Appendix

A

The Resolvent ϱ

In Definition 1.2.1 the resolvent ϱ is defined as the solution of an integral equation. In the deconvolution model, ϱ allows us to express $F \in \mathcal{F}_{[0,\infty)}$ in terms of $h_F \in \mathcal{H}$ (see Lemma 1.2.2). This is used in Chapter 3 and 4 in order to derive properties of the MLE $\hat{F}_n$ from the corresponding ones of an estimator of h_0. In this appendix we derive recursive explicit formulas for the resolvent for a class of noise densities g with compact support.

Recall that the resolvent ϱ is defined as the solution of the linear integral equation $[\varrho * g](t) = t$, $t \geq 0$ (see Definition 1.2.1), and that it is a continuous function on $[0, \infty)$. We now study the resolvent for the class of densities

$$g_n(y) \quad = \quad n(1-y)^{n-1}\mathbf{1}_{[0,1]}(y),\ n \in \mathbb{N}.$$

The following Lemma, together with Figure A.1 and A.2, illustrates the resolvent for the density g_2.

LEMMA A.1 (COMPUTATION ϱ AND ϱ' FOR g_2).
Let $g_2(y) = 2(1-y)\mathbf{1}_{[0,1]}(y)$ and $\varrho_m(t) = \varrho(t)|_{[m,m+1)}$ the restriction of ϱ on $[m, m+1)$ for $m \geq 0$. Then

$$\varrho_m(t) \quad = \quad \frac{1}{2}e^t \sum_{j=0}^{m} \frac{1}{j!} e^{-j}(j-t)^j \quad \text{and} \tag{A.1}$$

$$\varrho_m'(t) \quad = \quad \frac{1}{2}e^t \left(1 - \sum_{j=1}^{m} \frac{t}{j!} e^{-j}(j-t)^{j-1}\right). \tag{A.2}$$

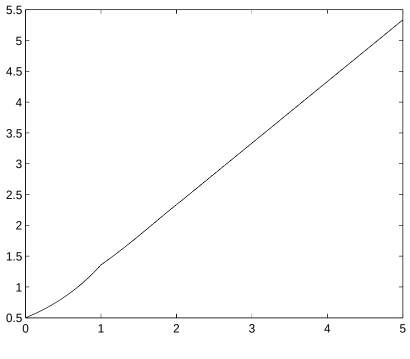

FIGURE A.1: Resolvent ϱ for g_2.

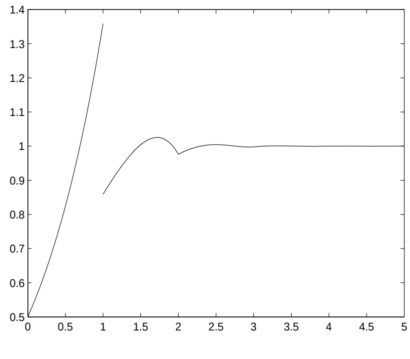

FIGURE A.2: Derivative ϱ' for g_2.

PROOF OF LEMMA A.1.
From Lemma 4.2 of Jongbloed (1995) we know that ϱ is differentiable on $(0, \infty) \setminus \{1\}$ and, for $t \geq 0$, a solution of the delay differential equation

$$-\varrho(t) + \varrho'(t) + \varrho(t-1) \quad = \quad 0,\ \varrho(0) \ = \ \frac{1}{2}. \tag{A.3}$$

Let $m = 0$. Then it is easily seen that $\varrho_0(t) = \exp(t)/2, t \in [0,1)$, is the solution of (A.3), using that $\varrho(t) = 0$ for all $t < 0$. Also for $m \geq 1$ we consider $\varrho(t)$ to be of the form $\varrho(t) = c(t)e^t$ for some function c and apply an induction argument. Assume that ϱ_{m-1} has a representation as in (A.1) and let $t \in [m, m+1)$. Then, implied by (A.3),

$$0 \;=\; e^t c'(t) + \varrho_{m-1}(t-1) \;=\; e^t c'(t) + \frac{1}{2} e^{t-1} \sum_{j=0}^{m-1} \frac{1}{j!} e^{-j} (j-t+1)^j$$

implying

$$c(t) \;=\; \frac{1}{2} e^{-1} \sum_{j=0}^{m-1} \frac{1}{j!} e^{-j} \frac{1}{j+1} (j+1-t)^{j+1} + c_0 \;=\; \frac{1}{2} \sum_{j=1}^{m} \frac{1}{j!} e^{-j} (j-t)^j + c_0$$

for some constant c_0. From $\varrho_m(m) = c(m)e^m = \varrho_{m-1}(m)$ (recall that ϱ is continuous on $[0,\infty)$) we get $c(m) = e^{-m}\varrho_{m-1}(m)$ which implies together with the previous display

$$c_0 \;=\; \frac{1}{2} \sum_{j=0}^{m-1} \frac{1}{j!} e^{-j} (j-m)^j - \frac{1}{2} \sum_{j=1}^{m} \frac{1}{j!} e^{-j} (j-m)^j \;=\; \frac{1}{2}.$$

Thus, for $t \in [m, m+1)$,

$$\varrho_m(t) \;=\; c(t)e^t \;=\; e^t \left(\frac{1}{2} \sum_{j=1}^{m} \frac{1}{j!} e^{-j} (j-t)^j + \frac{1}{2} \right) \;=\; \frac{1}{2} e^t \sum_{j=0}^{m} \frac{1}{j!} e^{-j} (j-t)^j$$

Differentiation then leads to

$$\begin{aligned}
\varrho_m'(t) \;&=\; \frac{1}{2} e^t \left(\sum_{j=0}^{m} \frac{1}{j!} e^{-j} (j-t)^j - \sum_{j=1}^{m} \frac{1}{j!} e^{-j} j (j-t)^{j-1} \right) \\
&=\; \frac{1}{2} e^t \left(1 + \sum_{j=1}^{m} e^{-j} (j-t)^{j-1} \left[\frac{1}{j!}(j-t) - \frac{1}{(j-1)!} \right] \right) \\
&=\; \frac{1}{2} e^t \left(1 - \sum_{j=1}^{m} e^{-j} (j-t)^{j-1} \frac{t}{j!} \right), \; t \in [m, m+1). \quad \square
\end{aligned}$$

For g_n, $n \geq 3$, the resolvent can be derived iteratively along the same line as in the previous lemma after replacing (A.3) by the appropriate integro-differential equation that characterizes ϱ according to the following Lemma.

Lemma A.2.
Let $g_n(y) = n(1-y)^{n-1}\mathbf{1}_{[0,1]}(y)$ *with* $n \geq 3$. *Then* ϱ *is* $(n-1)$*-times differentiable on* $(0,\infty)$*, satisfying, for* $t \geq 0$*,*

$$\sum_{\nu=0}^{n-1}(-1)^{n-1-\nu}\left[\prod_{\mu=0}^{n-1-\nu}(n-\mu)\right]\varrho^{(\nu)}(t) - (-1)^{n-1}\left[\prod_{\mu=0}^{n-1}(n-\nu)\right]\varrho(t-1) = 0, \tag{A.4}$$

with initial values

$$\varrho(0) = 1/n,\ \varrho^{(k)}(0+) \;=\; -\frac{1}{n}\sum_{\nu=0}^{k-1}(-1)^{k-\nu}\left[\prod_{\mu=0}^{k-\nu}(n-\nu)\right]\varrho^{(\nu)}(0+),\ k = 1,\ldots,n-2.$$

Proof.
Higher order differentiability of ϱ can be shown similar to Lemma 4.3 in Jongbloed (1995). Note that successive differentiation of $[\varrho * g_n](t) = t$, $t \geq 0$, yields first

$$g_n(0)\varrho(t) + \int_{t-1}^{t} g_n'(t-y)\varrho(y)\,dy \;=\; 1$$

and eventually, for $k = 1,\ldots,n-1$,

$$\sum_{\nu=0}^{k} g_n^{(k-\nu)}(0+)\varrho^{(\nu)}(t) - \sum_{\nu=0}^{k-1} g_n^{(k-\nu)}(1-)\varrho^{(\nu)}(t-1) + \int_{t-1}^{t} g_n^{(k+1)}(t-y)\varrho(y)\,dy \;=\; 0 \tag{A.5}$$

where $g_n^{(\nu)}(0+) = \lim_{h\downarrow 0} g_n^{(\nu)}(h)$ and $g_n^{(\nu)}(1-) = \lim_{h\uparrow 1} g_n^{(\nu)}(h)$, $\nu = 1,\ldots,n-1$. Since $g_n^{(n)}(y) = 0$ for all $y \in (0,1)$ and since, for $\nu = 1,\ldots,n-1$,

$$g_n^{(k)}(y) \;=\; (-1)^k\left[\prod_{\mu=0}^{k}(n-\mu)\right]\cdot(1-y)^{n-(k+1)}, y \in (0,1), \tag{A.6}$$

equation (A.5) implies (A.4) after choosing $k = n-1$.

Note that $t = 0$ in (A.5) results in the initial values $\varrho(0) = 1/n$ and, for $k = 1,\ldots,n-2$,

$$\begin{aligned}\varrho^{(k)}(0+) &= -\frac{1}{g(0)}\sum_{\nu=0}^{k-1} g_n^{(k-\nu)}(0+)\varrho^{(\nu)}(0+)\\ &= -\frac{1}{n}\sum_{\nu=0}^{k-1}(-1)^{k-\nu}\left[\prod_{\mu=0}^{k-\nu}(n-\nu)\right]\varrho^{(\nu)}(0+)\end{aligned}$$

where the last equality is due to (A.6). □

We close this appendix by computing ϱ on $[0,2)$ for g_3 where we cannot -contrary to the situation in Lemma A.1- expect an expression of the type $\varrho(t) = c(t)\exp(t)$.

EXAMPLE ($g_3(y) = 3(1-y)^2\mathbf{1}_{[0,1]}(y)$).
According to Lemma A.2 the resolvent is given for $t \in [0,1)$ as the solution of

$$6\varrho_0(t) - 6\varrho_0'(t) + 3\varrho_0''(t) \quad = \quad 0, \varrho(0) \;=\; \frac{1}{3}, \varrho'(0) \;=\; \frac{2}{3}.$$

The roots of the characteristic polynomial $c(\lambda) = 3\lambda^2 - 6\lambda + 6$ equal $\lambda_1 = 1+i$ and $\lambda_2 = 1-i$. Thus,

$$\begin{aligned}
\varrho_0(t) &= c_1\exp(t)(\cos(t) + i\sin(t)) + c_2\exp(t)(\cos(-t) + i\sin(-t)) \\
&= \exp(t)\left[(c_1+c_2)\cos(t) + i(c_1-c_2)\sin(t)\right], t \in [0,1),
\end{aligned}$$

for some constants c_1 and c_2. Using the initial value conditions yields first $c_1 = (1-i)/6$ and $c_2 = (1+i)/6$ and eventually

$$\varrho_0(t) = \exp(t)\left[\frac{1}{3}\cos(t) + \frac{1}{3}\sin(t)\right], t \in [0,1).$$

For $t \in [1,2)$ we can proceed iteratively, solving,

$$\begin{aligned}
0 &= 6\varrho_1(t) - 6\varrho_1'(t) + 3\varrho_1''(t) - 6\varrho_1(t-1), \\
\varrho_1(1) &= \exp(1)(\cos(1) + \sin(1))/3,\ \varrho_1'(1) = \varrho_0'(1) = 2/3\exp(1)\cos(1)
\end{aligned}$$

which results in

$$\begin{aligned}
\varrho_1(t) &= 0.692\exp(7/6t)\sin(\sqrt{23}/6t) \\
&\qquad -0.152\exp(7/6t)\cos(\sqrt{23}/6t) - 2\exp(t-1)\sin(t-1).
\end{aligned}$$

Even though one can continue in the same way for $t \geq 2$, trying to obtain closed expressions for ϱ as for g_2 presented above, we do not show the resolvent for $t > 2$ due to the more and more involved formulas.

B

Empirical Processes

Empirical process theory plays an important role throughout this document. The most fundamental definitions and concepts for a specific type of processes are summarized in this appendix: entropy, Glivenko Cantelli- and Donsker classes. An introduction to this field can be found in van der Vaart (1998, Chapter 19) whereas van der Vaart and Wellner (1996) provide a detailed overview of this subject. Results that are in particular useful for deriving properties of estimators defined as the maximizer of a criterion function are presented in van de Geer (2000).

The stochastic processes appearing in this document are of the following type. Let M denote a distribution function, M_n its empirical counterpart based on an iid sample from M of size n, and $\mathcal{A}$ a set of real valued functions defined on $\mathbb{R}$. Then define the processes

$$X_n(\alpha) \quad = \quad r_n \int_{(-\infty,\infty)} \alpha(z)\, d(M_n - M)(z) \tag{B.1}$$

for some deterministic sequence r_n and $\alpha \in \mathcal{A}$. We disregard any measurability issues and refer to the literature, for instance van der Vaart and Wellner (1996, Chapter 1), for more general, mathematically profound definitions of stochastic processes.
Whether processes of the type (B.1) converge almost surely or weakly depends on the size of the set $\mathcal{A}$ measured in terms of entropy (with or without bracketing) and the choice of the sequence r_n.

DEFINITION B.1 (ENTROPIES).
Let $\mathcal{A}$ be a class of functions endowed with a metric $\|\cdot\|$. A set of pairs of functions $\{[l_i, u_i], i = 1, \ldots, n\}$ are called δ-brackets for some $\delta > 0$ if $\|u_i - l_i\| \le \delta$ for all $i = 1, \ldots, n$ and if for any $\alpha \in \mathcal{A}$ one can find an $i_0 \in \{1, \ldots, n\}$ such that $l_{i_0} \le \alpha \le u_{i_0}$. For the smallest such possible n, $\log n$ is the $\delta-$entropy with bracketing of $\mathcal{A}$ with respect to $\|\cdot\|$, denoted by $H_{[]}(\delta, \mathcal{A}, \|\cdot\|)$.
Let $\delta > 0$ and assume there exists a set of functions $\{\alpha_i, i = 1 \ldots, n\}$ such that for all $\alpha \in \mathcal{A}$ there exists an α_{i_0} with $i_0 \in \{1, \ldots, n\}$ and $\|\alpha - \alpha_{i_0}\| \le \delta$. Then for the smallest such possible n, $\log n$ is called the δ-entropy of $\mathcal{A}$ with respect to $\|\cdot\|$ and denoted by $H(\delta, \mathcal{A}, \|\cdot\|)$.

Almost sure and weak convergence of the processes defined in (B.1) is stated below. Proofs can for instance be found in Section 19.2 of van der Vaart (1998).

THEOREM B.2 (GLIVENKO-CANTELLI).
Let $\mathcal{A}$ be a class of measurable functions on $\mathbb{R}$ and M a distribution function on $\mathbb{R}$ such that $H_{[]}(\delta, \mathcal{A}, L_1(M)) < \infty$ for all $\delta > 0$. Then, choosing $r_n \equiv 1$ in (B.1), X_n converges uniformly almost surely to zero, i.e.

$$\|X_n\|_{\mathcal{A}} \quad = \quad \sup_{\alpha \in \mathcal{A}} \left| \int_{[0,\infty)} \alpha(z)\, d(M_n - M)(z) \right| \;\to\; 0 \;\; a.s. \tag{B.2}$$

Let $A : \mathbb{R} \to \mathbb{R}$ be an envelope function of $\mathcal{A}$, i.e. $|\alpha(z)| \le A(z) < \infty$ for all $\alpha \in \mathcal{A}$ and all $z \in \mathbb{R}$ with $\int A(x)\, dM(x) < \infty$. Formula (B.2) is then also implied by the condition

$$\sup_Q \left(H_{[]}\left(\delta \|A\|_{L_2(Q)}, \mathcal{A}, L_2(Q)\right)\right) \quad < \quad \infty, \; \text{for all } \delta > 0$$

where the supremum is taken over all probability measures Q.

THEOREM B.3 (DONSKER).
Let $\mathcal{A}$ be a class of measurable functions on $\mathbb{R}$ and M a distribution function on $\mathbb{R}$ such that $J_{[\,]}(1, \mathcal{A}, L_2(M)) < \infty$ for

$$J_{[\,]}\left(\delta, \mathcal{A}, L_2(M)\right) \quad = \quad \int_0^\delta \sqrt{H_{[\,]}(\tau, \mathcal{A}, L_2(M))}\, d\tau.$$

Then $\mathcal{A}$ is $M-$Donsker, i.e. $X_n = \sqrt{n} \int \alpha\, d(M_n - M)$ converges to a tight limit process L in the space $\ell^\infty(\mathcal{A})$, the space of all bounded functions $\nu : \mathcal{A} \to \mathbb{R}$. Alternatively, $\mathcal{A}$ is M-Donsker if

$$\int_0^1 \sqrt{\sup_Q H_{[\,]}\left(\tau \|A\|_{L_2(Q)}, \mathcal{A}, L_2(Q)\right)}\, d\tau \quad < \quad \infty$$

for A being an envelope function of $\mathcal{A}$ satisfying $\int A^2(x)\, dM(x) < \infty$ and where the supremum is taken over all probability measures Q.

The limit process L in the previous Donsker theorem is an $M-$Brownian Motion, a Gaussian process with mean zero and covariance function

$$\mathrm{Cov}(L(\alpha_1), L(\alpha_2)) \quad = \quad \int \alpha_1 \alpha_2\, dM - \int \alpha_1\, dM \cdot \int \alpha_2\, dM,\ \ \alpha_1, \alpha_2 \in \mathcal{A}.$$

C

Various Entropies

Entropy calculations of certain classes of functions appear in most of the chapters. If the class consists of elements satisfying some monotonicity or smoothness conditions, we can apply known results for computing its entropy (see Section C.1). Often it is helpful to interpret the class as a transformation or composition of classes with known entropies. In Section C.2 we describe how to relate the entropy of the class of interest to these individual entropies.

C.1 Special Classes of Functions

Lemma C.1.1 (Monotone/Bounded Variation).
Let $\mathcal{P}$ be a class of either bounded monotone functions on $\mathbb{R}$ or a class of uniformly bounded variation. Then in both cases there exists a positive constant A such that for all $\delta > 0$ and Q being a probability measure

$$H_{[\,]}\left(\delta, \mathcal{P}, L_2(Q)\right) \quad \leq \quad A\delta^{-1}. \tag{C.1.1}$$

Proof.
See Lemma 3.8 and (2.6) in van de Geer (2000). □

Lemma C.1.2 (Vapnik-Chervonenkis Classes).
Let $\mathcal{P}$ be a Vapnik-Chervonenkis class with index $V(\mathcal{P}) \in (0, \infty)$ (as defined in Chapter 19 of van der Vaart (1998)). Then there exists a constant K such that for any $r \geq 1$ and $0 < \delta < 1$

$$\sup_Q \left(\exp\left(H\left(\delta\|P\|_{L_r(Q)}, \mathcal{P}, L_r(Q)\right)\right)\right) \quad \leq \quad K \cdot V(\mathcal{P})(16e)^{V(\mathcal{P})}\left(\frac{1}{\delta}\right)^{r\left(V(\mathcal{P})-1\right)}$$

where P denotes an envelope of $\mathcal{P}$, i.w. $|p| \leq P$ for all $p \in \mathcal{P}$, and where the supremum is taken over all probability measures Q

Proof.
See Lemma 19.15 in van der Vaart (1998). □

C.2 Transformation and Composition

Lemma C.2.1 (Reciprocal Class).
Let $\mathcal{P}$ be a class of functions such that p^{-1} is well-defined for all $p \in \mathcal{P}$ and let $\mathcal{P}^{-1} = \{1/p : p \in \mathcal{P}\}$. If $|p| \geq c$ for all $p \in \mathcal{P}$ and some constant $c > 0$ then we have for all $\delta > 0$ and probability measures Q

$$H\left(\delta, \mathcal{P}^{-1}, L_2(Q)\right) \quad \leq \quad H\left(c^2\delta, \mathcal{P}, L_2(Q)\right). \tag{C.2.1}$$

If $\mathcal{P}$ is a class of functions with $p \geq c > 0$ for all $p \in \mathcal{P}$, statement (C.2.1) *remains valid after replacing the entropies by the corresponding ones with brackets.*

Proof.
Let $\delta > 0$ and Q be a probability measure. Let $\{d_i : i = 1, \dots, K\}$ be a $c^2\delta$-covering set of $\mathcal{P}$ with respect to $L_2(Q)$. Without loss of generality we can assume that $|d_i| \geq c$ for all $i = 1, \dots, K$. Then for all $p \in \mathcal{P}$ one can find a $j \in \{1, \dots, K\}$ such that $\|p - d_j\|_{L_2(Q)} \leq c^2\delta$. Using this j yields

$$\left\|\frac{1}{d_j} - \frac{1}{p}\right\|_{L_2(Q)} \quad = \quad \left\|\frac{p - d_j}{pd_j}\right\|_{L_2(Q)} \quad \leq \quad \frac{1}{c^2}c^2\delta \;=\; \delta.$$

Let $\left\{\left[\, a_i^l, a_i^u \,\right] : i = 1, \dots, K\right\}$ be a set of $c^2\delta$-brackets for $\mathcal{P}$, where $\mathcal{P}$ now contains positive functions bounded from below by $c > 0$. Then the set $\left\{\left[\, 1/a_i^u, 1/a_i^l \,\right] : i = 1, \dots, K\right\}$ acts as a set of δ-brackets for $\mathcal{P}^{-1}$ since we can assume that $0 < c \leq a_i^l \leq a_i^u$ for all $i = 1 \dots, K$. □

Lemma C.2.2 (Sum of Classes).
Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be classes of functions. Then the entropy with brackets for the class $\mathcal{P}_1 \oplus \mathcal{P}_2 = \{p_1 + p_2 : p_1 \in \mathcal{P}_1,\, p_2 \in \mathcal{P}_2\}$ satisfies

$$H_{[\,]}(\delta, \mathcal{P}_1 \oplus \mathcal{P}_2, L_2(Q)) \quad \leq \quad H_{[\,]}(\delta/2, \mathcal{P}_1, L_2(Q)) + H_{[\,]}(\delta/2, \mathcal{P}_2, L_2(Q))$$

for any $\delta > 0$ and probability measure Q.

Proof.
Let $\delta > 0$ and Q be a probability measure. A set of $\delta/2$ brackets for $\mathcal{P}_1$ with respect to $L_2(Q)$ is given by $\left\{\left[a_i^l, b_i^u\right] : i = 1, \dots, k_1\right\}$ and one for $\mathcal{P}_2$ by $\left\{\left[b_j^l, b_j^u\right] : i = 1, \dots, k_2\right\}$. Define $c_{ij}^l = a_i^l + b_j^l$ and $c_{ij}^u = a_i^u + b_j^u$ for all combinations of i and j. We then obtain $a_i^l + b_j^l \leq p_1 + p_2 \leq a_i^u + b_j^u$ for $p_\nu \in \mathcal{P}_\nu, \nu = 1, 2$, if $a_i^l \leq p_1 \leq a_i^u$ and $b_j^l \leq p_2 \leq b_j^u$ for some $(i, j) \in \{1, \dots, k_1\} \times \{1, \dots, k_2\}$. Also, by the triangle inequality,

$$\|c_{ij}^u - c_{ij}^l\|_{L_2(Q)} \quad = \quad \|a_i^u + b_j^u - a_i^l - b_j^l\|_{L_2(Q)} \;\leq\; \frac{\delta}{2} + \frac{\delta}{2} \;= \delta$$

implying $H_{[\,]}(\delta, \mathcal{P}_1 \oplus \mathcal{P}_2, L_2(Q)) \leq \log(k_1 k_2)$. □

Lemma C.2.3 (Product of Classes).
Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be nonnegative, pointwise bounded classes of functions, i.e. $0 \leq p_1 \leq c_1$ for all $p_1 \in \mathcal{P}_1$ and $0 \leq p_2 \leq c_2$ for all $p_2 \in \mathcal{P}_2$ for constants $c_1 > 0$ and $c_2 > 0$. Let $\mathcal{P}_1 \odot \mathcal{P}_2 = \{p_1 \cdot p_2 : p_1 \in \mathcal{P}_1,\, p_2 \in \mathcal{P}_2\}$. Then we have for all $\delta > 0$ and Q being a probability measure

$$H_{[\,]}(\delta, \mathcal{P}_1 \odot \mathcal{P}_2, L_2(Q)) \quad \leq \quad H_{[\,]}(\delta/(2c_2), \mathcal{P}_1, L_2(Q)) + H_{[\,]}(\delta/(2c_1), \mathcal{P}_2, L_2(Q)).$$

PROOF.
Let $\delta > 0$, Q a probability measure and let $\{[a_i^l, a_i^u] : i = 1, \ldots, k_1\}$ be a set of $\delta/(2c_2)$-brackets for $\mathcal{P}_1$ with respect to $L_2(Q)$ and $\{[b_j^l, b_j^u] : i = 1, \ldots, k_2\}$ be one of $\delta/(2c_1)$-brackets for $\mathcal{P}_2$.

The functions $c_{ij}^l = a_i^l b_j^l$ and $c_{ij}^u = a_i^u b_j^u$ for all combinations of i and j serve as δ-brackets for $\mathcal{P}_1 \odot \mathcal{P}_2$ since $a_i^l b_j^l \leq p_1 p_2 \leq a_i^u b_j^u$ for $p_\nu \in \mathcal{P}_\nu$ for $\nu = 1, 2$ if $a_i^l \leq p_1 \leq p_i^u$ and $b_j^l \leq p_2 \leq b_j^u$ for some $i \in \{1, \ldots, k_1\}$ and $j \in \{1, \ldots, k_2\}$. Moreover, by the triangle inequality,

$$\begin{aligned} \|c_{ij}^u - c_{ij}^l\|_{L_2(Q)} &= \|a_i^u b_j^u - a_i^u b_j^l + a_i^u b_j^l - a_i^l b_j^l\|_{L_2(Q)} \\ \leq \; c_1 \|b_j^u - b_j^l\|_{L_2(Q)} + c_2 \|a_i^u - a_i^l\|_{L_2(Q)} &\leq c_1 \frac{\delta}{2c_1} + c_2 \frac{\delta}{2c_2} = \delta. \end{aligned}$$

Hence $H_{[\,]}(\delta, \mathcal{P}_1 \odot \mathcal{P}_2, L_2(Q)) \leq \log(k_1 k_2)$. $\square$

Note that Pollard (1990, pages 22-23) shows similar statements as in Lemma C.2.2 and Lemma C.2.3 using entropies without bracketing.

D

Consistency of $\hat{F}_n$

In Chapter 2 empirical process theory is used to obtain almost sure uniform strong consistency for $\hat{F}_n$ if F_0 is continuous (see Theorem 3.1.5). There, consistency of $\hat{F}_n$ is deduced from consistency of $\hat{h}_n$. In Section D.1 an alternative argument is presented to derive the same property of $\hat{F}_n$ immediately, without first considering $\hat{h}_n$. This in particular means that the resolvent is not used in this approach. In doing so, we follow the lines of Groeneboom and Wellner (1992) where one can find a proof of consistency in deconvolution models with symmetric noise density g.

D.1 Alternative Proof of Theorem 3.1.5

Let $F_0 \in \mathcal{F}_{[0,\infty)}$ be continuous and assume the same model and notation as introduced in Section 2.1. Before stating an alternative proof of Theorem 3.1.5, we list three lemmas that will be needed for it and whose proofs can be found in Section D.2.

Lemma D.1.1.
For each $n \in \mathbb{N}$ the following inequality holds:

$$\int_{[0,\infty)} \frac{h_0(z)}{\hat{h}_n(z)} \, dH_n(z) \quad = \quad \int_{[0,\infty)} \frac{\int_{[0,z]} g(z-x) \, dF_0(x)}{\int_{[0,z]} g(z-x) \, d\hat{F}_n(x)} \, dH_n(z) \ \leq \ 1. \tag{D.1.1}$$

Replacing in (D.1.1) the functions H_n and $\hat{F}_n$ by H_0 and F_0, respectively, yields a limiting form of (D.1.1) as stated in the following Lemma.

Lemma D.1.2.
Let F be a (sub)distribution function. Then the inequality

$$\int_{[0,\infty)} \frac{h_0(z)}{h_F(z)} \, dH_0(z) \quad = \quad \int_{[0,\infty)} \frac{\int_{[0,z]} g(z-x) dF_0(x)}{\int_{[0,z]} g(z-x) dF(x)} \, dH_0(z) \ \leq \ 1 \tag{D.1.2}$$

holds if and only if $F = F_0$.

Let $(F_n)_{n=1}^{\infty}$ be a sequence of distribution functions on $[0,\infty)$ converging vaguely to a function F; a right continuous, nondecreasing function that takes values in $[0,1]$.

Define for $\varepsilon > 0$ the set

$$B_\varepsilon \quad = \quad \{z \in [0,\infty) : \, h_F(z) \geq \varepsilon^2/2\}. \tag{D.1.3}$$

For $\mu > 0$ define the finite set $\tilde{D}_\mu$ by

$$\tilde{D}_\mu \quad = \quad \{x \in [0,\infty) : F(x) - F(x-) > \mu\} \ = \ \{d_1, d_2, \ldots d_D\}$$

with $F(x-) = \lim_{h \downarrow 0} F(x-h)$ and some $D \in \mathbb{N}$. Also define for $\nu > 0$

$$\tilde{D}_{\mu,\nu} \quad = \quad \bigcup_{d_i \in \tilde{D}_\mu} [\, d_i - \nu 2^{-(i+1)}, \, d_i + \nu 2^{-(i+1)} \,]. \tag{D.1.4}$$

Then $\tilde{D}_{\mu,\nu}$ is closed as the finite union of closed intervals and has Lebesgue measure less than or equal to ν.

Lemma D.1.3.
Let $0 < \varepsilon \leq (\sqrt[3]{28g(0)}\,)^{-1}$ *and define* $D_\varepsilon = \tilde{D}_{\varepsilon^5,\varepsilon}$. *Then there exists a constant* $c > 0$ *and an integer* n_0 *such that*

$$\sup_{B_\varepsilon \cap D_\varepsilon^c} \left| \frac{h_0(z)}{h_{F_n}(z)} - \frac{h_0(z)}{h_F(z)} \right| \quad \leq \quad c\,\varepsilon,\, n \geq n_0.$$

Proof of Theorem 3.1.5.
Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space supporting a sequence $Z_1, Z_2, \ldots$ of iid random variables such that Z_1 has density h_0. By the Glivenko-Cantelli theorem we know that there exists a set $B \in \mathcal{A}$ with probability equal to one, such that for all $\omega \in B$ we have $\sup_z |H_n(z,\omega) - H_0(z)| \to 0$ as $n \to \infty$. Fix $\omega \in B$.

Choose an arbitrary subsequence $(n_l)_{l=1}^\infty \subset (n)_{n=1}^\infty$. By the Helly compactness theorem there exists a further subsequence $(n_k)_{k=1}^\infty \subset (n_l)_{l=1}^\infty$ such that $\hat{F}_{n_k}(\,\cdot\,,\omega)$ converges vaguely to a function F. It remains to show that $F = F_0$ regardless the initially chosen subsequence $(n_k)_{k=1}^\infty$ which will be done by showing (D.1.2) for this limit function F using (D.1.1). Then we can conclude that the sequence $\hat{F}_n(\,\cdot\,,\omega)$ itself converges weakly to F_0. And since F_0 is continuous this pointwise convergence can be strengthened to uniform convergence due to the boundedness and monotonicity of F_0. (In what follows we omit the ω in the notation.)

Choose $0 < \varepsilon < \min\{(\sqrt[3]{28g(0)}\,)^{-1}, (\sqrt[4]{g(0)}\,)^{-1}\}$ such that the sets

$$A_\varepsilon \quad = \quad \Big\{z \in [0,\infty) : \int_0^z g(z-y)\,dF_0(y) > \varepsilon\Big\} \; = \; \Big\{z \in [0,\infty) : \; h_0(z) > \varepsilon\Big\}$$

and B_ε (defined as in (D.1.3), using the current limit function F) are nonempty which is possible since h_0 is a density and h_F a subdensity not identically equal to zero, otherwise one could construct a contradiction to (D.1.1).

Let $D_\varepsilon = \tilde{D}_{\varepsilon^5,\varepsilon}$ (also now in terms of the current limit function F) to define the set

$$C_\varepsilon \quad = \quad A_\varepsilon \cap \overset{\circ}{B}_\varepsilon \cap D_\varepsilon^c.$$

Note that

- A_ε is open since h_0 is continuous on $[\,0,\infty)$, B_ε is closed since h_F is upper semicontinuous (seen by (2.1.2)), and D_ε^c is open by definition so that C_ε is open,
- the function $I_{C_\varepsilon} h_0/h_F$ is lower semicontinuous and
- $\hat{h}_{n_k}(z) \geq \varepsilon^2/4$ for all $z \in \overset{\circ}{B}_\varepsilon \cap D_\varepsilon^c$ and sufficiently large k (see (D.2.2)).

Lemma D.1.3 now implies that there exist constants $c_1 > 0$ and $k_0 \in \mathbb{N}$ such that

$$\sup_{z \in C_\varepsilon} \left| \frac{h_0(z)}{\hat{h}_{n_k}(z)} - \frac{h_0(z)}{h_F(z)} \right| \quad < \quad c_1 \varepsilon$$

for all $k \geq k_0$ and hence, by Lemma D.1.1,

$$\int_{C_\varepsilon} \frac{h_0(z)}{h_F(z)} \, dH_{n_k}(z) \quad \leq \quad \int_{C_\varepsilon} \left(\frac{h_0(z)}{\hat{h}_{n_k}(z)} + c_1 \varepsilon \right) dH_{n_k}(z) \ \leq \ 1 + c_1 \varepsilon, \ k \geq k_0.$$

Since H_n converges weakly to H_0 we get by a version of the Portmanteau theorem for lower semicontinuous, bounded functions (see Pollard, 1984, p.73)

$$\liminf_{k \to \infty} \int_{C_\varepsilon} \frac{h_0(z)}{h_F(z)} \, dH_{n_k}(z) \quad \geq \quad \int_{C_\varepsilon} \frac{h_0(z)}{h_F(z)} \, dH_0(z)$$

and hence

$$\int_{C_\varepsilon} \frac{h_0(z)}{h_F(z)} \, dH_0(z) \quad \leq \quad 1 + c_1 \varepsilon. \tag{D.1.5}$$

In order to deduce (D.1.2) from (D.1.5) we show that for small ε the set C_ε is large in a certain sense so that (D.1.2) can be interpreted as a limit version of (D.1.5) for ε tending to zero.

First note, that (D.2.1) implies $h_{n_k}(z) \leq h_F(z) + 7g(0)\varepsilon^5 \leq 3\varepsilon^2$ for $z \in B_{2\varepsilon}^c$ since $\varepsilon \leq (\sqrt[3]{28g(0)}\,)^{-1} \leq (\sqrt[3]{7g(0)}\,)^{-1}$.
Moreover, we need to show that $D_\varepsilon^c \cap \mathring{B}_\varepsilon \supset D_\varepsilon^c \cap B_{2\varepsilon}$. Let $z \in D_\varepsilon^c \cap B_{2\varepsilon}$. Then $z \in D_\varepsilon^c$ and $h_F(z) \geq 2\varepsilon^2 > \varepsilon^2/2 + 3g(0)\varepsilon^5$ since $\varepsilon < \sqrt[3]{1/(2g(0))}$. Again due to the choice of ε, i.e. $\varepsilon \leq \sqrt[4]{1/g(0)}$, the size of a jump of h_F is smaller than ε, so that there exist a $\delta > 0$ and an open set $O = (z - \delta, z + \delta) \subset D_\varepsilon^c$ such that $h_F(x) > \varepsilon^2/2 + g(0)\varepsilon^5 \geq \varepsilon^2/2$ for all $x \in O$. Thus the set O is contained in $D_\varepsilon^c \cap B_\varepsilon$ and z belongs to $D_\varepsilon^c \cap \mathring{B}_\varepsilon$.

Now we get due to (D.1.1) (note that $A_\varepsilon \cap D_\varepsilon^c \cap B_{2\varepsilon}^c$ is open)

$$\begin{aligned} 1 \quad &\geq \quad \liminf_{k \to \infty} \int_{[0,\infty)} \frac{h_0(z)}{\hat{h}_{n_k}(z)} \, dH_{n_k}(z) \geq \liminf_{k \to \infty} \int_{A_\varepsilon \cap D_\varepsilon^c \cap B_{2\varepsilon}^c} \frac{h_0(z)}{\hat{h}_{n_k}(z)} \, dH_{n_k}(z) \\ &\geq \quad \frac{\varepsilon}{3\varepsilon^2} \liminf_{k \to \infty} \int_{A_\varepsilon \cap D_\varepsilon^c \cap B_{2\varepsilon}^c} dH_{n_k}(z) \geq \frac{1}{2\varepsilon} \int_{A_\varepsilon \cap D_\varepsilon^c \cap B_{2\varepsilon}^c} dH_0(z). \end{aligned}$$

This implies, now using that $D_\varepsilon^c \cap \left(\mathring{B}_\varepsilon\right)^c \subset D_\varepsilon^c \cap B_{2\varepsilon}^c$, and $h_0(z) \leq g(0)$ for all z

$$
\begin{aligned}
\int_{C_\varepsilon} dH_0(z) \;&=\; \int_{A_\varepsilon \cap D_\varepsilon^c} dH_0(z) - \int_{A_\varepsilon \cap D_\varepsilon^c \cap \left(\overset{\circ}{B}_\varepsilon\right)^c} dH_0(z) \\
&\geq\; \int_{A_\varepsilon \cap D_\varepsilon^c} dH_0(z) - \int_{A_\varepsilon \cap D_\varepsilon^c \cap B_{2\varepsilon}^c} dH_0(z) \\
&\geq\; \int_{A_\varepsilon} dH_0(z) - \int_{D_\varepsilon} h_0(z)\, dz - 3\varepsilon \\
&\geq\; \int_{A_\varepsilon} dH_0(z) - g(0)\cdot\varepsilon - 3\varepsilon \;=\; \int_{A_\varepsilon} dH_0(z) - c_2\varepsilon
\end{aligned}
\tag{D.1.6}
$$

for a positive constant c_2.

For the last part of this proof define $C_{0,\varepsilon} = A_\varepsilon \cap \tilde{D}_{0,\varepsilon}^c \cap \overset{\circ}{B}_\varepsilon$ (see (D.1.4) for the definition of $\tilde{D}_{0,\varepsilon}$). The inequalities (D.1.5) and (D.1.6) remain valid after replacing C_ε by $C_{0,\varepsilon}$ since $C_{0,\varepsilon} \subset C_\varepsilon$ which implies

$$
\int_{C_{0,\varepsilon}} \frac{h_0(z)}{h_F(z)}\, dH_0(z) \;\;\leq\;\; \int_{C_\varepsilon} \frac{h_0(z)}{h_F(z)}\, dH_0(z) \;\leq\; 1 + c_1\varepsilon. \tag{D.1.7}
$$

And secondly, the decomposition $C_\varepsilon = C_{0,\varepsilon} \cup [C_\varepsilon \cap C_{0,\varepsilon}^c]$ and the fact that Lebesgue measure of $C_\varepsilon \cap C_{0,\varepsilon}^c$ is smaller than that of $D_\varepsilon^c \cap \tilde{D}_{0,\varepsilon}$ lead to, with $c_3 > 0$,

$$
\begin{aligned}
\int_{C_{0,\varepsilon}} dH_0(z) \;&=\; \int_{C_\varepsilon} dH_0(z) - \int_{C_\varepsilon \cap C_{0,\varepsilon}^c} dH_0(z) \\
&\geq\; \int_{A_\varepsilon} dH_0(z) - c_2\varepsilon - g(0) \int_{C_\varepsilon \cap C_{0,\varepsilon}^c} dz \\
&\geq\; \int_{A_\varepsilon} dH_0(z) - c_2\varepsilon - g(0) \int_{D_\varepsilon^c \cap \tilde{D}_{0,\varepsilon}} dz \\
&\geq\; \int_{A_\varepsilon} dH_0(z) - c_2\varepsilon - g(0) \int_{\tilde{D}_{0,\varepsilon}} dz \;\geq\; \int_{A_\varepsilon} h_0(z)\, dz - c_3\varepsilon.
\end{aligned}
\tag{D.1.8}
$$

For small $\varepsilon > 0$ the set A_ε can be seen as a union of open interval with $\mathbf{1}_{A_\varepsilon} \uparrow \mathbf{1}_{A_0}$ as $\varepsilon \downarrow 0$. Furthermore by the definition of $C_{0,\varepsilon}$ it can be seen that $C_{0,\varepsilon_1} \supset C_{0,\varepsilon_2}$ as $\varepsilon_1 < \varepsilon_2$ and hence it follows from (D.1.8) that $I_{C_{0,\varepsilon}}(z) \uparrow I_{A_0}(z)$ for H_0-almost all z if $\varepsilon \downarrow 0$. Thus by the monotone convergence theorem inequality (D.1.7) can be transformed to

$$
\begin{aligned}
1 \;&\geq\; \limsup_{\varepsilon\downarrow 0} \int_0^\infty I_{C_{0,\varepsilon}}(z) \frac{h_0(z)^2}{h_F(z)}\, dz \\
&=\; \int_0^\infty \limsup_{\varepsilon\downarrow 0} I_{C_{0,\varepsilon}}(z) \frac{h_0(z)^2}{h_F(z)}\, dz \;=\; \int_0^\infty \mathbf{1}_{A_0}(z) \frac{h_0(z)^2}{h_F(z)}\, dz \;=\; \int_0^\infty \frac{h_0(z)^2}{h_F(z)}\, dz
\end{aligned}
$$

so that Lemma D.1.2 finishes the proof. □

D.2 Proofs and Technical Lemmas

Proof of Lemma D.1.1.
From Theorem 2.2.1 we know that $\int_{[x,\infty)} g(z-x)/\hat{h}_n(z)\,dH_n(z) \le 1$ for $x \ge 0$. Integration with respect to H_0 implies

$$\begin{aligned} 1 &\ge \int_{x\in[0,\infty)} \int_{z\in[x,\infty)} \frac{g(z-x)}{\hat{h}_n(z)}\,dH_n(z)\,dH_0(x) \\ &= \int_{z\in[0,\infty)} \frac{1}{\hat{h}_n(z)} \int_{x\in[0,z]} g(z-x)\,dH_0(x)\,dH_n(z) = \int_{[0,\infty)} \frac{h_0(z)}{h_F(z)}\,dH_n(z). \quad \square \end{aligned}$$

Note that Lemma D.1.1 can also be shown along the same lines as the proof of the Fenchel characterization (see Theorem 2.2.1) by considering

$$\lim_{\varepsilon\downarrow 0} \varepsilon^{-1} \Big(\Psi_n((1-\varepsilon)\hat{F}_n + \varepsilon F_0) - \Psi_n(\hat{F}_n) \Big) \quad \le \quad 0.$$

Proof of Lemma D.1.2.
One can easily see that using $F = F_0$ in (D.1.2) leads to equality. Now, assume that (D.1.2) holds and define $I(j)$ for a subdensity function j on $[0,\infty)$ as

$$I(j) \quad = \quad \int_0^\infty \frac{h_0(z)^2}{j(z)}\,dz.$$

Then $I(j)$ is minimized by h_0, i.e. $I(h_0) = \min_{\{j\,:\; j \text{ is a subdensity}\}} I(j)$, since

$$I(j) + 1 \;\ge\; \int_0^\infty \Big\{ \frac{h_0(z)^2}{j(z)} + j(z) \Big\}\,dz \;\ge\; \int_0^\infty \big\{ h_0(z) + h_0(z) \big\}\,dz \;=\; I(h_0) + 1$$

by minimizing the integrand pointwisely. Note that $I(h_0) = 1$ by definition of I. This implies $I(j) = 1$ for all subdensities j on $[0,\infty)$ due to (D.1.2) and the previous display.

To finish the proof assume that $F \ne F_0$ which implies that $h_F(z) \ne h_0(z)$ for z in an interval of positive length. Hence $I(h_F) > I(h_0) = 1$ contradicting $I(h_F) = 1$ and implying $F = F_0$. $\square$

Lemma D.2.1.
Assume vague convergence of F_n to F and let $\varepsilon > 0$. Then, for $D_\varepsilon = \tilde{D}_{\varepsilon^5,\varepsilon}$, this implies

(i) $\limsup_{n\to\infty} \sup_{D_\varepsilon^c} (F_n(x) - F(x)) \le 4\varepsilon^5$ *and*

(ii) $\liminf_{n\to\infty} \inf_{D_\varepsilon^c} (F_n(x) - F(x-)) \ge -4\varepsilon^5$

where $F(x-) = \lim_{h\downarrow 0} F(x-h)$.

PROOF.
Let $\varepsilon > 0$ and $x > 0$. Note that $F_n(x) \to F(x)$ as $n \to \infty$ whenever x is a continuity point of F. If F is discontinuous at x then one can find a small $\delta > 0$ such that $x+\delta$ and $x-\delta$ are continuity points of F in such a way that $F(x+\delta) - F(x) < \varepsilon^5/2$ and $F(x-\delta) - F(x-) > -\varepsilon^5/2$ by the right continuity of F. Since we have for n sufficiently large $F_n(x+\delta) - F(x+\delta) < \varepsilon^5/2$ and $F_n(x-\delta) - F(x-\delta) > -\varepsilon^5/2$ we get

$$\begin{aligned} F_n(x) - F(x) &\leq (F_n(x+\delta) - F(x+\delta)) + (F(x+\delta) - F(x)) < \varepsilon^5 \text{ and} \\ F_n(x) - F(x-) &\geq (F_n(x-\delta) - F(x-\delta)) + (F(x-\delta) - F(x-)) > \varepsilon^5 \end{aligned}$$

for sufficiently large n.

In order to derive uniform results on D_ε^c we consider the big discontinuities of F (those which are contained in D_ε) and the small ones separately. For that define a finite grid G with points $x_0 = 0$ and $x_{i+1} = \inf\{x \in D_\varepsilon^c : F(x) \geq F(x_i) + \varepsilon^5/2\}$ for $i > 0$ and add the boundary points of D_ε to G. Rename the resulting $k+1$ elements of G as $0 = x_0 \leq x_1 \leq \cdots \leq x_k$. By the definition of G we have for $i = 1, \ldots, k$

$$F(x_i) - F(x_{i-1}) \leq \frac{\varepsilon^5}{2} + \varepsilon^5 \leq 2\varepsilon^5.$$

Due to the finiteness of G we have for large n, say $n \geq n_0$,

$$\sup_G (F_n(x) - F(x)) \leq \varepsilon^5 \quad \text{and} \quad \inf_G (F_n(x) - F(x-)) \geq -\varepsilon^5.$$

Let $x \in D_\varepsilon^c$. Then there exists a $\mu \in \{1, \ldots, k-1\}$ such that $x \in [x_\mu, x_{\mu+1})$. And thus, for $n \geq n_0$,

$$\begin{aligned} &F_n(x) - F(x) \leq F_n(x_{\mu+1}) - F(x_\mu) \\ &\quad = F_n(x_{\mu+1}) - F(x_{\mu+1}) + F(x_{\mu+1}) - F(x_\mu) \leq \varepsilon^5 + 2\varepsilon^5 \leq 4\varepsilon^5 \end{aligned}$$

and

$$\begin{aligned} &F_n(x) - F(x-) \geq F_n(x_\mu) - F(x_{\mu+1}) \\ &\quad = F_n(x_\mu) - F(x_\mu-) + F(x_\mu-) - F(x_\mu) + F(x_\mu) - F(x_{\mu+1}) \\ &\quad \geq -\varepsilon^5 - \varepsilon^5 - 2\varepsilon^5 = -4\varepsilon^5 \end{aligned}$$

which finishes the proof since we obtain bounds independent of x. □

LEMMA D.2.2.
Let F_n converge vaguely to its limit function F and define the sequence $(\Phi_n)_{n=1}^\infty$ by $\Phi_n(z) = -\int_0^z g'(w) F_n(z-w)\, dw$ for $z \geq 0$. Then Φ_n converges on $[0, \infty)$ uniformly to $\Phi(z) = -\int_0^z g'(z-w) F(w)\, dw$.

PROOF.
Let $z \in [0, \infty)$. Then for all $w > 0$ we get that $-g'(w)F_n(z-w)\mathbf{1}_{[0,z]}(w)$ converges almost everywhere to $g'(w)F(z-w)\mathbf{1}_{[0,z]}(w)$ as $n \to \infty$ since $F_n(x) \to F(x)$ almost everywhere. Hence, by the dominated convergence theorem, Φ_n converges pointwisely to Φ on $[0, \infty)$. The functions Φ_n as well as Φ are bounded and monotone nondecreasing so that this pointwise convergence can be strengthened to uniform convergence on $[0, \infty)$. □

PROOF OF LEMMA D.1.3.
Let $0 < \varepsilon < (\sqrt[3]{28g(0)})^{-1}$. Then there exists an $n_0 \geq 0$ such that $F_n(x) - F(x) \leq 6\varepsilon^5$ and $F_n(x) - F(x-) \geq -5\varepsilon^5$ for all $x \in D_\varepsilon^c$ and for all $n \geq n_0$ by Lemma D.2.1. This implies $F_n(x) - F(x) = (F_n(x) - F(x-)) + (F(x-) - F(x)) \geq -5\varepsilon^5 - \varepsilon^5 = -6\varepsilon^5$, for $x \in D_\varepsilon^c$, and hence

$$\sup_{D_\varepsilon^c} |\, F_n(x) - F(x) \,| \quad \leq \quad 6\varepsilon^5$$

for $n \geq n_0$. Using representation (2.1.2) and Lemma D.2.2 leads to

$$\begin{aligned} \sup_{D_\varepsilon^c} |\, h_{F_n}(z) - h_F(z) \,| &\leq \sup_{B_\varepsilon \cap D_\varepsilon^c} g(0)\, |F_n(z) - F(z)| + \sup_{[0,\infty)} |\Phi_n(z) - \Phi(z)| \\ &\leq \; 6g(0)\varepsilon^5 + g(0)\varepsilon^5 \; = \; 7g(0)\varepsilon^5 \end{aligned} \tag{D.2.1}$$

for sufficiently large n. This implies that we get

$$h_{F_n}(z) \geq \varepsilon^2/2 - 7g(0)\varepsilon^5 \quad \geq \quad \varepsilon^2/4 \tag{D.2.2}$$

for all $z \in B_\varepsilon \cap D_\varepsilon^c$ by the choice of ε. Thus,

$$\begin{aligned} &\sup_{B_\varepsilon \cap D_\varepsilon^c} \left| \frac{h_0(z)}{h_{F_n}(z)} - \frac{h_0(z)}{h_F(z)} \right| \\ &= \quad \sup_{B_\varepsilon \cap D_\varepsilon^c} \frac{h_0(z)}{h_{F_n}(z) h_F(z)} \, |h_{F_n}(z) - h_F(z)| \leq \frac{8g(0)}{\varepsilon^4} 7g(0)\varepsilon^5 \; = \; c\,\varepsilon \end{aligned}$$

for some positive constant c and for n sufficiently large. □

Bibliography

Aarts, L., Groeneboom, P., and Jongbloed, G. (2007). Estimating the upper support point in deconvolution. *Scandinavian Journal of Statistics*, 34:552–568.

Bach, G. (1967). Size distribution of particles derived from the size distribution of their sections. In: Stereology, H. Elias.

Bazaraa, M. S. and Shetty, C. (1979). *Nonlinear Programming: Theory and Algorithms.* Wiley, New York.

Bruckner, A. M., Bruckner, J. B., and Thomson, B. S. (1997). *Real Analysis.* Prentice-Hall, New Jersey.

Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence of deconvolving a density. *Journal of the American Statistical Association*, 83:1184–1986.

Cordy, C. B. and Thomas, D. R. (1997). Deconvolution of a distribution function. *Journal of the American Statistical Association*, 92:1459–1465.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1–38.

Devroye, L. (1989). Consistent deconvolution in density estimation. *The Canadian Journal of Statistics*, 17:235–239.

Diggle, P. J. and Hall, P. (1993). A fourier approach to nonparametric deconvolution of a density estimate. *Journal of the Royal Statistical Society, Series B*, 55:523–531.

Donauer, S., Groeneboom, P., and Jongbloed, G. (2009a). Asymptotic distribution theory for the MLE in deconvolution problems with decreasing kernel. in preparation.

Donauer, S., Groeneboom, P., and Jongbloed, G. (2009b). Global rate results for the MLE in a class of deconvolution models. *Statistics and Probability Letters*, 79:519–524.

Fan, J. (1991a). Global behavior of deconvolution kernel estimates. *Statistica Sinica*, 1:541–551.

Fan, J. (1991b). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, 19:1257–1272.

Fan, J. and Koo, J.-Y. (2002). Wavelet deconvolution. *IEEE Transactions on Information Theory*, 48:734–747.

Fox, G. Q. (1988). A morphometic analysis of synaptic vesicle distributions. *Brain Reseach*, 475:103–117.

Geskus, R. and Groeneboom, P. (1996). Asymptotically optimal estimation of smooth functionals for interval censoring, part 1. *Statistica Neerlandica*, 50:69–88.

Geskus, R. and Groeneboom, P. (1997). Asymptotically optimal estimation of smooth functionals for interval censoring, part 2. *Statistica Neerlandica*, 51:201–219.

Groeneboom, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probability Theory and Related Fields*, 81:79–109.

Groeneboom, P. (1996). Lectures on inverse problems, Lectures on Probability Theory and Statistics. Ecole d'Ete de Probabilite de Sait Flour XXIV.

Groeneboom, P. (2009). Integral equations and inverse statistical problems. To appear in the "Zürich series for advanced mathematics", European Mathematical Society Publishing House.

Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2001). Estimation of a convex function: Characterizations and asymptotic theory. *The Annals of Statistics*, 29:1653–1698.

Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2008a). The support reduction algorithm for computing nonparametric function estimates in mixture models. *Scandinavian Journal of Statistics*, 35:285–399.

Groeneboom, P., Maathuis, M. H., and Wellner, J. A. (2008b). Current status data with competing risks: Consistency and rates of convergence of the MLE. *The Annals of Statistics*, 36:1031–1063.

Groeneboom, P., Maathuis, M. H., and Wellner, J. A. (2008c). Current status data with competing risks: Limiting distribution of the MLE. *The Annals of Statistics*, 36:1064–1089.

Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhäuser Verlag, Boston.

Groeneboom, P. and Wellner, J. A. (1999). Computing Chernoff's distribution. *Journal of Computational and Graphical Statistics*, 10:388–400.

Hall, P. and Lahiri, S. N. (2008). Estimation of distributions, moments and quantiles in deconvolution problems. *The Annals of Statistics*, 36:2110–2134.

Hesse, C. H. (1995). Distribution function estimation from noisy observations. *Publ. Inst. Stat. Paris Sud*, 39:21–35.

Jongbloed, G. (1995). *Three Statistical Inverse Problems.* PhD thesis, Technische Universiteit Delft.

Jongbloed, G. (1998a). Exponential deconvolution: Two asymptotically equivalent estimators. *Statistica Neerlandica*, 52:6–17.

Jongbloed, G. (1998b). The iterative convex minorant algorithm for nonparametric estimation. *Journal of Computational and Graphical Statistics*, 7:310–321.

Jongbloed, G. and van der Meulen, F. (2008). Estimating a concave distribution function from data corrupted with additive noise. *to appear in The Annals of Statistics.*

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *The Annals of Statistics*, 18:191–219.

Kress, R. (1989). *Linear Integral Equations.* Springer, Berlin.

Liebow, C. and Rothman, S. (1973). Distribution of zymogen granule size. *American Journal of Physiology*, 225:258–262.

Liu, M. C. and Taylor, R. L. (1989). A consistent nonparametric density estimator for the deconvolution problem. *The Canadian Journal of Statistics*, 17:427–438.

Luenberger, D. G. (2005). *Linear and Nonlinear Programming.* Springer, New York.

Maestrini, C., Maerlotti, M., Vighi, M., and Malaguti, M. (1992). Second phase volume fraction and rubber particle size determination in rubber-thoughened polymers. *Journal of Mathematical Science*, 27(22):5994–6016.

Masry, E. and Rice, J. A. (1992). Gaussian deconvolution via differentiation. *The Canadian Journal of Statistics*, 20:9–21.

Mendelsohn, J. and Rice, J. (1982). Deconvolution of microfluorometric histograms with B-splines. *Journal of the American Statistical Association*, 77:748–753.

Ohser, J. and Sandau, K. (2000). Considerations about the estimation of the size distribution in Wicksell's corpuscle problem. Lecture Notes in Physics, Springer, Berlin.

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer Verlag New York.

Pollard, D. (1990). *Empirical Processes: Theory and Applications.* Hayward, California.

Prakasa Rao, B. (1969). Estimation of a unimodal density. *Sankhya A*, 31:23–36.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988). *Order Restricted Statiatical Inference*. John Wiley & Sons, New York.

Sjöström, L., Björntorp, P., and Vrana, J. (1971). Microscopic fat cell size measurements on frozen-cut adipose tissue in comparison with automatic determinations of osmium-fixed fat cells. *Journal of Lipid Research*, 12:521–530.

Starck, J.-L. and Bijaoui, A. (1994). Filtering and deconvolution by the wavelet transform. *Signal Processing*, 35:195–211.

Starck, J.-L. and Murtagh, F. (2002). Deconvolution in astronomy: A review. *Atronomical Society of the Pacific*, 114:1051–1069.

van de Geer, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *The Annals of Statistics*, 21:14–44.

van de Geer, S. (1996). Rates of convergence for the maximum likelihood estimator in mixture models. *Nonparametric Statistics*, 6:293–310.

van de Geer, S. (2000). *Empirical Processes in M-Estimation*. Cambridge University Press, Cambridge.

van de Geer, S. (2003). Asymptotic theory for maximum likelihood in nonparametric mixture models. *Computational Statistics & Data Analysis*, 41:453–464.

van der Vaart, A. W. (1991). On differentiable functionals. *The Annals of Statistics*, 19:178–204.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

van der Vaart, A. W. and Wellner, J. A. (2000). Preservation thoerems for Glivenko-Cantelli and uniform Glivenko-Cantelli classes. In High Dimensional Probability II, Evarist Gine, David Mason, and Jon A. Wellner, editors, Birkhäuser, Boston.

van Es, B., Gugushvili, S., and Spreij, P. (2007). A kernel type nonparametric density estimator for decompounding. *Bernoulli*, 13:672–694.

van Es, B., Jongbloed, G., and van Zuijlen, M. (1998). Isotonic inverse estimators for nonparametric deconvolution. *The Annals of Statistics*, 26:2395–2406.

Wicksell, S. (1925). The corpuscle problem. *Biometrika*, 17:84–99.

Wright, S. J. (1997). *Primal Dual Interior Point Methods*. SIAM Publications.

Zhang, C.-H. (1990). Fourier methods for estimating mixing densities and distributions. *The Annals of Statistics*, 18:806–831.

Summary

Asymptotics in Deconvolution Models

-Approximating Perfect Knowledge-

Huge amounts of data can nowadays easily be stored. But the simple availability of detailed information does not automatically lead to more precise descriptions, conclusions or predictions of a quantity of interest. In order to tap the potential of the data, one needs to choose a suitable mathematical model.
Natural restrictions sometimes prevent us from observing a specific quantity. For instance, think of observations that can be decomposed into signal plus noise and where the object of interest is the uncorrupted signal. Knowing this specific setting and interpreting the available data as observations of the signal, introduces an error that could be avoided. It can be accurately taken into account by using a model that is specifically designed for situations where one can only observe a quantity that is related to the quantity of interest by some known relation.
One specific example of such an inverse model is discussed in this thesis: the *deconvolution model*. There, an observation is the sum of the variable of interest and some independent random error.
We focus on the *asymptotics* of nonparametric distribution function estimators. In the deconvolution setting we aimed at deriving asymptotic properties of the maximum likelihood estimator (MLE), and in particular the pointwise limit distribution of the estimator evaluated at a fixed point, originally given as a conjecture in Groeneboom and Wellner (1992).

In Chapter 2-5 we study the MLE in a class of deconvolution models with decreasing noise densities satisfying certain smoothness conditions. The only very implicit characterization of this estimator in terms of Fenchel optimality conditions, makes it hard to straightforwardly derive its properties. Nevertheless, we prove various (global and local) asymptotic results for the MLE.
We succeeded in showing that the MLE is a well defined, piecewise constant estimator that only has points of jump at observation points and can be computed using iterative

methods as can be found in the literature. It converges uniformly to the underlying distribution function F_0 (provided that F_0 is continuous) and is therefore consistent with respect to the uniform metric. Assuming bounded support of the noise density, its last point of jump is shown to stay away from the upper support point of the sampling density. Moreover, the MLE converges at rate $n^{-1/3}$ to F_0, globally with respect to the L_2-metric and locally uniformly in a neighborhood of fixed and shrinking length around some x_0. The latter result implies that also the distance between two successive points of jump of the MLE converges at rate $n^{-1/3}$.

Apart from the intrinsic interest of the asymptotic results, they are important ingredients of our strategy to derive the pointwise limit distribution. Under the conjecture that specific functionals of the MLE converge at rate $n^{-1/2}$ to their true values, we discuss how to derive that the limit equals the derivative of the concave majorant at zero of a Brownian motion with negative quadratic drift. We do believe that the conjecture holds since it can be verified in specific cases, but is still being investigated.

Samenvatting

Asymptotiek in Deconvolutie Modellen

-Benadering van Perfecte Kennis-

Grote hoeveelheden data kunnen tegenwoordig eenvoudig worden opgeslagen. Echter, de beschikbaarheid van veel informatie leidt niet automatisch tot betere beschrijvingen, conclusies of voorspellingen. Om goed gebruik te maken van de data, moet een passend wiskundig model worden gekozen.
Er zijn situaties waarin we een specifieke grootheid waarin we geïnteresseerd zijn niet precies kunnen observeren. Denk bijvoorbeeld aan data die opgevat kunnen worden als de som van een signaal en ruis, waarbij we geïnteresseerd zijn in eigenschappen van het niet verstoorde signaal. Het is niet juist om de beschikbare data in deze situatie te interpreteren als het signaal zelf. Er dient een model te worden gebruikt dat de gemeten data opvat als realisatie van een stochastisch mechanisme dat via een zekere relatie samenhangt met de te bepalen eigenschappen van het niet verstoorde signaal.
Een specifiek voorbeeld van een dergelijk invers model wordt bestudeerd in dit proefschrift: het *deconvolutie model*. In dit model is een observatie gelijk aan de som van een grootheid waarvan we het stochastisch gedrag willen weten en een onafhankelijke fout. We richten ons hierbij op de *asymptotiek* van de niet-parametrische maximum likelihood schatter (MLE) voor de verdelingsfunctie. In het bijzonder zijn we geïnteresseerd in de puntsgewijze limietverdeling van de schatter geëvalueerd in een vast punt. In Groeneboom en Wellner (1992) is een vermoeden geformuleerd over deze limietverdeling.

In Hoofdstuk 2-5 bestuderen we de MLE in een klasse van deconvolutiemodellen, waarbij de ruisdichtheid dalend is en voldoet aan zekere gladheidscondities. De impliciete karakterisering van deze schatter in termen van de Fenchel optimaliteitscondities wordt gebruikt om eindige steekproef- en verscheidene (globale en lokale) asymptotische resultaten voor de MLE af te leiden.
Er wordt aangetoond dat de MLE een goed gedefinieerde, stuksgewijs constante schatter is die alleen sprongpunten heeft in observatiepunten. Tevens worden uit de literatuur bekende iteratieve optimalisatiemethoden gebruikt om de schatter te berekenen. De

schatter convergeert uniform in kans naar de onderliggende verdelingsfunctie F_0 (onder de voorwaarde dat F_0 continu is). Onder de aanname dat de support van de ruisdichtheid compact is, wordt bewezen dat het laatste sprongpunt asymptotisch in kans wegblijft van het rechter randpunt van de support. Bovendien convergeert de MLE met snelheid $n^{-1/3}$ naar F_0, zowel globaal in de L_2-metriek als ook lokaal uniform in een omgeving met vaste en kleiner wordende lengte rond een x_0. Uit dit laatste resultaat wordt afgeleid dat de afstand tussen twee opeenvolgende sprongpunten van de MLE van stochastische orde $n^{-1/3}$ is.
Afgezien van het belang van deze asymptotische resultaten op zichzelf, zijn ze belangrijke ingrediënten voor een strategie om de puntsgewijze limietverdeling van de MLE af te leiden. Aannemende dat specifieke functionalen van de MLE met snelheid $n^{-1/2}$ convergeren naar hun limietwaarden, leiden we de asymptotische puntsgewijze verdeling af. De aanname waaronder deze asymptotische verdeling wordt afgeleid, wordt ondersteund door het feit dat ze in specifieke gevallen geverifieerd kan worden. Een bewijs van de $n^{-1/2}$ convergentie van de functionalen is echter nog niet geleverd. Dat zal meer onderzoek vergen.

Curriculum Vitae

The author of this thesis was born on October 10, 1977 in Hannover, Germany. From 1990 to 1997 she attended the Gymnasium in Mellendorf, Germany. Academic studies in Mathematics with a minor in Statistics at the Universität Ulm, Germany, followed from 1997 to 2003 and were completed with the degree Diplom-Mathematikerin. The thesis 'Stochastische Modelle für Garantiekosten-Prognosen - Dichteschätzung in Theorie und Praxis' summarizes the work on an industrial project on estimating the warranty costs for trucks at the Daimler Chrysler AG and was supervised by Prof. Dr. Ulrich Stadtmüller. During the academic year 2001 to 2002, Stefanie participated in the Masters program at the University of Wisconsin-Milwaukee, USA, and received a Master of Science in Mathematics. Collaboration with Prof. Dr. Jay Beder resulted in a thesis entitled 'Aliasing in irregular fractions of the 2^n factorial design'.
During almost all these years the author has been working as a Teaching Assistant, in Germany as well as in the United States. After graduating in 2003 she continued teaching as a lecturer at various student help centers in Ulm, Germany, and studied at the Freie Theologische Ausbildungsstätte in Kirchberg, Germany, for three months in 2004. At the end of the same year the PhD research under the supervision of Prof. Dr. Piet Groeneboom and Prof. Dr. Geurt Jongbloed at the Vrije Universiteit Amsterdam, Netherlands, was started.
Stefanie has accepted a position as Akademische Rätin at the Gottfried Wilhelm Leibniz Universität Hannover, Germany.