A BERRY–ESSEEN BOUND FOR M-ESTIMATORS

V. Bentkus M.Bloznelis F.Götze

Abstract. We prove a Berry–Esseen bound for general M-estimators under optimal regularity conditions on the score function and the underlying distribution. As an application we obtain Berry–Esseen bounds for the sample median, the L_p-median, $p > 1$ and Huber's estimator of location.

1. Introduction and Main Results

Let X denote a random variable taking values in a measurable space $(\mathcal{X}, \mathcal{F})$ with distribution F. For an open subset T of R, let $f : \mathcal{X} \times T \to R$ be a jointly measurable function. Let $X_1, \dots, X_N$ be independent observations drawn from the distribution F. A random variable $t_N = t_N(X_1, \dots, X_N)$ which minimizes the function

$$Q_N(t) = N^{-1} \sum_{i=1}^{N} f(X_i, t)$$

is called M-estimator. If $\mathbf{E}\,\big|f(X,t)\big| < \infty$, for $t \in T$, and the function $t \to \mathbf{E}\, f(X,t)$ has a unique minimum point, say t_0, then t_N may be used as an estimator of t_0.

The usual requirement for t_N is that it estimates $t_0 \in T$ consistently. Furthermore, under certain conditions $N^{1/2}(t_N - t_0)$ is asymptotically normal, see Huber (1964), Serfling (1980). For recent advances concerning the asymptotic normality we refer to Pollard (1984, 1985) and Hoffmann-Jorgensen (1994).

Our aim is to establish a general Berry–Esseen bound for the statistics $N^{1/2}(t_N - t_0)$ under minimal regularity conditions on the the score function $f(x,t)$. Hence we shall estimate

$$\Delta_N := \sup_{u \in R} \Big| \mathbf{P}\big\{ N^{1/2}(t_N - t_0)/b < u \big\} - \Phi(u) \Big|,$$

1991 *Mathematics Subject Classification.* Primary 60F05; secondary 62F12.

Key words and phrases. Berry–Esseen bound, Maximum likelihood estimator, M-estimator, sample median, sample quantile, Huber's estimator, convergence rate, Central Limit Theorem.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

where Φ denotes the standard normal distribution function and b is a scale parameter defined in (1.3) below.

Any M-estimator $t_N = t_N(X_1, \dots, X_N)$ is a symmetric function of its arguments $X_1, \dots, X_N$, i.e., it is a symmetric statistic. Berry–Esseen bounds for general symmetric statistics were obtained by van Zwet (1984), Friedrich (1988) and Bolthausen Götze and (1993). These results are applicable (see van Zwet (1984) and Bentkus Götze and van Zwet (1994)) to a wide class of statistics. They apply to M-estimators if the score function f is sufficiently smooth since in this case an explicit approximate solution of the minimization problem is available. However, in a number of important examples the score function is not sufficiently smooth and direct proofs are needed.

Berry–Esseen bounds for M- and related (maximum likelihood, minimum contrast) estimators were considered by Linnik and Mitrofanova (1963), Michel and Pfanzagl (1971), Pfanzagl (1971), Chibisov (1972), Bickel (1974), Matsuda (1983), Radavičius (1990), Paulauskas (1995), etc. Pfanzagl (1971) obtained the bound $\Delta_N = O(N^{-1/2})$ for a minimum contrast estimator in the case where the score function is twice differentiable in t and the second derivative $f''(x,t) = \dfrac{\partial^2}{\partial t^2} f(x,t)$ satisfies Lipschitz condition. In particular cases the smoothness condition on the score function were relaxed in Matsuda (1983) and Radavičius (1990). Matsuda (1983) considered the maximum likelihood (ML) estimator of the shift parameter of a density function and obtained an estimate of Δ_N assuming that the score function is two times differentiable. Radavičius (1990) obtained the bound $\Delta_N = O(N^{-1/2})$ for a ML estimator in the case where the score function $f(x,t)$ is differentiable and convex in t and the derivative $f'(x,t) = \dfrac{\partial}{\partial t} f(x,t)$ satisfies a Lipschitz condition.

We prove a Berry–Esseen bound for general M-estimators. To formulate the result we need notations related to *smoothness,* (*asymptotic*) *convexity* and *consistency.*

Smoothness. Let there exist a neighborhood, say

$$V = (t_0 - \delta, t_0 + \delta) \subset T,$$

of t_0 such that, for any $t \in V$,

$$f'(x,t) = \lim_{\tau \to 0} \big(f(x, t+\tau) - f(x,t) \big) / \tau \quad \text{exists}, \tag{1.1}$$

for F-almost all $x \in \mathcal{X}$. In other words, for any $t \in V$, there is a set $U_t \subset \mathcal{X}$ of F–measure zero such that, for all $x \notin U_t$, $f'(x,t)$ exists.

Introduce the mean values

$$L_1(t) := \mathbf{E}\, f'(X,t) \qquad \text{and} \qquad L_2(t) := \mathbf{E}\, \big(f'(X,t) \big)^2$$

and assume that the function L_1 is differentiable, L_2 is continuous at the point t_0,

$$L_1(t_0) = 0, \quad L_1'(t_0) > 0 \qquad \text{and} \qquad L_2(t_0) > 0.$$

Furthermore, we require

$$\bigl|L_1(t) - L_1'(t_0)\cdot(t-t_0)\bigr| \le C|t-t_0|^{1+\alpha}, \qquad \bigl|L_2(t) - L_2(t_0)\bigr| \le D|t-t_0|^{\beta}, \quad (1.2)$$

for $t \in V$, with some $\alpha > 0$, $\beta > 0$ and $C > 0$, $D > 0$.

We shall denote

$$a := L_1'(t_0), \qquad \sigma^2 := L_2(t_0) \qquad \text{and} \qquad b = \sigma/a. \quad (1.3)$$

The *sample median* may serve as an illustration for the smoothness condition (1.1). Assume that $\mathcal{X} = R$, that $f(x,t) = |x-t|$, and that a continuous distribution function F has a unique median, say t_0. Then t_N is the sample median and it is a consistent estimator of t_0. For any given x, the function $f(x,t)$ is not differentiable in t. Nevertheless f satisfies (1.1) with the exceptional (one point) set $U_t = \{t\}$. Condition (1.2) is satisfied if, for $t \in V$, the distribution F has a density, say p, and $p(t) - p(t_0) = O(|t-t_0|^\alpha)$ as $t \to t_0$.

Convexity. There exist $\delta > 0$ and $B > 0$ such that

$$\mathbf{P}\Bigl\{Q_N(t) \ \text{ is convex on } \ (t_0-\delta, t_0+\delta)\Bigr\} \ge 1 - B\,N^{-1/2}. \quad (1.4)$$

In minimization problems the convexity is a natural requirement and this property of Q_N is sufficient for all our applications. In a number of interesting examples such as Huber's robust estimator of location, the sample median, the L_p-median, $p > 1$, the function Q_N is merely convex. Moreover, the law of large numbers yields the asymptotic convexity of Q_N in the case considered by Pfanzagl (1971) as well.

Consistency. Let, for some $\delta > 0$ and $A > 0$,

$$\mathbf{P}\bigl\{\,|t_N - t_0| \ge \delta\bigr\} \le A\,N^{-1/2}. \quad (1.5)$$

Assumptions (1.4) and (1.5) allow to reduce the problem to a purely convex case. Indeed, the probability in (1.5) and $BN^{-1/2}$ in (1.4) are of the same order as the remainder term in the Berry-Esseen bound and hence may be neglected. In many interesting cases the consistency condition (1.5) holds for $\delta = c\sqrt{\ln N/N})$ with a sufficiently large constant $c = c(f,F)$. Our results are applicable in these cases and for our purposes it suffices to choose the parameter δ in (1.4) of the same order.

Denote

$$\beta_3 = \sup\Bigl\{\beta_3(t) : |t-t_0| \le 6\,b\sqrt{\ln N/N}\Bigr\}, \qquad \text{where} \ \ \beta_3(t) = \mathbf{E}\,\bigl|f'(X,t)\bigr|^3.$$

Theorem 1.1. *Suppose that* (1.1), (1.2), (1.4) *and* (1.5) *are satisfied with some* $\delta \geq 6\, b\, \sqrt{\ln N/N}$, *some* A, B, C, D, *and with* $\alpha = \beta = 1$.

Then

$$\Delta_N = \sup_{u \in R} \Big| \mathbf{P}\big\{ b^{-1} N^{1/2} (t_N - t_0) < u \big\} - \Phi(u) \Big| \\ \leq c\, N^{-1/2} \big(A + B + \beta_3/\sigma^3 + a^{-1} b\, C + b\, D\, \sigma^{-2} \big), \tag{1.6}$$

for N *satisfying*

$$\frac{N}{\ln N} \geq c \left(a^{-2} b^2 C^2 + \sigma^{-4} b^2 D^2 + 1 \right), \tag{1.7}$$

where c *denotes an universal constant.*

Throughout $c, c_1, c_2, \ldots$ will denote generic absolute constants. We write $c(T_1, T_2, \ldots)$ when the constant depends on $T_1, T_2, \ldots$.

The bound of Theorem 1.1 is explicit. Our notation might suggest that we are considering a fixed underlying distribution F and fixed values of the parameters δ, A, B, C and D. In fact F and $\delta = \delta_N$, $A = A_N, \ldots$ may depend on N.

For the sample mean $t_N = N^{-1}(X_1 + \cdots + X_N)$ Theorem 1.1 yields a bound for Δ_N which is asymptotically equivalent to the classical Berry–Esseen bound. Indeed, t_N, the minimizer of $Q_N(t) = 2^{-1} N^{-1} \sum_{i=1}^N (X_i - t)^2$, is the sample mean and in this case Theorem 1.1 implies

$$\limsup_N \Delta_N N^{1/2} < c\, \mathbf{E} \left| X - \mathbf{E}\, X \right|^3 / \sigma^3,$$

provided that $\sigma^2 = \mathbf{E}\, (X - \mathbf{E}\, X)^2 > 0$.

Conditions (1.1) and (1.2) are satisfied when the score function and the underlying distribution in combination are sufficiently smooth. Thus, if the distribution F is arbitrary we need to impose stronger regularity conditions on the function f, and vice versa. For instance, conditions (1.1) and (1.2) hold for arbitrary F if f is twice differentiable in $t \in V$ and, for some $0 < \alpha \leq 1$,

$$\big| f''(x,t) - f''(x,s) \big| \leq h(x) \cdot |t - s|^{\alpha}, \tag{1.8}$$

for $s, t \in V$ and $x \in \mathcal{X}$. In this case (1.2) is satisfied with α as in (1.8) and with $\beta = 1$. Here h is a measurable function such that $\mathbf{E}\, h^2(X) < \infty$. Evidently, (1.8) is satisfied for $f(x,t) = (x - t)^2$. Thus, Theorem 1.1 yields the Berry–Esseen bound for the sample mean without any conditions on the smoothness of F.

An example when the smoothness of F becomes important is the sample median. In this case the score function $f(x,t) = |x - t|$ satisfies (1.1) whenever F is continuous. Furthermore, condition (1.2) is satisfied with $\alpha = \beta = 1$ if F has a bounded density function, say p, and

$$\big| p(t) - p(t_0) \big| \leq c(F)\, |t - t_0|, \tag{1.9}$$

for $t \in V$. If, moreover, $p(t_0) > 0$, then Theorem 1.1 yields the Berry–Esseen bound $O(N^{-1/2})$, see Proposition 2.1 in Section 2.

Theorem 1.1 is a consequence of a more general result given in the next theorem. Denote

$$\overline{\beta}(t) = \mathbf{E}\left|f'(X,t)\right|^3 \mathrm{I\!I}\Big\{\left|f'(X,t)\right| \le \sigma\sqrt{N}\Big\} + \sigma\sqrt{N}\mathbf{E}\left|f'(X,t)\right|^2 \mathrm{I\!I}\Big\{\left|f'(X,t)\right| > \sigma\sqrt{N}\Big\}$$

and let

$$\overline{\beta} = \sup\Big\{\overline{\beta}(t) : |t - t_0| \le 6\, b\,\sqrt{\ln N/N}\Big\}.$$

Theorem 1.2. *Suppose that* (1.1), (1.2), (1.4) *and* (1.5) *are satisfied with some* $\delta \ge 6b\sqrt{\ln N/N}$, *some* A, B, C, D *and with some* $\alpha, \beta \in (0,1]$.

Then there exists an absolute constant c *such that*

$$\Delta_N \le c\,N^{-1/2}(A + B + \overline{\beta}/\sigma^3) + c\,a^{-1}b^{\alpha}CN^{-\alpha/2} + c\,b^{\beta}D\sigma^{-2}N^{-\beta/2}, \tag{1.10}$$

for N *such that*

$$\frac{N}{\ln N} \ge c\Big(9^{1/\alpha}a^{-2/\alpha}b^2C^{2/\alpha} + 9^{1/\beta}b^2D^{2/\beta}\sigma^{-4/\beta} + 1\Big). \tag{1.11}$$

Theorem 1.2 is applicable to a large class of M-estimators. In particular Theorem 1.2 yields the Berry–Esseen bound $O(N^{-1/2})$ for the sample median, Huber's estimator of location and for the sample L_p-median, $p > 1$, under optimal smoothness conditions on the distribution F. These examples are described in Section 2. For M-estimators defined by means of smooth score functions, i.e., twice differentiable functions satisfying (1.8), Theorem 1.2 yields the following result.

For a function f satisfying (1.8) denote

$$\varepsilon = \Big(\frac{a}{4\mathbf{E}\,h(X)}\Big)^{1/\alpha}$$

and put $\varepsilon = \infty$ if $\mathbf{E}\,h(X) = 0$.

Corollary 1.3. *Suppose that the smoothness condition* (1.8) *is satisfied for* $s, t \in V = (t_0 - \delta', t_0 + \delta')$ *with some function* h *and some* $\alpha \in (0,1]$. *Suppose that the consistency condition* (1.5) *holds with some* A *and some* δ *such that*

$$6b\sqrt{\ln N/N} \le \delta \le \min\{\varepsilon; \delta'\}.$$

Then there exists an absolute constant c *such that*

$$\Delta_N \le c\,N^{-1/2}(A + B + b\sigma^2D + \overline{\beta}/\sigma^3) + c\,N^{-\alpha/2}a^{-1}b^{\alpha}C, \tag{1.12}$$

for N satisfying

$$\frac{N}{\ln N} \geq c\left(9^{1/\alpha} a^{-2/\alpha} b^2 C^{2/\alpha} + b^2 D^2 \sigma^{-4} + 1\right) \tag{1.13}$$

with

$$B = c\, N^{-1/2}\Big(a^{-2}(\xi + \eta\big(\mathbf{E}\, h(X)\big)^{-2}\Big), \quad C = \mathbf{E}\, h(X), \quad D = c\,(\sigma^2 + \xi + \eta)(1 + \delta^{1+2\alpha}),$$

where $\xi = \mathbf{E}\left(f''(X, t_0)\right)^2$ *and* $\eta = \mathbf{E}\, h^2(X)$.

In order to compare our result with that of Pfanzagl (1971) let us consider the case when $\beta_3(t_0) < \infty$, the smoothness condition (1.8) is satisfied with $\alpha = 1$ and the consistency condition (1.5) holds with some $A < \infty$ which is independent of N. In this case Corollary 1.3 yields $\Delta_N = O(N^{-1/2})$. A similar asymptotic result, but under a bit more restrictive conditions was proved by Pfanzagl (1971) for minimum contrast estimators. In particular, instead of the consistency condition (1.5), Pfanzagl (1971) assumed that the distribution F belongs to the parametric family defining the minimum contrast estimator.

Observe, that Corollary 1.3 yields the Berry–Esseen bound $O(N^{-\alpha/2})$ if the smoothness condition (1.8) is satisfied with a parameter α. The order $O(N^{-\alpha/2})$ depends on the smoothness of f in an optimal way. To see this consider an example. Let t_N be the minimizer of the function $N^{-1}\sum_{i=1}^N \big|X_i - t\big|^{2+\alpha}$ and assume that X has a symmetric distribution such that $\mathbf{P}\{|X| = 1\} = \mathbf{P}\{X = 0\} = 1/2$. Then, see Example 2.8 in Section 2,

$$\liminf_N N^{\alpha/2}\, \Delta_N > 0.$$

In this case the score function $f(x,t) = |x - t|^{2+\alpha}$ satisfies (1.8) with the exponent α and Corollary 1.3 yields $\Delta_N = O(N^{-\alpha/2})$.

In Corollary 1.3 we do not need to impose the asymptotic convexity condition (1.4) since it is implied by the smoothness condition (1.8). In this case the function Q_N is twice differentiable and hence it is convex on the interval $(t_0 - \delta, t_0 + \delta)$ provided that

$$Q_N''(t) = N^{-1} \sum_{i=1}^N f''(X_i, t) \geq 0, \qquad t \in (t_0 - \delta, t_0 + \delta). \tag{1.14}$$

The smoothness condition (1.8) and the law of large numbers for $\sum_{i=1}^N f''(X_i, t_0)$ imply that the probability of (1.14) tends to 1, provided that $\delta \leq \varepsilon$. Hence (1.4) is fulfilled.

The smoothness condition (1.8) can be replaced by the weaker one: for some $H \geq 0$ and $p, q > 1$,

$$\mathbf{E}\left|f''(X,t) - f''(X,s)\right|^p \leq H\left|t - s\right|^q, \tag{1.15}$$

for $s, t \in [t_0 - \delta', t_0 + \delta']$. Condition (1.15) allows to estimate the oscillation of the random process $Q''_N(t)$, see, e.g., Talagrand (1990). Such an estimate together with the law of large numbers for $\sum_{i=1}^N f''(X_i, t_0)$ implies (1.4) since the probability of (1.14) tends to 1 as $N \to \infty$, for δ sufficiently small (but independent on N). Furthermore, (1.15) implies that there exists a version of the random process $Q''_N(t)$ with continuous sample paths. Hence, in presence of (1.15) the only smoothness requirement on f is that, for $t \in [t_0 - \delta', t_0 + \delta']$,

$$f''(x,t) = \lim_{\tau \to 0} \big(f'(x, t+\tau) - f'(x,t)\big)/\tau \quad \text{exists,} \tag{1.16}$$

for F-almost all $x \in \mathcal{X}$.

Corollary 1.4. *Suppose that* (1.15) *and* (1.16) *are satisfied with some* δ', H *and some* $p, q > 1$. *Suppose that* (1.5) *holds with some* A *and* δ *such that*

$$6\, b\, \sqrt{\ln N/N} \le \delta \le \min\Big\{ \big(2^{-1} a H^{-1/p}\big)^{p/q};\ \delta' \Big\}.$$

Then there exists an absolute constant c *such that, for* N *satisfying* (1.13),

$$\begin{aligned} \Delta_N &\le c\, N^{-1/2}(A + bD\sigma^{-2} + \overline{\beta}/\sigma^3) + c\, N^{-q/2p} a^{-1} b^{q/p} C \\ &\quad + c(p) N^{-\gamma} a^{-p} \varkappa + c(p,q) N^{-\gamma/p} a^{-1} H^{1/p}(1 + 2\delta), \end{aligned}$$

where $\gamma = \min\{p-1; p/2\}$, $r = p/(p-1)$, $\varkappa = \mathbf{E}\,\big|f''(X, t_0)\big|^p$,

$$C = H^{1/p}, \quad D = \big(\varkappa + H\, \delta^q\big)^{1/p} \beta_r^{1/r}, \quad \beta_r = \sup\Big\{ \mathbf{E}\,\big|f'(X,t)\big|^r :\ |t - t_0| \le \delta \Big\}.$$

Remark. Our approach allows to construct bounds for Δ_N even under more general conditions than (1.1) and (1.2). For instance, in Theorem 1.2 it suffices to assume that conditions (1.1) and (1.2) hold for $t \in V'$, where V' is a dense subset of V. In particular the Berry–Esseen bound $O(N^{-1/2})$ holds for the sample median in the cases when F might have atoms in any neighborhood of t_0. More precisely, it is sufficient to require that F is representable as a sum $F_0 + F_1$, where the measure F_0 has a positive density satisfying (1.9) in a neighborhood of the median t_0, and where F_1 is a discrete measure which assigns probabilities $p_i > 0$ to points x_i such that

$$\sum_{i:\, |x_i - t_0| < u} p_i = O(u^{-2}) \qquad \text{as} \quad u \to 0.$$

Another possibility to weaken (1.2) is to replace $|t - t_0|^{1+\alpha}$ and $|t - t_0|^\beta$ by $\phi_1(|t-t_0|)$ and $\phi_2(|t - t_0|)$ with functions $\phi_1(\tau) = o(\tau)$ and $\phi_2(\tau) = o(1)$, as $\tau \to 0$.

The rest of the paper is organized as follows. In the Section 2 we consider various applications of the main result and give several examples. In Section 3 we prove Theorems 1.1, 1.2 and Corollaries 1.3 and 1.4. Calculations related to the applications are postponed to Section 4.

2. Applications and Examples

In this section we apply Theorem 1.2 to the *sample median*, the *sample quantiles, Huber's estimator of location* and to the L_p*-median,* $p>1$. These statistics are M-estimators with special score functions. Hence in each particular case we have a fixed score function f and our aim is to find minimal smoothness conditions on the distribution F which ensure a Berry–Esseen bound. The examples are arranged in order of increasing smoothness of the score function.

Let F denote the distribution function of a real random variable X. If F is absolutely continuous, p will denote a density function of F. Recall that

$$\Delta_N=\sup_{u\in R}\Big|\mathbf{P}\big\{b^{-1}\,N^{1/2}(t_N-t_0)<u\big\}-\Phi(u)\Big|$$

with $b=\sigma/a$ to be specified in each particular example.

2.1. The sample median. Assume that F has the unique median t_0. Let $f(x,t)=|x-t|$. A random variable t_N which minimizes $Q_N(t)$ is called the sample median. In this case $b=1/\big(2p(t_0)\big)$.

Proposition 2.1. *Assume that F has a density p on $(t_0-\delta,t_0+\delta)$. Suppose that $p(t_0)>0$ and that, for some $\alpha\in(0,1]$ and $H>0$,*

$$|p(t)-p(t_0)|\le H|t-t_0|^{\alpha},\qquad\qquad |t-t_0|<\delta. \tag{2.1}$$

Then there exists an absolute constant c such that

$$\Delta_N\le c\,N^{-\alpha/2}Hp^{-1-\alpha}(t_0)+c\,N^{-1/2}, \tag{2.2}$$

for N satisfying

$$\frac{N}{\ln N}>\frac{c}{\mathbf{P}^2\ \ t_0-\delta<X<t_0}+\frac{c}{\mathbf{P}^2\ \ t_0<X<t_0+\delta}+\frac{c\,36^{1/\alpha}H^{2/\alpha}}{p^{2+2/\alpha}(t_0)}\,.$$

In particular, the Berry-Esseen bound $O(N^{-1/2})$ holds if F is twice differentiable at the point t_0 and $p(t_0)>0$.

Next we give an example which shows that the optimal rate assuming condition (2.1) is $O(N^{-\alpha/2})$.

Example 2.2. Let $\alpha\in(0,1]$ and let F be a distribution with the density $p(t)=c_\alpha\big(1-|t|^\alpha\big)$, for $|t|\le 1$, where $c_\alpha=(\alpha+1)/(2\alpha)$. It is easy to see that F has a unique median $t_0=0$ and $p(0)=c_\alpha$. Condition (2.1) of Proposition 2.1 is satisfied with exponent α and $H=2\,c_\alpha$. Hence (2.2) yields $\Delta_N=O(N^{-\alpha/2})$. The rate $O(N^{-\alpha/2})$ can not be improved since $\liminf_N N^{\alpha/2}\Delta_N>0$.

2.2. Sample quantiles. Let $0 < p < 1$ and let $f(u) = |u| + (1 - 2p)\,u$. A random variable t_N which minimizes the function $t \to N^{-1} \sum_{i=1}^{N} f(t - X_i)$ is a sample p-quantile. Theorem 1.2 applies to the sample p-quantile and yields a Berry–Esseen bound like (2.2) under the same regularity condition as in Proposition 2.1.

Using a different approach and a somewhat more restrictive condition on the smoothness of F a Berry–Esseen bound $O(N^{-1/2})$ for sample quantiles, $0 < p < 1$, was obtained in Reiss (1974), see also Serfling (1980). In particular, Reiss (1974) and Serfling (1980) assumed that the distribution function F is twice differentiable in a neighborhood of t_0.

For X non degenerate, the (sample) L_p-median, $p > 1$ is defined as the minimizer t_N of $Q_N(t)$ with the score function $f(x,t) = |x - t|^p$. In this case the function $f(x,t) = |x - t|^p$ is strictly convex and differentiable in t. Hence, the unique L_p-median $t_0 = \operatorname{argmin}_t \mathbf{E}\left|X - t\right|^p$ exists provided that $\mathbf{E}\,|X|^p < \infty$. Furthermore,

$$a = p(p-1)\mathbf{E}\,|X - t_0|^{p-2} \qquad \text{and} \qquad \sigma^2 = p^2 \mathbf{E}\,|X - t_0|^{2(p-1)}.$$

2.3. The L_p-median, $1 < p < 2$.

Proposition 2.3. *Assume that for some $\delta > 0$, $C > 0$ and $\beta \in (2 - p, 1]$,*

$$|F(t) - F(t_0)| < C|t - t_0|^\beta, \qquad |t - t_0| < \delta. \tag{2.3}$$

Suppose that $\mathbf{E}\,|X|^{(p-1)(p+\beta)} < \infty$. Then $\Delta_N = O(N^{-(p+\beta-2)/2})$.

Observe, that conditions of the proposition ensure that $a < \infty$ and $\sigma^2 < \infty$.

Assuming the smoothness condition (2.3) the rate $O(N^{-(p-1)/2})$ is unimprovable. In order to prove the rate $O(N^{-(p-1+\beta)/2})$ with some $\beta > 0$ we need to impose a stronger regularity condition on F.

Proposition 2.4. *Assume that F has a density p on $(t_0 - \delta, t_0 + \delta)$ such that, for some $C > 0$ and $\beta \in (0, 2 - p)$,*

$$|p(t) - p(t_0)| < C|t - t_0|^\beta, \qquad |t - t_0| < \delta. \tag{2.4}$$

Then the condition $\mathbf{E}\,|X|^{(p-1)(1+p+\beta)} < \infty$ implies $\Delta_N = O(N^{-(p+\beta-1)/2})$.

One may expect that if F has a density on $(t_0 - \delta, t_0 + \delta)$ satisfying

$$|p(t) - p(t_0)| < C|t - t_0|^{2-p}, \qquad |t - t_0| < \delta, \tag{2.5}$$

the Berry–Esseen bound $O(N^{-1/2})$ holds. However this condition is not sufficient.

Proposition 2.5. *Suppose that* (2.5)*. holds. Moreover, assume that the function*

$$t \to |t - t_0|^{p-3} \big| p(t) - p(t_0) \big|$$

is integrable in a neighborhood of t_0*. If* $\mathbf{E}\,|X|^{3(p-1)} < \infty$ *then* $\Delta_N = O(N^{-1/2})$*.*

The smoothness condition of Proposition 2.5 is close to optimal. In Section 4 we shall show that, if $\mathbf{E}\,|X|^{3(p-1)} < \infty$ and F has a density p such that $p(t_0) > 0$ and

$$p(t_0 + u) = p(t_0) + u^{2-p}, \qquad p(t_0 - u) = p(t_0) - u^{2-p}, \qquad\qquad 0 < u < \delta,$$

for some $\delta > 0$, then

$$\liminf_N N^{1/2} \ln^{-1} N \, \Delta_N > 0. \tag{2.6}$$

2.4. Huber's estimator of location. Assume that X is non degenerate. Let $k > 0$. The Huber (1964) estimator of location is the minimizer of $Q_N(t) = N^{-1} \sum_{i=1}^N \rho_k(X_i - t)$, where

$$\rho_k(t) = \frac{t^2}{2}, \quad \text{for} \quad |t| \le k, \qquad \text{and} \qquad \rho_k(t) = k|t| - \frac{k^2}{2}, \quad \text{for} \quad |t| > k.$$

It is easy to see that

$$\text{there exists a unique } \; t_0 \; \text{ such that } \; \mathbf{E}\, \rho_k'(X - t_0) = 0 \tag{2.7}$$

and

$$\mathbf{P}\{|X - t_0| < k\} > 0 \tag{2.8}$$

provided that k is sufficiently large. In particular, the parameters

$$a = \mathbf{P}\{|X - t_0| < k\}, \;\; \sigma^2 = k^2 \mathbf{P}\{|X - t_0| > k\} + \mathbf{E}\,|X - t_0|^2 \mathbb{I}\{|X - t_0| \le k\}$$

are positive.

Proposition 2.6. *Assume that* (2.7) *and* (2.8) *hold. Assume that, for* $t_1 = t_0 + k$ *and* $t_2 = t_0 - k$*,*

$$\big| F(t_i + h) - F(t_i) \big| \le H\, |h|^\alpha, \qquad\qquad |h| < \delta, \qquad i = 1, 2, \tag{2.9}$$

with some $H > 0$*,* $\alpha \in (0, 1]$ *and* $\delta > 0$*. Then*

$$\Delta_N \le c\, N^{-\alpha/2} H\, k^{2\alpha} / a^{1+\alpha} + c\, N^{-1/2} \Big(k^3/\sigma^3 + k/\big(a\,\sigma\big) \Big) + o\big(N^{-1/2}\big).$$

Observe that the derivative $t \to \rho_k''(t - t_0)$ is discontinuous at the points $t_0 + k$ and $t_0 - k$. Therefore, in order to obtain a Berry–Esseen bound we need to impose the smoothness condition (2.9). In particular, the bound $O(N^{-1/2})$ holds if F is differentiable at $t_0 + k$ and $t_0 - k$.

2.5. The L_p-median, $2 < p < 3$.

Proposition 2.7. *Assume that* $\mathbf{E}\,|X|^{(p-1)(2+\beta)} < \infty$ *with some* $\beta \in (0, p-2]$. *Then* $\Delta_N = O(N^{-\beta/2})$.

The following example shows that $O(N^{-(p-2)/2})$ is the optimal rate without any assumption on the smoothness of F.

Example 2.8. Let X be a random variable with a symmetric distribution such that $\mathbf{P}\{|X| = 1\} = \mathbf{P}\{X = 0\} = 1/2$. Then $t_0 = 0$, $a = p(p-1)/2$ and $\sigma^2 = p^2/2$. An application of Proposition 2.7 yields $\Delta_N = O(N^{-(p-2)/2})$. The rate is optimal since it is easy to show that $\liminf_N N^{(p-2)/2}\Delta_N > 0$.

In order to obtain the improved bound $O(N^{-(p-2+\gamma)/2})$ with some $\gamma > 0$ we need to impose a regularity condition on F.

Proposition 2.9. *Assume that, for some* $C > 0$, $\gamma \in (0, 3-p)$ *and* $\delta > 0$,

$$|F(t) - F(t_0)| < C|t - t_0|^\gamma, \qquad |t - t_0| < \delta. \tag{2.10}$$

Suppose that $\mathbf{E}\,|X|^{(p-1)(p+\gamma)} < \infty$. *Then* $\Delta_N = O\big(N^{-(p-2+\gamma)/2}\big)$.

One may expect that the smoothness condition

$$\big|F(t) - F(t_0)\big| < C|t - t_0|^{3-p}, \qquad |t - t_0| < \delta \tag{2.11}$$

is sufficient for proving bound $O(N^{-1/2})$. However this is not the case. As in Proposition 2.5 we need to impose a stronger condition on F.

Proposition 2.10. *Suppose that* (2.11) *holds with some* $\delta > 0$ *and* $C > 0$. *Moreover, assume that the function*

$$t \to |t - t_0|^{p-4}\big|F(t) - F(t_0)\big|$$

is integrable in a neighborhood of t_0. *If* $\mathbf{E}\,|X|^{3(p-1)} < \infty$ *then* $\Delta_N = O(N^{-1/2})$.

2.6. The L_p-median, $p \geq 3$. If $p \geq 3$ then the Berry–Esseen bound for the (sample) L_p-median holds without conditions on the smoothness of F, since in this case the score function satisfies the smoothness condition (1.8) with $\alpha = 1$. Hence Corollary (1.3) yields the bound $O(N^{-\beta/2})$ provided that $\mathbf{E}\,|X|^{(p-1)(2+\beta)} < \infty$, for $0 < \beta \leq 1$.

3. Proofs

Proof of Theorem 1.1. Theorem 1.1 is a particular case of Theorem 1.2.

Proof of Corollary 1.3. Corollary 1.3 is a consequence of Theorem 1.2. In what follows we show that conditions of Theorem 1.2 are satisfied. Observe that

$$|f''(x,t)| \leq |f''(x,t_0)| + \delta^\alpha h(x), \qquad \text{for} \quad |t - t_0| < \delta.$$

Hence $\dfrac{\partial}{\partial t}\,\mathbf{E}\, f'(X,t) = \mathbf{E}\, f''(X,t)$. In particular, $a = \mathbf{E}\, f''(X,t_0)$. Calculations show that (1.2) holds with α as in (1.8) and with

$$\beta = 1, \qquad C = \mathbf{E}\, h(X), \qquad D = c\,(\sigma^2 + \xi + \eta)(1 + \delta^{1+2\alpha}).$$

It remains to verify condition (1.4). Since $a = \mathbf{E}\, f''(X,t_0)$, an application of Chebyshev's inequality yields

$$\mathbf{P}\big\{Q_N''(t_0) < a/2\big\} \le N^{-1} 4a^{-2}\mathbf{E}\,\big(f''(X,t_0) - \mathbf{E}\, f''(X,t_0)\big)^2 < N^{-1} c a^{-2} \xi. \tag{3.1}$$

If $\mathbf{E}\, h(X) = 0$ (1.4) immediately follows from (3.1) and (1.8). If $\mathbf{E}\, h(X) > 0$ an application of Chebyshev's inequality yields

$$\mathbf{P}\Big\{N^{-1}\sum_{i=1}^N h(X_i) > 2\mathbf{E}\, h(X)\Big\} \le \frac{\mathbf{E}\ \ h(X) - \mathbf{E}\, h(X)\ \ ^2}{N\ \ \mathbf{E}\, h(X)\ \ ^2} < \frac{c\,\eta}{N\ \ \mathbf{E}\, h(X)\ \ ^2}. \tag{3.2}$$

By (1.8),

$$\sup_{|t-t_0|<\delta} \big|Q_N''(t) - Q_N''(t_0)\big| \le \delta^\alpha N^{-1} \sum_{i=1}^N h(X_i).$$

Combining this inequality with (3.1) and (3.2) we obtain that

$$\mathbf{P}\Big\{\inf_{|t-t_0|<\delta} Q_N''(t) > 0\Big\} \ge 1 - c\, N^{-1}\Big(a^{-2}\xi + \eta\big(\mathbf{E}\, h(X)\big)^{-2}\Big)$$

provided that $\delta \le \big(4\,\mathbf{E}\, h(X)/a\big)^{-1/\alpha}$. Hence (1.4) follows with

$$B = cN^{-1/2}\Big(a^{-2}(\xi + \eta\big(\mathbf{E}\, h(X)\big)^{-2}\Big),$$

which completes the proof.

Proof of Corollary 1.4. Corollary 1.4 is a consequence of Theorem 1.2. In what follows we check the conditions of Theorem 1.2. By Lemma 3.1 below, the condition (1.4) is satisfied with a constant

$$B = N^{1/2}\Big(c(p) N^{-\gamma} a^{-p} \varkappa + c(p,q) N^{-\gamma/p} a^{-1} H^{1/p}(1 + 2\delta)\Big).$$

Furthermore,

$$|f''(X,t)| \le g(X), \quad g(X) := |f''(X,t_0)| + \sup_{t_0-\delta\le s,\, t\le t_0+\delta} \big(f''(X,s) - f''(X,t)\big).$$

It is shown in the proof of Lemma 3.1 that $\mathbf{E}\, g(X) < \infty$. Hence $\dfrac{\partial}{\partial t}\mathbf{E}\, f'(X,t) = \mathbf{E}\, f''(X,t)$, for $|t - t_0| < \delta$.

It remains to verify condition (1.2). Using (1.15) one may show that the condition (1.2) is satisfied with

$$\alpha = q/p, \quad \beta = 1 \qquad \text{and} \quad C = H^{1/p}, \quad D = \big(\varkappa + H\delta^q\big)^{1/p}\beta_r^{1/r}.$$

An application of Theorem 1.2 completes the proof.

Proof of Theorem 1.2. Without loss of generality we may assume that all random variables are defined on a common probability space $\big(\Omega, \mathcal{P}\big)$. Furthermore, we may and shall assume that $0 \in T$ and $t_0 = 0$.

In view of (1.11) it suffices to bound Δ_N for N such that

$$\frac{\ln N}{N} \geq c\left(9^{1/\alpha}C^{2/\alpha}b^2a^{-2/\alpha} + 9^{1/\beta}D^{2/\beta}b^2\sigma^{-4/\beta} + 1\right). \tag{3.3}$$

The proof of the theorem will consist of two steps. First, using the asymptotic convexity condition (1.4) we shall show that the probability $\mathbf{P}\big\{t_N < t\big\}$ is close to $\mathbf{P}\big\{0 < Q_N'(t)\big\}$, for $t \in (-\delta, \delta)$, i.e., that

$$\mathbf{P}\big\{0 < Q_N'(t)\big\} - \frac{A}{\sqrt{N}} - \frac{B}{\sqrt{N}} \leq \mathbf{P}\big\{t_N < t\big\} \leq \mathbf{P}\big\{0 \leq Q_N'(t)\big\} + \frac{A}{\sqrt{N}} + \frac{B}{\sqrt{N}}\,. \tag{3.4}$$

To estimate the probability $\mathbf{P}\big\{0 \leq Q_N'(t)\big\}$ in the second step we may apply the classical Berry–Esseen bound since, for given t, the random variable $Q_N'(t) = N^{-1}\sum_{i=1}^N Y_i(t)$ is the sum of real random variables $Y_i(t) = f'(X_i, t)$.

To verify (3.4) we shall use a simple but useful observation concerning convex functions, cf. Huber (1964). Let $t \to g(t)$ be a convex function. If t_* is a minimizer of g and $g'(t)$ exists then

$$0 < g'(t) \quad\Rightarrow\quad t_* < t \quad\Rightarrow\quad 0 \leq g'(t), \tag{3.5}$$

which means simply that $g(t)$ is nondecreasing, for $t \geq t_*$, and it is non increasing, for $t \leq t_*$. The monotone derivative g' of the convex function g exists everywhere except at most countable set of points, see Schvartz (1967). To make (3.5) valid for all t, extend (if necessary) the function $g'(t)$ so that the extension $\overline{g}'(t)$ is monotone and $\overline{g}'(t_*) = 0$, for all minimizers t_* of g. The asymptotic convexity condition (1.4) allows to apply (3.5) to the function $t \to Q_N(\omega, t)$ and to derive (3.4). Indeed, by (1.4) and (1.5), the events

$$\mathcal{A}_\delta = \big\{\omega : Q_N(t,\omega) \text{ is convex in } t, \text{ for } t \in (-\delta,\delta)\big\}, \qquad \mathcal{B}_\delta = \big\{\omega : |t_N| < \delta\big\}$$

satisfy

$$\mathbf{P}\big\{\mathcal{A}_\delta\big\} > 1 - N^{-1/2}\,B \qquad \text{and} \qquad \mathbf{P}\big\{\mathcal{B}_\delta\big\} > 1 - N^{-1/2}\,A.$$

Hence, (3.4) reduces to the verification of

$$0 < \overline{Q}'_N(\omega, t) \;\;\Rightarrow\;\; t_N(\omega) < t \;\;\Rightarrow\;\; 0 \le \overline{Q}'_N(\omega, t),$$

for $\omega \in \mathcal{A}_\delta \cap \mathcal{B}_\delta$, where $\overline{Q}'_N$ denotes the extension of the derivative of the convex function Q_N. But, for fixed t, the random variables $\overline{Q}'_N(t)$ and $Q'_N(t)$ coincide almost surely, and (3.4) follows.

Let us apply the Berry–Esseen bound to $Q'_N(t)$. Denote

$$Y(t) = f'(X, t), \qquad m_t = \mathbf{E}\,Y(t), \qquad \sigma_t = \Big(\mathbf{E}\,\big(Y(t) - \mathbf{E}\,Y(t)\big)^2\Big)^{1/2}$$

and $\eta = Y(t) - m_t$,

$$\hat\beta_t = \mathbf{E}\,|\eta|^3 \mathbb{I}\big\{|\eta| < \sigma_t\sqrt{N}\big\} + \sigma_t\sqrt{N}\mathbf{E}\,|\eta|^2 \mathbb{I}\big\{|\eta| \ge \sigma_t\sqrt{N}\big\}.$$

We have

$$\Big|\mathbf{P}\Big\{0 < \sum_{i=1}^N Y_i(t)\Big\} - \Phi\big(\sqrt{N} m_t/\sigma_t\big)\Big| \le c\,N^{-1/2}\hat\beta_t/\sigma_t^3, \tag{3.6}$$

see, e.g., Theorem 8 of Chapter V in Petrov (1975).

Observe that

$$m_t \;=\; \mathbf{E}\,Y(t) \;=\; L_1(t) \qquad \text{and} \qquad \sigma_t^2 \;=\; L_2(t) - L_1^2(t).$$

It follows from (1.2) that, for $|t| \le 6\,b\sqrt{\ln N/N}$,

$$\big|m_t - a\,t\big| \le C|t|^{1+\alpha} \le C\big(6\,b\sqrt{\ln N/N}\big)^{1+\alpha} \le \sigma/2$$

and

$$\frac{1}{2}\,\sigma^2 \le \sigma_t^2 \le \frac{3}{2}\,\sigma^2, \tag{3.7}$$

provided that (3.3) holds. It follows from these inequalities that

$$\hat\beta_t/\sigma_t^3 \le c\,\overline{\beta}/\sigma^3 + c\,N^{-1}, \quad \text{for} \quad |t| \le 6\,b\sqrt{\ln N/N}, \tag{3.8}$$

provided that (3.3) holds.

Recall that $\Delta_N = \sup_u \Delta_N(u)$, where

$$\Delta_N(u) = \Big|\mathbf{P}\big\{b^{-1}\,N^{-1/2} t_N < u\big\} - \Phi(u)\Big|.$$

In what follows we show that $\Delta_N(u)$ does not exceed the right hand side of (1.10). We shall consider separately two cases: $|u| < 6\sqrt{\ln N}$ and $|u| \geq 6\sqrt{\ln N}$. In the case $|u| \geq 6\sqrt{\ln N}$ the bound to prove follows from the estimate

$$\mathbf{P}\big\{|t_N| \geq 6\,b\sqrt{\ln N/N}\big\} \leq c\,N^{-1/2}\big(1 + A + B + \overline{\beta}(t_1)/\sigma^3 + \overline{\beta}(t_2)/\sigma^3\big), \tag{3.9}$$

where

$$t_1 = 6\,b\sqrt{\ln N/N} \qquad \text{and} \qquad t_2 = -6\,b\sqrt{\ln N/N}.$$

Indeed, (3.9) implies

$$\begin{aligned}\Delta_N(u) &\leq N^{-1/2}c\left(1 + A + B + \overline{\beta}(t_1)/\sigma^3 + \overline{\beta}(t_2)/\sigma^3\right) + 2\Phi(-6\sqrt{\ln N})\\ &\leq N^{-1/2}c\left(1 + A + B + \overline{\beta}(t_1)/\sigma^3 + \overline{\beta}(t_2)/\sigma^3\right).\end{aligned}$$

Let us prove (3.9). It follows from (3.4)–(3.6) and (3.8) that the probability $\mathbf{P}\big\{|t_N| \geq 6\,b\sqrt{\ln N/N}\big\}$ does not exceed

$$c\,N^{-1/2}\big(1+A+B+\overline{\beta}(t_1)/\sigma^3+\overline{\beta}(t_2)/\sigma^3\big)+\Big[\Phi\big(\sqrt{N}m_{t_2}/\sigma_{t_2}\big)+1-\Phi\big(\sqrt{N}m_{t_1}/\sigma_{t_1}\big)\Big].$$

Using (1.2) and (3.3) it is easy to show that

$$\sqrt{N}m_{t_2}/\sigma_{t_2} \leq -2\sqrt{\ln N} \qquad \text{and} \qquad \sqrt{N}m_{t_1}/\sigma_{t_1} \geq 2\sqrt{\ln N}.$$

Hence the quantity in the brackets $[\cdots]$ is bounded by $c\,N^{-1/2}$ and we obtain (3.9).

It remains to estimate $\Delta_N(u)$, for $|u| \leq 6\sqrt{\ln N}$. Let $t = N^{-1/2}u\,b$. It follows from (3.4)–(3.6) and (3.8) that

$$\Delta_N(u) \leq N^{-1/2}c\big(1 + A + B + \overline{\beta}(t)/\sigma^3\big) + R, \qquad R = \Big|\Phi\big(\sqrt{N}m_t/\sigma_t\big) - \Phi(u)\Big|. \tag{3.10}$$

Thus the theorem would follow if we show that

$$R \leq c\left(a^{-1}b^{\alpha}CN^{-\alpha/2} + b^{\beta}D\sigma^{-2}N^{-\beta/2} + N^{-1/2}\right).$$

By the Lagrange mean value it is sufficient to verify that the difference $d := \sqrt{N}m_t/\sigma_t - u$ satisfies

$$|d| \leq c\,Ca^{-1}b^{\alpha}N^{-\alpha/2}u^{1+\alpha} + D\sigma^{-2}b^{\beta}N^{-\beta/2}u^{1+\beta} + N^{-1/2}u. \tag{3.11}$$

Write

$$d = \sqrt{N}(m_t - at)/\sigma_t - \frac{\sqrt{N}at(\sigma_t - \sigma)}{\sigma\cdot\sigma_t}\,.$$

Substituting

$$\sigma_t - \sigma = (\sigma_t + \sigma)^{-1}(\sigma_t^2 - \sigma^2) = (\sigma_t + \sigma)^{-1}\big(L_2(t) - \sigma^2 - L_1^2(t)\big)$$

and using the estimate (3.7) we obtain

$$|d| \leq \sqrt{2}\sigma^{-1}\sqrt{N}C|t|^{1+\alpha} + \sigma^{-3}a\sqrt{N}|t|\big(D|t|^{\beta} + L_1(t)^2\big). \tag{3.12}$$

By (3.3) we may estimate

$$\sigma^{-3}a\sqrt{N}|t|L_1^2(t) = u\sigma^{-2}L_1^2(t) < uN^{-1/2}, \quad \text{for} \quad |t| \leq 6\,b\sqrt{\ln N/N}.$$

This together with (3.12) implies (3.11) and completes the proof of the theorem.

Lemma 3.1. *Assume that* (1.15) *holds and that* $\mathbf{E}\,\big|f''(X,t_0)\big|^p < \infty$. *Assume that*

$$\delta \le \min\big\{ \big(2^{-1} a H^{-1/p}\big)^{p/q};\ \delta' \big\}.$$

Then there exist constants $c(p)$ *and* $c(p,q)$ *such that*

$$\mathbf{P}\Big\{ \inf\big\{ Q_N''(t):\quad |t-t_0|\le\delta \big\} < 0 \Big\} \le \frac{c(p)\,\varkappa}{N^{\gamma}\,a^p} + \frac{c(p,q)\,H^{1/p}\,(1+2\delta)}{N^{\gamma/p} a},$$

where γ *and* $\varkappa$ *are as in Corollary* 1.4.

Proof of Lemma 3.1. Denote

$$Z(t) = f''(X,t) \qquad \text{and} \qquad Z_i(t) = f''(X_i,t), \qquad i = 1,2,\dots,N.$$

By (1.15) and Hölder's inequality,

$$\mathbf{E}\,\big|Z(t) - Z(t_0)\big| \le \big(H|t-t_0|^q\big)^{1/p}.$$

Hence, for $|t-t_0|\le\delta$,

$$\mathbf{E}\,Z(t) \ge \mathbf{E}\,Z(t_0) - \big(H|t-t_0|^q\big)^{1/p} \ge a - \big(H\delta^q\big)^{1/p} > a/2.$$

We have

$$Q_N''(t) = \mathbf{E}\,Z(t) - N^{-1}\sum_{i=1}^N \big(\mathbf{E}\,Z_i(t) - Z_i(t)\big) > 0$$

if

$$M := \sup_{|t-t_0|\le\delta} S_N(t) < a/2, \tag{3.13}$$

where we denote

$$S_N(t) = N^{-1}\sum_{i=1}^N \big(\mathbf{E}\,Z_i(t) - Z_i(t)\big).$$

Further, $M \le M_1 + M_2$, with

$$M_1 = S_N(0), \quad \text{and} \quad M_2 = \sup\big\{ S_N(t) - S_N(s):\quad t_0-\delta\le s,t\le t_0+\delta \big\}.$$

By Chebyshev's inequality,

$$\mathbf{P}\{M\ge a/2\} \le \mathbf{P}\{M_1\ge a/4\} + P\{M_2\ge a/4\} \le 4^p a^{-p}\mathbf{E}\,M_1^p + 4a^{-1}\mathbf{E}\,M_2. \tag{3.14}$$

Using well known inequalities, see e.g. Petrov (1975), for p-th moments of sums of centered random variables we obtain

$$\mathbf{E}\, M_1^p \le c(p) N^{1-p} \mathbf{E}\, \big| f''(X,0) \big|^p = c(p) N^{1-p} \varkappa, \qquad \text{for} \quad 1 < p \le 2,$$

and

$$\mathbf{E}\, M_1^p \le c(p) N^{-p/2} \mathbf{E}\, \big| f''(X,0) \big|^p = c(p) N^{-p/2} \varkappa, \qquad \text{for} \quad p > 2.$$

Thus

$$\mathbf{E}\, M_1^p \le c(p) N^{-\gamma} \varkappa, \qquad\qquad p > 1. \tag{3.15}$$

Similarly,

$$\mathbf{E}\, \big| S_N(t) - S_N(s) \big|^p \le c(p) N^{-\gamma} \mathbf{E}\, \Big| \big(Z(t) - Z(s) \big) - \mathbf{E}\, \big(Z(t) - Z(s) \big) \Big|^p, \quad p > 1.$$

Hence, by (1.15),

$$\mathbf{E}\, \big| S_N(t) - S_N(s) \big|^p \le c(p) N^{-\gamma} H |s-t|^q.$$

This estimate yields

$$\mathbf{E} \sup_{t_0-\delta \le s,t \le t_0+\delta} \Big(S_N(s) - S_N(t) \Big) \le c(p) N^{-\gamma/p} H^{1/p} \int_0^{2\delta} \Big(\frac{2\delta}{u^{p/q}} \Big)^{1/p} du,$$

see, e.g., Theorem 1.2 in Talagrand (1990). We have

$$\mathbf{E}\, M_2 \le c(p,q) N^{-\gamma/p} H^{1/p} (1 + 2\delta).$$

Combining this inequality with (3.14) and (3.15) we see that the probability of (3.13) is larger than

$$1 - c(p) N^{-\gamma} a^{-p} \varkappa + c(p,q) N^{-\gamma/p} a^{-1} H^{1/p} (1 + 2\delta),$$

thereby proving the lemma.

4. Appendix

In the appendix we outline proofs of the propositions and examples of section 2 omitting tedious computations. Detailed proofs are contained in an extended version of the paper which appears as a Preprint of SFB 343, 1995, Universität Bielefeld.

The propositions are corollaries of the general result provided as Theorem 1.2. Therefore we check the conditions of this theorem.

Proof of Proposition 2.1. It suffices to verify conditions of Theorem 1.2. Let us check condition (1.2).

For $t \in (t_0 - \delta, t_0 + \delta)$, we have

$$L_1(t) = \mathbf{P}\{X < t\} - \mathbf{P}\{X > t\}, \qquad L_1'(t) = 2p(t), \qquad L_2(t) = 1.$$

Hence $a = 2p(t_0)$ and $\sigma^2 = L_2(t_0) = 1$. Since $L_1(t_0) = 0$, the mean value theorem yields

$$L_1(t) = L_1(t) - L_1(t_0) = L_1'\big(t_0 + \tau(t - t_0)\big)(t - t_0), \qquad 0 \le \tau \le 1.$$

Hence, the inequality

$$\big|L_1(t) - a(t - t_0)\big| \le 2H|t - t_0|^{1+\alpha}, \qquad |t - t_0| < \delta,$$

is a consequence of

$$\big|L_1'(t') - L_1'(t_0)\big| < 2H|t' - t_0|^{\alpha}, \qquad |t' - t_0| < \delta.$$

But this is exactly (2.1). Therefore, condition (1.2) is satisfied with the exponent α and with $\beta = 1$, $C = 2H$ and $D = 0$.

It remains to verify condition (1.5). Let $X^{(1)} \le X^{(2)} \le \dots \le X^{(N)}$ denote the order statistic of the sample $X_1, X_2, \dots, X_N$. It is easy to see that if

$$X^{([N/2]-1)} < X^{([N/2])} < X^{([N/2]+1)} < X^{([N/2]+2)} \tag{4.1}$$

then $t_N = X^{([N/2]+1)}$ if N is odd and $X^{([N/2])} \le t_N \le X^{([N/2]+1)}$ if N is even. Using Bernstein's inequality, see e.g. Petrov (1975), one may show that

$$\mathbf{P}\Big\{\omega : \quad t_0 - \delta < X^{([N/2]-1)} \le X^{([N/2]+2)} < t_0 + \delta\Big\} > 1 - N^{-1} \tag{4.2}$$

provided that

$$\frac{N}{\ln N} \ge c\,\mathbf{P}^{-2}\{t_0 - \delta < X < t_0\} + c\,\mathbf{P}^{-2}\{t_0 < X < t_0 + \delta\}. \tag{4.3}$$

Since F is absolutely continuous on $(t_0 - \delta, t_0 + \delta)$, the bounds (4.2) and (4.1) together imply (1.5) with a constant $A = N^{-1/2}$ provided that (4.3) holds. An application of Theorem 1.2 completes the proof.

Proof of the statements of Example 2.2. It is easy to see that $t_0 = 0$, $a = (\alpha + 1)/\alpha$, $\sigma^2 = 1$ and that conditions of Theorem 1.2 are satisfied with

$$\delta = 1, \quad b = \alpha/(1 + \alpha), \quad \overline{\beta} = 1, \quad A = B = D = 0 \quad \text{and} \quad C = (\alpha + 1)/\alpha.$$

Let $u = b^{-1}$, and write $t = N^{-1/2}ub = N^{-1/2}$. We shall show that $\liminf_N N^{\alpha/2}\,\Delta_N(u) > 0$, where $\Delta_N(u) = \bigl|\mathbf{P}\bigl\{b^{-1}\,N^{1/2}t_N < u\bigr\} - \Phi(u)\bigr|$.

It follows from (3.4), (3.6) and (3.8) that

$$\begin{aligned}\Delta_N(u) &\geq \Bigl|\Phi\Bigl(N^{1/2}\,m_t/\sigma_t\Bigr) - \Phi(u)\Bigr| - cN^{-1/2}\bigl(1 + A + B + \overline{\beta}(t)/\sigma^3\bigr) \\ &\geq \Bigl|\Phi\Bigl(N^{1/2}\,m_t/\sigma_t\Bigr) - \Phi(u)\Bigr| - cN^{-1/2}. \qquad (4.4)\end{aligned}$$

Here $m_t = L_1(t)$ and $\sigma_t = L_2(t) - L_1^2(t)$. Observe that in our case $\sigma_t = 1$. A simple calculation yields

$$m_t = \mathbf{P}\{X < t\} - \mathbf{P}\{X > t\} = uN^{-1/2} - \alpha^{-1}t^{1+\alpha}.$$

Hence

$$\Bigl|N^{1/2}m_t/\sigma_t - u\Bigr| = \alpha^{-1}N^{-\alpha/2}.$$

An application of the mean value theorem implies

$$\bigl|\Phi\bigl(N^{1/2}m_t/\sigma_t\bigr) - \Phi(u)\bigr| \geq c(\alpha)N^{-\alpha/2},$$

with some $c(\alpha) > 0$. This inequality combined with (4.4) completes the proof.

Proof of Proposition 2.3. It suffices to check the conditions of Theorem 1.2. Observe that the condition (2.3) ensures that $a = p(p-1)\mathbf{E}\,|X - t_0|^{p-2} < \infty$. Calculations show that, for small h (one may take, e.g., $|h| < \delta/20$),

$$\bigl|L_1(t_0 + h) - ah\bigr| < c(p,\beta,F)\,|h|^{\beta+p-1}$$

and

$$\bigl|L_2(t_0 + h) - L_2(t_0)\bigr| \leq c(p,\beta,F)\,|h|. \qquad (4.5)$$

Hence the first inequality of (1.2) is satisfied with an exponent $\alpha = \beta + p - 2$ and the second one with an exponent 1.

It remains to check condition (1.5). Since the function $t \to |x - t|^p$ is strictly convex, we have

$$\mathbf{P}\bigl\{t_N \leq t_0 - \delta'\bigr\} \leq \mathbf{P}\bigl\{0 \leq Q_N'(t_0 - \delta')\bigr\}, \qquad \mathbf{P}\bigl\{t_N \geq t_0 + \delta'\bigr\} \leq \mathbf{P}\bigl\{0 \geq Q_N'(t_0 + \delta')\bigr\}.$$

The random variable $Q_N(t)$ is the sum $\sum_{i=1}^N Y_i(t)$ of real i.i.d. random variables $Y_i(t) = f'(X_i, t)$. An application of Chebyshev's inequality yields

$$\begin{aligned}\mathbf{P}\bigl\{0 \geq Q_N'(t_0 + \delta')\bigr\} &= \mathbf{P}\Bigl\{-L_1(t_0 + \delta') \geq Q_N'(t_0 + \delta') - L_1(t_0 + \delta')\Bigr\} \\ &\leq \frac{\mathbf{E}\;\; Q_N'(t_0 + \delta') - L_1(t_0 + \delta')\;\;^2}{L_1^2(t_0 + \delta')} \leq 2N^{-1}\,\frac{L_2(t_0 + \delta')}{L_1^2(t_0 + \delta')}\,.\end{aligned}$$

Similarly

$$\mathbf{P}\{0 \le Q'_N(t_0 - \delta')\} \le 2N^{-1} \frac{L_2(t_0-\delta')}{L_1^2(t_0-\delta')}\,.$$

Observe that

$$\mathbf{E}\, Y_1(t_0 - \delta') = L_1(t_0 - \delta') < 0 = L_1(t_0) < L_1(t_0 + \delta') = \mathbf{E}\, Y_1(t_0 + \delta'), \qquad \delta' > 0,$$

since the function L_1 is strictly increasing.

We obtain

$$\mathbf{P}\{|t_N - t_0| \ge \delta'\} \le \frac{2\, L_2(t_0-\delta')}{N\, L_1^2(t_0-\delta')} + \frac{2\, L_2(t_0+\delta')}{N\, L_1^2(t_0+\delta')} \le \frac{c(f, F, \delta')}{N}, \quad \delta' > 0. \tag{4.6}$$

The choice $\delta' = \delta/20$ and an application of Theorem 1.2 completes the proof.

Proof of Proposition 2.4. To prove Proposition 2.4 it suffices to check the conditions of Theorem 1.2. As in the proof of Proposition 2.3 the estimate (4.6) implies (1.5). It remains to verify condition (1.2).

Since F has a bounded density p, for $|t - t_0| < \delta$, the function L_1 is differentiable and $L_1'(t) = p(p-1)\mathbf{E}\,|X - t|^{p-2}$, for $|t - t_0| < \delta$. Calculations show that, for small h (one may take $|h| < \delta/20$), the estimate (4.5) holds and

$$\big|L_1'(t_0 + h) - L_1'(t_0)\big| \le c(p, F)|h|^{p-1+\beta}. \tag{4.7}$$

Since $L_1(t_0) = 0$, this inequality yields

$$\big|L_1(t_0 + h) - ah\big| \le c(p, F)|h|^{p+\beta}.$$

Hence the first inequality of (1.2) is satisfied with the exponent $\alpha = p + \beta - 1$ and the second one with the exponent 1. This proves the proposition.

Proof of Proposition 2.5. The proof proceeds along the lines of the proof of Proposition 2.4. Minor differences appear in verifying condition (4.7) while estimating the integral

$$\int_{c|h|\le|x|\le c\delta} \big|x - t_0\big|^{p-3}\big|p(x) - p(t_0)\big|dx \le c|h|^{p-2+\beta}dx, \qquad \beta < 2 - p,$$

by means of (2.4). Such bound fails to hold, for $\beta = 2 - p$. Hence in addition to the condition (2.5) we need to impose an integrability condition. This completes the proof.

Proof of (2.6). We shall show that

$$\liminf_N\, N^{1/2}\, \ln^{-1} N\, \Delta_N(u) > 0, \qquad \text{with} \quad u = b^{-1}.$$

Assume without loss of generality that $t_0 = 0$. As in proof of Proposition 2.3 the estimate (4.6) implies (1.5) (with some A independent of N). Furthermore, $\mathbf{E}\,|X|^{3(p-1)} < \infty$ yields $\overline{\beta} < \infty$. Therefore, as in proof of the statements of Example 2.2, it suffices to show that, for some $c(f,F) > 0$,

$$\Big|\Phi\big(N^{1/2}\, m_t/\sigma_t\big) - \Phi(u)\Big| \ge c(f,F) N^{-1/2} \ln N, \qquad \text{for large} \quad N. \tag{4.8}$$

Here $m_t = L_1(t)$, $\sigma_t^2 = L_2(t) - L_1^2(t)$ and $t = N^{-1/2} u\, b = N^{-1/2}$.

Let us prove (4.8). Calculations that, for small h, the inequality (4.5) holds and

$$\big|L_1(h) - ah - hr(h)\big| < c(f,F)|h|^2, \tag{4.9}$$

where the function r is such that

$$\big|r(h)\big| \ge c(f,F)|h| \cdot \big|\ln|h|\big|. \tag{4.10}$$

Take $h = t$. It follows from (3.7), (4.5) and (4.9) that, for large N,

$$\begin{aligned}\Big|\ \frac{\sqrt{N}\, m_t}{\sigma_t} - u - \frac{r(t)}{\sigma}\ \Big| &\le \sqrt{N}\sigma^{-1}|L_1(t)| \cdot \big|\sigma/\sigma_t - 1\big| + \sqrt{N}\sigma^{-1}\Big|L_1(t) - a\,t - t\,r(t)\Big| \\ &\le c(p,F) N^{-1/2}. \end{aligned} \tag{4.11}$$

This bound together with (4.10) gives

$$\Big|\sqrt{N}\, m_t/\sigma_t - u\Big| > c(f,F) N^{-1/2} \ln N,$$

for sufficiently large N. The last inequality combined with the Lagrange mean value theorem implies (4.8) and completes the proof.

Proof of Proposition 2.6. It suffices to verify the conditions of Theorem 1.2. Let us consider condition (1.2).

Since $\mathbf{P}\{X = t_0 - k\} = \mathbf{P}\{X = t_0 + k\} = 0$, it is easy to show that the function L_1 is differentiable at t_0 and

$$\begin{aligned}&\big|L_1(t_0 + h) - \mathbf{P}\big\{|X - t_0| < k\big\} \cdot h\big| \\ &\quad \le |h|\Big(\big|F(t_0 + k + h) - F(t_0 + k)\big| + \big|F(t_0 - k + h) - F(t_0 - k)\big|\Big), \qquad |h| < \delta.\end{aligned}$$

Invoking (2.9) we obtain

$$\big|L_1(t) - L_1(t_0) - a(t - t_0)\big| \le 2H|t - t_0|^{1+\alpha}, \qquad |t - t_0| < \delta. \tag{4.12}$$

Furthermore, since

$$\Big|\big(\rho_k'(x-t)\big)^2-\big(\rho_k'(x-s)\big)^2\Big|\le 2k|t-s|, \qquad x,s,t\in R,$$

we obtain

$$\big|L_2(t)-L_2(t_0)\big|\le 2k|t-t_0|.$$

This inequality together with (4.12) yields (1.2) with an exponent α, with $\beta=1$, $C=2H$ and $D=2k$.

Let us verify the consistency condition (1.5). Inequality (2.8) implies that the function L_1 is strictly increasing in a neighborhood of t_0, i.e., that, for any $\delta'>0$,

$$L_1(t_0-\delta')<L_1(t_0)=0<L_1(t_0+\delta').$$

An inspection of the proof of the estimate (4.6) shows that (4.6) remains valid (with, e.g. $\delta'=\delta/2$) in this case as well. Hence condition (1.5) holds with $A=N^{-1/2}c(f,F)$. Observe also that $\sigma^2\le k^2$ and that $\overline{\beta}<k^3$. An application of Theorem 1.2 completes the proof.

Proof of Proposition 2.7. The proposition follows immediately from Corollary 1.3.

Proof of the statements of Example 2.8. We shall prove that

$$\liminf_N \Delta_N(u)N^{(p-2)/2}>0, \qquad \text{with} \quad u=b^{-1}.$$

It was shown in the proof of Corollary 1.3 that the consistency condition (1.5) is satisfied with $A=N^{-1/2}c(f,F)$. Furthermore, it is easy to see that in our case $t_0=0$ and $\overline{\beta}<\infty$. Therefore, see the proof of (2.6), the problem reduces to showing that

$$\Big|\Phi\big(N^{1/2}\,m_t/\sigma_t\big)-\Phi(u)\Big|\ge c(p)N^{-(p-2)/2} \qquad \text{as} \qquad N\to\infty$$

with some $c(p)>0$. Here t, m_t and σ_t are as in proof of (2.6).

Obviously, $\big|L_2(t)-L_2(0)\big|<c(p)|t|$. Furthermore, simple calculations show that

$$\big|L_1(t)-at-tr(t)\big|\le c(p)t^2,$$

where the function $r(t)=p\,2^{-1}|t|^{p-2}\mathrm{sign}(t)$ satisfies

$$r(t)=r\big(N^{-1/2}\big)\ge c(p)N^{-(p-2)/2}.$$

The rest of the proof is similar to the proof of (2.6), and we omit it. The proof is complete.

Proofs of Propositions 2.9 *and* 2.10. The proofs go along the lines of the proof of Proposition 2.4. and resp. 2.5.

References

Arcones, M. A., *Distributional convergence of M-estimators under unusual rates*, Stat. and Probab. Letters **21** (1994), 271-280.

Bentkus V., Götze F., Paulauskas V. and Račkauskas A., *The accuracy of Gaussian approximation in Banach spaces*, Preprint 90 -100, SFB 343, Universität Bielefeld (In Russian: Itogi Nauki i Tehniki, ser. Sovr. Probl. Matem., Moskva, VINITI **81** (1991), 39–139; to appear in English in Encyclopedia of Mathematical Sciences, Springer), 1990.

Bentkus. V., Götze, F. and van Zwet, W., *An Edgeworth expansion for symmetric Statistics* (1994), Preprint 94 - 020, SFB 343, Universität Bielefeld.

Bickel, P.J., *Edgeworth expansions in non parametric statistics*, Ann. Statist. **2** (1974), 1–20.

Bolthausen E. and Götze F., *The rate of convergence for multivariate sampling statistics*, Ann. Statist. **21** (1993), 1692–1710.

Chibisov, D. M., *On the normal approximation for a certain class of statistics*, Sixth Berkeley Symposium on Math. Statist. and Probab. ((L. LeCam and J. Neyman), eds.), vol. 1, University of California Press, Berkeley, 1972, pp. 153–174.

Esseen, C.G., *Fourier analysis of distribution functions*, Acta Math. **77** (1945), 1–125.

Friedrich, K. O., *A Berry–Esseen bound for functions of independent random variables*, Ann. Statist. **17** (1989), 170–183.

Hoffmann-Jorgensen, *Probability with a view toward statistics*, vol. II, Chapman & Hall, New–York, 1994.

Huber, P.J., *Robust estimation of a location parameter*, Ann. Math. Statist. **35** (1964), 73–101.

Huber, P. J., *The behavior of maximum likelihood estimates under nonstandard conditions.*, Fifth Berkeley Symposium on Math. Statist. and Probab. ((L. LeCam and J. Neyman), eds.), vol. 1, University of California Press, Berkeley, 1967, pp. 221–233.

Linnik, Yu. V. and Mitrofanova, N. M., *On the asymptotic distribution of maximum likelihood estimates (in Russian)*, Doklady Akademii Nauk SSSR **149** (1963), 518–520.

Matsuda, T., *The Berry–Esseen bound for maximum likelihood estimates of translation parameter of truncated distribution*, Osaka Math. J. **20** (1983), 185-195.

Michel, R. and Pfanzagl, J, *The accuracy of the normal approximation for minimum contrast estimates*, Z. Wahrsch. verw. Gebiete **18** (1971), 73–84.

Paulauskas, V., *Rates of convergence in the asymptotic normality for some local maximum estimators. II.* (1995), Preprint 95-5, Dept. of Mathematics, Vilnius University.

Petrov, V.V., *Sums of independent random variables*, Springer, New York (1975).

Pfanzagl, J., *The Berry-Esseen bound for minimum contrast estimates*, Metrika **17** (1971), 82–91.

Pollard, D., *New ways to prove central limit theorems*, Econometric Theory **1** (1985), 295–313.

Pollard, D., *Convergence of Stochastic Processes*, Springer, New–York, 1984.

Radavičius, M., *Rate of convergence of maximal likelihood density estimate to a normal law*, Lithuanian Math. J. **30** (1990), 54-60.

Reiss, R.-D., *On the accuracy of the normal approximation for quantiles*, Ann. Probab. **2** (1974), 741–744.

Serfling R.J., *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.

Schwartz, L., *Analyse Mathématique.*, vol. 1, Herman, Paris, 1967.

Talagrand, M., *Sample boundedness of stochastic processes under increment conditions*, Ann. Probab. **18** (1990), 1–149.

van Zwet, W.R., *A Berry-Esseen bound for symmetric statistics*, Z. Wahrsch. verw. Gebiete **66** (1984), 425–440.