P. Kříž; Josef Štěpán

A note on almost sure convergence and convergence in measure

Commentationes Mathematicae Universitatis Carolinae, Vol. 55 (2014), No. 1, 29--40

Persistent URL: http://dml.cz/dmlcz/143566

A note on almost sure convergence and convergence in measure

P. Kříž, J. Štěpán

Abstract. The present article studies the conditions under which the almost everywhere convergence and the convergence in measure coincide. An application in the statistical estimation theory is outlined as well.

Keywords: convergence in measure; almost sure convergence; pointwise compactness; Lusin property; strongly consistent estimators

Classification: Primary 28A20; Secondary 62F12

1. Introduction

Recall the basic concepts from estimation theory.

1. Consider $\{(\Omega, \mathcal{F}, P_\theta), \theta \in \Theta\}$ a parametric family of probability spaces, where Θ is an arbitrary set, $\phi : \Theta \to \mathbb{R}$ a parametric map.
2. To estimate $\phi(\theta)$ we are allowed to perform a sequence of observations $X_n : (\Omega, \mathcal{F}) \to (Z, \mathcal{Z})$ (a Hausdorff topological space with its Borel σ-algebra)
3. and construct a sequence of measurable maps $t_n : Z^{\mathbb{N}} \to \mathbb{R}$, called a weakly consistent and strongly consistent estimator of $\phi(\theta)$ if
   $t_n(X_1, X_2, \dots) \to \phi(\theta)$ in P_θ-probability $\forall \theta \in \Theta$
   and
   $t_n(X_1, X_2, \dots) \to \phi(\theta)$ P_θ-almost surely $\forall \theta \in \Theta$
   respectively.
4. In some cases we prefer to choose estimators $\{t_n\}$ in a suitable set K of measurable maps $t_n : Z^{\mathbb{N}} \to \mathbb{R}$.
   For all this to make sense in mathematical statistics, the choice should be $t_n(x_1, \dots, x_n, x_{n+1}, \dots) = t_n(x_1, \dots, x_n)$.
5. The Bayesian approach further considers the parametric space Θ endowed with a σ-algebra $\mathcal{T}$ and an apriori probability measure F.

Our aim is to study circumstances under which a weakly consistent estimator is a strong one automatically. We shall start with an individual probability ν and

The work of Pavel Kříž was supported by the grants GA UK 2673/2011 and SVV 265 315.

try to endow a sequence of measurable functions $\{f_n\}$ by an additional property that makes true rather exotic implication

$$f_n \text{ converges in } \nu\text{-measure} \quad \Longrightarrow \quad f_n \text{ converges } \nu\text{-a.s.}$$

Our results presented in Sections 2, 3 and 4 complement those by Ionescu Tulcea in [5] and [6]. Section 5 is a comeback to the estimation problem stated above.

2. Preliminaries

Consider a Hausdorff space S and denote by $C(S)$ and $\mathcal{B}(S)$ the space of continuous and Borel measurable real functions defined on S, respectively. Further denote by ν a Radon probability on S. For $D \subset S$ and $K \subset \mathcal{B}(S)$ denote the restriction $K|_D = \{f|_D : f \in K\}$.

Definition 2.1. A set $K \subset \mathcal{B}(S)$ is called *uniformly ν-Lusin* if

$$\forall \epsilon > 0 \quad \exists D_\epsilon \subset S \text{ compact}: \quad \nu(D_\epsilon) \geq 1 - \epsilon, \quad K|_{D_\epsilon} \subset C(D_\epsilon).$$

Recall that any countable K is uniformly ν-Lusin and that the uniformly ν-Lusin property is closed under countable unions.

Definition 2.2. A σ-compact set $D \subset S$, such that $\nu(D) = 1$ and for any $f, g \in K \subset \mathcal{B}(S)$ the equivalence

$$f = g \quad \nu - \text{a. s.} \quad \Longleftrightarrow \quad f|_D = g|_D$$

holds, will be referred to as an *SU-set for ν and K* (set of uniqueness).

Theorem 2.3. *Consider a uniformly ν-Lusin set $K \subset \mathcal{B}(S)$. Then there exists an SU-set for ν and K.*

Proof: For each $n \in \mathbb{N}$ consider the compact $D_{1/n}$ from Definition 2.1 and define a compact $M_n = \operatorname{supp}(\nu|_{D_{1/n}}) \subset D_{1/n}$, i.e. the support of ν restricted to $D_{1/n}$. Clearly $\nu(M_n) > 1 - \frac{1}{n}$. Now put $D = \bigcup_n M_n$ and observe that D is the SU-set. Obviously D is a σ-compact set with $\nu(D) = 1$ and

$$f = g \quad \nu - \text{a.s.} \quad \Longrightarrow \quad f|_{M_n} = g|_{M_n} \quad \nu|_{M_n} - \text{a.e. } \forall n \quad \Longrightarrow$$
$$\Longrightarrow \quad f|_{M_n} = g|_{M_n} \quad \forall n \quad \Longrightarrow \quad f|_D = g|_D.$$

The argument for the second implication reads:
Since $f|_{M_n}, g|_{M_n} \in C(M_n)$, we have

$$Q = \{x \in M_n : f|_{M_n} \neq g|_{M_n}\} \text{ is open in } M_n \text{ and } \nu|_{M_n}(Q) = 0.$$

Because M_n is the support of $\nu|_{D_{1/n}}$, we have $Q = \emptyset$, hence $f|_{M_n} = g|_{M_n}$.
The converse, i.e. $f|_D = g|_D \Longrightarrow f = g \quad \nu - \text{a.s.}$, holds trivially. □

Having a Hausdorff topological space we shall consider the spaces

$$C_b(S) \subset C(S) \subset \mathcal{B}(S) \subset S^{\mathbb{R}},$$

where $C_b(S)$ is the Banach space of real, bounded, continuous functions defined on S ($C_b(S) = C(S)$ if S is a compact space). The space $S^{\mathbb{R}}$ will be endowed with its product topology (the topology of the pointwise convergence) denoted as $T_p = T_{p,S^{\mathbb{R}}}$. The Banach space $C_b(S)$ will be considered in its weak topology $T_w = T_{w,C_b(S)}$. Finally, having a Radon probability ν on S we shall also deal with the topology $T_\nu = T_{\nu,\mathcal{B}(S)}$ of the convergence in ν-measure on the space $\mathcal{B}(S)$.

Agree to write $T_{p,A}, T_{w,A}$ and $T_{\nu,A}$ for the corresponding subspace topologies, where $A \subset S^{\mathbb{R}}, A \subset C_b(S)$ and $A \subset \mathcal{B}(S)$, respectively.

Let us describe convergence in the respective topologies in more detail. Having real functions f_n, f from $S^{\mathbb{R}}$ (resp. $C_b(S), \mathcal{B}(S)$) recall that $f_n \to f$ w.r.t. T_p iff $f_n(s) \to f(s)$ for all $s \in S$. Next, the convergence $f_n \to f$ w.r.t. T_w is defined as $x(f_n) \to x(f)$ for all continuous linear functionals x on $C_b(S)$. Recall that for S being a compact Hausdorff space, $f_n \to f$ w.r.t. T_w iff the sequence $\{f_n\} \subset C(S)$ is bounded in supremum norm and $f_n(s) \to f(s)$ for each $s \in S$. This characterisation of weak convergence follows from Riesz representation theorem (see e.g. [3, p. 265, Corollary IV.6.4]). Finally, recall that $f_n \to f$ w.r.t. T_ν iff

$$\forall \epsilon > 0 : \lim_{n\to\infty} \nu\{s : |f_n(s) - f(s)| > \epsilon\} = 0.$$

Note that for a compact Hausdorff space S, the following relations between the topologies hold:

$$T_{\nu,C(S)} \quad \subset \quad T_{p,C(S)} \quad \subset \quad T_{w,C(S)}. \tag{2.1}$$

The first inclusion follows from the fact that the convergence in measure ν implies the convergence ν-almost surely. The second inclusion follows from the fact that coordinate projections $\pi_s(f) := f(s)$ are continuous linear functionals.

If S is a compact Hausdorff space stress that by saying that a subset $K \subset C(S)$ is sequentially relatively weakly compact, we mean that any sequence in K has a subsequence converging weakly to a limit in $C(S)$. Moreover, if the weak limits are in K, the set K is called sequentially weakly compact. Recall that the sequential relative weak compactness and the relative weak compactness (i.e. the weak closure is weakly compact) coincide by Eberlein-Smuljan theorem (see [3, p. 430, Theorem V.6.1]).

Finally denote

$$SC(K) = \{f \in \mathcal{B}(S) : \quad \exists f_n \in K, \quad f_n \to f \text{ on } S\}$$

the set of sequential cluster points of K w.r.t. the T_p-topology in $\mathcal{B}(S)$ and agree that a Radon probability measure ν on S is called self-supporting if $\operatorname{supp}(\nu) = S$.

3. T_p compactness and almost sure convergence

We shall suggest some conditions, under which the Cauchy property in ν measure implies the convergence ν-almost surely.

Theorem 3.1. *Let $K \subset \mathcal{B}(S)$ be a relatively sequentially $T_{p,S^{\mathbb{R}}}$-compact, hence a relatively sequentially $T_{p,\mathcal{B}(S)}$-compact set. Assume that $SC(K)$ is a set, which is uniformly ν-Lusin (for example $SC(K) \subset C(S)$). Denote by D an SU-set for ν and $SC(K)$. If $\{f_n\} \subset K$ is a Cauchy sequence in ν-measure, then there is $f \in \mathcal{B}(S)$ such that $f|_D \in SC(K|_D)$ and $f_n \to f$ ν-a.s.*

Proof: If $\{f_n\}$ is a Cauchy sequence in ν-measure, then $SC(f_1|_D, f_2|_D, \ldots)$ is a singleton set $\{f_0\}$. The reasoning is as follows:

$$\begin{aligned} &f_{n_k}|_D \to g \in \mathcal{B}(D) \text{ on } D, \quad f_{m_k}|_D \to h \in \mathcal{B}(D) \text{ on } D \quad \Rightarrow \\ &g = h \quad \nu-\text{a.s., because } f_n|_D \text{ is Cauchy in measure } \nu \quad \Rightarrow \\ &g = h \text{ on } D \quad \text{by Definition 2.2.} \end{aligned}$$

Further assume that $f_n|_D \nrightarrow f_0$ on D. Then there is a neighbourhood $O(f_0) \in T_{p,\mathcal{B}(D)}$ and a subsequence $\{f_{n_k}|_D\}$ such that $f_{n_k}|_D \notin O(f_0)$. Consider a subsequence $\{f_{n_{k_i}}|_D\}$ such that $f_{n_{k_i}}|_D \to g_0$ on D. We get $f_0 \neq g_0$ distinct functions in $SC(f_1|_D, f_2|_D, \ldots)$, which is a contradiction.

Finally, observe that $f_n|_D \to f_0$ on D and thus $f_n \to f := f_0 \cdot I_D$ ν-a.s. and $f|_D = f_0 \in SC(K|_D)$. □

Example 3.2. To illustrate the importance of the uniform ν-Lusin property, recall the well-known example of a sequence convergent in measure that does not converge a.s. Consider the probability space $([0,1], \mathcal{B}([0,1]), \lambda_{[0,1]})$. Let k_n and v_n satisfy $n = k_n + 2^{v_n}$, $0 \le k_n < 2^{v_n}$. Define

$$f_n(x) = \begin{cases} 1, & \text{if } x \in [k_n 2^{-v_n}, (k_n+1)2^{-v_n}] \\ 0, & \text{otherwise.} \end{cases}$$

Then $\{f_n\}$ is a sequence that is relatively sequentially T_p compact and convergent in measure $\lambda_{[0,1]}$. However, it does not converge $\lambda_{[0,1]}$ - almost surely and $SC(f_1, f_2 \ldots) = \{I_{\{x\}} : x \in [0,1]\} \cup \{0\}$, which is not a uniformly $\lambda_{[0,1]}$-Lusin set.

Theorem 3.3. *Let $K \subset \mathcal{B}(S)$ be a sequentially T_p-compact, uniformly ν-Lusin set and D an SU-set for ν and K. If $\{f_n\} \subset K$ is a Cauchy sequence in ν-measure, then there is $f \in \mathcal{B}(S)$ such that $f|_D \in K|_D$ and $f_n \to f$ ν-a.s. Moreover $T_{p,K|_D} = T_{\nu,K|_D}$ and both topologies are compact metrizable.*

Proof: The sequential T_p-compactness implies $SC(K) \subset K$ and the existence of ν-a.s. limit follows by Theorem 3.1.

Definition 2.2 implies that $T_{\nu,K|_D}$ is a Hausdorff topology and the equality of the topologies thus follows by Proposition 463C, p. 539 in [4]. Their metrizability follows by the metrizability of $T_{\nu,K|_D}$, and thus the compactness is implied by the sequential compactness of $T_{p,K|_D}$. □

Recall that the sequential compactness of K w.r.t. T_p in the theorem above can be replaced by the countable compactness w.r.t. T_p (by application of Corollary 463K, p. 545 in [4]).

Corollary 3.4. *Let S be a compact Hausdorff space, $K \subset C(S)$ a bounded (sequentially) relatively weakly compact set in $C(S)$ and ν a self-supporting Radon probability on S. Denote by $\overline{K}^p$, $\overline{K}^w$ and $\overline{K}^\nu$ the $T_{p,S^{\mathbb{R}}}$, $T_{w,C(S)}$ and $T_{\nu,C(S)}$-closure of K, respectively. Then*

$$\overline{K}^p = \overline{K}^w = \overline{K}^\nu =: \overline{K} \quad \text{and} \quad T_{p,\overline{K}} = T_{w,\overline{K}} = T_{\nu,\overline{K}}.$$

As a consequence,

$$\{f_n\} \subset K \quad \nu\text{-Cauchy} \quad \Longrightarrow \quad \exists f \in \overline{K} : f_n \to f \quad \nu\text{-a.s.}, \quad f_n \to f \quad \text{weakly}$$

and there are $g_n \in \mathrm{co}(f_1, f_2, \dots)$ such that $g_n \to f$ uniformly on S.

By $\mathrm{co}(K)$, denote the set of all convex combinations of elements of K.

Proof: First, we show that $\overline{K}^p = \overline{K}^w$. We shall start with the inclusion $\overline{K}^w \subset \overline{K}^p$:

$$f \in \overline{K}^w \quad \Longrightarrow \quad f \in C(S) \text{ and } f_\alpha \to f \text{ for some net } f_\alpha \in K \text{ in } T_{w,C(S)}\text{-topology}$$
$$\Longrightarrow \quad f_\alpha \to f \text{ w.r.t. } T_{p,C(S)} \quad \Longrightarrow \quad f_\alpha \to f \text{ w.r.t. } T_{p,S^{\mathbb{R}}} \quad \Longrightarrow \quad f \in \overline{K}^p.$$

Note that the second implication follows from (2.1). The argument for the second inclusion $\overline{K}^p \subset \overline{K}^w$ is:

$$f \in \overline{K}^p \quad \Longrightarrow \quad f_\alpha \to f \text{ for some net } f_\alpha \in K \text{ in } T_{p,S^{\mathbb{R}}}\text{-topology}$$
$$\Longrightarrow \quad f_\beta \to f \text{ in } T_{w,C(S)} \text{ topology for some subnet } f_\beta \quad \Longrightarrow \quad f \in \overline{K}^w.$$

Indeed, since K is relatively $T_{w,C(S)}$ compact set (= relatively sequentially $T_{w,C(S)}$ compact) there exists a subnet f_β convergent w.r.t. $T_{w,C(S)}$. Moreover, the weak limit of f_β is also a pointwise limit of f_β, which equals the pointwise limit of f_α.

The same argument applies to arbitrary $A \subset \overline{K}^p$, resulting in $\overline{A}^p = \overline{A}^w$. Hence, $T_{p,\overline{K}^p} = T_{w,\overline{K}^w}$.

Next, recall that $\overline{K}^p$ is a T_w-compact and thus a T_p-compact set. We apply Theorem 3.3 with $D = S$ and the subsequent comment to verify that $T_{p,\overline{K}^p} = T_{\nu,\overline{K}^p}$ and both topologies are compact metrizable.

Further note that $T_{\nu,C(S)}$ is a metrizable topology, because the ν-equivalence class of an $f \in C(S)$ is a singleton set. Due to the compactness of $T_{\nu,\overline{K}^p}$, the set $\overline{K}^p$ is a closed superset of K w.r.t. $T_{\nu,C(S)}$-topology. Hence, $\overline{K}^\nu \subset \overline{K}^p$. On the other hand, the set $\overline{K}^\nu$ is closed w.r.t. the $T_{p,S^{\mathbb{R}}}$-topology because $T_{p,\overline{K}^p} = T_{\nu,\overline{K}^p}$. Hence, $\overline{K}^p \subset \overline{K}^\nu$ and we get $\overline{K}^p = \overline{K}^\nu$.

The rest follows by Theorem 3.3 because $T_{\nu,\overline{K}} = T_{p,\overline{K}} = T_{w,\overline{K}}$. The existence of convex combinations g_n follows for example by [3, p. 422, Corollary V.3.14]. □

Remark 3.5. Consider a compact set $D \subset S$ and a uniformly bounded relatively T_w-compact set $\{t_n\}$ in $C(S)$. Denote

$$m_D = \{\mu : \mu \text{ is a Radon probability measure, } \operatorname{supp}(\mu) = D\}$$

and observe that Corollary 3.4 yields the following implication: If there exists a $\nu \in m_D$ such that $\{t_n\}$ is a ν-Cauchy sequence, then $\{t_n\}$ is μ-Cauchy for all $\mu \in m_D$ and there exists $t \in C(S)$ such that $t_n \to t$ pointwise on D, hence $t_n \to t$ μ-a.s. for all $\mu \in m_D$.

Example 3.6. Assume $S = \{0,1\}^{\mathbb{N}}$ with ν a Radon probability measure on S. Define a continuous mapping $s : S \to [0,1]$ by

$$s(x) = \sum_{k=1}^{\infty} 2^{-k} x_k, \quad \text{where } x = (x_1, x_2, \dots).$$

Denote

$$[y]_n = \frac{k}{2^n} \quad \text{if} \quad y \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right), \quad y \in [0,1]$$

and put

$$h_n(y) = \frac{1}{\sum_{k=1}^{n} 2^{-k}} [y]_n, \quad y \in [0,1].$$

Since obviously $h_n(y) \to y$ on $[0,1]$, the set $\{h_n \circ s : n \in \mathbb{N}\}$ is T_p sequentially relatively compact with continuous (and thus uniformly ν-Lusin) cluster points w.r.t. T_p. By Theorem 3.1, the convergence of $h_n \circ s$ in ν-measure and the convergence ν-almost surely coincide.

Further note that

$$h_n(s(x_1, x_2, \dots)) = \frac{1}{\sum_{k=1}^{n} 2^{-k}} \sum_{k=1}^{n} 2^{-k} x_k.$$

The weighted averages $h_n \circ s$ are equicontinuous w.r.t. T_p and thus $\{h_n \circ s : n \in \mathbb{N}\}$ is a uniformly bounded relatively weakly compact subset of $C(S)$. The equivalence of the two modes of convergence of $h_n \circ s$ stated above follows also by Corollary 3.4.

Our results presented up to now augment to some extent those by Ionescu Tulcea [5], [6] (see also [4, 463C and 463F]). To make some comparison, remark that the essential part of our reasoning, to be found in the statement of Theorem 2.3, reads as follows:
For arbitrary Radon probability ν on S and some $K \subset \mathcal{B}(S)$ there is a Borel set $D \subset S$ such that $\nu(D) = 1$ and

$\nu\{x \in S : f(x) \neq g(x)\} > 0$ happens iff f, g are functions, that are distinct on D.

The above property postulated for example in [4, Theorem 463F] is verified for sets K that are uniformly ν-Lusin by an analysis of the support set for ν.

4. When a set $K \subset \mathcal{B}(S)$ is T_p-compact and uniformly Lusin?

Throughout the present section we shall assume that $K \subset \mathcal{B}(S)$ is a bounded set. Consider the following pair of conditions:

(A)
For arbitrary $F \subset K$ countable and a sequence $\{s_0, s_1, s_2, \dots\} \subset S$ such that $\lim_n f(s_n) = f(s_0)$ exists for all $f \in F$ the convergence is quasi-uniform on F, i.e. for every $\epsilon > 0$ and n_0 there exists a finite number of indices $n_1, \dots, n_k \geq n_0$ such that for each $f \in F$,

$$\min_{i=1,\dots,k} |f(s_{n_i}) - f(s_0)| < \epsilon.$$

(B)
For arbitrary sequence $\{f_n\} \subset K$ and a countable set $A \subset S$ there exists a cluster point of $\{f_n|_A\}$ in $C(A)$ w.r.t. the pointwise convergence.

If S is a compact Hausdorff space and $K \subset C(S)$ a bounded set, then, denoting by $\overline{K}^w$ and $\overline{K}^p$ the closure of K in the $T_{w,C(S)}$ and $T_{p,C(S)}$ topology respectively,

(4.1)
$$\begin{aligned}
(A) &\Longleftrightarrow K \text{ is a sequentially relatively weakly compact subset of } C(S)\\
&\Longleftrightarrow K \text{ is a sequentially relatively compact set in the } T_{p,C(S)}\text{-topology}\\
&\Longleftrightarrow \overline{K}^w \text{ is a compact set in the } T_{w,C(S)}\text{-topology}\\
&\Longleftrightarrow \overline{K}^p \text{ is a compact set in the } T_{p,C(S)}\text{-topology}\\
&\Longleftrightarrow (B)
\end{aligned}$$

are true equivalences.

The former four equivalences are simply the content of Theorem IV.6.14 in [3], p. 269, the latter one is verified in a more general form by Asanov and Veličko (1981) in [1]. A direct verification of the equivalence (A) $\Longleftrightarrow$ (B) might be of some interest.

(A) $\Rightarrow$ (B): Assume $\{f_n\} \subset K$ and a countable set $A \subset S$. The condition (A) implies the sequential relative compactness of K in the $T_{p,C(S)}$-topology. The sequential relative compactness ensures the existence of a subsequence $\{f_{n_k}\} \subset \{f_n\}$ and a function $f \in C(S)$ such that $f_{n_k} \to f$ on S. Hence, $f|_A$ is a cluster point of $\{f_n|_A\}$ in $C(A)$.

(B) $\Rightarrow$ (A): Assume $A = \{s_0, s_1, \dots\} \subset S$, and $f(s_n) \to f(s_0)$ for all $f \in F$ (a countable subset of K). Consider the (metrizable) product topology on $\mathbb{R}^A$ and apply (B) to prove that the closure $\overline{F|_A}$ of $F|_A$ is a compact subset of $C(A)$ in the product topology of $\mathbb{R}^A$. For a while assume that

$$f(s_n) \to f(s_0) \quad \forall f \in \overline{F|_A} \tag{4.2}$$

holds. Define mappings $\hat{s}_n : \overline{F|_A} \to \mathbb{R}$ by

$$\hat{s}_n(f) = f(s_n) \quad \forall f \in \overline{F|_A}, \quad \forall n \in \{0, 1, 2, \dots\}.$$

Clearly, $\hat{s}_n$ are continuous mappings on the compact set $\overline{F|_A}$ for each $n = 0, 1, \dots$. Moreover, by (4.2), $\hat{s}_n \to \hat{s}_0$ on $\overline{F|_A}$. It follows by Theorem IV.6.11 in [3] that $\hat{s}_n \to \hat{s}_0$ quasi-uniformly on $\overline{F|_A}$ and thus quasi-uniformly on F.

It remains to verify (4.2). Take $f \in \overline{F|_A}$ arbitrary and assume that $f(s_n) \nrightarrow f(s_0)$, i.e. that there is a subsequence $\{f(s_{n_k})\}$ such that $|f(s_{n_k}) - f(s_0)| > \epsilon$ for all k and some $\epsilon > 0$. However, by compactness of S there is a subnet $\{t_\alpha\}$ of $\{s_{n_k}\}$ such that $t_\alpha \to s_{00} \in S$. Note that

$$g(s_{00}) = \lim g(t_\alpha) = \lim g(s_{n_k}) = g(s_0) \quad \text{for arbitrary } g \in F.$$

Let $g_n \to f$ on A for some $g_n \in F$ and note that

$$f(s_{00}) = \lim g_n(s_{00}) = \lim g_n(s_0) = f(s_0) \quad \text{if } s_{00} \in A.$$

Consider a countable set $H = A \cup \{s_{00}\}$ and correctly define $h : H \to \mathbb{R}$ by

$$h(s) = f(s) \quad \text{if } s \in A \quad \text{and} \quad h(s_{00}) = f(s_0) \text{ if } s_{00} \notin A.$$

All this implies that $g_n(s) \to h(s)$ for all $s \in H$ and the condition (B) proves that $h \in C(H)$. Hence,

$$\lim f(t_\alpha) = f(s_{00}) = f(s_0),$$

which is a contradiction. Thus, (4.2) is proved.

We need some simple sufficient conditions to identify (relatively) sequentially T_p-compact and uniformly ν-Lusin sets $K \subset \mathcal{B}(S)$.

Theorem 4.1. *Assume that $K \subset \mathcal{B}(S)$ is a set that is bounded and sequentially separable in T_p-topology of $\mathcal{B}(S)$ such that either* (A) *or* (B) *holds. Then K is uniformly ν-Lusin and there is a σ-compact set $D \subset S$ such that*

(4.3) *$\nu(D) = 1, K|_D$ is a sequentially relatively T_p-compact subset of $\mathcal{B}(D)$; and*

(4.4) *the sequential T_p-closure of $K|_D$ in $\mathcal{B}(D)$ is a uniformly ν-Lusin set.*

Proof: Consider $F \subset K$ a countable sequentially T_p-dense set in K. Obviously F is uniformly ν-Lusin, hence there are increasing compacts $D_n \subset S$ such that

$$\nu(D_n) \geq 1 - \frac{1}{n} \quad \text{and} \quad f|_{D_n} \in C(D_n) \quad \forall f \in F.$$

Then $F|_{D_n} \subset K|_{D_n}$ are bounded subsets of $C(D_n)$ that satisfy condition (A) or equivalently (B) and therefore these sets are relatively weakly sequentially compact in $C(D_n)$ for arbitrary $n \in \mathbb{N}$.

Consider now a function $f \in K$ and a sequence $\{f_k\} \subset F$ such that $f_k \to f$

on S. By the diagonal procedure we exhibit a subsequence $\{f_{n_j}\}$ and functions $g_n \in C(D_n)$ such that

$$f_{n_j} \to g_n \quad \text{weakly on } C(D_n) \quad \Rightarrow \quad f_{n_j} \to g_n \quad \text{on } D_n$$
$$\Rightarrow f|_{D_n} = g_n \quad \Rightarrow \quad f|_{D_n} \in C(D_n)$$

and we have proved that K is uniformly ν-Lusin set.
Finally, put $D = \bigcup_1^\infty D_n$ and consider an arbitrary sequence $\{f_k|_D\} \subset K|_D$. Then for each n, the sequence

$\{f_k|_{D_n}\} \subset K|_{D_n}$ satisfies (A) and therefore, according to (4.1),
it has a subsequence that is T_p-convergent to a $g_n \in C(D_n)$.

By the diagonal procedure again we find a subsequence $\{f_{n_j}\}$ such that

$$f_{n_j}|_{D_n} \to g_n \quad \text{on} \quad D_n \quad \forall n.$$

Hence there is a $g \in \mathcal{B}(D)$ such that $f_{n_j}(s) \to g(s)$ for all $s \in D$ and therefore

$K|_D$ is a sequentially relatively T_p-compact subset of $\mathcal{B}(D)$.

Similarly, as in the proof that K is uniformly ν-Lusin, any sequential T_p-cluster point g of $K|D$ is constructed in the following manner:

$g|_{D_n} = g_n \in C(D_n)$, hence the sequential T_p-closure of $K|_D$ in $\mathcal{B}(D)$
is uniformly ν-Lusin set.

□

Let g be an arbitrary T_p-cluster point of $K|_D$ in $\mathcal{B}(S)$, where $D = \bigcup_1^\infty D_n$ is a σ-compact set constructed above. Assuming that S is a completely regular space there are $g_n^e \in C(D)$ such that $g_n^e|_{D_n} = g_n$, hence $g_n^e \to g$ on D and consequently $g \in B_1(D)$, where $B_1(D)$ denotes the space of functions of the first Baire category on D. According to (4.3)

$K|_D$ is a sequentially relatively T_p-compact subset of $B_1(D)$.

Theorem 4.2. *Let $K \subset \mathcal{B}(S)$ be a bounded set that is sequentially separable in T_p-topology of $\mathcal{B}(S)$ such that either* (A) *or* (B) *holds. If $\{f_n\} \subset K$ is a Cauchy sequence in ν measure then there is $f \in \mathcal{B}(S)$ such that $f_n \to f$ ν-a.s.*

Proof: Consider the σ-compact set D constructed in Theorem 4.1. Obviously, $K|_D$ is a sequentially separable subset of $\mathcal{B}(D)$ that satisfies either the condition (A) or the condition (B). It follows by Theorem 4.1 that $K|_D$ is a relatively sequentially T_p-compact set in $\mathcal{B}(D)$ such that the set of its sequential cluster points $SC(K|_D)$ is uniformly $\nu|_D$-Lusin. The rest follows by Theorem 3.1. □

Example 4.3. Consider the Helly space H of non-decreasing functions $f : [0,1] \to [0,1]$. Clearly, each $f \in H$ has at most countably many points of discontinuity. Recall that H is bounded, sequentially separable in $B_1[0,1]$ (H is separable and first countable space) and (sequentially) T_p-compact in $B_1[0,1]$ (see [7, M, p. 164]).

However, H fulfils neither condition (A) nor (B). To see this, consider

$$s_n = 1 - \frac{1}{n}, s_0 = 1; f_n(x) = x^n, \text{ with } A = \{s_0, s_1, \dots\}; F = \{f_1, f_2, \dots\}.$$

If we consider the Lebesgue measure λ on $[0,1]$, then clearly H is not uniformly λ-Lusin. In spite of this, the assertion of Theorem 4.2 holds for H, as we are able to demonstrate that

$$\begin{aligned}&\{f_n\} \subset H \text{ a Cauchy sequence in } \lambda \text{ measure} \Longrightarrow\\ &\exists f \in H \text{ such that } f_n \to f \ \lambda\text{-a.s.}\end{aligned}$$

The argument reads as follows:
There is a subsequence $\{f_{n_k}\} \subset \{f_n\}$ such that $f_{n_k} \to f \in H$ pointwise. Hence, $f_n \to f$ in λ-measure.

Let x_0 be a continuity point of f and fix $\epsilon > 0$. Then

1. $\exists \delta > 0 : |y - x_0| < \delta \Rightarrow |f(y) - f(x_0)| < \frac{\epsilon}{4}$;

and

2. $\exists N > 0 \quad \forall n \geq N : \lambda\{x : |f_n(x) - f(x)| \geq \frac{\epsilon}{4}\} < \delta$.

Thus

3. $\forall n \geq N \quad \exists\, x_{L,n} \in (x_0 - \delta, x_0), x_{H,n} \in (x_0, x_0 + \delta)$:
$|f_n(x_{L,n}) - f(x_{L,n})| < \frac{\epsilon}{4}, |f_n(x_{H,n}) - f(x_{H,n})| < \frac{\epsilon}{4}$.

We conclude

4. $n \geq N \Rightarrow |f_n(x_0) - f(x_0)| \leq |f_n(x_{H,n}) - f(x_0)| +$
$|f(x_0) - f_n(x_{L,n})| \leq |f_n(x_{H,n}) - f(x_{H,n})| + |f(x_{H,n}) - f(x_0)| +$
$|f(x_0) - f(x_{L,n})| + |f(x_{L,n}) - f_n(x_{L,n})| \leq \epsilon$,
where the first inequality follows by the monotonicity of f_n.

Thus $f_n(x_0) \to f(x_0)$ at each continuity point x_0 of f.

5. Application in estimation theory

The results above might be applied to prove the strong consistency of weakly consistent estimators based on their analytical properties. Come back to the estimation theoretic setting 1, 2 and 3 from Section 1 with $S = Z^{\mathbb{N}}$. Denote μ_θ the image measures on S defined by $\mu_\theta(B) = P_\theta[(X_1, X_2, \dots) \in B]$.

Theorem 5.1. *Recall the Bayesian approach* (*see paragraph* 5 *from Section* 1). *Consider a bounded weakly consistent estimator* $\{t_n\}$ *of* $\phi(\theta)$ *such that the sequence satisfies condition* (A) *or* (B) *and an arbitrary* (*Bayesian*) *probability* *distribution F on Θ. Then*

$$\left(t_n(X_1, X_2, \dots) \to \phi(\theta) \quad P_\theta - \text{a.s.} \right) \quad \textit{for } F\textit{-almost all } \theta \in \Theta.$$

Proof: Set $\nu = \int \mu_\theta F(d\theta)$. Obviously, $\{t_n\}$ is a Cauchy sequence in ν-measure. Thus by Theorem 4.2 we obtain $f \in \mathcal{B}(S)$ such that $t_n \to f$ ν-a.s.. The P_θ-a.s. convergence for F-almost all θ follows by the definition of ν. □

Note that for a countable Θ we obtain strong consistency of t_n by taking F with positive mass in each $\theta \in \Theta$.

As shown below, to apply Corollary 3.4, no measure on Θ is needed.

Theorem 5.2. *Assume $K = \{t_n\} \subset C(S)$, where $S = Z^{\mathbb{N}}$. Further assume that there are compacts $S_j \subset S$ for $j = 1, 2, \dots$ such that $K|_{S_j}$ is uniformly bounded, relatively weakly compact subset of $C(S_j)$ for each j and $P_\theta(\bigcup S_j) = 1$ for all $\theta \in \Theta$. Then $\{t_n\}$ is weakly consistent estimator of $\phi(\theta)$ if and only if it is strongly consistent estimator of the parametric function.*

Proof: Apply Corollary 3.4 individually for each $\nu = P_\theta$ and $S = S_j$. □

Recall that the relative weak compactness in $C(S)$ can be verified by (4.1).

6. Conclusion

In conclusion we have to admit that our choice to construct a strongly consistent estimator $\{t_n\}$ as a T_p-compact sequence has some limitations as far as applications to mathematical statistics are concerned. In a very special case of Corollary 3.4 and of the subsequent Remark 3.5 ($\{t_n\} \subset C(S)$ where S is a compact space) we get a strongly consistent estimator that is not fit to distinguish probability distributions supported by the same compact support.

On the other hand, consider $S = \{0,1\}^{\mathbb{N}}$, denote by $p_n : S \to \{0,1\}$ the coordinate projections and by Θ the set of all Borel probability measures μ on S such that p_n converge in μ-measure, denoting by $\phi(\mu)$ the corresponding μ-probability limit. The sequence $\{p_n\}$ being trivially a weakly consistent estimator of $\phi(\mu)$ is obviously not a strongly consistent one. However, in the Bayesian setting of Theorem 5.1 it is easy to construct *almost* strongly consistent subsequence estimator $\{p_{n_k}\}$. More precisely, for arbitrary Bayesian probability distribution F (on the Borel σ-algebra of the set Θ endowed by the topology of the weak convergence of Borel probability measures on S) the barycentric measure $\nu = \int \mu F(d\mu)$ is in Θ and therefore there is a subsequence $\{p_{n_k}\}$ and $\Theta_0 \subset \Theta$ with $F(\Theta_0) = 1$ such that

(6.1) $\{p_{n_k}\}$ is a strongly consistent estimator for the parametric function $\phi|_{\Theta_0}$.

It follows that there is a Borel map $t : S \to \{0,1\}$ ($t = \limsup_k p_{n_k}$) such that

$$t = \phi(\mu) \quad \mu\text{-almost surely for } \mu \in \Theta_0.$$

Recall that such a map t, not necessarily Borel measurable, is called a probability limit identification function (PLIF) for Θ_0. While there is (under the continuum hypothesis) a PLIF for the universum Θ (see [9]), there is no Borel PLIF for this set (see [2]). Thus, the above discussion might serve to further research on how to construct Borel PLIFs for some *small* sets $\Theta_0 \subset \Theta$ started in [8].

Observe that to construct the convergent subsequence $\{p_{n_k}\}$ in (6.1) we need a very complex information about the barycentric probability $\nu = \int \mu F(d\mu)$ so that (6.1) is in fact only an existence statement. Obviously, the constructions of weakly consistent estimators $\{t_n\}$ of $\phi|_{\Theta_0}$ where $\Theta_0 \subset \Theta$ that are automatically strongly consistent might have some merits.

References

[1] Asanov M.O., Veličko N.V., *Kompaktnye množestva v $C_p(X)$*, Comment. Math. Univ. Carolinae **22** (1981), 255–266.

[2] Blackwell D., *There are no Borel SPLIFs*, Ann. Probability **8** (1980), 1189–1190.

[3] Dunford N., Schwartz J.T., *Linear Operators Part I: General Theory*, John Wiley & Sons, Inc., New Jersey, 1988.

[4] Fremlin D.H., *Measure Theory, Vol 4, Topological Measure Spaces*, Colchester: Torres Fremlin, 2003.

[5] Ionescu Tulcea A., *On pointwise convergence, compactness and equicontinuity I*, Z. Wahrscheinlichkeitstheorie und verw. Gebiete **26** (1973), 197–205.

[6] Ionescu Tulcea A., *On pointwise convergence, compactness and equicontinuity II*, Advances in Math. **12** (1974), 171–177.

[7] Kelley J.L., *General Topology*, Springer, New York, 1975.

[8] Kříž P., *How to construct Borel measurable PLIFs?*, WDS'11 Proc. of Contr. Papers, Part I, (2011), 43–48.

[9] Štěpán J., *The probability limit identification function exists under the continuum hypothesis*, Ann. Probability **1** (1973), 712–715.

Charles University, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague 8, Czech Republic

E-mail: pavel-kriz@post.cz
stepan@karlin.mff.cuni.cz

(*Received* November 30, 2012*, revised* May 30, 2013)