AUSTRIAN JOURNAL OF STATISTICS
Volume 35 (2006), Number 2&3, 289–298

Estimator Consistency in a General Setup

Petr Lachout
Charles University, Prague, Czech Republic

Abstract: Many problems considered and investigated in statistics follow a general schema. Observed data are generated by a model containing randomness and determined via a collection of parameters. We are interested in future behavior of observed system. Therefore, a convenient estimation procedure for unknown parameters becomes to be the crucial task. This schedule often leads to derivation of an optimization problem that solution is a reasonable estimator of required parameters.

We are discussing behavior of such an estimator with a relatively general background. The setup is illustrated on a linear regression model.

Keywords: Optimization Problem, Estimator, Strong Consistency, Weak Convergence of Probability Measures.

1 Introduction

We consider a general scheme of parameter estimation in this paper. Our task is to estimate true value of a model parameter. As parameter we allow reals, real vectors, real functions, etc. Simply, parameter is a member of a metric space. True value of parameter is estimated with an ε-solution of an optimization problem. We present assumptions under which such an estimator is consistent in almost sure sense. We do not require measurability of the estimator. Hence, almost sure convergence used throughout the paper must be understood according to definition in Vaart and Wellner (1996). Thus, it can happen that an estimator is almost surely consistent, although it is not consistent in probability.

As an illustration of our general result we consider M-estimator in linear regression model. There is a vast literature on the linear regression model, e.g. Chen and Wu (1988); Dodge and Jurečková (2000); Jurečková (1980, 1985); Jurečková and Sen (1996); Knight (1998); Leroy and Rousseeuw (1987); Rockafellar and Wets (1998), etc. But, the results usually assume unique minimizer, regressors are supposed i.i.d. or deterministic, errors are i.i.d., errors are independent with regressors, estimator must be measurable. Our paper requires no prescribed structure for observations and errors. We only assume weak convergence of their common empirical measure. Also, we allow non-uniqness of the estimator. The non-measurability problem is overcome by using a general scheme and definition of almost sure convergence without any measurability assumption, see Vaart and Wellner (1996).

2 General Result

We consider a general scheme of parameter estimation in this paper. Our task is to estimate true value of an important parameter. Let us denote it θ_0. We suppose to know the set, say Θ, of all possible values of this parameter. Also we have known a parameterized family of probability measures $\mathcal{P}_\Theta = \{\mu_\theta \mid \theta \in \Theta\}$ defined on a metric space $\mathcal{Y}$.

We observe $X_1, X_2, \ldots$ belonging into a metric space $\mathcal{X}$. From observed data we construct probability measures $\mu_n(\cdot \mid X_1, \ldots, X_n)$ on $\mathcal{Y}$. These measures play rule of estimators for the true probability measure μ_{θ_0}.

The true parameter θ_0 is estimated by an ε_n-estimator $\hat{\theta}_n \in \Theta$, i.e. fulfilling for all $\theta \in \Theta$

$$\mathsf{L}(\mu_n(\cdot \mid X_1, \ldots, X_n)\,; \hat{\theta}_n) < \mathsf{L}(\mu_n(\cdot \mid X_1, \ldots, X_n)\,; \theta) + \varepsilon_n\,, \tag{1}$$

where L is a given distance between measures and parameters.

Now, let us formalize the schema in a list of assumptions.

Assumption A: Spaces $\mathcal{X}$, $\mathcal{Y}$, Φ are metric spaces, $\Theta \subset \Phi$ is nonempty and $\mathcal{F} \subset \{f : \mathcal{Y} \to \mathbb{R} \mid f \text{ is measurable}\}$, possibly empty set.

Assumption B: $\varepsilon_n > 0$ for any $n \in \mathbb{N}$ and $\lim_{n\to\infty} \varepsilon_n = 0$.

Assumption C: For any $\theta \in \Theta$, μ_θ is a Borel probability measure on $\mathcal{Y}$. We denote $\mathcal{P}_\Theta = \{\mu_\theta \mid \theta \in \Theta\}$.

Assumption D: For any $k \in \mathbb{N}$, we observe a random variable $X_k \in \mathcal{X}$.

Assumption E: For any $n \in \mathbb{N}$, $x_1, \ldots, x_n \in \mathcal{X}$, $\mu_n(\cdot \mid x_1, \ldots, x_n)$ is a Borel probability measure on $\mathcal{Y}$.
We denote $\mathcal{P}_{emp} = \{\mu_n(\cdot \mid x_1, \ldots, x_n) \mid x_1, \ldots, x_n \in \mathcal{X}, n \in \mathbb{N}\}$.

Assumption F: The function $\mathsf{L} \;:\; (\mathcal{P}_{emp} \cup \mathcal{P}_\Theta) \times \Theta \to \mathbb{R}$ is non-negative.

Assumption G: θ_0 is a minimizer of the function $\mathsf{L}(\mu_{\theta_0}\,; \cdot)$.

Assumption H: $\lim_{n\to\infty} \mathsf{L}(\nu_n\,; \theta_0) = \mathsf{L}(\mu_{\theta_0}\,; \theta_0)$ whenever $\nu_n \in \mathcal{P}_{emp}$, $\nu_n \xrightarrow[n\to+\infty]{w} \mu_{\theta_0}$, $\forall\, f \in \mathcal{F} : \lim_{n\to+\infty} \int f(y)\nu_n(\mathsf{d}y) = \int f(y)\mu_{\theta_0}(\mathsf{d}y)$.

Assumption I: There is a compact set $K \subset \Theta$ such that

1. $\theta_0 \in K$.
2. $\liminf_{n\to+\infty} \mathsf{L}(\nu_n\,; \theta_n) \geq \mathsf{L}(\mu_{\theta_0}\,; \theta)$ whenever $\nu_n \in \mathcal{P}_{emp}$, $\nu_n \xrightarrow[n\to+\infty]{w} \mu_{\theta_0}$, $\forall\, f \in \mathcal{F} : \lim_{n\to+\infty} \int f(y)\nu_n(\mathsf{d}y) = \int f(y)\mu_{\theta_0}(\mathsf{d}y)$, $\theta, \theta_n \in K$, $\theta_n \xrightarrow[n\to+\infty]{} \theta$.
3. For any sequence of probability measures $\nu_n \in \mathcal{P}_{emp}$, $\nu_n \xrightarrow[n\to+\infty]{w} \mu_{\theta_0}$, $\forall\, f \in \mathcal{F}$, $\lim_{n\to+\infty} \int f(y)\nu_n(\mathsf{d}y) = \int f(y)\mu_{\theta_0}(\mathsf{d}y)$ we have

$$\liminf_{n\to+\infty} \inf_{\theta\in\Theta\setminus K} \mathsf{L}(\nu_n\,; \theta) > \mathsf{L}(\mu_{\theta_0}\,; \theta_0).$$

These assumptions ensure the existence and consistency of the estimator.

Lemma 1 *Under B and F, an estimator $\hat{\theta}_n \in \Theta$ fulfilling (1) exists for any $n \in \mathbb{N}$.*

Proof. Let $n \in \mathbb{N}$. Accordingly to Assumptions F and B,

$$0 \leq \inf_{\theta \in \Theta} \mathsf{L}(\mu_n(\cdot \,|\, X_1, \ldots, X_n)\,; \theta) < \infty \quad \text{and} \quad \varepsilon_n > 0\,.$$

Hence, an $\hat{\theta}_n \in \Theta$ fulfilling (1) exists. □

We have to recall a few from topological terminology.

Definition 1 *For a sequence η_n, $n \in \mathbb{N}$ in a metric space $\mathcal{W}$, we denote the set of its cluster points by $\mathsf{Ls}\,(\eta_n, n)$, i.e.*

$$\mathsf{Ls}\,(\eta_n, n) = \left\{ \psi \in \mathcal{W} \;\middle|\; \exists \textit{ subsequence s.t. } \lim_{n \to +\infty} \eta_{k_n} = \psi \right\}.$$

Definition 2 *We say that a sequence η_n, $n \in \mathbb{N}$ in a metric space $\mathcal{W}$ is compact if each its subsequence possesses at least one cluster point.*

Compact sequence in metric space possesses an equivalent description.

Lemma 2 *Let η_n, $n \in \mathbb{N}$ be a sequence in a metric space $\mathcal{W}$. Then, the following statements are equivalent:*

1. *The sequence is compact.*
2. *There is a compact $L \subset \mathcal{W}$ such that $\eta_n \in L$ for all $n \in \mathbb{N}$.*
3. *The set $\{\eta_n \;|\; n \in \mathbb{N}\} \cup \mathsf{Ls}\,(\eta_n, n)$ is compact.*

General topology concept and a proof of Lemma 2 can be found in any textbook on topology, e.g. Kelley (1955).

Now, we proceed to the main theorem of our paper.

Theorem 1 *Let $\Omega_0 \subset \Omega$, $\mathsf{Prob}\,(\Omega_0) = 1$ be such that for all $\omega \in \Omega_0$*

$$\mu_n(\cdot \,|\, X_1(\omega), \ldots, X_n(\omega)) \xrightarrow[n \to +\infty]{w} \mu_{\theta_0}$$

$$\forall\, f \in \mathcal{F} : \lim_{n \to +\infty} \int f(y) \mu_n(\mathsf{d}y \,|\, X_1(\omega), \ldots, X_n(\omega)) = \int f(y) \mu_{\theta_0}(\mathsf{d}y)$$

and Assumptions A–I be fulfilled. Then $\hat{\theta}_n \in \Theta$ fulfilling (1) exists for any $n \in \mathbb{N}$ and for all $\omega \in \Omega_0$ the sequence $\hat{\theta}_n(\omega)$, $n \in \mathbb{N}$ is compact and

$$\emptyset \neq \mathsf{Ls}\left(\hat{\theta}_n(\omega), n\right) \subset \mathsf{argmin}\,\{\mathsf{L}(\mu_{\theta_0}\,; \theta) \;|\; |\theta \in \Theta\}\,.$$

Proof. Existence of an ε_n-estimator is proved in previous Lemma 1.

For fixed $\omega \in \Omega_0$, the situation became to be deterministic.

1. Assumption H implies

$$\lim_{n\to+\infty} \mathsf{L}(\mu_n(\cdot \mid X_1(\omega),\dots,X_n(\omega))\,;\theta_0) = \mathsf{L}(\mu_{\theta_0}\,;\theta_0).$$

2. Property (1) and Assumptions B, H, I imply

$$\begin{aligned}
&\limsup_{n\to+\infty} \mathsf{L}(\mu_n(\cdot \mid X_1(\omega),\dots,X_n(\omega))\,;\hat\theta_n(\omega)) \le \\
&\le \lim_{n\to+\infty} \Big[\mathsf{L}(\mu_n(\cdot \mid X_1(\omega),\dots,X_n(\omega))\,;\theta_0) + \varepsilon_n\Big] \\
&= \mathsf{L}(\mu_{\theta_0}\,;\theta_0) < \liminf_{n\to+\infty} \inf_{\theta\in\Theta\setminus K} \mathsf{L}(\mu_n(\cdot \mid X_1(\omega),\dots,X_n(\omega))\,;\theta).
\end{aligned}$$

   Therefore, ε_n-estimator $\hat\theta_n(\omega) \in K$ for all $n \in \mathbb{N}$ sufficiently large.

3. We have shown that $\hat\theta_n(\omega) \in K$ for all $n \in \mathbb{N}$ sufficiently large. Hence the sequence $\hat\theta_n(\omega)$, $n \in \mathbb{N}$ is compact and possesses at least one cluster point.

4. Let $\eta \in \Phi$ be a cluster point of the sequence $\hat\theta_n(\omega)$, $n \in \mathbb{N}$ and $\lim_{n\to+\infty} \hat\theta_{k_n}(\omega) = \eta$. Then $\eta \in K$ and employing property (1) and Assumptions B, H, G, I, we receive

$$\begin{aligned}
&\mathsf{L}(\mu_{\theta_0}\,;\theta_0) \le \mathsf{L}(\mu_{\theta_0}\,;\eta) \le \\
&\le \liminf_{n\to+\infty} \mathsf{L}(\mu_{k_n}(\cdot \mid X_1(\omega),\dots,X_{k_n}(\omega))\,;\hat\theta_{k_n}(\omega)) \\
&\le \lim_{n\to+\infty} \Big[\mathsf{L}(\mu_{k_n}(\cdot \mid X_1(\omega),\dots,X_{k_n}(\omega))\,;\theta_0) + \varepsilon_{k_n}\Big] \\
&\le \mathsf{L}(\mu_{\theta_0}\,;\theta_0).
\end{aligned}$$

   Hence,

$$\mathsf{L}(\mu_{\theta_0}\,;\eta) = \mathsf{L}(\mu_{\theta_0}\,;\theta_0)$$

   and, therefore, $\eta \in \mathsf{argmin}\,\{\mathsf{L}(\mu_{\theta_0}\,;\theta) \mid \theta \in \Theta\}$. □

Our proof treats any trajectory separately. Therefore, we do not need measurability of $\mu_n(\cdot \mid x_1,\dots,x_n)$ with respect to $x_1,\dots,x_n \in \mathcal{X}$. Also, our definition of the ε_n-estimator does not require measurability. Thus, it can happen that the estimator is not a random variable.

3 Linear Regression

As an example illustrating the theory presented in the first section of the paper we will discuss a linear regression model. Where, unknown regression coefficients are estimated by an ε_n-M-estimator.

We observe random couples $(Y_1, X_1), \dots, (Y_n, X_n)$ connected by a linear regression model

$$Y_i = X_i^\top \beta_0 + \varepsilon_i \quad \forall\, i = 1, \dots, n\,. \tag{2}$$

Where $Y_i \in \mathbb{R}$, $X_i \in \mathbb{R}^d$ are observed random variables, $\varepsilon_i \in \mathbb{R}$ are random errors which are not observed and $\beta_0 \in \Theta \subset \mathbb{R}^d$ is deterministic but unknown parameter.

Parameter set Θ expresses our a prior information, knowledge about parameters. For example, we know that some functions of parameters are nonnegative or having precise value, e.g. some parameters are nonnegative or bounded by a value, some linear combinations of parameters are nonnegative or having precise value, etc.

Considering relative frequencies of a sequence $(y_1, x_1), \dots, (y_n, x_n)$, we receive a probability measure $\eta_{n;\,Y,X}$, i.e. for any Borel subset A of $\mathbb{R}^{d+1}$

$$\eta_{n;\,Y,X}(A \,|\, (y_1, x_1), \dots, (y_n, x_n)) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}[(y_i, x_i) \in A].$$

Hence the empirical probability distribution of observed sample is

$$\eta_{n;\,Y,X}(\cdot \,|\, (Y_1, X_1), \dots, (Y_n, X_n)).$$

Unknown regression coefficients are estimated by an ε_n-M-estimator based on a loss function defined by the formula

$$\mathsf{L}(\mu\,; \beta) = \int \rho(y - x^\top \beta) \mu(\mathsf{d}y, \mathsf{d}x). \tag{3}$$

Especially, for empirical distribution we receive

$$\begin{aligned}
&\mathsf{L}(\eta_{n;\,Y,X}(\cdot \,|\, (Y_1, X_1), \dots, (Y_n, X_n))\,; \beta) = \\
&= \int \rho(y - x^\top \beta) \eta_{n;\,Y,X}(\mathsf{d}y, \mathsf{d}x \,|\, (Y_1, X_1), \dots, (Y_n, X_n)) \\
&= \frac{1}{n} \sum_{i=1}^{n} \rho(Y_i - X_i^\top \beta).
\end{aligned}$$

An ε_n-M-estimator is $\hat{\beta}_n \in \Theta$ fulfilling for all $\beta \in \Theta$

$$\begin{aligned}
&\mathsf{L}(\eta_{n;\,Y,X}(\cdot \,|\, (Y_1, X_1), \dots, (Y_n, X_n))\,; \hat{\beta}_n) < \\
&\quad < \mathsf{L}(\eta_{n;\,Y,X}(\cdot \,|\, (Y_1, X_1), \dots, (Y_n, X_n))\,; \beta) + \varepsilon_n.
\end{aligned} \tag{4}$$

Now, the studied situation is fully described and we can proceed to assumptions.

Again, we consider relative frequencies of $(x_1, e_1), \dots, (x_n, e_n)$ forming a probability measure $\eta_{n;\,X,\varepsilon}$, i.e. for any Borel subset A of $\mathbb{R}^{d+1}$

$$\eta_{n;\,X,\varepsilon}(A \,|\, (x_1, e_1), \dots, (x_n, e_n)) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}[(x_i, e_i) \in A].$$

Hence the joint empirical probability distribution of covariates and errors is

$$\eta_{n;\,X,\varepsilon}(\cdot \,|\, (X_1, \varepsilon_1), \dots, (X_n, \varepsilon_n))\,.$$

For consistency, we assume

Assumption J: $\Theta \subset \mathbb{R}^d$ is a closed subset.

Assumption K: There is a Borel measure $\nu_{X,\varepsilon}$ defined on $\mathbb{R}^{d+1}$ such that

$$\eta_{n;\,X,\varepsilon}(\cdot \mid (X_1,\varepsilon_1),\ldots,(X_n,\varepsilon_n))) \xrightarrow[n\to+\infty]{w} \nu_{X,\varepsilon} \text{ a.s.}$$

Assumption L: For any $\beta \in \Theta$

$$\int \rho(e)\,\nu_{X,\varepsilon}(\mathsf{d}x,\mathsf{d}e) \le \int \rho(e + x^\top(\beta_0-\beta))\,\nu_{X,\varepsilon}(\mathsf{d}x,\mathsf{d}e).$$

Assumption M: Function $\rho : \mathbb{R} \to \mathbb{R}$ is nonnegative and continuous.

Assumption N: There is a function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ which is continuous, nondecreasing and fulfilling:

1. $\rho(t) \le \psi(|t|)$ for all $t \in \mathbb{R}$.
2. Let for all $t > 0$ $\int \psi(|e| + t\|x\|)\nu_{X,\varepsilon}(\mathsf{d}x,\mathsf{d}e) < +\infty$.
3. Let for all $t > 0$ $\frac{1}{n}\sum_{i=1}^n \psi(|\varepsilon_i| + t\|X_i\|) \xrightarrow[n\to+\infty]{a.s.} \int \psi(|e| + t\|x\|)\nu_{X,\varepsilon}(\mathsf{d}x,\mathsf{d}e)$.

Assumption O: There is $\Delta > 0$ such that $\mathsf{H} = \inf\{\rho(t) \mid |t| > \Delta, t \in \mathbb{R}\} > \mathsf{L}(\mu_{\beta_0};\beta_0)$.

Assumption P: Let $\nu_{X,\varepsilon}(\{(x,e) \in \mathbb{R}^{d+1} \mid x^\top\gamma \ne 0\}) = 1$ for all $\gamma \in \mathbb{R}^d$, $\gamma \ne 0$.

Theorem 2 *If Assumptions B, J-P are fulfilled then $\hat\beta_n \in \Theta$ fulfilling (4) exists for any $n \in \mathbb{N}$ and there is $\Omega_0 \subset \Omega$, $\mathsf{Prob}(\Omega_0) = 1$ such that for all $\omega \in \Omega_0$ the sequence $\hat\beta_n(\omega)$, $n \in \mathbb{N}$ is compact and*

$$\emptyset \ne \mathsf{Ls}\left(\hat\beta_n(\omega), n\right) \subset \mathsf{argmin}\left\{\mathsf{L}(\mu_{\beta_0};\theta) \mid |\beta \in \Theta\right\}.$$

Proof. We will show that this theorem is a particular case of Theorem 1.

We set $\mathcal{X} = \mathcal{Y} = \mathbb{R}^{d+1}$, $\Phi = \mathbb{R}^d$ and for any $\beta \in \Theta$ the measure μ_β is defined by

$$\mu_\beta(A) = \int \mathbb{I}[(e + x^\top\beta, x) \in A]\,\nu_{X,\varepsilon}(\mathsf{d}x,\mathsf{d}e).$$

As pointed functions we set

$$\mathcal{F} = \left\{(y,x) \mapsto \psi(|y - x^\top\beta_0| + t\|x\|) \mid t > 0\right\}.$$

Consider a sequence $\gamma_k \in \mathbb{R}^d$, $\|\gamma_k\| = 1$ for all $k \in \mathbb{N}$ such that

$$\nu_{X,\varepsilon}(\{(x,e) \mid k\,|x^\top\gamma_k| \ge \Delta + |e|\}) < \inf_{\|\gamma\|=1} \nu_{X,\varepsilon}(\{(x,e) \mid k\,|x^\top\gamma| \ge \Delta + |e|\}) + \frac{1}{k}.$$

The sequence lays in a compact, therefore, it contains a convergent subsequence

$$\lim_{j\to+\infty}\gamma_{k_j}=\hat{\gamma},\ \|\hat{\gamma}\|=1.$$

Hence

$$\bigcup_{J=1}^{+\infty}\bigcap_{j=J}^{+\infty}\left\{(x,e)\ \middle|\ k_j\left|x^\top\gamma_{k_j}\right|\geq\Delta+|e|\right\}\supset\left\{(x,e)\ \middle|\ x^\top\hat{\gamma}\neq 0\right\}.$$

Because of Assumptions P and σ-additivity of the measure $\nu_{X,\varepsilon}$, we have

$$\lim_{\kappa\to+\infty}\inf_{\|\gamma\|=1}\nu_{X,\varepsilon}(\left\{(x,e)\ \middle|\ \kappa\left|x^\top\gamma\right|\geq\Delta+|e|\right\})=1.$$

Now, using Assumptions O, we are able to find Γ such that

$$\mathsf{H}\,\nu_{X,\varepsilon}(\left\{(x,e)\ \middle|\ \Gamma\left|x^\top\gamma\right|\geq\Delta+|e|\right\})>\mathsf{L}(\mu_{\beta_0}\,;\beta_0)\quad\forall\,\|\gamma\|=1\,.$$

Then, we define the required compact as

$$K=\{\beta\in\Theta\ |\ \|\beta-\beta_0\|\leq\Gamma\}\,.$$

Hence, Assumptions A-E are fulfilled.

1. Loss function is non-negative since ρ is non-negative by Assumption M.

   For $\beta\in\Theta$ and $\mu\in\mathcal{P}_{emp}$,

$$\mathsf{L}(\mu\,;\beta)=\frac{1}{n}\sum_{i=1}^{n}\rho(y_i-x_i^\top\beta)\in\mathbb{R}$$

   For $\beta,\gamma\in\Theta$,

$$\mathsf{L}(\mu_\gamma\,;\beta)=\int\rho(e+x^\top(\gamma-\beta))\ \nu_{X,\varepsilon}(\mathsf{d}x,\mathsf{d}e)\in\mathbb{R}$$

   because of Assumption N we have

$$\mathsf{L}(\mu_\gamma\,;\beta)\leq\int\psi(|e|+\|\gamma-\beta\|\|x\|)\ \nu_{X,\varepsilon}(\mathsf{d}x,\mathsf{d}e)<+\infty.$$

   Therefore, Assumption F is valid.

2. Assumption G is fulfilled because β_0 is a minimizer of the function $\mathsf{L}(\mu_{\beta_0}\,;\bullet)$ according to Assumption L.

3. Let $\nu_n\in\mathcal{P}_{emp}$, $\beta,\beta_n\in\Theta$, $\nu_n\xrightarrow[n\to+\infty]{w}\mu_{\beta_0}$, for any $t>0$

$$\lim_{n\to+\infty}\int\psi(|y-x^\top\beta_0|+t\|x\|)\nu_n(\mathsf{d}y,\mathsf{d}x)=\int\psi(|y-x^\top\beta_0|+t\|x\|)\mu_{\beta_0}(\mathsf{d}y,\mathsf{d}x)$$

and $\beta_n \xrightarrow[n\to+\infty]{} \beta$.

Let us fix $\varepsilon > 0$. Then there exist T, Q and a compact $\bar{K} \subset \mathbb{R}^{d+1}$ such that

$$\|\beta_n - \beta_0\| \le T \quad \text{for all } n \in \mathbb{N},$$
$$\int_{\psi(|e|+T\|x\|)>Q} (\psi(|e| + T\|x\|) - Q)\, \nu_{X,\varepsilon}(\mathsf{d}x, \mathsf{d}e) < \varepsilon,$$
$$\nu_n(\mathbb{R}^{d+1} \setminus \bar{K}) < \frac{\varepsilon}{Q} \quad \text{for all } n \in \mathbb{N}.$$

$$\begin{aligned}
\mathsf{L}(\nu_n\,;\beta_n) &= \int \rho(y - x^\top\beta_n)\nu_n(\mathsf{d}y, \mathsf{d}x) = \\
&= \int \min\{Q, \rho(y - x^\top\beta)\}\nu_n(\mathsf{d}y, \mathsf{d}x) + \\
&\quad + \int \min\{Q, \rho(y - x^\top\beta_n)\} - \min\{Q, \rho(y - x^\top\beta)\}\nu_n(\mathsf{d}y, \mathsf{d}x) + \\
&\quad + \int_{\rho(y-x^\top\beta_n)>Q} \big(\rho(y - x^\top\beta_n) - Q\big)\, \nu_n(\mathsf{d}y, \mathsf{d}x).
\end{aligned}$$

(a) The function $\min\{Q, \rho\}$ is bounded and continuous. Therefore,

$$\int \min\{Q, \rho(y - x^\top\beta)\}\nu_n(\mathsf{d}y, \mathsf{d}x) \xrightarrow[n\to+\infty]{} \int \min\{Q, \rho(y - x^\top\beta)\}\mu_{\beta_0}(\mathsf{d}y, \mathsf{d}x).$$

(b) The second term fulfills

$$\begin{aligned}
&\left|\int \min\{Q, \rho(y - x^\top\beta_n)\} - \min\{Q, \rho(y - x^\top\beta)\}\nu_n(\mathsf{d}y, \mathsf{d}x)\right| < \\
&< 2\varepsilon + \left|\int_{\bar{K}} \min\{Q, \rho(y - x^\top\beta_n)\} - \min\{Q, \rho(y - x^\top\beta)\}\nu_n(\mathsf{d}y, \mathsf{d}x)\right| \\
&\le 2\varepsilon + \sup_{(y,x)\in\bar{K}} \big|\min\{Q, \rho(y - x^\top\beta_n)\} - \min\{Q, \rho(y - x^\top\beta)\}\big| \\
&\xrightarrow[n\to+\infty]{} 2\varepsilon
\end{aligned}$$

because ρ is continuous, Assumption M, and, hence, uniformly continuous on each compact set.

(c) The third term is smaller than ε since

$$\begin{aligned}
0 &\le \int_{\rho(y-x^\top\beta_n)>Q} \big(\rho(y - x^\top\beta_n) - Q\big)\, \nu_n(\mathsf{d}y, \mathsf{d}x) \\
&\le \int_{\psi(|y-x^\top\beta_0|+T\|x\|)>Q} \big(\psi(|y - x^\top\beta_0| + T\|x\|) - Q\big)\, \nu_n(\mathsf{d}y, \mathsf{d}x) \\
&= \int \psi(|y - x^\top\beta_0| + T\|x\|)\nu_n(\mathsf{d}y, \mathsf{d}x) -
\end{aligned}$$

$$
\begin{aligned}
&- \int \min\{Q, \psi(|y - x^\top \beta_0| + T\|x\|)\} \nu_n(\mathsf{d}y, \mathsf{d}x) \\
\xrightarrow[n\to+\infty]{} &\int_{\psi(|y-x^\top\beta_0|+T\|x\|)>Q} \big(\psi(|y - x^\top\beta_0| + T\|x\|) - Q\big)\, \mu_{\beta_0}(\mathsf{d}y, \mathsf{d}x) \\
= &\int_{\psi(|e|+T\|x\|)>Q} (\psi(|e| + T\|x\|) - Q)\, \nu_{X,\varepsilon}(\mathsf{d}x, \mathsf{d}e) < \varepsilon.
\end{aligned}
$$

We have proved

$$
\lim_{n\to+\infty} \mathsf{L}(\nu_n\,; \beta_n) = \int \rho(y - x^\top\beta)\mu_{\beta_0}(\mathsf{d}y, \mathsf{d}x) = \mathsf{L}(\mu_{\beta_0}\,; \beta).
$$

Thus, Assumption H and the second part of Assumption I are verified.

4. Evidently, $\beta_0 \in K$. Therefore, the third part of Assumption I must be shown, only.

   Let $\nu_n \in \mathcal{P}_{emp}$, $\nu_n \xrightarrow[n\to+\infty]{w} \mu_{\beta_0}$ and $\varepsilon > 0$.

   Then, there is a bounded open set $G \subset \mathbb{R}^{d+1}$ such that $\nu_{X,\varepsilon}(G) > 1 - \varepsilon$.

   For $\beta \in \Theta$, $\|\beta - \beta_0\| > \Gamma$ we receive following chain of inequalities.

$$
\begin{aligned}
\mathsf{L}(\nu_n\,; \beta) &= \int \rho(y - x^\top\beta)\nu_n(\mathsf{d}y, \mathsf{d}x) \\
&\geq \int_{|y-x^\top\beta|>\Delta} \rho(y - x^\top\beta)\nu_n(\mathsf{d}y, \mathsf{d}x) \\
&\geq \mathsf{H}\,\nu_n\left(\left\{(y, x) \;\middle|\; |y - x^\top\beta| > \Delta\right\}\right) \\
&\geq \mathsf{H}\,\nu_n\left(\left\{(y, x) \;\middle|\; |x^\top(\beta - \beta_0)| > \Delta + |y - x^\top\beta_0|\right\}\right) \\
&\geq \mathsf{H}\,\nu_n\left(\left\{(y, x) \;\middle|\; \Gamma\left|x^\top\frac{\beta - \beta_0}{\|\beta - \beta_0\|}\right| > \Delta + |y - x^\top\beta_0|\right\}\right).
\end{aligned}
$$

   For properly chosen sequence of γ_n, $\|\gamma_n\| = 1$ and its cluster point $\hat{\gamma}$, we have

$$
\begin{aligned}
&\liminf_{n\to+\infty} \inf_{\beta\notin K} \mathsf{L}(\nu_n\,; \beta) \geq \\
&\geq \mathsf{H} \liminf_{n\to+\infty} \inf_{\|\gamma\|=1} \nu_n\left(\left\{(y, x) \;\middle|\; \Gamma\left|x^\top\gamma\right| > \Delta + |y - x^\top\beta_0|\right\}\right) \\
&\geq \mathsf{H} \liminf_{n\to+\infty} \nu_n\left(\left\{(y, x) \;\middle|\; \Gamma\left|x^\top\gamma_n\right| > \Delta + |y - x^\top\beta_0|\right\}\right) \\
&\geq \mathsf{H} \liminf_{n\to+\infty} \nu_n\left(\left\{(y, x) \;\middle|\; \Gamma\left|x^\top\hat{\gamma}\right| > \Delta + |y - x^\top\beta_0| + \Gamma\left|x^\top(\gamma_n - \hat{\gamma})\right|,\ (x, y - x^\top\beta_0) \in G\right\}\right) \\
&\geq \mathsf{H} \liminf_{n\to+\infty} \nu_n\left(\left\{(y, x) \;\middle|\; \Gamma\left|x^\top\hat{\gamma}\right| > (1+\varepsilon)\Delta + |y - x^\top\beta_0|,\ (x, y - x^\top\beta_0) \in G\right\}\right) \\
&\geq \mathsf{H}\,\mu_{\beta_0}(\left\{(y, x) \;\middle|\; \Gamma\left|x^\top\hat{\gamma}\right| > (1+\varepsilon)\Delta + |y - x^\top\beta_0|,\ (x, y - x^\top\beta_0) \in G\right\}) \\
&= \mathsf{H}\,\nu_{X,\varepsilon}(\left\{(x, e) \;\middle|\; \Gamma\left|x^\top\hat{\gamma}\right| > (1+\varepsilon)\Delta + |e|,\ (x, e) \in G\right\}) \\
&> \mathsf{H}\,\nu_{X,\varepsilon}(\left\{(x, e) \;\middle|\; \Gamma\left|x^\top\hat{\gamma}\right| > (1+\varepsilon)\Delta + |e|\right\}) - H\varepsilon.
\end{aligned}
$$

   Letting ε vanish we have

$$
\begin{aligned}
\liminf_{n\to+\infty} \inf_{\beta\notin K} \mathsf{L}(\nu_n\,; \beta) &\geq \mathsf{H}\,\nu_{X,\varepsilon}(\left\{(x, e) \;\middle|\; \Gamma\left|x^\top\hat{\gamma}\right| \geq \Delta + |e|\right\}) \\
&> \mathsf{L}(\mu_{\beta_0}\,; \beta_0)\,.
\end{aligned}
$$

Thus, the rest of Assumption I is verified. Assumptions A - I are valid. Existence of Ω_0, $\text{Prob}\,(\Omega_0) = 1$ such that on it weak convergence of empirical measures and convergence of integrals for all functions from $\mathcal{F}$ follows Assumption K, monotonicity of ψ and Assumption N. □

We see that this setup covers both linear regression with random covariate X and the case of covariate X lead by a deterministic design. Nevertheless, it allows more general structure.

Acknowledgments

The author is grateful to an anonymous referee for his comments and suggestions helping to correct and improve the paper. The research has been partially supported by the Grant Agency of the Czech Republic under Grant 201/05/2340 and by the project MSM 0021620839 financed by MSMT.

References

Chen, X., and Wu, Y. (1988). Strong consistency of m-estimates in linear model. *J. Multivariate Analysis*, *27,1*, 116-130.

Dodge, Y., and Jurečková, J. (2000). *Adaptive Regression*. New York: Springer-Verlag.

Jurečková, J. (1980). Asymptotic representation of m-estimators of location. *Math. Operat. Stat. Sec. Stat.*, *11,1*, 61-73.

Jurečková, J. (1985). Representation of M-estimators with the second-order asymptotic distribution. *Statistics & Decision*, *3*, 263-276.

Jurečková, J., and Sen, P. (1996). *Robust Statistical Procedures*. New York: John Wiley & Sons, Inc.

Kelley, J. (1955). *General Topology*. New York: D. van Nostrand Comp.

Knight, K. (1998). Limiting distributions for l_1-regression estimators under general conditions. *Ann. Statist.*, *26,2*, 755-770.

Leroy, A., and Rousseeuw, P. (1987). *Robust Regression and Outlier Detection*. New York: John Wiley & Sons.

Rockafellar, T., and Wets, R.-B. (1998). *Variational Analysis*. Berlin: Springer-Verlag.

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

Author's address:

Petr Lachout
Department of Probability and Statistics
Charles University of Prague
Sokolovská 83
CZ-186 75 Praha 8
Czech Republic

E-mail: lachout@karlin.mff.cuni.cz