Rostock. Math. Kolloq. **57**, 3–51 (2003)

Subject Classification (AMS)
**62J02,62F12**

Fiedrich Liese, Igor Vajda

# On $\sqrt{n}$–Consistency and Asymptotic Normality of Consistent Estimators in Models with Independent Observations[1]

ABSTRACT. The paper presents relatively simple verifiable conditions for $\sqrt{n}$-consistency and asymptotic normality of M-estimators of vector parameters in a wide class of statistical models. The conditions are established for the $M$-estimators with absolutely continuous $\rho$-function of locally bounded variation, and for the class of models including e.g. the linear and the nonlinear regression, the generalized linear models and the proportional hazards models as special cases. The conditions are verified on $L_1$ and $L_2$ estimators embedded into a continuum of their alternative versions, as well as on one new class of M-estimators of parameters of exponential families which are shown to be robust in the sense of bounded gross-error sensitivity. Comparisons with known conditions for special models indicate that the present general conditions are not too restrictive in special situations and that sometimes they are even weaker than the previously published special conditions.

## 1 Introduction and basic concepts

We consider a general parametric statistical model with independent observations. In other words, for every $n \in \mathbb{N}$ we consider a random sample $\mathbf{Y}_n = (Y_1, \ldots, Y_n)'$ of independent real valued observations,

$$\mathbf{Y}_n \sim G(y_1, \ldots, y_n) = \prod_{i=1}^{n} G(y_i|i, \theta_0), \tag{1.1}$$

where $\theta_0$ is a true value of a parameter $\theta = (\theta_1, \ldots, \theta_m)' \in \Theta$ for open $\Theta \subset \mathbb{R}^m$ and

$$\mathcal{G}_1 = \{G(y|1, \theta) : \theta \in \Theta\}, \ldots, \mathcal{G}_n = \{G(y|n, \theta) : \theta \in \Theta\} \tag{1.2}$$

are given families of distribution functions (briefly *distributions*) possibly depending on the sample size $n$. This means that we admit the triangular observation schemes $(Y_1, \ldots, Y_n) = (Y_1^{(n)}, \ldots, Y_n^{(n)})$. Important particular versions of this model are discussed in Section 2.

[1]Supported by the grant A 1075101.

We study a general $M$-estimator of the unknown true parameter $\theta_0$ in the above considered model. This estimator is defined as a sequence of $\Theta$-valued measurable functions $\hat{\theta}_n = \hat{\theta}_n(\mathbf{Y}_n)$ minimizing on $\Theta$ the random functions

$$M_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \rho(Y_i - \varphi_i(\theta)), \tag{1.3}$$

where $\rho : \mathbb{R} \mapsto \mathbb{R}$ is a given function called *criterion function* and $\varphi_1 : \Theta \mapsto \mathbb{R}, \ldots, \varphi_n : \Theta \mapsto \mathbb{R}$ are given functions called *locators*. The locators may depend on the sample size $n$, i. e. we admit triangular schemes of locators

$$(\varphi_1, \ldots, \varphi_n) = (\varphi_1^{(n)}, \ldots, \varphi_n^{(n)}). \tag{1.4}$$

Since the $M$-estimator under consideration is defined by the criterion function and locators, we use the symbols

$$\hat{\theta}_n \sim \langle \rho; \varphi_1, \ldots, \varphi_n \rangle \qquad \text{or briefly} \;\; \hat{\theta}_n \sim \langle \rho; \varphi_i \rangle. \tag{1.5}$$

We are interested in the asymptotic properties of $M$-estimators $\hat{\theta}_n \sim \langle \rho; \varphi_i \rangle$ when the sample size $n$ tends to infinity. Therefore, unless otherwise explicitly stated, all asymptotic relations, formulas and properties are automatically considered for $n \to \infty$.

Our attention is restricted to the $M$-estimators $\hat{\theta}_n \sim \langle \rho; \varphi_i \rangle$ with criterion functions $\rho$ absolutely continuous on bounded intervals of $\mathbb{R}$ (briefly, absolutely continuous on $\mathbb{R}$). This means that there exists a measurable function $\psi : \mathbb{R} \mapsto \mathbb{R}$ satisfying the condition

$$\psi(y) = \frac{d\rho(y)}{dy} \quad \text{a.e.} \tag{1.6}$$

with respect to the Lebesgue measure on $\mathbb{R}$ and absolutely integrable on bounded intervals. We shall consider a right-continuous extension of $\psi$ on $\mathbb{R}$ which is (up to a constant $\rho(0)$ playing no role in the definition of $M$-estimator $\hat{\theta}_n$ (cf. (1.3)) one-one related to $\rho$ and satisfies for all $a$, $b \in \mathbb{R}$ the relation

$$\rho(b) - \rho(a) = \int_{(a,b]} \psi(y)dy \tag{1.7}$$

(the so-called fundamental theorem of calculus for Lebesgue integrals, cf. Theorem 18.16 in Hewitt and Stromberg [9]). Here, and in the sequel,

$$\int_{(a,b]} = -\int_{(b,a]} \quad \text{if } b < a. \tag{1.8}$$

The right-continuous function $\psi : \mathbb{R} \mapsto \mathbb{R}$ characterizes a *sensitivity* of the $M$-estimator $\hat{\theta}_n \sim \langle \rho; \varphi_i \rangle$ to small deviations of observations $Y_1, \ldots, Y_n$ (an appropriately normed version of $\psi$ is an *influence function* of the $M$-estimator, see Huber [12] or Hampel et al [8]). Due to the one-one relation between the criterion function $\rho$ and the sensitivity function $\psi$ mentioned above, we can replace the representation of $M$-estimators (1.5) by

$$\hat{\theta}_n \sim \langle \psi; \varphi_1, \ldots, \varphi_n \rangle \quad \text{or, briefly} \quad \hat{\theta}_n \sim \langle \psi; \varphi_i \rangle. \tag{1.9}$$

Our theory is restricted to the estimators $\hat{\theta}_n \sim \langle \psi; \varphi_i \rangle$ with sensitivity $\psi$ of a locally bounded variation. This means that $\psi$ is a difference of two nondecreasing functions $\psi^+$ and $\psi^-$ which are assumed to be continuous from the right. This theory presents conditions for $\sqrt{n}-$ consistency and asymptotic normality of $\hat{\theta}_n$ in terms of the sum $\psi^{\pm} = \psi^+ + \psi^-$.

As is indicated by the title of the paper, our main results are restricted to the $M$-estimators $\hat{\theta}_n \sim \langle \psi; \varphi_i \rangle$ which are consistent in the standard sense

$$\hat{\theta}_n \xrightarrow{P} \theta_0 \, . \tag{1.10}$$

We present conditions on the sensitivity function $\psi$, locators $\varphi_i$ and the model (1.1) under which $\hat{\theta}_n$ is $\sqrt{n}-$consistent in the sense

$$\lim_{y\to\infty} \lim_{n\to\infty} \mathsf{P}\left(\sqrt{n}\left\|\hat{\theta}_n - \theta_0\right\| > y\right) = 0 \tag{1.11}$$

and asymptotically normal in the sense

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N(0, V) \tag{1.12}$$

and under which the variance-covariance $m \times m$ matrix $V$ can be explicitly evaluated.

These main results are presented in the next Section 2. The conditions on the sensitivity function $\psi$, locators $\varphi_i$ and the model (1.1) are formulated as regularity conditions (R1)–(R4+). Important particular versions of the general model (1.1) and sufficient conditions for (R1)–(R4+) are in Section 3.

The consistency (1.10) in reasonably general classes of $M$-estimators (1.9) and models (1.1) is a difficult problem. Sufficient conditions have been established e.g. in Yohai and Maronna [31], Zhao and Chen [32], Hjort and Pollard [10], Liese and Vajda [18]-[21], Zhao [33], Arcones [1]-[2] and some other references therein. Presentation of such conditions would increase the complexity and size of the paper above bearable bounds. Therefore we refer in this respect to the mentioned literature and restrict ourselves to the verification of consistency only in special cases illustrating applicability of the main result of Section 2.

In Sections 4 and 5 we illustrate the applicability of the general results of Sections 2 and 3 to special classes of $M$-estimators (1.9) and models (1.1). Particular attention is payed to the class of $M$-estimators with the criterion functions

$$\rho(y) = \rho_\beta(y) = \beta\, y\, I_{[0,\infty)}(y) - (1-\beta)\, y\, I_{(-\infty,0)}(y), \quad 0 < \beta < 1, \tag{1.13}$$

introduced by Koenker and Basset [16] and later used by many authors (e. g. Portnoy [24], Koul and Saleh [17], Jurečková and Sen [14], Hallin and Jurečková [7]).

In Section 6 are proofs of main results of Section 2. The proofs employ some general results and techniques of van der Vaart and Wellner [30], in particular their Theorems 3.2.2 and 3.2.5. The proofs use also the methods developed in [21].

The present paper differs from[21] in a considerably simpler formulations and proofs of results, and in application of these results to different special models (1.1) and/or estimators (1.9). It also differs from the classical literature studying the consistency (1.10) and the asymptotic normality (1.12) of the estimators $\hat{\theta}_n$ defined as solutions of the equations

$$\sum_{i=1}^{n} \psi(Y_i - \varphi_i(\theta))\, \nabla\varphi_i(\theta) = 0 \tag{1.14}$$

on $\Theta$ when the locators $\varphi_i(\theta)$ are differentiable on $\Theta$ with gradients $\nabla\varphi_i(\theta)$ (see the monographs of Serfling [27], [12], Singer and Sen [28], [14], and references therein). Obviously, our $M$-estimators $\hat{\theta}_n \sim \langle\psi;\varphi_i\rangle$ coincide with solutions of (1.14) only in special cases, e. g. if the sensitivity $\psi$ is monotone on $\mathbb{R}$ (i. e. the criterion function $\rho$ is convex) and the locators $\varphi_i(\theta)$ are linear in $\theta$. This takes place e. g. if $\theta \in \Theta = \mathbb{R}$ is the location parameter, $\varphi_i(\theta) = \theta$ and

$$\rho(y) = \begin{cases} y^2 & \text{for} \quad |y| \le k, \\ 2k|y| - k^2 & \text{for} \quad |y| > k, \end{cases}$$

which is the situation studied by Huber [11]. The results about asymptotic normality of solutions $\hat{\theta}_n$ of (1.14), based on the ideas and techniques of [11, 12], are thus disjoint with our results except the relatively rare situations when solutions of (1.14) minimize the function $M_n(\theta)$ of (1.3). Such situations are trivial from the point of view of our theory which primarily intends to bring results about $M$-estimators $\hat{\theta}_n \sim \langle\rho;\varphi_i\rangle$ where either $\rho(y)$ is not convex in $y \in \mathbb{R}$ or $\varphi_i(\theta)$ are not linear in $\theta \in \Theta$, i. e. about situations not covered by the classical Huber-type theories.

## 2 Main results

In this section we consider an arbitrary model (1.1) and an arbitrary $M$-estimator $\hat{\theta}_n \sim \langle\psi;\varphi_i\rangle$ (equivalently, $\widehat{\theta}_n \sim \langle\rho;\varphi_i\rangle$, see (1.5) and (1.9)) with the variation of $\psi$ *locally bounded*, i. e. bounded on bounded intervals of $\mathbb{R}$. This means that there exist nondecreasing functions $\psi^+$, $\psi^- : \mathbb{R} \mapsto \mathbb{R}$ with the property

$$\psi = \psi^+ - \psi^-. \tag{2.1}$$

We define on $\mathbb{R}$ the nondecreasing function

$$\psi^\pm = \psi^+ + \psi^-. \tag{2.2}$$

**Definition 2.1** *We say that the locators $\varphi_i$ are adapted to the model if*

$$\mathsf{E}\psi(Y_i - \varphi_i(\theta_0)) = 0, \quad i \in \mathbb{N}. \tag{2.3}$$

*We say that the estimator $\hat{\theta}_n$ is adapted to the model if the locators are adapted in the sense of* (2.3) *and the estimator is consistent in the sense of* (1.10).

In the rest of paper we consider the following conditions of regularity of the estimator $\hat{\theta}_n$ in the model (1.1).

(R1) The second moments (variances if (2.3) holds)

$$\sigma_i^2 = \mathsf{E}\left[\psi(Y_i - \varphi_i(\theta_0))\right]^2 \tag{2.4}$$

are uniformly bounded in the mean, i. e.,

$$\sup_{n\in\mathbb{N}} \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 < \infty. \tag{2.5}$$

(R2) The gradients

$$\dot{\varphi}_i(\theta) = \left(\frac{\partial}{\partial\theta_1}, \ldots, \frac{\partial}{\partial\theta_m}\right)' \varphi_i(\theta), \quad \theta \in \Theta,\ i \in \mathbb{N} \tag{2.6}$$

exist and are locally bounded and locally Lipschitz in the sense that one can find a closed ball

$$B = B_\delta(\theta_0) = \{\mathbf{y} \in \mathbb{R}^m : \|\mathbf{y} - \theta_0\| \le \delta\} \tag{2.7}$$

and a constant $\lambda > 0$ possibly depending on $B$, such that $B \subset \Theta$ and

$$\|\dot{\varphi}_i(\theta)\| \le \lambda, \quad \theta \in B,\ i \in \mathbb{N} \tag{2.8}$$

$$\|\dot{\varphi}_i(\theta) - \dot{\varphi}_i(\tilde{\theta})\| \le \lambda\|\theta - \widetilde{\theta}\|, \quad \theta, \widetilde{\theta} \in B,\ i \in \mathbb{N}. \tag{2.9}$$

(R3) There exists $\tau_0 > 0$ such that the functions

$$H_i(t) = \mathsf{E}\,\psi(Y_i - \varphi_i(\theta_0) + t), \quad i \in \mathbb{N} \tag{2.10}$$

are differentiable on the interval $(-\tau_0, \tau_0)$ and the derivatives

$$h_i(t) = \frac{d}{dt} H_i(t), \quad i \in \mathbb{N} \tag{2.11}$$

satisfy the condition

$$\lim_{\tau\downarrow 0} \sup_{n\in\mathbb{N}} \frac{1}{n} \sum_{i=1}^{n} \omega(h_i, \tau) = 0 \tag{2.12}$$

where $\omega(h_i,\tau) = \sup_{|t|\le\tau} |h_i(0) - h_i(t)|$, $0 < \tau < \tau_0$, is the modulus of continuity of $h_i(t)$ in the neighborhood of $t = 0$. Further, the variances $\sigma_i^2$ from (R1), gradients $\dot{\varphi}_i$ from (R2) and functions $h_i$ from (R3) satisfy

$$\Sigma_n = \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2 \dot{\varphi}_i(\theta_0)\, \dot{\varphi}_i(\theta_0)' \to \Sigma, \tag{2.13}$$

$$\Phi_n = \frac{1}{n}\sum_{i=1}^{n} h_i(0) \dot{\varphi}_i(\theta_0)\, \dot{\varphi}_i(\theta_0)' \to \Phi \tag{2.14}$$

where the $m \times m$ matrices $\Sigma$ and $\Phi$ are positive definite.

(R4) There exist constants $\tau_0 > 0$ and $\kappa$ such that the function (2.2) satisfies for all $0 < \tau < \tau_0$ the relation

$$\sup_{n\in\mathbb{N}} \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}\left[\psi^{\pm}(X_i+\tau) - \psi^{\pm}(X_i-\tau)\right]^2 < \kappa \tag{2.15}$$

where $X_i = Y_i - \varphi(\theta_0)$.

(R4+) There exist constants $\tau_0 > 0$ and $q > 0$ and $\kappa$ such that the function (2.2) satisfies for all $0 < \tau < \tau_0$ the relation

$$\sup_{n\in\mathbb{N}} \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}\left[\psi^{\pm}(X_i+\tau) - \psi^{\pm}(X_i-\tau)\right]^2 < \kappa\tau^q \tag{2.16}$$

where $X_i = Y_i - \varphi(\theta_0)$.

Sufficient conditions for (R3), (R4) and (R4+) will be studied in the next section. Here we formulate the main result of the paper. We remind that the asymptotic relations are considered for $n \to \infty$ unless otherwise stated.

**Theorem 2.2** *If the estimator $\widehat{\theta}_n \sim \langle\psi;\varphi_i\rangle$ is adapted to the model (1.1) in the sense of Definition 2.1 and satisfies the regularity conditions (R1) – (R4) then it is $\sqrt{n}$–consistent in the sense of (1.11).*

**Theorem 2.3** *Let the estimator $\widehat{\theta}_n \sim \langle\psi;\varphi_i\rangle$ be adapted to the model (1.1) in the sense of Definition 2.1 and satisfy the regularity conditions (R1) – (R4) and (R4+). If*

$$n^{-1/2}\sum_{i=1}^{n} \psi(Y_i - \varphi_i(\theta_0))\, \dot{\varphi}_i(\theta_0) \xrightarrow{\mathcal{L}} N(0,\Sigma) \tag{2.17}$$

*then the estimator $\hat{\theta}_n$ is asymptotically normal in the sense of (1.12) with the variance-covariance matrix*

$$V = \Phi^{-1}\,\Sigma\,\Phi^{-1}. \tag{2.18}$$

The proofs of Theorem 2.2 and 2.3 are deferred to Section 6. Here we present a sufficient condition for the condition (2.17) of Theorem 2.3.

**Proposition 2.4** *If the assumptions (2.3) and (2.13) hold and for some $\gamma > 0$,*

$$\sup_{i\in\mathbb{N}} \mathsf{E}\|\psi(Y_i - \varphi_i(\theta_0))\,\dot{\varphi}(\theta_0)\|^{2+\gamma} < \infty \tag{2.19}$$

*then the asymptotic normality condition (2.17) holds.*

**Proof:** Clear from the Lyapunov central limit theorem. □

## 3 Results under restricted generality

In this section we restrict in different ways the generality of the model (1.1) and also the generality of the $M$-estimator $\hat{\theta}_n \sim \langle\psi;\varphi_i\rangle$ studied in the previous section. We study sufficient conditions for the assumptions of Theorems 2.2 and 2.3 under this restricted generality.

**Definition 3.1** *The general statistical model with independent observations defined by (1.1) is said to be*

(i) regression model *if there are given sets $\mathcal{X} \subset \mathbb{R}^k$, $T \subset \mathbb{R}$, and a mapping $\phi : \mathcal{X}\times\Theta \mapsto T$, and if for $1 \le i \le n$ are given realizations $\mathbf{x}_i$ of $\mathbf{x} = (x_1, \ldots, x_k)' \in \mathcal{X}$ and families of distributions $\mathcal{F}_i = \{F_i(y|\vartheta) : \vartheta \in T\}$, both possibly depending on $n$, such that*

$$G(y|i,\theta) = F_i(y|\phi(\mathbf{x}_i,\theta)) \quad \text{for } 1 \le i \le n \quad \text{and} \quad \theta \in \Theta; \tag{3.1}$$

(ii) homogeneous regression model *if it satisfies (i) and*

$$\mathcal{F}_i = \mathcal{F} = \{F(y|\vartheta) : \vartheta \in T\} \quad \text{for } 1 \le i \le n \tag{3.2}$$

*where the family of distributions $\mathcal{F}$ depends neither on $i$ nor on $n$;*

(iii) linear regression model *if it satisfies (i), $\mathcal{X}$ belongs to the same Euclidean space $\mathbb{R}^m$ as $\Theta$ and*

$$\phi(\mathbf{x},\theta) = \mathbf{x}'\theta \quad \text{for } \mathbf{x} \in \mathcal{X} \text{ and } \theta \in \Theta; \tag{3.3}$$

(iv) regression model with additive errors *if it satisfies (i) and $\mathcal{F}_i$ are location families not depending on $n$, i. e. if $T = \mathbb{R}$ and*

$$\mathcal{F}_i = \{F_i(y-\vartheta) : \vartheta \in \mathbb{R}\}, \quad 1 \le i \le n, \tag{3.4}$$

*for a sequence of parent distributions $F_1(y)$, $F_2(y), \ldots$ not depending on $n$.*

The combinations of properties (ii) – (iv) of regression models are admitted. In this manner we obtain the following important special cases.

**Example 3.2** Homogeneous regression with additive errors. This means the *standard nonlinear regression* where the observations are defined by formula

$$Y_i = \phi(\mathbf{x}_i, \theta_0) + \mathcal{E}_i, \quad 1 \le i \le n, \tag{3.5}$$

and the additive errors $\mathcal{E}_i$ are i.i.d. by the parent $F$ of the location family $\mathcal{F} = \{F(y-\vartheta) : \vartheta \in \mathbb{R}\}$ satisfying simultaneously the assumptions (3.2) and (3.4).

**Example 3.3** Homogeneous linear regression with additive errors. This means the *standard linear regression* where $\mathcal{X} \subset \mathbb{R}^m$ and

$$Y_i = \mathbf{x}_i'\theta_0 + \mathcal{E}_i, \quad 1 \le i \le n, \tag{3.6}$$

where the additive errors $\mathcal{E}_i$ satisfy the conditions of Example 3.2.

**Example 3.4** The general homogeneous regression leads to independent observations

$$Y_i \sim F(y|\phi(\mathbf{x}_i, \theta_0)), \quad 1 \le i \le n, \tag{3.7}$$

specified by a $k \times n$ matrix

$$\mathbf{X}_n = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \tag{3.8}$$

of regressors and a family of distributions $\mathcal{F} = \{F(y|\vartheta) : \vartheta \in T\}$. If $\mathcal{F}$ is a location family then we obtain the standard nonlinear regression of Example 3.2.

**Example 3.5** The homogeneous linear regression in general differs from the standard linear regression. It has been called *pseudolinear regression* in Liese and Vajda [20]. Here the independent observations

$$Y_i \sim F(y|\mathbf{x}_i'\theta_0), \quad 1 \le i \le n, \tag{3.9}$$

are specified by the matrix (3.8) and by a family of distributions $\mathcal{F} = \{F(y|\vartheta) : \vartheta \in T\}$. If $T = \mathbb{R}$ and $\mathcal{F}$ is a location family then the pseudolinear regression reduces to the standard linear regression of Example 3.3. If $\mathcal{F}$ is an exponential family then the pseudolinear regression model reduces to the *generalized linear model*. As an example of the generalized linear regression we can consider the *Cox model* where $\mathcal{F}$ consists of the exponential distributions $F(y|\vartheta) = 1 - \exp\{\vartheta \ln(1 - F(y))\}$, $\vartheta \in \mathbb{R}$, for a given distribution $F(y) = F(y|1)$ differentiable on the support $(0, \infty)$ (then $\Lambda(y) = -\ln(1 - F(y))$ is a cumulative hazard function).

Next we study the adaptation condition (2.3) in the homogeneous regression models and standard nonlinear regression models introduced above. This condition means in fact that

$$\int \psi(y - \varphi_i(\theta))\, dG(y|i, \theta) = 0 \quad \text{for all} \ \ \theta \in \Theta \text{ and } i \in \mathbb{N}. \tag{3.10}$$

In the homogeneous regression models the adaptation (3.10) reduces to evaluation of solutions $a(\vartheta)$ of the system of equations

$$\int \psi(y-a)\,dF(y|\vartheta)=0, \quad \vartheta\in T, \tag{3.11}$$

in the real variable $a\in\mathbb{R}$. Indeed, by (3.1) and (3.2), (3.10) holds provided

$$\varphi_i(\theta)=a(\phi(\mathbf{x}_i,\theta)) \quad \text{if} \quad \int \psi(y-a(\vartheta))\,dF(y|\vartheta)=0, \quad \vartheta\in T. \tag{3.12}$$

In the standard nonlinear regression of Example 3.2 with an error distribution $F(y)$, the adaptation condition (3.12) further simplifies into

$$\varphi_i(\theta)=\phi(\mathbf{x}_i,\theta)+b(F) \quad \text{if} \quad b(F)=a(0), \quad \text{i.e.} \quad \int \psi(y-b(F))\,dF(y)=0. \tag{3.13}$$

An $M$-estimator $\hat\theta_n\sim\langle\psi;\phi(\mathbf{x}_i,\theta)+c\rangle$ with a fixed $c\in\mathbb{R}$ is in fact adapted to all nonlinear regression models (3.5) with error distributions $F$ restricted by the condition $b(F)=c$. However, this condition may not be easily verifiable for some functions $\psi$. In order to obtain an $M-$estimator adapted to the standard nonlinear regression models (3.5) with an arbitrary error distribution $F$, it suffices to extend the parameter space $\Theta$ into $\Theta^*=\Theta\times\mathbb{R}$ and replace $\varphi_i(\theta)=\phi(\mathbf{x}_i,\theta)+c$ by

$$\varphi_i^*(\theta^*)=\phi(\mathbf{x}_i,\theta)+b \quad \text{for } \theta^*=(\theta,b)\in\Theta^*,$$

i. e. to consider the $M$-estimator

$$\hat\theta_n^*=(\widehat\theta_n,\hat b_n)\sim\langle\psi;\phi(\mathbf{x}_i,\theta)+b\rangle \tag{3.14}$$

of the extended true parameter $\theta_0^*=(\theta_0,b_0)$ where $b_0=b(F)$. The validity of (2.3) for $\hat\theta_n^*$, i. e. the validity of (2.3) with $\varphi_i(\theta_0)$ replaced by $\varphi_i^*(\theta_0^*)=\phi(\mathbf{x}_i,\theta_0)+b_0$ is obvious.

Now we present simple conditions which imply the assumptions (R4) and (R4+) of Theorems 2.2 and 2.3 for particular versions of the $M$-estimators (1.9) and general model (1.1).

**Proposition 3.6** *If both components $\psi^+$ and $\psi^-$ of the decomposition (2.1) are Lipschitz on $\mathbb{R}$ then the $M$-estimator $\hat\theta_n\sim\langle\psi;\varphi_i\rangle$ satisfies the regularity condition (R4+) in the general model (1.1).*

**Proof:** Under the assumptions of this proposition the function $\psi^\pm$ defined in (2.2) satisfies the Lipschitz condition

$$|\psi^\pm(y_1)-\psi^\pm(y_2)|\le C|y_1-y_2|$$

for some constant $C$ and all $y_1$, $y_2\in\mathbb{R}$. Therefore the expression in the brackets of (2.15) is bounded above by $(2\tau)^2$. This means that (2.16) with $\kappa=4C^2$, $q=2$ and arbitrary $\tau>0$ holds for the model (1.1). □

The next result is an alternative to Proposition 3.6. In this result we assume that $\psi$ is absolutely continuous on $\mathbb{R}$. Similarly as in (1.6), this means that there exists a measurable and locally absolutely integrable function $\dot{\psi} : \mathbb{R} \mapsto \mathbb{R}$ satisfying the condition

$$\dot{\psi}(y) = \frac{d\psi(y)}{dy} \quad \text{a.e.} \tag{3.15}$$

with respect to the Lebesgue measure on $\mathbb{R}$. Then, similarly as in (1.7), for every $y \in \mathbb{R}$

$$\psi(y) = \psi(0) + \int_{(0,y]} \dot{\psi}(s) ds \quad (\text{cf. } (1.8))$$

and, moreover,

$$\psi^+(y) = \psi^+(0) + \int_{(0,y]} \dot{\psi}(s)\, I(\dot{\psi}(s) > 0)\, ds$$

and

$$\psi^-(y) = \psi^-(0) - \int_{(0,y]} \dot{\psi}(s)\, I(\dot{\psi}(s) < 0)\, ds$$

for the components of the decomposition (2.1). Therefore (2.2) implies that for every $y \in \mathbb{R}$

$$\psi^\pm(y) = \psi^\pm(0) + \int_{(0,y]} |\dot{\psi}(s)|\, ds \quad (\text{cf. } (1.8)). \tag{3.16}$$

Obviously, if $\dot{\psi}$ is bounded a.e. on $\mathbb{R}$ then it follows from the formulas above that $\psi^+$ and $\psi^-$ are Lipschitz on $\mathbb{R}$ so that Proposition 3.6 is applicable. Therefore the next result is interesting only in situations where $\dot{\psi}$ is unbounded.

**Proposition 3.7** *Let $\psi$ be absolutely continuous on $\mathbb{R}$ with an a. e. derivative $\dot{\psi}$. The $M$-estimator $\hat{\theta}_n \sim \langle \psi; \varphi_i \rangle$ satisfies the regularity condition* (R4+) *in the general model* (1.1) *if one of the following conditions holds:*

(i) *$\dot{\psi}$ is square integrable on $\mathbb{R}$;*

(ii) *for $X_i = Y_i - \varphi_i(\theta_0)$ and some $\varepsilon > 0$*

$$C := \sup_{n \in \mathbb{N}} \frac{1}{n} \sum_{i=1}^{n} \mathsf{E} \sup_{|s| \le \varepsilon} (\dot{\psi}(X_i + s))^2 < \infty; \tag{3.17}$$

(iii) *$\dot{\psi} = \dot{\psi}_1 + \dot{\psi}_2$ where $\dot{\psi}_1$ satisfies (i) and $\dot{\psi}_2$ satisfies (ii).*

**Proof:** By (3.16) and Schwarz' inequality, for every $y \in \mathbb{R}$ and $\tau > 0$

$$\begin{aligned} [\psi^\pm(y+\tau) - \psi^\pm(y-\tau)]^2 &\le \left( \int_{(y-\tau, y+\tau]} |\dot{\psi}(s)| ds \right)^2 \\ &\le 2\tau \int_{(y-\tau, y+\tau]} (\dot{\psi}(s))^2 ds \\ &\le 2\tau \int_{\mathbb{R}} (\dot{\psi}(s))^2 ds =: A_1(\dot{\psi}). \end{aligned} \tag{3.18}$$

Therefore, if $\dot{\psi}$ is square integrable then (2.16) holds for $q = 1$, all $\tau > 0$ and

$$\kappa = 2 \int_{\mathbb{R}} (\dot{\psi}(s))^2 \, ds.$$

Further, by (3.18),

$$\left[\psi^{\pm}(y+\tau) - \psi^{\pm}(y-\tau)\right]^2 \leq 4\tau^2 \sup_{|s|\leq\tau} (\dot{\psi}(y+s))^2 =: A_2(\dot{\psi}).$$

Therefore, if (ii) holds then (2.16) holds for $q = 2$, $\kappa = 4C^2$ and all $0 < \tau \leq \varepsilon$. Finally, from the above inequalities we see that

$$\left[\int_{y-\tau}^{y+\tau} |\dot{\psi}(t)| dt\right]^2 \leq \min(A_1(\dot{\psi}), A_2(\dot{\psi})).$$

From here and

$$\left[\int_{y-\tau}^{y+\tau} |\dot{\psi}_1(t) + \dot{\psi}_2(t)| dt\right]^2 \leq 2\left[\int_{y-\tau}^{y+\tau} |\dot{\psi}_1(t)| dt\right]^2 + 2\left[\int_{y-\tau}^{y+\tau} |\dot{\psi}_2(t)| dt\right]^2$$

we obtain the statement in (iii). □

The following proposition presents similar conditions as Proposition 3.7 for the estimator $\widehat{\theta}_n \sim \langle \psi, \varphi_i \rangle$ for nonexplosive $\psi^{\pm}$.

**Definition 3.8** *We say that a nondecreasing function $\xi : \mathbb{R} \mapsto \mathbb{R}$ is explosive if there exists $\tau > 0$ such that*

$$\sup_{y\in\mathbb{R}} [\xi(y+\tau) - \xi(y-\tau)] = \infty.$$

Thus $\psi^{\pm}$ is nonexplosive if for every $\tau > 0$

$$C(\tau) := \sup_{y\in\mathbb{R}} \left[\psi^{\pm}(y+\tau) - \psi^{\pm}(y-\tau)\right] < \infty. \tag{3.19}$$

Clearly, $C(\tau)$ is nondecreasing in the domain $\tau > 0$ with $C(0) \geq 0$. Nonexplosive $\psi^{\pm}$ satisfies the inequalities

$$\left[\psi^{\pm}(y+\tau) - \psi^{\pm}(y-\tau)\right]^2 \leq C(\tau) \left[\psi^{\pm}(y+\tau) - \psi^{\pm}(y-\tau)\right] \tag{3.20}$$

and

$$\left[\psi^{\pm}(y+\tau) - \psi^{\pm}(y-\tau)\right]^2 \leq (C(\tau))^2. \tag{3.21}$$

**Proposition 3.9** *Every $M-$estimator $\widehat{\theta}_n \sim \langle \psi, \varphi_i \rangle$ with nonexplosive $\psi^{\pm}$ satisfies the regularity assumption* (R4) *in the model* (1.1)*. If there exist constants $\tau_0, q > 0$ and $\kappa$ such that for $X_i = Y_i - \varphi_i(\theta_0)$ and all $0 < \tau < \tau_0$*

$$\sup_{n\in\mathbb{N}} \frac{1}{n} \sum_{i=1}^{n} \mathsf{E}\left[\psi^{\pm}(X_i+\tau) - \psi^{\pm}(X_i-\tau)\right] < \kappa\tau^q \tag{3.22}$$

*then it satisfies also* (R4+)*.*

**Proof:** The first assertion is clear from (2.15) and (3.21). The second assertion follos from (2.16), (3.20) and (3.22). □

**Proposition 3.10** *Let $\psi^{\pm}$ of the estimator of Proposition 3.9 be piecewise constant with finitely many jumps of sizes $\Delta_k > 0$ at points $t_k$, and let for some fixed $\varepsilon > 0$ the neighborhoods $N_k(\tau) = (t_k - \tau, t_k + \tau)$, $0 < \tau < \varepsilon$, be disjoint for different $k$. If the distribution functions $F_i(y)$ of $X_i$ in (3.22) have densities in the union $U(\varepsilon) = \cup_k N_k(\tau)$ and*

$$C := \sup_{y \in \mathbb{N}} \frac{1}{n} \sum_{i=1}^{n} \sup_{y \in U(\varepsilon)} f_i(y) < \infty, \tag{3.23}$$

*then (3.22) holds for $\tau_0 = \varepsilon/2$, $q = 1$, and $\kappa = 4C \sum_k \Delta_k$.*

**Proof:** If $\tau \leq \varepsilon/2$ then

$$\psi^{\pm}(y+\tau) - \psi^{\pm}(y-\tau) = \begin{cases} \Delta_k & \text{if} \quad y \in N_k(\tau) \\ 0 & \text{otherwise} \end{cases}$$

and

$$\int_{N_k(\tau)} dF_i(y) \leq F(t_k + 2\tau) - F(t_k - 2\tau) \leq 4\tau \sup_{y \in U(\varepsilon)} f_i(y).$$

Therefore

$$\mathsf{E}[\psi^{\pm}(X_i + \tau) - \psi^{\pm}(X_i - \tau)] \leq \sum_k \Delta_k \int_{N_k(\tau)} dF_i(y).$$

The desired result follows from here. □

Our last result is concerning estimators $\widehat{\theta}_n \sim \langle \psi, \varphi_i \rangle$ with nonexplosive $\psi$ in the general regression models where $G(y|i,\theta) = F_i(y|\phi(\mathbf{x}_i,\theta))$ and $\varphi_i(\theta) = a(\phi(\mathbf{x}_i,\theta))$, see (3.1) and (3.12). We use the notation

$$\vartheta_i = \phi(\mathbf{x}_i, \theta) \qquad \text{and} \qquad a_i = a(\vartheta_i) \tag{3.24}$$

In this notation the functions $H_i(t)$ of (2.10) are given by the formula

$$H_i(t) = \int \psi(y - a_i + t) dF(y|\vartheta_i), \quad t \in \mathbb{R}, \tag{3.25}$$

in the general regression model (3.1). In the simplified notation

$$F_i(y) = F(y + a_i|\vartheta_i), \qquad F_{i,s}(y) = F(y + a_i - s|\vartheta_i) \tag{3.26}$$

it holds

$$H_i(t) = \int \psi(y) dF_i(y - t), \quad t \in \mathbb{R}, \tag{3.27}$$

so that, for $s \neq 0$,

$$\frac{1}{s}[H_i(t+s) - H_i(t)] = \int \psi(y) d\Phi_{i,s,t}(y) \tag{3.28}$$

where

$$\Phi_{i,s,t}(y) = \frac{F_{i,s}(y-t) - F_i(y-t)}{s}, \quad y \in \mathbb{R}.$$

Let us consider $\psi^{\pm} = \psi^{+} + \psi^{-}$ and suppose that for some $\tau > 0$

$$\psi^{+},\ \psi^{-} \in L_1(F_{i,t}) \quad \text{for all } i \in \mathbb{N} \text{ and all } |t| \leq \tau. \tag{3.29}$$

Here and in the sequel, $L_1(G)$ denotes the Banach space of functions absolutely integrable with respect to the measure defined on $\mathbb{R}$ by a nondecreasing and right continuous function $G : \mathbb{R} \mapsto \mathbb{R}$. We assume nonexplosive $\psi^{\pm}$ defined by the condition (3.19).

**Proposition 3.11** *Let an $M$-estimator $\widehat{\theta}_n \sim \langle \psi, a(\mathbf{x}_i, \theta) \rangle$ with non-explosive $\psi^{\pm}$ be adapted to the general regression model (3.1). Further, let $\psi^{+}$, $\psi^{-} \in L_1(F_{i,t})$ for some $\tau > 0$ and all $|t| \leq \tau$ and $i \in \mathbb{N}$, let all distributions $F_{i,s}$, $i \in \mathbb{N}$, $s \in \mathbb{R}$, be differentiable on $\mathbb{R}$ with derivatives $f_{i,s}$, and put $f_i = f_{i,0}$.*

(I) *If*

$$\sup_{|s| \leq \tau} f_{i,s} \in L_1(\psi^{\pm}) \tag{3.30}$$

*then the convolutions $H_i(t)$ are absolutely continuous on $(-\tau/2, \tau/2)$, with a. e. derivatives*

$$h_i(t) = -\int f_i(y-t) d\psi(y), \quad i \in \mathbb{N}. \tag{3.31}$$

(II) *If $f_i$ are locally Lipschitz in sense that for every $y \in \mathbb{R}$*

$$|f_i(y-t) - f_i(y)| \leq \lambda_i(y)\,|t|, \quad t \in (-\tau, \tau), \tag{3.32}$$

*and both $f_i$ and $\lambda_i$ belong to $L_1(\psi^{\pm})$ then the previous condition (3.30) is satisfied. If, moreover,*

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \lambda_i \in L_1(\psi^{\pm}) \tag{3.33}$$

*then $\widehat{\theta}_n$ satisfies the regularity condition* (R3) *for $\tau_0 = \tau/2$.*

**Proof:** Let $|s| \leq \tau/2$, $|t| \leq \tau/2$ and $i$ be arbitrary fixed. Then

$$\int d\Phi_{i,s,t}(y) = 0, \tag{3.34}$$

$$\Phi_{i,s,t}(\infty) = \lim_{y \to \infty} \Phi_{i,s,t}(y) = 0 \tag{3.35}$$

and, by (3.29),

$$\int |\psi(y)| d\Phi_{i,s,t}(y) < \infty. \tag{3.36}$$

Hence, by (3.27) and the Fubini theorem,

$$\begin{aligned}
\frac{1}{s}[H_i(t+s) - H_i(t)] &= \int\int I(0 < x \leq y) d\psi(x) d\Phi_{i,s,t}(y) \\
&= \int\int I(0 < x \leq y) d\Phi_{i,s,t}(y) d\psi(x) \\
&= \int\int I(x \leq y < \infty) d\Phi_{i,s,t}(y) d\psi(x) \\
&= \int (\Phi_{i,s,t}(\infty) - \Phi_{i,s,t}(x)) d\psi(x).
\end{aligned}$$

Therefore, by (3.34)–(3.36),

$$\frac{1}{s}[H_i(t+s) - H_i(t)] = -\int \Phi_{i,t,s}(y) d\psi(y). \tag{3.37}$$

Since

$$\lim_{s\to 0} \Phi(i,t,s)(y) = f_i(y-t) \quad \text{a.e.}$$

and since (3.30) justifies interchange of the integral and $\lim_{s\to 0}$ in (3.37), assertion (I) is proved. The first part of assertion (II) follows from the inequality

$$f_i \leq \sup_{|s|\leq \tau} f_{i,s} \leq f_i + \lambda_i \tau,$$

and the second part follows from the first part and from the fact that, under (3.31) and (3.32),

$$\frac{1}{n}\sum_{i=1}^{n} \sup_{|t|\leq\tau} |h_i(t) - h_i(0)| \leq \frac{\tau}{n}\sum_{i=1}^{n} \int \lambda_i(y) d\psi(y).$$

Indeed, under (3.33) the $\limsup_n$ of the right-hand side tends to zero as $\tau \downarrow 0$. □

For bounded sensitivities $\psi$ the assumptions of Proposition 3.9 simplify in sense that (3.29) is automatically satisfied.

For the standard nonlinear regression model with an absolutely continuous error distribution $F$ and the same $b(F)$ as in (3.13), the condition (3.29) simplifies into

$$\psi^+,\ \psi^- \in L_1(F(y - b(F))) \quad \text{and} \quad \lim_{a\to\infty} \psi^{\pm}(a) \sup_{|y|\geq a} f(y) = 0, \tag{3.38}$$

where $f$ is the derivative of $F$. Further, (3.30) takes on the form

$$\sup_{|s|\leq\tau} f(y - b(F) - s) \in L_1(\psi^{\pm}), \tag{3.39}$$

the Lipschitz condition (3.32) is in this case

$$|f(y - b(F) - t) - f(y - b(F))| \leq \lambda(y)\,|t|, \tag{3.40}$$

and the remaining conditions of assertion (II) reduce to $f,\, \lambda \in L_1(\psi^{\pm})$.

Applicability of the results of this section is illustrated in the next sections.

## 4 $L_{1+\alpha}$–estimators

Let us start with two examples.

**Example 4.1** Perhaps the best known of all $M$-estimators is the $L_2$-*estimator*

$$\hat{\theta}_n \sim \langle \psi(y) = y;\, \varphi_i \rangle. \tag{4.1}$$

Here $\rho(y) = y^2/2$ and the decomposition (2.1) and formula (2.2) are trivial in the sense that $\psi^- \equiv 0$ and $\psi^+(y) = \psi^{\pm}(y) = \psi(y) = y$. Since $\rho(y) = y^2/2$, it follows from the definition of $\hat{\theta}_n$ that, in any model (1.1), $\hat{\theta}_n$ minimizes the $L_2$-distance between observations $\mathbf{Y}_n = (Y_1, \ldots, Y_n)'$ and locators $\boldsymbol{\varphi}_n(\theta) = (\varphi_1(\theta), \ldots, \varphi_n(\theta))'$,

$$\hat{\theta}_n = \arg\min_{\Theta}\, \|\mathbf{Y}_n - \boldsymbol{\varphi}_n(\theta)\|_2, \tag{4.2}$$

where $\|\cdot\|_2$ denotes the $L_2$-norm. The rule (2.3) for adaptation of locators reduces to a *mean value rule* $\varphi_i(\theta_0) = \mathsf{E}Y_i$, i. e. the formula (3.10) for locators takes on the form

$$\varphi_i(\theta) = \int y dG(y|i, \theta), \quad \theta \in \Theta, \tag{4.3}$$

where $G(y|i, \theta)$ are the distributions of model (1.1). Similarly, the particular adaptation rules (3.12) and (3.13) reduce to

$$\varphi_i(\theta) = a(\phi(\mathbf{x}_i, \theta)) \quad \text{for} \quad a(\vartheta) = \int y dF(y|\vartheta)$$

and

$$\varphi_i(\theta) = \phi(\mathbf{x}_i, \theta) + \int y dF(y),$$

respectively.

In the regression models with additive errors, (4.2) represents a *least squared error* criterion. Due to the simplicity of both, the criterion function $\|\mathbf{Y}_n - \boldsymbol{\varphi}_n(\theta)\|_2$ and the universal adaptation rule (4.3), the $L_2$-estimators play a fundamental role in the statistical practice as well as in the theory. The linearity of $\psi(y)$, placing these estimators into the center of interest of the linear statistics, makes the asymptotic theory of these estimators relatively easy. This theory has been developed into considerable details, see e. g. Rao [25].

**Example 4.2** Another well known $M$-estimator is the $L_1$-*estimator*

$$\hat{\theta}_n \sim \langle \psi(y) = 1 - 2I(y < 0);\, \varphi_i \rangle. \tag{4.4}$$

Here, as before, $I(\cdot)$ denotes the indicator of events. The $\psi$-function of the $L_1$-estimator is an example where the decompositions (2.1), (2.2) are trivial in the sense that $\psi^- \equiv 0$ and $\psi(y) = \psi^+(y) = \psi^{\pm}(y)$ has a jump of size 2 at $y = 0$. Since $\rho(y) =| y|$, it follows from the definition 3.1 that, in any model (1.1), $\hat{\theta}_n$ minimizes the $L_1$-distance between observations $\mathbf{Y}_n = (Y_1, \ldots, Y_n)$ and locators $\boldsymbol{\varphi}_n(\theta) = (\varphi_1(\theta), \ldots, \varphi_n(\theta))$,

$$\hat{\theta}_n = \arg\min_{\Theta} \|\mathbf{Y}_n - \boldsymbol{\varphi}_n(\theta)\|_1, \tag{4.5}$$

where $\|\cdot\|_1$ denotes the $L_1$-norm. The general rule (2.3) for adaptation of locators reduces to the *median rule*, $\varphi_i(\theta_0) = \mathrm{med} Y_i$, i. e. (3.10) takes on the form

$$\varphi_i(\theta) = \mathrm{med} G(y|i, \theta), \quad \theta \in \Theta, \tag{4.6}$$

where

$$\mathrm{med} G(y|i, \theta) = \inf\{y \in \mathbb{R} : G(y|i, \theta) \geq 1/2\}$$

denotes the median of $G(y|i, \theta)$. Similarly, the special adaptation rules (3.12) and (3.13) reduce to

$$\varphi_i(\theta) = a(\phi(\mathbf{x}_i, \theta)) \quad \text{for} \quad a(\vartheta) = \mathrm{med} F(y|\vartheta)$$

and

$$\varphi_i(\theta) = \phi(\mathbf{x}_i, \theta) + \mathrm{med} F(y).$$

In the regression models with additive errors, (4.2) represents a *least absolute error* criterion. Due to the relative simplicity of both, the criterion function $\|\mathbf{Y}_n - \boldsymbol{\varphi}_n(\theta)\|_1$ and the universal adaptation rule (4.6), the $L_1$-estimators play an important role in the statistical practice as well as in the theory (see e. g. Serfling [27], Dodge [4], Farenbrother [6], Ronchetti [26], Pollard [22], Knight [15] and references therein).

The $L_1$-or $L_2$-estimators $\hat{\theta}_n$ can be embedded into various families of estimators $\hat{\theta}_n^{(\alpha)}$ with a parameter $\alpha \in \mathbb{R}$ controlling finite-sample-size properties, such as rejection regions and variances-covariances of deviations $\hat{\theta}_n^{(\alpha)} - \theta_0$, or asymptotic properties like influence curves and relative efficiencies.

In this section we study the family of *quantile* $L_{1+\alpha}$-*estimators*

$$\hat{\theta}_n^{(\alpha)} \sim \langle \psi(y) = 1 + \alpha - 2I(y < 0);\, \varphi_i \rangle, \quad -1 < \alpha < 1, \tag{4.7}$$

where $\hat{\theta}_n^{(0)}$ is the $L_1$-estimator of Example 4.2. The $\psi$-functions of (4.7) differ from the $\psi$-function of (4.4) by a constant shift $\alpha$ : if $\alpha > 0$ then the sensitivity is suppressed in the domain $y < 0$ and enhanced in the domain $y \geq 0$, while for $\alpha < 0$ the opposite is true. Note that for the extremal $\alpha = 1$ or $\alpha = -1$ we obtain in (4.7) sensitivities concentrated only on $y \geq 0$ or $y < 0$, respectively. The corresponding quantile $L_2$- and $L_0$-estimators are legitimate particular cases of the $M$-estimators studied in this paper (one of them is studied at the end of this section). Notice that the quantile $L_2$-estimator differs from the usual $L_2$-estimator of Example 4.1.

Since $\rho(y) = y\,\psi(y) = \rho_\alpha(y)$ where

$$\rho_\alpha(y) = (1+\alpha)\,y\,I(y>0) - (1-\alpha)\,yI(y<0),$$

the definition of $M$-estimator implies that

$$\hat{\theta}_n^{(\alpha)} = \arg\min_{\Theta} \left( (1+\alpha)\,\|\mathbf{Y}_n - \boldsymbol{\varphi}_n(\theta)\|_1^+ + (1-\alpha)\,\|\mathbf{Y}_n - \boldsymbol{\varphi}_n(\theta)\|_1^- \right), \tag{4.8}$$

where

$$\|\mathbf{Y}_n - \boldsymbol{\varphi}_n(\theta)\|_1^{+(-)} = \sum_{i=1}^{n} |Y_i - \varphi_i(\theta)|^{+(-)}$$

and

$$|y|^+ = |y|\,I(y>0) \quad \text{and} \quad |y|^- = |y|\,I(-y>0).$$

Thus we see that the criterion (4.8) differs from (4.5) in that the criterion function takes the values $|Y_i - \varphi_i(\theta)|$ with different weights $1+\alpha$ or $1-\alpha$, depending on whether $Y_i - \varphi_i(\theta)$ is positive or negative. Since the above defined $\rho_\alpha(y)$ is twice larger than $\rho_\beta(y)$ of (1.12) for $\beta = (1+\alpha)/2$, the quantile $L_1$–estimators $\widehat{\theta}_n^{(\alpha)} \sim \langle \rho_\alpha; \varphi_i \rangle$ coincide with the estimators $\widehat{\theta}_n^{(\beta)} \sim \langle \rho_\beta; \varphi_i \rangle$ where $\rho_\beta$ is given by (1.13) for $\beta = (1+\alpha)/2$. If these estimators are applied in the regression models then they are called *regression quantiles.*

For the $\psi$-function defined in (4.7), and for arbitrary $\varphi \in \mathbb{R}$ and arbitrary distribution function $G(y)$,

$$\int \psi(y-\varphi)dG(y) = 1 + \alpha - 2G(\varphi).$$

Consequently, the general rule (3.10) for adaptation of locators reduces into the $(1+\alpha)/2$-*quantile rule*

$$\varphi_i(\theta) = G^{-1}\left((1+\alpha)/2 \,|\, i, \theta\right), \quad \theta \in \Theta, \tag{4.9}$$

where

$$G^{-1}(\beta) = \inf\left\{y \in \mathbb{R} : G(y) \geq \beta\right\}, \quad 0 < \beta < 1, \tag{4.10}$$

is the quantile function of $G(y)$. From (3.12) or (3.13) we obtain the special adaptation rules

$$\varphi_i(\theta) = F^{-1}\left((1+\alpha)/2 \,|\, \phi(\mathbf{x}_i, \theta)\right) \quad \text{or} \quad \varphi_i(\theta) = \phi(\mathbf{x}_i, \theta) + F^{-1}((1+\alpha)/2).$$

The second of these rules is used in the standard nonlinear regression model (3.5). It cannot be used if the error distribution $F(y)$ is unknown. By (3.14), in this case one can consider an extended $L_{1+\alpha}$-estimator $(\hat{\theta}_n^{(\alpha)}, \hat{b}_n^{(\alpha)})$ of the extended true parameter $(\theta_0,\, b_0 = F^{-1}((1+\alpha)/2))$ adapted by the rule

$$\varphi_i(\theta, b) = \phi(\mathbf{x}_i, \theta) + b, \quad (\theta, b) \in \widetilde{\Theta}, \quad \widetilde{\Theta} = \Theta \times \mathbb{R}. \tag{4.11}$$

But

$$Y_i = \phi(\mathbf{x}_i, \theta_0) + b_0 + \tilde{\mathcal{E}}_i, \quad \tilde{\mathcal{E}}_i \sim \tilde{F}(y) = F(y + b_0),$$

where $\tilde{F}^{-1}((1+\alpha)/2) = 0$, and $\phi(\mathbf{x}_i, \theta) + b$ is a special case of a general function $\tilde{\phi}(\mathbf{x}_i, \tilde{\theta})$ of $(m+1)$-dimensional parameter $\tilde{\theta} \in \tilde{\Theta}$, $\tilde{\Theta} \subset \mathbb{R}^{m+1}$ open. Therefore $(\hat{\theta}_n^{(\alpha)}, \hat{b}_n^{(\alpha)})$ is a special case of a general $L_{1+\alpha}$-estimator $\hat{\tilde{\theta}}_n$ of true $\tilde{\theta}_0 \in \tilde{\Theta}$ in the model

$$Y_i = \tilde{\phi}(\mathbf{x}_i, \widetilde{\theta}_0) + \tilde{\mathcal{E}}_i, \quad \tilde{\mathcal{E}}_i \sim \tilde{F}(y), \quad \tilde{F}^{-1}((1+\alpha)/2) = 0. \tag{4.12}$$

All conditions imposed in this model on $\tilde{F}(y)$ and $\tilde{\phi}(\mathbf{x}_i, \widetilde{\theta})$ easily transform into conditions on distribution $F(y) = \tilde{F}(y - F^{-1}((1-\alpha)/2))$ of the errors $\mathcal{E}_i$ in the model (3.5) and on $\phi(\mathbf{x}_i, \theta) + b$. Similarly, all properties of the estimator $\hat{\tilde{\theta}}_n^{(\alpha)}$ straightforward transform into properties of the particular version $(\hat{\theta}_n^{(\alpha)}, \hat{b}_n^{(\alpha)})$. Hence, in the standard nonlinear (and linear) regression with an unknown error distribution, it suffices to investigate the estimators (4.7) under the assumption

$$F^{-1}((1+\alpha)/2) = 0, \tag{4.13}$$

using the adaptation rule

$$\varphi_i(\theta) = \phi(\mathbf{x}_i, \theta), \quad \theta \in \Theta, \tag{4.14}$$

for $\Theta \subset \mathbb{R}^m$ open and $m \geq 2$.

The estimators $(\hat{\theta}_n^{(\alpha)}, \hat{b}_n^{(\alpha)})$, $\alpha \in (-1, 1)$, with the adaptation rule (4.11), have been introduced into the literature by Koenker and Basset [16]. As said above, these estimators, called regression quantiles, coincide with $(\theta_n^{(\beta)}, \hat{b}_n^{(\beta)})$ defined by the criterion functions (1.13) for $\beta = (1+\alpha)/2 \in (0, 1)$. Koenker and Basset established the asymptotic normality of these estimators in the standard linear regression (3.6) with an unknown distribution $F(y)$. Jurečková and Procházka [13] extended their result to the standard nonlinear regression (3.5) with an unknown $F(y)$. In this section we study the estimators (4.7) under the restrictions (4.13), (4.14). As argued above, our study covers as particular cases the estimators $(\hat{\theta}_n^{(\alpha)}, \hat{b}_n^{(\alpha)})$ of $(m-1)$-dimensional parameter $\theta_0$ and $b_0 = F^{-1}((1-\alpha)/2)$ in the model (3.5) free of the restriction (4.13).

We shall obtain asymptotic normality of the estimators $\hat{\theta}_n^{(\alpha)}$, $\alpha \in (-1, 1)$, defined by (4.7) and (4.13), from Theorem 2.1 under the assumption (4.13). To this end we assume the following.

(a) $\hat{\theta}_n^{(\alpha)}$ is consistent in the sense of (1.10).

(b) One can find a closed ball $B \subset \Theta$ of a radius $\delta > 0$ centered at $\theta_0$ on which there exist the gradients

$$\dot{\phi}(\mathbf{x}_i, \theta) = \left(\frac{\partial}{\partial \theta_1}, \dots \frac{\partial}{\partial \theta_m}\right)' \phi(\mathbf{x}_i, \theta), \qquad i \in \mathbb{N},$$

and a constant $\lambda$ possibly depending on $B$, such that

$$\|\dot{\phi}(\mathbf{x}_i, \theta)\| \le \lambda \quad \text{and} \quad \|\dot{\phi}(\mathbf{x}_i, \theta) - \dot{\phi}(\mathbf{x}_i, \tilde{\theta})\| \le \lambda \|\theta - \tilde{\theta}\|$$

for all $\theta, \tilde{\theta} \in B,\ i \in \mathbb{N}$, i.e. the regularity condition (R2) holds.

(c) It holds

$$\Psi_n = \frac{1}{n} \sum_{i=1}^{n} \dot{\phi}(\boldsymbol{x}_i, \theta)\, \dot{\phi}(\boldsymbol{x}_i, \theta)' \to \Psi,$$

where the $m \times m$ matrix $\Psi$ is positive definite.

(d) The error distribution function $F(y)$ is differentiable on an interval $(-\tau, \tau)$ and the derivative $f(y)$ of $F(y)$ is continuous at $y = 0$ with $f(0) > 0$.

**Theorem 4.3** *If the conditions (a) – (d) hold and the error distribution satisfies* (4.13) *then the estimators $\hat{\theta}_n^{(\alpha)}$ defined by* (4.7) *and* (4.14) *are asymptotically normal in the sense*

$$\sqrt{n} \left(\hat{\theta}_n^{(\alpha)} - \theta_0\right) \xrightarrow{\mathcal{L}} N\left(0, \frac{1-\alpha^2}{4f^2(0)} \Psi^{-1}\right), \tag{4.15}$$

*where $f(0) > 0$ is defined by (d) and the positive definite matrix $\Psi$ is defined by (c).*

**Proof:** Let $\alpha \in (-1, 1)$ and $F(y)$ satisfying (4.13) fulfil assumptions (a) – (d). We shall verify that $\hat{\theta}_n^{(\alpha)}$ satisfies all assumptions of Theorem 2.3. By Propositions 3.9, 3.10, and (d), $\hat{\theta}_n^{(\alpha)}$ satisfies the regularity condition (R4+). By (2.10), (4.13) and (4.14), if $t \in \mathbb{R}$ then

$$H_i(t) = \int \psi(y+t)dF(y) = 1 + \alpha - 2F(-t), \quad i \in \mathbb{N}.$$

Consequently, by (d), the estimators $\hat{\theta}_n^{(\alpha)}$ satisfy the regularity condition (R3) of Theorem 2.3 for $h_i(t) = 2f(-t)$ and $\tau_0 = \tau$. As to the remaining conditions, (2.3) was clarified above, the consistency was assumed in (a), (R2) was assumed in (b) and (R1) holds because

$$\begin{aligned} \sigma_i^2 &= \int \psi^2(y)dF(y) = (1+\alpha)^2 \int_0^\infty dF(y) + (1-\alpha)^2 \int_{-\infty}^0 dF(y) \\ &= (1+\alpha)^2(1-(1+\alpha)/2) + (1-\alpha)^2(1+\alpha)/2 \\ &= 1-\alpha^2 \quad \text{for all } i \in \mathbb{N}. \end{aligned}$$

Conditions (2.13), (2.14) of condition (R3) follow from (c) for

$$\Sigma = (1-\alpha^2)\,\Psi \quad \text{and} \quad \Phi = 2f(0)\,\Psi.$$

Since the functions $\psi(t)$ as well as the gradients $\dot{\varphi}_i$ are bounded (see (b)), the remaining condition (2.17) of Theorem 2.3 holds by Proposition 2.4. The desired relation (4.15) thus follows from Theorem 2.3. □

By what has been said above, the following assertion about an arbitrary error distribution $F(y)$ follows from Theorem 4.3. In this assertion, and in the rest of section, we put

$$\beta = \frac{1+\alpha}{2}, \quad \beta \in (0,1). \tag{4.16}$$

**Corollary 4.4** *Let $\alpha \in (-1,1)$ be arbitrary, and let $\beta$ be given by (4.16). If conditions (a) – (d) hold with $F(y)$ replaced by $\tilde{F}(y) = F(y - F^{-1}(\beta))$ then the above specified $L_{1+\alpha}$-estimator $(\hat{\theta}_n^{(\alpha)}, \hat{b}_n^{(\alpha)})$ is asymptotically normal in the sense*

$$\sqrt{n}\left[(\hat{\theta}_n^{(\alpha)}, \hat{b}_n^{(\alpha)}) - (\theta_0, F^{-1}(\beta))\right] \xrightarrow{\mathcal{L}} N\left(0, \frac{\beta(1-\beta)}{f^2(F^{-1}(\beta))}\tilde{\Psi}^{-1}\right) \quad \text{as } n \to \infty \tag{4.17}$$

*for $f(y) = dF(y)/dy$ and the matrix*

$$\tilde{\Psi} = \begin{pmatrix} \Psi & , & 0 \\ 0 & , & 1 \end{pmatrix},$$

*where $\Psi$ is given by (c).*

The asymptotic laws (4.15), (4.17) have been established for the $L_1$-estimator, where $\beta = 1/2$, as well as for the general $L_{1+\alpha}$-estimator under various conditions, see e. g. Pollard [22], Jurečková and Procházka [13] and other cited there. Let us compare the present conditions for these laws with the conditions assumed in the two cited papers.

Pollard [22] assumed (6.7) so that his conditions can be compared with those of Theorem 4.3. He studied the $L_1$-estimator $\hat{\theta}_n^{(0)}$ in the standard linear regression, where (b) is automatically fulfilled and the matrices considered in (c) are

$$\Psi_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\,\mathbf{x}_i'.$$

For these matrices, (c) is a classical condition of regression analysis. As shown on p. 189 of Pollard [22], this condition is somewhat stronger than what is assumed in his Theorem 1. On the other hand, our condition (d) is slightly weaker than the assumption that $F(y)$ is continuously differentiable in an interval $(-\tau, \tau)$ with the derivative $f(y)$ positive on $(-\tau, \tau)$, which appears in the mentioned Theorem 1. The consistency of $\hat{\theta}_n^{(0)}$ assumed in (a) takes place under (c) and (d). This can be proved by applying the Convexity Lemma on p. 187 of Pollard [22].

Thus, as to the $L_1$-estimators in linear models, the conditions obtained from Theorem 2.3 are comparable with previously published ones, obtained by methods tailor-designed for these estimators and models. In this sense the comparison with [22] demonstrates that Theorem 2.3 is not trivial.

Jurečková and Procházka [13] studied the same estimator and model as Corollary 4.4. The conditions (b), (c) of this corollary are the same as (b), (c) in Theorem 4.3. The condition (d) is changed in the sense that $F(y)$ is differentiable in an interval $(F^{-1}(\beta)-\tau, F^{-1}(\beta)+\tau)$ with the derivative $f(y)$ continuous at $y = F^{-1}(\beta)$ and $f(F^{-1}(\beta)) > 0$. The consistency of $(\hat{\theta}_n^{(\alpha)}, \hat{b}_n^{(\alpha)})$ required in (a) follows under (b), (c), (d) by the same method as used above for the consistency of $\hat{\theta}_n^{(0)}$. Jurečková and Procházka assumed, in addition to (b), (c), (d), that $\phi(\mathbf{x}, \theta)$ is strictly monotone in each component of $\theta$, twice differentiable in each of these components, with the first and second derivatives uniformly bounded on $\mathcal{X} \times \Theta$, and that the above mentioned $f(y)$ is symmetric about $y = 0$, bounded on $\mathbb{R}$ and differentiable on $(F^{-1}(\beta)-\tau, F^{-1}(\beta)+\tau)$. Moreover, they assumed that $\mathcal{X} \subset \mathbb{R}^k$ and $\Theta \subset \mathbb{R}^m$ are compact, and that the regression functions $\phi(\mathbf{x}_i, \theta)$ and gradients $\dot{\phi}(\mathbf{x}_i, \theta)$ satisfy some additional conditions.

Obviously, here one can deduce a stronger conclusion in favour of Theorem 2.3 than formulated in the context of the simpler $L_1$-estimator above. On the other hand, it is clear that the results obtained from Theorem 2.3 cannot always be as strong as the results achievable for special $M$-estimators and models. This can be illustrated by a reference to [15], where the $L_1$-estimator is studied in a standard linear regression with error distribution $F(y)$. The author proved an asymptotic law similar to (4.15) even in situations where the derivative $f(y)$ of $F(y)$ is discontinuous at the median of $F(y)$. To this end, by exploiting special features of the $\psi$-function defined in (4.4), and special properties of linear models, he formulated asymptotic normality conditions different from (c), (d) in Proposition 4.1, and also from the conditions considered in the previous literature. Example 4.6 below illustrates that a similar non-applicability of our theory may take place also for other $M$-estimators.

**Remark 4.5** By (4.17), the asymptotic relative efficiency in the class of quantile $L_{1+\alpha}$-estimators depends on the function $\Gamma(\beta) = \beta(1-\beta)/f^2(F^{-1}(\beta))$; if $\beta_0 = \arg\min_{\beta\in(0,1)} \Gamma(\beta)$ then the estimator with $\alpha = 2\beta_0 - 1$ is relatively most efficient (cf. (4.16)). By the l'Hospital rule, if $f$ has differentiable tails with a derivative $\dot{f}$ then, for $\beta \to 0$ and $\beta \to 1$,

$$\lim \Gamma(\beta) = \lim \frac{1-2\beta}{2\dot{f}(F^{-1}(\beta))} = \infty$$

provided $\dot{f}(y) \uparrow 0$ for $y \to \infty$ and $\dot{f}(y) \downarrow 0$ for $y \to -\infty$. In this typical case the indices $\alpha$ of all relatively most efficient estimators are bounded away from $-1$ and 1. If $\dot{f}$ is continuous on $\mathbb{R}$ then at least one such relatively most efficient $L_{1+\alpha}$-estimator exists.

**Example 4.6** Let the error distribution be exponential, $F(y) = (1 - e^{-y})\, I(y > 0)$. Then $f(F^{-1}(\beta)) = 1 - \beta$. In this case $\Gamma(\beta) = \beta/(1-\beta)$ is increasing on $(0,1)$, so that one can expect that the extremal quantile $L_0$-estimator $\hat{\theta}_n^{(-1)}$ maximizes the asymptotic relative efficiency in the class of estimators $\hat{\theta}_n^{(\alpha)}$, $\alpha \in [-1,1]$. According to (4.7), the adapted version of this estimator is defined by

$$\hat{\theta}_n^{(-1)} = \underset{\theta \in \Theta}{\arg\min} \sum_{i=1}^{n} |Y_i - \phi(\mathbf{x}_i, \theta)|\, I(Y_i < \phi(\mathbf{x}_i, \theta)).$$

Here

$$H(t) = 2(e^t - 1)\, I(t < 0),$$

and

$$h(t) = 2e^t\, I(t < 0).$$

We see that the regularity condition (R4) does not hold. Consequently, Theorem 2.3 is not applicable to $\hat{\theta}_n^{(-1)}$, i. e. (4.17) is not guaranteed for $\alpha = -1$ ($\beta = 0$). In fact, since $\Gamma(0) = 0$, one can expect in this case a higher rate of consistency than $\sqrt{n}$ obtained in (4.17). The higher rate of consistency can be easily verified if $\Theta = \mathbb{R}$ and $\phi(\mathbf{x}, \theta) = \theta$, i. e. if $Y_i = \theta_0 + \mathcal{E}_i$ where $\mathcal{E}_i$ are exponentially distributed errors. Then

$$\hat{\theta}_n^{(-1)} = \min\{Y_1, \ldots, Y_n\}$$

so that

$$\mathsf{P}\left(n(\hat{\theta}_n^{(-1)} - \theta_0) > t\right) = e^{-t}, \quad t \in \mathbb{R},$$

i. e. $\hat{\theta}_n^{(-1)}$ is consistent of the order $n$.

## 5 $L_{2+\alpha}$–estimators

In the statistical literature, the classical $L_2$-estimator (4.1) has been embedded to many families of $M$-estimators. These can usually be interpreted as families of $L_{2+\alpha}$-estimators

$$\langle \psi_\alpha;\, \varphi_i \rangle, \quad \alpha \in \mathbb{R}, \tag{5.1}$$

with $\psi_\alpha(y)$ continuous at $\alpha = 0$ and $\psi_0$ coinciding with $\psi(y)$ of (4.1), i. e. satisfying for all $y \in \mathbb{R}$ the relations

$$\lim_{\alpha \to 0} \psi_\alpha(y) = \psi_0(y) \quad \text{and} \quad \psi_0(y) = y. \tag{5.2}$$

In other words, the family of estimators can be rearranged so that $\alpha = 0$ leads to the $L_2$-estimator.

**Example 5.1** The *Huber estimators* (see e. g. [12]) form a family of the type (5.1) with

$$\psi_\alpha(y) = \int_0 yI(-|\alpha|^{-1} < s < |\alpha|^{-1})ds \quad \text{for } \alpha \neq 0, \tag{5.3}$$

extended to $\alpha = 0$ in accordance with (5.2). Here the decomposition (2.1) and formula (2.2) are trivial in the sense that $\psi_\alpha^- \equiv 0$ and $\psi_\alpha^+ = \psi_\alpha^\pm = \psi_\alpha$. The *skipped mean* is defined by

$$\psi_\alpha(y) = y\, I(-|\alpha|^{-1} < y < |\alpha|^{-1}) \quad \text{for} \quad \alpha \neq 0$$

and extended by (5.2). If $\alpha \neq 0$ then $\psi_\alpha^+(y)$ coincide with Huber's (5.3), $\psi_\alpha^-(y) = I(y \geq |\alpha|^{-1}) - I(y < -|\alpha|^{-1})$ and

$$\rho_\alpha(y) = \alpha^{-2} - (\alpha^{-2} - y^2)\, I(-|\alpha|^{-1} < y < \alpha).$$

For more details about this and the next example we refer to [8]. The *Tukey biweight* is defined by

$$\psi_\alpha(y) = y(\alpha^{-2} - y^2)^2\, I(-|\alpha|^{-1} < y < |\alpha|^{-1}) \quad \text{for } \alpha \neq 0,$$

where, for $\alpha \neq 0$,

$$\begin{aligned} \psi_\alpha^+ &= \int_0^y I\left(-\left(\sqrt{3}|\alpha|\right)^{-1} < s < \left(\sqrt{3}|\alpha|\right)^{-1}\right) ds, \\ \psi_\alpha^-(y) &= \int_0^y I\left(s > \left(\sqrt{3}|\alpha|\right)^{-1}\right) d\psi(s) - \int_0^y I\left(s < -\left(\sqrt{3}|\alpha|\right)^{-1}\right) d\psi(s), \end{aligned}$$

and

$$\rho_\alpha(y) = \frac{1}{6|\alpha|^6} - \frac{(\alpha^{-2} - y^2)}{6}\, I\left(-|\alpha|^{-1} < y < |\alpha|^{-1}\right).$$

Portnoy [23] and independently Vajda [29] studied the family of $L_{2+\alpha}$ estimators defined by

$$\psi_\alpha(y) = y\, e^{-(\alpha y)^2} \quad \text{for} \quad \alpha \neq 0 \tag{5.4}$$

with

$$\rho_\alpha(y) = \frac{1}{2\alpha^2}\left(1 - e^{-(\alpha y)^2}\right) \quad \text{for} \quad \alpha \neq 0.$$

As is shown in the second reference, the estimators defined by (5.4) can be obtained from a minimum distance rule applied to $\alpha^2$-divergences of theoretical and empirical distributions.

For the $L_{2+\alpha}$-estimators with $\alpha \neq 0$ studied in this example, there is no universal adaptation rule similar to the $(1+\alpha)/2$-quantile rule (4.9) of previous section, or to the mean value rule (4.3) applicable when $\alpha = 0$. One general adaptation rule applicable to these estimators is given in the next proposition.

**Proposition 5.2** *Consider an $M$-estimator $\hat{\theta}_n \sim \langle \psi; \varphi_i \rangle$ with a monotone $\psi(y)$, skew-symmetric about $y = 0$, in the standard nonlinear regression model (3.5) with an error distribution $F(y)$ satisfying the condition $\psi \in L_1(F)$. If $F(y) - F(0)$ is skew-symmetric about $y = 0$ (i. e. if the errors are symmetrically distributed about zero) then the locators are adapted by the rule*

$$\varphi_i(\theta) = \phi(\mathbf{x}_i, \theta), \quad i \in \mathbb{R}. \tag{5.5}$$

*This adaptation is unique unless there exists a constant $b \in \mathbb{R}$ such that*

$$\psi(y - b) = \psi(y) \quad F - a.s. \tag{5.6}$$

**Proof:** The skew-symmetries of $\psi$ and $F$ imply that

$$\int \psi(y) dF(y) = 0.$$

By (3.13), this means that (5.5) is an adaptation rule. If $b \neq 0$ then the monotonicity of $\psi$ implies that $\psi(y - b) - \psi(y)$ does not change sign on $\mathbb{R}$. Therefore

$$\int \psi(y - b) dF(y) \neq \int \psi(y) dF(y) = 0$$

unless (5.6) holds. By (3.13), this implies the uniqueness of the rule (5.5). □

The skew-symmetry of the above considered sensitivity functions $\psi$ about 0 means that the sensitivity of the corresponding estimators to errors in data is symmetrically distributed about 0. In the rest of this section we study one class of $L_{2+\alpha}$-estimators with sensitivity functions $\psi_\alpha$ skew-asymmetric about 0. Such estimators are convenient when errors in data are asymmetrically distributed. As an example we may consider the situation when nonnegative data $X_i$ are transformed into $Y_i = \ln X_i$ for fitting a symmetric location model on $\mathbb{R}$. Then an error $\varepsilon$ in data $X_i$ leads to an error $\varepsilon\, e^{-Y_i}$ in data $Y_i$, which is exponentially decreasing with increasing values of $Y_i$. This partially motivates the following steps.

Let us study the family of *exponential $L_{2+\alpha}$-estimators*

$$\hat{\theta}_n^{(\alpha)} \sim \langle \psi(y) = y\, e^{\alpha y};\, \varphi_i \rangle\,, \quad \alpha \in \mathbb{R}, \tag{5.7}$$

where $\hat{\theta}_n^{(0)}$ is the $L_2$-estimator of Example 4.1. Here

$$\rho(y) = \begin{cases} \dfrac{e^{\alpha y}(\alpha y - 1) + 1}{\alpha^2} & \text{if} \quad \alpha \neq 0 \\ y^2 & \text{if} \quad \alpha = 0. \end{cases} \tag{5.8}$$

A strong additional motivation for the estimators (5.7) is a relatively simple adaptation, in the sense of (3.13), to the generalized regression models (3.5) with exponential parent families $\mathcal{F}$, in particular to the generalized linear models mentioned in Example 3.5. The only estimator with this property studied so far in the literature seems to be the classical MLE. Thus the class (5.7) deserves to be investigated in detail.

By (3.13), the adaptation of $\widehat{\theta}_n^{(\alpha)}$ to the regression model (3.1) with a parent family $\mathcal{F} = \{F(y|\vartheta) : \vartheta \in y\}$ reduces to solution of equations (3.12), which are now of the form

$$\int (y-a)\, e^{\alpha(y-a)} dF(y|\vartheta) = 0, \quad \vartheta \in T. \tag{5.9}$$

We restrict ourselves to the homogeneous regression model (3.2) with exponential families $\mathcal{F}$ in the natural form (cf. Brown [3]), i.e. with densities

$$f(y|\vartheta) = e^{\vartheta y - c(\vartheta)} \sim F(y|\vartheta), \quad \vartheta \in T, \tag{5.10}$$

with respect to a $\sigma$-finite measure $\nu$ on $\mathbb{R}$, where

$$T = \left\{\vartheta \in \mathbb{R} : 0 < \int e^{\vartheta y} d\nu(y) < \infty\right\} \quad \text{and} \quad c(\vartheta) = \ln \int e^{\vartheta y} d\nu(y). \tag{5.11}$$

Here $T$ is convex, and $c(\vartheta)$ is a *cumulant generating function* convex on $T$.

For families $\mathcal{F}$ in a natural form, the distributions figuring in (1.1) are given by

$$G(y|i,\theta) \sim g(y|i,\theta) = f(y)|\phi(\mathbf{x}_i,\theta) = e^{\phi(\mathbf{x}_i,\theta)\, y - c(\phi(\mathbf{x}_i,\theta))} \tag{5.12}$$

for all $y$ from the support of $\nu$, and all $i \in \mathbb{N}$ and $\theta \in \Theta$. If $\phi(\mathbf{x},\theta) = \mathbf{x}'\theta$ then we obtain generalized linear models with *natural link functions* (see e. g. Fahrmeir and Kaufmann [5]) where

$$G(y|i,\theta) \sim g(y|i,\theta) = f(y|\mathbf{x}_i'\theta) = e^{\mathbf{x}_i'\theta y - c(\mathbf{x}_i'\theta)} \tag{5.13}$$

for all $y$ from the support of $\nu$ and all $i \in \mathbb{N}$ and $\theta \in \Theta$. The exponential families are assumed to be nontrivial in the sense that $\nu$ is not concentrated in one point, that $T$ has a nonempty interior, and that all values $\mathbf{x}_i'\theta$ or $\phi(\mathbf{x}_i',\theta)$ are in this interior.

In a nontrivial exponential family $\mathcal{F}$, the cumulant generating function $c(\vartheta)$ is strictly convex and infinitely differentiable on the interior $T^0$ of $T$, with derivatives

$$\dot{c}(\vartheta) = \frac{dc(\vartheta)}{d\vartheta} \quad \text{and} \quad \ddot{c}(\vartheta) = \frac{d^2c(\vartheta)}{d\vartheta^2} \tag{5.14}$$

satisfying for all $\vartheta \in T^0$ the equalities

$$\int (y-\dot{c}(\vartheta))\, e^{\vartheta y - c(\vartheta)} d\nu(y) = 0 \tag{5.15}$$

and

$$\int (y-\dot{c}(\vartheta))^2\, e^{\vartheta y - c(\vartheta)} d\nu(y) = \ddot{c}(\vartheta). \tag{5.16}$$

The derivative $\dot{c}(\vartheta)$ of the strictly convex function $c(\vartheta)$ is increasing on the interior $T^0$, and the second derivative $\ddot{c}(\vartheta)$ is positive on $T^0$. By (5.15) and (5.16), $\dot{c}(\vartheta)$ is the mean in $\mathcal{F}$,

$$\mu(\vartheta) = \int y f(y|\vartheta) d\nu(y), \tag{5.17}$$

and $\ddot{c}(\vartheta)$ is the variance or, equivalently, the *Fisher information* of $\mathcal{F}$, i. e.

$$\mu(\vartheta) = \dot{c}(\vartheta) \quad \text{and} \quad \mathcal{I}(\vartheta) = \ddot{c}(\vartheta), \quad \vartheta \in T^0. \tag{5.18}$$

Moreover, for each $\vartheta \in T^0$, $a = \dot{c}(\vartheta)$ is the unique solution of the equations

$$\int (y-a)\, e^{\vartheta y - c(\vartheta)} d\nu(y) = 0 \tag{5.19}$$

and

$$\int (y-a)^2\, e^{\vartheta y - c(\vartheta)} d\nu(y) = \ddot{c}(\vartheta). \tag{5.20}$$

For simplicity, we study the important particular case where $T = \mathbb{R}$. Then in the homogeneous regression models (3.1) under consideration, equations in (5.9) reduce to

$$\int (a-y)\, e^{\alpha(y-a)+\vartheta y - c(\vartheta)} d\nu(y) = 0, \quad \vartheta \in \mathbb{R}, \tag{5.21}$$

which can be obtained from equations (5.19) with $\vartheta$ replaced by $\vartheta + \alpha$. Therefore, given any $\alpha \in \mathbb{R}$,

$$a(\vartheta) = \dot{c}(\vartheta + \alpha), \quad \vartheta \in \mathbb{R}, \tag{5.22}$$

are the unique solutions of equations (5.21). According to (3.12), this means that the *pseudoadditive rule*

$$\varphi_i(\theta) = \dot{c}\left(\phi(\mathbf{x}_i, \theta) + \alpha\right), \quad \theta \in \Theta, \tag{5.23}$$

leads to the adaptation of exponential $L_{2+\alpha}$-estimators to the exponential homogeneous regression models under consideration in the sense of (3.12), i. e. the adapted versions of the estimators (5.7) are

$$\widehat{\theta}_n^{(\alpha)} \sim \langle y\, e^{\alpha y};\, \dot{c}(\phi(\mathbf{x}_i, \theta) + \alpha) \rangle, \quad \alpha \in \mathbb{R}. \tag{5.24}$$

Replacing $\phi(\mathbf{x}_i, \theta)$ by the scalar product $\mathbf{x}_i'\theta$ we obtain from (5.23), (5.24) corresponding formulas for the exponential $L_{2+\alpha}$-estimators adapted to generalized linear models with natural link functions.

Let us look at the restrictions which Theorem 2.3 imposes on the estimators (5.24) and the respective exponential regression models. We start with sufficient conditions for (R3) and (R4+). The decomposition of the $\psi$-function figuring in (5.24) is as follows

$$\psi^+(y) = \begin{cases} y\, e^{\alpha y}\, I(\alpha y + 1 \geq 0) + \dfrac{1}{2}\, I(\alpha y + 1 \geq 0) & \text{if} \quad \alpha \neq 0, \\ y & \text{if} \quad \alpha = 0 \end{cases} \tag{5.25}$$

and

$$\psi^-(y) = \begin{cases} -y\, e^{\alpha y}\, I(\alpha y + 1 < 0) + \dfrac{1}{2}\, I(\alpha y + 1 \geq 0) & \text{if} \quad \alpha \neq 0, \\ 0 & \text{if} \quad \alpha = 0. \end{cases} \tag{5.26}$$

The part $\psi^-(y)$ is non-explosive, Lipschitz and bounded and square integrable on $\mathbb{R}$. The other part $\psi^+(y)$ is not so nice – it is explosive, non-Lipschitz, unbounded and square non-integrable on $\mathbb{R}$. To satisfy (R4+) we shall need Proposition 3.7.

**Proposition 5.3** *The estimator $\widehat{\theta}_n$ defined by (5.7) fulfils in the model under consideration for all $\alpha \in \mathbb{R}$ the condition 2.12 in the regularity condition* (R3). *If the expectations $\mu(\vartheta)$ and Fisher informations $\mathcal{I}(\vartheta)$ defined in (5.18) satisfy, for*

$$\phi_i = \phi(\mathbf{x}_i, \theta_0), \quad i \in \mathbb{N}, \tag{5.27}$$

*and some $\alpha \in \mathbb{R}$, the inequalities*

$$\sup_{i \in \mathbb{N}} \left[\mu(\phi_i + 2\alpha) - \mu(\phi_i + \alpha)\right]^2 < \infty \tag{5.28}$$

*and*

$$\sup_{n \in \mathbb{N}} \frac{1}{n} \sum_{i=1}^n \sup_{|t| \leq 2|\alpha|} \mathcal{I}(\phi_i + t) < \infty, \tag{5.29}$$

*then the corresponding estimator $\widehat{\theta}_n^{(\alpha)}$ defined by (5.7) fulfills also the regularity conditions* (R1) *and* (R4+).

**Proof:** (I) For every $\alpha \in \mathbb{R}$, the derivative

$$\dot{\psi}^+(y) = (\alpha y + 1)\, e^{\alpha y}\, I(\alpha y + 1 \geq 0) \tag{5.30}$$

of $\psi^+(y)$ is nondecreasing on $\mathbb{R}$ if $\alpha \geq 0$, and nonincreasing if $\alpha < 0$. Consequently,

$$\Psi_\tau(y) := \sup_{|s| \leq \tau} \left|\dot{\psi}^+(y+s)\right| = \dot{\psi}^+(y + \tau\, \mathrm{sgn}\alpha), \quad y \in \mathbb{R}, \ \tau > 0, \tag{5.31}$$

where

$$\mathrm{sgn}\alpha = \begin{cases} 1 & \text{if} \quad \alpha \geq 0, \\ -1 & \text{if} \quad \alpha < 0. \end{cases}$$

By using the relation

$$|\dot{\psi}^+(y) - \dot{\psi}(y)| \leq \sup_{t \in \mathbb{R}} |\dot{\psi}^-(t)| \leq 1/e^2, \quad y \in \mathbb{R},$$

we find that (3.17) is equivalent to the condition

$$\sup_{n \in \mathbb{N}} \frac{1}{n} \sum_{i=1}^n \mathsf{E}\left[\dot{\psi}(Y_i - \dot{c}(\phi_i + \alpha) + \tau\, \mathrm{sgn}\alpha)\right]^2 < \infty \quad \text{for some } \tau > 0, \tag{5.32}$$

where $\dot{\psi}(y) = (\alpha y + 1)\, e^{\alpha y}$. By the Taylor Theorem the difference in the brackets of (5.28) equals $\alpha \ddot{c}(\phi_i + \alpha_i)$ for some $\alpha_i \in \mathbb{R}$. Using (5.19), (5.20), we obtain that the expectation of (5.31) is equal to

$$\begin{aligned} & e^{2|\alpha|\tau + b_i(\alpha)} \left[\alpha^2 \ddot{c}(\phi_i + 2\alpha) + (\alpha^2 \ddot{c}(\phi_i + \alpha_i) + |\alpha|\, \tau + 1)^2\right] \\ \leq\; & e^{2|\alpha|\tau} \left[\alpha^2 \ddot{c}(\phi_i + 2\alpha) + (\alpha^2 \ddot{c}(\phi_i + \alpha_i) + |\alpha|\, \tau + 1)^2\right], \end{aligned}$$

where

$$\begin{aligned} 0 \;\leq\; & b_i(\alpha) = c(\phi_i + 2\alpha) - c(\phi_i) - 2\alpha \dot{c}(\phi_i + 2\alpha) \\ \leq\; & 2|\alpha| \sup_{|t| \leq 2|\alpha|} \mathcal{I}(\phi_i + t) \end{aligned} \tag{5.33}$$

because $c(\vartheta)$ is convex and $\ddot{c}(\vartheta) = \mathcal{I}(\vartheta) > 0$. Therefore, if (5.28) and (5.29) hold then (5.32) holds too, and Proposition 3.7 implies that (R4+) holds.

(II) Using (5.33) and the notation of part (I), we get from the definition of $\sigma_i^2$ in (2.4) and from (5.23),

$$\begin{aligned} \sigma_i^2 &= \mathsf{E}\left[\psi(Y_i - \dot{c}(\phi_i + \alpha))\right]^2 \\ &= e^{b_i(\tau)} \left[\ddot{c}(\phi_i + 2\alpha) + (\dot{c}(\phi_i + 2\alpha) - \dot{c}(\phi_i + \alpha))^2\right]. \end{aligned} \tag{5.34}$$

By (5.33), the assumptions (5.28) and (5.29) imply the inequality (2.5) required in (R1).

(III) Using (5.33), we get from the formula for $H_i(t)$ in (2.10) and from (5.23),

$$\begin{aligned} H_i(t) &= \mathsf{E}\, \psi(Y_i - \dot{c}(\phi_i + \alpha) + t) \\ &= t\, e^{\alpha t + \tilde{b}_i(\alpha)}, \quad t \in \mathbb{R}, \end{aligned}$$

where (cf. (5.33))

$$\tilde{b}_i(\alpha) = c(\phi_i + \alpha) + c(\phi_i) - \alpha \dot{c}(\phi_i + \alpha) \leq 0 \tag{5.35}$$

due to the convexity of $c(\vartheta)$. This function is differentiable on $\mathbb{R}$ with the derivative

$$h_i(t) = (\alpha t + 1)\, e^{\alpha t + \tilde{b}_i(\alpha)}, \quad t \in \mathbb{R}. \tag{5.36}$$

Since $\tilde{b}_i(\alpha) \leq 0$, it holds for all $i \in \mathbb{N}$

$$\begin{aligned} |h_i(t) - h_i(0)| &\leq \left|(\alpha t + 1)\, e^{\alpha t} - 1\right| e^{\tilde{b}_i(\alpha)} \leq |(\alpha t + 1)\, e^{\alpha t} - 1| \\ &= \left|\alpha t + (\alpha t + 1)\,(e^{\alpha t} - 1)\right|. \end{aligned}$$

Therefore the condition 2.12 is satisfied on the infinite interval $(-\tau_0, \tau_0) = \mathbb{R}$ even if the conditions (5.28) and/or (5.29) fail to hold. □

Combining Proposition 5.3 and Theorem 2.3, we obtain the following assertion, in which we use for $\alpha \in \mathbb{R}$ and $i \in \mathbb{N}$ the constants $b_i(\alpha)$, $\tilde{b}_i(\alpha)$ defined by (5.33), (5.35) and

$$\sigma_i^2(\alpha) = e^{b_i(\alpha)} \left[\mathcal{I}(\phi_i + 2\alpha) + (\mu(\phi_i + 2\alpha) - \mu(\phi_i + \alpha))^2\right] \tag{5.37}$$

for $\phi_i$ and $\mu(\vartheta)$, $\mathcal{I}(\vartheta)$ defined by (5.27) and (5.18). We also use the formulas

$$\dot{\varphi}_i(\theta_0) = \mathcal{I}(\phi_i + \alpha)\, \dot{\phi}_i, \quad \text{for} \quad \dot{\phi}_i = \dot{\phi}(\mathbf{x}_i, \theta_0) \tag{5.38}$$

for the gradients $\dot{\varphi}_i(\theta_0)$ considered in conditions (2.8), (2.9) of (R2), provided the derivatives

$$\dot{\phi}(\mathbf{x}_i, \theta) = \left(\frac{\partial}{\partial \theta_1}, \ldots, \frac{\partial}{\partial \theta_m}\right)' \phi(\mathbf{x}_i, \theta), \quad i \in \mathbb{N}, \tag{5.39}$$

exist in an open ball $B \subset \Theta$ centered at $\theta_0$. We restrict ourselves to the exponential models (3.7) which satisfy (R2), i.e. for which the last condition holds and the gradients (5.39) satisfy (2.8) and (2.9).

**Theorem 5.4** *Let for some $\alpha \in \mathbb{R}$ the estimator $\widehat{\theta}_n \sim \langle \psi, \varphi_i \rangle$ defined by (5.7) and (5.23) satisfy (R2) and the conditions (5.27) and (5.28) in a homogeneous regression model with exponential parent family (3.2). Further, let*

$$\Sigma_n := \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2(\alpha)\, (\mathcal{I}(\phi_i + \alpha))^2\, \phi_i \dot{\phi}_i' \to \Sigma \tag{5.40}$$

*and*

$$\Phi_n := \frac{1}{n} \sum_{i=1}^{n} e^{\tilde{b}_i(\alpha)} (\mathcal{I}(\phi_i + \alpha))^2 \phi_i\, \dot{\phi}_i' \to \Phi, \tag{5.41}$$

*where the matrices $\Sigma$ and $\Phi$ are positive definite. Finally, let*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(Y_i - \mu(\phi_i + \alpha))\, \mathcal{I}(\phi_i + \alpha)\, \dot{\phi}_i \xrightarrow{\mathcal{L}} N(0, \Sigma). \tag{5.42}$$

*If $\widehat{\theta}_n^{(\alpha)}$ is consistent then it is asymptotically normal in the sense*

$$\sqrt{n}(\widehat{\theta}_n^{(\alpha)} - \theta_0) \xrightarrow{\mathcal{L}} N(0,\, \Phi^{-1} \Sigma\, \Phi^{-1}). \tag{5.43}$$

**Proof:** Since $\widehat{\theta}_n^{(\alpha)}$ is consistent and satisfies (5.23), it is adapted to the model under consideration. By Proposition 5.3, (5.40) and (5.41), it satisfies the regularity conditions (R1), (R3) and (R4+). The remaining regularity condition (R2) is assumed. Since (5.42) means in the present situation the same as (2.17) all assumptions of Theorem 2.3 are satisfied. Therefore (5.43) follows from Theorem 2.3. □

Let us look at the special case

$$\mathcal{X} \subset \mathbb{R}^m, \quad \phi(\mathbf{x}, \theta) = \mathbf{x}'\theta \quad \text{and} \quad \alpha = 0, \tag{5.44}$$

i. e. the $L_2$-estimator $\widehat{\theta}_n^{(0)}$ of a true parameter $\theta_0 \in \Theta = \mathbb{R}$ in a generalized linear model with natural link function. Then (3.1) reduces to

$$G(y|i,\theta) \sim g(y|i,\theta) = e^{\mathbf{x}_i'\theta - c(\mathbf{x}_i'\theta)}, \quad \theta \in \mathbb{R}^m, \ i \in \mathbb{N}, \tag{5.45}$$

further $\phi_i = \mathbf{x}_i'\theta_0$ in (5.27), the gradients of (5.39) are given by formula

$$\dot{\phi}_i = \dot{\phi}(\mathbf{x}_i, \theta_0) = \mathbf{x}_i, \quad \theta_0 \in \mathbb{R}^m, \ i \in \mathbb{N}, \tag{5.46}$$

and the $\psi$-function is linear, $\psi(y) = y$. The conditions (5.28), (5.29) and (R2) take place if

$$\sup_{i \in \mathbb{N}} \|\mathbf{x}_i\| < \infty. \tag{5.47}$$

Further, (5.37) implies that

$$e^{\tilde{b}_i(0)} = 1 \quad \text{and} \quad \sigma_i^2(0) = \mathcal{I}(\mathbf{x}_i'\theta_n), \quad i \in \mathbb{N},$$

in the conditions (5.40), (5.41) of Theorem 5.4 so that they reduce to

$$\Sigma_n = \frac{1}{n}\sum_{i=1}^{n} (\mathcal{I}(\mathbf{x}_i'\theta_0))^3 \mathbf{x}_i\mathbf{x}_i' \to \Sigma \tag{5.48}$$

and

$$\Phi_n = \frac{1}{n}\sum_{i=1}^{n} (\mathcal{I}(\mathbf{x}_i'\theta_0))^2 \mathbf{x}_i\mathbf{x}_i' \to \Phi \tag{5.49}$$

for some positive definite matrices $\Sigma$ and $\Phi$. The remaining condition of Theorem 5.4 takes on the form

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} (Y_i - \mu(\mathbf{x}_i'\theta_0))\, \mathcal{I}(\mathbf{x}_i'\theta_0)\mathbf{x}_i \xrightarrow{\mathcal{L}} N(0, \Sigma) \tag{5.50}$$

for $\Sigma$ figuring in (5.48). We shall show that (5.50) follows from (5.47) and (5.48). Indeed, then $\vartheta_i = \mathbf{x}_i'\theta_0$ and $\mathcal{I}_i = \mathcal{I}(\vartheta_i)$ are uniformly bounded for $i \in \mathbb{N}$. Hence if $t \to 0$ then, uniformly for $i \in \mathbb{N}$,

$$c(\vartheta_i + t\,\mathcal{I}_i) = c(\vartheta_i) + \mu(\vartheta_i)\, t\,\mathcal{I}_i + \mathcal{I}_i^2\,\frac{t^2}{2} + o(t^2).$$

Further, for every $\xi \in \mathbb{R}$,

$$\mathsf{E}\exp\{(Y_i - \mu(\vartheta_i))\,\mathcal{I}_i\,\xi/\sqrt{n}\} = c(\vartheta_i + \xi\,\mathcal{I}_i/\sqrt{n}) - c(\vartheta_i)\,\xi\,\mathcal{I}_i/\sqrt{n}.$$

It follows from here that the moment generating functions $M_n(\tau) = \mathsf{E}\exp\{Z_n'\tau\}$, $\tau = (\tau_1, \dots, \tau_m) \in \mathbb{R}^m$, of the random variables

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - \mu(\varphi_i))\, \mathcal{I}_i\, \mathbf{x}_i, \quad n \in \mathbb{N},$$

converge under (5.47) and (5.48) pointwise to

$$M(\tau) = \exp\left\{\frac{1}{2}\tau\, \Sigma\, \tau'\right\},$$

which suffices for (5.50). Therefore the following statement holds.

**Corollary 5.5** *Let a generalized linear model* (5.45) *satisfy* (5.47) – (5.49)*. If the $L_2$-estimator $\widehat{\theta}_n^{(0)}$ of a true parameter $\theta_0 \in \mathbb{R}^m$ is consistent, then it is asymptotically normal in sense of* (5.43)*, where $\Sigma$ and $\Phi$ are the matrices appearing in* (5.48) *and* (5.49)*.*

Note that under the weak convergence of probability measures

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{\mathbf{x}_i} \Rightarrow \mu$$

of Dirac's probability measures $\delta_{\mathbf{x}_i}$ on the regressor space $\mathcal{X}$, the conditions (5.48), (5.49) hold for

$$\Sigma = \int_{\mathcal{X}} \mathcal{I}(\mathbf{x}'\theta_0)^2 \mathbf{x}\mathbf{x}'\mu(d\mathbf{x}), \qquad \Phi = \int_{\mathcal{X}} \mathcal{I}(x'\theta_0)\mathbf{x}\mathbf{x}'\mu(d\mathbf{x}).$$

Similarly, the conditions (5.40), (5.41) hold but the formulas for the limit matrices are more complicated. Let us also note that in the generalized linear models of Corollary 5.5, none of the estimators $\widehat{\theta}_n^{(\alpha)}$, $\alpha \in \mathbb{R}$, is in general the MLE. Below is studied a special where $\widehat{\theta}_n^{(0)}$ is the MLE.

A similar asymptotic normality result as presented by Corollary 5.5 has been proved for the MLE in generalized linear models with natural link functions in Theorem 3 of Fahrmeir and Kaufmann [5]. The conditions of that theorem are weaker but less easily verifiable than the conditions (5.47) – (5.49) of Corollary 5.5, and the theorem does not provide the asymptotic variance-covariance matrix. Therefore the two results are not directly comparable.

The power of Theorem 2.3 has been verified in Section 4 by an application to the linear and nonlinear regression models. Another verification of this power can be obtained by an application to the model with observations $Y_i$ with distributions from a natural exponential family. In this special case the $L_{2+\alpha}$-estimators

$$\hat{\vartheta}_n^{(\alpha)} \sim \langle t\, e^{\alpha t};\, \dot{c}(\vartheta + \alpha)\rangle, \quad \alpha \in \mathbb{R}, \tag{5.51}$$

estimate a true value $\vartheta_0 \in \mathbb{R}$ of the parameter of these distributions and $\hat{\vartheta}_n^{(0)}$ is the MLE. The estimators (5.51) are special cases of estimators (5.24) obtained for $\Theta = \mathbb{R}$ and $\phi(\mathbf{x}, \theta) = \theta$, so that $\phi_i = \vartheta_0$ and $\dot{\phi}_i = 1$ in the formulas above. Consequently, in (5.33) and (5.35).

$$\begin{aligned} b_i(\alpha) &= c(\vartheta_0 + 2\alpha) - c(\vartheta_0) - 2\alpha\dot{c}(\vartheta_0 + \alpha) =: b(\alpha), \\ \tilde{b}_i(\alpha) &= c(\vartheta_0 + \alpha) - c(\vartheta_0) - \alpha\dot{c}(\vartheta_0 + \alpha) =: \tilde{b}(\alpha), \end{aligned}$$

and in (5.37)

$$\sigma_i^2(\alpha) = e^{b(\alpha)}\left[\mathcal{I}(\vartheta_0 + 2\alpha) + (\mu(\vartheta_0 + 2\alpha) - \mu(\vartheta_0 + \alpha))^2\right] =: \sigma^2(\alpha).$$

Therefore (5.48) and (5.49) hold for

$$\Sigma_n = \Sigma = \sigma^2(\alpha)\,(\mathcal{I}(\vartheta_0 + \alpha))^2$$

and

$$\Phi_n = \Phi = e^{\tilde{b}(\alpha)}(\mathcal{I}(\vartheta_0 + \alpha))^2,$$

so that in (5.43) we have

$$\Phi^{-1}\Sigma\,\Phi^{-1} = \frac{e^{c(\vartheta_0+2\alpha)-c(\vartheta_0)}[\mathcal{I}(\vartheta_0 + 2\alpha) + (\mu(\vartheta_0 + 2\alpha) - \mu(\vartheta_0 + \alpha))^2]}{[e^{c(\vartheta_0+\alpha)-c(\vartheta_0)}\mathcal{I}(\vartheta_0 + \alpha)]^2} = s^2(\alpha). \tag{5.52}$$

The assumptions of Theorem 5.4 hold except the consistency which is clarified in the next proposition where we assume $T = \mathbb{R}$ for simplicity.

**Proposition 5.6** *For every exponential family under consideration, the estimators $\hat{\vartheta}_n^{(\alpha)}$ defined by (5.51) are consistent, with values uniquely given by the formula*

$$\mu(\tilde{\vartheta}_n^{(\alpha)} + \alpha) = \frac{\sum_{i=1}^n Y_i\, e^{\alpha Y_i}}{\sum_{i=1}^n e^{\alpha Y_i}}, \quad n \in \mathbb{N}, \tag{5.53}$$

*for $\mu(t) = \dot{c}(t)$ strictly increasing on $\mathbb{R}$.*

**Proof:** Let $\alpha \in \mathbb{R}$ and $n \in \mathbb{N}$ be arbitrary fixed. By definition, $\hat{\vartheta}_n^{(\alpha)}$ minimizes

$$M_n(\vartheta) = \sum_{i=1}^n \rho(Y_i - \mu(\vartheta)), \tag{5.54}$$

where $\rho(t)$ is given by (5.8). If $\alpha = 0$ then the assertion is obvious. Suppose that $\alpha \neq 0$. Since $\mu(t)$ is infinitely differentiable on $\mathbb{R}$, we can consider the derivatives

$$\begin{aligned} \dot{M}_n(\vartheta) = \frac{d}{d\vartheta}M_n(\vartheta) &= \dot{\mu}(\vartheta)\sum_{i=1}^n \psi(Y_i - \mu(\vartheta)) \\ &= \dot{\mu}(\vartheta)\, e^{-\alpha\mu(\vartheta)}\sum_{i=1}^n (Y_i - \mu(\vartheta))\, e^{-\alpha Y_i} \end{aligned}$$

and

$$\ddot{M}_n(\vartheta) = \frac{d^2}{d\vartheta^2} M_n(\vartheta) = \ddot{\mu}(\vartheta) \sum_{i=1}^{n} \psi(Y_i - \mu(\vartheta)) + (\dot{\mu}(\vartheta))^2 Z_n(\vartheta),$$

where

$$\begin{aligned} Z_n(\vartheta) &= \sum_{i=1}^{n} \dot{\psi}(Y_i - \mu(\vartheta)) \\ &= \alpha \sum_{i=1}^{n} \psi(Y_i - \mu(\vartheta)) + \sum_{i=1}^{n} e^{\alpha(Y_i - \mu(\vartheta))}. \end{aligned}$$

By (5.18), $\dot{\mu}(\vartheta)$ is the Fisher information $\mathcal{I}(\vartheta) > 0$ for all $\vartheta \in \mathbb{R}$. Therefore $\hat{\vartheta}_n^{(\alpha)}$ given by (5.53) is the only solution of the equation $\dot{M}_n(\vartheta) = 0$. Further,

$$\begin{aligned} \ddot{M}_n(\hat{\vartheta}_n^{(\alpha)}) &= (\mathcal{I}(\hat{\vartheta}_n^{(\alpha)}))^2 Z_n(\hat{\vartheta}_n^{\alpha}) \\ &= (\mathcal{I}(\hat{\vartheta}_n^{(\alpha)}))^2 \sum_{i=1}^{n} e^{\alpha(Y_i - \mu(\hat{\vartheta}_n^{(\alpha)}))} > 0, \end{aligned}$$

so that $\hat{\vartheta}_n^{(\alpha)}$ is a unique local minimum of $M_n(\vartheta)$ on $\mathbb{R}$. We shall prove the relation

$$M_n(\vartheta) \geq M_n(\hat{\vartheta}_n^{(\alpha)}), \quad \vartheta \in \Theta, \tag{5.55}$$

which implies that $\hat{\vartheta}_n^{(\alpha)}$ is a unique global minimum of $M_n(\vartheta)$ on $\mathbb{R}$, i. e. that the second half of Proposition 5.6 is valid. By (5.54) and (5.8), for every $\vartheta \in \mathbb{R}$,

$$M_n(\vartheta) = \frac{1}{\alpha^2} \left( 1 - \Gamma_n(\vartheta) \sum_{i=1}^{n} e^{\alpha Y_i} \right)$$

where

$$\Gamma_n(\vartheta) = \left( 1 + \alpha \left[ \mu(\vartheta + \alpha) - \mu(\hat{\vartheta}_n^{(\alpha)} + \alpha) \right] \right) e^{-\alpha\mu(\vartheta+\alpha)}.$$

Therefore (5.55) holds if

$$\Gamma_n(\vartheta) \leq \Gamma_n(\hat{\vartheta}_n^{(\alpha)}) = e^{-\alpha\mu(\hat{\vartheta}_n^{(\alpha)}+\alpha)}, \quad \theta \in \Theta,$$

i. e. if $\Delta_n(\vartheta) = \mu(\vartheta + \alpha) - \mu(\hat{\vartheta}_n^{(\alpha)} + \alpha)$ satisfies the relation

$$1 + \alpha\Delta_n(\vartheta) \leq e^{\alpha\Delta_n(\vartheta)}, \quad \vartheta \in \Theta.$$

This completes the proof of the second half of Proposition 5.6. The first part (consistency of $\hat{\vartheta}_n^{(\alpha)}$) follows, via the strict monotonicity and continuity of $\mu(t)$, from the fact that, by (5.53) and the law of large numbers,

$$\mu(\hat{\vartheta}_n^{(\alpha)} + \alpha) \xrightarrow{P} \mu(\vartheta + \alpha) \quad \text{as } n \to \infty.$$

□

By combining Proposition (5.6) with what has been said before, we obtain the following result.

**Proposition 5.7** *Assume $T = \mathbb{R}$ for the exponential family (5.10). The estimators (5.51) of a true parameter $\vartheta_0 \in \mathbb{R}$ are explicitly given by formula (5.53). They are consistent and asymptotically normal in the sense that, for all $\alpha \in \mathbb{R}$,*

$$\sqrt{n}(\hat{\vartheta}_n^{(\alpha)} - \vartheta_0) \xrightarrow{\mathcal{L}} N(0, s^2(\alpha)) \quad as\ n \to \infty, \tag{5.56}$$

*where $s^2(\alpha)$ is given by (5.52).*

The asymptotic normality result (5.56) was obtained from the theory of Section 2 under the same generality as it can be obtained by a direct analysis of the concrete class of estimators

$$\hat{\vartheta}_n^{(\alpha)} = \mu^{-1}\left(\frac{\sum_{i=1}^n Y_i e^{\alpha Y_i}}{\sum_{i=1}^n e^{\alpha Y_i}}\right) - \alpha, \quad \alpha \in \mathbb{R}. \tag{5.57}$$

None of the assumptions of this theory imposed a superfluous restriction on the model or $\alpha$. Again, this verifies in some sense that the general theory is strong enough to deal with concrete situations.

In the rest of section we study the exponential $L_{2+\alpha}$-estimators of parameters of two well known exponential families.

**Example 5.8** Let the family (3.2) be standard normal with a location parameter $\vartheta \in \mathbb{R}$. Then

$$c(\vartheta) = \frac{\vartheta^2}{2}, \quad \mu(\vartheta) = \vartheta, \quad \mathcal{I}(\vartheta) = 1, \quad \vartheta \in \mathbb{R}, \tag{5.58}$$

and the dominating measure $\nu$ is the standard normal probability measure. By (5.57), the exponential $L_{2+\alpha}$-estimates are given by the formula

$$\hat{\vartheta}_n^{(\alpha)} = \frac{\sum_{i=1}^n Y_i\, e^{\alpha Y_i}}{\sum_{i=1}^n e^{\alpha Y_i}} - \alpha, \quad \alpha \in \mathbb{R}, \tag{5.59}$$

and, by (5.52) and (5.58), $s^2(\alpha) = 1 + \alpha^2$. By Proposition 5.7, the estimators (5.59) are asymptotically normal with asymptotic mean 0 and asymptotic variances $1 + \alpha^2$. If instead of the standard normal law $f(y|\vartheta_0)$ under consideration, the observations are governed by

$$(1 - \varepsilon)\, f(y|\vartheta_0) + \varepsilon\, f(y|\vartheta_0, \sigma), \quad 0 < \varepsilon < 1, \tag{5.60}$$

where $f(y|\vartheta_0, \sigma)$ is a normal density with location $\vartheta_0$ and scale $\sigma > 0$, then

$$b(\vartheta_0|\alpha, \varepsilon) = \frac{\varepsilon\, \vartheta_0 \sigma(\sigma - 1) \exp\left\{\frac{1}{2}\left[(\vartheta_0\sigma + \alpha)^2 - \vartheta_0^2\right]\right\}}{(1 - \varepsilon)\exp\{2\vartheta_0\alpha\} + \varepsilon\sigma \exp\left\{\frac{1}{2}\left[(\vartheta_0\sigma + \alpha)^{2-\vartheta_0^2}\right]\right\}}$$

is the asymptotic bias of $\hat{\vartheta}_n^{(\alpha)}$. By a suitable choice of $\alpha \neq 0$, this bias can be held at a considerably lower levels over an a priori expected domain of $\vartheta_0$ than is the level due to the MLE $\hat{\vartheta}_n^{(0)}$.

**Example 5.9** Let the family (3.2) be Poisson with a parameter $\vartheta = \ln\lambda \in \mathbb{R}$. Then

$$c(\vartheta) = \mu(\vartheta) = \mathcal{I}(\vartheta) = e^{\vartheta}, \quad \vartheta \in \mathbb{R}, \tag{5.61}$$

and the dominating measure $\nu$ on $\mathbb{R}$ is finite and discrete,

$$\nu = \sum_{k=0}^{\infty} \frac{\delta_k}{k!},$$

where $\delta_k$ is the Dirac measure concentrating the mass 1 at the point $k \in \mathbb{R}$. In this case, by (5.57), the exponential $L_{2+\alpha}$-estimates are given by the formula

$$\hat{\vartheta}_n^{(\alpha)} = \ln \frac{\sum_{i=1}^n Y_i\, e^{\alpha Y_i}}{\sum_{i=1}^n e^{\alpha Y_i}} - \alpha \tag{5.62}$$

and, by (5.52) and (5.58),

$$s^2(\alpha) = \exp\left\{e^{\vartheta_0+\alpha}(e^{\alpha}-1) - \vartheta_0\right\}\left[1 + e^{\vartheta_0}(e^{\alpha}-1)^2\right]. \tag{5.63}$$

Therefore, by Proposition 5.7, the estimators defined by (5.62) are asymptotically normal with asymptotic mean 0 and asymptotic variances (5.63). This means that

$$\sqrt{n}\left(e^{\hat{\vartheta}_n^{(\alpha)}} - e^{\vartheta_0}\right) = e^{\vartheta_0}\sqrt{n}\left(e^{\hat{\vartheta}_n - \vartheta_0} - 1\right)$$

tends in law to

$$N\left(0, \exp\left\{e^{\vartheta_0+\alpha}(e^{\alpha+1}) + \vartheta_0\right\}\left[1 + e^{\vartheta_0}(e^{\alpha}-1)^2\right]\right),$$

i. e., that the exponential $L_{2+\alpha}$-estimators $\hat{\lambda}_n^{(\alpha)}$ of $\lambda_0 = e^{\vartheta_0}$ are asymptotically normal in the sense

$$\sqrt{n}\left(\hat{\lambda}_n^{(\alpha)} - \lambda_0\right) \xrightarrow{\mathcal{L}} N\left(0, \lambda_0 \exp\{\lambda_0 e^{\alpha}(e^{\alpha}-1)\}\left[1 + \lambda_0(e^{\alpha}-1)^2\right]\right).$$

If instead of the Poisson distribution $F(y|\vartheta)$ under consideration the observations are distributed by

$$(1-\varepsilon)\, F(y|\vartheta) + \varepsilon G(y), \quad 0 < \varepsilon < 1, \tag{5.64}$$

where

$$G(y) = \zeta(2)^{-1} \sum_{k=1}^{\infty} \frac{1}{k^2}\, I(y > k)$$

and $\zeta(s)$, $s > 1$, is the Riemann function, then the asymptotic bias of the MLE $\hat{\vartheta}_n^{(0)}$ is infinite for arbitrarily small $\varepsilon$. Indeed, if $\tilde{Y}_i$ are observations i.i.d. by (5.64) then

$$E\,\tilde{Y}_i = (1-\varepsilon)\, e^{\vartheta_0} + \varepsilon\zeta(2)^{-1} \sum_{k=1}^{\infty} \frac{k}{k^2} = \infty.$$

On the other hand, the asymptotic bias $b(\vartheta_0|\alpha,\varepsilon)$ of every estimator $\hat{\vartheta}_n^{(\alpha)}$ with $\alpha < 0$ satisfies the relation

$$\lim_{\varepsilon \downarrow 0} b(\vartheta_0|\alpha,\varepsilon) = 0 \quad \text{for every } \vartheta_0 \in \mathbb{R}.$$

The results in the last two examples demonstrate that in the class of $L_{2+\alpha}$-estimators one can find more robust alternatives to the $L_2$-estimator (MLE). The price payed for the robustness is a larger asymptotic variance when the observations are not contaminated.

## 6 Proof of Theorems 2.2 and 2.3

Unless otherwise explicitly stated, we consider in this section arbitrary model (1.1) and $M$-estimator $\hat{\theta}_n \sim \langle \psi; \varphi_i \rangle$ where $\psi$ can be decomposed as the difference (2.1) of two nondecreasing functions $\psi^+$ and $\psi^-$. We suppose for simplicity that both these functions are right-continuous. Then also their sum $\psi^{\pm}$ introduced in (2.2) and $\psi$ itself are right continuous. We shall formulate a series of auxiliary statements leading to the proofs of the Theorems 2.2 and 2.3. All statements refer to the concepts and conditions introduced in Sections 1 and 2. Most of these statements are technical but some of them are interesting also from the statistical point of view.

If $\xi : \mathbb{R} \mapsto \mathbb{R}$ is nondecreasing and right continuous then there exists unique measure $\mu_\xi$ on the Borel subsets of $\mathbb{R}$ associated with $\xi$ and satisfying relation $\mu(a, b]) = \xi(b) - \xi(a)$ for all real numbers $a < b$. If $\phi : \mathbb{R} \mapsto \mathbb{R}$ is measurable then the Lebesgue-Stieltjes integral is defined as the Lebesgue integral for the associated measure, e. g.

$$\int_{(a,b]} \phi(s)\, d\xi(s) = \int_{(a,b]} \phi(s)\, \mu_\xi(ds).$$

If $\eta$ is another monotone right continuous function then the bivariate Lebesgue–Stieltjes integral

$$\int_{(a,b]^2} \phi(s,t)\, d\xi(s)\, d\eta(t) \tag{6.1}$$

can be defined by means of the associated measure $\mu_\xi \otimes \mu_\eta$ on the Borel subsets of $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. For locally bounded functions $\phi(s)$ (e. g. for linear combinations of monotone functions), and for differences $\xi(s) = \xi^+(s) - \xi^-(s)$ of two nondecreasing right-continuous functions, one can define the Lebesgue–Stieltjes integral

$$\int_{(a,b]} \phi(s)\, d\xi(s) = \int_{(a,b]} \phi(s)\, d\xi^+(s) - \int_{(a,b]} \phi(s)\, d\xi^-(s).$$

If $\eta = \eta^+ - \eta^-$ is a similar difference then one can similarly extend the bivariate Lebesgue–Stieltjes integrals (6.1). Using the bounded measurable function

$$\phi(s,t) = I(a < t \le b)\, I(a < s \le t) = I(a < s \le b) I(s \le t \le b)$$

defined by means of the indicator function $I(\cdot)$, and employing equalities of the type

$$\int I(a < s \le t)\, d\xi(s) = \xi(t) - \xi(a),$$

one obtains from the Fubini theorem the *per partes* rule

$$\int_{(a,b]} \eta(s)\, d\xi(s) + \int_{(a,b]} \xi(s-)\, d\eta(s) = \xi(b)\,\eta(b) - \xi(a)\,\eta(a) \tag{6.2}$$

for Lebesgue–Stieltjes integrals. In this rule, $\xi(s-)$ denotes the left continuous version of $\xi(s)$.

Our first statement is concerning the criterion function $\rho$ satisfying according to (1.7) for all $y \in \mathbb{R}$ the relation

$$\rho(y) = \rho(0) + \int_{(0,y]} \psi(s)ds. \tag{6.3}$$

**Proposition 6.1** *For all $y$, $t \in \mathbb{R}$ holds the generalized Taylor formula*

$$\rho(y+t) = \rho(y) + \psi(y)\, t + R(y,t) \tag{6.4}$$

*where the remainder is*

$$R(y,t) = \int_y^{y+t} (y+t-s)\, d\psi(s). \tag{6.5}$$

**Proof:** By (1.7) and the *per partes* rule (6.2),

$$\begin{aligned} \rho(y+t) - \rho(y) &= \int_{(y,y+t]} \psi(s)\, ds \\ &= \psi(y)\, t + \int_{(y,y+t]} (y+t-s)\, d\psi(s). \end{aligned}$$

□

By applying the generalized Taylor formula (6.4) in (1.3) we obtain

$$\begin{aligned} M_n(\theta) - M_n(\theta_n) &= \frac{1}{n}\sum_{i=1}^n [\rho(Y_i - \varphi_i(\theta)) - \rho(Y_i - \varphi_i(\theta_0))] \\ &= \frac{1}{n}\sum_{i=1}^n \psi(X_i)\, t_i + \frac{1}{n}\sum_{i=1}^n R(X_i, t_k) \end{aligned} \tag{6.6}$$

where $X_i = Y_i - \varphi_i(\theta_0)$ and $t_i = \varphi_i(\theta) - \varphi_i(\theta_0)$. The first sum in the last row is linear in $t_i$. Therefore we are interested in the behavior of the expected remainders $\mathsf{E}R(X_i, t)$ in a neighborhood of $t = 0$.

**Proposition 6.2** *Let the regularity condition* (R3) *hold and let $X_i = Y_i - \varphi_i(\theta_0)$. Then the expectations $\mathsf{E}R(X_i, t)$ are locally quadratic in the sense that, for the functions $h_i : (-\tau_0, \tau_0) \mapsto \mathbb{R}$ introduced in* (R3) *and all $0 < \tau < \tau_0$*

$$\sup_{|t|\le\tau} \left| \mathsf{E}R(X_i,\, t) - h_i(0)\, \frac{t^2}{2} \right| \le \frac{t^2}{2}\omega(h_i, \tau). \tag{6.7}$$

**Proof:** Consider $t \in (-\tau_0, \tau_0)$. If (R3) holds then, by the Fubini theorem and (6.4),

$$\begin{aligned}
\mathsf{E}R(X_i,t) &= \int_0^t \mathsf{E}\psi(X_i+s)\,ds - t\mathsf{E}\psi(X_i) \\
&= \int_0^t [H_i(s)-H_i(0)]ds \\
&= \int_0^t\int_0^s h_i(u)duds = \int_0^t (t-u)h_i(u)du \\
&= \int_0^t (t-u)h_i(0)du + \int_0^t (t-u)[h_i(u)-h_i(0)]du.
\end{aligned}$$

The rest is clear from here and from the definition of $\omega(h_i,\tau)$. □

The next result estimates fluctuations of the remainders $R(X_i,t)$ around $\mathsf{E}\,R(X_i,t)$.

**Proposition 6.3** *If the regularity condition* (R4+) *holds then for* $\tau_0$, $q$ *and* $\kappa$ *considered in* (R4+), *and for* $X_i = Y_i - \varphi_i(\theta_0)$ *and all* $0<\tau<\tau_0$,

$$\sup_{n\in\mathbb{N}} \frac{1}{n}\sum_{i=1}^n \mathsf{E}\sup_{|t|\le\tau}(R(X_i,t))^2 < \kappa\,\tau^{2+q}. \tag{6.8}$$

**Proof:** Let $y\in\mathbb{R}$ be arbitrary fixed. By substitution $y+t\mapsto t$ and the convention (1.8), it follows from (6.5)

$$R(y,t) = \int_{(0,t]}(t-s)\,d\psi(s+y) = \int_{(t^-,t^+]}|t-s|\,d\psi(s+y)$$

where $t^- = \min\{0,t\}$ and $t^+ = \max\{0,t\}$. Hence for every $t\in\mathbb{R}$

$$\begin{aligned}
|R(y,t)| &= \left|\int_{(t^-,t^+]}|t-s|\,d\psi^+(s) - \int_{(t^-,t^+]}|t-s|\,d\psi^-(s)\right| \\
&\le \left|\int_{(t^-,t^+]}|t-s|\,d\psi^{\pm}(y+s)\right| \quad \text{(cf. 2.2)} \\
&\le |t|\left[\psi^{\pm}(y+t^+) - \psi^{\pm}(y-t^-)\right] \\
&\le |t|\left[\psi^{\pm}(y+|t|) - \psi^{\pm}(y-|t|)\right].
\end{aligned}$$

Consequently,

$$\sup_{|t|\le\tau}(R(y,t))^2 \le t^2\left[\psi^{\pm}(y+\tau)-\psi^{\pm}(y-\tau)\right]^2$$

and (6.8) follows from (R4+). □

Next follows an important technical result which is sharper than a similar result in [21] and which is proved by a different method. Consider closed balls $B_\gamma \subset \mathbb{R}^m$ of diameters

$0 < \gamma \leq \delta$, $\delta \leq \infty$, centered at $0 \in \mathbb{R}^m$, and a sequence $S_1(\mathbf{u})$, $S_2(\mathbf{u}), \dots$ of continuous independent zero-mean random processes $(S_i(\mathbf{u}) : \mathbf{u} \in B_\delta)$ with $S_1(0) = S_2(0) = \cdots = 0$. For given $0 < \gamma \leq \delta$ and $n \in \mathbb{N}$, we shall estimate the expected modulus of continuity

$$\Omega_n(\gamma) = \mathsf{E} \sup_{u \in B_\gamma} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_i(u) \right| \tag{6.9}$$

of the normalized sum at $\mathbf{u} = 0$. A useful estimate will be obtained by means of the theory of empirical processes, in particular by the results in Chapter 2 of [30]. We suppose that for some $\delta > 0$

$$|S_i(\mathbf{u}) - S_i(\widetilde{\mathbf{u}})| \leq \Lambda_i \|\mathbf{u} - \widetilde{\mathbf{u}}\| \quad \text{for all} \quad \mathbf{u}, \widetilde{\mathbf{u}} \in B_\delta,\ i \in \mathbb{N}, \tag{6.10}$$

and

$$\sup_{n \in \mathbb{N}} \mathsf{E} \frac{1}{n} \sum_{i=1}^{n} L_n^2 < \infty \quad \text{for} \quad L_n = \left( \frac{1}{n} \sum_{i=1}^{n} \Lambda_i^2 \right)^{1/2} \tag{6.11}$$

Set

$$\Gamma_n(\gamma) = \sup_{\mathbf{u} \in B_\gamma} \left[ \frac{1}{n} \sum_{i=1}^{n} S_i^2(\mathbf{u}) \right]^{1/2}.$$

**Proposition 6.4** *Suppose that $S_1(\mathbf{u}), S_2(\mathbf{u}), \dots$ are continuous , independent zero-mean stochastic processes continuous on $B_\delta$ with $S_i(0) = 0$. If the condition (6.10) and (6.11) hold then there exists a universal constants $K$ and $\kappa(d)$ such that for $\gamma \leq \delta$*

$$\mathsf{E} \sup_{\|\mathbf{u}\| \leq \gamma} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_i(\mathbf{u}) \right| \leq \gamma \mathsf{E} \left[ L_n \Gamma \left( d, \frac{2\Gamma_n(\gamma)}{\delta L_n} \right) \right] \tag{6.12}$$

*and*

$$\mathsf{E} \sup_{\|\mathbf{u}\| \leq \delta} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_i(\mathbf{u}) \right| \leq \delta \Gamma\left(d, 2\right) \mathsf{E} L_n \tag{6.13}$$

*where*

$$\Gamma(d, s) = 2K \int_0^s \sqrt{|\ln(\kappa(d) t^d)|} dt.$$

*For every $0 < \alpha < 1$ there exists a constant $C(\alpha, d)$ such that*

$$\mathsf{E} \sup_{\|\mathbf{u}\| \leq \gamma} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_i(\mathbf{u}) \right| \leq C(\alpha, d) \gamma^\alpha \Gamma^\alpha(d, \gamma) \mathsf{E} L_n^{1-\alpha}. \tag{6.14}$$

**Proof:** Suppose $\varepsilon_1, \dots, \varepsilon_n$ are independent binary random variables taking on the values 1 and $-1$ with equal probability 1/2. Assume that for $n = 1, 2, \dots$ the set $A_n \subset \mathbb{R}^n$ is bounded with respect to the Euclidean distance $\|\cdot\|_n$ on $\mathbb{R}^n$. Denote by $\mathbf{N}(\varepsilon, A_n)$ the minimal number of balls of radius $\varepsilon > 0$ covering $A_n$. Then by Corollary 2.2.8 in [30] there is a universal constant $K$ such that

$$\mathsf{E} \sup_{(a_1,...,a_n)\in A_n} \left|\sum_{i=1}^{n} a_i\varepsilon_i\right| \le K \int_0^\infty \sqrt{\ln \mathbf{N}\left(\frac{\varepsilon}{2}, A_n\right)} d\varepsilon. \tag{6.15}$$

The symmetrization Lemma 2.3.6 of [30] yields

$$\mathsf{E} \sup_{\|\mathbf{u}\|\le\gamma} \left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_i(\mathbf{u})\right| \le \mathsf{E}\left(\mathsf{E}_\varepsilon \sup_{\|\mathbf{u}\|\le\gamma} \left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_i(\mathbf{u})\varepsilon_i\right|\right) \tag{6.16}$$

where the $\varepsilon_1, ..., \varepsilon_n$ are independent Bernoulli variables which are independent of the processes $S_1(\mathbf{u}), ..., S_n(\mathbf{u})$ and take on the values 1 and $-1$ with probability 1/2. The symbol $\mathsf{E}_\varepsilon$ denotes the expectation w.r.t. $\varepsilon_1, ..., \varepsilon_n$. To estimate the right hand term we suppose that the processes $S_1(\mathbf{u}), ..., S_n(\mathbf{u})$ and random variables $\varepsilon_1, ..., \varepsilon_n$ are defined on a product space, say $(\Omega_1 \times \Omega_2, \mathfrak{F}_1 \otimes \mathfrak{F}_2, \mathbb{P}_1 \times \mathbb{P}_2)$ where the processes depend on $\omega_1 \in \Omega_1$ and the binary variables depend on $\omega_2 \in \Omega_2$. Fix $\omega_1 \in \Omega_1$ and introduce

$$A_{n,\gamma}(\omega_1) = \left\{\frac{1}{\sqrt{n}} S_1(\mathbf{u},\omega_1), ..., \frac{1}{\sqrt{n}} S_n(\mathbf{u},\omega_1), \mathbf{u} \in B_\gamma\right\} \subseteq \mathbb{R}^n.$$

For fixed $\omega_1$ we estimate the entropy number appearing in (6.15). The Lipschitz condition (6.10) implies that for every $\varepsilon$-net for $B_\gamma$ there is an $L_n\varepsilon-$net for $A_{n,\gamma}(\omega_1)$. For $\gamma \le \delta$ the entropy number of $B_\gamma$ does not exceed $\kappa(d)(\frac{\gamma}{\varepsilon})^d$ where $\kappa(d)$ is a constant depending on $d$ only. As the diameter of $A_{n,\gamma}$ does not exceed, $2\Gamma_n(\gamma)$ we have

$$\mathbf{N}\left(\frac{\varepsilon}{2}, A_{n,\gamma}\right) \le \begin{cases} \kappa(d)\left[\frac{\gamma}{\varepsilon}\right]^d (2L_n)^d & \text{for} \quad \varepsilon \le 4\Gamma_n(\gamma) \\ 1 & \text{for} \quad \varepsilon > 4\Gamma_n(\gamma) \end{cases} \tag{6.17}$$

and

$$\begin{aligned} \mathsf{E} \sup_{(a_1,...,a_n)\in A_{n,\delta}} \left|\sum_{i=1}^{n} a_i\varepsilon_i\right| &\le K \int_0^{4\Gamma_n(\gamma)} \sqrt{\left|\ln\left[\kappa(d)\left(\frac{2L_n\gamma}{\varepsilon}\right)^d\right]\right|} d\varepsilon \\ &= K2\delta L_n \int_0^{2\Gamma_n(\gamma)/(\delta L_n)} \sqrt{|\ln(\kappa(d)t^d)|} dt. \end{aligned}$$

To complete the proof we set

$$\Gamma(d,s) = 2K \int_0^s \sqrt{|\ln(\kappa(d)t^d)|} dt,$$

and obtain (6.12). To prove (6.13) it suffices to observe that the assumption (6.10) yields $\Gamma_n(\gamma)/(\delta L_n) \le 1$. Using the inequality

$$\ln x \le \frac{x^{1-\alpha}}{1-\alpha} \quad \text{for} \quad x \ge 1 \text{ and } 0 < \alpha < 1$$

we find a constant $C(d,\alpha)$ such that

$$2K\int_0^s \sqrt{|\ln(\kappa(d)t^d)|}dt \le C(d,\alpha)s^{1-\alpha},$$

which proves the statement (6.14). □

In the next proposition we assume that the adaptation condition (2.3) and the regularity conditions (R1), (R2) and (R4) hold. We introduce the local parameter

$$\mathbf{u} = \theta - \theta_0 \in B_\delta = \{\mathbf{u} \in \mathbb{R}^m : \|\mathbf{u}\| \le \delta\}$$

where $\delta > 0$ is the same as in the definition (2.7) of the ball $B$ in (R2). In the proposition we study the zero-mean version $D_n(\mathbf{u}) - \mathsf{E}\, D_n(\mathbf{u})$ of the random process $(D_n(\mathbf{u}) : \mathbf{u} \in B_\delta)$ defined by

$$D_n(\mathbf{u}) = \sqrt{n}\,(M_n(\theta_0+\mathbf{u}) - M_n(\theta_0))\,. \tag{6.18}$$

To simplify formulas we use the notations

$$\varphi_i = \varphi_i(\theta_0), \quad \xi_i(\mathbf{u}) = \varphi_i(\theta_0+\mathbf{u}) - \varphi_i, \quad X_i = Y_i - \varphi_i, \quad \dot\varphi_i = \dot\varphi_i(\theta_0). \tag{6.19}$$

By (1.3) and (6.4),

$$\begin{aligned} D_n(\mathbf{u}) &= \frac{1}{\sqrt{n}}\left(\sum_{i=1}^n [\rho(X_i - \xi_i(\mathbf{u})) - \rho(X_i)]\right) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(X_i)\,\xi_i(\mathbf{u}) + \frac{1}{\sqrt{n}}\sum_{i=1}^n R_i(\mathbf{u}) \end{aligned}$$

where we put for simplicity

$$R_i(\mathbf{u}) = R(X_i, \xi_i(\mathbf{u})), \quad \mathbf{u} \in B_\delta.$$

It follows that

$$D_n(\mathbf{u}) = \mathcal{L}_n(\mathbf{u}) + \mathcal{D}_n(\mathbf{u}) + \mathcal{R}_n(\mathbf{u}), \quad \mathbf{u} \in B_\delta,$$

where the linear term $\mathcal{L}_n$, deviation $\mathcal{D}_n$ and remainder $\mathcal{R}_n$ are given by

$$\begin{aligned} \mathcal{L}_n(\mathbf{u}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(X_i)\,(\dot\varphi_i'\mathbf{u}), \qquad (6.20) \\ \mathcal{D}_n(\mathbf{u}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(X_i)[\xi_i(\mathbf{u}) - \dot\varphi_i'\mathbf{u}], \\ \mathcal{R}_n(\mathbf{u}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n R_i(\mathbf{u}). \end{aligned}$$

Since $\widehat{\theta}_n$ is adapted, (2.3) implies $\mathsf{E}\mathcal{L}_n(\mathbf{u}) = \mathsf{E}\mathcal{D}_n(\mathbf{u}) = 0$, so that

$$D_n(\mathbf{u}) - \mathsf{E}\, D_n(\mathbf{u}) = \mathcal{L}_n(\mathbf{u}) + \mathcal{D}_n(\mathbf{u}) + \mathcal{S}_n(\mathbf{u}), \tag{6.21}$$

where

$$\mathcal{S}_n(\mathbf{u}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_i(\mathbf{u}) \quad \text{for} \quad S_i(\mathbf{u}) = R_i(\mathbf{u}) - \mathsf{E}\, R_i(\mathbf{u}). \tag{6.22}$$

**Proposition 6.5** *Let $B_\gamma$ be a zero centered ball of radius $\gamma$ and let the adaption condition* (2.3) *and the regularity conditions* (R1), (R2) *and* (R4) *hold. Then for every $0 < \alpha < 1$ and the above considered processes $\mathcal{L}_n(\mathbf{u})$, $\mathcal{D}_n(\mathbf{u})$, $\mathcal{S}_n(\mathbf{u})$ defined on $B_\delta$ there exist constants $c_0$, $c_1$, $c_2$ such that, for all $n \in \mathbb{N}$,*

$$\mathsf{E} \sup_{\mathbf{u}\in B_\gamma} |\mathcal{L}_n(\mathbf{u})| \le c_0\gamma \quad \textit{if} \quad 0 < \gamma \le \delta, \tag{6.23}$$

$$\mathsf{E} \sup_{\mathbf{u}\in B_\gamma} |\mathcal{D}_n(\mathbf{u})| \le c_1\gamma^2 \quad \textit{if} \quad 0 < \gamma \le \delta, \tag{6.24}$$

$$\mathsf{E} \sup_{\mathbf{u}\in B_\gamma} |\mathcal{S}_n(\mathbf{u})| \le c_2\gamma \quad \textit{if} \quad 0 < \gamma \le \delta. \tag{6.25}$$

*If in addition* (R4+) *holds then for every $0 < \alpha < 1$ there exist a constants $c_3$ and $q > 0$ such that, for all $n \in \mathbb{N}$,*

$$\mathsf{E} \sup_{\mathbf{u}\in B_\gamma} |\mathcal{S}_n(\mathbf{u})| \le c_3\gamma^{1+\alpha q/2} \quad \textit{if} \quad 0 < \gamma \le \delta.$$

**Proof:** The regularity condition (2.8) implies

$$|\varphi_i(\theta_0 + \mathbf{u}) - \varphi_i(\theta_0)| \le \lambda \|\mathbf{u}\| \le \lambda\gamma \qquad \text{for} \quad \gamma \le \delta \tag{6.26}$$

and

$$\varphi_i(\theta_0 + \mathbf{u}) - \varphi_i(\theta_0) - \dot{\varphi}_i'\mathbf{u} = \int_0^1 [\dot{\varphi}_i'(\theta_0 + s\mathbf{u}) - \dot{\varphi}_i'(\theta_0)]\mathbf{u} ds. \tag{6.27}$$

Hence by the Lipschitz continuity of $\dot{\varphi}_i$ required in (2.9)

$$\begin{aligned} |\varphi_i(\theta_0 + \mathbf{u}) - \varphi_i - \dot{\varphi}_i'\mathbf{u}| &\le \|\mathbf{u}\| \int_0^1 \lambda \|\mathbf{u}\|\, s ds \\ &= \frac{\lambda}{2} \|\mathbf{u}\|^2 . \end{aligned} \tag{6.28}$$

By the definition of $\mathcal{L}_n(\mathbf{u})$, for every $\gamma \le \delta$

$$\mathsf{E} \sup_{\mathbf{u}\in B_\gamma} |\mathcal{L}_n(\mathbf{u})| \le \gamma \mathsf{E} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i)\, \dot{\varphi}_i \right\| .$$

By assumption (2.3) we have $\mathsf{E}\psi(X_i) = 0$. Hence it follows from the independence of $X_i$ that

$$\left(\mathsf{E}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(X_i)\,\dot{\varphi}_i\right\|\right)^2 \le \frac{1}{n}\sum_{i=1}^{n}\|\dot{\varphi}_i\|^2\,\mathsf{E}\psi^2(X_i).$$

We see from (3.10) and (2.5) that (6.23) holds for $c_0 = \sqrt{\lambda C}$, where $\lambda$ is the constant figuring in (3.10) and $C$ is the supremum in (2.5). Similarly, by the definition of $\mathcal{D}_n(\mathbf{u})$ and (6.28),

$$\left(\mathsf{E}\sup_{\mathbf{u}\in B_\gamma}|\mathcal{D}_n(\mathbf{u})|\right)^2 \le \frac{\lambda\gamma^2}{2n}\sum_{i=1}^{n}\mathsf{E}(\psi(X_i))^2.$$

To prove (6.25) we shall apply Proposition 6.4. The independent zero-mean processes $(S_i(\mathbf{u}) : \mathbf{u}\in B_\gamma)$, $i\in\mathbb{N}$, satisfy all assumptions of Proposition 6.4. Indeed, since $\xi_i(\mathbf{u}) = \varphi_i(\theta_0 + \mathbf{u}) - \varphi_i$, it holds $S_i(0) = 0$. Using similar arguments as in the proof of Proposition 6.3 we get that the modulus of the function

$$|t-s|I(t^- < s \le t^+) - |\widetilde{t}-s|I(\widetilde{t}^- < s \le \widetilde{t}^+)$$

is for every $t,\widetilde{t}\in\mathbb{R}$ bounded above by $|t-\widetilde{t}|$. Therefore using similar arguments as in the mentioned proof, we obtain that for all $t,\widetilde{t}\in(-\tau,\tau)$

$$\left|R(y,t) - R(y,\widetilde{t})\right| \le |t-\widetilde{t}|\left[\psi^{\pm}(y+\tau) - \psi^{\pm}(y-\tau)\right].$$

It follows from here that the processes $(R_i(\mathbf{u}) : \mathbf{u}\in B_\gamma)$ satisfy for all $\mathbf{u}$, $\widetilde{\mathbf{u}}\in B_\gamma$ the inequalities

$$|R_i(\mathbf{u}) - R_i(\widetilde{\mathbf{u}})| \le \left[\psi^{\pm}(X_i - \varphi_i + \tau_i) - \psi^{\pm}(X_i - \varphi_i - \tau_i)\right]|\xi_i(\mathbf{u}) - \xi_i(\widetilde{\mathbf{u}})|$$

where

$$\tau_i = \tau_i(\mathbf{u},\widetilde{\mathbf{u}}) = \max\left\{|\xi_i(\mathbf{u})|,\ |\xi_i(\widetilde{\mathbf{u}})|\right\}.$$

We get from (6.26) $\tau_i(\mathbf{u},\widetilde{\mathbf{u}}) \le \lambda\delta$, so that the monotonicity of $\psi^{\pm}$ implies

$$\begin{aligned}|R_i(\mathbf{u}) - R_i(\widetilde{\mathbf{u}})| &\le Z_i|\xi_i(\mathbf{u}) - \xi_i(\widetilde{\mathbf{u}})| \\ &= Z_i\left|\int_0^1 [\dot{\varphi}_i(\theta_0 + \widetilde{\mathbf{u}} + s(\mathbf{u}-\widetilde{\mathbf{u}}))]'[\mathbf{u}-\widetilde{\mathbf{u}}]ds\right|\end{aligned}$$

where

$$Z_i = \psi^{\pm}(X_i - \varphi_i + \lambda\delta_0) - \psi^{\pm}(X_i - \varphi_i - \lambda\delta_0).$$

Hence by (6.10)

$$|R_i(\mathbf{u}) - R_i(\widetilde{\mathbf{u}})| \le \lambda Z_i\,\|\mathbf{u}-\widetilde{\mathbf{u}}\|.$$

Thus the zero-mean versions $S_i(\mathbf{u}) = R_i(\mathbf{u}) - \mathsf{E}\,R_i(\mathbf{u})$ satisfy the inequalities

$$|S_i(\mathbf{u}) - S_i(\widetilde{\mathbf{u}})| \le \widetilde{Z}_i\,\|\mathbf{u}-\widetilde{\mathbf{u}}\| \qquad \text{where} \qquad \widetilde{Z}_i = \lambda Z_i + \lambda\mathsf{E}\,Z_i.$$

Note that

$$\mathsf{E}(\widetilde{Z}_i)^2 \le 4\lambda^2 \mathsf{E}(Z_i)^2. \tag{6.29}$$

The statement (6.25) now follows from (6.13) with

$$\mathsf{E}L_n = \sup_n \mathsf{E}\left(\frac{1}{n}\sum_{i=1}^n (\widetilde{Z}_i)^2\right)^{1/2} \le 2\lambda^2 \left(\frac{1}{n}\sum_{i=1}^n \mathsf{E}Z_i{}^2\right)^{1/2}$$

because (R4) guarantees that the right-hand terms are bouded by a constant. □

In the following result we use the above considered ball $B_\delta$, and also similar balls $B_\gamma$ centered at $0 \in \mathbb{R}^m$ with arbitrary $\gamma > 0$.

**Proposition 6.6** *(van der Vaart and Wellner). Let $\widehat{\theta}_n$ be consistent. If there exist constants $0 < \delta_0 \le \delta$ and $\kappa_1, \kappa_2 > 0$ such that*

$$\liminf_{n\to\infty} \inf_{\mathbf{u}\in B_{\delta_0}} \left(\frac{1}{\sqrt{n}}\,\mathsf{E}\,D_n(\mathbf{u}) - \kappa_1 \|\mathbf{u}\|^2\right) \ge 0 \tag{6.30}$$

*and*

$$\limsup_{n\to\infty} \mathsf{E} \sup_{u\in B_\gamma} |D_n(\mathbf{u}) - \mathsf{E}\,D_n(\mathbf{u})| \le \kappa_2\gamma \quad \textit{for all} \quad 0 < \gamma \le \delta_0 \tag{6.31}$$

*then the estimator $\hat{\theta}_n$ under consideration is $\sqrt{n}$-consistent in the sense of (1.11).*

**Proof:** See Theorem 3.2.5 of [30]. □

**Proposition 6.7** *Let the estimator satisfy the adaption condition 2.3, the regularity conditions (R2)-(R4) and the condition (2.14) of Theorem (2.3). Then for every $\mathbf{u} \in B_\delta$ and the matrix $\Phi_n$ defined in (2.14),*

$$\sup_{\mathbf{u}\in B_\delta} \left| \frac{1}{\sqrt{n}}\,\mathsf{E}\,D_n(\mathbf{u}) - \frac{1}{2}\,\mathbf{u}'\Phi_n\mathbf{u} \right| \le \frac{(\lambda\delta)^2}{2n} \sum_{i=1}^n \omega(h_i, \lambda\delta), \tag{6.32}$$

*where $\lambda$ is the constant from the regularity condition (2.9). Furthermore, (6.30) holds for some $\delta_0$ and some $\kappa_1 > 0$.*

**Proof:** By (6.21),

$$\frac{1}{\sqrt{n}}\,\mathsf{E}\,D_n(\mathbf{u}) - \frac{1}{2}\sum_{i=1}^n h_i(0)\,\xi_i^2(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^n \left[\mathsf{E}\,R(Y_i - \varphi_i,\, \xi_i(\mathbf{u})) - \frac{1}{2}h_i(0)\xi_i^2(\mathbf{u})\right],$$

where $\|\xi_i(\mathbf{u})\| \le \lambda\delta$. Relation (6.32) follows from here and from Proposition 6.2. To complete the proof we note that by (6.28) and $(a-b)^2 \le 2a^2 + 2b^2$ it holds

$$\xi_i^2(\mathbf{u}) \ge (\dot{\varphi}_i(\mathbf{u}))^2 - \frac{1}{2}\left(\frac{\lambda}{2}\,\|\mathbf{u}\|^2\right)^2.$$

□

**Proof of Theorem 2.2** Clear from Propositions 6.5-6.7. □

Introduce

$$\tilde{D}_n(\mathbf{v}) = \sqrt{n}\, D_n(\mathbf{v}/\sqrt{n}), \quad \widetilde{\mathcal{L}}_n(\mathbf{v}) = \sqrt{n}\, \mathcal{L}_n(\mathbf{v}/\sqrt{n})$$

and, similarly, also $\tilde{\mathcal{D}}_n(\mathbf{v})$, $\tilde{\mathcal{R}}_n(\mathbf{v})$ and $\widetilde{\mathcal{S}}_n(\mathbf{v}) = \tilde{\mathcal{R}}_n(\mathbf{v}) - \mathsf{E}\tilde{\mathcal{R}}_n(\mathbf{v})$, for $\mathbf{v} \in B_r$ and all sufficiently large $n$.

**Proposition 6.8** *If all assumptions of Proposition 6.5 hold then for every closed ball $B_r$,*

$$\lim_{n\to\infty} \mathsf{E} \sup_{\mathbf{v}\in B_r} |\tilde{\mathcal{D}}_n(\mathbf{v})| = \lim_{n\to\infty} \mathsf{E} \sup_{\mathbf{v}\in B_r} |\widetilde{\mathcal{S}}_n(\mathbf{v})| = 0. \tag{6.33}$$

*Consequently,*

$$\sup_{\mathbf{v}\in B_r} \left| \tilde{D}_n(\mathbf{v}) - \mathsf{E}\, \tilde{D}_n(\mathbf{v}) - \widetilde{\mathcal{L}}_n(\mathbf{v}) \right| \xrightarrow{P} 0 \quad \text{as} \quad n \to \infty. \tag{6.34}$$

**Proof:** By Proposition 6.5, for all $r > 0$

$$\mathsf{E} \sup_{\mathbf{v}\in B_r} |\tilde{\mathcal{D}}_n(\mathbf{v})| \le \sqrt{n}\, c_1 \left(\frac{r}{\sqrt{n}}\right)^2 \quad \text{and} \quad \mathsf{E} \sup_{\mathbf{v}\in B_r} |\widetilde{\mathcal{S}}_n(\mathbf{v})| \le \sqrt{n}\, c_2 \left(\frac{r}{\sqrt{n}}\right)^{1+\alpha q/2}.$$

(6.33) is clear from here. □

In the following lemma we consider

$$Z = (Z_1, \ldots, Z_m)' \sim N(0, \Sigma), \tag{6.35}$$

where $\Sigma$ is the matrix defined by (2.13).

**Proposition 6.9** *If the assumptions of Theorem 2.3 hold then for every $r > 0$, the distribution of the process $(\widetilde{\mathcal{L}}_n(\mathbf{v}) : \mathbf{v} \in B_r)$ tends weakly to the distribution of $(\mathbf{v}'Z : \mathbf{v} \in B_r)$.*

**Proof:** For a fixed $\mathbf{v} \in B_r$, $\mathbf{v}'\Sigma_n\mathbf{v}$ is the variance of the vector $\widetilde{\mathcal{L}}_n(\mathbf{v})$, where $\Sigma_n$ is defined in (2.13). By (2.17),

$$\widetilde{\mathcal{L}}_n(\mathbf{v}) \xrightarrow{\mathcal{L}} N(0, \mathbf{v}'\Sigma\mathbf{v}) \quad \text{as} \quad n \to \infty.$$

The stated convergence follows from the fact that $\widetilde{\mathcal{L}}_n(\mathbf{v})$ is linear in $\mathbf{v}$. □

In the next lemma and its proof, we consider the matrices $\Phi$ and $\Phi_n$ defined in (2.14) and the random vector $Z$ defined by (6.35).

**Proposition 6.10** *If the assumptions of Proposition 6.7 hold then for every closed ball $B_r$,*

$$\lim_{n\to\infty} \sup_{\mathbf{v}\in B_r} \left| \mathsf{E}\, \tilde{D}_n(\mathbf{v}) - \frac{1}{2}\mathbf{v}'\Phi\,\mathbf{v} \right| = 0 \tag{6.36}$$

*and the process*

$$\tilde{D}(\mathbf{v}) = \frac{1}{2}\,\mathbf{v}'\Phi\,\mathbf{v} - \mathbf{v}'Z\,, \quad \mathbf{v} \in \mathbb{R}^m, \tag{6.37}$$

*is minimized at the unique* $\Phi^{-1}Z$*, i.e.*

$$\Phi^{-1}Z = arg\ min_{\mathbf{v}\in\mathbb{R}^m}\ \tilde{D}(\mathbf{v}). \tag{6.38}$$

**Proof:** By Proposition 6.7, for every $B_r$ under consideration

$$\sup_{\mathbf{v}\in B_r}\left|\mathsf{E}\,\tilde{D}_n(\mathbf{v}) - \frac{1}{2}\,\mathbf{v}'\Phi_n\mathbf{v}\right| \le \frac{(\lambda r)^2}{2n}\sum_{i=1}^{n}\omega(h_i, r/\sqrt{n})$$

and, by (2.12), the right hand side tends to zero as $n \to \infty$. The relation (6.37) follows from the easily verifiable formula

$$\tilde{D}(\mathbf{v}) = \frac{1}{2}\left\|\Phi^{1/2}\mathbf{v} - \Phi^{-1/2}Z\right\| - Z'\Phi Z,$$

where $\Phi^{1/2}$ is the symmetric root of the matrix $\Phi$. □

**Proof of Theorem 2.3.** Define a random sequence

$$\hat{\mathbf{v}}_n = \sqrt{n}(\widehat{\theta}_n - \theta_0), \quad n \in \mathbb{N}.$$

By definition of $\tilde{D}_n(\mathbf{v})$, for each $n \in \mathbb{N}$,

$$\widetilde{\mathbf{v}}_n = \operatorname*{arg\,min}_{\mathbf{v}\in\mathbb{R}^m}\ \tilde{D}_n(\mathbf{v}). \tag{6.39}$$

By Propsition 6.7, $\widehat{\theta}_n$ is $\sqrt{n}$-consistent, so that the sequence of distributions of $\widetilde{\mathbf{v}}_n$ is tight. By (6.34) and (6.36), for every closed ball $B_r$, the distribution of the process $(\tilde{D}_n(v) : v \in B_r)$ converges weakly to the distribution of $(\tilde{D}(v) : v \in B_r)$ defined by (6.37) and satisfying (6.38). By the *argmax continuous mapping* Theorem 3.2.2 of [30], this implies

$$\hat{v}_n \xrightarrow{\mathcal{L}} \Phi^{-1}Z = N(0, \Phi^{-1}\Sigma\Phi) \quad \text{as } n \to \infty,$$

which proves (1.12) and (2.18). □

## References

[1] **Arcones, M. A. :** *M-estimators converging to a stable limit.* J. Multivariate Anal. 74, 193–221 (2000)

[2] **Arcones, M. A. :** *Asymptotic distribution of regression M-estimators.* J. Statist. Plann. Inference 97, 235–261 (2001)

[3] **Brown, L. D. :** *Fundamentals of Statistical Exponential Families.* IMS Lecture Notes Vol. 9 (1986)

[4] **Dodge, Y. :** *An introduction to statistical data analysis $L_1$-norm based.* In: *Statistical Data Analysis Based on the $L_1$ norm and Related Methods*, (Y. Dodge, ed.), Amsterdam, North Holland, pp. 1–22 (1988)

[5] **Fahrmeier, L.**, and **Kaufmann, H. :** *Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models .* Ann. Statist. 13, 342–368 (1985)

[6] **Farenbrother, R. W. :** *The historical development of the $L_1$ and $L_\infty$ estimation procedures.* In: *Statistical Data Analysis Based on the $L_1$-norm and Related Methods*, (Y. Dodge, ed.), Amsterdam, North Holland, pp. 37–64 (1987)

[7] **Hallin, M.**, and **Jurečková, J. :** *Optimal tests for autoregressive models based on autoregression rank scores.* Ann. Statist. 27, 1385–1414 (1999)

[8] **Hampel, F. R., Rousseeuw, P. J., Ronchetti, E. M.**, and **Stahel, W. A. :** *Robust Statistics. The Approach Based on Influence Functions.* Wiley, New York (1986)

[9] **Hewitt, H.**, and **Stromberg, K. :** *Real and Abstract Analysis.* Springer–Verlag, Berlin, Heidelberg, New York (1965)

[10] **Hjort, N. L.**, and **Pollard, D. :** *Asymptotic for Minimizers of Convex Processes.* Statistical Research Report, University of Oslo (1993)

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

[12] **Huber, P. J. :** *Robust Statistics.* Wiley, New York (1981)

[13] **Jurečková, J.**, and **Procházka, B. :** *Regression quantiles and trimmed squares estimator in nonlinear regression model.* Nonparametric Statistics 3, 201–222 (1994)

[14] **Jurečková, J.**, and **Sen, P. K. :** *Robust Statistical Inference: Asymptotics and Interrelations.* Wiley, New York (1996)

[15] **Knight, K. :** *Limiting Distributions for $L_1$-regression estimators under general conditions.* Ann. Statist. 26, 755–770 (1998)

[16] **Koenker, R.**, and **Basset, G. :** *Regression quantiles.* Econometrica 46, 33–500 (1978)

[17] **Koul, H. L.**, and **Saleh, A. K. E. :** *Autoregression quantiles and related rank scores processes.* Ann. Statist. 23, 670–689 (1995)

[18] **Liese, F.**, and **Vajda, I. :** *Consistency of $M$-estimates in general regression models.* J. Multivariate Anal. 50, 93–114 (1994)

[19] **Liese, F.**, and **Vajda, I. :** *Necessary and sufficient conditions for consistency of generalized $M$-estimates.* Metrika 42, 291–324 (1995)

[20] **Liese, F.**, and **Vajda, I. :** *$M$-estimators of structural parameters in pseudolinear models.* Appl. Statist. 20, 1514–1533 (1999)

[21] **Liese, F.**, and **Vajda, I. :** *A general asymptotic theory of $M$-estimators.* Math. Methods Statist. (submitted) (2002)

[22] **Pollard, D. :** *Asymptotics for least absolute deviation regression estimates.* Econometric Theory 7, 186–199 (1991)

[23] **Portnoy, S. :** *Robust estimation in dependent situations.* Ann. Statist. 5, 22–43 (1977)

[24] **Portnoy, S. :** *Asymptotic behavior of regression quantiles in nonstationary cases.* J. Multivariate Anal. 38, 100–113 (1991)

[25] **Rao, C. R. :** *Linear Statistical Inference and Its Appliations.* 2nd ed., Wiley, New York (1973)

[26] **Ronchetti, E. :** *Bounded influence in regression: A review.* In: Statistical Data Analysis Based on the $L_1$ norm and Related Methods, (Y. Dodge, ed.), Amsterdam, North Holland, pp. 65–80 (1988)

[27] **Serfling, R. J. :** *Approximation Theorems of Mathematical Statistics.* Wiley, New York (1980)

[28] **Sen, P. K.**, and **Singer, J. M. :** *Large Sample Methods in Statistics.* Chapman & Hall, London (1993)

[29] **Vajda, I. :** *Efficiency and robustness control via distorted maximum likelihood estimation.* Kybernetika 22, 47–67 (1986)

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

[31] **Yohai, V. J.**, and **Maronna, R. A. :** *Asymptotic behaviour of $M$-estimates for the linear model.* Ann. Statist. 7, 258–268 (1979)

[32] **Zhao, L. C., Rao, C. R.**, and **Chen, X. :** *A note on the consistency of $M$-estimates in linear models.* Collection: Stochastic Processes, 359–367, Springer, New York (1993)

[33] **Zhao, L. C. :** *Some contributions to M-estimators in linear models.* J. of Statist. Planning and Inference 88, 189-203 (2000)

**received:** September 16, 2002

**Authors:**

Prof. Dr. Friedrich Liese
Universität Rostock
Fachbereich Mathematik
18051 Rostock
Germany

e-mail: friedrich.liese@mathematik.uni-rostock.de

Prof. Dr. Igor Vajda
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodàrenskou věži 4
CZ-182 08 Praha
Czech Republic

e-mail: vajda@utia.cas.cz