Stability of Stochastic Programming Problems

W. Römisch

Humboldt-Universität Berlin
Institut für Mathematik

http://www-iam.mathematik.hu-berlin.de/~romisch

Tutorial

9th International Conference on Stochastic Programming

Berlin (Germany), August 25-31, 2001

Introduction

Consider the stochastic programming model

$$\min\{\int_{\Xi} f_0(\xi,x)\mu(d\xi) : x \in M(\mu)\} \qquad (1)$$

$$M(\mu) := \{x \in X : \int_{\Xi} f_j(\xi,x)\mu(d\xi) \le 0, j = 1,...,d\}$$

where f_j from $\Xi \times I\!R^m$ to the extended reals $\overline{I\!R}$ are normal integrands, X is a nonempty closed subset of $I\!R^m$, Ξ is a closed subset of $I\!R^s$ and μ is a Borel probability measure on Ξ.

(Recall that f_j is a normal integrand if it is Borel measurable and $f_j(\xi,.)$ is lower semicontinuous for each $\xi \in \Xi$.)

We denote by $\mathcal{P}(\Xi)$ the set of all Borel probability measures on Ξ and by $v(\mu)$ and $S_\varepsilon(\mu)$ the optimal value and the (ε-approximate) solution set ($\varepsilon \ge 0$) of (1), i.e.,

$$\begin{aligned}
v(\mu) &= \inf_{x\in M(\mu)} \int_{\Xi} f_0(\xi,x)\mu(d\xi),\\
S_\varepsilon(\mu) &= \{x \in M(\mu) : \int_{\Xi} f_0(\xi,x)\mu(d\xi) \le v(\mu)+\varepsilon\},\\
S(\mu) &= S_0(\mu) = \arg\min_{x\in M(\mu)} \int_{\Xi} f_0(\xi,x)\mu(d\xi).
\end{aligned}$$

Since the underlying probability distribution μ is often incompletely known in applied models, the **stability behaviour** of the stochastic program when changing (perturbing, estimating, approximating) $\mu \in \mathcal{P}(\Xi)$ is important.

Here, stability refers to **(quantitative) continuity properties** of the optimal value function $v(.)$ and of the set-valued mapping $S_\varepsilon(.)$ at μ, where both are regarded as mappings given on certain subset of $\mathcal{P}(\Xi)$ equipped with some **convergence** of probability measures and some **probability metric**, respectively.
(The corresponding subset of probability measures is determined such that certain moment conditions are satisfied that are related to growth properties of the integrands f_j with respect to ξ.)

Examples:
two-stage stochastic programs,
chance constrained stochastic programs.

Literatur

Surveys: Dupačová 90, Schultz 00

70s: Kall, Kaňková (78), Marti, Wets

Kaňková 80,..., Dupačová 84,...,
Wets 83, 89, Birge/Wets 86,
Kall 87, Robinson/Wets 87, Römisch/Wakolbinger 87,
Dupačová/Wets 88, Vogel 88, 92,94, King 89,
King/Wets 90, King/Rockafellar 93,
Salinetti 81, 89, Shapiro 89, 91, 94, 95, 99,
Ermoliev/Norkin 91, Lucchetti/Wets 93,
Römisch/Schultz 91, 93, 96, Schultz 92, 95, 96,
Artstein 94, Artstein/Wets 94, 95, Wang 95,
Pflug 96, 99, Pflug/Ruszczyński/Schultz 98, 99,
Fiedler/Römisch 95, Dentcheva/Römisch 00,
Gröwe 97, Henrion 00, Henrion/Römisch 99, 00,
Rachev/Römisch 00,......

Weak convergence in $\mathcal{P}(\Xi)$

$$\begin{aligned}\mu_n \to_w \mu \quad \text{iff} \quad & \int_\Xi f(\xi)\mu_n(d\xi) \to \int_\Xi f(\xi)\mu(d\xi) \\ & (\forall f \in C_b(\Xi)), \\ \text{iff} \quad & \mu_n(\{\xi \le z\}) \to \mu(\{\xi \le z\}) \\ & \text{if } \mu(\{\xi \le \cdot\}) \text{ is continuous at } z.\end{aligned}$$

Probability metrics on $\mathcal{P}(\Xi)$

Monographs: Rachev 91, Rachev/Rüschendorf 98

Metrics with ζ-structure:

$$d_{\mathcal{F}}(\mu,\nu) = \sup\{|\int_\Xi f(\xi)(\mu-\nu)(d\xi)| \, : \, f \in \mathcal{F}\}$$

where $\mathcal{F}$ is an appropriate set of measurable functions from Ξ to $\overline{\mathbb{R}}$ and μ, ν are probability measures in some set $\mathcal{P}_{\mathcal{F}}$ on which $d_{\mathcal{F}}$ is finite.

Examples:

(a) $\mathcal{F}$ is a class of locally Lipschitzian functions on Ξ,

(b) $\mathcal{F} = \{\chi_B : B \in \mathcal{B}\}$, $\mathcal{B}$ is a class of Borel subsets of Ξ.

It is possible to associate certain canonical sets $\mathcal{F}$ and, hence, canonical metrics $d_{\mathcal{F}}$ to specific classes of stochastic programs.

Example 1:

(two-stage model with simple recourse)

$m = s = 1,\ d = 0,\ f_0(\xi, x) := \max\{\xi - x, 0\}$,

$\Xi := I\!R,\ X := [-1, 1]\ (:= I\!R)$,

$\mu := \delta_0$ (unit mass at 0),

$\mu_n := (1 - \frac{1}{n})\delta_0 + \frac{1}{n}\delta_{n^2}\,,\ n \in I\!N$.

$v(\mu) = 0,\ S(\mu) = [0, 1]\ (= [0, \infty))$,

$v(\mu_n) = n - \frac{1}{n}\ (= -\infty),\ S(\mu_n) = \{1\}\ (= \emptyset)$

$(n \in I\!N)$.

Note: $\mu_n \to^w \mu$, but first order moments do not converge !

Example 2:

(linear chance constrained model)

$m = s = d = 1,\ X := (-\infty, 0],\ \Xi = I\!R$,

$f_0(\xi, x) := x,\ f_1(\xi, x) := \frac{3}{4} - \chi_{(-\infty, x]}(\xi)$,

$\mu := \frac{1}{2}\delta_0 + \frac{1}{2}\delta_{-1}$,

$\mu_n := (\frac{1}{2} + \frac{1}{n})\delta_{\frac{1}{n}} + (\frac{1}{2} - \frac{1}{n})\delta_{-1}\ (n \in I\!N)$.

$v(\mu) = 0,\ S(\mu) = \{0\}$,

$v(\mu_n) = \infty,\ S(\mu_n) = \emptyset\ (n \in I\!N)$.

Note: $\mu_n \to^w \mu$, but distribution functions do not converge uniformly !

Quantitative Stability

Let $\mathcal{U}$ be some nonempty subset of $I\!R^m$, and

$$\begin{aligned} \mathcal{F}_{\mathcal{U}} &:= \{f_j(.,x) \,:\, x \in X \cap cl\,\mathcal{U}, j = 0,\dots,d\}, \\ \mathcal{P}_{\mathcal{F},\mathcal{U}} &:= \{\nu \in \mathcal{P}(\Xi) : \int_{\Xi} \inf_{\substack{x\in X \\ \|x\|\le r}} f_j(\xi,x)\nu(d\xi) > -\infty\,, \forall r > 0\,, \\ &\qquad \sup_{x\in X\cap cl\,\mathcal{U}} \Big|\int_{\Xi} f_j(\xi,x)\nu(d\xi)\Big| < \infty\,,\ j = 0,\dots,d\}, \end{aligned}$$

and the probability (pseudo-) metric on $\mathcal{P}_{\mathcal{F},\mathcal{U}}$:

$$d_{\mathcal{F},\mathcal{U}}(\mu,\nu) = \sup_{x\in X\cap cl\,\mathcal{U}} \max_{j=0,\dots,d} \Big|\int_{\Xi} f_j(\xi,x)(\mu-\nu)(d\xi)\Big|\,.$$

Lemma:
The functions $(x,\nu) \mapsto \int_{\Xi} f_j(\xi,x)\nu(d\xi)$ are lower semicontinuous on $X \times \mathcal{P}_{\mathcal{F},\mathcal{U}}$.

Localized concepts for optimal values and solution sets:

$$\begin{aligned} v_{\mathcal{U}}(\nu) &= \inf\{\int_{\Xi} f_0(\xi,x)\nu(d\xi) : x \in M(\nu) \cap cl\,\mathcal{U}\}, \\ S_{\mathcal{U}}(\nu) &= \{x \in M(\nu) \cap cl\,\mathcal{U} : \int_{\Xi} f_0(\xi,x)\nu(d\xi) = v_{\mathcal{U}}(\nu)\}. \end{aligned}$$

A nonempty set $\mathcal{S} \subseteq I\!R^m$ is called a complete local minimizing (CLM) set of (1) with respect to $\mathcal{U}$ if $\mathcal{U} \subseteq I\!R^m$ is open and $\mathcal{S} = S_{\mathcal{U}}(\mu) \subset \mathcal{U}$. Clearly, sets of global minimizers are CLM sets and it holds $S_{\mathcal{U}}(\mu) = S(\mu)$ if $S(\mu) \subset \mathcal{U}$.

Theorem 1: (Rachev/Römisch 00)
Assume that $S(\mu)$ is nonempty and $\mathcal{U} \subset \mathbb{R}^m$ is an open bounded neighbourhood of $S(\mu)$, and that $\mu \in \mathcal{P}_{\mathcal{F},\mathcal{U}}$.
If $d \geq 1$, let the function $x \mapsto \int_\Xi f_0(\xi, x)\mu(d\xi)$ be Lipschitz continuous on $X \cap cl\,\mathcal{U}$, and, let the function $(x, y) \mapsto d(x, M_y(\mu))$ be locally Lipschitz continuous at each $(\bar{x}, 0)$, $\bar{x} \in S(\mu)$.
Then there exist constants $L, \delta > 0$ such that

$$\begin{aligned} |v(\mu) - v_{\mathcal{U}}(\nu)| &\leq L d_{\mathcal{F},\mathcal{U}}(\mu, \nu) \\ \emptyset \neq S_{\mathcal{U}}(\nu) &\subseteq S(\mu) + \Psi(L d_{\mathcal{F},\mathcal{U}}(\mu, \nu))\mathbb{B} \end{aligned}$$

holds for all $\nu \in \mathcal{P}_{\mathcal{F},\mathcal{U}}$ and that
$S_{\mathcal{U}}(\nu)$ is a CLM set w.r.t. $\mathcal{U}$ whenever $\nu \in \mathcal{P}_{\mathcal{F},\mathcal{U}}$ and $d_{\mathcal{F},\mathcal{U}}(\mu, \nu) < \delta$.
Here $\Psi(\eta) := \eta + \psi^{-1}(\eta)$ and
$\psi(\tau) := \min\{\int_\Xi f_0(\xi, x)\mu(d\xi) - v(\mu) : d(x, S(\mu)) \geq \tau, x \in M(\mu) \cap cl\,\mathcal{U}\}$ $(\eta, \tau \in \mathbb{R}_+)$, and $M_y(\mu) := \{x \in X : \int_\Xi f_j(\xi, x)\mu(d\xi) \leq y_j, j = 1, ..., d\}$.

The function ψ is the *growth* or *conditioning* function of (1) on $\mathcal{U}$. ψ and Ψ are lower semicontinuous on $\mathbb{R}_+$; ψ is nondecreasing and Ψ is increasing, both vanish at 0 and $\psi^{-1}(t) := \sup\{\tau \in \mathbb{R}_+ : \psi(\tau) \leq t\}$.

(Proof by appealing to stability results of Klatte 87, 94 and Rockafellar/Wets 97.)

Theorem 1 shows that $d_{\mathcal{F},\mathcal{U}}$ plays the role of a minimal probability metric for (1) implying quantitative stability.

Furthermore, notice that Theorem 1 remains valid when bounding $d_{\mathcal{F},\mathcal{U}}$ from above by another distance and when reducing the set $\mathcal{P}_{\mathcal{F},\mathcal{U}}$ to a subset on which this distance is defined and finite.

Such a distance is called a canonical probability metric d_{ca} associated with (1), if it has the structure $d_{\mathcal{F}}$ generated by some class $\mathcal{F} = \mathcal{F}_{ca}$ of functions from Ξ to $\overline{\mathbb{R}}$ such that $\mathcal{F}_{ca}$ contains the functions $Cf_j(\cdot, x)$ for each $x \in X \cap cl\,\mathcal{U}$, $j = 0, \ldots, d$ and some normalizing constant $C > 0$, and that the functions in $\mathcal{F}_{ca}$ have the same analytical properties as $f_j(\cdot, x)$, $j = 0, \ldots, d$.

Typical analytical properties defining canonical classes $\mathcal{F}_{ca}$, which are relevant in stochastic programming, are piecewise Lipschitz continuity properties.

Example: (Fortet-Mourier metrics)
Let $p \geq 1$, $\xi_0 \in \Xi$ and consider the following class of continuous functions from Ξ to $I\!R$

$$\begin{aligned}\mathcal{F}_p := \{f \quad : \quad & |f(\xi) - f(\tilde{\xi})| \leq \\ & \max\{1, \|\xi - \xi_0\|^{p-1}, \|\tilde{\xi} - \xi_0\|^{p-1}\}\|\xi - \tilde{\xi}\|, \forall \xi, \tilde{\xi} \in \Xi\}\end{aligned}$$

and the corresponding probability metric generated by $\mathcal{F}_p$ and defined on $\mathcal{P}_p(\Xi)$:

$$\begin{aligned}\zeta_p(\mu, \nu) \quad &:= \quad d_{\mathcal{F}_p}(\mu, \nu) = \sup_{f \in \mathcal{F}_p} \left| \int_\Xi f(\xi)(\mu - \nu)(d\xi) \right| \\ \mathcal{P}_p(\Xi) \quad &:= \quad \{\nu \in \mathcal{P}(\Xi) : \int_\Xi \|\xi\|^p \nu(d\xi) < \infty\}\end{aligned}$$

Convex case and $d := 0$:

Assume that $f_0(\xi, \cdot)$ is convex on $\mathbb{R}^m$ $\forall \xi \in \Xi$.

Theorem 2:
Assume that $S(\mu)$ is nonempty and $\mathcal{U} \subset \mathbb{R}^m$ is an open bounded neighbourhood of $S(\mu)$, and that $\mu \in \mathcal{P}_{\mathcal{F},\mathcal{U}}$.
Then there exist constants $L, \bar{\varepsilon} > 0$ such that

$$\begin{aligned} |v(\mu) - v(\nu)| &\leq d_{\mathcal{F},\mathcal{U}}(\mu, \nu) \quad \text{and} \\ \emptyset \neq S(\nu) &\subset S(\mu) + \Psi(d_{\mathcal{F},\mathcal{U}}(\mu, \nu))\mathbb{B} \end{aligned}$$

whenever $\nu \in \mathcal{P}_{\mathcal{F},\mathcal{U}}$ with $d_{\mathcal{F},\mathcal{U}}(\mu, \nu) < \bar{\varepsilon}$, and that it holds for any $\varepsilon \in (0, \bar{\varepsilon})$

$$D_H(S_\varepsilon(\mu), S_\varepsilon(\nu)) \leq \frac{L}{\varepsilon} d_{\mathcal{F},\mathcal{U}}(\mu, \nu)$$

whenever $\nu \in \mathcal{P}_{\mathcal{F},\mathcal{U}}$, $d_{\mathcal{F},\mathcal{U}}(\mu, \nu) < \varepsilon$.
Here $\Psi(\eta) := \eta + \psi^{-1}(2\eta)$, $\eta \geq 0$, ψ is the conditioning function of Theorem 1 and D_H is the Hausdorff distance of nonempty closed subsets of $\mathbb{R}^m$.

Proof using a perturbation result by Rockafellar/Wets 97.

Linear two-stage stochastic programs

We consider the linear two-stage stochastic program with fixed recourse

$$\min\{cx + \int_\Xi q(\xi)y(\xi)\mu(d\xi) \ : \ Wy(\xi) = h(\xi) - T(\xi)x,$$

$$y(\xi) \geq 0, x \in X\}$$

where $c \in \mathbb{R}^m$, $X \subseteq \mathbb{R}^m$ is a polyhedron, Ξ is a polyhedron in $\mathbb{R}^s$, W is an $(r, \overline{m})$-matrix, $\mu \in \mathcal{P}(\Xi)$, and $q(\xi) \in \mathbb{R}^{\overline{m}}$, $h(\xi) \in \mathbb{R}^r$ and the (r, m)-matrix $T(\xi)$ depend affine linearly on $\xi \in \Xi$.

Denoting by $\Phi(q(\xi), h(\xi) - T(\xi)x)$ the value of the optimal second stage decision, the above problem may be rewritten equivalently as a minimization problem with respect to the first stage decision x.
Defining the integrand $f_0 : \Xi \times \mathbb{R}^m \to \overline{\mathbb{R}}$ by

$$f_0(\xi, x) = \begin{cases} cx + \Phi(q(\xi), h(\xi) - T(\xi)x), \\ \qquad\qquad h(\xi) - T(\xi)x \in \operatorname{pos} W, q(\xi) \in D, \\ +\infty, \ \text{otherwise}, \end{cases}$$

$$\begin{aligned} \operatorname{pos} W &:= \{Wy : y \in \mathbb{R}^{\overline{m}}_+\}, \\ D &:= \{u \in \mathbb{R}^{\overline{m}} : \{z \in \mathbb{R}^r : W'z \leq u\} \neq \emptyset\} \\ \Phi(u, t) &:= \inf\{uy : Wy = t, y \geq 0\} \ \ ((u, t) \in \mathbb{R}^{\overline{m}} \times \mathbb{R}^r), \end{aligned}$$

the equivalent minimization problem takes the form

$$\min\{\int_\Xi f_0(\xi, x)\mu(d\xi) : \ x \in X\}. \tag{2}$$

Assumptions:

(A1) There holds $h(\xi) - T(\xi)x \in \mathrm{pos}\, W$ and $q(\xi) \in D$ for each pair $(\xi, x) \in \Xi \times X$ (*relatively complete recourse* and *dual feasibility*).

(A2) $\mu \in \mathcal{P}(\Xi)$ has a finite second order moment.

Theorem 3:

Let (A1) and (A2) be satisfied and let $S(\mu)$ be nonempty and $\mathcal{U}$ be an open, bounded neighbourhood of $S(\mu)$.

Then there exist constants $L, \bar{\varepsilon} > 0$ such that

$$\begin{aligned} |v(\mu) - v(\nu)| &\leq L\zeta_2(\mu,\nu) \\ \emptyset \neq S(\nu) &\subseteq S(\mu) + \Psi(L\zeta_2(\mu,\nu))\mathbb{B} \end{aligned}$$

whenever $\nu \in \mathcal{P}_2(\Xi)$ and $\zeta_2(\mu,\nu) < \bar{\varepsilon}$, where Ψ is defined as in Theorem 2.

Furthermore, it holds for any $\varepsilon \in (0, \bar{\varepsilon})$

$$D_H(S_\varepsilon(\mu), S_\varepsilon(\nu)) \leq \frac{L}{\varepsilon}\zeta_2(\mu,\nu)$$

whenever $\nu \in \mathcal{P}_2(\Xi)$, $\zeta_2(\mu,\nu) < \varepsilon$.

Chance constrained stochastic programs

$$\min\{cx : x \in X, \mu(\{\xi \in \Xi : T(\xi)x \geq h(\xi)\}) \geq p\}$$

where $c \in \mathbb{R}^m$, X is a polyhedron in $\mathbb{R}^m$, Ξ a polyhedron in $\mathbb{R}^s$, $p \in (0,1)$, $\mu \in \mathcal{P}(\Xi)$, and $h(\xi) \in \mathbb{R}^r$ and the (r,m)-matrix $T(\xi)$ depend affine linearly on $\xi \in \Xi$.
We set $d = 1$, $f_0(\xi, x) = cx$, $f_1(x,\xi) = p - \chi_{H(x)}(\xi)$, where $H(x) = \{\xi \in \Xi : T(\xi)x \geq h(\xi)\}$, and obtain

$$\begin{aligned} \mathcal{P}_{\mathcal{F},\mathcal{U}}(\Xi) &= \mathcal{P}(\Xi)\,, \\ d_{\mathcal{F},\mathcal{U}}(\mu,\nu) &= \sup_{x \in X \cap cl\, \mathcal{U}} |\mu(H(x)) - \nu(H(x))| \; (\mu, \nu \in \mathcal{P}(\Xi)) \end{aligned}$$

The sets $H(x)$ are polyhedra with a uniformly bounded number of faces. Canonical metric:

$$\begin{aligned} d_{ph,k}(\mu,\nu) := \sup\{|\mu(P) - \nu(P)| : P \text{ polyhedron} \\ \text{with at most } k \text{ faces}\} \end{aligned}$$

Theorem 4:
Let $S(\mu)$ be nonempty and $\mathcal{U} \subseteq \mathbb{R}^m$ be an open bounded neighbourhood of $S(\mu)$, and $\mu \in \mathcal{P}(\Xi)$.
Let the function $(x,y) \mapsto d(x, M_y(\mu))$ be locally Lipschitz continuous at each $(\bar{x}, 0)$, $\bar{x} \in S(\mu)$.
Then there exist constants $L > 0$, $\delta > 0$ and $k \in \mathbb{N}$ such that

$$\begin{aligned} |v(\mu) - v_{\mathcal{U}}(\nu)| &\leq L\, d_{ph,k}(\mu,\nu) \\ \emptyset \neq S_{\mathcal{U}}(\nu) &\subseteq S(\mu) + \Psi(L d_{ph,k}(\mu,\nu))\mathbb{B} \end{aligned}$$

and $S_{\mathcal{U}}(\nu)$ is a CLM set w.r.t. $\mathcal{U}$ whenever $\nu \in \mathcal{P}(\Xi)$ and $d_{ph,k}(\mu,\nu) < \delta$.
Here, Ψ is defined as in Theorem 1.

Empirical Approximations

Let $\xi_1, \xi_2, ..., \xi_n, ...$ be i.i.d. random vectors in $\mathbb{R}^s$ (on $(\Omega, \mathcal{A}, \mathbb{P})$) with common probability distribution $\mu \in \mathcal{P}_{\mathcal{F},\mathcal{U}}$. We consider the empirical measures $\mu_n(.) = \frac{1}{n}\sum_{i=1}^{n} \delta_{\xi_i(.)}$ $(n \in \mathbb{N})$ and the empirical approximations of (1)

$$\min\{\frac{1}{n}\sum_{i=1}^{n} f_0(\xi_i(\cdot), x) : x \in X\,,$$
$$\frac{1}{n}\sum_{i=1}^{n} f_j(\xi_i(\cdot), x) \leq 0, j = 1, \dots, d\}.$$

Then $v_{\mathcal{U}}(\mu_n(\cdot))$ and $S_{\mathcal{U}}(\mu_n(\cdot))$ are measurable.

A class $\mathcal{F}$ is called permissible if the mappings $d_{\mathcal{F}}(\mu, \mu_n(.))$ from Ω to $\mathbb{R}$ are measurable.
$\mathcal{F}$ is called a μ-Glivenko-Cantelli class if
$\mathbb{P} - \lim_{n\to\infty} d_{\mathcal{F}}(\mu, \mu_n(\cdot)) = 0$.

Ky Fan metric in $\mathcal{X}(\mathbb{R})$:
$\kappa(\mathcal{X}, \mathcal{Y}) := \inf\{\eta \geq 0 : \mathbb{P}(|\mathcal{X} - \mathcal{Y}| > \eta) \leq \eta\}$.

Theorem 5:
Let the assumptions of Theorem 1 be satisfied and $\mathcal{F}_{\mathcal{U}}$ be permissible for μ. Then it holds each $n \in I\!N$

$$\begin{aligned} \kappa(v(\mu), v_{\mathcal{U}}(\mu_n(\cdot))) &\leq \max\{1, L\}\kappa(d_{\mathcal{F},\mathcal{U}}(\mu_n(\cdot), \mu), 0) \\ \kappa(\sup_{x \in S_{\mathcal{U}}(\mu_n(\cdot))} d(x, S(\mu)), 0) &\leq \Psi(\kappa(d_{\mathcal{F},\mathcal{U}}(\mu_n(\cdot), \mu), 0)), \end{aligned}$$

where $L > 0$ and Ψ are as in Theorem 1.
Moreover, for $I\!P$-almost all $\omega \in \Omega$ the set $S_{\mathcal{U}}(\mu_n(\omega))$ is a CLM set of (1) w.r.t. $\mathcal{U}$ for sufficiently large $n \in I\!N$.

Whether (a rate of) convergence of $(d_{\mathcal{F}}(\mu_n(\cdot), \mu))$ is available, depends on the size of the class $\mathcal{F}$ measured in terms of covering or bracketing numbers.

Let $\mathcal{F}$ be a subset of the normed space $L_p(\Xi, \mu)$ (for some $p \geq 1$) equipped with the usual norm $\|\cdot\|_p$. The covering number $N(\varepsilon, \mathcal{F}, L_p(\Xi, \mu))$ is the minimal number of open balls $\{g \in L_p(\Xi, \mu) : \|g - f\|_p < \varepsilon\}$ needed to cover $\mathcal{F}$.
Given two functions f_1 and f_2 from $L_p(\Xi, \mu)$, the set $[f_1, f_2] := \{f \in L_p(\Xi, \mu) : f_1(\xi) \leq f(\xi) \leq f_2(\xi)$ for μ-almost all $\xi \in \Xi\}$ is called an ε-bracket if $\|f_1 - f_2\|_p < \varepsilon$.
Then the bracketing number $N_{[\,]}(\varepsilon, \mathcal{F}, L_p(\Xi, \mu))$ is the minimal number of ε-brackets needed to cover $\mathcal{F}$.
A class $\mathcal{F} \subset L_1(\Xi, \mu)$ is a μ-Glivenko-Cantelli class if $N_{[\,]}(\varepsilon, \mathcal{F}, L_1(\Xi, \mu)) < \infty$ for each $\varepsilon > 0$.

Theorem 6:
Let the assumptions of Theorem 1 be satisfied and $\mathcal{F}_{\mathcal{U}}$ be uniformly bounded and permissible for μ. Assume that either of the following conditions holds for some constants $r \geq 1$, $R \geq 1$ and $\varepsilon \in (0,1)$:
(i) $N(\varepsilon, \mathcal{F}_{\mathcal{U}}, L_2(\Xi, \nu)) \leq (\frac{R}{\varepsilon})^r$ for any discrete $\nu \in \mathcal{P}(\Xi)$ with finite support ,
(ii) $N_{[\,]}(\varepsilon, \mathcal{F}_{\mathcal{U}}, L_2(\Xi, \mu)) \leq (\frac{R}{\varepsilon})^r$.
Then the following rates of convergence

$$
\begin{aligned}
\kappa(v(\mu), v_{\mathcal{U}}(\mu_n(\cdot))) &= O((\log n)^{\frac{1}{2}} n^{-\frac{1}{2}}) \\
\kappa(\sup_{x \in S_{\mathcal{U}}(\mu_n(\cdot))} d(x, S(\mu)), 0) &= O(\Psi((\log n)^{\frac{1}{2}} n^{-\frac{1}{2}}))
\end{aligned}
$$

are valid, where Ψ is as in Theorem 1.

Examples:

The class $\mathcal{F}_{ph,k} := \{\chi_P : P$ polyhedron with at most k faces$\}$ satisfies (i) of Theorem 6.

The class $\mathcal{F}_{lts} := \{f_0(\cdot, x) : f_0$ is defined as for two-stage models satisfying (A1), $x \in X \cap cl\,\mathcal{U}\}$ satisfies the property

$$
N_{[\,]}(\varepsilon K_p, \mathcal{F}_{lts}, L_p(\Xi, \mu)) \leq C\varepsilon^{-m},
$$

for each $0 < \varepsilon < 1$, $p \geq 1$, some $C > 0$ depending only on m and the diameter of $X \cap cl\,\mathcal{U}$ and some $K_p > 0$ depending on the $2p$-th order moment of μ.
Hence, (ii) is satisfied if $\int_\Xi \|\xi\|^4 \mu(d\xi) < \infty$ and Theorem 6 applies if Ξ is bounded.