Theory of Stochastic Processes
Vol. 29 (45), no. 2, 2025, pp. 1–29

D. FERGER

MINIMIZERS OF CONVEX U-PROCESSES AND THEIR DOMAINS OF ATTRACTION

In this paper, we study the minimizers of convex U-processes and their domains of attraction. U-processes arise in various statistical contexts, particularly in M-estimation, where estimators are defined as minimizers of certain objective functions. Our main results establish necessary and sufficient conditions for the distributional convergence of these minimizers, identifying a broad class of normalizing sequences that go beyond the standard square-root asymptotics with normal limits. We show that the limit distribution belongs to exactly one of the four classes introduced by Smirnov [20]. These results do not only extend Smirnov's theory but also generalize existing asymptotic theories for M-estimators, including classical results by Huber [12] and extensions to higher-degree U-statistics. Furthermore, we analyze the domain of attraction for each class, providing alternative characterizations that determine which types of statistical estimators fall into a given asymptotic regime.

1. Introduction

Let $X_1, \ldots, X_n, n \in \mathbb{N}$, be independent random variables defined on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in some measurable space $(S, \mathcal{S})$ and joint distribution $P = \mathbb{P} \circ X_i^{-1}$. Consider a map $h : S^l \times \mathbb{R} \to \mathbb{R}$ such that $h(\mathbf{x}, \cdot) : \mathbb{R} \to \mathbb{R}$ is convex for every $\mathbf{x} = (x_1, \ldots, x_l) \in S^l$ and $h(\cdot, t) : S^l \to \mathbb{R}$ is $\mathcal{S}^l$-Borel measurable and symmetric for each $t \in \mathbb{R}$. It induces an *U-process*

$$U_n^0(t) := \binom{n}{l}^{-1} \sum_{1 \le i_1 < \ldots < i_l \le n} h(X_{i_1}, \ldots, X_{i_l}, t), \ t \in \mathbb{R}, \tag{1}$$

with pertaining mean-function

$$U^0(t) = \mathbb{E}[h(X_1, \ldots, X_l, t)] = \int_{S^l} h(\mathbf{x}, t) dP^l(\mathbf{x}), \ t \in \mathbb{R}, \tag{2}$$

provided the integral in (2) exists, i.e.

$$\int_{S^l} |h(\mathbf{x}, t)| dP^l(\mathbf{x}) < \infty. \tag{3}$$

In statistics many parameters of interests are defined or can be represented as minimizing points of U^0 and in this case minimizers of U_n^0 are reasonable estimators for these. According to Huber [13], p.44, it is an advantage to replace U^0 with the function

$$U(t) := \int_{S^l} h(\mathbf{x}, t) - h(\mathbf{x}, t_0) dP^l(\mathbf{x}), \ t \in \mathbb{R}, \tag{4}$$

where t_0 is any fixed real number, e.g. $t_0 = 0$. If U^0 exists, then so does U and the respective sets of all minimizing points coincide. As to the existence of U let $D^+h(\mathbf{x}, t)$ and $D^-h(\mathbf{x}, t)$ denote the right-and left derivative, respectively, of $h(\mathbf{x}, \cdot)$ at point t. We will always assume that

(A0) $\int_{S^l} |D^{\pm} h(\mathbf{x}, t)| P^l(d\mathbf{x}) < \infty \quad \forall\, t \in \mathbb{R}$.

2000 *Mathematics Subject Classification.* Primary: 60F05,62E20; Secondary: 62G20.
Key words and phrases. M-estimators, convex U-processes, distributional convergence, domains of attraction.

Conclude from Proposition 2.1 in Ferger [5] (with $\mathfrak{X} = S^l$ and $Q = P^l$ there) that then $U(t)$ exists and is real-valued for every real t and that the map $U : \mathbb{R} \to \mathbb{R}$ is convex. Now the above mentioned advantage is that according to Proposition 2.2 in [5] condition (A0) in fact is strictly weaker than the integrability condition (3).

Moreover, by Proposition 2.4 in [5]

$$D^{\pm}U(t) = \int_{S^l} D^{\pm}h(\mathbf{x},t)P^l(d\mathbf{x}) = \mathbb{E}[D^{\pm}h(\mathbf{X},t)] \quad \forall\, t \in \mathbb{R}, \tag{5}$$

where $\mathbf{X} := (X_1, \ldots, X_l)$. According to Proposition 2.5 in Ferger [5] the set $\operatorname{Argmin}(U)$ of all minimizing points of U is given by

$$\begin{aligned} \operatorname{Argmin}(U) &= \{m \in \mathbb{R} : U(m) = \inf_{t\in\mathbb{R}} U(t)\} \\ &= \{m \in \mathbb{R} : D^-U(m) \le 0 \le D^+U(m)\} \\ &= \{m \in \mathbb{R} : \mathbb{E}[D^-h(\mathbf{X},m)] \le 0 \le \mathbb{E}[D^+h(\mathbf{X},m)]\}. \end{aligned} \tag{6}$$

Introduce the empirical measure

$$P_n := \binom{n}{l}^{-1} \sum_{1\le i_1<\ldots<i_l\le n} \delta_{(X_{i_1},\ldots,X_{i_l})},$$

where $\delta_{\mathbf{x}}$ denotes the Dirac-measure at point $\mathbf{x} = (x_1, \ldots, x_l) \in S^l$. If the integrating measure P^l in the definition of $U(t)$ is replaced by P_n, then one obtains

$$U_n(t) = \binom{n}{l}^{-1} \sum_{1\le i_1<\ldots<i_l\le n} h(X_{i_1},\ldots,X_{i_l},t) - h(X_{i_1},\ldots,X_{i_l},t_0),\ t \in \mathbb{R}.$$

Obviously, U_n differs from U_n^0 only up to a constant, whence both functions possess the same minimizing points. Assume that $h(\mathbf{x},t) \to \infty$ as $|t| \to \infty$ for all $\mathbf{x} \in S^l$. Then U_n is *coercive*, i.e., $U_n(t) \to \infty$ as $|t| \to \infty$. In this case the set $\operatorname{Argmin}(U_n)$ of all minimizing points of U_n is a non-empty compact interval. From Proposition 2.5 of Ferger [5] it follows that

$$\emptyset \ne \operatorname{Argmin}(U_n) = \{t \in \mathbb{R} : V_n^-(t) \le 0 \le V_n^+(t)\}, \tag{7}$$

where

$$V_n^{\pm}(t) = \binom{n}{l}^{-1} \sum_{1\le i_1<\ldots<i_l\le n} D^{\pm}h(X_{i_1},\ldots,X_{i_l},t). \tag{8}$$

In the important special case $l = 1$ the above statistics simplify to

$$V_n^{\pm}(t) = n^{-1}\sum_{i=1}^{n} D^{\pm}h(X_i,t).$$

First, we let m_n be the smallest minimizer of U_n (or equivalently of U_n^0). Later on we will drop this assumption. Since by Proposition 2.1 in Ferger [5] (with $Q = P_n$) the stochastic process U_n is convex, Corollary 3.3 in Ferger [6] ensures that m_n is a real random variable. Our main results give several necessary and sufficient conditions for the distributional convergence of $(m_n - m)/a_n$ with some positive sequence (a_n) converging to zero, which can be determined. It is also shown that the distribution function of the limit variable can belong to exactly one of four possible classes, which are specified precisely. These classes coincide with those of Smirnov [20]. All this we do under the following assumptions:

(A1) $D^+U(m) = 0 = D^-U(m)$.

If we put $V(t) := D^+U(t)$, then (A1) can be rewritten as $V(m-) = 0 = V(m)$, because $D^-U(t) = D^+U(t-) = V(t-)$ by Theorem 24.1 in Rockafellar [18]. From (A1) one can deduce with (6) that m is a minimizing point of U.

(A2) $\mathbb{E}[D^+h(\mathbf{X}, m)^2] < \infty$

If (A2) holds, then by Jensen's inequality and Fubini's theorem

$$\zeta := \int_S \mathbb{E}[D^+h(x, X_2, \ldots, X_l, m)]^2 P(dx) \leq \mathbb{E}[D^+h(\mathbf{X}, m)^2] < \infty, \quad l \geq 2,$$

is a finite non-negative real number. If $l = 1$, then $\zeta := \mathbb{E}[D^+h(X_1, m)^2] \in [0, \infty)$. It is supposed that actually

(A3) ζ is positive: $\zeta > 0$.

Observe that $K(\mathbf{x}) := D^+h(\mathbf{x}, m)$ is symmetric and measurable. Let $K_1(x) := \mathbb{E}[K(x, X_2, \ldots, X_l]), x \in \mathbb{R}$, be the *first associated function* of the kernel K. It follows from (A1) that $\zeta = \mathrm{Var}[K_1(X_1)]$ and so $\sigma^2 := l^2\zeta$ is the variance of the first term in the Hoeffding-decomposition of the U-statistic with kernel K, confer e.g. Koroljuk and Borovskich [14] or Serfling [19].

(A4) There exist points $t_0 < 0 < t_1$ such that $\mathbb{E}[\big(D^+h(\mathbf{X}, m+t) - D^+h(\mathbf{X}, m)\big)^2] < \infty$ for $t \in \{t_0, t_1\}$.

In the following theorem (a_n) is a sequence of positive real constants with $a_n \to 0$. Further, given a map $H : \mathbb{R} \to \mathbb{R}$ the set of all continuity points of H is denoted by C_H. In case that H is an increasing function, then $\tilde{H}$ defined by $\tilde{H}(x) := H(x+)$ is increasing as well, but in addition is right continuous and has left limits (rcll). Moreover, $\tilde{H} = H$ on C_H, confer Chung [4], pp.1-5. We say that a distribution function is *degenerated at zero*, if it corresponds to the Dirac measure at zero. Finally, let Φ_σ denote the distribution function of a centered normal random variable $N(0, \sigma^2)$ with variance σ^2.

Theorem 1.1. *Let H_n denote the distribution function of m_n. Assume that (A0)-(A4) hold. Recall that $V = D^+U$ and $\sigma^2 = l^2\zeta$. Then the following statements (9) and (10) are equivalent:*

$$H_n(a_n x + m) \to H(x) \quad \forall\, x \in C_H, \tag{9}$$

where H is some sub distribution function.

$$\delta_n(x) := \sqrt{n}V(m + a_n x) \to \delta(x) \in [-\infty, \infty] \quad \forall\, x \in D, \tag{10}$$

where D is some dense subset of $\mathbb{R}$.

In both cases, $H(x) = \Phi_\sigma(\delta(x))$ for all $x \in \mathbb{R}$ and $D = C_H$.

Moreover, if H is a distribution function, which is not degenerated at 0, then H belongs to exactly one of the following four disjoint classes.

$$\textbf{\textit{class 1}}: \; H(x) = \begin{cases} 0 & , \quad x < 0 \\ \Phi_\sigma\Big(cx^\alpha\Big) & , \quad x > 0 \end{cases} \qquad (c, \alpha > 0). \tag{11}$$

$$\textbf{\textit{class 2}}: \; H(x) = \begin{cases} \Phi_\sigma(-c|x|^\alpha) & , \quad x < 0 \\ 1 & , \quad x > 0 \end{cases} \qquad (c, \alpha > 0). \tag{12}$$

$$\textbf{\textit{class 3}}: \; H(x) = \begin{cases} \Phi_\sigma(-c|x|^\alpha) & , \quad x < 0 \\ \Phi_\sigma(dx^\alpha) & , \quad x > 0 \end{cases} \qquad (c, d, \alpha > 0). \tag{13}$$

$$\textbf{\textit{class 4}}: \; H(x) = \begin{cases} 0 & , \quad x < -c_1 \\ \frac{1}{2} & , \quad -c_1 < x < c_2 \\ 1 & , \quad x > c_2 \end{cases} \qquad (c_1, c_2 \geq 0, \max\{c_1, c_2\} > 0). \tag{14}$$

If Y is a real random variable with distribution function H, then the convergence in (9) is equivalent to the distributional convergence

$$\frac{m_n - m}{a_n} \xrightarrow{\mathcal{D}} Y. \tag{15}$$

We will see that our Theorem 1.1 is the source of a very rich asymptotic theory with unusual normalizing sequences (a_n) going far beyond the square root asymptotic normality. And even if H is only a sub distribution function, then the limit Y above is an extended random variable with values in $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\} = [-\infty, \infty]$ and the distributional convergence takes place in the (metric) space $\overline{\mathbb{R}}$.

In the literature estimators like m_n are called M-estimators. Apparently the first rigorous analysis of such estimators in case $l = 1$ goes back to Huber [12], who considers $S = \mathbb{R}$ and $h(x,t) = \phi(t-x)$ with $\phi : \mathbb{R} \to \mathbb{R}$ a convex function. Brøns et. al. [3] generalize Huber's idea to so-called ψ-means, where ψ corresponds to D^+h. Both authors show square root asymptotic normality. In a next step Habermann [10], Niemiro [17] and Hjort and Pollard [11] consider more general spaces S and extend the approach form the real line to the euclidian space $\mathbb{R}^d$. Here, too, the authors show convergence in distribution to a (multivariate) normal law with rate $\sqrt{n}$. Finally, Bose [2] extends these results to the case $l \geq 2$.

The paper is organized as follows: In section 2 we prove Theorem 1.1. Section 3 is dedicated to characterizing the domains of attraction of the four possible classes. This means that -alternatively to (10)- other equivalent conditions are specified to ensure that the convergence (9) holds and in addition it is indicated to which class H belongs. In both sections basic ideas of Smirnov [20] are used. In section 4 we extend our results for the smallest minimizer to general minimizing points of U_n. Section 5 is concerned with smooth derivatives. Namely, if $V = D^+U$ has left-and right derivatives at point m, then this results in a square root asymptotic with limits of normal type. Section 6 deals with the huge class of estimators based on mappings h of the shape $h(\mathbf{x}, t) = \phi(t - k(\mathbf{x}))$, where $\phi : \mathbb{R} \to \mathbb{R}$ is a convex function and $k : S^l \to \mathbb{R}$ is any symmetric and measurable map. The special choice $\phi(t) = t(1_{\{t \geq 0\}} - \alpha)$ and $k(x) = x$ yields Smirnov's [20] theory for α-quantile estimators, $\alpha \in (0,1)$. In fact, a good deal more examples are presented and discussed. The last section 7 (Appendix) contains several technical results, which are used in our proofs.

2. Proof of Theorem 1.1

The following observation is fundamental:

$$H_n(a_n x + m) = \mathbb{P}(m_n \leq m + a_n x) = \mathbb{P}(D^+U_n(m + a_n x) \geq 0), \tag{16}$$

where the last equality follows from Theorem 1.1 in Ferger [6].

For $\mathbf{i} = (i_1, \ldots, i_l)$ let $\mathbf{X_i} := (X_{i_1}, \ldots, X_{i_l})$. Then

$$D^+U_n(m + a_n x) = \binom{n}{l}^{-1} \sum_{\mathbf{i}} D^+h(\mathbf{X_i}, m + a_n x),$$

where the summation $\sum_{\mathbf{i}}$ extends over all $\mathbf{i}$ with $1 \leq i_1 < i_2 < \ldots < i_l \leq n$. Here, the number of summands is equal to $\binom{n}{l}$, whence

$$
\begin{aligned}
(17) \quad & \{D^+U_n(m + a_n x) \geq 0\} \\
= \; & \{\binom{n}{l}^{-1} \sum_{\mathbf{i}} (D^+h(\mathbf{X_i}, m + a_n x) - D^+U(m + a_n x)) \geq -D^+U(m + a_n x)\} \\
= \; & \{-\sqrt{n}\binom{n}{l}^{-1} \sum_{\mathbf{i}} (D^+h(\mathbf{X_i}, m + a_n x) - D^+U(m + a_n x)) \leq \sqrt{n} D^+U(m + a_n x)\} \\
= \; & \{-L_n(x) \leq \delta_n(x)\},
\end{aligned}
$$

where

$$L_n(x) = \sqrt{n}\binom{n}{l}^{-1} \sum_{\mathbf{i}} (D^+h(\mathbf{X_i}, m + a_n x) - D^+U(m + a_n x)).$$

By inserting the term $D^+h(\mathbf{X_i}, m)$ it follows that

$$
\begin{aligned}
(18) \qquad & L_n(x) = \sqrt{n}\binom{n}{l}^{-1} \sum_{\mathbf{i}} D^+h(\mathbf{X_i}, m) \\
+ \; & \sqrt{n}\binom{n}{l}^{-1} \sum_{\mathbf{i}} (D^+h(\mathbf{X_i}, m + a_n x) - D^+h(\mathbf{X_i}, m) - D^+U(m + a_n x)).
\end{aligned}
$$

One can easily see from the definition that $D^+h(\mathbf{x}, m)$ is symmetric in $\mathbf{x}$. Moreover, $\mathbb{E}[D^+h(\mathbf{X_i}, m)] = D^+U(m) = 0$ by (5) and assumption (A1). Thus the first summand in the decomposition (18) is a normalized and centered U-statistic with kernel $D^+h(\mathbf{x}, m)$. Consequently, (A3) allows us to apply the Central Limit Theorem for U-statistics, confer, e.g., Proposition 4.2.1 in Koroljuk and Borovskich [14], which yields:

$$\sqrt{n}\binom{n}{l}^{-1} \sum_{\mathbf{i}} D^+h(\mathbf{X_i}, m) \xrightarrow{\mathcal{D}} N(0, \sigma^2). \tag{19}$$

Let the second summand in the decomposition (18) be denoted by $N_n(x)$, i.e.

$$N_n(x) := \sqrt{n}\binom{n}{l}^{-1} \sum_{\mathbf{i}} (D^+h(\mathbf{X_i}, m + a_n x) - D^+h(\mathbf{X_i}, m) - D^+U(m + a_n x)). \tag{20}$$

Another application of (5) and (A1) gives that $N_n(x)$ is also a centered and normalized U-statistic with symmetric kernel

$$K_n(\mathbf{x}) \equiv K_{n,x}(\mathbf{x}) = D^+h(\mathbf{x}, m + a_n x) - D^+h(\mathbf{x}, m) - D^+U(m + a_n x).$$

By Lemma A on p.183 in Serfling [19] we have that for all $n \in \mathbb{N}$ and every $x \in \mathbb{R}$:

$$
\begin{aligned}
Var[N_n(x)] \; & \leq \; l \, Var[K_n(\mathbf{X})] \\
& = \; l \, Var[D^+h(\mathbf{X}, m + a_n x) - D^+h(\mathbf{X}, m)] \\
& \leq \; l \, \mathbb{E}[(D^+h(\mathbf{X}, m + a_n x) - D^+h(\mathbf{X}, m))^2] \qquad (21)
\end{aligned}
$$

If $x = 0$, then $N_n(x) = N_n(0) = 0$. For the case $x \neq 0$ we will use the subsequence criterion and that every positive null sequence contains a subsequence that converges to zero from above. Assume that $x > 0$. Let $(n_k)_{k \in \mathbb{N}}$ be a subsequence of the natural numbers. Then $(a_{n_k})_{k \in \mathbb{N}}$ contains a subsequence $(a_{n_{k_l}})_{l \in \mathbb{N}}$ such that $a_{n_{k_l}} \downarrow 0$ as $l \to \infty$. Let us write $b_l = a_{n_{k_l}}$ for short. Since $h(\mathbf{x}, \cdot)$ is a convex function, it follows that $D^+h(\mathbf{x}, \cdot)$ is an increasing and right-continuous function on $\mathbb{R}$ for each $\mathbf{x} \in S^l$, confer Theorem 24.1 in Rockafellar [18]. Therefore $Z_l := D^+h(\mathbf{X}, m + b_l x) - D^+h(\mathbf{X}, m) \geq 0$ and because of $m + b_l x \downarrow m$ the squares $Z_l^2 = (D^+h(\mathbf{X}, m + b_l x) - D^+h(\mathbf{X}, m))^2 \downarrow 0, l \to \infty$. According

to (100) of Lemma 7.1 in the appendix the Z_l^2 are $\mathbb{P}$-integrable for eventually all $l \in \mathbb{N}$. Thus an application of the Monotone Convergence Theorem yields that

$$\mathbb{E}[(D^+h(\mathbf{X}, m + b_l x) - D^+h(\mathbf{X}, m))^2] \to 0, l \to \infty. \tag{22}$$

Conclude from (21) that $Var[N_{n_{k_l}}(x)] \to 0$, whence by the subsequence criterion actually $Var[N_n(x)] \to 0$ as $n \to \infty$ for all $x > 0$. Next, assume that $x < 0$. Then $m + b_l x \uparrow m$, whence $0 \geq Z_l \uparrow D^-h(X, m) - D^+h(\mathbf{X}, m) =: Z$ upon noticing that $D^+h(\mathbf{x}, m - 0) = D^-h(\mathbf{x}, m)$ by Theorem 24.1 in Rockafellar [18]. Now, $-Z \geq 0$ and by linearity

$$\mathbb{E}[-Z] = \mathbb{E}[D^+h(\mathbf{X}, m)] - \mathbb{E}[D^-h(\mathbf{X}, m)] = D^+U(m) - D^-U(m) = 0,$$

where the second last equality holds by (5) and the last equality by assumption (A1). Thus $Z = 0$ $\mathbb{P}$-a.s. and since $Z_l^2 \downarrow Z^2 \overset{a.s.}{=} 0$ another application of the Monotone Convergence Theorem shows that (22) is true also for every negative x. With the same arguments as for positive x we obtain that $Var[N_n(x)] \to 0$ as $n \to \infty$ for all $x < 0$. In summary, $\mathbb{E}[N_n(x)] = 0$ for all $n \in \mathbb{N}$ and $Var[N_n(x)] \to 0, n \to \infty$ for all $x \in \mathbb{R}$. So, we can infer that $N_n(x) \xrightarrow{\mathbb{P}} 0$ for every $x \in \mathbb{R}$. Recall that

$$L_n(x) = \sqrt{n}\binom{n}{l}^{-1} \sum_{\mathbf{i}} D^+h(\mathbf{X_i}, m) + N_n(x) \quad \forall\, x \in \mathbb{R}.$$

Hence (19) and Slutsky's theorem ensure that

$$L_n(x) \xrightarrow{\mathcal{D}} N(0, \sigma^2) \quad \forall\, x \in \mathbb{R}. \tag{23}$$

The limit law (23) makes it easy to prove the following result.

Theorem 2.1. *If (A0)-(A4) are satisfied, then*

$$R_n(x) := H_n(a_n x + m) - \Phi_\sigma(\delta_n(x)) \to 0 \quad \forall\, x \in \mathbb{R}.$$

Proof. By equations (16) and (17) we have that

$$H_n(a_n x + m) = \mathbb{P}(-L_n(x) \leq \delta_n(x)) = G_{n,x}(\delta_n(x)) \tag{24}$$

for all $x \in \mathbb{R}$, where $G_{n,x}$ denotes the distribution function of $-L_n(x)$. It follows from (23) that $G_{n,x}(u) \to \Phi_\sigma(u), n \to \infty$, for every $u \in \mathbb{R}$. By Pólya's theorem one actually has that $||G_{n,x} - \Phi_\sigma|| := \sup_{u \in \mathbb{R}} |G_{n,x}(u) - \Phi_\sigma(u)| \to 0$. Since by (24)

$$|R_n(x)| = |G_{n,x}(\delta_n(x)) - \Phi_\sigma(\delta_n(x))| \leq ||G_{n,x} - \Phi_\sigma||,$$

the assertion follows. □

From Theorem 2.1, in turn, one easily obtains the equivalence of (9) and (10) in Theorem 1.1 and the relation between H and δ. We call this *the first part of Theorem 1.1*. Later on this knowledge is used to prove *the second part*.

Proof. (First part of Theorem 1.1) Assume (9) holds, where w.l.o.g. the function is rcll, because otherwise we can replace H with $\tilde{H}$, confer the discussion in front of Theorem 1.1. Then by Theorem 2.1

$$\Phi_\sigma(\delta_n(x)) = (H_n(a_n x + m) - H(x)) - R_n(x) + H(x) \to H(x) \quad \forall\, x \in C_H.$$

Since Φ_σ is invertible, it follows that $\delta_n(x) \to \Phi_\sigma^{-1}(H(x))$ for every $x \in C_H$. Thus (10) holds with $\delta(x) = \Phi_\sigma^{-1}(H(x)), x \in \mathbb{R}$, and $D = C_H$, which is known to be dense in $\mathbb{R}$.

Conversely, assume that (10) is true. Then $H_n(a_n x + m) = R_n(x) + \Phi_\sigma(\delta_n(x)) \to \Phi_\sigma(\delta(x))$ for all $x \in D$ by Theorem 2.1. Now, the functions G_n given by $G_n(x) := H_n(a_n x + m), x \in \mathbb{R}$, are increasing for every $n \in \mathbb{N}$, because the constants a_n are positive. This also applies to H with $H(x) = \Phi_\sigma(\delta(x))$. To see this first note that the functions δ_n are increasing, because $V = D^+U$ is increasing. Therefore, the limit function δ is increasing as well. This shows that in fact H is increasing. So, we have

that $G_n(x) \to H(x)$ for all x in a dense subset D and that all involved functions are increasing. Then Lemma 5.74 in Witting and Müller-Funk [22] says that the convergence holds for all $x \in C_H$, which is (9). (Notice that in that Lemma 5.74 it is required that the set D is also denumerable. However, checking the (short) proof shows that countability is not needed.) □

Note that under (A1), for all $n \in \mathbb{N}, \delta_n(x) \leq 0$ and $\delta_n(x) \geq 0$ according as $x \leq 0$ or $x \geq 0$. Thus by taking the limit $n \to \infty$ it follows that $\delta : \mathbb{R} \to [-\infty, \infty]$ has the following properties: it is increasing (as we saw in the proof above), non-positive on $(-\infty, 0]$, non-negative on $[0, \infty)$ and in particular $\delta(0) = 0$. In the following we will show that δ can indeed have only four different forms, see (33)-(36) below. We take the first step on the way there with the following result:

Proposition 2.1. *Assume that in Theorem 1.1 the limit H in (9) is a distribution function not degenerated at zero. Then the limit δ in (10) satisfies the following functional equations:*

$$\delta(x) = \sqrt{k}\, \delta(\alpha_k x) \quad \forall\, x \in \mathbb{R} \quad \forall\, k \in \mathbb{N}, \tag{25}$$

where the α_k are positive constants given by $\alpha_k = \lim_{n\to\infty} \frac{a_{nk}}{a_n}$.

Proof. It should be remembered that the first part of Theorem 1.1 has already been proved and therefore can be used in what follows. For each fixed $k \in \mathbb{N}, (nk)_{n\in\mathbb{N}}$ is a subsequence of the natural numbers. From (10) of Theorem 1.1 it follows that $\delta_{nk}(x) = \sqrt{nk}V(m + a_{nk}x) \to \delta(x)$ for every $x \in C_H$, whence

$$\sqrt{n}V(m + a_{nk}x) \to \frac{\delta(x)}{\sqrt{k}} =: \delta^{(k)}(x) \quad \forall\, x \in C_H.$$

Since $a_{nk} \to 0, n \to \infty$, another application of Theorem 1.1 (first part) yields that

$$H_n(a_{nk}x + m) \to \Phi_\sigma(\delta^{(k)}(x)) =: H^{(k)}(x) \quad \forall\, x \in C_H = C_{H^{(k)}}. \tag{26}$$

Here, the equality of the continuity-sets follows from $C_\delta = C_{\delta^{(k)}}$. By the convergence (26) and by the convergence in (9) of Theorem 1.1, the Convergence of Strict Types (confer Theorem 25, p.265 in Fristedt and Gray [7]) says that $\alpha_k = \lim_{n\to\infty} \frac{a_{nk}}{a_n} > 0$ and that $H^{(k)}(x) = H(\alpha_k x)$ for all $x \in \mathbb{R}$. Using the identity $H(x) = \Phi_\sigma(\delta(x)), x \in \mathbb{R}$, from Theorem 1.1 we obtain that $\Phi_\sigma(\delta(x)/\sqrt{k}) = \Phi_\sigma(\delta(\alpha_k x)), x \in \mathbb{R}$, which finally results in $\delta(x) = \sqrt{k}\, \delta(\alpha_k x)$ for all $x \in \mathbb{R}$, because Φ_σ is invertible. □

Note that $\alpha_1 = 1$ no matter how the sequence $(a_n)_{n\in\mathbb{N}}$ looks like. So, for $k = 1$ the functional equation (25) reduces to $\delta(x) = \delta(x)$, a tautology. In the analysis of the functional equations (25) we distinguish 3 possible cases.

Case 0: Suppose there exists some natural $k \geq 2$ such that $\alpha_k > 1$.

(i) Let $x > 0$. Then $\alpha_k x > x$ and thus $\delta(\alpha_k x) \geq \delta(x)$ by monotonicity of δ. As a consequence

$$\delta(x) \leq 0 \quad \text{or} \quad \delta(x) = \infty, \tag{27}$$

because otherwise $0 < \delta(x) < \infty$ and therefore $1 \leq \frac{\delta(\alpha_k x)}{\delta(x)} = \frac{1}{\sqrt{k}}$, where the last equality follows from (25) by division with $\delta(x) \in (0, \infty)$. It follows that $\sqrt{k} \leq 1$ in contradiction to $k \geq 2$. Next, we will show that $\delta(x) > 0$ for every $x > 0$. In fact, assume that there is some $x_0 > 0$ such that

$$\delta(x_0) \leq 0. \tag{28}$$

Then actually $\delta(x) \leq 0$ for all $x > 0$. To see this fix some $x > 0$ and put

$$y_l := \alpha_k^l x_0, \quad l \in \mathbb{N}_0.$$

Since $\alpha_k > 1$ and x_0 is positive, it follows that $y_l \uparrow \infty$ as $l \to \infty$. In particular, $y_l \geq x$ for some sufficiently large l. Consequently,

$$\delta(x) \leq \delta(y_l). \tag{29}$$

But

$$\delta(y_l) = \delta(\alpha_k \alpha_k^{l-1} x_0) = \delta(\alpha_k y_{l-1}) = k^{-1/2}\delta(y_{l-1}) = \ldots = k^{-l/2}\delta(y_0) = k^{-l/2}\delta(x_0).$$

Note that this formula applies to every $\alpha_k > 0$ and every $x_0 \in \mathbb{R}$. For later use we write down the basic equality:

$$\delta(y_l) = k^{-l/2}\delta(x_0). \tag{30}$$

Now it follows from (29),(30) and (28) that $\delta(x) \leq 0$ for all $x > 0$ and therefore $H(x) = \Phi_\sigma(\delta(x)) \leq 1/2$ for all $x > 0$ in contradiction to H is a distribution function. Thus we have shown that $\delta(x) > 0$ for all $x > 0$, which by (27) ensures that

$$\delta(x) = \infty \quad \forall\, x > 0. \tag{31}$$

(ii) Let $x < 0$. This case can be treated analogously as in (i) above. One obtains:

$$\delta(x) = -\infty \quad \forall\, x < 0. \tag{32}$$

Since $H(x) = \Phi_\sigma(\delta(x))$ it follows from (31) and (32) that $H(x) = 1$ for all $x > 0$ and $H(x) = 0$ for all $x < 0$, which means that H is a distribution function degenerated at zero in contrast to our assumption. This means that **Case 0 cannot occur**.

Case 1: There exists some natural $k \geq 2$ such that $\alpha_k < 1$.

(i) Let $x < 0$. Then $\alpha_k x > x$ and by (25), $\frac{1}{\sqrt{k}}\delta(x) = \delta(\alpha_k x) \geq \delta(x)$. As a consequence (+) $\delta(x) \geq \sqrt{k}\delta(x)$. We obtain that

$$\delta(x) \leq 0 \quad \text{or} \quad \delta(x) = \infty,$$

because otherwise $0 < \delta(x) < \infty$ and therefore $1 \geq \sqrt{k}$ by division in (+), a contradiction to $k \geq 2$. The case $\delta(x) = \infty$ is impossible for each $x < 0$, because $\delta(x) \leq \delta(0) = 0 < \infty$. So it remains the case (a) $\delta(x) \leq 0 \quad \forall\, x < 0$. Here we make a further case distinction, namely:

(a1) $\delta(x) > -\infty \quad \forall\, x < 0$.

(a2) There exists some $x_0 < 0$ such that $\delta(x_0) = -\infty$. We claim that then $\delta(x) = -\infty$ actually for all $x < 0$. To see this assume that there is some $x < 0$ with $\delta(x) > -\infty$. Put $y_l := \alpha_k^l x_0, l \in \mathbb{N}_0$. Since $\alpha_k \in (0,1)$ and $x_0 < 0$, it follows that $y_l \uparrow 0, l \to \infty$, whence $y_l \geq x$ for some l sufficiently large. An application of (30) gives that $\delta(x_0) = k^{l/2}\delta(y_l) \geq k^{l/2}\delta(x) > -\infty$, so that $\delta(x_0) > -\infty$ as well in contradiction to our assumption.

So with a view to (a), only the two cases remain in total:

(a1) $-\infty < \delta(x) \leq 0 \quad \forall\, x < 0$.

(a2) $\delta(x) = -\infty \quad \forall\, x < 0$.

Similarly, for $x > 0$ you get the only possible two cases:

(a3) $0 \leq \delta(x) < \infty \quad \forall\, x > 0$.

(a4) $\delta(x) = \infty \quad \forall\, x > 0$.

Of the 4 possible combinations, (a2) and (a4) are ruled out because the associated function $H(x) = \Phi_\sigma(\delta(x))$ is the same as in Case 0. This leaves the following 3 combinations:

(1) $\delta(x) = -\infty$ if $x < 0$ and $0 \leq \delta(x) < \infty$ if $x > 0$.

(2) $-\infty < \delta(x) \leq 0$ if $x < 0$ and $\delta(x) = \infty$ if $x > 0$.

(3) $-\infty < \delta(x) \leq 0$ if $x < 0$ and $0 \leq \delta(x) < \infty$ if $x > 0$.

Now the solutions of the functional equations (25) under the respective constraints (1)-(3)are:

$$\textbf{(1)}:\ \delta(x)=\begin{cases}-\infty & , \quad x<0\\ cx^{\alpha} & , \quad x>0\end{cases}\qquad (c,\alpha>0). \tag{33}$$

$$\textbf{(2)}:\ \delta(x)=\begin{cases}-c|x|^{\alpha} & , \quad x<0\\ \infty & , \quad x>0\end{cases}\qquad (c,\alpha>0). \tag{34}$$

$$\textbf{(3)}:\ \delta(x)=\begin{cases}-c|x|^{\alpha} & , \quad x<0\\ dx^{\alpha} & , \quad x>0\end{cases}\qquad (c,d,\alpha>0). \tag{35}$$

These solutions can be found in Smirnov [20] on pp.105-106, who in turn refers to Gnedenko [8].

Case 2: There exists some integer $k\geq 2$ such that $\alpha_k=1$. Then $(*)$ $\delta(x)=\sqrt{k}\delta(x)$ for all $x\in\mathbb{R}$. Assume that $\delta(x)\in\mathbb{R}\setminus\{0\}$. Then division by $\delta(x)$ in $(*)$ yields $1=\sqrt{k}$, in contradiction to $k\geq 2$. Thus $\delta(x)\in\{-\infty,0,\infty\}$. Recall that $\delta(0)=0$ and that δ is increasing. It follows that:

$$\textbf{(4)}:\ \delta(x)=\begin{cases}-\infty & , \quad x<-c_1\\ 0 & , \quad -c_1<x<c_2\\ \infty & , \quad x>c_2\end{cases}\qquad (c_1,c_2\geq 0,\max\{c_1,c_2\}>0). \tag{36}$$

Here, $c_1,c_2\geq 0$ results from $\delta(0)=0$. Finally, $H=\Phi_\sigma\circ\delta$ by Theorem 1.1 (first part), which is degenerated at zero if $c_1=0=c_2$. This is a contradiction to our assumption and thus $\max\{c_1,c_2\}>0$.

From (33)-(36) we immediately obtain the four possible classes (11)-(14) by another application of Theorem 1.1 (first part). This completes our proof of Theorem 1.1.

3. Domains of attraction of the limits H

Now that we have specified all possible limit distributions, the next step is to develop general conditions on a particular pair (P,h) such that (9) holds with limit H a distribution function not degenerated at 0. In that case (P,h) is said to be in the *domain of attraction of* H, symbolically: $(P,h)\in\mathcal{D}(H)$. More specifically, we would like to show how to find constants a_n so that $(m_n-m)/a_n$ converges in distribution, as well as determining the limit itself.

According to (5) we have that

$$V(t)=D^+U(t)=\int_{S^l}D^+h(\mathbf{x},t)P^l(d\mathbf{x})=\mathbb{E}[D^+h(\mathbf{X},t)],\quad t\in\mathbb{R}. \tag{37}$$

Recall that Condition (A1) has the form $V(m)=0=V(m-)$. Notice that here m need not to be uniquely determined. However, if this is the case, then there is an equivalent characterisation:

Lemma 3.1. *Suppose $V(m)=0=V(m-)$. Then m is unique if and only if*

$$V(x)<0\quad\forall\, x<m\qquad \textit{and}\qquad V(x)>0\quad\forall\, x>m. \tag{38}$$

Proof. Assume that m is unique, but there exists some $x<m$ such that $V(x)\geq 0$ or there exists some $y>m$ such that $V(y)\leq 0$. In the first case it follows that $0\leq V(x)\leq V(m)=0$, because V is increasing. Thus for (every) $z\in(x,m)$ we know that $V(z)=0$, whence $V(z-)=0$ as well. This means that z satisfies (A1), which contradicts the uniqueness of m. In the second case one can argue in the same way.

Conversely, assume that (38) holds, but there exists some $m'\neq m$ such that $V(m')=0=V(m'-)$. However, $V(m')=0$ immediately contradicts (38). □

The following theorems give necessary and sufficient conditions for $(P,h)\in\mathcal{D}(H)$, where H runs through the **classes 1-4**.

Theorem 3.1. *Assume that conditions (A0)-(A4) are satisfied. Then in order that $(P,h)\in\mathcal{D}(H)$ with $H\in$ **class 1** it is necessary and sufficient that m is unique,*

$$\frac{V(m+t)}{V(m-t)}\to 0,\quad t\downarrow 0, \tag{39}$$

and

$$\frac{V(m+\tau t)}{V(m+t)}\to \tau^{\alpha},\quad t\downarrow 0,\quad \forall\,\tau>0. \tag{40}$$

A possible choice for the constants in (9) is $a_n:=V^{-1}(1/\sqrt{n})-m\downarrow 0$. In this case, α in (40) and $c=1$ are the parameters in (11).

Proof. Sufficiency: Recall that V is increasing and right continuous with left limits (rcll). By (38) the function V is not flat at m and therefore V^{-1} is continuous at 0 by Lemma 1.18 in Witting [21]. Infer from that continuity and the definition of (a_n) that $a_n\downarrow 0$. Moreover, since $m+a_n=V^{-1}(1/\sqrt{n})$, it follows that

$$V(m+a_n-)=V(V^{-1}(1/\sqrt{n})-)\le\frac{1}{\sqrt{n}}\le V(V^{-1}(1/\sqrt{n}))=V(m+a_n). \tag{41}$$

Consequently, we have $\frac{1}{V(m+a_n)}\le\sqrt{n}\le\frac{1}{V(m+a_n-)}$ and therefore

$$\frac{V(m+a_nx)}{V(m+a_n)}\le\sqrt{n}V(m+a_nx)\le\frac{V(m+a_nx)}{V(m+a_n-)}. \tag{42}$$

Deduce from (40) with $\tau=x>0$ and $t=a_n\downarrow 0$ that

$$\lim_{n\to\infty}\frac{V(m+a_nx)}{V(m+a_n)}=x^{\alpha}\quad\forall\,x>0. \tag{43}$$

Fix some $\epsilon\in[0,1/2]$. Another application of (40) with $\tau=x/(1-\epsilon)$ and $t=a_n(1-\epsilon)\downarrow 0$ yields

$$\lim_{n\to\infty}\frac{V(m+a_nx)}{V(m+a_n(1-\epsilon))}=\big(\frac{x}{1-\epsilon}\big)^{\alpha}\quad\forall\,x>0\quad\forall\,\epsilon\in[0,1/2]. \tag{44}$$

Put $f_n(\epsilon):=V(m+a_nx)/V(m+a_n(1-\epsilon))$ and $f(\epsilon):=\big(\frac{x}{1-\epsilon}\big)^{\alpha}$ with $x>0$ fixed. Observe that f_n and f are increasing and in addition f is continuous. By Pólya's Theorem the sequence (f_n) converges uniformly on $[0,1/2]$, whence one can interchange the limits, i.e.

$$\lim_{\epsilon\downarrow 0}\lim_{n\to\infty}f_n(\epsilon)=\lim_{n\to\infty}\lim_{\epsilon\downarrow 0}f_n(\epsilon)=\lim_{n\to\infty}\frac{V(m+a_nx)}{V(m+a_n-)}.$$

Thus by taking the limit $\epsilon\downarrow 0$ in (44) we arrive at

$$\lim_{n\to\infty}\frac{V(m+a_nx)}{V(m+a_n-)}=x^{\alpha}\quad\forall\,x>0. \tag{45}$$

From (42) in combination with (43) and (45) it follows that

$$\delta_n(x)=\sqrt{n}V(m+a_nx)\to x^{\alpha}\quad\forall\,x>0. \tag{46}$$

Next, consider $x<0$. Then $V(m+a_nx)<0$ by (38) and $\sqrt{n}\ge 1/V(m+a_n)$ by (41) and therefore

$$\sqrt{n}V(m+a_nx)\le\frac{V(m+a_nx)}{V(m+a_n)}=\frac{V(m+a_nx)}{V(m-a_nx)}\cdot\frac{V(m-a_nx)}{V(m+a_n)}. \tag{47}$$

Here, on the right side the first factor is negative for all $n\in\mathbb{N}$ and its inverse converges to zero by condition (39) with $t=-a_nx\downarrow 0$. Thus the first factor converges to $-\infty$, whereas the second factor converges to $(-x)^{\alpha}>0$ by condition (40) with $\tau=-x>0$ and $t=a_n\downarrow 0$. So, the upper bound in (47) converges to $-\infty$, which finally shows that

$$\delta_n(x)=\sqrt{n}V(m+a_nx)\to-\infty\quad\forall\,x<0. \tag{48}$$

Infer from (46) and (48) that condition (10) of Theorem 1.1 is fulfilled, where δ is as in (33) with $c = 1$ and $D = \mathbb{R} \setminus \{0\}$. Consequently (9) follows with $H \in$ **class 1**.

Necessity: By assumption and Theorem 1.1 there exists some sequence $a_n \to 0$ such that

$$\delta_n(x) = \sqrt{n}V(m + a_n x) \to \delta(x) \quad \forall\, x \in C_H = \mathbb{R} \setminus \{0\} \tag{49}$$

with δ as in (33). Assume that m is not unique, i.e. $V(m') = 0$ for some $m' \neq m$. If $m' < m$, then by monotonicity $V(u) = 0$ for every $u \in [m', m]$. It follows that for $x < 0$, $m + a_n x \in [m', m]$ and thus $\delta_n(x) = \sqrt{n}V(m + a_n x) = 0$ for eventually all $n \in \mathbb{N}$, in contradiction to $\delta_n(x) \to \delta(x) = -\infty$. If $m' > m$, then one obtains in the same way that for $x > 0$, $\delta_n(x) = 0$ for eventually all $n \in \mathbb{N}$, in contradiction to $\delta_n(x) \to \delta(x) = cx^\alpha > 0$. This shows that m is unique.

For the derivation of (39) consider at first an arbitrary fixed $t > 0$. W.l.o.g. we may assume that (a_n) is strictly decreasing because otherwise extract a strictly decreasing subsequence. Then there exists some $n = n(t) \in \mathbb{N}$ such that $a_{n+1} \leq t \leq a_n$ and therefore

$$V(m + a_{n+1}) \leq V(m + t) \leq V(m + a_n). \tag{50}$$

Notice that $V(m - t) < 0 < V(m + t)$ by (38). It follows that

$$\frac{\sqrt{n+1}}{\sqrt{n}} \cdot \frac{\delta_n(1)}{\delta_{n+1}(-1)} \leq \frac{V(m+t)}{V(m-t)} \leq \frac{\sqrt{n}}{\sqrt{n+1}} \cdot \frac{\delta_{n+1}(1)}{\delta_n(-1)}. \tag{51}$$

Note that $n = n(t) \uparrow \infty$ as $t \downarrow 0$. Clearly the first ratio of the lower bound converges to 1, whereas by (49) the numerator of the second ratio converges to $c > 0$ and the denominator converges to $-\infty$. Consequently the lower bound converges to zero. This also applies to the upper bound, whence condition (39) follows.

Moreover, for each $\tau > 0$ we obtain

$$\frac{\sqrt{n}}{\sqrt{n+1}} \cdot \frac{\delta_{n+1}(\tau)}{\delta_n(1)} \leq \frac{V(m+\tau t)}{V(m+t)} \leq \frac{\sqrt{n+1}}{\sqrt{n}} \cdot \frac{\delta_n(\tau)}{\delta_{n+1}(1)}. \tag{52}$$

Thus another application of (49) results in (40). □

Theorem 3.2. *Assume that conditions (A0)-(A4) are satisfied. Then $(P, h) \in \mathcal{D}(H)$ with $H \in$ **class 2** if and only if m is unique and*

$$\frac{V(m-t)}{V(m+t)} \to 0, \quad t \downarrow 0, \tag{53}$$

and

$$\frac{V(m-\tau t)}{V(m-t)} \to \tau^\alpha, \quad t \downarrow 0, \quad \forall\, \tau > 0. \tag{54}$$

A possible choice for the constants in (9) is $a_n := m - V^{-1}(-1/\sqrt{n}) \downarrow 0$ is. Then α in (54) is the exponent in (12) and $c = 1$ there.

Proof. Analogously to the above proof. □

Theorem 3.3. *Assume that conditions (A0)-(A4) are satisfied. Then $(P, h) \in \mathcal{D}(H)$ with $H \in$ **class 3** if and only if m is unique and*

$$\frac{V(m+t)}{V(m-t)} \to A < 0, \quad t \downarrow 0, \tag{55}$$

and

$$\frac{V(m+\tau t)}{V(m+t)} \to \tau^\alpha, \quad t \downarrow 0, \quad \forall\, \tau > 0. \tag{56}$$

A possible choice for the constants in (9) is $a_n := V^{-1}(1/\sqrt{n}) - m \downarrow 0$. *Moreover,* α *in (56) is the exponent in (13) and the constants there are* $c = -1/A > 0$ *with* A *from (55) and* $d = 1$.

Proof. If-part: As in the proof of Theorem 3.1 one shows that $\delta_n(x) = \sqrt{n}V(m+a_nx) \to x^\alpha$ for all $x > 0$. Consider $x < 0$. Then

$$\sqrt{n}V(m + a_nx) = \sqrt{n}V(m + a_n(-x))\Big(\frac{V(m + a_n(-x))}{V(m + a_nx)}\Big)^{-1}.$$

Here the first factor converges to $(-x)^\alpha$ and the second factor to A^{-1} by condition (55).

We obtain that

$$\delta_n(x) \to \delta(x) = \begin{cases} -c|x|^\alpha & , \quad x < 0 \\ x^\alpha & , \quad x > 0, \end{cases}$$

where $c = -1/A > 0$. Thus Theorem 1.1 ensures that $(P, h) \in \mathcal{D}(H)$ with $H \in$ **class 3**.

Only-if-part: By Theorem 1.1 it follows that

$$\delta_n(x) \to \delta(x) = \begin{cases} -c|x|^\alpha & , \quad x < 0 \\ dx^\alpha & , \quad x > 0, \end{cases}$$

where c and d are positive constants. Therefore uniqueness of m can be shown as in the proof of Theorem 3.1 upon noticing that $\delta(x) \neq 0$ for all $x \neq 0$. Moreover, conditions (55) with $A = -d/c < 0$ and (56) follow from (51) and (52). □

Theorem 3.4. *Suppose that (A0)-(A4) hold with* m *being unique. Then* $(P, h) \in \mathcal{D}(H)$, *where* $H \in$ ***class 4*** *if and only if*

$$\sqrt{n}V(a_n^*x + m) \to \delta(x) = \begin{cases} -\infty & , \quad x < -c_1 \\ 0 & , \quad -c_1 < x < c_2 \\ \infty & , \quad x > c_2. \end{cases} \tag{57}$$

Here, $c_1, c_2 \geq 0$ *with* $\max\{c_1, c_2\} > 0$ *and*

$$a_n^* = \big(V^{-1}(1/\sqrt{n}) - V^{-1}(-1/\sqrt{n})\big)/(c_1 + c_2) \downarrow 0.$$

Proof. From the proof of Theorem 3.1 we already know that uniqueness of m entails continuity of V^{-1} at 0. Thus $a_n^* \downarrow 0$. Further, by Theorem 1.1 the relation $(P, h) \in \mathcal{D}(H)$ with $H \in$ **class 4** is true if and only if there is some sequence $a_n \to 0$ such that

$$\sqrt{n}V(a_nx + m) \to \delta(x). \tag{58}$$

Let $x = c_2 + \epsilon$ with $\epsilon \in (0, \infty)$. By (58) there exists some $n_0 = n_0(\epsilon) \in \mathbb{N}$ such that $\sqrt{n}V(a_nx+m) \geq 1$ and therefore $V(a_nx+m) \geq 1/\sqrt{n}$ for all $n \geq n_0$. For these n it follows that $a_nx + m \geq V^{-1}(1/\sqrt{n})$, by which we arrive at $a_n(c_2 + \epsilon) \geq V^{-1}(1/\sqrt{n}) - m =: \alpha_n$ for all $n \geq n_0$.

Let $x = -c_1 + \epsilon$ with $\epsilon \in (0, c_1 + c_2)$. Using (58) we find an integer $n_1 = n_1(\epsilon)$ such that $|\sqrt{n}V(a_nx + m)| \leq 1$, whence $V(a_nx + m) \geq -1/\sqrt{n}$ for every $n \geq n_1$. As a consequence one obtain the following equivalent relations: $a_nx + m \geq V^{-1}(-1/\sqrt{n}) \Leftrightarrow a_nx \geq -m + V^{-1}(-1/\sqrt{n}) \Leftrightarrow a_n(c_1 - \epsilon) \leq m - V^{-1}(-1/\sqrt{n}) =: \beta_n$.

If $x = c_2 - \epsilon, \epsilon \in (0, c_1 + c_2)$, then there exists a $n_2 \in \mathbb{N}$ such that for all $n \geq n_2$ the following implications hold:

$|\sqrt{n}V(a_nx + m)| \leq 1/2 \Rightarrow V(a_nx + m) \leq \frac{1}{2}\frac{1}{\sqrt{n}} < \frac{1}{\sqrt{n}} \Rightarrow a_nx + m < V^{-1}(1/\sqrt{n}) \Rightarrow a_n(c_2 - \epsilon) < \alpha_n$.

Similarly, one treats the case $x = -c_1 - \epsilon, \epsilon \in (0, \infty)$. To summarize, we obtain that for each $\epsilon \in (0, c_1 + c_2)$ there is some $N = N(\epsilon) \in \mathbb{N}$ such that

$$a_n(c_2 - \epsilon) < \alpha_n \leq a_n(c_2 + \epsilon), \quad a_n(c_1 - \epsilon) \leq \beta_n < a_n(c_1 + \epsilon) \quad \forall\, n \geq N.$$

If you add both inequalities and then divide by $2a_n$, the result with $d := (c_1 + c_2)/2$ is

$$d - \epsilon \leq \frac{\alpha_n + \beta_n}{2a_n} \leq d + \epsilon \quad \forall\, n \geq N. \tag{59}$$

Define

$$\theta_n := \frac{d}{\frac{\alpha_n+\beta_n}{2a_n}} - 1.$$

Then on the one hand it follows from (59) that $-\frac{\epsilon}{d+\epsilon} \leq \theta_n \leq \frac{\epsilon}{d-\epsilon}$, whence $\theta_n \to 0$. On the other hand a simple rearrangement of the formula for θ_n gives

$$a_n = \frac{\alpha_n + \beta_n}{2d}(1 + \theta_n) = a_n^*(1 + \theta_n).$$

Thus (58) is the same as

$$\sqrt{n}V(a_n^*(1 + \theta_n)x + m) \to \delta(x). \tag{60}$$

Our proof is complete, once we have shown that (60) is equivalent to (57). To do this, introduce $\delta_n^*(x) := \sqrt{n}V(a_n^*x + m)$ and $\delta_n^{**}(x) := \sqrt{n}V(a_n^*(1 + \theta_n)x + m)$. Assume that (57) holds. Consider $x \in (-c_1, c_2)$. By assumption $\delta_n^*(x) \to \delta(x) = 0$. Observe that δ_n^* and δ both are increasing and that δ is continuous. Therefore Pólya's theorem ensures that δ_n^* converges to δ uniformly on every compact set $K \subseteq (-c_1, c_2)$. As a consequence $\delta_n^*(x_n) \to \delta(x)$ for every sequence (x_n) with $x_n \to x$. Since $x_n := (1 + \theta_n)x \to x$, we receive that $\delta_n^{**}(x) = \delta_n^*(x_n) \to \delta(x)$. Let $x > c_2$. For some $z \in (c_2, x)$ we have that $x_n \geq z$ for eventually all $n \in \mathbb{N}$. Conclude from $\delta_n^{**}(x) = \delta_n^*(x_n) \geq \delta_n^*(z) \to \infty$ that $\delta_n^{**}(x) \to \infty$. Similarly, one obtains for $x < -c_1$ that $\delta_n^{**}(x) \to -\infty$. This shows that (60) is fulfilled. For the reverse conclusion notice that $\delta_n^*(x) = \delta_n^{**}(y_n)$ with $y_n = x/(1+\theta_n) \to x$, so that we can proceed as above. □

4. Extension to general minimizers of U_n

So far we only considered the smallest minimizing point m_n of U_n. Let m_n^* be the largest minimizing point of U_n, which by Corollary 3.3 in Ferger [6] is a real random variable. If $H_n^*(x) := \mathbb{P}(m_n^* < x)$, then

$$H_n^*(a_n x + m) = \mathbb{P}(m_n^* < m + a_n x) = \mathbb{P}(D^- U_n(m + a_n x) > 0),$$

where the second equality holds by Theorem 1.1 of Ferger [6]. (Notice that H_n^* is left continuous with right limits.) From here on, we can proceed in the same way as in the proof of Theorem 1.1 (first part), replacing D^+h with D^-h in assumptions (A2)-(A4). More precisely, the assumptions are now:

(B0) = (A0).
(B1) = (A1).
(B2) $\mathbb{E}[D^-h(X_1, \ldots, X_l, m)^2] < \infty$.
(B3) Put

$$\zeta^- := \int_S [\int_{S^{l-1}} D^-h(x_1, x_2, \ldots, x_l, m) dP^{l-1}(x_2, \ldots, x_l)]^2 dP(x_1)$$

for $l \geq 2$, and $\zeta^- := \int_S D^-h(x, m)^2 P(dx)$ for $l = 1$. We require that $\zeta^- > 0$. (Finiteness of ζ^- follows from (B2).)

(B4) There exist points $s_0 < 0 < s_1$ such that

$$\mathbb{E}[\big(D^-h(X_1, \ldots, X_l, m + t) - D^-h(X_1, \ldots, X_l, m)\big)^2] < \infty$$

for $t \in \{s_0, s_1\}$.

In complete analogy to the proof of Theorem 1 (part one), we obtain the following:

Interim Conclusion Assume that m is a real number such that (B0)-(B4) hold. Then the following statements (61) and (62) are equivalent:

$$H_n^*(a_n x + m) \to H^*(x) \quad \forall\, x \in C_{H^*}, \tag{61}$$

where H^* is a sub distribution function.

$$\delta_n^*(x) := \sqrt{n} D^- U(m + a_n x) \to \delta^*(x) \in [-\infty, \infty] \quad \forall\, x \in D^*, \tag{62}$$

where D^* is some dense subset of $\mathbb{R}$.

In both cases, $H^*(x) = \Phi_{\sigma^*}(\delta^*(x))$ and $D^* = C_{H^*}$. Moreover, $\sigma^{*2} = l^2 \zeta^-$.

Now that we have looked at the smallest and largest minimizer of U_n, let's look in general at arbitrary measurable minimizer $\tilde{m}_n \in \operatorname{Argmin}(U_n)$ a.s. By (7) every such minimizer is completely characterized by the inequalities

$$\int_{S^l} D^- h(\mathbf{x}, \tilde{m}_n) P_n(d\mathbf{x}) \le 0 \le \int_{S^l} D^+ h(\mathbf{x}, \tilde{m}_n) P_n(d\mathbf{x}) \text{ a.s.} \tag{63}$$

Theorem 4.1. *Let $\tilde{m}_n$ be a real random variable, which minimizes U_n a.s. Then all results from section 1 apply verbatim for $\tilde{m}_n$ and its distribution function $\tilde{H}_n(x) := \mathbb{P}(\tilde{m}_n \le x), x \in \mathbb{R}$.*

Proof. First notice the basic relation

$$H_n^* \le \tilde{H}_n \le H_n, \tag{64}$$

which holds to be true, because $m_n \le \tilde{m}_n \le m_n^*$ a.s.

We prove that the first part of Theorem 1.1 holds for m_n replaced by $\tilde{m}_n$. To this end observe that by Lemma 7.2 in the appendix the validity of assumptions (A0)-(A4) entail that of (B0)-(B4). From Theorem 2.1 it is known that

$$R_n(x) := H_n(a_n x + m) - \Phi_\sigma(\delta_n(x)) \to 0 \quad \forall\, x \in \mathbb{R}. \tag{65}$$

Since (B0)-(B4) hold, it follows analogously that

$$R_n^*(x) := H_n^*(a_n x + m) - \Phi_{\sigma^*}(\delta_n^*(x)) \to 0 \quad \forall\, x \in \mathbb{R}.$$

By Lemma 7.3 in the appendix, $\delta_n^*(x) = \delta_n(x)$ for all $x \in D_1$ and for all $n \in \mathbb{N}$. Moreover, $\zeta^- = \zeta$ by Lemma 7.2 in the appendix, whence $\sigma^* = \sigma$. Thus

$$R_n^*(x) = H_n^*(a_n x + m) - \Phi_\sigma(\delta_n(x)) \to 0 \quad \forall\, x \in D_1. \tag{66}$$

Put $\tilde{R}_n(x) := \tilde{H}_n(a_n x + m) - \Phi_\sigma(\delta_n(x)), x \in \mathbb{R}$. Then by the equality in (66) and by (64) it follows that $R_n^*(x) \le \tilde{R}_n(x) \le R_n(x)$ for all $x \in D_1$. Thus (65) and (66) make us to apply the sandwich-theorem resulting in $\tilde{R}_n(x) \to 0$ for each $x \in D_1$.

Now, assume that

$$\tilde{H}_n(a_n x + m) \to H(x) \quad \forall\, x \in C_H, \tag{67}$$

where H is increasing and rcll. From Lemma 7.3 in the appendix we know that $D_1 = A^c$, where A is a countable set. Then $D := (A \cup D_H)^c = D_1 \cap C_H$ is dense. If follows that

$$\Phi_\sigma(\delta_n(x)) = (\tilde{H}_n(a_n x + m) - H(x)) - \tilde{R}_n(x) + H(x) \to H(x) \quad \forall\, x \in D$$

and therefore

$$\delta_n(x) \to \delta(x) \quad \forall\, x \in D \tag{68}$$

with $\delta(x) = \Phi_\sigma^{-1}(H(x))$ as desired.

Conversely, assume that (68) holds. By Theorem 1.1, $H_n(a_n x + m) \to H(x) = \Phi_\sigma(\delta(x))$ for all $x \in C_H$. According to Lemma 7.3, $\delta_n^* \to \delta(x)$ for all $x \in D_2$, where D_2 is dense. Recall that $\sigma^* = \sigma$. Thus the Interim Conclusion says that $H^*(a_n x + m) \to H(x)$

for all $x \in C_H$. Hence the inequalities (64) and another application of the sandwich-theorem yield the convergence (67).

As to the second part observe that δ satisfies the functional equations (2.1). This can be shown in the same way as for the sequence (m_n). And the solutions of these have been derived in section 2. □

5. Square-root asymptotics

All our conditions on V describing the asymptotic behaviour of the distributions of the estimators $\tilde{m}_n$ are of local type, namely: How does V behave in a small neighborhood of m? The most simple case is when it is differentiable at m as in the following proposition. Its first part generalizes the results of Huber [12] and of Brøns et.al. [3]. The statement of the second part is new.

Proposition 5.1. *Let (A0)-(A4) be satisfied. Assume that V is differentiable at m with $V'(m) > 0$. Then*

$$\sqrt{n}(\tilde{m}_n - m) \xrightarrow{\mathcal{D}} N(0, \tau^2) \quad \text{with } \tau^2 = \frac{l^2\zeta}{V'(m)^2}. \tag{69}$$

If $V'(m) = 0$, then

$$\sqrt{n}(\tilde{m}_n - m) \xrightarrow{\mathcal{D}} Y \quad \text{in } \overline{\mathbb{R}}, \tag{70}$$

where $\mathbb{P}(Y = -\infty) = \mathbb{P}(Y = \infty) = \frac{1}{2}$.

Proof. Let $a_n = 1/\sqrt{n}$. Then by $V(m) = 0$

$$\delta_n(x) = \sqrt{n}V(m + \frac{x}{\sqrt{n}}) = \frac{V(m + \frac{x}{\sqrt{n}}) - V(m)}{\frac{x}{\sqrt{n}}}x \quad \forall\, 0 \neq x \in \mathbb{R}. \tag{71}$$

Recall that $\delta_n(0) = 0$. Therefore, $\delta_n(x) \to V'(m) \cdot x = \delta(x)$ for all $x \in \mathbb{R}$ and the assertions follow from Theorem 1.1. Indeed, $H(x) = \Phi_\sigma(V'(m)x)$ is the distribution function of $N(0, \sigma^2/(V'(m))^2)$, if $V'(m) > 0$. This shows (69). If $V'(m) = 0$, then $H(x) = 1/2$ for all $x \in \mathbb{R}$, which is the distribution function of Y. This shows the second part (70). □

In fact, Proposition 5.1 admits a slight generalization, which may be useful sometimes.

Remark 5.1. Assume that V has only left- and right derivatives $D^-V(m)$ and $D^+V(m)$ at point m, which both are positive. Then it follows from (71) and Theorem 1.1 that

$$\sqrt{n}(\tilde{m}_n - m) \xrightarrow{\mathcal{D}} Y,$$

where $Y = 1_{\{N<0\}} \frac{\sigma}{D^-V(m)} N + 1_{\{N\geq 0\}} \frac{\sigma}{D^+V(m)} N$ and $N = N(0,1)$.

In case that $D^-V(m) = 0 < D^+V(m)$ the equations (71) show that $\delta_n(x) \to \delta(x)$ for all $x \in \mathbb{R}$, where $\delta(x) = 0$ for $x < 0$ and $\delta(x) = D^+V(m) \cdot x$ for $x \geq 0$. Therefore by Theorem 1.1,

$$\sqrt{n}(\tilde{m}_n - m) \xrightarrow{\mathcal{D}} Y \quad \text{in } \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}, \tag{72}$$

where Y has sub distribution function $H(x) = 1_{\{x<0\}} \frac{1}{2} + 1_{\{x\geq 0\}} \Phi_\sigma(D^+V(m)x)$. Notice that $\mathbb{P}(Y = -\infty) = \frac{1}{2}$ and $\mathbb{P}(Y = \infty) = 0$. Moreover, $\mathbb{P}(Y \in (-\infty, 0]) = 0$ and $\mathbb{P}(0 < Y \leq x) = \Phi_\sigma(D^+V(m)x) - 1/2$.

In case that $D^-V(m) > 0 = D^+V(m)$ the convergence (72) holds, where Y has sub distribution function $H(x) = 1_{\{x\geq 0\}} \frac{1}{2} + 1_{\{x<0\}} \Phi_\sigma(D^-V(m)x)$. Here, $\mathbb{P}(Y = \infty) = \frac{1}{2}$ and $\mathbb{P}(Y = -\infty) = 0$.

Finally, if only $D^+V(m)$ exists and is positive, then

$$\mathbb{P}(\sqrt{n}(\tilde{m}_n - m) \leq x) \to \Phi_\sigma(D^+V(m)x) \quad \text{for all } x \geq 0. \tag{73}$$

Similarly, if only $D^-V(m)$ exists and is positive, then

$$\mathbb{P}(\sqrt{n}(\tilde{m}_n - m) \le x) \to \Phi_\sigma(D^-V(m)x) \quad \text{for all } x \le 0. \tag{74}$$

These last two results (73) and (74) follow analogously to the proof of Theorem 1.1 (first part).

Next, we think that the basic functions h are generated as follows: Let $\psi : S^l \times \mathbb{R} \to \mathbb{R}$ be a bivariate function such that $\psi(\mathbf{x}, \cdot)$ is monotone increasing for each $\mathbf{x} \in S^l$. W.l.o.g. we can assume that $\psi(\mathbf{x}, \cdot)$ is rcll, because othererwise use instead its rcll version $\tilde{\psi}(\mathbf{x}, \cdot)$, compare the section in front of Theorem 1.1. Define

$$h(\mathbf{x}, t) := \int_0^t \psi(\mathbf{x}, s)ds.$$

Then $h(\mathbf{x}, \cdot)$ is convex for every $\mathbf{x} \in S^l$, confer e.g. section 1.6 in Niculescu and Persson [16]. Moreover Theorem 24.2 in Rockafellar [18] says that $D^+h(\mathbf{x}, t) = \psi(\mathbf{x}, t+) = \psi(\mathbf{x}, t)$ by right continuity and $D^-h(\mathbf{x}, t) = \psi(\mathbf{x}, t-)$. Then

$$\zeta = \int_S \mathbb{E}[\psi(x, X_2, \ldots, X_l, m)]^2 P(dx),$$

if $l \ge 2$ and $\zeta = \mathbb{E}[\psi(X_1, m)^2]$, if $l = 1$. From (37) it follows that

$$V(t) = \int_{S^l} \psi(\mathbf{x}, t)P^l(d\mathbf{x}) = \mathbb{E}[\psi(\mathbf{X}, t)], \quad t \in \mathbb{R}. \tag{75}$$

Similarly, $V_n^\pm$ in (8) can be rewritten as

$$V_n^+(t) = \binom{n}{l}^{-1} \sum_{\mathbf{i}} \psi(\mathbf{X_i}, t) \quad \text{and} \quad V_n^-(t) = \binom{n}{l}^{-1} \sum_{\mathbf{i}} \psi(\mathbf{X_i}, t-).$$

Here, recall that $\mathbf{X_i} = (X_{i_1}, \ldots, X_{i_l})$ and that the summation $\sum_{\mathbf{i}}$ extends over all l-tuples $\mathbf{i} = (i_1, \ldots, i_l)$ with $1 \le i_1 < i_2 < \ldots < i_l \le n$. The following result is rather just a reformulation of Proposition 5.1. It enables us to compare it with a known result in the literature.

Corollary 5.1. *Let $\tilde{m}_n$ be a real random variable such that*

$$V_n^-(\tilde{m}_n) \le 0 \le V_n^+(\tilde{m}_n) \quad a.s. \tag{76}$$

Suppose that V in (75) is continuous and strictly increasing on $\mathbb{R}$ and differentiable at m. Moreover, assume that:

(0) *$\psi(\cdot, t)$ and $\psi(\cdot, t-)$ are P^l-integrable for all $t \in \mathbb{R}$.*
(1) *$V(m) = 0$.*
(2) *$\mathbb{E}[\psi(\boldsymbol{X}, m)^2] < \infty$.*
(3) *$\zeta > 0$.*
(4) *There exist reals $t_0 < 0 < t_1$ such that $\mathbb{E}[\big(\psi(\boldsymbol{X}, m + t) - \psi(\boldsymbol{X}, m)\big)^2] < \infty$ for $t \in \{t_0, t_1\}$.*

If $V'(m) > 0$, then

$$\sqrt{n}(\tilde{m}_n - m) \xrightarrow{\mathcal{D}} N(0, \tau^2) \quad \text{with } \tau^2 = \frac{l^2\zeta}{V'(m)^2}.$$

If $V'(m) = 0$, then

$$\sqrt{n}(\tilde{m}_n - m) \xrightarrow{\mathcal{D}} Y \quad \text{in } \overline{\mathbb{R}},$$

where Y is uniformly distributed on $\{-\infty, \infty\}$.

Proof. From (0)-(4) it follows that (A0)-(A4) are fulfilled. Furthermore, $\tilde{m}_n \in \operatorname{Argmin}(U_n)$ a.s. by (7). Thus the assertion follows from Proposition 5.1. □

The above corollary generalizes Theorem A on p.251 in Serfling [19] in many respects. Firstly, there the length $l = 1$. Secondly, the estimator $\tilde{m}_n$ must satisfy the equality (*) $V_n^+(\tilde{m}_n) = 0$. This assumption is stronger than our (76). Indeed, if $V_n^+(\tilde{m}_n) = 0$, then by monotonicity $V_n^-(\tilde{m}_n) \leq V_n^+(\tilde{m}_n) = 0$, whence (76) is fulfilled. But, if for instance $\psi(x,t) = 1_{\{x \leq t\}} - \alpha$ with $x, t \in \mathbb{R}$ and $\alpha \in (0,1)$, then $V_n^+(t) = F_n(t) - \alpha$, where F_n is the empirical distribution function pertaining to the sample $X_1, \ldots, X_n$. Consequently, an estimator $\tilde{m}_n$ satisfying (*) cannot exist, whenever $\alpha \notin \{k/n : 1 \leq k \leq n-1\}$, because (*) reduces to $F_n(\tilde{m}_n) = \alpha$. In particularly, in case that α is an irrational number, then for every $n \in \mathbb{N}$ no estimator exists. In contrast, the inequalities (76) simplify to $F_n(\tilde{m}_n-) \leq \alpha \leq F_n(\tilde{m}_n)$, which means that $\tilde{m}_n$ is an empirical α-quantile. Thirdly, it is required that $\mathbb{E}[\psi(\mathbf{X}, t)^2]$ is finite for t in a neighborhood of m, whereas we require this only for $t = m$. As to our condition (4) notice that it follows from that third assumption by the c_r-inequality, see Lemma 7.4 in the Appendix below. Fourthly, the function $\mathbb{E}[\psi(\mathbf{X}, t)^2]$ must be continuous at $t = m$. Fifthly, a further assumption in Theorem A in Serfling [19] is that $\tilde{m}_n$ is strongly consistent, which by Lemma A on p.249 in Serfling is guaranteed if $\psi(x,t)$ is continuous in t in a neighborhood of m. Finally, we have a result on distributional convergence even in the case that $V'(m) = 0$ (and actually if only the one-sided derivatives exist, confer Remark 5.1.)

6. Examples in statistics

In statistical applications the bivariate functions h typically is of the form

$$h(\mathbf{x}, t) = \phi(t - k(\mathbf{x})), \tag{77}$$

where $\phi : \mathbb{R} \to \mathbb{R}$ is convex and $k : S^l \to \mathbb{R}$ is symmetricc and measurable. W.l.o.g. we may assume that $\phi(0) = 0$, because shifts along the ordinate leave minimizer unchanged. Then ϕ admits the representation $\phi(t) = \int_0^t \varphi(s)ds, t \in \mathbb{R}$, where $\varphi = D^+\phi$ is increasing and rcll. Conversely, starting from a function φ that is increasing and rcll we can introduce $\phi(t) := \int_0^t \varphi(s)ds, t \in \mathbb{R}$. It follows that ϕ is convex and that $D^+\phi(t) = \varphi(t)$ and $D^-\phi(t) = \varphi(t-)$, whence ϕ is differentiable at t, if (and only if) φ is continuous at t. If φ is strictly increasing, then ϕ is strictly convex. Let $R := \mathbb{P} \circ k(\mathbf{X})^{-1}$ be the distribution of $k(\mathbf{X})$ and F the corresponding distribution function.

Recall that m minimizes U if and only if

$$V(m-) \leq 0 \leq V(m). \tag{78}$$

In case that V is continuous (as for instance when φ is continuous) this is the same as $V(m) = 0$. Notice that $D^+h(\mathbf{x}, t) = \varphi(t - k(\mathbf{x})+) = \varphi(t - k(\mathbf{x}))$ and $D^-h(\mathbf{x}, t) = \varphi(t - k(\mathbf{x})-)$. Using the Change of variable formula we see that (A0) is satisfied if and only if

$$\int_{\mathbb{R}} |\varphi(t - x\pm)| R(dx) < \infty \text{ for all } t \in \mathbb{R}. \tag{79}$$

In that case by (5)

$$V(t) = \int_{S^l} \varphi(t - k(\mathbf{x}))P^l(d\mathbf{x}) = \int_{\mathbb{R}} \varphi(t - x)R(dx) \quad \forall\, t \in \mathbb{R}. \tag{80}$$

Deduce from (8) that

$$V_n^+(t) = \binom{n}{l}^{-1} \sum_{1 \leq i_1 < \ldots < i_l \leq n} \varphi(t - k(X_{i_1}, \ldots, X_{i_l}))$$

and

$$V_n^-(t) = \binom{n}{l}^{-1} \sum_{1 \leq i_1 < \ldots < i_l \leq n} \varphi(t - k(X_{i_1}, \ldots, X_{i_l})-)$$

By (7) a point $\tilde{m}_n$ minimizes U_n if and only if

(81)
$$\sum_{1\le i_1<\ldots<i_l\le n} \varphi(\tilde{m}_n - k(X_{i_1},\ldots,X_{i_l})-) \le 0 \le \sum_{1\le i_1<\ldots<i_l\le n} \varphi(\tilde{m}_n - k(X_{i_1},\ldots,X_{i_l})).$$

If φ is continuous, then V_n^+ and V_n^- coincide and therefore $\tilde{m}_n$ solves the equation

$$\sum_{1\le i_1<\ldots<i_l\le n} \varphi(t - k(X_{i_1},\ldots,X_{i_l})) = 0,\ t\in\mathbb{R}. \tag{82}$$

Another use of the Change of variable formula shows that (A2) holds exactly when

$$\int_{\mathbb{R}} \varphi(m-x)^2 R(dx) < \infty. \tag{83}$$

Further, ζ in (A3) can be rewritten as $\zeta = \int_{\mathbb{R}} \varphi(m-x)^2 R(dx)$, if $l=1$ and

$$\zeta = \int_S \mathbb{E}[\varphi(m - k(x, X_2,\ldots,X_l))]^2 P(dx),\ \text{if } l\ge 2. \tag{84}$$

Finally, (A4) reduces to

$$\exists\, t_0<0<t_1 : \int_{\mathbb{R}} \big(\varphi(m+t-x)-\varphi(m-x)\big)^2 R(dx) < \infty \quad \text{for}\ \ t\in\{t_0,t_1\}. \tag{85}$$

The existence of a minimizing point m with the property (A1) is not a matter of course. Our next result gives general conditions which ensure the existence. Here, recall that F denotes the distribution function of R.

Lemma 6.1. *Let ϕ be strictly convex. Suppose that in addition*

(1) *ϕ is coercive with $\varphi(0-)\le 0\le\varphi(0)$*
or
(2) *ϕ is even and $Z := k(X_1,\ldots,X_l)$ is symmetric (not necessarily at zero).*

Morover, assume that ϕ is differentiable on $\mathbb{R}\setminus\{0\}$ or equivalently that φ is continuous on $\mathbb{R}\setminus\{0\}$. If F is continuous, then there exists a unique $m\in\mathbb{R}$ satisfying condition (A1). It fulfills the equation

$$\int_{\mathbb{R}\setminus\{t\}} \varphi(t-x)R(dx) = 0,\ t\in\mathbb{R}. \tag{86}$$

For ϕ differentiable on the entire real line the continuity assumption on F can be dropped and the integrals in (86) are over the entire space $\mathbb{R}$.

If (2) holds, then m is the center of symmetry of Z.

Proof. The assertion follows from Lemma 2.12 in Ferger [5] with $Q=R$ and $v=\phi$ there. □

In our first class of examples we consider very smooth functions φ.

Example 6.1. (**smooth case 1**) Assume that ζ in (84) is positive. Let φ be differentiable with a bounded derivative φ' and strictly increasing with $\varphi(0)=0$. If $\int x^2 R(dx) = \mathbb{E}[k(\mathbf{X})^2] < \infty$ and if R is not a discrete probability distribution, then there exists a unique minimizer m of U, such that

$$\sqrt{n}(\tilde{m}_n - m) \xrightarrow{\mathcal{D}} N(0,\tau^2) \quad \text{with } \tau^2 = \frac{l^2\zeta}{(\int \varphi'(m-x)R(dx))^2} \in (0,\infty).$$

Here, the integral $\int\varphi'(m-x)R(dx)$ in the denominator is positive and finite. If $\varphi'>0$, then the non-discreteness assumption on R can be dropped.

Proof. Since φ in particular is continuous, it follows that ϕ is differentiable. ϕ is strictly convex, because φ is strictly increasing. To see coercivity of ϕ consider $t > 0$. Choose a fixed point $\bar{t} \in (0, t)$. Then

$$\phi(t) = \int_0^t \varphi(s)ds = \int_0^{\bar{t}} \varphi(s)ds + \int_{\bar{t}}^t \varphi(s)ds \geq \varphi(\bar{t})(t - \bar{t}).$$

Notice that by strict monotonicity $\varphi(\bar{t}) > \varphi(0) = 0$. Taking the limit in the above inequality results in $\phi(t) \to \infty$ as $t \to \infty$. Similarly one shows that $\phi(t) \to \infty$ as $t \to -\infty$. Hence ϕ is coercive. Thus Lemma 6.1 guarantees the existence of an unique $m \in \mathbb{R}$ satisfying assumption (A1). Next we check that the remaining A-assumptions are fulfilled. Let $c \in \mathbb{R}$ with $\varphi' \leq c$. To see (79) observe that

$$|\varphi(t - x\pm)| = |\varphi(t - x)| = |\varphi(t - x) - \varphi(0)| \leq c|t - x| \leq c(|t| + |x|).$$

Consequently, $\int |\varphi(t - x\pm)|R(dx) \leq c(|t| + \int |x|R(dx)) < \infty$. Similarly, one obtains that $\int \varphi(m - x)^2 R(dx) \leq 2c^2(m^2 + \int x^2 R(dx)) < \infty$ and that $\int_{\mathbb{R}} (\varphi(m + t - x) - \varphi(m - x))^2 R(dx) \leq 2c^2 t^2 < \infty$, which yields (83) and (85). So all assumptions (A0)-(A4) are fulfilled. Further, by the Monotone Convergence Theorem V is continuous and it is increasing by the monotonicity of the integral. In fact, V is strictly increasing, because otherwise there are two real points $s < t$ such that $\int \varphi(t - x) - \varphi(s - x)R(dx) = 0$. Therfore $1 = R(\{x \in \mathbb{R} : \varphi(t - x) = \varphi(s - x)\}) = R(\emptyset)$, because φ is strictly increasing.

Furthermore, V is differentiable at m with $V'(m) = \int \varphi'(m - x)R(dx)$ by the Differentiation lemma. Here, $V'(m) > 0$ because otherwiese $\int \varphi'(m - x)R(dx) = 0$. Since $\varphi' \geq 0$ as φ is increasing, it follows that $R(A) = 1$, where $A = \{x \in \mathbb{R} : \varphi'(m - x) = 0\}$. Now, φ actually is strictly increasing, whence $B := \{t \in \mathbb{R} : \varphi'(t) = 0\}$ is denumerable. But $A = \{m - t : t \in B\}$, whence A is a denumerable subset of $\mathbb{R}$ with $R(A) = 1$. This means that R is a discrete probability distribution in contradiction to our assumption. Thus we have verified all conditions of Proposition 5.1, which yields the asymptotic normality.

If actually $\varphi' > 0$, then $A = \emptyset$ in contradiction to $R(A) = 1$. This shows the last part of the proposition. □

As a simple consequence of Example 6.1 we obtain the Central Limit Theorem (U-CLT) for non-degenerate U-statistics.

Example 6.2. (**U-statistics**) Let $\phi(t) = t^2$, so that $\varphi(t) = 2t, \varphi' = 2$ and $V(t) = 2(t - \mu)$, where $\mu = \int yR(dy) = \mathbb{E}[k(\mathbf{X})]$. Therefore, $m = \mu$, because V is continuous. Further, it follows from (82) that

$$\tilde{m}_n = \binom{n}{l}^{-1} \sum_{1 \leq i_1 < \ldots < i_l \leq n} k(X_{i_1}, \ldots, X_{i_l}),$$

i.e. $\tilde{m}_n$ is the U-statistic with kernel k. Recall the kernel K in (A3):

$$K(\mathbf{x}) := D^+ h(\mathbf{x}, m) = D^+ \phi(m - k(\mathbf{x})) = \varphi(m - k(\mathbf{x})) = -2(k(\mathbf{x}) - \mu).$$

It follows that $K_1(x) = -2(k_1(x) - \mu), x \in \mathbb{R}$, where k_1 is the first associate function of k. In (A3) it was explained that $\zeta = \mathrm{Var}(K_1(X_1)) = 4\mathrm{Var}(k_1(X_1))$, so that $\tau^2 = \frac{1}{4} l^2 \zeta = l^2 \mathrm{Var}(k_1(X_1))$. Thus Example 6.1 yields:

$$\sqrt{n}(\tilde{m}_n - m) = \sqrt{n}\{\binom{n}{l}^{-1} \sum_{1 \leq i_1 < \ldots < i_l \leq n} k(X_{i_1}, \ldots, X_{i_l}) - \mu\} \to N(0, l^2 \mathrm{Var}(k_1(X_1)).$$

This is in accordance with the Central Limit Theorem for (non-degenerate) U-statistcs.

If $l = 1$, then $\tilde{m}_n$ reduces to the arithmetic mean $\overline{X}_n = \frac{1}{n} \sum_{i=1}^n k(X_i)$ and if in addition $S = \mathbb{R}$ and k is equal to the identity, then $\tilde{m}_n$ further simplifies to the arithmetic mean of the data $X_1, \ldots, X_n$.

It should be mentioned that Example 6.1 does not give a new proof of the U-CLT, because the U-CLT is used in the proof of Theorem 1.1, which is the mother of all our examples.

The "generating" function $\varphi(t) = t$ yields the arithmetic mean (for $l = 1$), while $\varphi(t) = 1_{\{t\geq 0\}} - \frac{1}{2}$ gives the median. Sigmoid functions are a compromise between these two extremes.

Example 6.3. (**sigmoid φ**) Let $S = \mathbb{R}, l = 1$ and $k(x) = x$, whence F is the distribution function of the (real-valued) data $X_1, \ldots, X_n$.

(1) Let Φ be the distribution function of $N(0,1)$. Consider $\varphi = \Phi - \frac{1}{2}$ and $F(x) = \Phi(x-\mu)$ with $\mu \in \mathbb{R}$. Then $m = \mu$ by Lemma 6.1 and $\tilde{m}_n$ is the unique solution of the equation $\sum_{i=1}^n \Phi(t - X_i) = \frac{n}{2}, t \in \mathbb{R}$. An application of Example 6.1 yields

$$\sqrt{n}(\tilde{m}_n - \mu) \xrightarrow{\mathcal{D}} N(0, \frac{\pi}{3}).$$

(2) Let C be the distribution function of the standard Cauchy-distribution. Assume that $\varphi = C - \frac{1}{2}$ and $F(x) = C(x-a)$ with $a \in \mathbb{R}$. Similarly as in (1) above, $m = a$ and $\tilde{m}_n$ is the unique solution of the equation $\sum_{i=1}^n C(t - X_i) = \frac{n}{2}, t \in \mathbb{R}$. Another application of Example 6.1 yields

$$\sqrt{n}(\tilde{m}_n - a) \xrightarrow{\mathcal{D}} N(0, \frac{\pi^2}{3}).$$

Even though the class of estimators in Example 6.1 is already really big, it does not contain the L_p-estiamtor at least for $p \neq 2$. However, our next result includes also this one.

Example 6.4. Suppose that (79), (83) and (85) hold and that $\zeta > 0$. Let φ be continuous and strictly increasing, concave on $[0,\infty)$, convex on $(-\infty, 0)$ with $\varphi(0) = 0$ and differentiable on $\mathbb{R}\setminus\{0\}$. Then there exists a unique minimizer $m \in \mathbb{R}$ of U. If in addition

$$\frac{1}{t}(\varphi(t) - \varphi(-t))R([m, m+t]) \to 0, \ t \downarrow 0, \tag{87}$$

$$\frac{1}{t}(\varphi(t) - \varphi(-t))R((m-t, m]) \to 0, \ t \downarrow 0, \tag{88}$$

and R is not discrete, then the integral $\int_{\mathbb{R}\setminus\{m\}} \varphi'(m-x)R(dx)$ is a positive and finite real number and

$$\sqrt{n}(\tilde{m}_n - m) \xrightarrow{\mathcal{D}} N(0, \tau^2) \quad \text{with } \tau^2 = \frac{l^2\zeta}{(\int_{\mathbb{R}\setminus\{m\}} \varphi'(m-x)R(dx))^2} \in (0,\infty).$$

If actually $\varphi' > 0$ on $\mathbb{R}\setminus\{0\}$, then the non-discreteness assumption can be omitted.

In case that conversely φ is convex on $[0,\infty)$ and concave on $(-\infty, 0)$, then all statements remain valid.

As to (weak) sufficient conditions for the validity of (87) and (88) we refer to Remark 6.2 below.

Proof. The existence of an unique $m \in \mathbb{R}$ with property (A1) follows exactly as in the above proof. And also that V is continuous and strictly increasing. By Proposition 5.1 it remains to show that V is differentiable at m with a positive derivative. To see this

observe that for every $t > 0$,

$$\begin{aligned}\frac{V(m+t)-V(m)}{t} &= \int_{\mathbb{R}} \frac{\varphi(m+t-x)-\varphi(m-x)}{t} R(dx)\\ &= \int_{(-\infty,m)} \frac{\varphi(m+t-x)-\varphi(m-x)}{t} R(dx)\\ &+ \int_{[m,m+t]} \frac{\varphi(m+t-x)-\varphi(m-x)}{t} R(dx)\\ &+ \int_{(m+t,\infty)} \frac{\varphi(m+t-x)-\varphi(m-x)}{t} R(dx) =: I + II + III.\end{aligned}$$

If $x \in (-\infty, m)$, then $0 < m - x < m + t - x$, whence

$$1_{(-\infty,m)}(x)\frac{\varphi(m+t-x)-\varphi(m-x)}{t} \uparrow 1_{(-\infty,m)}(x)\varphi'(m-x),\ t \downarrow 0$$

as φ is concave on $[0,\infty)$, confer Theorem 23.1 of Rockafellar [18]. Therefore

$$I \uparrow \int_{(-\infty,m)} \varphi'(m-x)R(dx), t \downarrow 0,$$

by the Monotone Convergence Theorem upon noticing that the functions on the left side are integrable for all t according to (79). (Later on we will see that the integral on right side is finite.) If $x \in [m, m+t]$, then the integrand of the second integral II is greater than or equal zero and less than or equal $(\varphi(t) - \varphi(-t))/t$. Thus

$$0 \le II \le \frac{1}{t}(\varphi(t) - \varphi(-t))R([m, m+t]) \to 0,\ t \downarrow 0 \ \text{ by (87)}.$$

As to the third integral III notice that if $x > m + t$, then $0 > m + t - x > m - x$. Recall that the difference quotients by convexity of φ on $(-\infty, 0)$ are increasing in the variable $t > 0$, confer Theorem 23.1 of Rockafellar [18]. Consequently

$$\begin{aligned}0 \le 1_{(m+t,\infty)}(x)\frac{\varphi(m+t-x)-\varphi(m-x)}{t} &\le \frac{\varphi(m+t-x)-\varphi(m-x)}{t}\\ &\le \frac{\varphi(m+t^*-x)-\varphi(m-x)}{t^*} \quad \forall\, t \in (0, t^*],\end{aligned}$$

where t^* is any fixed positive real number. Thus the Dominated Convergence Theorem yields that

$$III \to \int_{(m,\infty)} \varphi'(m-x)R(dx) \in [0,\infty),\ t \downarrow 0.$$

This shows that V is right differentiable at m. For the derivation of left differentiability we start with

$$\begin{aligned}\frac{V(m-t)-V(m)}{-t} &= \int_{\mathbb{R}} \frac{\varphi(m-t-x)-\varphi(m-x)}{-t} R(dx)\\ &= \int_{(-\infty,m-t]} \frac{\varphi(m-t-x)-\varphi(m-x)}{-t} R(dx)\\ &+ \int_{(m-t,m]} \frac{\varphi(m-t-x)-\varphi(m-x)}{-t} R(dx)\\ &+ \int_{(m,\infty)} \frac{\varphi(m-t-x)-\varphi(m-x)}{-t} R(dx) =: A + B + C.\end{aligned}$$

Since

$$0 \le 1_{(-\infty,m-t]}(x)\frac{\varphi(m-t-x)-\varphi(m-x)}{-t} \le \frac{\varphi(m-t^*-x)-\varphi(m-x)}{-t^*} \quad \forall\, t \in (0, t^*],$$

it follows by the Dominated Convergence Theorem that

$$A \to \int_{(-\infty,m)} \varphi'(m-x)R(dx),\ t \downarrow 0,$$

where the integral is a finite real number as announced above. The summand B can be treated as the summand II above resulting with (88) in $B \to 0$. Finally, with the help of the Monotone Convergence Theorem one shows that $C \uparrow \int_{(m,\infty)} \varphi'(m-x)R(dx)$, where it is already known that the integral here is finite. Summing up we arrive at $V'(m) = \int_{\mathbb{R}\setminus\{m\}} \varphi'(m-x)R(dx)$. That this quantity is in fact positive follows as in the proof of Proposition 6.1. Here, in the definitions of the sets A and B one simply has to replace $\mathbb{R}$ by $\mathbb{R} \setminus \{m\}$. If convex and concave reverse their roles, the derivation of $V'(m) = \int_{\mathbb{R}\setminus\{m\}} \varphi'(m-x)R(dx)$ follows similarly. □

Remark 6.1. If in Examples 6.1 and 6.4 above the length $l = 1$, then the condition $\zeta > 0$ there is automatically fulfilled provided R is not equal to the Dirac-measure at point m. To see this recall that $\zeta = \int \varphi(m-x)^2 R(dx)$ in case $l = 1$. Assume that $\zeta = 0$. Then $1 = R(\{x \in \mathbb{R} : \varphi(m-x) = 0\}) = R(\{m\})$, because φ is strictly increasing and $\varphi(0) = 0$. It follows that $R = \delta_m$ in contradiction to our assumption on R.

Remark 6.2. (1) Assume that F has left- and right derivatives at m. Then in particularly, F is continuous at m and thus condition (87) is fulfilled, because

$$\frac{1}{t}(\varphi(t) - \varphi(-t))R([m, m+t]) = (\varphi(t) - \varphi(-t))\frac{F(m+t) - F(m)}{t} \to 0, t \downarrow 0.$$

Similarly, one sees that condition(88) is satisfied as well.

(2) If φ in Example 6.4 is differentiable at zero with $\varphi'(0) < \infty$ and if F is continuous at m, then (87) and (88) are again fulfilled. If actually $\varphi'(0) = 0$, then F need not be continuous at m.

Example 6.4 enables us to handle the L_p-estimator.

Example 6.5. (L_p**-estimator**) For $p > 1$ let $\phi(t) = |t|^p, t \in \mathbb{R}$. ($p = 1$ corresponds to the median and will be examined later on.) Then ϕ is differentiable on $\mathbb{R}$ with continuous derivative $\varphi(t) = p\ \mathrm{sign}(t)\ |t|^{p-1}$. Moreover, if $1 < p < 2$, then φ is concave on $[0, \infty)$, convex on $(-\infty, 0)$ and $\varphi(0)$. If $p \geq 2$, then concave and convex reverse their roles. It follows with the c_r-inequality in Lemma 7.4 below that

$$\int |\varphi(t-x)|R(dx) = p \int |t-x|^{p-1} R(dx) < \infty\ \forall\, t \in \mathbb{R},$$

if and only if $\int |x|^{p-1} R(dx) < \infty$. Conclude from Lemma 6.1 that U has a unique minimizing point $m \in \mathbb{R}$ with $V(m) = 0 = V(m-)$, which satisfies the equation

$$\int_{(-\infty,m]} (m-x)^{p-1} R(dx) = \int_{(m,\infty)} (x-m)^{p-1} R(dx).$$

Another application of the c_r-inequality yields that

$$\int \varphi(t-x)^2 R(dx) = p^2 \int |t-x|^{2(p-1)} R(dx) \text{ is finite for all } t \in \mathbb{R} \tag{89}$$

provided $\int |x|^{2(p-1)} R(dx) < \infty$. From (89) it follows that (83) is fulfilled, but also (85) upon noticing that $(\varphi(m+t-x) - \varphi(m-x))^2 \leq 2(\varphi(m+t-x)^2 + \varphi(m-x)^2)$.

Finally, assume that F is left- and right-differentiable at m. Then by Remark 6.2(1) the conditions (87) and (88) are fulfilled. In case $p > 2$ these conditions are satisfied for

any distribution function F according to Remark 6.2(2). (The case $p = 2$ is treated in Example 6.1.)

In summary we obtain: If $\int |x|^{p-1}R(dx)$ is finite, then the minimizer m of the function $t \mapsto \mathbb{E}[|t - k(\mathbf{X})|^p]$ exists and is uniquely determined. Let $\tilde{m}_n$ be the L_p-estimator, i.e. $\tilde{m}_n$ solves the equation

$$\sum_{\mathbf{i}} 1_{\{k(\mathbf{X_i})<t\}}(t - k(\mathbf{X_i}))^{p-1} = \sum_{\mathbf{i}} 1_{\{k(\mathbf{X_i})>t\}}(k(\mathbf{X_i}) - t)^{p-1},\ t \in \mathbb{R}.$$

If actually $\int |x|^{2(p-1)}R(dx) < \infty$, then

$$\sqrt{n}(\tilde{m}_n - m) \xrightarrow{\mathcal{D}} N(0, \tau^2) \text{ with } \tau^2 = \frac{l^2\zeta}{p^2(p-1)^2\left(\int_{\mathbb{R}\setminus\{m\}} |m - x|^{p-2}R(dx)\right)^2}.$$

Notice that Example 6.4 in particularly guarantees the existence of the integral in the denominator. In case $l = 1$ the variance τ^2 is equal to

$$\frac{\int |m - x|^{2(p-1)}R(dx)}{(p-1)^2\left(\int_{\mathbb{R}\setminus\{m\}} |m - x|^{p-2}R(dx)\right)^2}.$$

If in addition $S = \mathbb{R}$ and $k(x) = x$, then we obtain the result of Hjort and Pollard [11].

In our next example, φ is not continuous, but has a jump at point zero. Recall that F is the distribution function of $k(\mathbf{X})$.

Example 6.6. (**U-quantiles**) For every fixed $\alpha \in (0, 1)$ consider $\phi(t) = t(1_{\{t\geq 0\}} - \alpha)$. It follows that $\varphi(t) = 1_{\{t\geq 0\}} - \alpha$ and therefore

$$V(t) = F(t) - \alpha.$$

Thus (78) yields that all minimizers m of U are exactly those satisfing the inequalities $F(m-) \leq \alpha \leq F(m)$, i.e. Argmin$(U)$ is equal to the set of all α-quantiles of F. Let

$$F_n(x) := \binom{n}{l}^{-1} \sum_{1\leq i_1<\ldots<i_l\leq n} 1_{\{k(X_{i_1},\ldots,X_{i_l})\leq x\}},\ x \in \mathbb{R}.$$

Notice that F_n is the empirical distribution function pertaining to the sample

$$\{k(X_{i_1}, \ldots, X_{i_l}) : 1 \leq i_1 < \ldots < i_l \leq n\}$$

of size $N = \binom{n}{l}$. By (81) the induced estimator $\tilde{m}_n$ is an α-quantile of F_n, i.e. $F_n(\tilde{m}_n-) \leq \alpha \leq F_n(\tilde{m}_n)$. In general it is not unique. A conventional choice is $F_n^{-1}(\alpha)$, the smallest α-quantile. The condition (10) becomes

$$\delta_n(x) = \sqrt{n}(F(m + a_n x) - \alpha) \to \delta(x) \quad \forall\, x \in D. \tag{90}$$

It follows easily that (A0)-(A4) are fulfilled. Here, (A1) holds, if F is continuous at m and (A3) holds in case $l = 1$, because then $\zeta = \alpha(1 - \alpha)$. If $l \geq 2$, then $\zeta = \int_S \mathbb{P}(k(x, X_2, \ldots, X_l) \leq m)^2 P(dx) - \alpha^2$, whence $\zeta = 0$ leads to $\mathbb{P}(k(x, X_2, \ldots, X_l) \leq m) = \alpha$ for P-almost every $x \in S$, which usually results in a contradiction. Assume that (90) is valid with $\delta(x) \to \pm\infty$ as $x \to \pm\infty$. Then by Theorem 1.1 we know that (9) holds with H being a distribution function, that belongs to the four classes **class1**-**class4**. Infer from Theorems 3.1-3.3 that m is uniquely determined in case that H lies in one of the first 3 classes. If $\delta(x)$ is bounded on $[0, \infty)$ or on $(-\infty, 0]$, then H is a sub distribution function and we obtain distributional convergence to a $\overline{\mathbb{R}}$-valued random variable with distribution function (for the extended real line) $H(x) = \Phi_\sigma(\delta(x))$.

In Example 4.3 of Ferger [5] one finds distribution functions F such that $a_n = n^{-\frac{1}{2\beta}}$ with some $\beta > 0$ and H has the shape (11), (12) or (13). Another F there yields $a_n = 1/\log(n)$ and a limit H as in (14). As a further example consider

$$F(x) := \frac{1}{2} + \mathrm{sign}(x)|x|(\log|x|)^2 \text{ for all } x \in [-\epsilon, \epsilon] \setminus \{0\} \tag{91}$$

with $\epsilon > 0$ sufficiently small. Notice that F is continuos, if $F(0) := \frac{1}{2}$. Infer that $m = 0$ is the unique median of F. It is not differentiable at 0 and $F'(x) \to \infty$ as $t \to 0$. One easily checks with Theorem 3.3 that F lies in **class3**. In fact, if $a_n := n^{-\frac{1}{2}}(\log\sqrt{n})^{-2}$, then for all $x > 0$ and for eventually all $n \in \mathbb{N}$ we obtain:

$$\delta_n(x) = \sqrt{n}(F(a_n x) - \frac{1}{2}) = x\Big(\frac{-\log\sqrt{n} - 2\log\log\sqrt{n} + \log(x)}{\log\sqrt{n}}\Big)^2 \to x$$

Since $V(t) = -V(-t)$ for every $t \in [-\epsilon, \epsilon]$, it follows that $\delta_n(x) \to x$ for all $x \in \mathbb{R}$. Consequently,

$$\sqrt{n}(\log\sqrt{n})^2\tilde{m}_n \xrightarrow{\mathcal{D}} N(0, l^2\zeta).$$

If F is differentiable at m, then so is $V = F - \alpha$ and we obtain the square root asymptotics of Proposition 5.1 and Remark 5.1 with $V'(m) = F'(m)$. In particularly,

$$\sqrt{n}(\tilde{m}_n - m) \xrightarrow{\mathcal{D}} N(0, l^2\zeta/F'(m)^2), \tag{92}$$

where $l = 1$ yields the well-known limit $N(0, \alpha(1-\alpha)/F'(m)^2)$.

The above Example 6.6 in the special case $S = \mathbb{R}, l = 1$ and $k(x) = x$ contains the whole theory on quantile-estimators of Smirnov's [20] fundamental paper. It includes his equivalent conditions that the underlying distribution function F of the data lies in the repective domain of attraction (symbolically: $F \in \mathcal{D}(H)$). In fact, in Theorem 3.1- 3.4 the function V is given by $V(t) = F(t) - \alpha$. In contrast to Smirnov [20] we also specify the normalizing sequence (a_n). For instance, it follows from Theorems 3.1 and 3.3 that a possible choice is $a_n = F^{-1}(\alpha + \frac{1}{\sqrt{n}}) - F^{-1}(\alpha)$ for $F \in \mathcal{D}(H)$ with $H \in$ **class1** or **class3**.

Another special case in Example 6.6 is $\alpha = \frac{1}{2}$ resulting in the median of F. Here, the corresponding ϕ can be rewritten as $\phi(t) = \frac{1}{2}|t|$. The factor $\frac{1}{2}$ in front of $|t|$ is obviously unneceessary and therefore will be ommitted in the sequel. So now we look at $h(\mathbf{x}, t) = |t - k(\mathbf{x})|$ and additionally consider $S = \mathbb{R}$.

Example 6.7. (**Hodges-Lehmann estimator**) Here, $k(x_1, \ldots, x_l) := l^{-1}(x_1 + \ldots + x_l)$. For $l = 2$ the estimator $\tilde{m}_n$ is a median of the so-called *Walsh-averages* $\{(X_i + X_j)/2 : 1 \le i < j \le n\}$. This is a variant of the Hodges-Lehmann estimator, where the Walsh-averages with $i = j$ are added to the sample.

If in the above example the function k more generally is given by $k(x_1, x_2) = \beta x_1 + (1-\beta)x_2$ with $\beta > 0$ a fixed constant, then the resulting median is a variant of the estimator due to **Maritz, Wu and Staudte** [15].

Example 6.8. (**Bickel-Lehmann estimator**) This estimator is defined as a median of the sample $\{|X_i - X_j| : 1 \le i < j \le n\}$ and corresponds to $k(x_1, x_2) = |x_1 - x_2|$.

In our next example $S = \mathbb{R}^2$.

Example 6.9. (**Theil-Sen estimator**) Let $X_i = (Y_i, Z_i), 1 \le i \le n$, and $\{k(X_i, X_j) := \frac{Z_i - Z_j}{Y_i - Y_j} : 1 \le i < j \le n\}$, which is a sample of well-defined real random variables, if for instance Y_1 is continuously distributed. Then the median of the sample is known to be a robust estimator for the slope β in the simple linear regression model $Z_i = \alpha + \beta Y_i$.

The function $\varphi(t) = 1_{\{t\geq 0\}} - \alpha$, which induces the α-quantile has a jump at point zero (and is continuous elsewhere). Let us more generally consider monotone increasing functions φ, continuous on $\mathbb{R}\setminus\{0\}$ and with a jump at zero. Assume that $\kappa+ := \varphi(0) \geq 0$ and $\kappa_- := \varphi(0-) \leq 0$ and that $\kappa_* := \min\{\kappa_+, -\kappa_-\} > 0$. Then $\kappa := \kappa_+ - \kappa_- > 0$ is the jump height. Notice that φ still satisfies the assumption of Lemma 6.1, whence there exists a unique $m \in \mathbb{R}$ with $V(m) = 0 = V(m-)$ so that (A1) is fulfilled. Moreover, $\int \varphi(m-x)^2 R(dx) \geq \kappa_*^2 > 0$, so that (A3) is fulfilled at least in the important special case $l = 1$. Introduce

$$\varphi_c := \varphi - \kappa_+ 1_{[0,\infty)} - \kappa_- 1_{(-\infty,0)}. \tag{93}$$

Observe that $\varphi_c(0+) = \varphi(0) - \kappa_+ = 0$ and $\varphi_c(0-) = \varphi(0-) - \kappa_- = 0$. Thus φ_c is continuous at zero and $\varphi_c(0) = 0$. Geometrically, the graph of φ_c is created from φ by moving the two branches to the left and right of the ordinate to the origin and connecting them there. Notice that φ meets the requirements (79) and (83) if and only if this is true for φ_c. Indeed, it is convenient to formulate our conditions in terms of φ_c rather than for φ. Moreover from (93) we obtain a decomposition of V as follows:

$$V(t) = \kappa F(t) + V_c(t) + \kappa_-, \tag{94}$$

where $V_c(t) = \int \varphi_c(t-x)R(dx)$. Thus δ_n has the following shape:

$$\delta_n(x) = \kappa\sqrt{n}(F(m+a_n x) - F(m)) + \sqrt{n}(V_c(m+a_n x) - V_c(m)). \tag{95}$$

Assume that

$$\bar{\delta}_n(x) := \sqrt{n}(F(m+a_n x) - F(m)) \to \bar{\delta}(x) \quad \forall\, x \in D, \tag{96}$$

where D is a dense subset of $\mathbb{R}$ and $\bar{\delta}$ belongs to one of the four classes in (33)-(36). In other words, $F \in \mathcal{D}(H)$. If φ_c satifies the smoothness-conditions of the φ's in Examples 6.1 or 6.4, then our proofs show that V_c is differentiable at m with $V_c'(m) = \int \varphi_c'(m-x)R(dx) \geq 0$. Suppose that

$$\sqrt{n}a_n \to \rho.$$

Then by (95)

$$\begin{aligned} \delta_n(x) &= \kappa\sqrt{n}(F(m+a_n x) - F(m)) + \sqrt{n}a_n \frac{V_c(m+a_n x) - V_c(m)}{a_n x} x \\ &\to \kappa\bar{\delta}(x) + \rho V_c'(m)x = \delta(x) \quad \text{for all } x \in D. \end{aligned} \tag{97}$$

Notice that $\rho \geq 0$.

Example 6.10. If $\rho = 0$, then all statements of Theorem 1.1 hold with $\delta = \kappa\bar{\delta}$.

If for instance F is as in (91), then $\sqrt{n}a_n = (\log\sqrt{n})^{-2} \to 0$. Or if $a_n = n^{-\frac{1}{2\beta}}$, then $\rho = 0$, whenever $\beta < 1$.

Example 6.11. Consider $0 < \rho < \infty$. Observe that for every $x > 0$ we have:

$$\frac{F(m+a_n x) - F(m)}{a_n x} = \frac{\bar{\delta}_n(x)}{\kappa\sqrt{n}a_n x} \quad \forall\, n \in \mathbb{N}.$$

Assume that $\bar{\delta}$ has the shape (33). Taking the limit $n \to \infty$ shows that F has a positive right derivative $D^+F(m) = \bar{\delta}(x)/(\kappa\rho x) > 0$, because $\bar{\delta}(x) > 0$. Rearranging the last equality leads to $\bar{\delta}(x) = \kappa\rho D^+F(m)x$ for all $x > 0$. Consequently, δ in (97) is given by $\delta(x) = \rho(\kappa D^+F(m) + V_c'(m))x$, if $x > 0$ and equal to $-\infty$, if $x < 0$. Thus with Theorem 1.1 and Slutsky's Theorem we arrive at

$$\sqrt{n}(\tilde{m}_n - m) \xrightarrow{\mathcal{D}} Y, \tag{98}$$

where

$$\mathbb{P}(Y \le x) = \begin{cases} 0 & , \quad x < 0 \\ \Phi_\sigma\Big((\kappa D^+F(m) + V_c'(m))x)\Big) & , \quad x > 0. \end{cases}$$

Analogously, we can treat the two cases $\bar{\delta}$ of shape (34) or (35) and obtain the $\sqrt{n}$-distributional convergence (98). Here in case (34),

$$\mathbb{P}(Y \le x) = \begin{cases} \Phi_\sigma\Big((\kappa D^-F(m) + V_c'(m))x)\Big) & , \quad x < 0 \\ 1 & , \quad x > 0. \end{cases}$$

and otherwise

$$\mathbb{P}(Y \le x) = \begin{cases} \Phi_\sigma\Big((\kappa D^-F(m) + V_c'(m))x)\Big) & , \quad x < 0 \\ \Phi_\sigma\Big((\kappa D^+F(m) + V_c'(m))x)\Big) & , \quad x > 0. \end{cases}$$

Thus, if F is differentiable at m, then

$$\sqrt{n}(\tilde{m}_n - m) \xrightarrow{\mathcal{D}} N(0, \tau^2) \quad \text{with } \tau^2 = \frac{l^2\zeta}{(\kappa D^-F(m) + V_c'(m))^2}.$$

Formally, for $\kappa = 0$ we obtain the asymptotic normality in Examples 6.1 and 6.4, because then $\varphi_c = \varphi$ and hence $V_c'(m)$ is equal to $\int \varphi'(m-x)R(dx)$ or $\int_{\mathbb{R}\setminus\{m\}} \varphi'(m-x)R(dx)$ according as φ is differentiable at 0 or not.

Finally, assume that $\bar{\delta}$ is of type (36). Then by (97)

$$\delta_n(x) \to \delta(x) = \begin{cases} -\infty & , \quad x < -c_1 \\ \rho V_c'(m)x & , \quad -c_1 < x < c_2 \\ \infty & , \quad x > c_2. \end{cases} \qquad (c_1, c_2 \ge 0, \max\{c_1, c_2\} > 0)$$

and by Theorem 1.1 the limit $H(x) = \Phi_\sigma(\delta(x))$ is a distribution function not degenerated at 0. By the second part of Theorem 1.1 H can only belong to the four classes **class1**-**class4**. This results in a contradiction when $V_c'(m)$ is positive. However, if (and only if) $V_c'(m) = 0$, then we obtain that

$$\frac{\tilde{m}_n - m}{a_n} \xrightarrow{\mathcal{D}} Y,$$

where Y is uniformly distributed on $\{-c_1, c_2\}$.

Example 6.12. Assume that $\rho = \infty$. If $V_c'(m) > 0$, then it follows from (97) that $\delta(x) = \infty$ and $\delta(x) = -\infty$ according as $x > 0$ or $x < 0$. Thus Theorem 1.1 yields that $(\tilde{m}_n - m)/a_n \xrightarrow{\mathbb{P}} 0$. However, if $V_c'(m) = 0$, then all statements of Theorem 1.1 hold with $\delta = \kappa\bar{\delta}$.

Example 6.13. Finally, we would like to consider convex functions ϕ, which are piecewise linear. The corresponding "generating" function φ is a step-function. For the sake of simplicity let us consider a three-step function $\varphi = \alpha 1_{(-\infty,0)} + \beta 1_{[0,r)} + \gamma 1_{[r,\infty)}$ with $r > 0$ and $\alpha < 0 < \beta < \gamma$. We obtain $V(t) = \alpha + (\beta - \alpha)F(t) + (\gamma - \beta)F(t - r)$. By Lemma 6.1 there exists a unique $m \in \mathbb{R}$ such that $V(m) = 0$. Therefore,

$$\delta_n(x) = \sqrt{n}V(m + a_n x) = (\beta - \alpha)\delta_{n,0}(x) + (\gamma - \beta)\delta_{n,r}(x),$$

where $\delta_{n,0}(x) = \sqrt{n}(F(m + a_n x) - F(m))$ and $\delta_{n,r}(x) = \sqrt{n}(F(m - r + a_n x) - F(m - r))$. If $\delta_{n,0}(x) \to \delta_0(x)$ and $\delta_{n,r}(x) \to \delta_r(x)$ for all $x \in D$, then

$$\delta_n(x) \to \delta(x) = (\beta - \alpha)\delta_0(x) + (\gamma - \beta)\delta_r(x) \text{ for all } x \in D.$$

Suppose for instance that $\delta_0(x)$ or $\delta_r(x)$ converge to $\pm\infty$ as $x \to \pm\infty$, then $\delta(x) \to \pm\infty$ as $x \to \pm\infty$, so that $H(x) = \Phi_\sigma(\delta(x))$ is a distribution function. If H is not degenerated

at 0, we obtain from Theorem 1.1 distributional convergence with H in one of the four classes.

7. Appendix

In this section, we derive results that may seem purely technical, but are extremely useful for our purposes.

Lemma 7.1. *If (A0) and (A1) hold, then*

$$\mathbb{E}[D^+h(X_1,\dots,X_l,m)^2] = \mathbb{E}[D^-h(X_1,\dots,X_l,m)^2]. \tag{99}$$

If (A4) holds, then

$$\mathbb{E}[\big(D^+h(X_1,\dots,X_l,m+t) - D^+h(X_1,\dots,X_l,m)\big)^2] < \infty \quad \forall\, t \in [t_0,t_1]. \tag{100}$$

Proof. Let $\mathbf{X} := (X_1,\dots,X_l)$. First notice that $D^+h(\mathbf{X},m)$ and $D^-h(\mathbf{X},m)$ both are integrable by (A0). Deduce from (A1) and (5) that $\mathbb{E}[D^+h(\mathbf{X},m) - D^-h(\mathbf{X},m)] = 0$, whence

$$D^+h(\mathbf{X},m) = D^-h(\mathbf{X},m)\ \mathbb{P}-\text{a.s}, \tag{101}$$

because $D^+h(\mathbf{x},\cdot) \geq D^-h(\mathbf{x},\cdot)$ for all $\mathbf{x} \in S^l$ by Theorem 24.1 in Rockafellar [18]. Consequently, $\mathbb{E}[D^+h(\mathbf{X},m)^2] = \mathbb{E}[D^-h(\mathbf{X},m)^2]$.

Let $t \in [0,t_1]$. By Theorem 24.1 in Rockafellar [18] $D^+h(\mathbf{x},\cdot)$ is increasing for each $\mathbf{x} \in S^l$. It follows that

$$0 \leq D^+h(\mathbf{X},m+t) - D^+h(\mathbf{X},m) \leq D^+h(\mathbf{X},m+t_1) - D^+h(\mathbf{X},m)$$

and thus $0 \leq \big(D^+h(\mathbf{X},m+t) - D^+h(\mathbf{X},m)\big)^2 \leq \big(D^+h(\mathbf{X},m+t_1) - D^+h(\mathbf{X},m)\big)^2$. Conclude from (A4) that

$$\mathbb{E}[\big(D^+h(\mathbf{X},m+t) - D^+h(\mathbf{X},m)\big)^2] < \infty \quad \forall t \in [0,t_1]$$

In the same fashion one treats the case $t \in [t_0,0]$. □

Lemma 7.2. *If (A0)-(A4) hold, then so do (B0)-(B4). Moreover, $\zeta = \zeta^-$.*

Proof. Assume that (A0)-(A4) are true. Obviously, (A0) and (A1) imply (B0) and (B1), respectively. Infer from (99) of Lemma 7.1 that (B2) is satisfied.

To see the validity of (B3) we first simplify our notation: $f^\pm(\mathbf{x}) := D^\pm h(\mathbf{x},m)$. A reformulation of (101) yields: $P^l(N) = 0$, where $N = \{\mathbf{x} \in S^l : f^+(\mathbf{x}) \neq f^-(\mathbf{x})\}$. An application of Fubini's theorem yields that $0 = \int_S P^{l-1}(N_{x_1})P(dx_1)$ with $N_{x_1} = \{(x_2,\dots,x_l) \in S^{l-1} : (x_1,x_2,\dots,x_l) \in N\}$. As the integrand is non- negative, so that $P^{l-1}(N_{x_1}) = 0$ P-almost surely, i.e. $P(E) = 1$, where $E = \{x_1 \in S : P^{l-1}(N_{x_1}) = 0\} = \{x_1 \in S : P^{l-1}(N^c_{x_1}) = 1\}$. Here,

$$\begin{aligned} N^c_{x_1} &= \{(x_2,\dots,x_l) \in S^{l-1} : (x_1,x_2,\dots,x_l) \notin N\} \\ &= \{(x_2,\dots,x_l) \in S^{l-1} : f^+(x_1,x_2,\dots,x_l) = f^-(x_1,x_2,\dots,x_l)\}. \end{aligned}$$

With these preparations we can conclude as follows:

$$\begin{aligned}
\zeta^- &= \int_S [\int_{S^{l-1}} f^-(x_1,\dots,x_l)dP^{l-1}(x_2,\dots,x_l)]^2 dP(x_1)\\
&= \int_E [\int_{N^c_{x_1}} f^-(x_1,\dots,x_l)dP^{l-1}(x_2,\dots,x_l)]^2 dP(x_1)\\
&= \int_E [\int_{N^c_{x_1}} f^+(x_1,\dots,x_l)dP^{l-1}(x_2,\dots,x_l)]^2 dP(x_1)\\
&= \int_S [\int_{S^{l-1}} f^+(x_1,\dots,x_l)dP^{l-1}(x_2,\dots,x_l)]^2 dP(x_1)\\
&= \zeta.
\end{aligned}$$

Especially (B3) holds true.

As to (B4) recall that $\mathbb{E}[D^{\pm}h(\mathbf{X},t)] = D^{\pm}U(t)$ for every $t \in \mathbb{R}$ by (5) and that the function U is convex. Thus by Theorem 24.1 of Rockafellar [18] D^+U and D^-U both are increasing. Moreover, $D^-U \le D^+U$ and the set $J := \{t \in \mathbb{R} : D^-U(t) < D^+U(t)\}$ is denumerable, confer Niculescu and Persson [16], p. 20. Consequently, the complement $\Delta := J^c = \{D^+U = D^-U\}$ is dense in $\mathbb{R}$ and $\mathbb{E}[D^+h(\mathbf{X},t) - D^-h(\mathbf{X},t)] = 0$ for all $t \in \Delta$, where the integrand is non-negative. Therefore

$$D^+h(\mathbf{X},t) = D^-h(\mathbf{X},t)\ \mathbb{P}\text{-a.s. for all } t \in \Delta. \tag{102}$$

By denseness we find points $u_0 \in [m+t_0, m)$ and $u_1 \in (m, m+t_1]$ such that $D^+h(\mathbf{X},u_0) = D^-h(\mathbf{X},u_0)$ and $D^+h(\mathbf{X},u_1) = D^-h(\mathbf{X},u_1)$ a.s. Put $s_0 := u_0 - m$ and $s_1 = u_1 - m$. Taking into account (101) we now have that

$$D^+h(\mathbf{X},m+s_0) = D^-h(\mathbf{X},m+s_0),\ D^+h(\mathbf{X},m+s_1) = D^-h(\mathbf{X},m+s_1)$$

and $D^+h(\mathbf{X},m) = D^-h(\mathbf{X},m)$ with probability one. Together with (100) this ensures (B4), because $t_0 \le s_0 < 0 < s_1 \le t_1$. □

Recall that $\delta_n(x) = \sqrt{n}D^+U(m+a_nx)$ and $\delta_n^*(x) = \sqrt{n}D^-U(m+a_nx)$.

Lemma 7.3.

(1) *There exists a subset $D_1 = A^c$, which is the complement of a countable set A such that $\delta_n^*(x) = \delta_n(x)$ for all $x \in D_1$ and for all $n \in \mathbb{N}$.*

(2) *Assume that*

$$\delta_n(x) \to \delta(x) \in [-\infty,\infty] \quad \forall\, x \in D,$$

where D lies dense in $\mathbb{R}$. Then there exists a dense subset D_2 with $D_2 \subseteq D_1$ such that

$$\delta_n^*(x) \to \delta(x) \in [-\infty,\infty] \quad \forall\, x \in D_2.$$

Proof. (1) Recall form the above proof that $J = \{t \in \mathbb{R} : D^-U(t) \ne D^+U(t)\}$ is denumerable. Notice that $\delta_n(x) \ne \delta_n^*(x)$ if and only if $x \in \{(y-m)/a_n : y \in J\} =: A_n$. Here, A_n is denumerable and so is $A := \cup_{n\in\mathbb{N}} A_n$. Thus $D_1 := A^c = \cap_{n\in\mathbb{N}} A_n^c$ is a dense subset of $\mathbb{R}$. Moreover, if $x \in D_1$, then $\delta_n(x) = \delta_n^*(x)$ for every $n \in \mathbb{N}$.

(2) Conclude from the assumption that $F_n(x) := \Phi_1(\delta_n(x)) \to \Phi_1(\delta(x)) =: F(x)$ for all $x \in D$. All involved functions are increasing. Therefore by Lemma 5.74 in Witting and Müller-Funk [22] one has that $F_n(x) \to F(x)$ for all $x \in C_F$. Since Φ_1 is invertible, we now know that $\delta_n(x) \to \delta(x)$ for all $x \in C_F$. By monotonicity of F the complement D_F of C_F is countable and so is $A \cup D_F$, whence $D_2 := (A \cup D_F)^c = D_1 \cap C_F$ is dense. Conclude with part (1) that $\delta_n^*(x) = \delta_n(x) \to \delta(x)$ for each $x \in D_2$. □

The following inequality is a well-known result from analysis, which we state for the sake of convenience.

Lemma 7.4. *(c_r-inequality)* $|u+v|^r \le c_r(|u|^r+|v|^r)$ *for all* $u, v \in \mathbb{R}$ *and for all* $r > 0$, *where* $c_r = 1$ *or* $c_r = 2^{r-1}$ *according as* $r \le 1$ *or* $r > 1$.

References

1. P. Billingsley, *Convergence of Probability Measures*, Second Edition, New York: John Wiley & Sons, 1999. https://doi.org/10.1002/9780470316962
2. A. Bose, *Bahadur reperesentation of M_m estimates*, Ann.Statist. **26** (1998), 771–777. https://doi.org/10.1214/aos/1028144859
3. H. K. Brøns, H. D. Brunk, W. E. Franck and D. L. Hanson, *Generalized Means and Associated Families of Distributions*, Ann.Math.Statist. **40** (1969), 339–355. https://doi.org/10.1214/aoms/1177697698
4. K. L. Chung, *A Course in Probability Theory*, Second Edition, San Diego: Academic Press, Inc., 1974.
5. D. Ferger, *Distributional Hyperspace Convergence of Argmin-Sets in Convex M-Estimation*, Theor. Probability and Math. Statist. **109** (2023), 3–35. https://doi.org/10.1090/tpms/1195
6. D. Ferger, *On semi-continuity and continuity of the smallest and largest minimizing point of real convex functions with applications in probability and statistics*, Eur. J. Math. **10**:14 (2024), 1–18. https://doi.org/10.1007/s40879-024-00728-2
7. B. Fristedt and L. Gray, *A Modern Approach to Probability Theory*, Boston: Birkhäuser, 1997. https://doi.org/10.1007/978-1-4899-2837-5
8. B. V. Gnedenko, *Some theorems on powers of distribution functions*, Učen, Zap. Moskov. Gos. Univ. **45** (1940), 61–72 (in Russian).
9. P. Gänssler and W. Stute, *Wahrscheinlichkeitstheorie*, Berlin, Heidelberg: Springer-Verlag, 1977.
10. S. J. Habermann, *Concavity and estimation*, Ann.Statist. **17** (1989), 1631–1661. https://doi.org/10.1214/aos/1176347385
11. N. L. Hjort and D. Pollard, *Asymptotics for minimisers of convex processes*, Preprint Yale University (1993), 1–24.
12. P. J. Huber, *Robust estimation of a location parameter*, Ann. Math. Statist. **35**, 73–101. https://doi.org/10.1214/aoms/1177703732
13. P. J. Huber, *Robust Statistics*, New York, Chichester, Brisbane, Toronto, Singapore: John Wiley & Sons, 1981.
14. V. S. Koroljuk and Yu. V. Borovskich, *Theory of U-Statistics*, Dordrecht: Springer-Science+Media, 1994. https://doi.org/10.1007/978-94-017-3515-5
15. J. S. Maritz, M. Wu and R. G. Staudte, Jr., *A Location Estimator Based on U-Statistics*, Ann. Statist. **5**, 779–786. https://doi.org/10.1214/aos/1176343900
16. C. P. Niculescu and L.-E. Persson, *Convex Functions and Their Applications: A Contemporary Approach*, New York: Springer Science+Business Media, Inc., 2006. https://doi.org/10.1007/0-387-31077-0
17. W. Niemiro, *Asymptotics For M-Estimators Defined By Convex Minimization*, Ann.Statist. **20** (1992), 1514–1533. https://doi.org/10.1214/aos/1176348782
18. R. T. Rockafellar, *Convex Analysis*, Princeton, NJ: Princeton Mathematical Series **28**, Princeton Univ. Press, 1970.
19. R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, New York: John Wiley & Sons, 1980. https://doi.org/10.1002/9780470316481
20. N. V. Smirnov, *Limit distributions for the terms of a variational series*, Mathematical Society Translations, **67**, 1952.
21. H. Witting, *Mathematische Statistik I*, Stuttgart: B.G. Teubner, 1985. https://doi.org/10.1007/978-3-322-90150-7
22. H. Witting and U. Müller-Funk, *Mathematische Statistik II*, Stuttgart: B.G. Teubner, 1995. https://doi.org/10.1007/978-3-322-90152-1

Fakultät Mathematik, Technische Universität Dresden, Zellescher Weg 12-14, D-01069 Dresden, Germany

E-mail address: dietmar.ferger@tu-dresden.de