} \hat{f}_1(t)/\hat{f}_0(t) = 1-p_2\ .$$

Since $f_1 \in \mathcal{E}^{f_0}$ we conclude $p_2 = 1$ and $f_1 \in \mathcal{E}^{f_0,1}$, so that identifiability now follows from theorem 2.14 (iii).

Hence, let $p_1, p_2 \in (0,1)$ from now on. If $\mu_1 = \nu_1$ as well as $\mu_2 = \nu_2$, then the first moment equation

$$(1-p_1)\nu_1 + p_1\mu_1 = (1-p_2)\nu_2 + p_2\mu_2 \ , \tag{2.4.17}$$

cf. (2.4.3), leads to $\mu_1 = \nu_1 = \mu_2 = \nu_2$ and thus

$$(1-p_1)\hat{f}_0(t) + p_1\hat{f}_1(t) = (1-p_2)\hat{f}_0(t) + p_2\hat{f}_2(t) \quad , \quad t \in \mathbb{R} \ . \tag{2.4.18}$$

As in this case $f_1, f_2 \in \mathcal{E}^{f_0,1}$, dividing (2.4.18) by $\hat{f}_0(t)$ and taking limits in t gives $p_1 = p_2$ and therewith moreover $f_1 = f_2$. As a result, we w.l.o.g. henceforth assume $\mu_1 \neq \nu_1$ in addition. By lemma 2.13,

$$\hat{f}_1(t)p_1\sin((\mu_1-\mu_2)t) = \hat{f}_0(t)[(1-p_1)\sin((\mu_2-\nu_1)t)-(1-p_2)\sin((\mu_2-\nu_2)t)] \tag{2.4.19}$$

for all $t \in \mathbb{R}$. On the one hand, if $f_1 \in \mathcal{E}^{f_0,2}$, then (2.4.19) yields

$$\lim_{t\to\infty} p_1 \sin((\mu_1 - \mu_2)t) = 0 \ .$$

As $p_1 \neq 0$, from corollary 2.9 it follows $\mu_1 = \mu_2$ and hence $\theta_1 = \theta_2$, using lemma 2.15. On the other hand, if $f_1 \in \mathcal{E}^{f_0,1}$, then (2.4.19) leads to

$$\lim_{t\to\infty} (1-p_1)\sin((\mu_2 - \nu_1)t) - (1-p_2)\sin((\mu_2 - \nu_2)t) = 0 \ .$$

Note that $\mu_2 \neq \nu_1$ and $\mu_2 \neq \nu_2$, since if one equality was true, then corollary 2.9 would give the validity of both, yielding $\mu_1 = \nu_1$ by (2.4.17). So, by corollary 2.10 we obtain $p_1 = p_2$ and $\nu_1 = \nu_2$ and thus $\mu_1 = \mu_2$ and $f_1 = f_2$ using (2.4.17) and (2.4.15). □

Taking a closer look at the second half of the proof above we even learn the following. Provided that $p_1, p_2 \in (0,1)$ and $\mu_1 \neq \nu_1$, the assumption $f_2 \in \mathcal{E}^{f_0}$ can be omitted. As a consequence we immediately obtain the following result, which will actually turn out to be highly valuable when estimating the model parameters in the location parameter extended setup.

Theorem 2.17. *Provided that $\theta^* \in (0,1) \times \mathcal{R}_2 \times \mathcal{E}_3^{f_0}$, mixture $\pi_h(\theta^*)$ is identifiable over the parameter space $(0,1) \times \mathbb{R}^2 \times \mathcal{E}_3$.*

Further, it is worth remembering that the mixture pdf h is a generalization of mixture g, cf. (2.2.1), so that from theorems 2.16 and 2.17 we straightforward derive the following results.

Corollary 2.18. *(i) In the original setup, see section 2.2, the parameter space*

$$\Big([0,1] \times \mathbb{R} \times \mathcal{E}_3^{f_0}\Big) \backslash \big\{(p,\mu,f) \mid \mu = 0 \ , \ f \in \mathcal{E}^{f_0,2}\big\}$$

is identifiable.

(ii) Provided that $\theta^* \in (0,1) \times \mathcal{R}_1 \times \mathcal{E}_3^{f_0}$, mixture $\pi_g(\theta^*)$ is identifiable over the parameter space $(0,1) \times \mathbb{R} \times \mathcal{E}_3$.

2.5. A scale parameter extension

In this section we want to study a scale parameter extended setup of mixture (2.2.1), i. e. we introduce the additional scale parameter σ to the known component pdf f_0. Hence, in the following,

$$\Theta = [0,1] \times \mathbb{R}_> \times \mathbb{R} \times \mathcal{E} \ ,$$

where $\mathbb{R}_> := \{x \in \mathbb{R} \mid x > 0\}$, and let

$$s(x;\theta) = (1-p) f_0(x/\sigma)/\sigma + p f(x-\mu) \tag{2.5.1}$$

with $\theta = (p, \sigma, \mu, f) \in \Theta$. In this context the pdf f_0 can w.l.o.g. be assumed to have a second order moment equal to one, provided that $f_0 \in \mathcal{E}_2$. Thus, each pdf

$$x \mapsto f_0(x/\sigma)/\sigma \quad , \quad \sigma \in \mathbb{R}_>$$

has a second order moment equal to σ^2. In fact,

$$\int_{\mathbb{R}} x^2 f_0(x/\sigma)/\sigma \, dx = \sigma^2 \int_{\mathbb{R}} y^2 f_0(y) \, dy = \sigma^2 \ , \tag{2.5.2}$$

using the substitution $y = x/\sigma$.

Lemma 2.19. *Let $\theta_1, \theta_2 \in [0,1] \times \mathbb{R}_> \times \mathcal{R}_1 \times \mathcal{E}_3$, $\theta_i = (p_i, \sigma_i, \mu_i, f_i)$, $i = 1, 2$. If $p_1 \in (0,1)$, $\mu_1 = \mu_2$, and $\pi_s(\theta_1) = \pi_s(\theta_2)$, then $\theta_1 = \theta_2$.*

Proof. Regarding the first three moments of mixture (2.5.1), the almost sure identity of $s(x;\theta_1)$ and $s(x;\theta_2)$, using $\mu_1 = \mu_2$ and (2.5.2), leads to

$$p_1\mu_1 = p_2\mu_1 \ , \tag{2.5.3}$$
$$(1-p_1)\sigma_1^2 + p_1(\xi_1 + \mu_1^2) = (1-p_2)\sigma_2^2 + p_2(\xi_2 + \mu_1^2) \ , \text{ and} \tag{2.5.4}$$
$$p_1(3\xi_1\mu_1 + \mu_1^3) = p_2(3\xi_2\mu_1 + \mu_1^3) \ . \tag{2.5.5}$$

As usual, ξ_1 and ξ_2 respectively denote the second order moments of f_1 and f_2. As $\mu_1 \neq 0$, (2.5.3) gives $p_1 = p_2$. Then, (2.5.5) and $p_1\mu_1 \neq 0$ lead to $\xi_1 = \xi_2$, yielding $\sigma_1 = \sigma_2$ by (2.5.4) as $p_1 \neq 1$. Finally, through

$$\begin{aligned} f_1(x) &= \frac{1}{p_1} s(x+\mu_1;\theta_1) - \frac{(1-p_1) f_0((x+\mu_1)/\sigma_1)}{\sigma_1 p_1} \\ &\overset{a.s.}{=} \frac{1}{p_2} s(x+\mu_2;\theta_2) - \frac{(1-p_2) f_0((x+\mu_2)/\sigma_2)}{\sigma_2 p_2} = f_2(x) \end{aligned}$$

we obtain $\theta_1 = \theta_2$. □

In order to derive identifiability results for (2.5.1) we will again make use of the mixture's Fourier transform $\hat{s}$ given by

$$\hat{s}(t; p, \sigma, \mu, f) = (1-p)\hat{f}_0(\sigma t) + p\exp(i\mu t)\hat{f}(t) \quad , \quad t \in \mathbb{R}$$

with

$$\begin{aligned} \operatorname{Re}\hat{s}(t; p, \sigma, \mu, f) &= (1-p)\hat{f}_0(\sigma t) + p\cos(\mu t)\hat{f}(t) \ , \\ \operatorname{Im}\hat{s}(t; p, \sigma, \mu, f) &= p\sin(\mu t)\hat{f}(t) \ , \end{aligned} \tag{2.5.6}$$

where $\hat{f}_0$ and $\hat{f}$ denote the Fourier transforms of f_0 and f, respectively.

Condition C3. For large $t \in \mathbb{R}$ it holds $\hat{f}(t) \neq 0$ and for all $\sigma \in \mathbb{R}_>$ we have

$$\lim_{t\to\infty} \frac{\hat{f}_0(\sigma t)}{\hat{f}(t)} = 0 \ .$$

Condition C4. For large $t \in \mathbb{R}$ it holds $\hat{f}_0(t) \neq 0$, for all $\sigma \in \mathbb{R}_>$

$$\lim_{t\to\infty} \frac{\hat{f}(t)}{\hat{f}_0(\sigma t)} = 0 \ ,$$

and each pair of differing scale parameters can be ordered such that

$$\lim_{t\to\infty} \frac{\hat{f}_0(\sigma_1 t)}{\hat{f}_0(\sigma_2 t)} = 0 \ .$$

Referring to section 2.4, cf. page 13 in particular, let us conformably introduce the sets $\mathcal{E}^{f_0,3}$, $\mathcal{E}^{f_0,4}$, $\tilde{\mathcal{E}}^{f_0} := \mathcal{E}^{f_0,3} \cup \mathcal{E}^{f_0,4}$, and $\tilde{\mathcal{E}}^{f_0}_3$.

Remark. The conditions C3 and C4 impose rather strong constraints since now even the choice of the scale parameter σ must not influence the convergence behavior of the component's Fourier transform ratio. Condition C4 moreover claims σ to influence the convergence rate of the known component's Fourier transform significantly, though. This holds true, for instance, whenever $\hat{f}_0(t)$ decreases at exponential rate.

Example 2.20. Let us consider densities φ and c, respectively denoting the standard Gaussian and the standard Cauchy pdf. The corresponding Fourier transforms are given by

$$\hat{\varphi}(t) = \exp(-t^2/2) \quad , \quad \hat{c}(t) = \exp(-|t|) \quad , \quad t \in \mathbb{R} \ .$$

For all $\sigma > 0$,

$$\lim_{t\to\infty} \frac{\hat{\varphi}(t)}{\hat{c}(\sigma t)} = \lim_{t\to\infty} \exp(-t^2/2 + \sigma|t|) = \lim_{t\to\infty} \exp(-t(t/2-\sigma)) = 0 \ .$$

Considering $\sigma_1, \sigma_2 > 0$ such that $\sigma_1 > \sigma_2$,

$$\lim_{t\to\infty} \frac{\hat{c}(\sigma_1 t)}{\hat{c}(\sigma_2 t)} = \lim_{t\to\infty} \exp(-(\sigma_1 - \sigma_2)|t|) = 0 ,$$

so that φ and c in fact meet regularity condition C4.

Theorem 2.21. *Provided that* $\theta^* \in (0,1) \times \mathbb{R}_> \times \mathcal{R}_1 \times \tilde{\mathcal{E}}_3^{f_0}$, *mixture* $\pi_s(\theta^*)$ *is identifiable over the parameter space* $[0,1] \times \mathbb{R}_> \times \mathbb{R} \times \mathcal{E}_3$.

Proof. Let $\theta_i = (p_i, \sigma_i, \mu_i, f_i)$, $i = 1, 2$, where $\theta_1 \in (0,1) \times \mathbb{R}_> \times \mathcal{R}_1 \times \tilde{\mathcal{E}}_3^{f_0}$ and $\theta_2 \in [0,1] \times \mathbb{R}_> \times \mathbb{R} \times \mathcal{E}_3$ such that $\pi_s(\theta_1) = \pi_s(\theta_2)$. By (2.5.6), for all $t \in \mathbb{R}$,

$$\begin{aligned}(1-p_1)\hat{f}_0(\sigma_1 t) - (1-p_2)\hat{f}_0(\sigma_2 t) + p_1 \cos(\mu_1 t)\hat{f}_1(t) &= p_2 \cos(\mu_2 t)\hat{f}_2(t) , \\ p_1 \sin(\mu_1 t)\hat{f}_1(t) &= p_2 \sin(\mu_2 t)\hat{f}_2(t) .\end{aligned}$$

Multiplying these equations by $\sin(\mu_2 t)$ and $\cos(\mu_2 t)$, respectively, yields

$$\begin{aligned}&((1-p_1)\hat{f}_0(\sigma_1 t) - (1-p_2)\hat{f}_0(\sigma_2 t)) \sin(\mu_2 t) \\ &= p_1 \hat{f}_1(t)(\sin(\mu_1 t)\cos(\mu_2 t) - \sin(\mu_2 t)\cos(\mu_2 t)) \\ &= p_1 \hat{f}_1(t)(\sin(\mu_1 t)\cos(-\mu_2 t) + \sin(-\mu_2 t)\cos(\mu_2 t)) \\ &= p_1 \hat{f}_1(t) \sin((\mu_1 - \mu_2)t) ,\end{aligned}$$

that is, for all $t \in \mathbb{R}$,

$$((1-p_1)\hat{f}_0(\sigma_1 t) - (1-p_2)\hat{f}_0(\sigma_2 t)) \sin(\mu_2 t) = p_1 \hat{f}_1(t) \sin((\mu_1 - \mu_2)t) . \tag{2.5.7}$$

Provided that condition C3 holds, dividing (2.5.7) by $\hat{f}_1(t)$ and taking limits in t gives

$$\lim_{t\to\infty} p_1 \sin((\mu_1 - \mu_2)t) = 0 .$$

As $p_1 > 0$, from corollary 2.9 it follows $\mu_1 = \mu_2$, so that by lemma 2.19 we obtain $\theta_1 = \theta_2$.

Otherwise, provided that condition C4 holds, we have to handle several cases differently. First, let us assume that $\sigma_1 \neq \sigma_2$. Then either (i) $\lim_t \hat{f}_0(\sigma_1 t)/\hat{f}_0(\sigma_2 t) = 0$ or (ii) $\lim_t \hat{f}_0(\sigma_2 t)/\hat{f}_0(\sigma_1 t) = 0$. In (ii), dividing (2.5.7) by $\hat{f}_0(\sigma_1 t)$ and taking limits in t gives

$$\lim_{t\to\infty} (1-p_1) \sin(\mu_2 t) = 0 ,$$

what promptly causes $\mu_2 = 0$ since $p_1 < 1$. Therewith, the first moment equation

$$p_1 \mu_1 = p_2 \mu_2 \tag{2.5.8}$$

yields $p_1\mu_1 = 0$, contradicting $p_1 > 0$, $\mu_1 \neq 0$. In (i), we in a similar way obtain

$$\lim_{t\to\infty} (1 - p_2) \sin(\mu_2 t) = 0 \ .$$

Considering $p_2 < 1$ we deduce a contradiction just as before. If we have $p_2 = 1$, though, then (2.5.8) gives $\mu_2 = p_1\mu_1$, so that by the third moment equation

$$p_1(3\xi_1\mu_1 + \mu_1^3) = p_2(3\xi_2\mu_2 + \mu_2^3)$$

we obtain

$$p_1(3\xi_1\mu_1 + \mu_1^3) = (3\xi_2 p_1\mu_1 + (p_1\mu_1)^3) = p_1(3\xi_1\mu_1 + \mu_1^3) + p_1\mu_1^3(p_1^2 - 1) \ ,$$

yielding

$$p_1\mu_1^3(p_1^2 - 1) = 0 \ .$$

This again leads to a contradiction according to the suppositions $p_1 \in (0, 1)$ and $\mu_1 \neq 0$. Therefore, finally, let $\sigma_1 = \sigma_2$. Then, dividing (2.5.7) by $\hat{f}_0(\sigma_1 t)$ and taking limits in t brings

$$\lim_{t\to\infty} (p_2 - p_1) \sin(\mu_2 t) = 0 \ . \tag{2.5.9}$$

As we have seen before, it is $\mu_2 \neq 0$, so that from corollary 2.9 it follows $p_2 = p_1 \neq 0$, again yielding $\mu_1 = \mu_2$ by (2.5.8). Lemma 2.19 concludes the proof. □

3. Estimation

We consider a real random variable X, representing any observable occurrence, which follows a probability distribution P having Lebesgue density h, cf. (2.4.1). Based on the observation of n independent realizations of X, in this section our goal is the derivation of an estimator $\hat{\theta}_n : \mathbb{R}^n \to [0,1] \times \mathbb{R}^2 \times \mathcal{E}$ for the unknown semiparametric model parameter θ. The way of deriving this estimator $\hat{\theta}_n$ will be strongly based on the symmetry of the unknown component pdf f as well as the identifiability results obtained in section 2.4. Properties of the estimator are worked for thereafter.

3.1. Estimating the model parameter

Let $X_1, X_2, \ldots, X_n$, $n \in \mathbb{N}$ be independent copies of X and let H denote the cdf corresponding to the distribution of X, i. e. if $\theta_\star = (p_\star, \nu_\star, \mu_\star, f_\star)$ denotes the *true* model parameter, then $x \mapsto H(x)$, $x \in \mathbb{R}$ is given by

$$
\begin{aligned}
H(x) = \int_{-\infty}^{x} h(y;\theta_\star)\, dy &= (1-p_\star)\int_{-\infty}^{x} f_0(y-\nu_\star)\, dy + p_\star \int_{-\infty}^{x} f_\star(y-\mu_\star)\, dy \\
&= (1-p_\star)\int_{-\infty}^{x-\nu_\star} f_0(z)\, dz + p_\star \int_{-\infty}^{x-\mu_\star} f_\star(z)\, dz \\
&= (1-p_\star) F_0(x-\nu_\star) + p_\star F(x-\mu_\star)\ ,
\end{aligned}
$$

at what the antiderivatives F_0 and F respectively correspond to f_0 and $f_\star$. Provided that $p_\star > 0$ we in particular have

$$
F(x) = H(x+\mu_\star)/p_\star - (1-p_\star)F_0(x-\nu_\star+\mu_\star)/p_\star\ . \tag{3.1.1}
$$

In the following we denote by $\vartheta = (p, \nu, \mu) \in (0,1] \times \mathbb{R}^2$ the Euclidean part of the model parameter θ. Motivated by (3.1.1) let us consider the functions

$$
\begin{aligned}
D_1(x;\vartheta) &= H(x+\mu)/p - (1-p)F_0(x-\nu+\mu)/p\ , \\
D_2(x;\vartheta) &= 1 - D_1(-x;\vartheta) = 1 - H(-x+\mu)/p + (1-p)F_0(-x-\nu+\mu)/p\ ,
\end{aligned} \tag{3.1.2}
$$

as well as

$$
D(x;\vartheta) = D_1(x;\vartheta) - D_2(x;\vartheta) \quad , \quad x \in \mathbb{R}\ . \tag{3.1.3}
$$

For fixed cdfs H and F_0 and any given parameter ϑ the function D can be interpreted as an indicator for the symmetry of the probability distribution induced by the cdf D_1. In fact, if the derivative of D_1 is symmetric about zero, then for all $x \in \mathbb{R}$ it follows $D_1(x) = 1 - D_1(-x)$, yielding $D_1(x;\vartheta) = D_2(x;\vartheta)$ and thus $D(x;\vartheta) = 0$.

Being measurable and bounded, D is square integrable with respect to any finite measure μ, say, what gives reason for considering its $L^2(\mu)$-norm

$$\|D(\cdot;\vartheta)\|_{\mu,2} = \int_{\mathbb{R}} |D(x;\vartheta)|^2 \, \mu(dx) \; . \tag{3.1.4}$$

In particular, Bordes and Vandekerkhove [3] propose to rate D by means of the $L^2(P)$-norm, leading to

$$d(\vartheta) := \int_{\mathbb{R}} |D(x;\vartheta)|^2 \, P(dx) \; , \tag{3.1.5}$$

which will in fact prove to be reasonable from a computational point of view in the following.

Let $\vartheta_\star = (p_\star, \nu_\star, \mu_\star)$ denote the Euclidean part of $\theta_\star$. As we will see now, in case of identifiability the function d can in fact be used to signalize the "correctness" of a considered parameter ϑ. For convenience, let again

$$\mathcal{R}_2 = \mathbb{R}^2 \backslash \{(x,x) \mid x \in \mathbb{R}\} \; .$$

Theorem 3.1. *If $f_0 \in \mathcal{E}_3$, H is strictly increasing on $\mathbb{R}$, and if $\theta_\star \in (0,1) \times \mathcal{R}_2 \times \mathcal{E}_3^{f_0}$, then*

$$d : (0,1) \times \mathcal{R}_2 \to [0,\infty) \; ,$$

defined as in (3.1.5), is a **discrepancy function** *with respect to $\vartheta_\star$, i. e. for all $\vartheta \in (0,1) \times \mathcal{R}_2$ it holds $d(\vartheta) \geq 0$ as well as $d(\vartheta) = 0$ if and only if $\vartheta = \vartheta_\star$.*

Proof. Obviously, $d(\vartheta) \geq 0$ for all ϑ in the domain of d. Therefore, let $\vartheta \in (0,1) \times \mathcal{R}_2$ be fixed such that $d(\vartheta) = 0$. As

$$d(\vartheta) = \int_{\mathbb{R}} |D(x;\vartheta)|^2 \, P(dx) = \int_{\mathbb{R}} |D(x;\vartheta)|^2 \, h(x;\theta_\star) \, dx$$

and since H is strictly increasing, we have $D(x;\vartheta) \cdot h(x;\theta_\star) = 0$ and $h(x;\theta_\star) > 0$ for almost all $x \in \mathbb{R}$ with respect to the Lebesgue measure λ. As therefore

$$\begin{aligned} &\lambda(\{x \in \mathbb{R} \mid D(x,\vartheta) \neq 0\}) \\ &\leq \lambda(\{x \in \mathbb{R} \mid D(x;\vartheta) \neq 0, \; h(x;\theta_\star) > 0\}) + \lambda(\{x \in \mathbb{R} \mid h(x;\theta_\star) = 0\}) = 0 \; , \end{aligned}$$

it follows

$$D(x;\vartheta) = 0 \tag{3.1.6}$$

for almost all $x \in \mathbb{R}$ with respect to λ. Finally, since $x \mapsto D(x, \vartheta)$ is continuous on $\mathbb{R}$, (3.1.6) actually holds true on the entire real line, giving

$$D_1(x;\vartheta) = D_2(x;\vartheta) = 1 - D_1(-x;\vartheta) \quad , \quad x \in \mathbb{R} \; . \tag{3.1.7}$$

In general, D_1 need not be a distribution function. Nevertheless, the Fourier transform ϕ_{D_1} exists since D_1 is of bounded variation, where in particular ϕ_{D_1} is a real function because of the symmetry (3.1.7). Further, as $f_0, f_\star \in \mathcal{E}_3$, D_1 has a third order moment, too, so that we can proceed with similar arguments as in the proof of theorem 2.16, i. e. from the equality

$$\pi(p, \nu, \mu, D_1) = \pi(p_\star, \nu_\star, \mu_\star, F_\star)$$

it finally follows that $\vartheta = \vartheta_\star$. □

Provided that $\theta_\star \in (0,1) \times \mathcal{R}_2 \times \mathcal{E}_3^{f_0}$, by theorem 3.1 the true Euclidean parameter $\vartheta_\star$ is the unique minimizer of the discrepancy function d on $(0,1) \times \mathcal{R}_2$. Even if $f_\star$ satisfies neither condition C1 nor C2, $\vartheta_\star$ keeps being a minimizer of d, no longer necessarily a unique one, though.

As the true distribution of X and hence cdf H and particularly the function D are assumed to be unknown, we cannot determine $\vartheta_\star$ through a deterministic minimization routine. Nevertheless, d can be used to derive a reasonable estimator $\hat{\vartheta}_n : \mathbb{R}^n \to (0,1) \times \mathbb{R}^2$ for the Euclidean parameter based on the independent copies $X_1, \ldots, X_n$. For this, we replace P and D in (3.1.5) by adequate estimators. A natural estimator for P, of course, is its empirical distribution $\hat{P}_n$, that is the sum of n Dirac measures with mass $1/n$ at each value of the randomly observed n-sample,

$$\hat{P}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i} \; . \tag{3.1.8}$$

The measure $\hat{P}_n$ is the uniquely identified probability measure on $\mathbb{R}$ corresponding to the empirical cdf $\hat{H}_n$ of X,

$$\hat{H}_n(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(-\infty, y]}(X_i) \quad , \quad y \in \mathbb{R} \; .$$

Since we need to find a minimizer of d or, more precisely, of an estimate $\hat{d}$ of d and since therefore we need to apply some optimization routine to $\hat{d}$, the function D has to be replaced by a differentiable estimator $\tilde{D}_n$, say. This is why we substitute H in (3.1.2) by the smoothed version

$$\tilde{H}_n(y) = \int_{-\infty}^{y} \tilde{h}_n(x)\, dx \quad , \quad y \in \mathbb{R} \; , \tag{3.1.9}$$

where $\tilde{h}_n$ constitutes some kernel density estimator for pdf $x \mapsto h(x;\theta_\star)$, giving

$$\tilde{D}_n(x;\vartheta) = \frac{1}{p}\Big(\tilde{H}_n(x+\mu) - (1-p)F_0(x-\nu+\mu) \\ + \tilde{H}_n(-x+\mu) - (1-p)F_0(-x-\nu+\mu)\Big) - 1 \,. \tag{3.1.10}$$

Thus, we finally come to the empirical discrepancy function

$$\hat{d}_n(\vartheta) = \int_{\mathbb{R}} |\tilde{D}_n(x;\vartheta)|^2 \, \hat{P}_n(dx) \\ = \frac{1}{n}\sum_{i=1}^{n} |\tilde{D}_n(X_i;\vartheta)|^2 \quad , \quad n \in \mathbb{N} \,. \tag{3.1.11}$$

As $\hat{d}_n$ is continuous on $(0,1] \times \mathbb{R}^2$, restricted to any compact subset Θ it attains its minimum value, so that a reasonable estimate of $\vartheta_\star$ is given by a corresponding (but in general not unique) minimizer,

$$\hat{\vartheta}_n = (\hat{p}_n, \hat{\nu}_n, \hat{\mu}_n) \in \arg\min_{\vartheta\in\Theta} \hat{d}_n(\vartheta) \,. \tag{3.1.12}$$

Thereupon, of course, a natural estimator for the unknown pdf $f_\star$ is induced by the derivative of D_1, replacing ϑ by $\hat{\vartheta}_n$ and h by the kernel density estimator $\tilde{h}_n$, yielding

$$\hat{f}_n(x) = \tilde{h}_n(x+\hat{\mu}_n)/\hat{p}_n - (1-\hat{p}_n)f_0(x-\hat{\nu}_n+\hat{\mu}_n)/\hat{p}_n \quad , \quad x \in \mathbb{R} \,,$$

what completes the semiparametric estimator

$$\hat{\theta}_n = (\hat{\vartheta}_n, \hat{f}_n) = (\hat{p}_n, \hat{\nu}_n, \hat{\mu}_n, \hat{f}_n) \,.$$

3.2. Empirical processes

The estimator $\hat{\vartheta}_n$ in (3.1.12) is said to be a *M-estimator* as it *minimizes* some given criterion function. This concept is common, e. g. in maximum likelihood estimation. In our case, however, the applied criterion function $\hat{d}_n$ is stochastic itself, containing the nonparametric cdf estimator $\tilde{H}_n$. This is why we first have to recall some theory of empirical processes before we are capable of studying further conditions of $\hat{\vartheta}_n$. For this, in the following we consider P-integrable functions f, say, and take a closer look at their expectation under both the empirical and the true distribution of X,

$$\mathbf{E}_{\hat{P}_n}[f] = \int f \, d\hat{P}_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i) \quad , \quad n \in \mathbb{N} \,, \\ \mathbf{E}_P[f] = \int f \, dP \,.$$

For convenience, let

$$E(f) := \mathbf{E}_P[f] \quad , \quad \hat{E}_n(f) := \mathbf{E}_{\hat{P}_n}[f] \quad , \quad n \in \mathbb{N} .$$

3.2.1. Glivenko-Cantelli theorem

Definition 3.2 (Glivenko-Cantelli). *A class $\mathcal{F}$ of P-integrable functions is said to be P-Glivenko-Cantelli if*

$$\|\hat{E}_n - E\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |\hat{E}_n(f) - E(f)| \longrightarrow 0 \quad a.s. \tag{3.2.1}$$

as n tends to infinity.

On a Glivenko-Cantelli class of functions the empirical expectation converges uniformly to the real expectation as the number of observations tends to infinity. At this one could be reminded of the empirical cdf $\hat{H}_n$ converging uniformly to the true cdf H of X. In fact, choosing $\mathcal{F} = \{\mathbf{1}_{(-\infty,y]} \mid y \in \mathbb{R}\}$, then, for all $y \in \mathbb{R}$,

$$\begin{aligned} \hat{E}_n(\mathbf{1}_{(-\infty,y]}) &= \frac{1}{n} \sum_{i=1}^{n} 1_{(-\infty,y]}(X_i) = \hat{H}_n(y; X_1, \ldots, X_n) , \\ E(\mathbf{1}_{(-\infty,y]}) &= \int \mathbf{1}_{(-\infty,y]} \, dP = P((-\infty, y]) = H(y) , \end{aligned} \tag{3.2.2}$$

so that the convergence (3.2.1) follows from the classical *Glivenko-Cantelli-Lemma*[1]. As a result, the set of indicator functions on the cells $(-\infty, y]$, $y \in \mathbb{R}$ is Glivenko-Cantelli.

We will now give a sufficient criterion for a class of functions being Glivenko-Cantelli. Let $\mathcal{F}$ be any set of P-integrable functions. Given two functions $l, u \in L^k(P)$, $l \leq u$, $k \in \mathbb{N}$, we define the bracket $[l, u]$ as the subset of functions $f \in \mathcal{F}$ satisfying $l \leq f \leq u$. A bracket $[l, u]$ having

$$\int (u - l)^k \, dP < \varepsilon^k$$

for some $\varepsilon > 0$, $k \in \mathbb{N}$ is referred to as an (ε, k)-*bracket*. Finally, the *bracketing number*

$$N(\varepsilon, k, \mathcal{F})$$

is defined as the minimum number of (ε, k)-brackets needed to cover the set $\mathcal{F}$.

Theorem 3.3 (Glivenko-Cantelli). *Every class $\mathcal{F}$ of P-integrable functions satisfying*

$$N(\varepsilon, 1, \mathcal{F}) < \infty \quad , \quad \varepsilon > 0$$

[1]Cf. van der Vaart [25], p. 266, theorem 19.1.

is P-Glivenko-Cantelli.

The proof is quite similar to the classical Glivenko-Cantelli-lemma. It can e. g. be found in van der Vaart and Wellner [26], p. 122, theorem 2.4.1.

Corollary 3.4. *Let $\mathcal{F} = \{f_\theta \mid \theta \in \Theta\}$ be a parametric Family indexed by a compact metric set Θ such that for all $x \in \mathbb{R}$ the mapping $\theta \mapsto f_\theta(x)$ is continuous and such that there exists a P-integrable envelope for the f_θ. Then, $\mathcal{F}$ is P-Glivenko-Cantelli.*

Proof. For $\theta \in \Theta$ let us denote by B^r_θ the open ball in Θ with center θ and radius r. Let us consider an arbitrary fixed $\theta_0 \in \Theta$. Defining pointwise

$$f_{B^r_{\theta_0}} := \inf_{\theta \in B^r_{\theta_0}} f_\theta \quad , \quad f^{B^r_{\theta_0}} := \sup_{\theta \in B^r_{\theta_0}} f_\theta$$

yields a bracket $Br^r_{\theta_0} := [f_{B^r_{\theta_0}}, f^{B^r_{\theta_0}}]$ such that for all $\theta \in B^r_{\theta_0}$ it is $f_\theta \in Br^r_{\theta_0}$. If $(r_n)_{n\in\mathbb{N}}$ is a sequence of decreasing radii, then due to the assumed continuity we have

$$f^{B^{r_n}_{\theta_0}}(x) - f_{B^{r_n}_{\theta_0}}(x) \downarrow 0 \quad , \quad x \in \mathbb{R}\, ,\ n \to \infty\, .$$

Thus and since there is a P-integrable envelope, by Lebesgue's dominated convergence theorem[2] this convergence holds true in $L_1(P)$, too, giving

$$\int (f^{B^{r_n}_{\theta_0}} - f_{B^{r_n}_{\theta_0}})dP \downarrow 0 \quad , \quad n \to \infty\, .$$

So, for all $\varepsilon > 0$ there exists a radius r_ε and an open ball $B^{r_\varepsilon}_{\theta_0}$ such that $Br^{r_\varepsilon}_{\theta_0}$ is an $(\varepsilon, 1)$-bracket.

For all $\theta \in \Theta$ let B^ε_θ denote the open ball around θ inducing the $(\varepsilon, 1)$-bracket Br^ε_θ. Trivially, Θ can be covered by the union $\bigcup_{\theta\in\Theta} B^\varepsilon_\theta$. By the compactness of Θ, though, there exists a finite subset $\tilde{\Theta} \subset \Theta$ such that

$$\Theta \subset \bigcup_{\theta\in\tilde{\Theta}} B^\varepsilon_\theta\, .$$

Hence, $\bigcup_{\theta\in\tilde{\Theta}} Br^\varepsilon_\theta$ is a finite cover of $\mathcal{F}$. It follows

$$N(\varepsilon, 1, \mathcal{F}) \leq \#(\tilde{\Theta}) < \infty\, .$$

As $\varepsilon > 0$ was arbitrarily chosen, the proof is concluded. □

[2] Cf. Bauer [1], p. 95, theorem 15.6.

3.2.2. Donsker's theorem and the functional Delta-method

In addition to the uniform convergence we are also interested in the asymptotic distribution of $\hat{E}_n$. Let $\mathcal{F}$ be some class of square P-integrable functions. Then, for all $n \in \mathbb{N}$, defining

$$\mathbb{E}_n(f) := \sqrt{n}(\hat{E}_n(f) - E(f)) \quad , \quad f \in \mathcal{F} ,$$

we regard

$$\mathbb{E}_n := (\mathbb{E}_n(f))_{f \in \mathcal{F}}$$

as a stochastic process indexed by the class $\mathcal{F}$. The process $\mathbb{E}_n$ is called the *empirical process* with respect to $\mathcal{F}$ and the probability measure P.

Definition 3.5 (Gaussian process). *Let T be an arbitrary index set. A stochastic process $\mathbb{G} = (\mathbb{G}(t))_{t \in T}$ is said to be Gaussian if any finite linear combination*

$$\sum_{i=1}^{n} \alpha_i \mathbb{G}(t_i) \quad , \quad n \in \mathbb{N} \, , \, \alpha_i \in \mathbb{R} \, , \, t_i \in \mathcal{F} \, , \, i = 1, \ldots, n$$

is normally distributed.

Definition 3.6 (Donsker). *A class $\mathcal{F}$ of square P-integrable functions is said to be P-Donsker if*

$$\mathbb{E}_n \rightsquigarrow \mathbb{G} ,$$

as n tends to infinity[3]*, where $\mathbb{G}$ is a zero mean Gaussian process with covariance functional $\rho : \mathcal{F}^2 \to \mathbb{R}$ given by*

$$\rho(f_i, f_j) = \mathbf{Cov}\left[\mathbb{G}(f_i), \mathbb{G}(f_j)\right] = E(f_i f_j) - E(f_i)E(f_j) \quad , \quad f_i, f_j \in \mathcal{F} .$$

A criterion for a class of functions to be P-Donsker can be given by a bracketing argument, too.

Theorem 3.7 (Donsker). *Every class $\mathcal{F}$ of square P-integrable functions satisfying*

$$\int_0^1 \sqrt{\log N(\varepsilon, 2, \mathcal{F})} \, d\varepsilon < \infty \tag{3.2.3}$$

is P-Donsker.

For a proof of Donsker's theorem cf. van der Vaart [25], p. 270, theorem 19.5, for instance.

[3] Here and in the following we use $\rightsquigarrow$ to express convergence in distribution.

Corollary 3.8. *If P is absolutely continuous with respect to the Lebesgue measure, then the class $\mathcal{F} = \{\mathbf{1}_{(-\infty,y]} \mid y \in \mathbb{R}\}$ is P-Donsker.*

Proof. For all $0 < \varepsilon \leq 1$ there exists $N \in \mathbb{N}$ such that

$$1/N < \varepsilon^2 \quad , \quad N \leq 2/\varepsilon^2 \, .$$

Defining $x_0 := -\infty$, $x_N := +\infty$ as well as x_k, $k = 1, \ldots, N-1$ such that

$$P((-\infty, x_k]) = k/N \, ,$$

which exist because of the absolute continuity of P, then the N brackets

$$[\mathbf{1}_{(-\infty,x_{N-1}]}, \mathbf{1}_{(-\infty,x_N)}] \quad , \quad [\mathbf{1}_{(-\infty,x_{k-1}]}, \mathbf{1}_{(-\infty,x_k]}] \quad , \quad k = 1, \ldots N-1$$

cover the set $\mathcal{F}$ and have $L_2(P)$-norm smaller than ε^2. In fact, e. g.

$$\int (\mathbf{1}_{(-\infty,x_k]} - \mathbf{1}_{(-\infty,x_{k-1}]})^2 \, dP = P((x_{k-1}, x_k]) = 1/N < \varepsilon^2 \, .$$

Thus, for all $0 < \varepsilon \leq 1$ we obtain $N(\varepsilon, 2, \mathcal{F}) \leq N \leq 2/\varepsilon^2$ and hence

$$\int_0^1 \sqrt{\log N(\varepsilon, 2, \mathcal{F})} \, d\varepsilon \leq \int_0^1 \sqrt{\log(2/\varepsilon^2)} \, d\varepsilon < \infty \, .$$

The finiteness of the integral above can for example be seen by the inequality $\sqrt{\log(2/x^2)} \leq x^{-0.8}$, $0 < x \leq 1$. □

Remark. Corollary 3.8 keeps its validity without requiring the absolute continuity of P, too. Constructing the grid of $(\varepsilon, 2)$-brackets becomes more complicated, though.

Corollary 3.9. *If P is absolutely continuous with respect to the Lebesgue measure, then the sequence of empirical processes*

$$(\sqrt{n}(\hat{F}_n(y) - F(y)))_{y \in \mathbb{R}} \quad , \quad n \in \mathbb{N} \, ,$$

where F and $\hat{F}_n$ denote the cdf and empirical cdf corresponding to P, respectively, converges in distribution to a zero mean Gaussian process $\mathbb{G}$ with autocovariance function $\rho : \mathbb{R}^2 \to \mathbb{R}$ given by

$$\rho(x, y) = \min(F(x), F(y)) - F(x)F(y) \quad , \quad x, y \in \mathbb{R} \, .$$

Proof. Regarding (3.2.2), for all $n \in \mathbb{N}$ and $y \in \mathbb{R}$,

$$\sqrt{n}(\hat{F}_n(y) - F(y)) = \sqrt{n}(\hat{E}_n(\mathbf{1}_{(-\infty,y]}) - E(\mathbf{1}_{(-\infty,y]})) = \mathbb{E}_n(\mathbf{1}_{(-\infty,y]}) \, .$$

As $\{\mathbf{1}_{(-\infty,y} \mid y \in \mathbb{R}\}$ is P-Donsker by corollary 3.8, it follows

$$\mathbb{E}_n \rightsquigarrow \mathbb{G} \quad , \quad n \to \infty \, ,$$

where the autocovariance function of $\mathbb{G}$ is given by

$$\begin{aligned}\rho(x,y) &= E(\mathbf{1}_{(-\infty,x]}\mathbf{1}_{(-\infty,y]}) - E(\mathbf{1}_{(-\infty,x]})E(\mathbf{1}_{(-\infty,y]})\\ &= \int \mathbf{1}_{(-\infty,\min(x,y)]}\, dP - \int \mathbf{1}_{(-\infty,x]}\, dP \int \mathbf{1}_{(-\infty,y]}\, dP\\ &= F(\min(x,y)) - F(x)F(y)\\ &= \min(F(x),F(y)) - F(x)F(y)\ ,\end{aligned}$$

see definition 3.6. □

As one is often interested in the convergence behavior of a random element $\phi(X)$, where the convergence of X is known, we finally need to work for a result concerning the transfer of weak convergence under adequate mappings ϕ.

Definition 3.10 (Hadamard-differentiable)**.** *Let D and E be normed linear spaces. A function $\phi : D \to E$ is said to be Hadamard-differentiable at $d_0 \in D$ if there exists a linear map $\phi'_{d_0} : D \to E$ such that*

$$t_n^{-1}\big(\phi(d_0 + t_n h_n) - \phi(d_0)\big) \longrightarrow \phi'_{d_0}(h) \quad , \quad n \to \infty$$

for all sequences $(t_n)_{n\in\mathbb{N}} \subset \mathbb{R}$, $t_n \to 0$ and $(h_n)_{n\in\mathbb{N}} \subset D$, $h_n \to h \in D$.

Example 3.11. Let us denote by $D(\mathbb{R})$ the *Skorohod Space*, the linear space of bounded *cadlag functions*[4] on $\mathbb{R}$, equipped with the general supremum norm. Further, for $\gamma > 0$, let $BV_\gamma(\mathbb{R}) \subset D(\mathbb{R})$ be the set of cadlag functions of total variation bounded by γ. Then, the map

$$\phi : D(\mathbb{R}) \times BV_\gamma(\mathbb{R}) \to \mathbb{R} \quad , \quad (f,g) \mapsto \int f\, dg$$

is Hadamard-differentiable at all pairs (f,g) such that f is of finite total variation and the derivative of ϕ at (f,g) is given by

$$\phi'_{(f,g)}(h_1,h_2) = \int f\, dh_2 + \int h_1\, dg\ , \tag{3.2.4}$$

see van der Vaart and Wellner [26], p. 382, lemma 3.9.17.

Remark. Note that whenever g is of bounded variation the Riemann-Stieltjes integral $\int f\, dg$ and hence the map ϕ is well-defined.

Corollary 3.12 (Chain rule)**.** *Let D, E, and F be normed linear spaces as well as $\psi : D \to E$ and $\phi : E \to F$ be functions. If ψ is Hadamard-differentiable at $d_0 \in D$ and ϕ is Hadamard-differentiable at $\psi(d_0) \in E$ with derivatives ψ'_{d_0} and $\phi'_{\psi(d_0)}$, respectively, then $\phi \circ \psi : D \to F$ is Hadamard-differentiable at d_0 with derivative*

$$(\phi \circ \psi)'_{d_0} = \phi'_{\psi(d_0)} \circ \psi'_{d_0}\ .$$

[4] A function being right-continuous with existing left limits (continue à droite, limitée à gauche).

Proof. Given arbitrary sequences $(t_n)_{n\in\mathbb{N}} \subset \mathbb{R}$, $t_n \to 0$ and $(h_n)_{n\in\mathbb{N}} \subset D$, $h_n \to h \in D$ we rewrite

$$t_n^{-1}\big(\phi(\psi(d_0 + t_n h_n)) - \phi(\psi(d_0))\big) = t_n^{-1}\big(\phi(\psi(d_0) + t_n k_n)) - \phi(\psi(d_0))\big) ,$$

where $k_n = t_n^{-1}\big(\psi(d_0 + t_n h_n) - \psi(d_0)\big)$. Due to the differentiability of ψ at d_0

$$k_n \longrightarrow \psi'_{d_0}(h) \quad , \quad n \to \infty .$$

Thus, applying the differentiability of ϕ at $\psi(d_0)$ concludes the proof. □

Theorem 3.13 (Delta-method)**.** *Let D and E be normed linear spaces and let $\phi : D \to E$ be Hadamard-differentiable at $d_0 \in D$. Let X_n, $n \in \mathbb{N}$, and X be random elements, X be separable, with values in D such that*

$$r_n(X_n - d_0) \rightsquigarrow X$$

for some sequence $(r_n)_{n\in\mathbb{N}} \subset \mathbb{R}$, $r_n \to \infty$. Then,

$$r_n(\phi(X_n) - \phi(d_0)) \rightsquigarrow \phi'_{d_0}(X) ,$$

where ϕ'_{d_0} is the Hadamard-derivative of ϕ at d_0, cf. definition 3.10. If ϕ'_{d_0} is continuous on the whole of D, then also

$$r_n(\phi(X_n) - \phi(d_0)) = \phi'_{d_0}(r_n(X_n - d_0)) + o_p(1) .$$

For the proof of theorem 3.13 see van der Vaart and Wellner [26], p. 374.

3.2.3. Smooth Estimates

Let $f : \mathbb{R} \to \mathbb{R}$ be any Lebesgue pdf with corresponding cdf F, empirical cdf $\hat{F}_n$, and smoothed empirical cdf $\tilde{F}_n$,

$$\tilde{F}_n(y) = \int_{-\infty}^{y} \tilde{f}_n(x)\, dx \quad , \quad y \in \mathbb{R} ,$$

where $\tilde{f}_n$ is a kernel density estimator for f,

$$\tilde{f}_n(x) = \frac{1}{nb_n} \sum_{i=1}^{n} k\left(\frac{x - X_i}{b_n}\right) \quad , \quad x \in \mathbb{R} ,$$

with kernel function[5] k and bandwidth $(b_n)_{n\in\mathbb{N}}$ such that $b_n \to 0$.

Assumption 3.14. (i) The pdf f is continuously differentiable and both f and its first order derivative, f' say, are bounded.

[5] A continuous Lebesgue density is referred to as a kernel function.

(ii) The kernel k is centered about zero, has a finite second order moment, and integrates to one, i. e.

$$\int x\,k(x)\,dx = 0 \quad , \quad \int x^2\,k(x)\,dx < \infty \quad , \quad \int k(x)\,dx = 1 \ .$$

Theorem 3.15. *If there exists a sequence $(c_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ such that $c_n \to \infty$ and $nb_nc_n/\log(n) \to \infty$ as n tends to infinity, then, under assumption 3.14,*

$$\sqrt{n}\|\tilde{F}_n - \hat{F}_n\|_\infty = O_{a.s.}\Big(\sqrt{b_nc_n|\log b_nc_n|} + \sqrt{\log\log n}\int_{c_n}^{\infty} k(x)\,dx + \sqrt{n}b_n^2\Big) \ .$$

The proof of theorem 3.15 can be found in Shorack and Wellner [22], see corollary 1, p. 766. With regard to this result, in the following we assume that some given sequence $(c_n)_{n\in\mathbb{N}}$ satisfies

$$c_n \longrightarrow \infty \quad , \quad nb_nc_n/\log(n) \longrightarrow \infty \quad , \quad \sqrt{b_nc_n|\log b_nc_n|} \longrightarrow 0 \quad , \quad \text{and}$$

$$\sqrt{\log\log n}\int_{c_n}^{\infty} k(x)\,dx \longrightarrow 0 \ . \tag{3.2.5}$$

Corollary 3.16. *If the bandwidth b_n gives $\sqrt{n}b_n^2 \to 0$, then, under assumption 3.14,*

$$\|\tilde{F}_n - \hat{F}_n\|_\infty = o_{a.s.}(n^{-1/2})$$

and

$$\|\tilde{F}_n - F\|_\infty = O_{a.s.}\big(\sqrt{(\log\log n)/n}\big) \ .$$

Proof. The first assertion is a straight consequence of theorem 3.15.

For all $x \in \mathbb{R}$,

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{(-\infty,x]}(X_i) \ ,$$

where the random variables $\mathbf{1}_{(-\infty,x]}(X_i)$, $i \in \mathbb{N}$ are independent and identically distributed, as this is true for the X_i's, with mean $F(x)$ and variance $F(x)(1-F(x))$.

Hence, by the *law of the iterated logarithm* due to Hartman-Wintner[6],

$$\begin{aligned}\limsup_{n\to\infty} \frac{\sqrt{n}}{\sqrt{\log\log n}} |\hat{F}_n(x) - F(x)| &= \limsup_{n\to\infty} \frac{\sum_{i=1}^n \big(\mathbf{1}_{(-\infty,x]}(X_i) - F(x)\big)}{\sqrt{n \log\log n}} \\ &= \sqrt{2F(x)(1-F(x))} \\ &\le \sqrt{2}\end{aligned}$$

almost surely, giving

$$\|\hat{F}_n - F\|_\infty = O_{a.s.}\big(\sqrt{(\log\log n)/n}\big) .$$

By means of triangle inequality

$$\|\tilde{F}_n - F\|_\infty \le \|\tilde{F}_n - \hat{F}_n\|_\infty + \|\hat{F}_n - F\|_\infty = o_{a.s.}(n^{-1/2}) + O_{a.s.}\big(\sqrt{(\log\log n)/n}\big)$$

and the second assertion follows. □

The next result by Silverman [23] provides the handling of the kernel density estimator $\tilde{f}_n$ and its derivative $\tilde{f}_n'$.

Theorem 3.17. *Under adequate assumptions, cf. Silverman [23], theorems A and C, both*

(i) $\|\tilde{f}_n - f\|_\infty = o_{a.s.}(1)$ *and*

(ii) $\|\tilde{f}_n' - f'\|_\infty = o_{a.s.}(1)$

hold.

Finally, theorem 3.17 (i) can be aggravated as follows.

Theorem 3.18. *Under assumption 3.14 and some further conditions on k and b_n, cf. Giné and Guillou [9], theorem 2.3,*

$$\|\tilde{f}_n - f\|_\infty = O_{a.s.}\big(\sqrt{|\log b_n|/(nb_n)}\big) + O(b_n) .$$

Proof. By Giné and Guillou [9],

$$\|\tilde{f}_n - \mathbf{E}\big[\tilde{f}_n\big]\|_\infty = O_{a.s.}\big(\sqrt{|\log b_n|/(nb_n)}\big) .$$

Moreover, for all $x \in \mathbb{R}$,

$$\begin{aligned}\mathbf{E}\big[\tilde{f}_n(x)\big] &= b_n^{-1} \int k\Big(\frac{x-y}{b_n}\Big) f(y)\, dy = \int k(z) f(x - b_n z)\, dz \\ &= \int k(z)\big(f(x) - b_n z f'(\xi_x^z)\big)\, dz \\ &= f(x) - b_n \int z f'(\xi_x^z) k(z)\, dz\end{aligned}$$

[6] Cf. e. g. Bauer [1], p. 272, theorem 31.1.

by a zero order Taylor expansion of f around x, where ξ_x^z lies in the line segment between x and $x - b_n z$. Hence, as f' is bounded, denoting by $\bar{k}_1$ the finite absolute first order moment of k, for some constant $c > 0$ it follows

$$\sup_{x\in\mathbb{R}} \left|\mathbf{E}\big[\tilde{f}_n(x)\big] - f(x)\right| \le \sup_{x\in\mathbb{R}} b_n \int |z||f'(\xi_x^z)|k(z)\,dz \le c\bar{k}_1 b_n \ ,$$

yielding

$$\|\mathbf{E}\big[\tilde{f}_n(x)\big] - f(x)\|_\infty = O(b_n) \ .$$

We conclude the proof by means of triangle inequality. □

Remark. For a detailed listing of the presuppositions to the kernel function k and the bandwidth b_n required for theorems 3.17 and 3.18 see assumption 3.24 (i)-(iv).

3.3. Consistency

In this section we will show that $\hat{\vartheta}_n$, cf. (3.1.12), is a consistent estimator for the Euclidean model parameter ϑ, that is we verify that $\hat{\vartheta}_n$ converges almost surely to $\vartheta_\star$ as the number n of observations tends to infinity.

For convenience, in the following we denote the true pdf corresponding to the distribution P of X by h, i. e.

$$h(x) = h(x;\vartheta_\star) \quad , \quad x \in \mathbb{R} \ .$$

Provided that h is differentiable we further denote its first order derivative by h'.

Lemma 3.19. *(i) The function $D : \mathbb{R}\times\Theta \to \mathbb{R}$, cf. (3.1.3), is bounded whenever Θ is a compact subset of $(0,1]\times\mathbb{R}^2$.*

(ii) The discrepancy function d, cf. (3.1.5), is continuous on $(0,1]\times\mathbb{R}^2$.

Proof. (i) From the compactness of Θ it follows that there exists $c > 0$ such that $p \ge c$ for all $\vartheta = (p,\nu,\mu) \in \Theta$. Hence, for all $x \in \mathbb{R}$, $\vartheta \in \Theta$,

$$\begin{aligned} |D(x;\vartheta)| &= \big|\big(H(x+\mu) - (1-p)F_0(x-\nu+\mu) \\ &\qquad + H(-x+\mu) - (1-p)F_0(-x-\nu+\mu)\big)/p - 1\big| \\ &\le 4/c + 1 \ . \end{aligned}$$

(ii) The mapping $\vartheta \mapsto D(x;\vartheta)^2$ is continuous on $(0,1]\times\mathbb{R}^2$ for all fixed $x \in \mathbb{R}$.

Let $\vartheta_0 \in (0,1]\times\mathbb{R}^2$ be fixed. In order to prove the continuity of d in ϑ_0 let us consider an arbitrary sequence $(\vartheta_n)_{n\in\mathbb{N}}$ in Θ such that $\vartheta_n \to \vartheta_0$ as n tends to infinity, where Θ is some compact subset of $(0,1]\times\mathbb{R}^2$ with $\vartheta_0 \in \Theta$. Due to the boundedness on Θ, see part (i), there exists a P-integrable envelope for the $D(x;\vartheta_n)^2$,

$n \in \mathbb{N}$, so that by Lebesgue's dominated convergence theorem the function series $(x \mapsto D(x;\vartheta_n)^2)_{n\in\mathbb{N}}$ converges to $x \mapsto D(x;\vartheta_0)^2$ in $L^1(P)$, too, yielding

$$|d(\vartheta_n) - d(\vartheta_0)| \le \int_{\mathbb{R}} |D(x;\vartheta_n)^2 - D(x;\vartheta_0)^2|\, P(dx) \longrightarrow 0 \quad , \quad n \to \infty \ ,$$

what concludes the proof. □

In the following we will give evidence of the empirical discrepancy function $\hat{d}_n$ being a consistent estimator for the real discrepancy d under adequate assumptions, what immediately provides the strong consistency of $d(\hat{\vartheta}_n)$ thereafter.

Assumption 3.20. Let b_n and k respectively denote the bandwidth and the kernel function of the kernel density estimator $\tilde{h}_n$, cf. (3.1.9). Then, the following holds.

(i) The pdf h is continuously differentiable and both h and h' are bounded.

(ii) The kernel function k meets assumption 3.14 (ii).

(iii) There exists a sequence $(c_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ which suffices (3.2.5).

(iv) The bandwidth b_n gives $\sqrt{n}b_n^2 \to 0$, $n \to \infty$.

Recall that assumption 3.20 provides the almost sure convergence of the smooth empirical cdf $\tilde{H}_n$.

Lemma 3.21. *Let $d \in \mathbb{N}$. If $(f_n)_{n\in\mathbb{N}}$, $f_n : \mathbb{R}^d \to \mathbb{R}$ is a function series such that*

$$\lim_{n\to\infty} \|f_n - f\|_\infty = 0$$

and if $g : \mathbb{R} \to \mathbb{R}$ is uniformly continuous on the codomain of f and the f_n, then

$$\lim_{n\to\infty} \|g \circ f_n - g \circ f\|_\infty = 0 \ .$$

Proof. Let $\varepsilon > 0$ be arbitrarily small. Due to the uniform continuity of g there exists $\delta_\varepsilon > 0$ such that, for all $n \in \mathbb{N}$ and $x_1, x_2 \in \mathbb{R}^d$, the implication

$$|f_n(x_1) - f(x_2)| < \delta_\varepsilon \quad \Longrightarrow \quad |g(f_n(x_1)) - g(f(x_2))| < \varepsilon \tag{3.3.1}$$

holds true. Since $(f_n)_{n\in\mathbb{N}}$ converges uniformly in x there moreover exists $n_0 \in \mathbb{N}$ such that

$$|f_n(x) - f(x)| < \delta_\varepsilon \quad , \quad n \ge n_0 \ , \ x \in \mathbb{R}^d \ ,$$

which combined with (3.3.1) yields

$$\sup_{x\in\mathbb{R}^d} |g(f_n(x) - g(f(x))| \le \varepsilon \quad , \quad n \ge n_0 \ .$$

The proof is concluded. □

Remark. Clearly, lemma 3.21 is transferable to the case where $f_n(x)$, $n \in \mathbb{N}$, $x \in \mathbb{R}^d$ is a random variable and the convergence $\|f_n - f\|_\infty \to 0$ holds true no better than almost surely.

Lemma 3.22. *Under assumption 3.20 the following holds.*

(i) If Θ *is a compact subset of* $(0,1] \times \mathbb{R}^2$*, then*

$$\sup_{\vartheta\in\Theta} |\hat{d}_n(\vartheta) - d(\vartheta)| = o_{a.s.}(1) \ .$$

(ii) If further $\vartheta_\star \in \Theta$*, then* $d(\hat{\vartheta}_n)$ *converges almost surely to zero.*

Proof. (i) For all $\vartheta \in \Theta$, $n \in \mathbb{N}$,

$$|\hat{d}_n(\vartheta) - d(\vartheta)| = \Big|\frac{1}{n}\sum_{i=1}^n \tilde{D}_n(X_i;\vartheta)^2 - \int D(x;\vartheta)^2\, P(dx)\Big|$$

$$\leq \frac{1}{n}\sum_{i=1}^n \big|\tilde{D}_n(X_i;\vartheta)^2 - D(X_i;\vartheta)^2\big| + \Big|\frac{1}{n}\sum_{i=1}^n D(X_i,\vartheta)^2 - \int D(x;\vartheta)^2\, P(dx)\Big| \ ,$$

where the last expression can be handled as follows.

For all $\vartheta = (p,\nu,\mu) \in \Theta$,

$$\begin{aligned}
&\sup_{x\in\mathbb{R}} |\tilde{D}_n(x;\vartheta) - D(x;\vartheta)| \\
&= \sup_{x\in\mathbb{R}} |\tilde{H}_n(x+\mu) - H(x+\mu) + \tilde{H}_n(-x+\mu) - H(-x+\mu)|/p \\
&\leq \sup_{x\in\mathbb{R}} |\tilde{H}_n(x+\mu) - H(x+\mu)|/p + \sup_{x\in\mathbb{R}} |\tilde{H}_n(-x+\mu) - H(-x+\mu)|/p \\
&= \tfrac{2}{p} \cdot \sup_{x\in\mathbb{R}} |\tilde{H}_n(x) - H(x)| \\
&\leq c\|\tilde{H}_n - H\|_\infty \ , \qquad (3.3.2)
\end{aligned}$$

where the constant $c > 0$ results from the compactness of Θ. As $c\|\tilde{H}_n - H\|_\infty$ is independent of the choice of ϑ and converges to zero by corollary 3.16,

$$\sup_{x\in\mathbb{R},\vartheta\in\Theta} |\tilde{D}_n(x;\vartheta) - D(x;\vartheta)| = o_{a.s.}(1) \ .$$

Since $D, \tilde{D}_n$, $n \in \mathbb{N}$ are bounded on $\mathbb{R} \times \Theta$, see lemma 3.19, and thus only take values in some compact interval in $\mathbb{R}$, the squaring $x \mapsto x^2$ is uniformly continuous on their joint codomain, so that regarding lemma 3.21 we further obtain

$$\sup_{x\in\mathbb{R},\vartheta\in\Theta} |\tilde{D}_n(x;\vartheta)^2 - D(x;\vartheta)^2| = o_{a.s.}(1) \ .$$

As the parameter space Θ is compact, $\vartheta \mapsto D(x;\vartheta)^2$, $\vartheta \in \Theta$ is continuous for every fixed $x \in \mathbb{R}$, and since there exists a P-integrable envelope for the $D(x;\vartheta)^2$

due to their boundedness on $\mathbb{R} \times \Theta$, by corollary 3.4 the parametric family

$$\mathcal{F}_\Theta := \{x \mapsto D(x;\vartheta)^2 \mid \vartheta \in \Theta\}$$

is P-Glivenko-Cantelli. Therefore,

$$\sup_{f \in \mathcal{F}_\Theta} |\hat{E}_n(f) - E(f)| = o_{a.s.}(1) ,$$

inducing

$$\sup_{\vartheta \in \Theta} \Big| \frac{1}{n} \sum_{i=1}^n D(X_i;\vartheta)^2 - \int D(x;\vartheta)^2 \, P(dx) \Big| = o_{a.s.}(1) .$$

In total, for all $\vartheta \in \Theta$,

$$\begin{aligned} \sup_{\vartheta \in \Theta} |\hat{d}_n(\vartheta) - d(\vartheta)| \le & \sup_{x \in \mathbb{R}, \vartheta \in \Theta} |\tilde{D}_n(x;\vartheta)^2 - D(x;\vartheta)^2| \\ & + \sup_{\vartheta \in \Theta} \Big| \frac{1}{n} \sum_{i=1}^n D(X_i;\vartheta)^2 - \int D(x;\vartheta)^2 \, P(dx) \Big| = o_{a.s.}(1) . \end{aligned}$$

(ii) As $d(\vartheta_\star) = 0$ and per definition $\hat{d}_n(\hat{\vartheta}_n) \le \hat{d}_n(\vartheta_\star)$, $n \in \mathbb{N}$,

$$d(\hat{\vartheta}_n) \le d(\hat{\vartheta}_n) - \hat{d}_n(\hat{\vartheta}_n) + \hat{d}_n(\vartheta_\star) - d(\vartheta_\star) \le 2 \sup_{\vartheta \in \Theta} |\hat{d}_n(\vartheta) - d(\vartheta)| ,$$

where the right-hand side converges almost surely to zero by part (i). □

The almost sure convergence of $d(\hat{\vartheta}_n)$ will now be essential to obtain the strong consistency of the estimator $\hat{\vartheta}_n$.

Theorem 3.23. *If $\theta_\star \in (0,1) \times \mathcal{R}_2 \times \mathcal{E}_3^{f_0}$, $f_0 \in \mathcal{E}_3$, and H is strictly increasing, then, under assumption 3.20, on every compact subset $\Theta \subset (0,1) \times \mathbb{R}^2$ with $\vartheta_\star \in \Theta$ the estimator $\hat{\vartheta}_n$ converges almost surely to the true Euclidean model parameter $\vartheta_\star$ as n tends to infinity.*

Proof. In order to conclude the strong consistency of $\hat{\vartheta}_n$ we make use of the necessary and sufficient condition

$$\lim_{n \to \infty} P(\sup_{m \ge n} \|\hat{\vartheta}_m - \vartheta_\star\|_1 > \varepsilon) = 0 \quad , \quad \varepsilon > 0 . \tag{3.3.3}$$

Therefore, let $\varepsilon > 0$ be arbitrarily small. Due to its continuity, cf. lemma 3.19, the discrepancy function d attains its minimum value on every compact subset of $(0,1] \times \mathbb{R}^2$. Hence, since

$$\Theta_\varepsilon := \{\vartheta \in \Theta \mid \|\vartheta - \vartheta_\star\| \ge \varepsilon\}$$

is compact, there exists $\vartheta_\varepsilon \in \Theta_\varepsilon$ such that

$$d(\vartheta) \geq d(\vartheta_\varepsilon) =: \delta_\varepsilon \quad , \quad \vartheta \in \Theta_\varepsilon \ .$$

This is why

$$\left\{ \sup_{m \geq n} \|\hat{\vartheta}_m - \vartheta_\star\| > \varepsilon \right\} \subset \left\{ \sup_{m \geq n} |d(\hat{\vartheta}_m)| \geq \delta_\varepsilon \right\} \ ,$$

yielding

$$P(\sup_{m \geq n} \|\hat{\vartheta}_n - \vartheta_\star\| > \varepsilon) \leq P(\sup_{m \geq n} |d(\hat{\vartheta}_m)| \geq \delta_\varepsilon) \quad , \quad n \in \mathbb{N} \ . \tag{3.3.4}$$

As $\vartheta_\varepsilon \neq \vartheta_\star$ and d is a discrepancy function with respect to $\vartheta_\star$, see theorem 3.1, it holds $\delta_\varepsilon > 0$, so that by lemma 3.22 (ii) the probability on the right-hand side of (3.3.4) converges to zero as n tends to infinity. The proof is concluded. □

3.4. Asymptotic normality

In addition to the strong consistency of the Euclidean part $\hat{\vartheta}_n$ we now want to establish the asymptotic normality of the semiparametric estimator $(\hat{\vartheta}_n, \hat{F}_n)$, where

$$\hat{F}_n(x) = \hat{H}_n(x + \hat{\mu}_n)/\hat{p}_n - (1 - \hat{p}_n)F_0(x - \hat{\nu}_n + \hat{\mu}_n)/\hat{p}_n \quad , \quad x \in \mathbb{R} \ .$$

In order to do so, we will derive the joint limiting distribution of

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta_\star, \hat{F}_n - F) \quad , \quad n \in \mathbb{N} \ .$$

Here, for all n, the empirical cdf $\hat{F}_n$ is regarded as a random element in the *Skorohod Space* $D(\mathbb{R})$, the normed linear space of bounded *cadlag* functions on $\mathbb{R}$, which we equip with the general supremum norm.

Provided that $\hat{d}_n$ is twice continuously differentiable, the residuals $\sqrt{n}(\hat{\vartheta}_n - \vartheta_\star)$ can be expressed in terms of the first and second order derivative of $\hat{d}_n$. In fact, defining

$$\dot{d}_n := \partial_\vartheta \hat{d}_n \quad , \quad \ddot{d}_n := \partial^2_{\vartheta^2} \hat{d}_n \quad , \quad n \in \mathbb{N} \ ,$$

a zero order Taylor expansion of $\dot{d}_n$ around $\vartheta_\star$ yields

$$\dot{d}_n(\hat{\vartheta}_n) = \dot{d}_n(\vartheta_\star) + (\hat{\vartheta}_n - \vartheta_\star)\ddot{d}_n(\bar{\vartheta}_n)$$

for some $\bar{\vartheta}_n \in \{\vartheta_\star + t(\hat{\vartheta}_n - \vartheta_\star) \mid t \in (0,1)\}$. If $\vartheta_\star \in \Theta$, then by the almost sure convergence of $\hat{\vartheta}_n$, cf. theorem 3.23, we expect $\hat{\vartheta}_n$ being a local minimizer in the interior of Θ, giving $\dot{d}_n(\hat{\vartheta}_n) = 0$ and thus

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta_\star)\ddot{d}_n(\bar{\vartheta}_n) = -\sqrt{n}\dot{d}_n(\vartheta_\star) \ . \tag{3.4.1}$$

Assumption 3.24. (i) Both h and f_0 are continuously differentiable on $\mathbb{R}$, h' is uniformly continuous, and all functions h, f_0, h', and f_0' are bounded.

(ii) The kernel function k is centered about zero, has a finite second order moment, and integrates to one. Further, k is differentiable, both k and its first order derivative, k' say, are uniformly continuous and of bounded variation, and k' has a finite first order moment. Also,

$$\Big\{x \mapsto k\Big(\frac{y-x}{z}\Big) \mid y \in \mathbb{R}, z \in \mathbb{R}_>\Big\}$$

is a bounded VC-class of functions[7].

(iii) For both $k_0 = k$ and $k_0 = k'$,

$$\int \sqrt{|x \log |x||}\, |dk_0(x)| < \infty \, .$$

Further, if k' admits $\psi_{k'}$ as modulus of continuity[8], then

$$\int_0^1 \sqrt{\log(y^{-1})}\, d\sqrt{\psi_{k'}(y)} < \infty \, .$$

(iv) The bandwidth $(b_n)_{n\in\mathbb{N}} \subset \mathbb{R}$, $b_n \downarrow 0$ satisfies $\sqrt{n} b_n^2 \to 0$,

$$\frac{\log n}{nb_n} \longrightarrow 0 \quad , \quad \frac{\log b_n^{-1}}{nb_n^3} \longrightarrow 0 \, , \tag{3.4.2}$$

as well as

$$\frac{\log b_n^{-1}}{nb_n} \longrightarrow 0 \quad , \quad \frac{\log b_n^{-1}}{\log\log n} \longrightarrow \infty \, , \tag{3.4.3}$$

and $b_n \leq c b_{2n}$ for some constant $c > 0$.

(v) There exists $(c_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ sufficing (3.2.5).

(vi) The parameter space $\Theta \subset (0,1) \times \mathbb{R}^2$ on which to estimate is compact.

(vii) We have $\theta_\star \in (0,1) \times \mathcal{R}_1 \times \mathcal{E}_3^{f_0}$, $f_0 \in \mathcal{E}_3$, and H is strictly increasing on $\mathbb{R}$.

Recall that assumption 3.24 simultaneously provides the strong consistency of the cdf $\tilde{H}_n$, pdf $\tilde{h}_n$, its derivative $\tilde{h}_n'$, and the Euclidean estimator $\hat{\vartheta}_n$. Part (iii), in particular, is required for theorem 3.17. The demand for a VC-class of functions is essential regarding theorem 3.18. Part (vii) ensures the continuity of the discrepancy function d which is necessary for the convergence of $\hat{\vartheta}_n$.

[7]Concerning VC-classes, named for V. Vapnik and A.Y. Čhervonenkis, cf. e. g. van der Vaart and Wellner [26], p. 141 et seq.

[8]For each function f, which is uniformly continuous, there exists $\psi_f : [0,\infty) \to [0,\infty)$ such that $|f(x_1) - f(x_2)| \leq \psi_f(|x_1 - x_2|)$, $x_1, x_2 \in \mathbb{R}$. The function ψ_f is referred to as modulus of continuity with respect to f.

Lemma 3.25. *Under assumption 3.24,*

$$\dot{d}_n(\vartheta_\star) = \frac{2}{n}\sum_{i=1}^{n} \hat{D}_n(X_i;\vartheta_\star)\cdot \partial_\vartheta D(X_i;\vartheta_\star) + o_{a.s.}\big(n^{-1/2}\big) \ ,$$

where $\hat{D}_n$ *is defined as* $\tilde{D}_n$*, cf. (3.1.10), replacing* $\tilde{H}_n$ *by* $\hat{H}_n$*.*

Proof. As

$$\dot{d}_n(\vartheta_\star) = \frac{2}{n}\sum_{i=1}^{n} \tilde{D}_n(X_i;\vartheta_\star)\cdot \partial_\vartheta \tilde{D}_n(X_i;\vartheta_\star) \tag{3.4.4}$$

let us define

$$S_n := \frac{2}{n}\sum_{i=1}^{n} \big(\tilde{D}_n(X_i;\vartheta_\star)\cdot \partial_\vartheta \tilde{D}_n(X_i;\vartheta_\star) - \hat{D}_n(X_i;\vartheta_\star)\cdot \partial_\vartheta D(X_i;\vartheta_\star)\big) \ .$$

By a zero addition

$$\begin{aligned} S_n &= \frac{2}{n}\sum_{i=1}^{n} \big(\tilde{D}_n(X_i;\vartheta_\star)\cdot \partial_\vartheta \tilde{D}_n(X_i;\vartheta_\star) - \tilde{D}_n(X_i;\vartheta_\star)\cdot \partial_\vartheta D(X_i;\vartheta_\star) \\ &\qquad + \tilde{D}_n(X_i;\vartheta_\star)\cdot \partial_\vartheta D(X_i;\vartheta_\star) - \hat{D}_n(X_i;\vartheta_\star)\cdot \partial_\vartheta D(X_i;\vartheta_\star)\big) \\ &= \frac{2}{n}\sum_{i=1}^{n} \Big(\tilde{D}_n(X_i;\vartheta_\star)\cdot \big(\partial_\vartheta \tilde{D}_n(X_i;\vartheta_\star) - \partial_\vartheta D(X_i;\vartheta_\star)\big) \\ &\qquad + \big(\tilde{D}_n(X_i;\vartheta_\star) - \hat{D}_n(X_i;\vartheta_\star)\big)\cdot \partial_\vartheta D(X_i;\vartheta_\star)\Big) \\ &= T_n^{(1)} + T_n^{(2)} \ , \end{aligned}$$

where

$$\begin{aligned} T_n^{(1)} &= \frac{2}{n}\sum_{i=1}^{n} \big(\tilde{D}_n(X_i;\vartheta_\star) - D(X_i;\vartheta_\star)\big)\cdot \big(\partial_\vartheta \tilde{D}_n(X_i;\vartheta_\star) - \partial_\vartheta D(X_i;\vartheta_\star)\big) \ , \\ T_n^{(2)} &= \frac{2}{n}\sum_{i=1}^{n} \big(\tilde{D}_n(X_i;\vartheta_\star) - \hat{D}_n(X_i;\vartheta_\star)\big)\cdot \partial_\vartheta D(X_i;\vartheta_\star) \ , \end{aligned}$$

minding that $D(x;\vartheta_\star) \equiv 0$.

For all $x \in \mathbb{R}$,

$$\begin{aligned} |\tilde{D}_n(x;\vartheta_\star) - D(x;\vartheta_\star)| &= p_\star^{-1}\big|\tilde{H}_n(x+\mu_\star) - H(x+\mu_\star) \\ &\qquad + \tilde{H}_n(-x+\mu_\star) - H(-x+\mu_\star)\big| \\ &\le 2p_\star^{-1}\|\tilde{H}_n - H\|_\infty \ . \end{aligned} \tag{3.4.5}$$

Similarly, we obtain

$$
\begin{aligned}
&|\partial_p \tilde{D}_n(x;\vartheta_\star) - \partial_p D(x;\vartheta_\star)| \le 2p_\star^{-2}\|\tilde{H}_n - H\|_\infty \ , \\
&|\partial_\nu \tilde{D}_n(x;\vartheta_\star) - \partial_\nu D(x;\vartheta_\star)| = 0 \ , \\
&|\partial_\mu \tilde{D}_n(x;\vartheta_\star) - \partial_\mu D(x;\vartheta_\star)| \le 2p_\star^{-1}\|\tilde{h}_n - h\|_\infty \ ,
\end{aligned}
$$

so that

$$
\begin{aligned}
\|T_n^{(1)}\|_1 &\le 8p_\star^{-2}\big(p_\star^{-1}\|\tilde{H}_n - H\|_\infty^2 + \|\tilde{H}_n - H\|_\infty \|\tilde{h}_n - h\|_\infty\big) \\
&= O_{a.s.}\big((\log\log n)/n\big) \\
&\quad + O_{a.s.}\big(\sqrt{(\log\log n)/n}\big)\big(O_{a.s.}\big(\sqrt{(\log b_n^{-1})/(nb_n)}\big) + O(b_n)\big) \\
&= o_{a.s.}(n^{-1/2}) + O_{a.s.}\Big(\frac{\sqrt{\log\log n}\sqrt{\log b_n^{-1}}}{n\sqrt{b_n}} + \frac{b_n\sqrt{\log\log n}}{\sqrt{n}}\Big) \ .
\end{aligned}
$$

by corollary 3.16 and theorem 3.18. Note that $\sqrt{n}b_n^2 = o(1)$ implies the convergence of $b_n \log\log n$ to zero. Hence, using (3.4.2),

$$
\begin{aligned}
\sqrt{n}\Big(\frac{\sqrt{\log\log n}\sqrt{\log b_n^{-1}}}{n\sqrt{b_n}} + \frac{b_n\sqrt{\log\log n}}{\sqrt{n}}\Big) &= \frac{\sqrt{\log\log n}\sqrt{\log b_n^{-1}}}{\sqrt{nb_n}} + o(1) \\
&= b_n\sqrt{\log\log n}\sqrt{\frac{\log b_n^{-1}}{nb_n^3}} + o(1) \\
&= o(1) \ .
\end{aligned}
$$

We conclude $T_n^{(1)} = o_{a.s.}(n^{-1/2})$.

Similar to (3.4.5), for all $x \in \mathbb{R}$,

$$
|\tilde{D}_n(x;\vartheta_\star) - \hat{D}_n(x;\vartheta_\star)| \le 2p_\star^{-1}\|\tilde{H}_n - \hat{H}_n\|_\infty \ .
$$

Further, $\partial_\vartheta D(x;\vartheta_\star)$ is bounded on $\mathbb{R}$ due to the boundedness of f_0, F_0, and h (for the exact computation of $\partial_\vartheta D(x;\vartheta_\star)$ cf. (3.4.9)-(3.4.11)), so that there exists $c > 0$ such that

$$
\|T_n^{(2)}\|_1 \le 4cp_\star^{-1}\|\tilde{H}_n - \hat{H}_n\|_\infty \ .
$$

Thus, corollary 3.16 leads to $T_n^{(2)} = o_{a.s.}(n^{-1/2})$, what concludes the proof. □

Lemma 3.26. *If $(\vartheta_n)_{n\in\mathbb{N}} \subset \Theta$ such that $\vartheta_n \to \vartheta_\star$, then, under assumption 3.24,*

(i) $\sup_{x\in\mathbb{R}} |D(x;\vartheta_n)| \longrightarrow 0$ *and*

(ii) $\sup_{x\in\mathbb{R}} |\partial_\vartheta D(x;\vartheta_n) - \partial_\vartheta D(x;\vartheta_\star)| \longrightarrow 0$

as n tends to infinity.

Proof. (i) The cdfs H and F_0 are Lipschitz continuous on $\mathbb{R}$ as h and f_0 are bounded,

i. e. there exist $c_1, c_2 > 0$ such that

$$|H(x_1)-H(x_2)| \le c_1|x_1-x_2| \quad , \quad |F_0(x_1)-F_0(x_2)| \le c_2|x_1-x_2| \quad , \quad x_1, x_2 \in \mathbb{R} \, .$$

Since $D(x;\vartheta_\star) = 0$, $x \in \mathbb{R}$,

$$\begin{aligned}|D(x;\vartheta_n)| &= |D(x;\vartheta_n) - D(x;\vartheta_\star)| \\ &\le |D(x;\vartheta_n) - D(x;p_\star,\nu_n,\mu_n)| + |D(x;p_\star,\nu_n,\mu_n) - D(x;\vartheta_\star)| \, .\end{aligned}$$

First, by the compactness of Θ and the boundedness of H and F_0,

$$D(x;\vartheta_n) - D(x;p_\star,\nu_n,\mu_n) = \left(p_n^{-1} - p_\star^{-1}\right) O(1) \, ,$$

giving

$$\sup_{x\in\mathbb{R}} |D(x;\vartheta_n) - D(x;p_\star,\nu_n,\mu_n)| \longrightarrow 0 \quad , \quad n \to \infty$$

by the convergence of ϑ_n. Second,

$$\begin{aligned}D(x;p_\star,\nu_n,\mu_n) - D(x;\vartheta_\star) = p_\star^{-1}\big(&H(x+\mu_n) - (1-p_\star)F_0(x-\nu_n+\mu_n) \\ &+ H(-x+\mu_n) - (1-p_\star)F_0(-x-\nu_n+\mu_n) \\ &- H(x+\mu_\star) + (1-p_\star)F_0(x-\nu_\star+\mu_\star) \\ &+ H(-x+\mu_\star) - (1-p_\star)F_0(-x-\nu_\star+\mu_\star)\big) \, ,\end{aligned}$$

so that by the Lipschitz continuity

$$\begin{aligned}|D(x;p_\star,\nu_n,\mu_n) - D(x;\vartheta_\star)| &\le p_\star^{-1}\big(2c_1|\mu_n - \mu_\star| \\ &\qquad + 2(1-p_\star)c_2|\mu_n - \nu_n - \mu_\star + \nu_\star|\big) \\ &\le p_\star^{-1} 2\big(c_1 + (1-p_\star)c_2\big)\|\vartheta_n - \vartheta_\star\|_1 \, .\end{aligned}$$

Hence,

$$\sup_{x\in\mathbb{R}} |D(x;p_\star,\nu_n,\mu_n) - D(x;\vartheta_\star)| \longrightarrow 0 \quad , \quad n \to \infty \, ,$$

concluding (i).

(ii) Additionally using the boundedness of h' and f_0', the uniform convergence of $\partial_\vartheta D(x;\vartheta_n)$ can be concluded likewise. □

Lemma 3.27. *Let $\bar{\vartheta}_n$ be as in (3.4.1). Then, under assumption 3.24,*

(i) $\ddot{d}_n(\bar{\vartheta}_n)$ converges almost surely to the deterministic 3×3 matrix

$$\boldsymbol{I} = 2 \int \partial_\vartheta D(x;\vartheta_\star)^t \cdot \partial_\vartheta D(x;\vartheta_\star) \, P(dx)$$

as $n \to \infty$.

(ii) The matrix $\boldsymbol{I}$ *is positive definite.*

Proof. (i) Taking the derivative of (3.4.4) gives

$$\ddot{d}_n(\vartheta) = \frac{2}{n}\sum_{i=1}^{n}\left(\partial_\vartheta \tilde{D}_n(X_i;\vartheta)^t \cdot \partial_\vartheta \tilde{D}_n(X_i;\vartheta) + \tilde{D}_n(X_i;\vartheta)\cdot \partial^2_{\vartheta^2}\tilde{D}_n(X_i;\vartheta)\right) .$$

With the particular argument $\bar{\vartheta}_n$ we rewrite

$$\ddot{d}_n(\bar{\vartheta}_n) = T_n^{(1)} + T_n^{(2)} + T_n^{(3)} , \tag{3.4.6}$$

where

$$\begin{aligned} T_n^{(1)} &= \frac{2}{n}\sum_{i=1}^{n}\tilde{D}_n(X_i;\bar{\vartheta}_n)\cdot \partial^2_{\vartheta^2}\tilde{D}_n(X_i;\bar{\vartheta}_n) , \\ T_n^{(2)} &= \frac{2}{n}\sum_{i=1}^{n}\left(\partial_\vartheta \tilde{D}_n(X_i;\bar{\vartheta}_n)^t \cdot \partial_\vartheta \tilde{D}_n(X_i;\bar{\vartheta}_n) - \partial_\vartheta D(X_i;\vartheta_\star)^t\cdot \partial_\vartheta D(X_i;\vartheta_\star)\right) , \\ T_n^{(3)} &= \frac{2}{n}\sum_{i=1}^{n}\partial_\vartheta D(X_i;\vartheta_\star)^t\cdot \partial_\vartheta D(X_i;\vartheta_\star) . \end{aligned}$$

Let us successively examine the convergence behavior of $T_n^{(1)}$, $T_n^{(2)}$, and $T_n^{(3)}$.

1. First, we conclude $\partial^2_{\vartheta^2}\tilde{D}_n(x;\bar{\vartheta}) = O_{a.s.}(1)$ through pointing out the asymptotic boundedness of all component. Basically, this follows from the compactness of Θ and from h, f_0, and their derivatives all being bounded. In fact, e. g.

   $$\begin{aligned} \partial^2_{\mu^2}\tilde{D}_n(x;\vartheta) = p^{-1}\big(&\tilde{h}_n'(x+\mu) - (1-p)f_0'(x-\nu+\mu) \\ &+ \tilde{h}_n'(-x+\mu) - (1-p)f_0'(-x-\nu+\mu)\big) \quad , \quad x\in\mathbb{R} , \end{aligned}$$

   where $\tilde{h}_n'$ denotes the first order derivative of $\tilde{h}_n$. As h' is bounded and $\tilde{h}_n'$ converges uniformly to h', cf. theorem 3.17 (ii),

   $$\|\tilde{h}_n'\|_\infty = \|h'\|_\infty + \|\tilde{h}_n' - h'\|_\infty = O(1) + o_{a.s.}(1) = O_{a.s.}(1) .$$

   The asymptotic boundedness of $\partial^2_{\mu^2}\tilde{D}_n(x;\vartheta)$ follows. The remaining components of $\partial^2_{\vartheta^2}\tilde{D}_n(x;\vartheta)$ can be handled likewise, most are easier, actually.

   Second, as $\bar{\vartheta}_n = \vartheta_\star + t_n(\hat{\vartheta}_n - \vartheta_\star)$ for some $t_n\in(0,1)$,

   $$\|\bar{\vartheta}_n - \vartheta_\star\| \le \|\hat{\vartheta}_n - \vartheta_\star\| \longrightarrow 0 \quad \text{a.s.} \quad , \quad n\to\infty$$

by theorem 3.23. Therefore,

$$\begin{aligned}\sup_{x\in\mathbb{R}}|\tilde{D}_n(x;\bar{\vartheta}_n)| &\leq \sup_{x\in\mathbb{R}}|\tilde{D}_n(x;\bar{\vartheta}_n)-D(x;\bar{\vartheta}_n)|+\sup_{x\in\mathbb{R}}|D(x;\bar{\vartheta}_n)|\\ &\leq c_2\|\tilde{H}_n-H\|_\infty+\sup_{x\in\mathbb{R}}|D(x;\bar{\vartheta}_n)|\\ &= o_{a.s.}(1) \end{aligned}\tag{3.4.7}$$

by corollary 3.16 and lemma 3.26 (i), where the constant c_2 results from the compactness of Θ, cf. (3.3.2). Using (3.4.7) and the asymptotic boundedness of $\partial^2_{\vartheta^2}\tilde{D}_n(x;\vartheta)$ we conclude

$$T_n^{(1)} = o_{a.s.}(1)\ .$$

2. Let us representatively consider the (1,3)-component of the matrix $T_n^{(2)}$, i. e.

$$\frac{2}{n}\sum_{i=1}^{n}\big(\partial_p\tilde{D}_n(X_i;\bar{\vartheta}_n)\cdot\partial_\mu\tilde{D}_n(X_i;\bar{\vartheta}_n)-\partial_pD(X_i;\vartheta_\star)\cdot\partial_\mu D(X_i;\vartheta_\star)\big)\ ,$$

which we in the following denote by ${}_{(1,3)}T_n^{(2)}$. For all $x\in\mathbb{R}$,

$$\partial_p\tilde{D}_n(x;\bar{\vartheta}_n)\cdot\partial_\mu\tilde{D}_n(x;\bar{\vartheta}_n)-\partial_pD(x;\vartheta_\star)\cdot\partial_\mu D(x;\vartheta_\star)=L_n^{(1)}(x)+L_n^{(2)}(x)\ ,$$

where

$$\begin{aligned}L_n^{(1)}(x)&=\partial_p\tilde{D}_n(x;\bar{\vartheta}_n)\cdot\big(\partial_\mu\tilde{D}_n(x;\bar{\vartheta}_n)-\partial_\mu D(x;\vartheta_\star)\big)\ ,\\ L_n^{(2)}(x)&=\partial_\mu D(x;\vartheta_\star)\cdot\big(\partial_p\tilde{D}_n(x;\bar{\vartheta}_n)-\partial_pD(x;\vartheta_\star)\big)\ .\end{aligned}$$

Note that both $\partial_p\tilde{D}_n(x;\bar{\vartheta}_n)$ and $\partial_\mu D(x;\vartheta_\star)$ are bounded. Through a zero addition similar to (3.4.7) we obtain

$$\begin{aligned}\sup_{x\in\mathbb{R}}\big(\partial_\mu\tilde{D}_n(x;\bar{\vartheta}_n)-\partial_\mu D(x;\vartheta_\star)\big)&\leq c_2\|\tilde{h}_n-h\|_\infty\\ &\quad+\sup_{x\in\mathbb{R}}\big(\partial_\mu D(x;\bar{\vartheta}_n)-\partial_\mu D(x;\vartheta_\star)\big)\\ &=o_{a.s.}(1)\end{aligned}$$

and

$$\begin{aligned}\sup_{x\in\mathbb{R}}\big(\partial_p\tilde{D}_n(x;\bar{\vartheta}_n)-\partial_pD(x;\vartheta_\star)\big)&\leq c_2^2\|\tilde{H}_n-H\|_\infty\\ &\quad+\sup_{x\in\mathbb{R}}\big(\partial_pD(x;\bar{\vartheta}_n)-\partial_pD(x;\vartheta_\star)\big)\\ &=o_{a.s.}(1)\end{aligned}$$

by theorem 3.17, corollary 3.16, and lemma 3.26 (ii). It follows $\|L_n^{(1)}\|_\infty =$

$o_{a.s.}(1)$ as well as $\|L_n^{(2)}\|_\infty = o_{a.s.}(1)$ and hence

$$|_{(1,3)}T_n^{(2)}| \leq 2\big(\|L_n^{(1)}\|_\infty + \|L_n^{(2)}\|_\infty\big) \longrightarrow 0 \quad \text{a.s.} \quad , \quad n \to \infty \ .$$

The remaining components of $T_n^{(2)}$ can be handled in a similar way, altogether giving

$$T_n^{(2)} = o_{a.s.}(1) \ .$$

3. The random matrices $\partial_\vartheta D(X_i;\vartheta_\star)^t \cdot \partial_\vartheta D(X_i;\vartheta_\star)$, $i \in \mathbb{N}$ are independent and identically distributed as the X_i's are. Moreover,

$$\mathbf{E}\left[\|\partial_\vartheta D(X_1;\vartheta_\star)^t \cdot \partial_\vartheta D(X_1;\vartheta_\star)\|_1\right] < \infty$$

due to the boundedness of h and f_0 and the compactness of Θ. By the strong law of large numbers it follows that, as n tends to infinity, almost surely

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \partial_\vartheta D(X_i;\vartheta_\star)^t \cdot \partial_\vartheta D(X_i;\vartheta_\star) &\longrightarrow \mathbf{E}\left[\partial_\vartheta D(X_1;\vartheta_\star)^t \cdot \partial_\vartheta D(X_1;\vartheta_\star)\right] \\ &= \int \partial_\vartheta D(x;\vartheta_\star)^t \cdot \partial_\vartheta D(x;\vartheta_\star)\, P(dx) \ .\end{aligned}$$

Hence,

$$T_n^{(3)} \longrightarrow \boldsymbol{I} \quad \text{a.s.} \quad , \quad n \to \infty \ .$$

With (3.4.6) and 1.-3. we conclude the convergence of $\ddot{d}_n(\bar{\vartheta}_n)$.

(ii) For any arbitrary $w \in \mathbb{R}^3$,

$$\begin{aligned} w^t \boldsymbol{I} w &= 2\int w^t\big(\partial_\vartheta D(x;\vartheta_\star)^t \cdot \partial_\vartheta D(x;\vartheta_\star)\big) w\, P(dx) \\ &= 2\int (\partial_\vartheta D(x;\vartheta_\star) w)^2\, P(dx) \geq 0 \ , \end{aligned} \tag{3.4.8}$$

giving the positivity of $\boldsymbol{I}$. Therefore, let w be such that $w^t\boldsymbol{I}w = 0$ in what follows.

As H is strictly increasing and thus $h > 0$ almost surely, (3.4.8) leads to

$$\partial_\vartheta D(x;\vartheta_\star) w = 0 \quad , \quad x \in \mathbb{R} \ ,$$

minding the continuity of $x \mapsto \partial_\vartheta D(x;\vartheta_\star)w$. Defining $\zeta \in \mathbb{R}\backslash\{0\}$ according to

$\zeta := \mu_\star - \nu_\star$, a more precise consideration of the partial derivatives gives

$$\begin{aligned}
\partial_p D(x;\vartheta_\star) &= p_\star^{-2}\big(F_0(x+\zeta) + F_0(-x+\zeta) - H(x+\mu_\star) - H(-x+\mu_\star)\big)\\
&= p_\star^{-2}\big(F_0(x+\zeta) + F_0(-x+\zeta)\\
&\quad - (1-p_\star)F_0(x+\zeta) - p_\star F(x)\\
&\quad - (1-p_\star)F_0(-x+\zeta) - p_\star F(-x)\big)\\
&= p_\star^{-1}\big(F_0(x+\zeta) + F_0(-x+\zeta) - F(x) - F(-x)\big)\\
&= p_\star^{-1}\big(F_0(x+\zeta) + F_0(-x+\zeta) - 1\big)\\
&= p_\star^{-1}\big(F_0(x+\zeta) - F_0(x-\zeta)\big)\\
&=: g_1(x) \; ,
\end{aligned} \tag{3.4.9}$$

$$\begin{aligned}
\partial_\nu D(x;\vartheta_\star) &= \frac{1-p_\star}{p_\star}\big(f_0(x+\zeta) + f_0(-x+\zeta)\big)\\
&= \frac{1-p_\star}{p_\star}\big(f_0(x+\zeta) + f_0(x-\zeta)\big)\\
&=: g_2(x) \; ,
\end{aligned} \tag{3.4.10}$$

$$\begin{aligned}
\partial_\mu D(x;\vartheta_\star) &= p_\star^{-1}\big(h(x+\mu_\star) - (1-p_\star)f_0(x-\nu_\star+\mu_\star)\\
&\quad + h(-x+\mu_\star) - (1-p_\star)f_0(-x-\nu_\star+\mu_\star)\big)\\
&= f_\star(x) + f_\star(-x)\\
&= 2f_\star(x)\\
&=: g_3(x) \; ,
\end{aligned} \tag{3.4.11}$$

satisfying

$$w_1 g_1(x) + w_2 g_2(x) + w_3 g_3(x) = 0 \quad , \quad x \in \mathbb{R} \; . \tag{3.4.12}$$

Referring to the Fourier transform of g_k as $\hat{g}_k$, $k = 1, \ldots, 3$,

$$\begin{aligned}
\hat{g}_1(t) &= \operatorname{sgn}(\zeta) p_\star^{-1} \hat{f}_0(t) \int_{-|\zeta|}^{|\zeta|} e^{itz}\, dz \; ,\\
\hat{g}_2(t) &= \frac{1-p_\star}{p_\star}\big(e^{-it\zeta} - e^{it\zeta}\big)\hat{f}_0(t) \; ,\\
\hat{g}_3(t) &= 2\hat{f}_\star(t) \quad , \quad t \in \mathbb{R} \; ,
\end{aligned}$$

at what $\hat{f}_\star$ and $\hat{f}_0$ respectively denote the Fourier transforms of $f_\star$ and f_0. In fact,

$$\begin{aligned}
p_\star \hat{g}_1(t) &= \int e^{-itx}\big(F_0(x+\zeta) - F_0(x-\zeta)\big)\,dx \\
&= \int e^{-itx}\Big(\int_{-\infty}^{x+\zeta} f_0(y)\,dy - \int_{-\infty}^{x-\zeta} f_0(y)\,dy\Big)dx \\
&= \operatorname{sgn}(\zeta)\int\int_{x-|\zeta|}^{x+|\zeta|} e^{-itx}\, f_0(y)\,dy\,dx \\
&= \operatorname{sgn}(\zeta)\int\int_{-|\zeta|}^{|\zeta|} e^{-itx}\, f_0(z+x)\,dz\,dx \\
&= \operatorname{sgn}(\zeta)\int_{-|\zeta|}^{|\zeta|}\int e^{-itx}\, f_0(z+x)\,dx\,dz \\
&= \operatorname{sgn}(\zeta)\int_{-|\zeta|}^{|\zeta|}\int e^{-it(y-z)}\, f_0(y)\,dy\,dz \\
&= \operatorname{sgn}(\zeta)\hat{f}_0(t)\int_{-|\zeta|}^{|\zeta|} e^{itz}\,dz\ ,
\end{aligned}$$

where the order of integration is interchangeable due to *Fubini's theorem*[9] as the function $(x,z)\mapsto e^{-itx}\, f_0(z+x)$ is λ^2-integrable on $\mathbb{R}\times[-|\zeta|,|\zeta|]$ for fixed $\zeta\in\mathbb{R}$. The remaining Fourier transforms $\hat{g}_2$ and $\hat{g}_3$ can be computed easily.

Let us consider the Fourier transform equation induced by (3.4.12), i. e.

$$w_1\hat{g}_1(t) + w_2\hat{g}_2(t) + w_3\hat{g}_3(t) = 0 \quad , \quad t\in\mathbb{R}\ . \tag{3.4.13}$$

First, note that the imaginary parts of $\hat{g}_1$ and $\hat{g}_3$ are equal to zero since $f_\star$ and f_0 are even functions,

$$\hat{g}_1(t) = \operatorname{sgn}(\zeta)p_\star^{-1}\hat{f}_0(t)\Big(\int_{-|\zeta|}^{|\zeta|}\cos(tz)\,dz + i\int_{-|\zeta|}^{|\zeta|}\sin(tz)\,dz\Big)\ , \tag{3.4.14}$$

and

$$\int_{-|\zeta|}^{|\zeta|}\sin(tz)\,dz = 0 \quad , \quad t\in\mathbb{R}\ .$$

Further,

$$\begin{aligned}
\hat{g}_2(t) &= \frac{1-p_\star}{p_\star}\big(\cos(-t\zeta) + i\sin(-t\zeta) - \cos(t\zeta) - i\sin(t\zeta)\big)\hat{f}_0(t) \\
&= \frac{-2(1-p_\star)}{p_\star}\,\sin(t\zeta)i\,\hat{f}_0(t) \quad , \quad t\in\mathbb{R}\ ,
\end{aligned}$$

[9] Cf. Bauer [1], p. 158, corollary 23.7, for instance.

so that the imaginary part of (3.4.13) gives

$$w_2 \frac{-2(1-p_\star)}{p_\star} \sin(t\zeta)\, \hat{f}_0(t) = 0 \quad , \quad t \in \mathbb{R} \,.$$

As $t \mapsto \hat{f}_0(t)$ is continuous with $\hat{f}_0(0) = 1$ and $\sin(t\zeta) \neq 0$ for small $t > 0$, there exists $t_0 \in \mathbb{R}$ such that $\sin(t_0\zeta)\hat{f}_0(t_0) \neq 0$, yielding $w_2 = 0$.

Considering the real part of (3.4.13), using (3.4.14), we obtain

$$w_1 \operatorname{sgn}(\zeta) p_\star^{-1} \hat{f}_0(t) \int_{-|\zeta|}^{|\zeta|} \cos(tz)\, dz + w_3 2\hat{f}_\star(t) = 0 \quad , \quad t \in \mathbb{R} \,. \tag{3.4.15}$$

As $f_\star \in \mathcal{E}_3^{f_0}$ by presupposition, i. e. $f_\star$ meets one of the conditions C1 and C2, we conclude the following.

If $\hat{f}_0(t)/\hat{f}_\star(t) \to 0$ holds, then $w_3 = 0$ by (3.4.15) as $t \mapsto \int_{-|\zeta|}^{|\zeta|} \cos(tz)dz$ is bounded. Further,

$$w_1 \operatorname{sgn}(\zeta) p_\star^{-1} \hat{f}_0(t) \int_{-|\zeta|}^{|\zeta|} \cos(tz)\, dz = 0 \quad , \quad t \in \mathbb{R}$$

in this case. As $\hat{f}_0(t) \int_{-|\zeta|}^{|\zeta|} \cos(tz)\, dz \neq 0$ for small $t > 0$, it also follows $w_1 = 0$.

If on the other hand $\hat{f}_\star(t)/\hat{f}_0(t) \to 0$, then

$$\lim_{t\to\infty} w_1 \operatorname{sgn}(\zeta) p_\star^{-1} \int_{-|\zeta|}^{|\zeta|} \cos(tz)\, dz = 0$$

by (3.4.15), giving $w_1 = 0$ as $t \mapsto \int_{-|\zeta|}^{|\zeta|} \cos(tz)\, dz$ is periodic and unequal to zero whenever $t \notin \{n\pi/|\zeta| \mid n \in \mathbb{N}\}$. As then $w_3 2\hat{f}_\star(t) = 0$, $t \in \mathbb{R}$, the particular argument $t = 0$ gives $w_3 = 0$.

In consequence, $w^t \boldsymbol{I} w = 0$ holds true if and only if $w = 0$, that is the matrix $\boldsymbol{I}$ is definite, too. □

For convenience, let us introduce the random sequence

$$(\mathcal{U}_n)_{n\in\mathbb{N}} \subset \mathbb{R}^3 \quad , \quad \mathcal{U}_n = (\mathcal{U}_n^{(1)}, \mathcal{U}_n^{(2)}, \mathcal{U}_n^{(3)})$$

defined by

$$\mathcal{U}_n = \frac{2}{n} \sum_{i=1}^{n} \hat{D}_n(X_i; \vartheta_\star) \cdot \partial_\vartheta D(X_i; \vartheta_\star) \,.$$

Lemma 3.25 points out the relevance of the process $(\mathcal{U}_n)_{n\in\mathbb{N}}$.

Lemma 3.28. *If* $(Y_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ *is a random sequence such that* $Y_n = o_p(1)$*, then*

$$Y_n\mathcal{U}_n = o_p(n^{-1/2}) \ .$$

Proof. As Θ is compact, $x \mapsto \partial_\vartheta D(x;\vartheta_\star)$ is bounded, and $D(x;\vartheta_\star) \equiv 0$ there exists $c > 0$ such that

$$|\mathcal{U}_n^{(k)}| \leq 2c\|\hat{H}_n - H\|_\infty \quad , \quad k = 1,2,3 \ .$$

As H is continuous, there exists a real random variable K such that $\sqrt{n}\|\hat{H}_n - H\|_\infty$, known as the *Kolmogorov-Smirnov statistic*[10], gives

$$\sqrt{n}\|\hat{H}_n - H\|_\infty \rightsquigarrow K \ .$$

Hence, by *Slutsky's theorem*[11],

$$\sqrt{n}|Y_n\mathcal{U}_n^{(k)}| \leq 2c\sqrt{n}|Y_n|\|\hat{H}_n - H\|_\infty \rightsquigarrow 0 \ ,$$

yielding $\sqrt{n}|Y_n\mathcal{U}_n^{(k)}| \rightsquigarrow 0$ and thus, as the limit is constant,

$$Y_n\mathcal{U}_n^{(k)} = o_p(n^{-1/2}) \quad , \quad k = 1,2,3 \ .$$

The convergence of $Y_n\mathcal{U}_n$ follows. □

Lemma 3.29. *Under assumption 3.24 the random sequence* $\sqrt{n}(\hat{\vartheta}_n - \vartheta_\star)$*,* $n \in \mathbb{N}$ *converges in distribution to a zero mean Gaussian random vector with covariance matrix* $\boldsymbol{I}^{-1}\boldsymbol{\Sigma}\boldsymbol{I}^{-1}$*, where* $\boldsymbol{I}$ *comes from lemma 3.27 and* $\boldsymbol{\Sigma}$ *is given in (3.4.19).*

Proof. By lemma 3.27 the matrix $\ddot{d}_n(\bar{\vartheta}_n)$ is regular for sufficiently large n with

$$\ddot{d}_n(\bar{\vartheta}_n)^{-1} \longrightarrow \boldsymbol{I}^{-1} \quad \text{a.s.} \quad , \quad n \to \infty \ .$$

Therefore, regarding (3.4.1),

$$\begin{aligned}
\sqrt{n}(\hat{\vartheta}_n - \vartheta_\star) &= -\sqrt{n}\dot{d}_n(\vartheta_\star)\ddot{d}_n(\bar{\vartheta}_n)^{-1} \\
&= \sqrt{n}\big(\mathcal{U}_n + o_{a.s.}(n^{-1/2})\big)\big(\boldsymbol{I}^{-1} + o_{a.s.}(1)\big) \\
&= \sqrt{n}\,\mathcal{U}_n\boldsymbol{I}^{-1} + \sqrt{n}\,\mathcal{U}_n o_{a.s.}(1) + o_{a.s.}(1) \\
&= \sqrt{n}\,\mathcal{U}_n\boldsymbol{I}^{-1} + o_p(1) \ , \qquad (3.4.16)
\end{aligned}$$

using lemmata 3.25 and 3.28. The $\mathcal{U}_n$ can be written in terms of a Riemann-Stieltjes integral according to

$$\mathcal{U}_n = 2\int \hat{D}_n(x;\vartheta_\star)\cdot\partial_\vartheta D(x;\vartheta_\star)\,\hat{H}_n(dx) \ .$$

[10] Cf. van der Vaart [25], p. 277, section 19.3 and particularly corollary 19.21.
[11] Cf. van der Vaart [25], p. 11, lemma 2.8.

Regarding $\hat{D}_n$ as a function of $\hat{H}_n$, e. g.

$$\hat{D}_n(x;\vartheta_\star) = D(x;\vartheta_\star,\hat{H}_n) = p_\star^{-1}\big(\hat{H}_n(x+\mu_\star) - (1-p_\star)F_0(x-\nu_\star+\mu_\star) \\ + \hat{H}_n(-x+\mu_\star) - (1-p_\star)F_0(-x-\nu_\star+\mu_\star)\big) - 1 \;,$$

let us introduce the maps $\psi : BV_1(\mathbb{R}) \to D(\mathbb{R})^3 \times BV_1(\mathbb{R})$ according to

$$\psi(F) = \begin{pmatrix} 2\cdot D(x;\vartheta_\star,F)\cdot \partial_\vartheta D(x;\vartheta_\star) \\ F \end{pmatrix}$$

and $\phi : D(\mathbb{R})^3 \times BV_1(\mathbb{R}) \to \mathbb{R}^3$ given by

$$\phi(G_1,G_2) = \int G_1\, dG_2 \quad , \quad G_1 \in D(\mathbb{R})^3 \;,\; G_2 \in BV_1(\mathbb{R}) \;,$$

where $BV_1(\mathbb{R})$ and $D(\mathbb{R})$ are defined as in example 3.11. We obtain

$$\mathcal{U}_n = (\phi\circ\psi)(\hat{H}_n) \quad , \quad n\in\mathbb{N} \;. \tag{3.4.17}$$

The map $\phi\circ\psi$ is Hadamard-differentiable at H. In fact, ψ is Hadamard-differentiable at H with derivative

$$\psi'_H(F(x)) = \begin{pmatrix} 2p_\star^{-1}\big(F(x+\mu_\star)+F(-x+\mu_\star)\big)\cdot\partial_\vartheta D(x;\vartheta_\star) \\ F(x) \end{pmatrix}$$

and concerning the differentiability of ϕ see example 3.11. Applying corollary 3.12 leads to

$$(\phi\circ\psi)'_H(F(x)) = \phi'_{\psi(H)}(\psi'_H(F(x))) \\ = 2p_\star^{-1}\int \big(F(x+\mu_\star)+F(-x+\mu_\star)\big)\cdot\partial_\vartheta D(x;\vartheta_\star)\,H(dx)$$

for all $F\in BV_1(\mathbb{R})$, where the first integral of the derivative of ϕ, cf. (3.2.4), drops out since $\psi(H) = (0,H)^t$ as $D(x;\vartheta_\star,H)\equiv 0$.

By corollary 3.9 there exists a Gaussian process $\mathbb{G}$ such that $\sqrt{n}(\hat{H}_n - H) \rightsquigarrow \mathbb{G}$ as $n\to\infty$. As $\phi\circ\psi$ is Hadamard-differentiable at H and as $\phi(\psi(H)) = 0$, by the Delta-method, theorem 3.13, we obtain

$$\sqrt{n}\,\mathcal{U}_n = \sqrt{n}\big(\phi(\psi(\hat{H}_n)) - \phi(\psi(H))\big) \rightsquigarrow (\phi\circ\psi)'_H(\mathbb{G}) \;,$$

where

$$\mathcal{G} := (\phi\circ\psi)'_H(\mathbb{G}(x)) = 2p_\star^{-1}\int \big(\mathbb{G}(x+\mu_\star)+\mathbb{G}(-x+\mu_\star)\big)\cdot\partial_\vartheta D(x;\vartheta_\star)H(dx) \tag{3.4.18}$$

is a three dimensional zero mean Gaussian random vector.[12] The covariance matrix

[12] Cf. Ibragimov and Rozanov [15], p. 12.

of $\mathcal{G}$ can be calculated as follows. Denoting the autocovariance function of the limit process $\mathbb{G}$ by ρ, cf. corollary 3.9, the autocovariance function of the Gaussian process $(\mathbb{G}(x+\mu_\star)+\mathbb{G}(-x+\mu_\star))_{x\in\mathbb{R}}$, noting that $\mathbb{G}$ has a zero mean, is given by

$$\begin{aligned}\bar\rho(x,y) &= \mathbf{E}\left[(\mathbb{G}(x+\mu_\star)+\mathbb{G}(-x+\mu_\star))(\mathbb{G}(y+\mu_\star)+\mathbb{G}(-y+\mu_\star))\right] \\ &= \mathbf{E}\left[\mathbb{G}(x+\mu_\star)\mathbb{G}(y+\mu_\star)\right] + \mathbf{E}\left[\mathbb{G}(x+\mu_\star)\mathbb{G}(-y+\mu_\star))\right] \\ &\quad + \mathbf{E}\left[\mathbb{G}(-x+\mu_\star)\mathbb{G}(y+\mu_\star)\right] + \mathbf{E}\left[\mathbb{G}(-x+\mu_\star)\mathbb{G}(-y+\mu_\star))\right] \\ &= \rho(x+\mu_\star, y+\mu_\star) + \rho(x+\mu_\star, -y+\mu_\star) \\ &\quad + \rho(-x+\mu_\star, y+\mu_\star) + \rho(-x+\mu_\star, -y+\mu_\star)\ .\end{aligned}$$

Therewith,

$$\mathcal{G} \sim \mathcal{N}(0, \mathbf{\Sigma})\ ,$$

where the covariance matrix $\mathbf{\Sigma}$ is defined by

$$\mathbf{\Sigma} = \frac{4}{p_\star^2}\iint \bar\rho(x,y)\,\partial_\vartheta D(x;\vartheta_\star)^t \partial_\vartheta D(y;\vartheta_\star)\ H(dx)H(dy)\ , \tag{3.4.19}$$

cf. Ibragimov and Rozanov [15], (4.13).

By (3.4.16) it follows that $\sqrt{n}(\hat\vartheta_n - \vartheta_\star)$ converges in distribution to a zero mean Gaussian random vector with covariance matrix $\boldsymbol{I}^{-1}\mathbf{\Sigma}\boldsymbol{I}^{-1}$. □

Theorem 3.30. *Under assumption 3.24 the random sequence*

$$\sqrt{n}\big(\hat\vartheta_n - \vartheta_\star, \hat F_n - F\big) \quad , \quad n \in \mathbb{N}$$

converges in distribution to a zero mean Gaussian process with covariance function $\Gamma_\star$ *given in (3.4.22).*

Proof. First, recall that $\tilde F_n$ and $\hat F_n$ are interchangeable at rate $\sqrt{n}$, cf. corollary 3.16, what allows to consider the sequence $\sqrt{n}(\tilde F_n - F)$ instead of $\sqrt{n}(\hat F_n - F)$. For all $x \in \mathbb{R}$,

$$\begin{aligned}F(x) &= p_\star^{-1}\big(H(x+\mu_\star) - (1-p_\star)F_0(x-\nu_\star+\mu_\star)\big)\ , \\ \tilde F_n(x) &= \hat p_n^{-1}\big(\tilde H_n(x+\hat\mu_n) - (1-\hat p_n)F_0(x-\hat\nu_n+\hat\mu_n)\big)\ .\end{aligned}$$

Regarding $\tilde F_n(x) = \tilde F_n(x;\hat\vartheta_n)$ as a function of ϑ, its first order derivative, for all $x \in \mathbb{R}$, is given by

$$\partial_\vartheta \tilde F_n(x;\vartheta)^t = \begin{pmatrix} p^{-2}\big(F_0(x-\nu+\mu) - \tilde H_n(x+\mu)\big) \\ p^{-1}(1-p)f_0(x-\nu+\mu) \\ p^{-1}\big(\tilde h_n(x+\mu) - (1-p)f_0(x-\nu+\mu)\big)\end{pmatrix}.$$

Hence, by a zero order Taylor expansion of $\tilde{F}_n$ around $\vartheta_\star$ we obtain

$$\begin{aligned}
&\sqrt{n}\big(\hat{F}_n(x) - F(x)\big)\\
&= \sqrt{n}\big(\tilde{F}_n(x;\hat{\vartheta}_n) - F(x)\big) + o_{a.s.}(1)\\
&= \sqrt{n}\big(\tilde{F}_n(x;\vartheta_\star) + \partial_\vartheta \tilde{F}_n(x;\bar{\vartheta}_n)(\hat{\vartheta}_n - \vartheta_\star)^t - F(x)\big) + o_{a.s.}(1)\\
&= \sqrt{n}\, p_\star^{-1}\big(\hat{H}_n(x+\mu_\star) - H(x+\mu_\star)\big)\\
&\quad + \sqrt{n}(\hat{p}_n - p_\star)(\bar{p}_n)^{-2}\big(F_0(x - \bar{\nu}_n + \bar{\mu}_n) - \tilde{H}_n(x+\bar{\mu}_n)\big)\\
&\quad + \sqrt{n}(\hat{\nu}_n - \nu_\star)(\bar{p}_n)^{-1}(1-\bar{p}_n) f_0(x - \bar{\nu}_n + \bar{\mu}_n)\\
&\quad + \sqrt{n}(\hat{\mu}_n - \mu_\star)(\bar{p}_n)^{-1}\big(\tilde{h}_n(x+\bar{\mu}_n) - (1-\bar{p}_n) f_0(x - \bar{\nu}_n + \bar{\mu}_n)\big)\\
&\quad + o_{a.s.}(1)\ .
\end{aligned}$$

for some $\bar{\vartheta}_n \in \{\vartheta_\star + t(\hat{\vartheta}_n - \vartheta_\star) \mid t \in (0,1)\}$, $n \in \mathbb{N}$. As almost surely $\bar{\vartheta}_n$ converges to $\vartheta_\star$, $\tilde{H}_n$ converges to H, $\tilde{h}_n$ converges to h, and $\sqrt{n}(\hat{\vartheta}_n - \vartheta_\star)$ is tight, cf. lemma 3.29, we have, uniformly in x,

$$\begin{aligned}
\sqrt{n}\big(\hat{F}_n(x) - F(x)\big) &= \sqrt{n}\, p_\star^{-1}\big(\hat{H}_n(x+\mu_\star) - H(x+\mu_\star)\big)\\
&\quad + \sqrt{n}(\hat{\vartheta}_n - \vartheta_\star)\boldsymbol{u} + o_p(1)\ ,
\end{aligned} \tag{3.4.20}$$

where the deterministic vector $\boldsymbol{u}$ is defined as

$$\boldsymbol{u} = \begin{pmatrix} p_\star^{-2}\big(F_0(x - \nu_\star + \mu_\star) - H(x+\mu_\star)\big) \\ p_\star^{-1}(1-p_\star) f_0(x - \nu_\star + \mu_\star) \\ f_\star(x) \end{pmatrix} .$$

Using (3.4.16) and (3.4.20) we obtain

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta_\star, \hat{F}_n(x) - F(x)) = \sqrt{n}\big(\mathcal{U}_n, \hat{H}_n(x+\mu_\star) - H(x+\mu_\star)\big)\boldsymbol{J} + o_p(1)$$

as an element in $\mathbb{R}^3 \times D(\mathbb{R})$, where

$$\boldsymbol{J} = \begin{pmatrix} \boldsymbol{I}^{-1} & \boldsymbol{I}^{-1}\boldsymbol{u} \\ 0 & p_\star^{-1} \end{pmatrix} . \tag{3.4.21}$$

Introducing the map $\varphi : BV_1(\mathbb{R}) \to D(\mathbb{R})$ according to

$$\varphi(F(x)) = F(x+\mu_\star) - H(x+\mu_\star)\ ,$$

which is Hadamard-differentiable at H with derivative

$$\varphi'_H(F(x)) = F(x+\mu_\star)\ ,$$

we rewrite

$$\big(\mathcal{U}_n, \hat{H}_n(x+\mu_\star) - H(x+\mu_\star)\big) = (\phi \circ \psi, \varphi)(\hat{H}_n)\ ,$$

cf. (3.4.17). By means of the Delta-method, theorem 3.13, as $\sqrt{n}\big(\hat{H}_n - H\big) \rightsquigarrow \mathbb{G}$, $(\phi \circ \psi, \varphi)(H) = 0$, and using (3.4.18), we conclude

$$\sqrt{n}\big(\mathcal{U}_n, \hat{H}_n(x+\mu_\star) - H(x+\mu_\star)\big) \rightsquigarrow \big((\phi \circ \psi)'_H, \varphi'_H\big)(\mathbb{G}) = \big(\mathcal{G}, \mathbb{G}(x+\mu_\star)\big) \ .$$

The covariance function of the process $\big(\mathcal{G}, \mathbb{G}(x+\mu_\star)\big)_{x\in\mathbb{R}}$ is given by

$$\Gamma(x,y) = \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\sigma}(y)^t \\ \boldsymbol{\sigma}(x) & \varrho(x,y) \end{pmatrix} ,$$

where the matrix $\boldsymbol{\Sigma}$ is defined in (3.4.19), the vector valued function $\boldsymbol{\sigma}$ is defined by

$$\begin{aligned} \boldsymbol{\sigma}(x) &= \mathbf{E}\left[\mathcal{G}\mathbb{G}(x+\mu_\star)\right] \\ &= \frac{2}{p_\star} \int \big(\rho(x+\mu_\star, y+\mu) + \rho(x+\mu_\star, -y+\mu_\star)\big)\partial_\vartheta D(y;\vartheta_\star)\, dy \ , \end{aligned}$$

and ϱ, the autocovariance function of the process $\big(\mathbb{G}(x+\mu_\star)\big)_{x\in\mathbb{R}}$, is

$$\varrho(x,y) = \rho(x+\mu_\star, y+\mu_\star) \ .$$

Recall that ρ is the covariance function of the process $\mathbb{G}$, cf. corollary 3.9. Altogether we conclude that the random process $\sqrt{n}\big(\hat{\vartheta}_n - \vartheta_\star, \hat{F}_n - F\big)$ converges in distribution to a zero mean Gaussian process with covariance function

$$\Gamma_\star(x,y) = \boldsymbol{J}^t\Gamma(x,y)\boldsymbol{J} \ , \tag{3.4.22}$$

where $\boldsymbol{J}$ is given in (3.4.21). □

4. Simulation study

Finally, we want to carry out a numerical validation of the estimator $\hat{\vartheta}_n$ suggested in section 3.1 using the free software environment `R` for statistical computing. The implementation of the estimator is straightforward. Based on the independent copies $X_1, \ldots, X_n$, given a kernel function k and the bandwidth b_n, the smooth empirical cdf $\tilde{H}_n$ can be written as

$$\begin{aligned}\tilde{H}_n(y) &= \int_{-\infty}^{y} \frac{1}{nb_n} \sum_{i=1}^{n} k\Big(\frac{x - X_i}{b_n}\Big)\, dx = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{y} \frac{1}{b_n}\, k\Big(\frac{x - X_i}{b_n}\Big)\, dx \\ &= \frac{1}{n} \sum_{i=1}^{n} K\Big(\frac{y - X_i}{b_n}\Big)\ ,\end{aligned}$$

denoting by K the antiderivative of k. Choosing a Gaussian kernel we use the function `pnorm` for the evaluation of K. This allows an easy bottom-up implementation of $\tilde{H}_n$, $\tilde{D}_n$, and $\hat{d}_n$. The minimization of $\hat{d}_n$ is done by a *quasi-Newton* method applying the function `optim` with option `method='BFGS'`. Further, the following calculations use the bandwidth rate $b_n = n^{-1.6}$. In combination with the Gaussian kernel this rate meets all the assumptions 3.20 (ii)-(iv), which ensure a consistent estimation. Also, it performed best during initial tests using the rates $n^{-\alpha}$ for $\alpha = 0.5, 1, 1.6, 2$.

Estimating a normal mixture

We assess the actual performance of the estimator $\hat{\vartheta}_n$ by way of comparison using the results obtained by Bordes et al. [4] when estimating the normal mixture

$$x \mapsto (1 - p)\varphi(x; 4) + p\varphi(x - \mu; 1)\ , \tag{4.0.1}$$

where $x \mapsto \varphi(x; \sigma^2)$ denotes the centered Gaussian pdf with variance σ^2. Here, in conformity with our model, the component pdf $f_0(x) = \varphi(x; 4)$ is assumed to be known. Given a n-sample drawn from mixture (4.0.1) they estimated the parameters p and μ for true values $(p_\star, \mu_\star) = (0.3, 3)$ and $(p_\star, \mu_\star) = (0.15, 3)$. The corresponding mixtures are illustrated in figure 4.1. Using sample sizes $n = 250$ and $n = 1000$ they each time performed 200 estimations. The estimates thus obtained are given in table 4.1, aggregated to their means and standard deviations.

We repeated these estimations in much the same manned except for applying the estimator $\hat{\vartheta}_n$ established in section 3.1, that is we additionally determine estimates for the second, latent location parameter ν with true value $\nu_\star = 0$. The performed

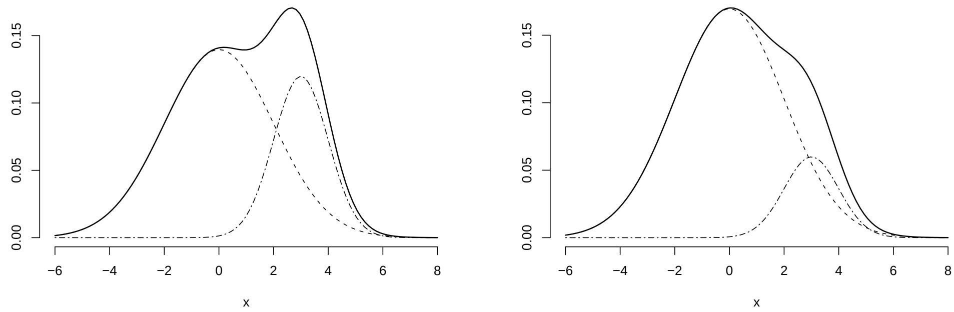

Figure 4.1.: Mixture (4.0.1) with true values $(p_\star, \mu_\star) = (0.3, 3)$ on the left-hand and $(p_\star, \mu_\star) = (0.15, 3)$ on the right-hand side. Each plot shows the mixture pdf (solid) with corresponding known (dashed) and unknown (two-dashed) component pdfs.

estimations are all based on the initial value $\vartheta_0 = (0.2, 0, 0)$, which was passed to the optimization routine `optim`. As previous implementations attested, however, the choice of this initial value does not influence the estimation method significantly.

The outcomes we obtained are much plausible as they nearly match those by Bordes et al. [4]. The estimates can be taken from table 4.2. At large, our extended method seems to be slightly inferior. This is as was expected, though, since simultaneously estimating an additional parameter means further efforts, as well. For a theoretical support we take a closer look at the asymptotic covariance matrix of the residuals $\sqrt{n}(\hat{\vartheta}_n - \vartheta_\star)$, provided by lemma 3.29. Based on the true parameter $(p_\star, \nu_\star, \mu_\star) = (0.3, 0, 3)$ numerical calculations give

$$\boldsymbol{I} = \begin{pmatrix} 10.200 & 1.649 & 1.469 \\ 1.649 & 0.314 & 0.180 \\ 1.469 & 0.180 & 0.308 \end{pmatrix} , \quad \boldsymbol{\Sigma} = \begin{pmatrix} 49.033 & 7.893 & 7.194 \\ 7.893 & 1.318 & 1.092 \\ 7.194 & 1.092 & 1.159 \end{pmatrix} ,$$

and therewith

$$\boldsymbol{I}^{-1}\boldsymbol{\Sigma}\boldsymbol{I}^{-1} = \begin{pmatrix} 1.996 & -6.248 & -4.287 \\ -6.248 & 28.756 & 13.459 \\ -4.287 & 13.459 & 16.995 \end{pmatrix} .$$

Canceling the second row and the second column of $\boldsymbol{I}$ and $\boldsymbol{\Sigma}$ leads to the asymptotic covariance matrix corresponding to the estimator applied by Bordes et al. [4]. Denoting the reduced matrices by $\bar{\boldsymbol{I}}$ and $\bar{\boldsymbol{\Sigma}}$ we obtain

$$\bar{\boldsymbol{\Sigma}}^{-1}\bar{\boldsymbol{\Sigma}}\bar{\boldsymbol{I}}^{-1} = \begin{pmatrix} 0.664 & -1.472 \\ -1.472 & 11.158 \end{pmatrix} .$$

Comparing the upper left and the lower right entries of these two covariance matrices in fact substantiates the inferiority of the extended estimator $\hat{\vartheta}_n$.

Table 4.1.: Estimates obtained by Bordes et al. [4]. The table gives the mean values and standard deviations (in brackets) of 200 estimates for true values $(p_\star, \mu_\star) = (0.15, 3)$ and $(p_\star, \mu_\star) = (0.3, 3)$ and sample sizes $n = 250$ and $n = 1000$.

	$\hat{p}_n$	$\hat{\mu}_n$
$p = 0.15$ / $n = 250$	0.165 (0.055)	2.878 (0.418)
$p = 0.15$ / $n = 1000$	0.154 (0.031)	2.944 (0.272)
$p = 0.30$ / $n = 250$	0.303 (0.057)	2.963 (0.226)
$p = 0.30$ / $n = 1000$	0.301 (0.029)	2.976 (0.131)

Table 4.2.: Estimates obtained using the estimator $\hat{\vartheta}_n$. The table gives the mean values and standard deviations (in brackets) of 200 estimates for true values $(p_\star, \nu_\star, \mu_\star) = (0.15, 0, 3)$ and $(p_\star, \nu_\star, \mu_\star) = (0.3, 0, 3)$ and sample sizes $n = 250$ and $n = 1000$.

	$\hat{p}_n$	$\hat{\nu}_n$	$\hat{\mu}_n$
$p = 0.15$ / $n = 250$	0.161 (0.021)	-0.024 (0.222)	2.838 (0.691)
$p = 0.15$ / $n = 1000$	0.157 (0.017)	-0.025 (0.095)	2.954 (0.242)
$p = 0.30$ / $n = 250$	0.303 (0.012)	-0.017 (0.238)	2.959 (0.359)
$p = 0.30$ / $n = 1000$	0.303 (0.009)	-0.009 (0.098)	2.990 (0.095)

5. Outlook

Throughout this thesis we have considered two generalized versions of the semiparametric mixture model treated by Bordes, Delmas, and Vandekerkhove [4] and Bordes and Vandekerkhove [3], namely an additional location parameter on the one hand and an additional scale parameter on the other hand was added to their mixture. We worked for conditions on the model parameter providing identifiability in these generalized setups and supplied an adapted estimator in the location parameter context, which was proved to be strongly consistent for its Euclidean part and asymptotically normal as a whole.

Of course, a next step should be the simultaneous consideration of both extensions, i. e. we only assume the "known" mixture component to belong to a predefined location-scale family. This provides more flexibility and is reasonable from a practical point of view. Absolutely, the crucial part will then be the derivation of suitable identifiability results for the corresponding mixture

$$m(x;p,\nu,\sigma,\mu,f) = (1-p)f_0\big((x-\nu)/\sigma\big)/\sigma + pf(x-\mu)\ .$$

So far it is unclear if observing the moments of m provides satisfactory conclusions as compared with lemma 2.15, for instance. Considering multivariate data such as repeated measurements could perhaps be a source of additional information, cf. Hall and Zhou [11] and more recently Kasahara and Shimotsu [16]. Also, regarding conditional mixtures based on the presence of covariates could assist, cf. Kitamura [17] and Henry, Kitamura, and Salanié [12]. After that, the estimator $\hat{\theta}_n$ can be adapted canonically. Provided that the mixture pdf induced by the true parameter $\theta_\star$ is identifiable, consistency results and the asymptotic normality of the extended estimator should follow by similar arguments as in sections 3.3 and 3.4.

Bibliography

[1] H. Bauer. *Wahrscheinlichkeitstheorie.* De-Gruyter-Lehrbuch. Walter de Gruyter, Berlin, New York, fifth edition, 2002.

[2] A.F. Beardon. *Complex Analysis: The Argument Principle in Analysis and Topology.* A Wiley-Interscience Publification. John Wiley & Sons, Chichester, New York, Singapore, 1979.

[3] L. Bordes and P. Vandekerkhove. Semiparametric two-component mixture model with a known component: An asymptotically normal estimator. *Preprint*, 2010.

[4] L. Bordes, C. Delmas, and P. Vandekerkhove. Semiparametric estimation of a two-component mixture model where one component is known. *Scandinavian Journal of Statistics*, vol. 33: 733–752, 2006.

[5] L. Bordes, S. Mottelet, and P. Vandekerkhove. Semiparametric estimation of a two-component mixture model. *The Annals of Satatistics*, vol. 34(3): 1204–1232, 2006.

[6] C. Corduneanu. *Almost Periodic Functions.* Number 22 in Interscience Tracts in Pure and Applied Mathematics. John Wiley & Sons, New York, London, Sydney, Toronto, 1961.

[7] I.R. Cruz-Medina, T.P. Hettmansperger, and H. Thomas. Semiparametric mixture models and repeated measures: The multinomial cut point model. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 53(3): 463–474, 2004.

[8] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39(1): 1–38, 1977.

[9] E. Giné and A. Guillou. Rates of strong uniform consistency for multivariate kernel density estimators. *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*, vol. 38(6): 907–921, 2002.

[10] P. Hall. On the non-parametric estimation of mixture proportions. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 43(2): 147–156, 1981.

[11] P. Hall and X.-H. Zhou. Nonparametric estimation of component distributions in a multivariate mixture. *The Annals of Statistics*, vol 31(1): 201–224, 2003.

[12] M. Henry, Y. Kitamura, and B. Salanié. Identifying finite mixtures in econometric models. Discussion Papers 0910-20, Columbia University, Department of Economics, 2010.

[13] T.P. Hettmansperger and H. Thomas. Almost nonparametric inference for repeated measures in mixture models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 62(4): 811–825, 2000.

[14] D.R. Hunter, S.W. Wang, and T.P. Hettmansperger. Inference for mixtures of symmetric distribution. *The Annals of Statistics*, vol. 35(1): 224–251, 2007.

[15] I.A. Ibragimov and Y.A. Rozanov. *Gaussian Random Processes*. Springer, New York, Heidelberg, Berlin, 1978.

[16] H. Kasahara and K. Shimotsu. Nonparametric identification and estimation of multivariate mixtures. Working Papers 1153, Queen's University, Department of Economics, 2007.

[17] Y. Kitamura. Nonparametric identifiability of finite mixtures. Unpublished manuscript, Yale University, Department of Economics, 2003.

[18] G.J. McLachlan and D. Peel. *Finite Mixture Models*. Wiley series in probability and statistics. John Wiley & Sons, New York, 2000.

[19] S. Newcomb. A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics*, vol. 8: 343–366, 1886.

[20] K. Pearson. Contributions to the theory of mathematical evolution. *Philosophical Transactions of the Royal Society of London A*, vol. 185: 71–110, 1894.

[21] R.C. Rao. The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 10(2): 159–203, 1948.

[22] A.R. Shorack and J.A. Wellner. *Empirical Processes with Applications to Statistics*. Wiley series in probability and mathematical statistics. John Wiley & Sons, New York, 1986.

[23] B.W. Silverman. Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.*, vol. 6(1): 177–184, 1978.

[24] D.M. Titterington. Minimum distance non-parametric estimation of mixture proportions. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 45(1): 37–46, 1983.

[25] A.W. van der Vaart. *Asymptotic Statistics*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, New York, 2000.

[26] A.W. van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer, New York, 2000.

A. Notation

A.1. Total and partial derivatives

If $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto f(x)$ is differentiable, then we denote its derivative with respect to the argument x by $\partial_x f$,

$$\partial_x f = \frac{d}{dx} f(x) \ .$$

Thus, $\partial_x f(x_0)$ gives the derivative of f evaluated at the particular argument x_0,

$$\partial_x f(x_0) = \frac{d}{dx} f(x)\big|_{x=x_0} \ .$$

Let $g : \mathbb{R}^n \to \mathbb{R}$ be totally differentiable. Then, the partial derivative of g with respect to the argument x_k, $k = 1, \ldots, n$, evaluated at $x_0 \in \mathbb{R}^n$, is denoted by

$$\partial_{x_k} g(x_0) \ .$$

The total derivative of g is defined by the $(1 \times n)$-matrix

$$\partial_x g(x_0) := \big(\partial_{x_1} g(x_0), \ldots, \partial_{x_n} g(x_0)\big) \ .$$

Hence, we regard the derivative of g as a row vector.

Finally, let $h : \mathbb{R}^n \to \mathbb{R}^m$ be totally differentiable. Rewriting

$$h(x) = (h_1(x), \ldots, h_m(x)) \quad , \quad x \in \mathbb{R}^n \ ,$$

where $h_k : \mathbb{R}^n \to \mathbb{R}$, $k = 1, \ldots, m$, the total derivative of h, evaluated at $x_0 \in \mathbb{R}^n$, is given by the $(m \times n)$-matrix

$$\partial_x h(x_0) = \begin{pmatrix} \partial_x h_1(x_0) \\ \vdots \\ \partial_x h_m(x_0) \end{pmatrix} \ .$$

A.2. Stochastic Landau symbols

Landau symbols, or the "big-oh" and "small-oh" notations, are common to express bounds and the asymptotic behavior of random sequences. Let $(r_n)_{n\in\mathbb{N}}, (X_n)_{n\in\mathbb{N}} \subset \mathbb{R}$, X_n, $n \in \mathbb{N}$ be stochastic.

We say that $X_n = o_{a.s.}(1)$ if X_n converges almost surely to zero as n tends to infinity, i. e.

$$\mathbf{P}\left[\lim_n X_n = 0\right] = 1 \ .$$

X_n is said to be $o_{a.s.}(r_n)$ if there exists another random sequence $(Y_n)_{n\in\mathbb{N}}$ such that

$$|X_n| \leq r_n Y_n \ , \ n \in \mathbb{N} \quad , \quad Y_n = o_{a.s.}(1) \ .$$

If r_n is positive for large n, then we obtain

$$X_n = o_{a.s.}(r_n) \quad \Longleftrightarrow \quad r_n^{-1} X_n = o_{a.s.}(1) \ .$$

If X_n is almost surely bounded, then we say $X_n = O_{a.s.}(1)$. Whenever $|X_n| \leq r_n$, $n \in \mathbb{N}$ almost surely we say that $X_n = O_{a.s.}(r_n)$. Of course, if r_n is positive for large n, then we likewise obtain

$$X_n = O_{a.s.}(r_n) \quad \Longleftrightarrow \quad r_n^{-1} X_n = O_{a.s.}(1) \ .$$

Similarly, we say that $X_n = o_p(1)$ if X_n converges stochastically to zero as n tends to infinity, i. e.

$$\lim_{n\to\infty} \mathbf{P}\left[|X_n| > \varepsilon\right] = 0 \quad , \quad \varepsilon > 0 \ ,$$

as well as $X_n = o_p(r_n)$ if $|X_n| \leq r_n Y_n$ for all $n \in \mathbb{N}$ and $Y_n = o_p(1)$, where again

$$X_n = o_p(r_n) \quad \Longleftrightarrow \quad r_n^{-1} X_n = o_p(1)$$

for positive r_n. If X_n is stochastically bounded (tight), that is for all $\varepsilon > 0$ there exists a compact K such that

$$\mathbf{P}\left[X_n \notin K\right] < \varepsilon \quad , \quad n \in \mathbb{N} \ ,$$

then we say $X_n = O_p(1)$. Further, we say $X_n = O_p(r_n)$ if $|X_n| \leq r_n Y_n$, $n \in \mathbb{N}$ and $Y_n = O_p(1)$. Once more, for positive r_n, we have

$$X_n = O_p(r_n) \quad \Longleftrightarrow \quad r_n^{-1} X_n = O_p(1) \ .$$

B. Declaration

I hereby declare that this thesis has been composed by myself and describes my own work unless otherwise indicated in the text. All sources of information have been specifically acknowledged.

Marburg, August 23, 2010

Daniel Hohmann