VU VRIJE UNIVERSITEIT AMSTERDAM

VU Research Portal

Asymptotics in Deconvolution Models

Donauer, S.

2009

document version
Publisher's PDF, also known as Version of record

Link to publication in VU Research Portal

citation for published version (APA)
Donauer, S. (2009). *Asymptotics in Deconvolution Models: Approximating Perfect Knowledge*.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl

Download date: 21. Jan. 2022

One

Introduction

This chapter provides some insight into the research area and the contributions to it made by this work without dwelling on the mathematical details presented in the subsequent chapters. The model studied in this thesis -the Deconvolution Model- is introduced in Section 1.1 among some examples of inverse models. A few general thoughts about estimation procedures in inverse models together with the definition of the Isotonic Inverse- and the Maximum Likelihood Estimator are given in Section 1.2. Moreover, the challenges of deriving their asymptotic behavior are pointed out in Section 1.3.

'... God is perfect in knowledge.' (Job 37:16 [NIV])

Contrary, human beings are always limited in their knowledge. This fact actually gives rise to the questions and models we are going to discuss in this thesis in a twofold way.

Statisticians collect information (data, observations) to describe, conclude about or predict a certain quantity of interest. The more information is available, the better the description, conclusion and prediction. An infinite amount of data would correspond to the situation of full knowledge which we unfortunately never meet according to the previous paragraph. Therefore, Statisticians examine this situation from a theoretical point of view: the available limited knowledge is used to study the behavior or changes of the description, conclusion or prediction as the amount of data increases. This procedure of *approximating perfect knowledge*, known as asymptotics, will get a lot of attention in the sequel.
Moreover, we consider models where the 'dilemma' of the limited knowledge leads to another challenge. Imagine you are interested in the time point when patients turn from HIV negative to HIV positive. Instead of this time point, which can never be observed exactly, one can obtain information about the last consultation before and the first one after that time point. Or imagine you want to study the growth of a three dimensional brain tumor over time. Here, the available information might come from the sizes of two dimensional projections on pictures done by an Magnetic Resonance Imaging (MRI). Or imagine you are interested in the body height of 14 year old girls living in Amsterdam. But maybe due to cost issues you are forced to use already available data about 16 year old girls. In all these examples, certain mechanisms prevent us from observing the quantity of interest. Whenever such a lack of observations of the variable of interest is present whereas at the same time observations of a related quantity can be obtained, we talk in this document about an inverse model. And one particular example is studied in this text in great detail.

1.1 Inverse Models

As the above examples already suggest, inverse models arise in various areas of applications. Biology, engineering, stereology, image- or survival analysis and ornithology are only a few examples. Hence, in order to be able to use realistic models that represent the special structure of these inverse models, there is a need to meet the mathematical challenges that come along with them. We describe two different inverse models to illustrate their diversity and proceed then by examing the *Deconvolution Model*.

Thick Slice Wicksell Model

How do Torpedos[1] navigate and communicate? It is known that they use an electric field generated by their electric organ. To understand how this organ functions, it is necessary to study the so-called synapses that govern the communication between nerve and target cells. Synaptic vesicles contained in the synapses store and/or release neurotransmitters that cause a reaction in the receiving postsynaptic cell. The amount of released transmitters influences the size of the response and the possible amount of stored (and therefore released) transmitters is naturally limited by the size of these vesicles. A change of its size through genetic or physiological alteration would influence the reaction and eventually the behavior of the Torpedo. Hence, one is interested in the size of these three dimensional synaptic vesicles to understand the functionality of the electric organ. In order to obtain information about the vesicles, Fox (1988) proposes to dissect slices of 118nm of tissue blocks of 2mm x 2mm x 2mm from the electric organ. With the help of an electronic microscope two dimensional spherical, membrane bounded profiles are uncovered as projections of the vesicles contained in the tissue. A naive and unrealistic (but in applications sometimes used) approach would be to interpret the size (radius) of the two dimensional profile as the size (radius) of the vesicle itself. This is motivated by the fact, that one would obtain a setting where direct observations are available and classical statistical techniques are applicable.

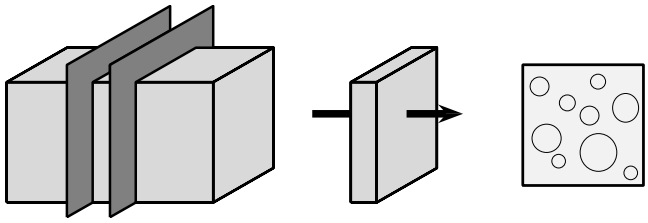

Figure 1.1: Obtaining observations in the thick slice Wicksell model.

However, still using the sizes of the two dimensional profiles, but seeing them as indirect observations of the sizes of the three dimensional vesicles, yields more realistic results. In this situation we use an inverse model, the thick slice Wicksell model (see Bach, 1967) as a generalization of the classical Wicksell model introduced in Wicksell (1925). In the more general model, observations are given by two dimensional projections obtained from projecting regularly shaped three dimensional objects contained in a slice of an opaque medium (see Figure 1.1). This procedure could be applied for studying the size of individual zymogen granules, i.e. intracellular secretory granules that play a role in

[1] Type of ray; sidderrog (Dutch); Zitterrochen (German)

secretion, see Liebow and Rothman (1973). The same holds for calculating the mean fast cell weight of adipose people. Assuming spherical fast cells, the diameters of them contained in a tissue slice can be examined by the thick slice Wicksell model as done in Sjöström et al. (1971). Polymer scientist (see Maestrini et al., 1992) want to investigate the toughness of two phase materials. For instance, the toughness of plastic is determined by the amount of small rubber particles. Since some production procedures do not allow to know the mixing proportion in advance, it has to be derived from the final product using observable projections of the rubber particles. In material science one might be also interested in generalizations by dealing with cubic particles (see Ohser and Sandau, 2000).

Deconvolution Model

What is the probability that a 14 year old girl in Amsterdam is smaller than 1.50m? The variable of interest is obviously the body height of the 14 year old girls. Consider for the sake of this example the situation that for some reason you only have access to data about the body height of 16 year olds. In a first simple and naive attempt one might assume a deterministic growth of 6cm for every girl between the age of 14 and 16 independent of their height at age 14. Hence, subtracting 6cm from each of the given observations will lead to a 'reconstructed' data set for the body height of 14 year old girls. Using the not very realistic assumption about the growth, we transformed the problem to a classical setup where direct observations of the variable of interest are available.
A more realistic approach would be to model the growth by a random variable Y whose distribution has certain properties, i.e. $X^{16} = X^{14} + Y$ where X^{14} and X^{16} denote the body height of 14 and 16 year olds, respectively, and $Y \geq 0$ the random growth, independent of X^{14}. Subtraction of Y from X^{16} is no longer possible. Therefore, observations of X^{14} can no longer be obtained explicitly. Hence we have to deal with indirect observations which can be done using results of this thesis.
This example is an inverse model, a Deconvolution Model, to be more specific. In deconvolution models the given observations are interpreted as the sum of the variable of interest and some independent error or noise. The random growth in the previous example corresponds to that noise variable. Measurement errors would be another example, viewing the measured length of an object as its true length plus the measurement error. In image analysis the blurring effect of a picture can also be modeled by such a noise variable. In digital communication, distortion occurs between the source of the signal and the receiver due to the transmission by for example a phone line. The noise would then correspond to that distortion.

We will use the following general notation for inverse models. The variable of interest (which cannot be observed) is denoted by X and possesses an unknown distribution function $F_0 \in \mathcal{F}$, the object we will study later on, where $\mathcal{F}$ denotes the set of all distribution functions on $\mathbb{R}$. However, throughout this document we restrict ourselves to the case where F_0 belongs to the subset $\mathcal{F}_{[0,\infty)}$ consisting of all distribution functions

on $[0,\infty)$. The density of the observable variable Z is then given by $h_0 = h_{F_0}$ which can be interpreted as $T(F_0)$ for some transformation

$$T : \mathcal{F}_{[0,\infty)} \to \mathcal{H},\ T(F) = h_F, \tag{1.1.1}$$

specified by the model.
In terms of the deconvolution models, one considers the situation of having a sample $Z_1,\dots,Z_n$ where each observation Z_i can be expressed as the sum $X_i + Y_i$, using independent samples $X_1,\dots,X_n$ and $Y_1,\dots,Y_n$. The random variables $X_1,\dots,X_n$ are iid and have an unknown distribution function F_0, whereas the noise density g of the iid random variables $Y_1,\dots,Y_n$ is known. Hence the observable sample $Z_1,\dots,Z_n$ follows the observation density

$$h_0(z) \;=\; h_{F_0}(z) \;=\; [g * dF_0](z) \;=\; \int_0^\infty g(z-x)\,dF_0(x),\ z \in \mathbb{R}, \tag{1.1.2}$$

the convolution of F_0 and g. Here, expressed in the framework of (1.1.1), the functional T is given by the convolution operator, i.e. $T(F) = g * dF$, where $g * dF$ is as in (1.1.2), so that $h_0 = T(F_0)$ is contained in $\mathcal{H} = \{h_F \,|h_F = T(F) = g * dF : F \in \mathcal{F}_{[0,\infty)}\}$.

1.2 Estimation Procedures

Our aim is to examine nonparametric estimators of the unknown F_0, f_0 or some functional thereof from an asymptotical point of view. All estimators are based on $Z_1,\dots,Z_n$. The structure of equation (1.1.2) reveals why this is referred to as *deconvolution* in the deconvolution model. Note that once an estimator for F_0 is derived, an estimator for f_0 could be defined using kernel smoothing. Vice versa, a density estimator for f_0 can be integrated to obtain an estimator for F_0. Functionals $K(F_0)$ are usually estimated by $K(\hat{F})$, where $\hat{F}$ is some estimator of F_0. In what follows, we will focus on estimating F_0 nonparametrically.

Different estimation procedures have been proposed in the literature for inverse models, all making use of the special structure of these models (references can be found in Section 2.1). We now introduce two concepts. The Isotonic Inverse Estimator (IIE) uses the explicit inverse transformation T^{-1} (if existent, see (1.1.1)). It can be seen as a special plug-in estimator, requiring no a priori shape constraint on $\mathcal{F}_{[0,\infty)}$ and, contrary to other plug-in procedures, no smoothing parameter. The nonparametric Maximum Likelihood Estimator (MLE) is given as the optimizer of a certain random criterion function and is studied in this text as an alternative estimator to the known IIE in the deconvolution setting.

Isotonic Inverse Estimator (IIE)

The Isotonic Inverse Estimator (IIE) can be seen as a nonparametric plug-in estimator that, contrary to the MLE, has an explicit representation. Assume the existence of an explicit inverse relation $T^{-1} : \mathcal{H} \to \mathcal{F}_{[0,\infty)}$ of (1.1.1). A reasonable plug-in estimator for F_0 would be given by $T^{-1}(\hat{H}_n^I) \in \mathcal{F}_{[0,\infty)}$ for an estimator $\hat{H}_n^I$ of H_0 contained in $\mathcal{H}$. This situation is illustrated in the left picture of Figure 1.2 where the oval represents $\mathcal{F}_{[0,\infty)}$ and the rather oddly shaped set on the right side the set $\mathcal{H}$, both subsets of bigger classes of functions given by the rectangles. It is this shape of $\mathcal{H}$ that might cause problems in practical applications. It could, for example, be a set of functions that satisfy certain order constraints or are smooth in some way. However, there exist examples where $\mathcal{H}$ can only be described implicitly so that choosing $\hat{H}_n^I \in \mathcal{H}$ is rather difficult.

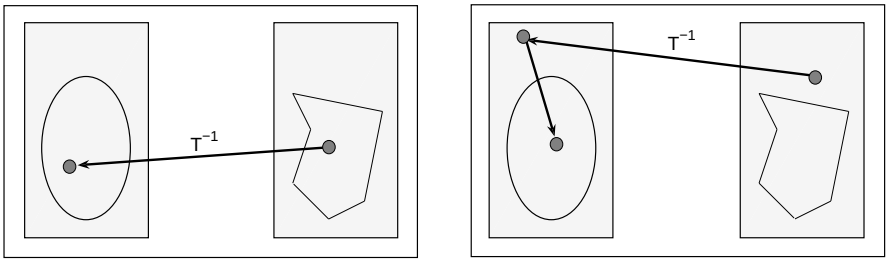

Figure 1.2: Illustration of the definition of the IIE.

The IIE $\tilde{F}_n$ is an *isotonized naive plug-in* estimator making no use of the shape of $\mathcal{H}$. The *naive plug-in* estimator is given by $F_n^{\text{naive}} = T^{-1}(\hat{H}_n^o)$ for some estimator $\hat{H}_n^o$ of H_0, not necessarily contained in the range of the operator T (assuming that T^{-1} can be extended to a larger set of functions containing $\hat{H}_n^o$, illustrated by the box in the right picture of Figure 1.2). The upper arrow in the right picture of Figure 1.2 corresponds to this step and already indicates that there is no guarantee that F_n^{naive} is actually contained in $\mathcal{F}_{[0,\infty)}$. In fact, often the monotonicity constraint of F_n^{naive} is violated. Therefore, $\tilde{F}_n$ is defined as the L_2-projection of F_n^{naive} onto the set of all distribution functions leading to a monotone estimator by definition. The second arrow in the same picture illustrates this. This projection coincides with the following isotonization, in particular useful for a noniterative computation of the estimator. Write (without caring about well-definedness)

$$F_n^{\text{naive}}(x) \;=\; \frac{\mathrm{d}}{\mathrm{dx}} \int_0^x F_n^{\text{naive}}(z)\, dz \;=\; \frac{\mathrm{d}}{\mathrm{dx}} \int_0^x \left[T^{-1}(\hat{H}_n^o)\right](z)\, dz,\ x \in [\,0, \infty).$$

Then it can be shown that the IIE can be written as

$$\tilde{F}_n(x) \;=\; \text{slogcom}\left\{ y \mapsto \int_0^y \left[T^{-1}(\hat{H}_n^o)\right](z)\, dz \right\}\Big|_{y=x} \tag{1.2.1}$$

where 'slogcom' denotes the right derivative (=slope) of the greatest convex minorant, i.e. of the greatest convex function that lies entirely below $y \mapsto \int_0^y \left[T^{-1}(\hat{H}_n^o)\right](z)\, dz$.

IIE in a Class of Deconvolution Models

For deconvolution models with decreasing kernels, the IIE has been studied in van Es et al. (1998). We illustrate its definition since it requires the computation of T^{-1} which will also be used in later chapters. Also, its asymptotic behavior will be compared to the one of the maximum likelihood estimator discussed in this document.

To compute T^{-1}, i.e. to express F in terms of h_F, we can invert $h_F = g * dF$ using a so-called resolvent function ϱ.

Definition 1.2.1 (Resolvent ϱ).
A function $\varrho : [\,0, \infty) \to [\,0, \infty)$ that solves the integral equation $[\varrho * g](x) = x$ for all $x \geq 0$ is called a resolvent (or reciprocal kernel) of g.

Note that for specific choices of g there exists an explicit formula for ϱ. For instance, in the exponential deconvolution model where $g(y) = \exp(-y)\mathbf{1}_{[\,0,\infty)}(y)$, the resolvent is given by $\varrho(x) = (1+x)\mathbf{1}_{[\,0,\infty)}(x)$. Choosing $g(y) = \mathbf{1}_{[\,0,1\,]}(y)$ yields $\varrho(x) = 1 + \lfloor x \rfloor$ for $x \geq 0$, where $\lfloor x \rfloor$ denotes the largest integer smaller or equal than x. However, these simple expressions for ϱ seem to be exceptional. More involved formulas using iterative representations are for example needed if $g_n(y) = n(1-y)^{n-1}\mathbf{1}_{[\,0,1\,]}(y)$ for $n = 2, 3, \ldots$ (see Appendix A for the cases $n = 2, 3$).

Lemma 1.2.2 (Inversion Formula).
Let ϱ be the resolvent as in Definition 1.2.1 for a density g that is bounded, decreasing and absolutely continuous. Then ϱ is differentiable on $(0, \infty)$ (except for at most at one point) and

$$F(x) \;=\; \frac{\mathrm{d}}{\mathrm{dx}} \left[\varrho * h_F\right](x) \;=\; [\varrho' * h_F](x) + \varrho(0) h_F(x),\; x \in [\,0, \infty). \tag{1.2.2}$$

Proof.
See Jongbloed (1995) for the differentiability of ϱ. Equality (1.2.2) is obtained after applying Leibniz rule of differentiation to

$$[\varrho * h_F](x) \;=\; [\varrho * g * dF](x) \;=\; [\mathrm{id} * dF](x) \;=\; \int_{[\,0,x\,]} F(y)\, dy. \quad \square$$

Due to (1.2.2) we have $T^{-1} : \mathcal{H} \to \mathcal{F}_{[0,\infty)}$, $T^{-1}(H) = [\varrho * dH]'$ with $\int_0^y \left[T^{-1}(H)\right](z)\,dz = \int_0^y [\varrho * dH]'(z)\,dz = [\varrho * dH](y)$ for $H \in \mathcal{H}$ and $y \geq 0$. Choosing $\hat{H}_n^o = H_n$ and extending T^{-1} to a class that contains H_n we obtain due to (1.2.1), for $x \geq 0$,

$$\tilde{F}_n(x) \quad = \quad \text{slogcom}\left\{y \mapsto [\varrho * dH_n](y)\right\}\Big|_{y=x}.$$

Example (Exponential Deconvolution).
Note that $[\varrho * dH_n](x) = \int_0^x (1 - x + z)\,dH_n(z)$ is, for $x \geq 0$, piecewise concave between successive observation points so that the IIE coincides with the 'slogcom' of the set $\{[\varrho * dH_n](z_i-), i = 1, \ldots, n\}$ with $[\varrho * dH_n](z_i-) = \lim_{h \downarrow 0}[\varrho * dH_n](z_i - h)$. This set is illustrated in Figure 1.3 together with its piecewise linear convex minorant using 50 observations and $F_0 = U[0, 5]$. Figure 1.4 shows the IIE $\tilde{F}_{50}$ as the derivative of this convex minorant together with F_0.

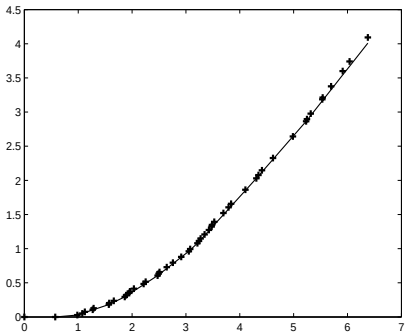

Figure 1.3: The convex minorant of $\{[\varrho * dH_{50}](z_i-), i = 1, \ldots, 50\}$.

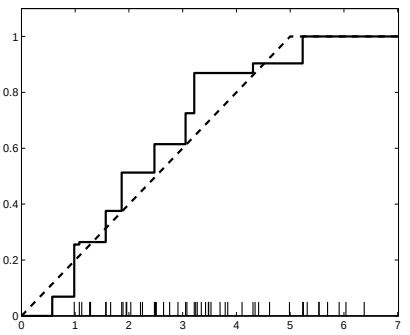

Figure 1.4: $\tilde{F}_{50}$ (solid), F_0 (dashed), Observations (marks).

Nonparametric Maximum Likelihood Estimator (MLE)

Formally, the Nonparametric Maximum Likelihood Estimator (MLE) $\hat{F}_n$ in inverse models is defined as the maximizer of

$$\Psi_n(F) \quad = \quad \int \log h_F(z)\,dH_n(z) \quad = \quad \int \log\left[T(F)\right](z)\,dH_n(z)$$

over the set $\mathcal{F}_{[0,\infty)}$, where H_n denotes the empirical distribution function corresponding to h_0. However, modifications of either the target function Ψ_n or the set $\mathcal{F}_{[0,\infty)}$ to optimize over, might be necessary to obtain a well defined optimization problem.

MLE in a Class of Deconvolution Models

For the deconvolution model, the MLE was first introduced by Groeneboom and Wellner (1992) together with a conjecture about its limit distribution in the case of decreasing noise densities g. That conjecture was in fact the initial motivation for us to examine this particular estimator. Jongbloed (1998a) studies the asymptotic behavior of the MLE for the exponential deconvolution model, i.e. $g \sim \text{Exp}(1)$, and verifies the conjecture in this specific case. However, the argument used there depends on the special structure of g and can therefore not be used for more general noise densities.
We now allow a broader class of noise densities g by assuming that $g : [\,0, \infty) \to [\,0, \infty)$ varies over the class of absolutely continuous, bounded and decreasing densities, allowing the representation

$$g(y) \;=\; -\int_y^\infty g'(w)\,dw \;=\; g(0) + \int_0^y g'(w)\,dw,\; y \geq 0. \tag{1.2.3}$$

Then it turns out that the MLE is well-defined (Chapter 2). It is a piecewise constant function (Section 2.1), computable via iterative procedures (Section 2.3), consistent (Section 3.1), possesses a cube-root-n behavior (Chapter 4) and satisfies the conjectured pointwise asymptotic limit under the assumption that certain functionals of $\hat{F}_n$ converge at rate $n^{-1/2}$ (Chapter 5).

Example (Exponential Deconvolution).
Choose g to be the standard exponential density. Using 100 observations generated according to $F_0 = U[\,0, 5\,]$, Figure 2.1 shows the MLE $\hat{F}_{100}$ in this exponential deconvolution model.

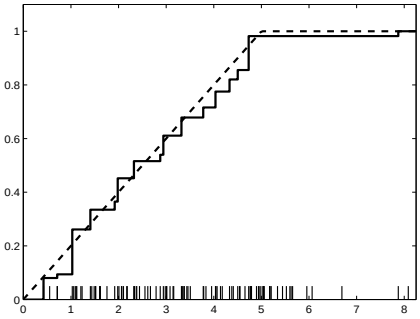

Figure 1.5: $\hat{F}_{100}$ (solid), F_0 (dashed), Observations (marks).

1.3 Asymptotic Behavior

We are interested in the behavior of the estimators as the number of observations tends to infinity. This behavior is described by the following three properties: consistency, rate of convergence and the limit distribution. Even though our main interest lies in the latter one, consistency and rates of convergence are needed to obtain the asymptotic distribution.

The asymptotic behavior of the IIE in the deconvolution model can be found in van Es et al. (1998), derived by using the explicit characterization of the estimator. Motivated by the conjectured pointwise limit behavior of the MLE in deconvolution settings in Groeneboom and Wellner (1992), we study in the coming chapters the three mentioned asymptotic properties for this estimator and compare them to the asymptotic behavior of the IIE.

Empirical process theory is used extensively (see Appendix B for notation and theorems needed in later chapters) to obtain the asymptotic properties. The biggest challenge is the lack of an explicit representation of the MLE. Therefore, all information about the estimator will be deduced from Fenchel optimality conditions (see Theorem 2.2.1); necessary and sufficient conditions that characterize the MLE as a solution of an optimization problem. Those optimality conditions depend very much on the structure of the special model, so that no general theory is available for deriving asymptotics of estimators in 'inverse settings'. They rather have to be derived for each model separately as done in the sequel for the deconvolution model.

For the rest of this section we focus on the MLE $\hat{F}_n$ and give a preview of the asymptotic results derived in Chapter 3-5, also in comparison to the IIE and the empirical distribution function F_n as a nonparametric estimator of F_0 in the case of directly observable data. As a model we choose $g(y) = 2(1-y)\mathbf{1}_{[0,1]}(y)$ and $F_0(x) = \frac{x}{5}\mathbf{1}_{[0,5]}(x) + \mathbf{1}_{(5,\infty)}(x)$. All pictures are produced using a maximum of $n = 5000$ observations due to computational time.

Consistency

To see whether the estimators $\hat{F}_n$ and $\hat{h}_n = h_{\hat{F}_n}$ converge to F_0 and $h_0 = h_{F_0}$, respectively, we first study $d_H(\hat{h}_n, h_0)$, the Hellinger distance d_H between $\hat{h}_n$ and h_0. Figure 1.6 illustrates realizations of $d_H(\hat{h}_n, h_0)$ as a function of the sample size n. As n grows, $d_H(\hat{h}_n, h_0)$ will be seen to converge almost surely to zero in Lemma 3.1.2. For $\hat{F}_n$ itself, we are mostly interested in the uniform convergence, analog to the Glivenko-Cantelli theorem for F_n and the unform consistency attained by the IIE. Figure 1.7 supports uniform convergence of $\hat{F}_n$, as it is shown in Theorem 3.1.5.

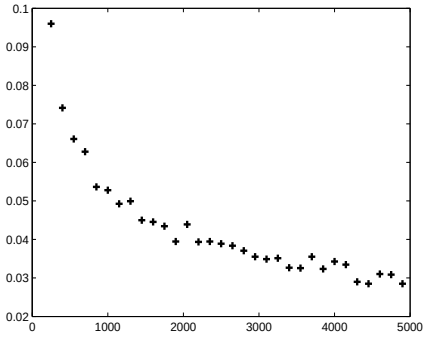

FIGURE 1.6: $d_H(\hat{h}_n, h_0)$.

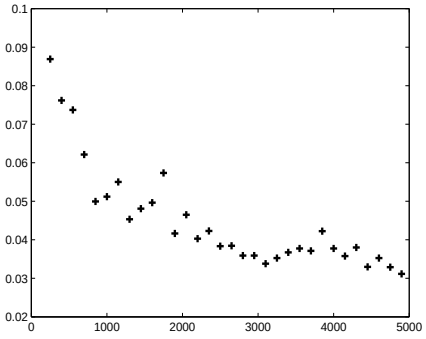

FIGURE 1.7: $\sup_{[0,\infty)} |\hat{F}_n(x) - F_0(x)|$.

Rates of Convergence

We consider rates of convergence of $\hat{h}_n$ and $\hat{F}_n$ from a global as well as local point of view. Although expecting a slower rate than $n^{-1/2}$ (which is the right scaling for the empirical distribution function F_n in classical settings), it turns out that we still obtain polynomial rates, i.e. rates of the type n^{α} for some negative α. The slope s of a linear regression of the points $\{(\log n, \log \hat{d}_n) : n \in N \subset \mathbb{N}\}$, for $\hat{d}_n$ being a measure of the distance of the MLE and h_0 or F_0, then reveals information about α. We will derive that $\alpha = -1/3$.

Figure 1.8 shows realized values of $\log d_H(\hat{h}_n, h_0)$ against $\log n$ for various values of n together with its linear fit whose slope is $s = 0.33$. In Theorem 4.1.1 this global cube-root-n behavior for $\hat{h}_n$ is established. The same behavior can be noticed for $\hat{F}_n$ with respect to the L_2-norm as illustrated in Figure 1.9 by plotting $\log \|\hat{F}_n - F_0\|_{L_2}$ versus $\log n$ with a linear fit with slope $s = 0.34$ and proven in Theorem 4.1.4. For local rate results we consider $\sup_{I_n} |\hat{F}_n(x) - F_0(x)|$ for $I_n = [\, x_0 - Mn^{-1/3}, x_0 + Mn^{-1/3}\,]$ for a fixed $x_0 > 0$ and some $M > 0$. For the illustration in Figure 1.10 we choose $x_0 = 2$ and $M = 10$ and obtain a slope of $s = 0.32$ for the linear regression of $\{\log n, \log \sup_{I_n} |\hat{F}_n(x) - F_0(x)|\}$ which is in line with the mathematical background presented in Lemma 4.3.3.

Functionals of $\hat{F}_n$ might converge at rate $n^{-1/2}$ to their true values. One particular example we are interested in, is $K_t(F) = \int_{x \in [\,0,t\,]} (g(t-x) - g(0))\, dF(x)$, $F \in \mathcal{F}_{[\,0,\infty)}$, $t > 0$ fixed, which will play a crucial role in Chapter 5. In Figure 1.11 we take $t = 2$ and plot 50 values of $K(\hat{F}_n) - K(F_0) = 2 \int_{[\,0,2\,]} (x-2)\, d(\hat{F}_n - F_0)(x)$ for various numbers of observations together with the functions $\pm 3n^{-1/2}$, giving some evidence of a rate of $n^{-1/2}$. A discussion of this rate is given in Section 5.2.

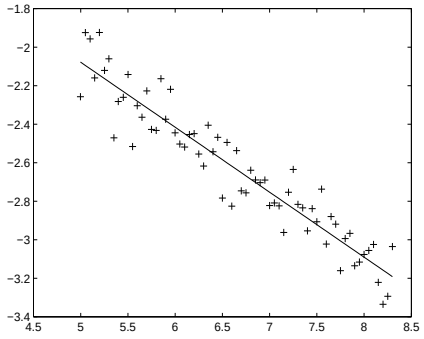

FIGURE 1.8: Global rate of $\hat{h}_n$, $s = 0.33$.

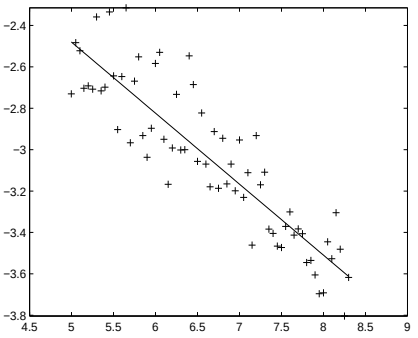

FIGURE 1.9: Global rate of $\hat{F}_n$, $s = 0.34$.

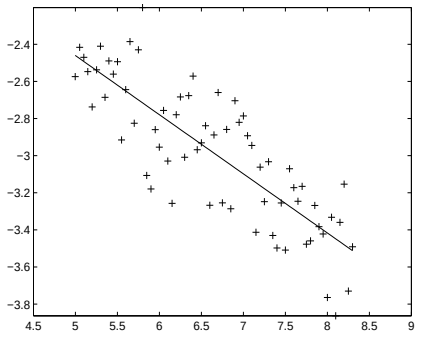

FIGURE 1.10: Local rate of $\hat{F}_n$, $s = 0.32$.

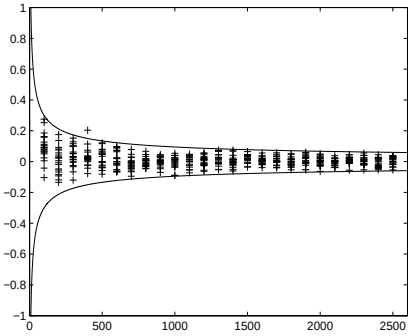

FIGURE 1.11: Rate of $K_2(F)$.

Pointwise Asymptotic Distribution

We study the asymptotic distribution of $\hat{F}_n$ at a fixed point, here $x_0 = 2$, as the equivalent to the central limit theorem that holds for F_n. Figure 1.12 shows a histogram based on 1500 values of $L_{3000} = (3000)^{-1/3}200|\hat{F}_{3000}(2) - F_0(2)|$ together with the (properly scaled) density of the so-called Chernoff distribution. In Theorem 5.1.3 we discuss how we can show that the limit distribution of $n^{1/3}c_0|\hat{F}_n(x_0) - F_0(x_0)|$, for some normalizing constant $c_0 > 0$, can indeed be described by Chernoff's distribution, just as the IIE.

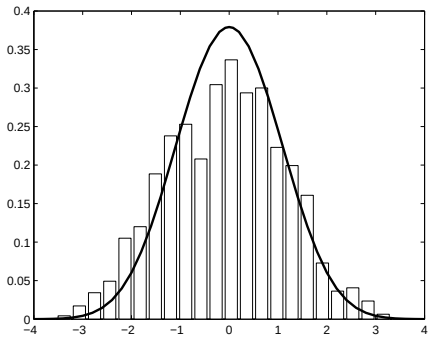

FIGURE 1.12: Histogram of L_{3000}.

1.4 Outline

This section gives an overview of the results presented and proven in this thesis.

Chapter 2 introduces the nonparametric maximum likelihood estimator (MLE). We establish its existence and uniqueness for a class of deconvolution models with decreasing noise densities g. As a solution of an optimization problem, the MLE is characterized in terms of Fenchel optimality conditions. A Newton procedure is proposed to compute the MLE and an illustration thereof is given.

Chapter 3 is devoted to consistency results. Almost sure consistency for $h_{\hat{F}_n}$ is derived with respect to the Hellinger metric. From that, pointwise almost sure consistency of $\hat{F}_n$ is deduced which is strengthened to a uniform result in case of F_0 being continuous. Furthermore, the behavior of the upper support point of the MLE is discussed.

Chapter 4 contains various rate results. It is proven that $h_{\hat{F}_n}$ converges with rate $n^{-1/3}$ in a global sense with respect to the Hellinger and the L_2-norm. Studying a decomposition of the Fenchel optimality conditions allows us to also establish local rate results of order $n^{-1/3}$ of $\hat{F}_n$ in a neighborhood of some fixed $x_0 > 0$. The same rate is also obtained for the difference of successive points of jump of $\hat{F}_n$.

Chapter 5 contains a discussion on the limit distribution of the MLE at a fixed point as conjectured in Groeneboom and Wellner (1992). There it is stated that the normalized process $n^{-1/3}c_{F_0}(\hat{F}_n(x_0) - F_0(x_0))$ for $x_0 > 0$ and some constant c_{F_0} converges in distribution to the derivative of the concave majorant at zero of a Brownian Motion with parabolic drift. In order to complete the proof that $\hat{F}_n(x_0)$ has indeed this distributional limit, smoothness of certain functionals of $\hat{F}_n$ in the sense that they converge at rate $n^{-1/2}$ to their true values, needs to be derived. An outline an of idea how this can be done is given too, as it is not yet fully established.

Appendix A contains explicit formulas for the resolvent ϱ for specific choices of g. A short overview of the fundamental theorems of empirical process theory used in this text is given in Appendix B. To streamline some of the technical sections, Appendix C collects results about entropy calculations that appear in various sections. In Appendix D an alternative proof for one of the consistency results presented in Chapter 3 is given.