myjournal manuscript No.
(will be inserted by the editor)

Robust Estimators for Estimating Discontinuous Functions

Christine H. Müller[⋆]

Carl von Ossietzky University Oldenburg, Postfach 2503, D - 26111 Oldenburg, Germany

Received: date / Revised version: date

Abstract We study the asymptotic behavior of a wide class of kernel estimators for estimating an unknown regression function. In particular we derive the asymptotic behavior at discontinuity points of the regression function. It turns out that some kernel estimators based on outlier robust estimators are consistent at jumps.

Key words Robust Estimators, Nonparametric Regression, Discontinuous Functions, Asymptotic Behavior.

1 Introduction

We consider the problem of estimating an unknown function $\mu : [a, b] \to \Re$ at a point $t_0 \in [a, b]$ by observations $y_{-NN}, \ldots, y_{NN}$ at points $t_{-NN}, \ldots, t_{NN} \in [a, b]$. We assume that the observation y_{nN} is a realization of a random variable $Y_{nN} = \mu(t_{nN}) + Z_{nN}$, $n = -N, \ldots, N$, and that the errors $Z_{-NN}, \ldots, Z_{NN}$ are independent each with distribution P. One possibility of estimating $\mu(t_0)$ is to use kernel estimators. The most well-known kernel estimator is the mean kernel estimator which is a consistent estimator for smooth functions (see e.g. Eubank 1988). However the mean kernel estimator has the disadvantage that discontinuities are smoothed. This is not the case if the mean is replaced by the median. Other estimators which preserve discontinuities are local estimators based on other outlier robust estimators which follow the majority of the data. In particular high breakdown point estimators have this property. For example estimators with high breakdown point are the least trimmed squares estimators (Rousseeuw and Leroy 1987) and

⋆ *This work was supported by the grant Mu 1031/4-1 of the Deutsche Forschungsgemeinschaft*

the Cauchy estimator (Huber 1984, Mizera and Müller 1999). The superiority of such estimators for estimating two-dimensional functions appearing in image denoising was demonstrated in Müller (1999).

For local estimators based on robust estimators not much is known about their asymptotic properties. For M-estimators with monotone and continuous score function we know from a result of Härdle and Gasser (1984) that they are consistent estimators for smooth functions μ. However, the consistency result does not hold for discontinuous or non-monotone score functions. For the median kernel estimator which has a discontinuous, monotone score function an alternative proof of consistency was given by Koch (1996) by approximating the discontinuous score function by continuous score functions. Chu et al. (1998) treat the asymptotic behavior of a special M-estimator with redescending score function. But this approach does not include the most commonly used M-estimators with redescending score function as the Cauchy estimator. Also open is the consistency of other local estimators based on robust estimators as the least trimmed squares estimators. Moreover up to now, all consistency results concern smooth functions μ.

Here we present the asymptotic behavior of local estimators based on robust estimators. In particular we study their asymptotic behavior at discontinuity points of μ.

2 The model and the estimators

To model discontinuities we assume that $\mu(t) = \mu_1(t)$ for $t \leq t_0$, and $\mu(t) = \mu_2(t)$ for $t > t_0$. Each function $\mu_1, \mu_2 : [a, b] \to \Re$ has two continuous derivatives. Since μ_1 and μ_2 can be different, we have a jump at the point t_0. Let $c = \mu_2(t_0) - \mu_1(t_0)$ be the size of the jump and let P_θ the distribution of any random variable given by $\theta + Z$ where Z has distribution P. Now set $\theta = \mu_1(t_0)$. Then $\mu_1(t_0) + Z$ has distribution P_θ and $\mu_2(t_0) + Z$ has distribution $P_{\theta+c}$. Let $P_{\theta,\alpha} = \alpha P_\theta + (1 - \alpha) P_{\theta+c}$ be a mixture of the distributions P_θ and $P_{\theta+c}$ with fixed $\alpha \in (0, 1)$.

As estimator for $\theta = \mu_1(t_0)$ based on $y_N = (y_{-NN}, \ldots, y_{NN})^\top$ we use any estimator $\hat{\theta}_N : \Re^{2N+1} \to \Re$ which is asymptotically linear at $P_{\theta,\alpha}$, i.e. which satisfies

$$\sqrt{2N+1}\left(\hat{\theta}_N(Y_N) - \hat{\theta}(P_{\theta,\alpha}) - \frac{1}{2N+1}\sum_{n=-N}^{N} \psi(Y_{nN})\right) \longrightarrow 0 \qquad (1)$$

in probability if $Y_N = (Y_{-NN}, \ldots, Y_{NN})$ and Y_{nN} has distribution $P_{\theta,\alpha}$ for $n = -N, \ldots, N$, where $\int \psi \, dP_{\theta,\alpha} = 0$ and $\hat{\theta}$ is a functional on probability measures. Similarly, condition (1) is satisfied as well if Y_{nN} has distribution P_θ for $t_{nN} \leq t_0$, Y_{nN} has distribution $P_{\theta+c}$ for $t_{nN} > t_0$, and $\sum_{n=-N}^{N} 1_{(-\infty,t_0]}(t_{nN}) = \lceil (2N+1)\alpha \rceil$. There are many examples of estimators satisfying this asymptotic linearity condition. The median, Fréchet differentiable estimators as many M-estimators like the Cauchy estimator (Clarke 1983) and least trimmed squares estimators (Bednarski and Clarke 1993) are among them. Hence the kernel estimators considered in this paper are estimators based on an asymptotically linear estimator and the rectangular kernel.

3 Asymptotic behavior

To derive the asymptotic behavior of kernel estimators based on asymptotically linear estimators we make the classical assumption of nonparametric regression that, with growing sample size N, the division of the interval $[a, b]$ becomes finer and the support of the kernel, the band, becomes smaller although it contains more and more division points. The idea behind classical asymptotic considerations in nonparametric regression is that the observations behave within the band asymptotically like identically distributed observations, i.e. that the observations behave like having all the distribution P_θ. Usually this fact is not used directly in classical proofs of nonparametric regression. But it can be used directly by applying the concept of contiguous distributions. If there is a jump within the band then the identical distribution is destroyed. However, if the proportion of the band on the lefthand and righthand side of the jump point t_0 is converging, say against α and $(1-\alpha)$, respectively, then the observations behave like Y_{nN} having distribution P_θ for $t_{nN} \le t_0$ and distribution $P_{\theta+c}$ for $t_{nN} > t_0$. Hence contiguous distributions can be used also in this situation.

If $Y_{nN} = \mu_1(t_{nN}) + Z_{nN}$ for $t_{nN} \le t_0$ and $Y_{nN} = \mu_2(t_{nN}) + Z_{nN}$ for $t_{nN} > t_0$ we denote the distribution of $Y_N = (Y_{-NN}, \ldots, Y_{NN})$ by

$$Q^N = \bigotimes_{t_{nN} \le t_0} P_{\mu_1(t_{nN})} \otimes \bigotimes_{t_{nN} > t_0} P_{\mu_2(t_{nN})}.$$

If $Y_{nN} = \theta + Z_{nN}$ for $t_{nN} \le t_0$ and $Y_{nN} = \theta + c + Z_{nN}$ for $t_{nN} > t_0$ we denote the distribution of $Y_N = (Y_{-NN}, \ldots, Y_{NN})$ by

$$P^N = \bigotimes_{t_{nN} \le t_0} P_\theta \otimes \bigotimes_{t_{nN} > t_0} P_{\theta+c},$$

where we set $\theta = \mu_1(t_0)$. The proof of the asymptotic distribution of our estimators bases mainly on the fact that Q^N is contiguous to P^N. Then due to the Third Lemma of LeCam (see for example Hájek and Šidák 1967, p. 208), we must only derive the asymptotic distribution of the estimators under P^N.

We make the following assumptions:

A) $t_{nN} = t_{0N} + \frac{n}{N\sqrt{2N+1}}$ for $n = -N, \ldots, N$, where $N_\alpha := \lceil (2N+1)\alpha \rceil = \sum_{n=-N}^{N} 1_{(-\infty, t_0]}(t_{nN})$.

B) $\mu_1(t) = m_1(t) + \theta$ and $\mu_2(t) = m_2(t) + \theta + c$, where m_1 and m_2 have two continuous derivatives and satisfy $m_1(t_0) = 0 = m_2(t_0)$.

C) The distribution P of the errors Z_{nN} is the normal distribution with expectation equal to zero or has a λ-density f satisfying the following conditions: f has three continuous derivatives and there exists $M > 0$ and $\epsilon > 0$ with

$$\left|\frac{f'(z)}{f(z)}\right| \leq M, \quad \left|\frac{f''(z)}{f(z)}\right| \leq M, \quad \sup_{|\delta|\leq\epsilon}\left|\frac{f'''(z+\delta)}{f(z)}\right| \leq M, \text{ for all } z \in \Re.$$

D) The estimator $\hat{\theta}_N$ satisfies property (1) for P^N, where $\int \psi \, dP_{\theta,\alpha} = 0$.

If f_θ and $f_{\theta+c}$ denotes the densities of P_θ and $P_{\theta+c}$, respectively, we have the following result on the asymptotic behavior of an asymptotically linear estimator $\hat{\theta}_N$ at jumps.

Theorem 1 *If the Assumptions A) - D) are satisfied, then the asymptotic distribution of* $\sqrt{2N+1}\left(\hat{\theta}_N(Y_N) - \hat{\theta}(P_{\theta,\alpha})\right)$ *under* Q^N *is a normal distribution with expectation*

$$\alpha^2 \int \psi_\theta \, f'_\theta \, d\lambda \, \mu'_1(t_0) - (1-\alpha)^2 \int \psi_{\theta+c} \, f'_{\theta+c} \, d\lambda \, \mu'_2(t_0)$$

and variance

$$\alpha \int \psi_\theta^2 \, dP_\theta + \ (1-\alpha) \int \psi_{\theta+c}^2 \, dP_{\theta+c},$$

where $\psi_\theta(y) = \psi(y) - \int \psi dP_\theta$.

If $\hat{\theta}(P_{\theta,\alpha}) = \theta$ then the estimator $\hat{\theta}_N$ is weakly consistent at jumps, since the convergence in distribution of $\sqrt{2N+1}(\hat{\theta}_N(Y_N) - \theta)$ implies always the convergence in probability of $\hat{\theta}_N(Y_N)$ to θ. However the property $\hat{\theta}(P_{\theta,\alpha}) = \theta$ is not often satisfied. The median satisfies this property only if the jump c is equal to zero or P is concentrated at one point and $\alpha > \frac{1}{2}$. But for the least trimmed squares estimator, this property is also satisfied if P has compact support, the jump c is large enough and $1-\alpha$ is smaller than the proportion of trimmed observations. Hence the least trimmed squares estimator is more often consistent at jumps than the median. Since α cannot be $\frac{1}{2}$, the consistency at jumps derived here is only a special form of consistency. Moreover, $\alpha \neq \frac{1}{2}$ implies that there is usually an asymptotic bias expressed by the expectation of the asymptotic normal distribution which is unequal to zero.

However, for finite sample sizes N, the condition $\alpha > \frac{1}{2}$ is no restriction. Since the rectangular kernel has always an odd number of support points, the proportion of points t_{nN} with $t_{nN} \leq t_0$ is always greater than $\frac{1}{2}$ if t_0 lies in the center of the kernel. Hence, for approximating the finite sample behavior of the estimators by the asymptotic distribution, the condition $\alpha > \frac{1}{2}$ is even useful.

If there is no jump at t_0, that is $c = 0$, then many estimators satisfy (1) and $\hat{\theta}(P_{\theta,\alpha}) = \theta$ and this holds for $\alpha \neq \frac{1}{2}$ and $\alpha = \frac{1}{2}$ as well. Hence Theorem 1 provides the weak consistency of a wide class of estimators. Also in this case, an asymptotic bias can appear. However this asymptotic bias is zero if $\alpha = \frac{1}{2}$ and $\mu_1'(t_0) = \mu_2'(t_0)$. This means that the center of the kernel is t_0 and that μ is not only continuous but also differentiable at t_0 which are the assumptions used by Eubank (1988, p. 147) for the mean kernel estimators and by Härdle and Gasser (1984) for the M-kernel estimators. While our results are more general in the direction that they concern a wide class of estimators and jumps, the results of Eubank, Härdle and Gasser are more general in the direction of different choices of the bandwidth. Our case concerns only the case that the bandwidth is chosen as $\lambda = N^{-1/3}$.

4 Proof of the theorem

The proof of the theorem bases on the following lemmata.

Lemma 1 *Under Assumptions A) and B) we have*

a) $|m_1(t_{nN})| = O\left(\frac{1}{\sqrt{N}}\right)$, $|m_2(t_{nN})| = O\left(\frac{1}{\sqrt{N}}\right)$, *for* $n = -N, \ldots, N$,

b)
$$\lim_{N\to\infty} \frac{1}{\sqrt{2N+1}} \sum_{n=-N}^{N} 1_{(-\infty,t_0]}(t_{nN})\, m_1(t_{nN}) = m_1'(t_0)\, (-\alpha^2),$$
$$\lim_{N\to\infty} \frac{1}{\sqrt{2N+1}} \sum_{n=-N}^{N} 1_{(t_0,\infty)}(t_{nN})\, m_2(t_{nN}) = m_2'(t_0)\, (1-\alpha)^2,$$

c)
$$\lim_{N\to\infty} \sum_{n=-N}^{N} 1_{(-\infty,t_0]}(t_{nN})\, m_1(t_{nN})^2 = m_1'(t_0)^2\, a_1,$$
$$\lim_{N\to\infty} \sum_{n=-N}^{N} 1_{(t_0,\infty)}(t_{nN})\, m_2(t_{nN})^2 = m_2'(t_0)^2\, a_2,$$

where $a_1,\, a_2 \in (0,\infty)$ *are given by*

$$a_1 = \lim_{N\to\infty} \sum_{n=-N}^{N} 1_{(-\infty,t_0]}(t_{nN})\, (t_{nN}-t_0)^2,$$
$$a_2 = \lim_{N\to\infty} \sum_{n=-N}^{N} 1_{(t_0,\infty)}(t_{nN})\, (t_{nN}-t_0)^2.$$

Proof
a) Taylor expansion and the property

$$\frac{n}{N\sqrt{2N+1}} + \frac{1}{\sqrt{2N+1}} - \frac{N_\alpha}{N\sqrt{2N+1}} < t_{nN} - t_0 \leq \frac{n}{N\sqrt{2N+1}} + \frac{1}{\sqrt{2N+1}} - \frac{N_\alpha}{N\sqrt{2N+1}} + \frac{1}{N\sqrt{2N+1}} \tag{2}$$

provide for example for m_1

$$|m_1(t_{nN})| = |m_1'(\tilde{t}_{nN})\ (t_{nN} - t_0)| = O\left(\frac{1}{\sqrt{N}}\right).$$

b) Taylor expansion and property (2) provide

$$\begin{aligned}
&\frac{1}{\sqrt{2N+1}} \sum_{n=-N}^{N} 1_{(-\infty,t_0]}(t_{nN}) m_1(t_{nN}) \\
&= \frac{1}{\sqrt{2N+1}} \sum_{n=-N}^{N} 1_{(-\infty,t_0]}(t_{nN}) \,(m_1'(t_0)(t_{nN} - t_0) \\
&\quad + \frac{1}{2}\, m_1''(\tilde{t}_{nN})(t_{nN} - t_0)^2\Big) \\
&= \frac{1}{\sqrt{2N+1}} \sum_{n=-N}^{N} 1_{(-\infty,t_0]}(t_{nN}) \left(m_1'(t_0)(t_{nN} - t_0) + O\left(\frac{1}{N}\right)\right).
\end{aligned}$$

Since

$$\lim_{N\to\infty} \frac{1}{\sqrt{2N+1}} \sum_{n=-N}^{N} 1_{(-\infty,t_0]}(t_{nN})\ (t_{nN} - t_0) = -\alpha^2, \tag{3}$$

the assertion is proved for m_1. For m_2 the assertion follows similarly by using

$$\lim_{N\to\infty} \frac{1}{\sqrt{2N+1}} \sum_{n=-N}^{N} 1_{(t_0,\infty)}(t_{nN})\ (t_{nN} - t_0) = (1-\alpha)^2$$

instead of property (3).
c) Taylor expansion provides for example for m_1

$$\begin{aligned}
&\sum_{n=-N}^{N} 1_{(-\infty,t_0]}(t_{nN}) m_1(t_{nN})^2 \\
&= \sum_{n=-N}^{N} 1_{(-\infty,t_0]}(t_{nN}) \left(m_1'(t_0)(t_{nN} - t_0) + \frac{1}{2} m_1''(\tilde{t}_{nN})(t_{nN} - t_0)^2\right)^2 \\
&= \sum_{n=-N}^{N} 1_{(-\infty,t_0]}(t_{nN}) \left(m_1'(t_0)^2 (t_{nN} - t_0)^2 + O\left(\frac{1}{N\sqrt{N}}\right)\right). \Box
\end{aligned}$$

Lemma 2 *Let be $g_1,\ g_2 : \Re \to \Re$ functions such that $\int g_1(y)^2 P_\theta(dy) < \infty$ and $\int g_2(y)^2 P_{\theta+c}(dy) < \infty$. Then we have under the Assumptions A) and B) with a_1 and a_2 from Lemma 1 c)*

$$\sum_{n=-N}^{N} 1_{(-\infty,t_0]}(t_{nN})\, g_1(Y_{nN})\, m_1(t_{nN})^2 + \sum_{n=-N}^{N} 1_{(t_0,\infty)}(t_{nN})\, g_2(Y_{nN})\, m_2(t_{nN})^2 \overset{N\to\infty}{\longrightarrow} \int g_1(y) P_\theta(dy)\, m_1'(t_0)^2\, a_1 + \int g_2(y) P_{\theta+c}(dy)\, m_2'(t_0)^2\, a_2$$

in probability under P^N.

Proof
Chebychev inequality and Lemma 1 c) provide the assertion. □

Proof of the Theorem
Since f is the density of the error distribution P, the density f_θ of P_θ has the form $f_\theta(y) = f(y-\theta)$. The same holds for the densities $f_{\theta+c}$, $f_{\mu_1(t_{nN})}$, $f_{\mu_2(t_{nN})}$ of $P_{\theta+c}$, $P_{\mu_1(t_{nN})}$, $P_{\mu_2(t_{nN})}$. Let be $g_\theta = -\frac{f_\theta'}{f_\theta}$ and $h_\theta = \frac{f_\theta''}{f_\theta}$ and denote by p_N the λ^N-density of P^N and by q_N the λ^N-density of Q^N. For using LeCam's third lemma (see e.g. Hájek and Šidák 1967, p. 208) we need a representation of $\log \frac{q_N}{p_N}$.

If the errors have a normal distribution, then we have for $t_{nN} \le t_0$

$$\begin{aligned}\log \tfrac{f_{\mu_1(t_{nN})}(y_{nN})}{f_\theta(y_{nN})} &= -\tfrac{1}{2\sigma^2}\left[(y_{nN}-\mu_1(t_{nN}))^2 - (y_{nN}-\theta)^2\right] \\ &= -\tfrac{1}{2\sigma^2}\left[-2(y_{nN}-\theta)m_1(t_{nN}) + m_1(t_{nN})^2\right] \\ &= g_\theta(y_{nN})\, m_1(t_{nN}) - \tfrac{1}{2}\tfrac{1}{\sigma^2} m_1(t_{nN})^2,\end{aligned}$$

and analogously for $t_{nN} > t_0$

$$\log \tfrac{f_{m_2(t_{nN})}(y_{nN})}{f_{\theta+c}(y_{nN})} = g_{\theta+c}(y_{nN})\, m_2(t_{nN}) - \tfrac{1}{2}\tfrac{1}{\sigma^2} m_2(t_{nN})^2.$$

If the errors have not a normal distribution, then Taylor expansion, Lemma 1 a) and Assumption C) provide for $t_{nN} \le t_0$

$$\begin{aligned}\frac{f_{\mu_1(t_{nN})}(y_{nN})}{f_\theta(y_{nN})} &= \frac{f(y_{nN}-\mu_1(t_{nN}))}{f_\theta(y_{nN})} = \frac{f_\theta(y_{nN}-m_1(t_{nN}))}{f_\theta(y_{nN})} \\ &= 1 - \frac{f_\theta'(y_{nN})}{f_\theta(y_{nN})}\, m_1(t_{nN}) + \frac{1}{2}\,\frac{f_\theta''(y_{nN})}{f_\theta(y_{nN})}\, m_1(t_{nN})^2 \\ &\quad - \frac{1}{6}\,\frac{f_\theta'''(y_{nN}+\tilde{m}_{nN})}{f_\theta(y_{nN})}\, m_1(t_{nN})^3 \\ &= 1 + g_\theta(y_{nN})\, m_1(t_{nN}) + \frac{1}{2}\, h_\theta(y_{nN})\, m_1(t_{nN})^2 + O\left(N^{-3/2}\right).\end{aligned}$$

Set $\rho(y) = \log(1+y) - y + \frac{1}{2}y^2$. Then ρ is strictly increasing for $y > -1$ and $\lim_{y\to 0} \frac{\rho(y)}{y^2} = 0$. Assumption C) and Lemma 1 a) imply that there exists a constant M_0 with

$$N \left| \rho \left(g_\theta(y_{nN})\, m_1(t_{nN}) + \frac{1}{2}\; h_\theta(y_{nN})\, m_1(t_{nN})^2 + O\left(N^{-3/2}\right)\right)\right|$$
$$\le N \;\max \left\{ \left| \rho\left(\frac{M_0}{\sqrt{N}}\right)\right|, \left| \rho\left(\frac{-M_0}{\sqrt{N}}\right)\right| \right\} \;\stackrel{N\to\infty}{\longrightarrow}\; 0.$$

This implies

$$\begin{aligned}
&\log\left(\frac{f_{\mu_1(t_{nN})}(y_{nN})}{f_\theta(y_{nN})}\right)\\
&= g_\theta(y_{nN})\, m_1(t_{nN}) + \frac{1}{2}\; h_\theta(y_{nN})\, m_1(t_{nN})^2 + O\left(N^{-3/2}\right)\\
&\quad -\frac{1}{2}\; g_\theta(y_{nN})^2\, m_1(t_{nN})^2 + O\left(N^{-3/2}\right)\\
&\quad +\rho\left(g_\theta(y_{nN})\, m_1(t_{nN}) + \frac{1}{2}\; h_\theta(y_{nN})\, m_1(t_{nN})^2 + O\left(N^{-3/2}\right)\right)\\
&= g_\theta(y_{nN})\, m_1(t_{nN}) + \frac{1}{2}\; h_\theta(y_{nN})\, m_1(t_{nN})^2 - \frac{1}{2}\; g_\theta(y_{nN})^2\, m_1(t_{nN})^2\\
&\quad +O\left(N^{-3/2}\right) + o(N^{-1}).
\end{aligned}$$

An analoguous expression holds for $t_{nN} > t_0$ where g_θ and h_θ are replaced by $g_{\theta+c}$ and $h_{\theta+c}$. Since $\int \frac{f''(y)}{f(y)} P(dy) = \int f''(y)dy = 0$, Lemma 2 implies

$$\begin{aligned}
&\sum_{n=-N}^{N} \big(1_{(-\infty,t_0]}(t_{nN})\; h_\theta(Y_{nN})\, m_1(t_{nN})^2\\
&\qquad +\; 1_{(t_0,\infty)}(t_{nN})\; h_{\theta+c}(Y_{nN})\, m_2(t_{nN})^2\big)\\
&\stackrel{N\to\infty}{\longrightarrow} \int h_\theta(y) P_\theta(dy)\; m_1'(t_0)^2\; a_1 + \int h_{\theta+c}(y) P_{\theta+c}(dy)\; m_2'(t_0)^2\; a_2 = 0
\end{aligned}$$

in probability under P^N. Thus we have

$$\begin{aligned}
&\log \frac{q_N(Y_N)}{p_N(Y_N)}\\
&= \sum_{n=-N}^{N} \big(1_{(-\infty,t_0]}(t_{nN})\; g_\theta(Y_{nN})\, m_1(t_{nN})\\
&\quad + 1_{(t_0,\infty)}(t_{nN})\; g_{\theta+c}(Y_{nN})\, m_2(t_{nN})\\
&\quad - \frac{1}{2}\; 1_{(-\infty,t_0]}(t_{nN})\; g_\theta(Y_{nN})^2\, m_1(t_{nN})^2\\
&\quad - \frac{1}{2}\; 1_{(t_0,\infty)}(t_{nN})\; g_{\theta+c}(Y_{nN})^2\, m_2(t_{nN})^2\Big)\\
&\quad + O(N^{-1/2}) + o_{P^N}(N^0).
\end{aligned}$$

This representation of $\log \frac{q_N(Y_N)}{p_N(Y_N)}$ holds also for normally distributed errors since, according to Lemma 2,

$$\sum_{n=-N}^{N} \left(1_{(-\infty,t_0]}(t_{nN}) \, \frac{1}{\sigma^2} \, m_1(t_{nN})^2 + 1_{(t_0,\infty)}(t_{nN}) \, \frac{1}{\sigma^2} \, m_2(t_{nN})^2 \right)$$

behaves in probability like

$$\sum_{n=-N}^{N} \big(1_{(-\infty,t_0]}(t_{nN}) \; g_\theta(Y_{nN})^2 \, m_1(t_{nN})^2 + 1_{(t_0,\infty)}(t_{nN}) \; g_{\theta+c}(Y_{nN})^2 \, m_2(t_{nN})^2 \big) \,.$$

Let a and b be arbitrary real numbers and set

$$W_{nN}(y) := a \, \frac{\psi(y)}{\sqrt{2N+1}} + b \, \big(1_{(-\infty,t_0]}(t_{nN}) \; g_\theta(y) \, m_1(t_{nN}) + 1_{(t_0,\infty)}(t_{nN}) \; g_{\theta+c}(y) \, m_2(t_{nN}) \big) \,.$$

We are now going to derive the asymptotic distribution of $W_N := \sum_{n=-N}^{N} W_{nN}(Y_{nN})$ under P^N by checking Lindeberg's condition. Since $\int g_\theta(y) P_\theta(dy) = -\int f_\theta'(y) dy = 0 = \int g_{\theta+c}(y) P_{\theta+c}(dy)$ and $\int \psi(y) \, P_{\theta,\alpha}(dy) = 0$, the expectation of W_N satisfies

$$\begin{aligned}
|\mathrm{E}(W_N)| &= \Bigg| \sum_{n=-N}^{-N+N_\alpha-1} \frac{a}{\sqrt{2N+1}} \int \psi(y) P_\theta(dy) + b \int g_\theta(y) P_\theta(dy) \, m_1(t_{nN}) \\
&\quad + \sum_{n=-N+N_\alpha}^{N} \frac{a}{\sqrt{2N+1}} \int \psi(y) P_{\theta+c}(dy) \\
&\quad + b \int g_{\theta+c}(y) P_{\theta+c}(dy) \, m_2(t_{nN}) \Bigg| \\
&= |a| \left| \frac{N_\alpha}{\sqrt{2N+1}} \int \psi(y) P_\theta(dy) + \frac{2N+1-N_\alpha}{\sqrt{2N+1}} \int \psi(y) P_{\theta+c}(dy) \right| \\
&\le |a| \, \sqrt{2N+1} \, \Bigg(\left| \frac{N_\alpha}{2N+1} - \alpha \right| \, \left| \int \psi(y) P_\theta(dy) \right| \\
&\quad + \left| \frac{2N+1-N_\alpha}{2N+1} - (1-\alpha) \right| \, \left| \int \psi(y) P_{\theta+c}(dy) \right| \\
&\quad + \left| \alpha \int \psi(y) P_\theta(dy) + (1-\alpha) \int \psi(y) P_{\theta+c}(dy) \right| \Bigg) \\
&\le |a| \, \sqrt{2N+1} \, \left(\frac{1}{2N+1} \left| \int \psi(y) P_\theta(dy) \right| + \frac{1}{2N+1} \left| \int \psi(y) P_{\theta+c}(dy) \right| \right. \\
&\quad \left. + \left| \int \psi(y) \, P_{\theta,\alpha}(dy) \right| \right) \\
&\overset{N\to\infty}{\longrightarrow} = 0.
\end{aligned}$$

By using Lemma 1 b) and c), we obtain for the variance S_{1N}^2 of $\sum_{n=-N}^{N} 1_{(-\infty,t_0]}(t_{nN})\, W_{nN}(Y_{nN})$

$$
\begin{aligned}
S_{1N}^2 &= \sum_{n=-N}^{-N+N_\alpha-1} \int (W_{nN}(y) - \mathrm{E}(W_{nN}(Y_{nN})))^2 \, P_\theta(dy) \\
&= \sum_{n=-N}^{-N+N_\alpha-1} \int \left(a \, \frac{\psi(y) - \int \psi \, dP_\theta}{\sqrt{2N+1}} + b \, g_\theta(y) \, m_1(t_{nN}) \right)^2 \, P_\theta(dy) \\
&= a^2 \, \frac{N_\alpha}{2N+1} \int \psi_\theta^2 \, dP_\theta + b^2 \int g_\theta^2 \, dP_\theta \sum_{n=-N}^{-N+N_\alpha-1} m_1(t_{nN})^2 \\
&\quad +2ab \int \psi_\theta \, g_\theta \, dP_\theta \, \frac{1}{\sqrt{2N+1}} \sum_{n=-N}^{-N+N_\alpha-1} m_1(t_{nN}) \\
&\overset{N\to\infty}{\longrightarrow} a^2 \, \alpha \int \psi_\theta^2 \, dP_\theta + b^2 \int g_\theta^2 \, dP_\theta \, m_1'(t_0)^2 \, a_1 \\
&\quad + 2ab \int \psi_\theta \, g_\theta \, dP_\theta \, m_1'(t_0) \, (-\alpha^2).
\end{aligned}
$$

An analoguous expression holds for the variance S_{2N}^2 of $\sum_{n=-N}^{N} 1_{(t_0,\infty)}(t_{nN})\ W_{nN}(Y_{nN})$ so that the variance $S_N^2 := S_{1N}^2 + S_{2N}^2$ of $\sum_{n=-N}^{N} W_{nN}(Y_{nN})$ converges to $s^2 := a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\sigma_{12}$, where

$$
\begin{aligned}
\sigma_1^2 &:= \alpha \int \psi_\theta^2 \, dP_\theta + \ (1-\alpha) \int \psi_{\theta+c}^2 \, dP_{\theta+c}, \\
\sigma_2^2 &:= \int g_\theta^2 \, dP_\theta \, m_1'(t_0)^2 \, a_1 + \int g_{\theta+c}^2 \, dP_{\theta+c} \, m_2'(t_0)^2 \, a_2, \\
\sigma_{12} &:= \int \psi_\theta \, g_\theta \, dP_\theta \, m_1'(t_0) \, (-\alpha^2) + \int \psi_{\theta+c} \, g_{\theta+c} \, dP_{\theta+c} \, m_2'(t_0) \, (1-\alpha)^2.
\end{aligned}
$$

Moreover, according to Lemma 1 a), there exists a constant M_0 such that with Lemma 1 b) and c)

$$
\begin{aligned}
&\sum_{n=-N}^{-N+N_\alpha-1} \int_{|W_{nN}(y)|>\epsilon S_N} (W_{nN}(y) - \mathrm{E}(W_{nN}(Y_{nN})))^2 \, P_\theta(dy) \\
&\leq \quad a^2 \, \frac{N_\alpha}{2N+1} \int_{|a\psi(y)|+|bg_\theta(y)\, M_0|>\epsilon S_N \sqrt{N}} \psi_\theta^2 \, dP_\theta \\
&\qquad + b^2 \int_{|a\psi(y)|+|bg_\theta(y)\, M_0|>\epsilon S_N \sqrt{N}} g_\theta^2 \, dP_\theta \sum_{n=-N}^{-N+N_\alpha-1} m_1(t_{nN})^2 \\
&\quad + \quad 2ab \int_{|a\psi(y)|+|bg_\theta(y)\, M_0|>\epsilon S_N \sqrt{N}} \psi_\theta \, g_\theta \, dP_\theta \, \frac{1}{\sqrt{2N+1}} \sum_{n=-N}^{-N+N_\alpha-1} m_1(t_{nN}) \\
&\overset{N\to\infty}{\longrightarrow} \quad 0.
\end{aligned}
$$

The same holds for the sum over $t_{nN} > t_0$. Hence Lindeberg's condition is satisfied so that the asymptotic distribution of W_N under P^N is a normal distribution with expectation 0 and variance s^2. Then, using the representation (1) of the estimators and the convergence of

$$\sum_{n=-N}^{N} \big(1_{(-\infty,t_0]}(t_{nN})\, g_\theta(Y_{nN})^2\, m_1(t_{nN})^2 \\ + 1_{(t_0,\infty)}(t_{nN})\, g_{\theta+c}(Y_{nN})^2\, m_2(t_{nN})^2\big)$$

towards σ_2^2 according to Lemma 2, the asymptotic distribution under P^N of

$$a\,\sqrt{2N+1}\left(\hat{\theta}_N(Y_N) - \hat{\theta}(P_{\theta,\alpha})\right) + b\;\log\frac{q_N(Y_N)}{p_N(Y_N)}$$

is a normal distribution with expectation $-b\;\frac{1}{2}\sigma_2^2$ and variance s^2. The Cramer-Wold device (see e.g. Serfling 1980, p. 18) and LeCam's third lemma (see e.g. Hájek and Šidák 1967, p. 208) provide the asymptotic distribution of $\sqrt{2N+1}\left(\hat{\theta}_N(Y_N) - \hat{\theta}(P_{\theta,\alpha})\right)$ under Q^N and thus the assertion of the theorem. □

References

1. T. Bednarski and B.R. Clarke. Trimmed likelihood estimation of location and scale of the normal distribution. *Austral. J. Statist.* **35**, (1993) 141–153.
2. C.K. Chu, I.K. Glad, F. Godtliebsen, J.S. Marron. Edge-preserving smoothers for image processing. *J. Amer. Statist. Assoc.* **93**, (1998) 526–541.
3. B.R. Clarke. Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations. *Ann. Statist.* **11**, (1983) 1196–1205.
4. R.L. Eubank. *Spline Smoothing and Nonparametric Regression.* Dekker, New York, 1988.
5. W. Härdle and T. Gasser. Robust nonparametric function fitting. *J. Roy. Statist. Soc. Ser. B* **46**, (1984) 42–51.
6. J. Hájek and Z. Šidák. *Theory of Rank Tests.* Academic Press, New York, 1967.
7. P.J. Huber. Finite sample breakdown of M- and P-estimators. *Ann. Statist.* **12**, (1984) 119–126.
8. I. Koch. On the asymptotic performance of median smoothers in image analysis and nonparametric regression. *Ann. Statist.* **24**, (1996) 1648–1666.
9. I. Mizera and C.H. Müller. Breakdown points and variation exponents of robust M-estimators in linear models. *Ann. Statist.* **27**, (1999) 1164–1177.
10. C.H. Müller. On the use of high breakdown point estimators in the image analysis. *Tatra Mountains Math. Publ.* **17**, (1999) 283–293.
11. P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection.* John Wiley, New York, 1987.
12. R.J. Serfling. *Approximation Theorems of Mathematical Statistics.* John Wiley, New York, 1980.