KYBERNETIKA — VOLUME 44 (2008), NUMBER 2, PAGES 259 – 276

STABILITY OF STOCHASTIC OPTIMIZATION PROBLEMS – NONMEASURABLE CASE

Petr Lachout

This paper deals with stability of stochastic optimization problems in a general setting. Objective function is defined on a metric space and depends on a probability measure which is unknown, but, estimated from empirical observations. We try to derive stability results without precise knowledge of problem structure and without measurability assumption. Moreover, ε-optimal solutions are considered.

The setup is illustrated on consistency of a ε-M-estimator in linear regression model.

Keywords: stability of stochastic optimization problem, weak convergence of probability measures, estimator consistency, metric spaces

AMS Subject Classification: 90C15, 90C31, 62F10

1. INTRODUCTION

We consider a general scheme of stochastic optimization problem in this paper. Our task is to derive stability of a given stochastic optimization problem. We work on a metric space and consider ε-solution of the stochastic optimization problem. We do not require measurability in this paper. Observed data, approximations and considered functions are assumed to be maps from probability space to a metric space, only. Therefore, we become to be out of the standard theory based on measurability assumption. The theory we employ is nicely explained in [20].

Working without measurability assumption, there are several concepts for convergences almost surely and in probability. These definitions together with relations between them can be found in [20], Chapter 1.9, pp. 52–56. For convenience, the reader can find relevant parts of this theory summarized in the Appendix.

We are working with the convergence almost surely. That means that we consider convergence for every $\omega \in \Omega$ separately. This notion of almost sure convergence is the weakest one. It can happen that an ε-optimal solution converges almost surely, although it does not converge in outer probability sense. This concept is investigated in [17] where the author calls it "sample-path optimization". We started this approach to sensitivity of stochastic optimization in [13].

As an illustration of our general result we consider ε-M-estimator in linear regression model. There is a vast literature on the linear regression model, e. g.

[1, 2, 7, 8, 9, 12, 14, 16, 18], 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, etc. Considered ε-M-estimator is a particular case of Asymptotically Optimal Estimators (AOE) introduced in [22]. Conditions under which the AOEs are consistent in probability are discussed in [22]. We present result on consistency almost surely.

Our paper requires no prescribed structure for observations and errors. We only assume weak convergence of their common "empirical measure". Also, we allow nonuniqness of the estimator and we do not require measurability of the estimator.

The last preliminary note concerns notation. For our purposes we need a bit stronger notion of the standard weak convergence of probability measures. Therefore we have to introduce a convenient notation.

Definition 1. Let μ, μ_n, $n \in \mathbb{N}$ be Borel probability measures on a metric space $\mathcal{Y}$ and $\mathcal{F} \subset \{f : \mathcal{Y} \to \mathbb{R} \mid f \text{ is measurable}\}$. We will say that μ_n converge $\mathcal{F}$-weakly to μ iff

$$\int_{\mathcal{Y}} f(y)\mu_n(\mathrm{d}y) \xrightarrow[n\to+\infty]{} \int_{\mathcal{Y}} f(y)\mu(\mathrm{d}y) \qquad \text{for every bounded continuous function } f : \mathcal{Y} \to \mathbb{R};$$

$$\int_{\mathcal{Y}} f(y)\mu_n(\mathrm{d}y) \xrightarrow[n\to+\infty]{} \int_{\mathcal{Y}} f(y)\mu(\mathrm{d}y) \qquad \text{for every } f \in \mathcal{F}.$$

We will denote the convergence by

$$\mu_n \xrightarrow[n\to+\infty]{\mathcal{F}-\mathrm{w}} \mu.$$

Easy observation shows that $\mu_n \xrightarrow[n\to+\infty]{\mathcal{F}-\mathrm{w}} \mu$ implies

$$\int_{\mathcal{Y}} f(y)\mu_n(\mathrm{d}y) \xrightarrow[n\to+\infty]{} \int_{\mathcal{Y}} f(y)\mu(\mathrm{d}y) \quad \text{for every } f \in \overline{\mathcal{F}},$$

where $\overline{\mathcal{F}}$ denotes closure in supremal norm of linear hull of union $\mathcal{F}$ and all bounded continuous functions.

For empty set of functions $\mathcal{F}$ we are receiving the standard weak convergence. Therefore we will accept notation $\mu_n \xrightarrow[n\to+\infty]{\mathrm{w}} \mu$ instead of $\mu_n \xrightarrow[n\to+\infty]{\emptyset-\mathrm{w}} \mu$.

Let us note, that this stronger version of the weak convergence takes place very often. For example, the strong law of large numbers for i.i.d. random variables can be rewritten as

$$\nu_n \xrightarrow[n\to+\infty]{\mathcal{H}-\mathrm{w}} \nu\,, \quad \text{where}$$

ν_n is the empirical measure defined from observations till time n, ν is the common distribution of observations and $\mathcal{H} = \{h : x \in \mathcal{X} \to |x|\}$.

Consequently, this strong law of large numbers can be written as

$$\int_{\mathcal{Y}} f(y)\mu_n(\mathrm{d}y) \xrightarrow[n\to+\infty]{} \int_{\mathcal{Y}} f(y)\mu(\mathrm{d}y)$$

for every continuous function $f : \mathcal{Y} \to \mathbb{R}$ fulfilling $|f(y)| \le A + B|y|$ for all $y \in \mathcal{Y}$ and convenient $A, B \in \mathbb{R}$.

2. GENERAL RESULT

We consider an optimization problem written in the form

$$\inf \{ \mathsf{f}\,(x \,|\, \mu_0) \ |\ x \in \mathcal{X} \}, \tag{1}$$

where $\mu_0 \in \mathcal{P}$. We suppose $\mathcal{X}$ to be a metric space, $\mathcal{P}$ be a family of probability measures defined on a metric space $\mathcal{Y}$ and $\mathsf{f} : \mathcal{X} \times \mathcal{P} \to \overline{\mathbb{R}}$, where $\overline{\mathbb{R}} = [-\infty, +\infty]$ denotes the extended real line.

The objective function is known up to unknown probability measure μ_0. We assume a procedure producing an estimation of μ_0. We suppose to observe $z_t \in \mathcal{Z}_t$ at any time $t \in \mathbb{N}$. Let us note, that the setting covers the most typical situation if one observes a sequence of data $w_1, w_2, \ldots, w_{k_t}$ belonging to a metric space $\mathcal{W}$. Hence, we group observations available at time $t \in \mathbb{N}$ in a vector $z_t = (w_1, w_2, \ldots, w_{k_t})$ and $\mathcal{Z}_t = \mathcal{W}^{k_t}$. From observed data we construct probability measure $\mu_t\,(\bullet \,|\, z_t)$ on $\mathcal{Y}$. These measures will play role of estimators for the "true" probability measure μ_0.

Let us introduce a denotation of objects of our interest. For a given function $f : \mathcal{X} \to \overline{\mathbb{R}}$ we are interested in minimal value, minimal solutions, and ε-minimal solutions

$$\varphi\,(f) = \inf \{ f(x) \ |\ x \in \mathcal{X} \}\ , \tag{2}$$

$$\Phi\,(f) = \{ x \in \mathcal{X} \ |\ f(x) = \varphi\,(f) \}\ . \tag{3}$$

Having $\varphi\,(f) \in \mathbb{R}$ we can deal with

$$\Psi\,(f;\varepsilon) = \{ x \in \mathcal{X} \ |\ f(x) \le \varphi\,(f) + \varepsilon \} \quad \forall\, \varepsilon > 0\,. \tag{4}$$

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

Assumption A1. $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}_t$, $t \in \mathbb{N}$ are metric spaces.

Assumption A2. $\mathcal{P}$ is a nonempty subset of all Borel probability measures on $\mathcal{Y}$ and $\mathcal{F} \subset \{ f : \mathcal{Y} \to \mathbb{R} \ |\ f \text{ is Borel measurable} \}$.
(The set $\mathcal{F}$ is allowed to be empty.)

Assumption A3. The function $\mathsf{f} \ :\ \mathcal{X} \times \mathcal{P} \to \overline{\mathbb{R}}$.

Assumption A4. $\varepsilon_t > 0$ for any $t \in \mathbb{N}$ and $\bar{\varepsilon} = \limsup_{t\to+\infty} \varepsilon_t < +\infty$.

Assumption A5. For any $t \in \mathbb{N}$, we observe $Z_t : \Omega \to \mathcal{Z}_t$.

Assumption A6. For any $t \in \mathbb{N}$, $z_t \in \mathcal{Z}_t$, $\mu_t(\bullet \mid z_t)$ is a Borel probability measure on $\mathcal{Y}$.
We denote $\mathcal{P}_{\text{emp}} = \{\mu_t(\bullet \mid z_t) \mid z_t \in \mathcal{Z}_t, t \in \mathbb{N}\}$.

Assumption A7. $\mu_0 \in \mathcal{P}$ and $\mathcal{P}_{\text{emp}} \subset \mathcal{P}$.

Assumption A8. $\varphi(\mathsf{f}(\bullet \mid \nu)) \in \mathbb{R}$ for every $\nu \in \mathcal{P}$.

Assumption A9. Whenever $\forall n \in \mathbb{N}\ \nu_n \in \mathcal{P}_{\text{emp}}$ and $\nu_n \xrightarrow[n\to+\infty]{\mathcal{F}-\text{w}} \mu_0$, then there is a convergent sequence $\tilde{\theta}_n \in \mathcal{X}$, $n \in \mathbb{N}$ such that

$$\lim_{n\to+\infty} \tilde{\theta}_n \in \Phi(\mathsf{f}(\bullet \mid \mu_0)) \quad \text{and} \quad \limsup_{n\to+\infty} \mathsf{f}\left(\tilde{\theta}_n \mid \nu_n\right) \le \varphi(\mathsf{f}(\bullet \mid \mu_0)) \,.$$

Assumption A10. There is a compact set $\mathsf{K} \subset \mathcal{X}$ such that

1. $\liminf\limits_{n\to+\infty} \mathsf{f}(\theta_n \mid \nu_n) \ge \mathsf{f}(\theta \mid \mu_0)$ whenever
   $\forall n \in \mathbb{N}\ \nu_n \in \mathcal{P}_{\text{emp}}$ and $\nu_n \xrightarrow[n\to+\infty]{\mathcal{F}-\text{w}} \mu_0$,
   $\forall n \in \mathbb{N}\ \theta_n \in \mathcal{X}$, $\theta_n \xrightarrow[n\to+\infty]{} \theta \in \mathsf{K}$.
2. For any sequence of probability measures $\nu_n \in \mathcal{P}_{\text{emp}}$, $\nu_n \xrightarrow[n\to+\infty]{\mathcal{F}-\text{w}} \mu_0$ and any open set $G \supset \mathsf{K}$ we have

$$\liminf_{n\to+\infty} \inf_{\theta\in\mathcal{X}\setminus G} \mathsf{f}(\theta \mid \nu_n) > \varphi(\mathsf{f}(\bullet \mid \mu_0)) + \bar{\varepsilon}\,.$$

In Assumption A5, Ω is just a set (of elementary events). Only later on, it will be considered as a probability space $(\Omega, \mathcal{A}, \mathsf{prob})$. Assumptions A9 and A10 can be expressed by means of lower and upper approximations treated in [15]. All of these assumptions ensure existence of ε-minimal solutions and their consistency.

Lemma 1. Let $\omega \in \Omega$ and assumptions A1 – A8 are fulfilled, then always $\Psi(\mathsf{f}(\bullet \mid \mu_t(\bullet \mid Z_t(\omega)));\varepsilon_t) \ne \emptyset$ for any $t \in \mathbb{N}$.

Proof. An ε-minimal solution exists since $\varphi(\mathsf{f}(\bullet \mid \mu_t(\bullet \mid Z_t(\omega)))) \in \mathbb{R}$, accordingly to Assumption A8. $\square$

We have to recall a few from topological terminology.

Definition 2. For a sequence η_n, $n \in \mathbb{N}$ in a metric space $\mathcal{W}$, we denote the set of its cluster points by $\mathsf{Ls}(\eta_n\,,\, n \in \mathbb{N})$, i.e.

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

Definition 3. 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 $\{\eta_n \mid n \in \mathbb{N}\} \cup \mathsf{Ls}\,(\eta_n\,,\, n \in \mathbb{N})$ is compact.

Lemma 3. Let η_n, $n \in \mathbb{N}$ be a sequence in a metric space $\mathcal{W}$ and $K \subset \mathcal{W}$ be a compact. If for every open set $G \supset K$ there is an $n_G \in \mathbb{N}$ such that $\eta_n \in G$ for all $n \in \mathbb{N}$, $n \geq n_G$, then the sequence is compact and $\mathsf{Ls}\,(\eta_n\,,\, n \in \mathbb{N}) \subset K$.

Other details and proofs can be found in any monograph on topology, e. g. [11]. These topological notions will be used in proof of following results.

Theorem 1. Let $\omega \in \Omega$ be such that

$$\mu_t\,(\bullet \,|\, Z_t(\omega)) \xrightarrow[t \to +\infty]{\mathcal{F}-\mathrm{w}} \mu_0$$

and Assumptions A1 – A10 be fulfilled. Then, $\Phi\,(\mathsf{f}\,(\bullet \,|\, \mu_0)) \cap \mathsf{K} \neq \emptyset$.

If $\hat{\theta}_t \in \Psi\,(\mathsf{f}\,(\bullet \,|\, \mu_t\,(\bullet \,|\, Z_t(\omega)))\,;\varepsilon_t)$ for any $t \in \mathbb{N}$ then the sequence $\hat{\theta}_t(\omega)$, $t \in \mathbb{N}$ is compact and

$$\emptyset \neq \mathsf{Ls}\left(\hat{\theta}_t(\omega)\,,\, t \in \mathbb{N}\right) \subset \Psi\,(\mathsf{f}\,(\bullet \,|\, \mu_0)\,;\bar{\varepsilon}) \cap \mathsf{K}\,. \tag{5}$$

Proof.

1. Assumption A9 implies existence of a sequence $\tilde{\theta}_t \in \mathcal{X}$, $t \in \mathbb{N}$ such that

$$\lim_{t \to +\infty} \tilde{\theta}_t \in \Phi\,(\mathsf{f}\,(\bullet \,|\, \mu_0))\;,$$
$$\limsup_{t \to +\infty} \mathsf{f}\left(\tilde{\theta}_t \,|\, \mu_t\,(\bullet \,|\, Z_t(\omega))\right) \leq \varphi\,(\mathsf{f}\,(\bullet \,|\, \mu_0))\;.$$

2. Assumption A10 implies that for every open set $G \supset \mathsf{K}$ we have

$$\limsup_{t \to +\infty} \mathsf{f}\left(\tilde{\theta}_t \,|\, \mu_t\,(\bullet \,|\, Z_t(\omega))\right) \leq \varphi\,(\mathsf{f}\,(\bullet \,|\, \mu_0)) \leq \varphi\,(\mathsf{f}\,(\bullet \,|\, \mu_0)) + \bar{\varepsilon}$$
$$< \liminf_{t \to +\infty} \inf_{\theta \in \mathcal{X} \setminus G} \mathsf{f}\,(\theta \,|\, \mu_t\,(\bullet \,|\, Z_t(\omega)))\;.$$

Therefore, $\tilde{\theta}_t \in G$ for all $t \in \mathbb{N}$ sufficiently large.

Hence according to Lemma 3,

$$\lim_{t\to+\infty} \tilde{\theta}_t \in \mathsf{K}\,.$$

We have found a point in $\mathsf{K} \cap \Phi\left(\mathsf{f}\left(\bullet \,|\, \mu_0\right)\right)$.

3. Assumptions A4, A9, A10 imply that for every open set $G \supset \mathsf{K}$ we have

$$\begin{aligned}
&\limsup_{t\to+\infty}\ \mathsf{f}\left(\hat{\theta}_t(\omega) \,|\, \mu_t\left(\bullet \,|\, Z_t(\omega)\right)\right) \\
&\le \limsup_{t\to+\infty}\left[\mathsf{f}\left(\tilde{\theta}_t \,|\, \mu_t\left(\bullet \,|\, Z_t(\omega)\right)\right) + \varepsilon_t\right] \\
&= \varphi\left(\mathsf{f}\left(\bullet \,|\, \mu_0\right)\right) + \bar{\varepsilon} < \liminf_{t\to+\infty}\ \inf_{\theta\in\mathcal{X}\setminus G} \mathsf{f}\left(\theta \,|\, \mu_t\left(\bullet \,|\, Z_t(\omega)\right)\right).
\end{aligned}$$

Therefore, $\hat{\theta}_t(\omega) \in G$ for all $t \in \mathbb{N}$ sufficiently large.

Hence according to Lemma 3 the sequence $\hat{\theta}_t(\omega)$, $t \in \mathbb{N}$ is compact and

$$\emptyset \neq \mathsf{Ls}\left(\hat{\theta}_t(\omega)\,,\, t \in \mathbb{N}\right) \subset K.$$

4. Let $\eta \in \mathsf{K}$ be a cluster point of the sequence $\hat{\theta}_t(\omega)$, $t \in \mathbb{N}$ and take a subnet with $\lim_{n\to+\infty} \hat{\theta}_{t_n}(\omega) = \eta$. According to Assumptions A4, A8, A9, A10, we receive

$$\begin{aligned}
\mathsf{f}\left(\eta \,|\, \mu_0\right) &\le \liminf_{n\to+\infty} \mathsf{f}\left(\hat{\theta}_{t_n}(\omega) \,|\, \mu_{t_n}\left(\bullet \,|\, z_{t_n}(\omega)\right)\right) \\
&\le \limsup_{n\to+\infty} \mathsf{f}\left(\hat{\theta}_{t_n}(\omega) \,|\, \mu_{t_n}\left(\bullet \,|\, z_{t_n}(\omega)\right)\right) \\
&\le \limsup_{n\to+\infty}\left[\mathsf{f}\left(\tilde{\theta}_{t_n} \,|\, \mu_{t_n}\left(\bullet \,|\, z_{t_n}(\omega)\right)\right) + \varepsilon_{t_n}\right] \\
&\le \varphi\left(\mathsf{f}\left(\bullet \,|\, \mu_0\right)\right) + \bar{\varepsilon}\,.
\end{aligned}$$

Hence,

$$\eta \in \Psi\left(\mathsf{f}\left(\bullet \,|\, \mu_0\right); \bar{\varepsilon}\right)\,. \qquad \square$$

Consider in addition Ω to be a probability space $(\Omega, \mathcal{A}, \mathsf{prob})$. Then, if the weak convergence take place for almost all $\omega \in \Omega$ we have a convergence in almost sure sense.

Theorem 2. Let $\Omega_0 \subset \Omega$ be such that $\mathsf{prob}_*\left(\Omega_0\right) = 1$ and for all $\omega \in \Omega_0$

$$\mu_t\left(\bullet \,|\, Z_t(\omega)\right) \xrightarrow[t\to+\infty]{\mathcal{F}-\mathrm{w}} \mu_0$$

and Assumptions A1 – A10 be fulfilled. If $\hat{\theta}_t \in \Psi\left(\mathsf{f}\left(\bullet \,|\, \mu_t\left(\bullet \,|\, Z_t(\omega)\right)\right); \varepsilon_t\right)$ for any $t \in \mathbb{N}$ then

$$\rho\left(\hat{\theta}_t\,,\ \Psi\left(\mathsf{f}\left(\bullet \,|\, \mu_0\right); \bar{\varepsilon}\right) \cap \mathsf{K}\right) \xrightarrow[t\to+\infty]{\text{a s}} 0\,, \tag{6}$$

where ρ is the metric of the metric space $\mathcal{X}$.

P r o o f. For (6) it is sufficient to show

$$\lim_{t\to+\infty} \rho\left(\hat{\theta}_t(\omega), \Psi\left(\mathsf{f}\left(\bullet \,|\, \mu_0\right); \bar{\varepsilon}\right) \cap \mathsf{K}\right) = 0\,,\ \text{for any } \omega \in \Omega_0.$$

Let us assume $\omega \in \Omega_0$ and $\Delta > 0$ such that

$$\limsup_{t\to+\infty} \rho\left(\hat{\theta}_t(\omega), \Psi\left(\mathsf{f}\left(\bullet \,|\, \mu_0\right); \bar{\varepsilon}\right) \cap \mathsf{K}\right) > \Delta\,.$$

Therefore, there is a subnet such that

$$\rho\left(\hat{\theta}_{t_n}(\omega), \Psi\left(\mathsf{f}\left(\bullet \,|\, \mu_0\right); \bar{\varepsilon}\right) \cap \mathsf{K}\right) > \Delta \text{ for all } n \in \mathbb{N}\,.$$

According to Theorem 1 the sequence $\hat{\theta}_{t_n}(\omega)$, $n \in \mathbb{N}$ is compact and, hence, its subsequence possesses a cluster point in $\Psi\left(\mathsf{f}\left(\bullet \,|\, \mu_0\right); \bar{\varepsilon}\right) \cap \mathsf{K}$. Any such cluster point η must fulfill $\rho\left(\eta, \Psi\left(\mathsf{f}\left(\bullet \,|\, \mu_0\right); \bar{\varepsilon}\right) \cap \mathsf{K}\right) \geq \Delta$.

Consequently, $\eta \notin \Psi\left(\mathsf{f}\left(\bullet \,|\, \mu_0\right); \bar{\varepsilon}\right) \cap \mathsf{K}$ which contradicts our assumption.

Thus, (6) is proved. □

Our proof treats any trajectory separately. Therefore, we do not need measurability of $\mu_t\left(\bullet \,|\, z_t\right)$ with respect to $z_t \in \mathcal{Z}_t$. Also, selection of the ε_t-minimal solution does not require measurability. Thus, it can naturally happen that the considered objects are not random variables.

Theorem 3. Let $\Omega_0 \subset \Omega$ be such that $\mathsf{prob}_*\left(\Omega_0\right) = 1$ and for all $\omega \in \Omega_0$

$$\mu_t\left(\bullet \,|\, Z_t(\omega)\right) \xrightarrow[t\to+\infty]{\mathcal{F}-\text{w}} \mu_0$$

and Assumptions A1 – A10 be fulfilled. If $\hat{\theta}_t \in \Psi\left(\mathsf{f}\left(\bullet \,|\, \mu_t\left(\bullet \,|\, Z_t(\omega)\right)\right); \varepsilon_t\right)$ for any $t \in \mathbb{N}$ is asymptotically measurable then

$$\rho\left(\hat{\theta}_t, \Psi\left(\mathsf{f}\left(\bullet \,|\, \mu_0\right); \bar{\varepsilon}\right) \cap \mathsf{K}\right) \xrightarrow[t\to+\infty]{\mathsf{prob}^*} 0\,. \tag{7}$$

P r o o f. Proof follows immediately from previous Theorem 2 and Lemma 9, in Appendix, since the limit is deterministic. □

Theorem 4. Let $\Omega_0 \subset \Omega$ be such that $\mathsf{prob}_*(\Omega_0) = 1$ and for all $\omega \in \Omega_0$

$$\mu_t\left(\bullet \,|\, Z_t(\omega)\right) \xrightarrow[t\to+\infty]{\mathcal{F}-\mathrm{w}} \mu_0$$

and Assumptions A1 – A10 be fulfilled. If $\hat{\theta}_t \in \Psi\left(\mathsf{f}\left(\bullet \,|\, \mu_t\left(\bullet \,|\, Z_t(\omega)\right)\right); \varepsilon_t\right)$ for any $t \in \mathbb{N}$ is strongly asymptotically measurable then

$$\rho\left(\hat{\theta}_t, \Psi\left(\mathsf{f}\left(\bullet \,|\, \mu_0\right); \bar{\varepsilon}\right) \cap \mathsf{K}\right) \xrightarrow[t\to+\infty]{\mathrm{a\,s}^*} 0\,, \tag{8}$$

$$\rho\left(\hat{\theta}_t, \Psi\left(\mathsf{f}\left(\bullet \,|\, \mu_0\right); \bar{\varepsilon}\right) \cap \mathsf{K}\right) \xrightarrow[t\to+\infty]{\mathsf{prob}^*} 0\,. \tag{9}$$

Proof. Proof follows immediately from previous Theorem 2 and Lemmas 7, 8, in Appendix, since the limit is deterministic. □

3. LINEAR REGRESSION

As an example illustrating the theory presented in the previous section we offer a linear regression model. Where, unknown regression coefficients are estimated by an ε_t-M-estimator.

We suppose to observe couples $(Y_1, X_1), (Y_2, X_2), \ldots, (Y_t, X_t)$ connected by a linear regression model

$$Y_i = X_i^\top \beta_0 + \mathrm{e}_i \quad \forall\; i = 1, 2, \ldots, t. \tag{10}$$

Where $Y_i : \Omega \to \mathbb{R}$, $X_i : \Omega \to \mathbb{R}^d$ are mappings, $\mathrm{e}_i : \Omega \to \mathbb{R}$ are unobserved mappings and $\beta_0 \in \Theta \subset \mathbb{R}^d$ is deterministic but unknown parameter.

Parameter set Θ expresses our a prior information about parameters. From the problem setting, we can 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.

As probability measures required in Assumption A6 we will employ empirical probability measure defined from observations. Let us define denotation of an empirical probability measure in a general case. Let $\mathcal{W} \neq \emptyset$ and $w_1, w_2, \ldots, w_t \in \mathcal{W}$. Then, the empirical probability measure is defined for any $A \subset \mathcal{W}$ as the relative number of observations hitting the set A, i. e. by the formula

$$\mathcal{E}_t\left(A \,|\, w_1, w_2, \ldots, w_t\right) = \frac{1}{t}\sum_{i=1}^{t} \mathbb{I}[w_i \in A]. \tag{11}$$

Let us recall that if $\mathcal{W}$ is a metric space then empirical probability measure restricted to Borel σ-algebra of $\mathcal{W}$ is a Borel probability measure.

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

$$\mathsf{f}\left(\beta \,|\, \mu\right) = \int \rho(y - x^\top \beta)\mu(\mathrm{d}y, \mathrm{d}x). \tag{12}$$

Especially, for empirical distribution of observations we receive

$$\begin{aligned}
&\mathsf{f}\left(\beta \,|\, \mathcal{E}_t\left(\bullet \,|\, (y_1,x_1),(y_2,x_2),\ldots,(y_t,x_t)\right)\right) \\
&= \int \rho(y - x^\top \beta)\ \mathcal{E}_t\left(\mathrm{d}y, \mathrm{d}x \,|\, (y_1,x_1),(y_2,x_2),\ldots,(y_t,x_t)\right) \\
&= \frac{1}{t}\sum_{i=1}^{t} \rho(y_i - x_i^\top \beta).
\end{aligned}$$

An ε_t-M-estimator is any $\hat{\beta}_t \in \Theta$ fulfilling for all $\beta \in \Theta$

$$\begin{aligned}
&\mathsf{f}\left(\hat{\beta}_t \,|\, \mathcal{E}_t\left(\bullet \,|\, (Y_1,X_1),(Y_2,X_2),\ldots,(Y_t,X_t)\right)\right) : \\
&\quad \le \mathsf{f}\left(\beta \,|\, \mathcal{E}_t\left(\bullet \,|\, (Y_1,X_1),(Y_2,X_2),\ldots,(Y_t,X_t)\right)\right) + \varepsilon_t.
\end{aligned} \tag{13}$$

Now, the studied situation is fully described and we are proceeding to assumptions. We introduce the following list of assumptions:

Assumption R1. $\Theta \subset \mathbb{R}^d$ is a closed subset.

Assumption R2. $\varepsilon_t > 0$ for any $t \in \mathbb{N}$ and $\limsup_{n\to+\infty} \varepsilon_t = \bar{\varepsilon}$.

Assumption R3. There are a Borel measure ν defined on $\mathbb{R}^{d+1}$ and $\Omega_1 \subset \Omega$ such that $\mathsf{prob}_*(\Omega_1) = 1$ and for all $\omega \in \Omega_1$

$$\mathcal{E}_t\left(\bullet \,|\, (X_1(\omega), \mathrm{e}_1(\omega)), (X_2(\omega), \mathrm{e}_2(\omega)), \ldots, (X_t(\omega), \mathrm{e}_t(\omega))\right) \xrightarrow[n\to+\infty]{\mathrm{w}} \nu \,.$$

Assumption R4. For any $\beta \in \Theta$

$$\int \rho(e)\,\nu\left(\mathrm{d}x, \mathrm{d}e\right) \le \int \rho(e + x^\top(\beta_0 - \beta))\,\nu\left(\mathrm{d}x, \mathrm{d}e\right).$$

Assumption R5. Function $\rho : \mathbb{R} \to \mathbb{R}$ is nonnegative and continuous.

Assumption R6. There are a function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ which is nondecreasing continuous and $\Omega_2 \subset \Omega$, $\mathsf{prob}_*(\Omega_2) = 1$ fulfilling:

1. For all $t \in \mathbb{R}$ $\quad \rho(t) \le \psi(|t|)$.
2. For all $t > 0$ $\quad \int \psi(|e| + t\|x\|)\nu\left(\mathrm{d}x, \mathrm{d}e\right) < +\infty$.
3. For all $t > 0$, $\omega \in \Omega_2$
   $\frac{1}{t}\sum_{i=1}^{t} \psi(|\mathrm{e}_i(\omega)| + t\|X_i(\omega)\|) \xrightarrow[n\to+\infty]{} \int \psi(|e| + t\|x\|)\nu\left(\mathrm{d}x, \mathrm{d}e\right)$.

Assumption R7. Denoting

$$\begin{aligned}
\mathsf{H}_\rho &= \liminf_{\Delta\to+\infty} \inf\left\{\rho(t) \;\middle|\; |t| > \Delta, t \in \mathbb{R}\right\}, \\
\mathsf{M} &= \inf\left\{\nu\left(\left\{(x,e) \in \mathbb{R}^{d+1} \;\middle|\; x^\top\gamma \ne 0\right\}\right) \;\middle|\; \|\gamma\| = 1, \gamma \in \mathbb{R}^d\right\},
\end{aligned}$$

we require $\mathsf{M} > 0$ and a balance

$$\mathsf{H}_\rho \mathsf{M} > \int \rho(e)\, \nu\, (\mathrm{d}x, \mathrm{d}e) + \bar{\varepsilon}.$$

Lemma 4. For any $\Delta > 0$ we have

$$\lim_{\kappa \to +\infty} \inf_{\|\gamma\|=1} \nu \left(\left\{ (x,e) \;\middle|\; \kappa \left| x^\top \gamma \right| \geq \Delta + |e| \right\} \right) = \mathsf{M}\,. \tag{14}$$

Proof. Let $\Delta > 0$.

1. Let $\gamma \in \mathbb{R}^d$ and $k \in \mathbb{N}$, then

$$\left\{ (x,e) \;\middle|\; k \left| x^\top \gamma \right| \geq \Delta + |e| \right\} \subset \left\{ (x,e) \;\middle|\; x^\top \gamma \neq 0 \right\}\,.$$

Consequently,

$$\limsup_{\kappa \to +\infty} \inf_{\|\gamma\|=1} \nu \left(\left\{ (x,e) \;\middle|\; \kappa \left| x^\top \gamma \right| \geq \Delta + |e| \right\} \right) \leq \mathsf{M}\,.$$

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

$$\nu \left(\left\{ (x,e) \;\middle|\; k \left| x^\top \gamma_k \right| \geq \Delta + |e| \right\} \right) < \inf_{\|\gamma\|=1} \nu \left(\left\{ (x,e) \;\middle|\; k \left| x^\top \gamma \right| \geq \Delta + |e| \right\} \right) + \frac{1}{k}\,.$$

The sequence belongs to a compact, therefore, it contains a subsequence $\gamma_{k_j}, j \in \mathbb{N}$ and $\hat{\gamma} \in \mathbb{R}^d$, $\|\hat{\gamma}\| = 1$ such that

$$\begin{aligned}
&\lim_{j \to +\infty} \gamma_{k_j} = \hat{\gamma}, \\
&\lim_{j \to +\infty} \nu \left(\left\{ (x,e) \;\middle|\; k_j \left| x^\top \gamma_{k_j} \right| \geq \Delta + |e| \right\} \right) \\
&\qquad = \liminf_{k \to +\infty} \nu \left(\left\{ (x,e) \;\middle|\; k \left| x^\top \gamma_k \right| \geq \Delta + |e| \right\} \right)\,.
\end{aligned}$$

Hence, we have inclusion

$$\begin{aligned}
&\left\{ (x,e) \;\middle|\; x^\top \hat{\gamma} \neq 0 \right\} \\
&\quad = \bigcup_{J=1}^{+\infty} \bigcap_{j=J}^{+\infty} \left\{ (x,e) \;\middle|\; \left| x^\top \gamma_{k_j} \right| \geq \left| x^\top (\hat{\gamma} - \gamma_{k_j}) \right| + \frac{1}{k_j} (\Delta + |e|) \right\} \\
&\quad \subset \bigcup_{J=1}^{+\infty} \bigcap_{j=J}^{+\infty} \left\{ (x,e) \;\middle|\; k_j \left| x^\top \gamma_{k_j} \right| \geq \Delta + |e| \right\}\,.
\end{aligned}$$

Because of σ-additivity of the measure ν, we have

$$\liminf_{\kappa \to +\infty} \inf_{\|\gamma\|=1} \nu \left(\left\{ (x,e) \;\middle|\; \kappa \left| x^\top \gamma \right| \geq \Delta + |e| \right\} \right)$$

$$
\begin{aligned}
&= \liminf_{k\to+\infty} \nu\left(\left\{(x,e) \;\middle|\; k\left|x^\top\gamma_k\right| \geq \Delta + |e|\right\}\right) \\
&= \lim_{j\to+\infty} \nu\left(\left\{(x,e) \;\middle|\; k_j\left|x^\top\gamma_{k_j}\right| \geq \Delta + |e|\right\}\right) \\
&\geq \nu\left(\bigcup_{J=1}^{+\infty}\bigcap_{j=J}^{+\infty}\left\{(x,e) \;\middle|\; k_j\left|x^\top\gamma_{k_j}\right| \geq \Delta + |e|\right\}\right) \\
&\geq \nu\left(\left\{(x,e) \;\middle|\; x^\top\hat\gamma \neq 0\right\}\right) \geq \mathsf{M}\,.
\end{aligned}
$$

□

Theorem 5. Let Assumptions R1 – R7 be fulfilled and set

$$
\mathsf{g}(\beta) = \int \rho(e + x^\top(\beta_0 - \beta))\,\nu\,(\mathrm{d}x, \mathrm{d}e) \quad \forall \beta \in \Theta\,. \tag{15}
$$

Then, $\beta_0 \in \Phi\,(\mathsf{g})$ and an estimator $\hat\beta_t(\omega)$ exists for every $t \in \mathbb{N}$, $\omega \in \Omega$. For every $\omega \in \Omega_0 = \Omega_1 \cap \Omega_2$ the sequence $\hat\beta_t(\omega)$, $t \in \mathbb{N}$ is compact and

$$
\emptyset \neq \mathsf{Ls}\left(\hat\theta_t(\omega)\,,\, t \in \mathbb{N}\right) \subset \Psi\,(\mathsf{g}; \bar\varepsilon)\;. \tag{16}
$$

If, moreover, the sequence $\hat\beta_t$, $t \in \mathbb{N}$ is asymptotically measurable then

$$
\mathsf{d}\left(\hat\beta_t, \Psi\,(\mathsf{g}\,; \bar\varepsilon)\right) \xrightarrow[t\to+\infty]{\mathsf{prob}^*} 0\,. \tag{17}
$$

If the sequence $\hat\beta_t$, $t \in \mathbb{N}$ is strongly asymptotically measurable then

$$
\mathsf{d}\left(\hat\beta_t, \Psi\,(\mathsf{g}\,; \bar\varepsilon)\right) \xrightarrow[t\to+\infty]{\mathrm{a\,s}^*} 0\,, \tag{18}
$$

$$
\mathsf{d}\left(\hat\beta_t, \Psi\,(\mathsf{g}\,; \bar\varepsilon)\right) \xrightarrow[t\to+\infty]{\mathsf{prob}^*} 0\,. \tag{19}
$$

Proof. We will show that this theorem is a particular case of Theorem 2. We set $\mathcal{X} = \mathcal{Y} = \mathbb{R}^{d+1}$,

$$
\begin{aligned}
&\mu_0\,(A) = \int \mathbb{I}[(e + x^\top\beta_0, x) \in A]\,\nu\,(\mathrm{d}x, \mathrm{d}e) \quad \forall A \text{ Borel subset of } \mathbb{R}^{d+1}\,, \\
&\mathcal{P}_{\mathrm{emp}} = \left\{\mathcal{E}_t\left(\bullet \,|\, (y_1, x_1), \ldots, (y_t, x_t)\right) \;\middle|\; y_1, \ldots, y_t \in \mathbb{R},\ x_1, \ldots, x_t \in \mathbb{R}^d,\ n \in \mathbb{N}\right\}\,, \\
&\mathcal{P} = \mathcal{P}_{\mathrm{emp}} \cup \{\mu_0\}\,, \\
&\mathcal{F} = \left\{(y, x) \mapsto \psi(|y - x^\top\beta_0| + t\|x\|) \;\middle|\; t > 0\right\}\,.
\end{aligned}
$$

We will show that Assumptions A1 – A6 are fulfilled.

1. Loss function is nonnegative, since ρ is nonnegative by Assumption R5, real and defined on $\Theta \times \mathcal{P}$. Particularly for $\beta \in \Theta$ and $\mu \in \mathcal{P}_{\mathrm{emp}}$

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

and for $\beta \in \Theta$,

$$\mathsf{f}\,(\beta\,|\,\mu_0) \quad = \quad \int \rho(e + x^\top(\gamma - \beta))\,\nu\,(\mathrm{d}x, \mathrm{d}e) \in \mathbb{R}$$

because of Assumption R6 we have

$$\mathsf{f}\,(\beta\,|\,\mu_0) \quad \leq \quad \int \psi(|e| + \|\gamma - \beta\|\|x\|)\,\nu\,(\mathrm{d}x, \mathrm{d}e) < +\infty.$$

Thus, Assumptions A3 and A8 are valid.

2. Let $\nu_n \in \mathcal{P}_{\mathrm{emp}}$.

   The measures possess an expression

$$\nu_n = \mathcal{E}_{k_n}\left(\bullet\,|\,(y_1^{(n)}, x_1^{(n)}), (y_2^{(n)}, x_2^{(n)}), \dots, (y_{k_n}^{(n)}, x_{k_n}^{(n)})\right)\,.$$

   Let us denote

$$\begin{aligned}
&\mathcal{H} = \left\{(x,e) \mapsto \psi(e| + t\|x\|)\ \mid t > 0\right\}\,,\\
&\mathrm{e}_i^{(n)} = y_i^{(n)} - s x_i^{(n)\top} s\beta_0 \quad \text{for all } i = 1, 2, \dots, k_n,\\
&\xi_n = \mathcal{E}_{k_n}\left(\bullet\,|\,(x_1^{(n)}, \mathrm{e}_1^{(n)}), (x_2^{(n)}, \mathrm{e}_2^{(n)}), \dots, (x_{k_n}^{(n)}, \mathrm{e}_{k_n}^{(n)})\right)\,.
\end{aligned}$$

   Then for functions f, g, we have

$$\begin{aligned}
&\int f(x,e)\xi_n(\mathrm{d}x, \mathrm{d}e) = \int f(x, y - x^\top\beta_0)\nu_n(\mathrm{d}y, \mathrm{d}x)\,,\\
&\int f(x,e)\nu\,(\mathrm{d}x, \mathrm{d}e) = \int f(x, y - x^\top\beta_0)\mu_0\,(\mathrm{d}y, \mathrm{d}x)\ ,\\
&\int g(y,x)\nu_n(\mathrm{d}y, \mathrm{d}x) = \int g(x^\top\beta_0 + e, x)\xi_n(\mathrm{d}x, \mathrm{d}e)\,,\\
&\int g(y,x)\mu_0\,(\mathrm{d}y, \mathrm{d}x) = \int g(x^\top\beta_0 + e, x)\nu\,(\mathrm{d}x, \mathrm{d}e)\ ,
\end{aligned}$$

   whenever the integral exist. Therefore, we have the equivalence

$$\nu_n \xrightarrow[n\to+\infty]{\mathcal{F}-\mathrm{w}} \mu_0 \quad \Longleftrightarrow \quad \xi_n \xrightarrow[n\to+\infty]{\mathcal{H}-\mathrm{w}} \nu\ .$$

3. Let $\nu_n \in \mathcal{P}_{\mathrm{emp}}$, $\beta, \beta_n \in \Theta$, $\nu_n \xrightarrow[n\to+\infty]{\mathcal{F}-\mathrm{w}} \mu_0$ 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

$$\begin{aligned}
&\|\beta_n - \beta_0\| \leq T \quad \text{for all } n \in \mathbb{N},\\
&\int_{\psi(|e| + T\|x\|) > Q} (\psi(|e| + T\|x\|) - Q)\,\nu\,(\mathrm{d}x, \mathrm{d}e) < \varepsilon,\\
&\nu_n(\mathbb{R}^{d+1} \setminus \bar{K}) < \frac{\varepsilon}{Q} \quad \text{for all } n \in \mathbb{N}.
\end{aligned}$$

Now, we are able to derive a convergence.

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

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

$$
\int \min\{Q, \rho(y - x^\top \beta)\}\nu_n(\mathrm{d}y, \mathrm{d}x) \xrightarrow[n\to+\infty]{} \int \min\{Q, \rho(y - x^\top \beta)\}\mu_0\,(\mathrm{d}y, \mathrm{d}x)\,.
$$

(b) The second term fulfills

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

because ρ is continuous, according to Assumption R5, and, hence, uniformly continuous on each compact set.

(c) The third term is smaller than ε since

$$
\begin{aligned}
&0 \le \int_{\rho(y - x^\top \beta_n) > Q} \big(\rho(y - x^\top \beta_n) - Q\big)\,\nu_n(\mathrm{d}y, \mathrm{d}x) \\
&\le \int_{\psi(|y - x^\top \beta_0| + T\|x\|) > Q} \big(\psi(|y - x^\top \beta_0| + T\|x\|) - Q\big)\,\nu_n(\mathrm{d}y, \mathrm{d}x) \\
&= \int \psi(|y - x^\top \beta_0| + T\|x\|)\nu_n(\mathrm{d}y, \mathrm{d}x) \\
&- \int \min\{Q, \psi(|y - x^\top \beta_0| + T\|x\|)\}\nu_n(\mathrm{d}y, \mathrm{d}x) \\
&\qquad \xrightarrow[n\to+\infty]{} \\
&\int_{\psi(|y - x^\top \beta_0| + T\|x\|) > Q} \big(\psi(|y - x^\top \beta_0| + T\|x\|) - Q\big)\,\mu_0\,(\mathrm{d}y, \mathrm{d}x) \\
&= \int_{\psi(|e| + T\|x\|) > Q} (\psi(|e| + T\|x\|) - Q)\,\nu\,(\mathrm{d}x, \mathrm{d}e) < \varepsilon.
\end{aligned}
$$

Thus, we proved

$$\lim_{n\to+\infty} \mathsf{f}\left(\beta_n \,|\, \nu_n\right) \quad = \quad \int \rho(y - x^\top \beta)\mu_0\left(\mathrm{d}y, \mathrm{d}x\right) = \mathsf{f}\left(\beta \,|\, \mu_0\right)\,.$$

Thus, Assumption A9 and the first part of Assumption A10 are verified. Actually, we proved more. We have proved continuity of the loss function at $\Theta \times \{\mu_0\}$.

4. The second part of Assumption A10 remained to be shown.
   Let $\nu_n \in \mathcal{P}_{\mathrm{emp}}$, $\nu_n \xrightarrow[n\to+\infty]{\mathrm{w}} \mu_0$.
   Then according to Assumption R7, there is a number $\Delta > 0$ such that

$$\inf_{|t|>\Delta} \rho(t) \cdot \mathsf{M} > \mathsf{f}\left(\beta_0 \,|\, \mu_0\right) + \bar{\varepsilon}\,.$$

   Hence according to Lemma 4, we are able to find Γ such that

$$\inf_{|t|>\Delta} \rho(t) \cdot \inf_{\|\gamma\|=1} \nu\left(\left\{(x,e) \;\middle|\; \Gamma\left|x^\top\gamma\right| \geq \Delta + |e|\,\right\}\right) \quad > \quad \mathsf{f}\left(\beta_0 \,|\, \mu_0\right) + \bar{\varepsilon}\,.$$

   Then, we define the required compact as

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

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

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

   For any $\delta > 0$, properly chosen sequence of γ_n, $\|\gamma_n\| = 1$ and its cluster point $\hat{\gamma}$, we have

$$\begin{aligned}
&\liminf_{n\to+\infty} \inf_{\beta\notin K} \mathsf{f}\left(\beta \,|\, \nu_n\right) \\
&\geq \inf_{|t|>\Delta} \rho(t) \cdot \liminf_{n\to+\infty} \inf\left\{\nu_n\left(\left\{(y,x) \;\middle|\; \Gamma\left|x^\top\gamma\right| > \Delta + |y - x^\top\beta_0|\,\right\}\right) \;\middle|\; \|\gamma\| = 1\,\right\} \\
&\geq \inf_{|t|>\Delta} \rho(t) \cdot \liminf_{n\to+\infty} \nu_n\left(\left\{(y,x) \;\middle|\; \Gamma\left|x^\top\gamma_n\right| > \Delta + |y - x^\top\beta_0|\,\right\}\right) \\
&\geq \inf_{|t|>\Delta} \rho(t) \cdot \liminf_{n\to+\infty} \nu_n\left(\left\{(y,x) \;\middle|\; \Gamma\left|x^\top\hat{\gamma}\right| > \Delta + |y - x^\top\beta_0| + \Gamma\left|x^\top(\gamma_n - \hat{\gamma})\right|\,\right\}\right)
\end{aligned}$$

$$\geq \inf_{|t|>\Delta} \rho(t) \cdot \liminf_{n\to+\infty} \nu_n \left(\left\{(y,x) \;\middle|\; \Gamma \left|x^\top \hat\gamma\right| > (1+\delta)\Delta + |y - x^\top \beta_0| \right\}\right)$$
$$\geq \inf_{|t|>\Delta} \rho(t) \cdot \mu_0 \left(\left\{(y,x) \;\middle|\; \Gamma \left|x^\top \hat\gamma\right| > (1+\delta)\Delta + |y - x^\top \beta_0| \right\}\right)$$
$$= \inf_{|t|>\Delta} \rho(t) \cdot \nu \left(\left\{(x,e) \;\middle|\; \Gamma \left|x^\top \hat\gamma\right| > (1+\delta)\Delta + |e| \right\}\right) .$$

Letting δ vanish we have

$$\begin{aligned}\liminf_{n\to+\infty} \inf_{\beta\notin K} \mathsf{f}\left(\beta \,|\, \nu_n\right) &\geq \inf_{|t|>\Delta} \rho(t) \cdot \nu \left(\left\{(x,e) \;\middle|\; \Gamma \left|x^\top \hat\gamma\right| \geq \Delta + |e| \right\}\right) \\ &> \mathsf{f}\left(\beta_0 \,|\, \mu_0\right) + \bar\varepsilon .\end{aligned}$$

Thus, the rest of Assumption A10 is verified.

Assumptions A1 – A10 are verified.

Assumptions R3 and R6 give us $\Omega_0 = \Omega_1 \cap \Omega_2$ with $\mathsf{prob}_*\left(\Omega_0\right) = 1$ such that for every $\omega \in \Omega_0$

$$\mathcal{E}_t\left(\bullet \,|\, (X_1(\omega), \mathrm{e}_1(\omega)), (X_2(\omega), \mathrm{e}_2(\omega)), \ldots, (X_t(\omega), \mathrm{e}_t(\omega))\right) \xrightarrow[n\to+\infty]{\mathcal{F}-\mathrm{w}} \nu .$$

All assumptions of Theorem 2 are valid, therefore, its assertion is also valid and that coincides with assertion of this theorem.

Additional results follow immediately from Theorems 3 and 4. □

Let us note that this setup covers both linear regression with random covariate X and, also, with covariate X lead by a deterministic design.

Considered ε-M-estimator is a particular case of Asymptotically Optimal Estimators (AOE). Definition of AOE together with a discussion on their consistency in probability is published in [22].

APPENDIX

This auxiliary section contains all necessary theory on nonmeasurable mappings. Definitions and relations are taken from [20], Chapter 1.9, pp. 52–56. All proofs can be found in [20], also.

Definition 4. For a set $B \subset \Omega$ its outer probability is defined as

$$\mathsf{prob}^*\left(B\right) = \inf\left\{\mathsf{prob}\left(A\right) \;\middle|\; B \subset A\,, A \in \mathcal{A}\right\} \tag{A.1}$$

and the inner probability is

$$\mathsf{prob}_*\left(B\right) = 1 - \mathsf{prob}^*\left(\Omega \setminus B\right) . \tag{A.2}$$

Definition 5. For a mapping $T : \Omega \to \overline{\mathbb{R}}$ the outer integral is defined as

$$\mathsf{E}^*\left[T\right] = \inf\left\{\mathsf{E}\left[U\right] \;\middle|\; U \geq T\,, U : \Omega \to \overline{\mathbb{R}} \quad \text{measurable and } \mathsf{E}\left[U\right] \text{ exists}\right\} \tag{A.3}$$

and the inner integral is

$$\mathsf{E}_*\left[T\right] = -\mathsf{E}^*\left[-T\right] . \tag{A.4}$$

Lemma 5. For any mapping $T : \Omega \to \overline{\mathbb{R}}$ there exists a measurable function $T^* : \Omega \to \overline{\mathbb{R}}$ with

1. $T^* \geq T$;
2. $T^* \leq U$ a s for every measurable $U : \Omega \to \overline{\mathbb{R}}$ with $U \geq T$.

The function T^* is called a minimal measurable majorant of T and is not uniquely defined.
The function $T_* = -(-T)^*$ is called a maximal measurable minorant of T.

Lemma 6. For any $T : \Omega \to \overline{\mathbb{R}}$

1. $\mathsf{E}^*[T] = \mathsf{E}[T^*]$ if one of these integral exists;
2. $\mathsf{E}_*[T] = \mathsf{E}[T_*]$ if one of these integral exists.

Definition 6. Let $\mathcal{X}$ be a metric space with metric d and $X_n, X : \Omega \to \mathcal{X}$, $n \in \mathbb{N}$ be arbitrary maps.

- X_n, $n \in \mathbb{N}$ converges in outer probability to X if for every $\varepsilon > 0$ $\mathsf{prob}^*(\mathsf{d}(X_n, X) > \varepsilon) \to 0$. We use notation $X_n \xrightarrow[n\to+\infty]{\mathsf{prob}^*} X$.
- X_n, $n \in \mathbb{N}$ converges almost uniformly to X if for every $\varepsilon > 0$ there exists a measurable set $A_\varepsilon \subset \Omega$ with $\mathsf{prob}(A_\varepsilon) \geq 1 - \varepsilon$ and $\mathsf{d}(X_n, X) \longrightarrow 0$ uniformly on A_ε, i. e. $\sup_{\omega \in A_\varepsilon} \mathsf{d}(X_n(\omega), X(\omega)) \longrightarrow 0$. We use notation $X_n \xrightarrow[n\to+\infty]{\mathrm{a\,u}} X$.
- X_n, $n \in \mathbb{N}$ converges outer almost surely to X if $\mathsf{d}(X_n, X)^* \longrightarrow 0$ almost surely for some versions of $\mathsf{d}(X_n, X)^*$. We use notation $X_n \xrightarrow[n\to+\infty]{\mathrm{a\,s}^*} X$.
- X_n, $n \in \mathbb{N}$ converges almost surely to X if $\mathsf{prob}_*(\lim_{n\to+\infty} \mathsf{d}(X_n, X) = 0) = 1$. We use notation $X_n \xrightarrow[n\to+\infty]{\mathrm{a\,s}} X$.

Definition 7. Let $\mathcal{X}$ be a metric space. The sequence $X_n : \Omega \to \mathcal{X}$, $n \in \mathbb{N}$ of arbitrary maps is called asymptotically measurable if

$$\mathsf{E}^*[f(X_n)] - \mathsf{E}_*[f(X_n)] \xrightarrow[n\to+\infty]{} 0 \tag{A.5}$$

for every bounded continuous function $f : \mathcal{X} \to \mathbb{R}$.

Definition 8. $X_n : \Omega \to \mathcal{X}$, $n \in \mathbb{N}$ of arbitrary maps is called strongly asymptotically measurable if

$$f(X_n)^* - f(X_n)_* \xrightarrow[n\to+\infty]{\mathrm{a\,s}} 0 \tag{A.6}$$

for every bounded continuous function $f : \mathcal{X} \to \mathbb{R}$.

Lemma 7. Let $\mathcal{X}$ be a metric space, $X : \Omega \to \mathcal{X}$ be Borel measurable and $X_n : \Omega \to \mathcal{X}$, $n \in \mathbb{N}$ be a sequence of arbitrary maps. Then,

- $X_n \xrightarrow[n\to+\infty]{\mathrm{a\,s}^*} X$ implies $X_n \xrightarrow[n\to+\infty]{\mathsf{prob}^*} X$;
- $X_n \xrightarrow[n\to+\infty]{\mathrm{a\,s}^*} X$ if and only if $X_n \xrightarrow[n\to+\infty]{\mathrm{a\,u}} X$.

Lemma 8. Let $\mathcal{X}$ be a metric space, $X : \Omega \to \mathcal{X}$ be Borel measurable and separable, and $X_n : \Omega \to \mathcal{X}$, $n \in \mathbb{N}$ be a sequence of arbitrary maps. Then, $X_n \xrightarrow[n\to+\infty]{\mathrm{a\,s}^*} X$ if and only if $X_n \xrightarrow[n\to+\infty]{\mathrm{a\,s}} X$ and X_n, $n \in \mathbb{N}$ is strongly asymptotically measurable.

Lemma 9. Let $\mathcal{X}$ be a metric space, $X : \Omega \to \mathcal{X}$ be Borel measurable and separable, and $X_n : \Omega \to \mathcal{X}$, $n \in \mathbb{N}$ be a sequence of arbitrary maps, which is asymptotically measurable. Then, $X_n \xrightarrow[n\to+\infty]{\mathrm{a\,s}} X$ implies $X_n \xrightarrow[n\to+\infty]{\mathsf{prob}^*} X$.

ACKNOWLEDGEMENT

The research has been partially supported by the Czech Science Foundation under Grants 201/05/2340, 201/08/0486 and by the Ministry of Education, Youth and Sports of the Czech Republic under project MSM 0021620839.

(Received October 29, 2007.)

REFERENCES

[1] X. R. Chen and Y. H. Wu: Strong consistency of M-estimates in linear model. J. Multivariate Anal. *27* (1988), 1, 116–130.
[2] Y. Dodge and J. Jurečková: Adaptive Regression. Springer-Verlag, New York 2000.
[3] J. Dupačová: Stability and sensitivity analysis for stochastic programming. Ann. Oper. Res. *27* (1990), 115–142.
[4] J. Dupačová and R. J.-B. Wets: Asymptotic behavior of statistical estimators and of optimal solutions of stochastic problems. Ann. Statist. *16* (1988), 4, 1517–1549.
[5] J. Hoffmann-Jørgensen: Probability with a View Towards to Statistics I, II. Chapman and Hall, New York 1994.
[6] P. J. Huber: Robust Statistics. Wiley, New York 1981.
[7] J. Jurečková: Asymptotic representation of M-estimators of location. Math. Oper. Statist. Sec. Statist. *11* (1980), 1, 61–73.
[8] J. Jurečková: Representation of M-estimators with the second-order asymptotic distribution. Statist. Decision *3* (1985), 263–276.
[9] J. Jurečková and P. K. Sen: Robust Statistical Procedures. Wiley, New York 1996.
[10] A. J. King and T. Rockafellar: Asymptotic theory for solutions in statistical estimation and stochastic programming. Math. Oper. Res. *18* (1993), 148–162.
[11] J. L. Kelley: General Topology. D. van Nostrand, New York 1955.
[12] K. Knight: Limiting distributions for L_1-regression estimators under general conditions. Ann. Statist. *26* (1998), 2, 755–770.

[13] P. Lachout: Stochastic optimization sensitivity without measurability. In: Proc. 15th Internat. Conference on Mathematical Methods in Economics and Industry (K. Cechlárová, M. Halická, V. Borbeľová, V. Lacko, eds.), Herľany, Slovakia, June 2007, pp. 131–136.

[14] P. Lachout and E. Liebscher, and S. Vogel: Strong convergence of estimators as ε_n-minimizers of optimization problems. Ann. Inst. Statist. Math. *57* (2005), 2, 291–313.

[15] P. Lachout and S. Vogel: On continuous convergence and epi-convergence of random functions. Part I: Theory and relations. Kybernetika *39* (2003), 1, 75–98.

[16] A. M. Leroy and P. J. Rousseeuw: Robust Regression and Outlier Detection. Wiley, New York 1987.

[17] S. M. Robinson: Analysis of sample-path optimization. Math. Oper. Res. *21* (1996), 3, 513–528.

[18] T. Rockafellar and R. J.-B. Wets: Variational Analysis. Springer-Verlag, Berlin 1998.

[19] R. Schultz: Some aspects of stability in stochastic programming. Ann. Oper. Research *100* (1996), 55–84.

[20] A. W. van der Vaart and J. A. Wellner: Weak Convergence and Empirical Processes. Springer, New York 1996.

[21] L. Wang and J. Wang: Limit distribution of statistical estimators for stochastic programs with dependent samples. Z. Angew. Math. Mech. *79* (1999), 4, 257–266.

[22] I. Vajda and M. Janžura: On asymptotically optimal estimates for general observations. Stochastic Proc. Appl. *72* (1997), 1, 27–45.

[23] A. Ruszczyński and A. Shapiro (eds.): Stochastic Programming. Elsevier, Amsterdam 2003.

Petr Lachout, Department of Probability and Statistics, Faculty of Mathematics and Physics – Charles University of Prague, Sokolovská 83, 186 75 Praha 8, and Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic.

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