[Turkish Journal of Mathematics](https://journals.tubitak.gov.tr/math)

Volume 42 | Number 4 Article 16

1-1-2018

Weak convergence of probability measures to Choquet capacity functionals

DIETMAR FERGER

Turkish Journal of Mathematics

http://journals.tubitak.gov.tr/math/

Research Article

Turk J Math
(2018) 42: 1747 – 1764

doi:10.3906/mat-1705-106

Weak convergence of probability measures to Choquet capacity functionals

Dietmar FERGER
Department of Mathematics, Technische Universität Dresden, Germany

Received: 26.05.2017 • **Accepted/Published Online:** 28.03.2018 • **Final Version:** 24.07.2018

Abstract: In the definition of weak convergence of probability measures it is assumed that the limit is a probability measure as well. We get rid of this assumption and require that the limit merely needs to be a Choquet-capacity functional. In terms of random variables this means that the distributional limit no longer is a random point, but a random closed set, namely one uniquely determined by the Choquet capacity. For our extended notion of weak convergence there is a counterpart of the portmanteau theorem. Moreover, we demonstrate basic relations to the theory of random closed sets with emphasis on weak convergence in hyperspace topologies including two correspondence theorems. Finally, the approach carries over to sequences of Choquet capacities.

Key words: Weak convergence, Choquet capacity, Fell topology, upper Vietoris topology, random closed set

1. Introduction

Let E with topology $\mathcal{G}$ be a locally compact second countable Hausdorff space (lcscH). In particular, by Uryson's Metrization Lemma there exists a metric d on E. The pertaining classes of all closed, compact, and Borel subsets are denoted by $\mathcal{F}, \mathcal{K}$, and $\mathcal{B}$, respectively. We make $\mathcal{F}$ a topological space by endowing it with the *Fell topology* τ_{Fell}, which is generated from a subbase

$$\{\mathcal{M}(K) : K \in \mathcal{K}\} \cup \{\mathcal{H}(G) : G \in \mathcal{G}\},$$

where $\mathcal{M}(D) := \{F \in \mathcal{F} : F \cap D = \emptyset\}$ is a *missing set* and $\mathcal{H}(D) := \{F \in \mathcal{F} : F \cap D \neq \emptyset\}$ is a *hitting set* for every subset D of E. From [10] we know that $(\mathcal{F}, \tau_{Fell})$ is compact, second countable, and Hausdorff. Convergence in the Fell topology has nice equivalent characterizations:

$$F_n \to F \text{ in } \tau_{Fell} \quad \Leftrightarrow \quad K - \lim_{n\to\infty} F_n = F \quad \Leftrightarrow \quad \delta(F_n, F) \to 0. \tag{1.1}$$

Here $K - \lim_{n\to\infty} F_n$ means the Painlevé–Kuratowski limit of (F_n) and δ is the Kuratowski metric; confer [17,19]. Let $\mathcal{B}_{Fell} := \sigma(\tau_{Fell})$ denote the Borel-σ algebra on $\mathcal{F}$ pertaining to τ_{Fell} and let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Then a measurable mapping $C : (\Omega, \mathcal{A}) \to (\mathcal{F}, \mathcal{B}_{Fell})$ is called a *random closed set in* E. For any random closed set C the set function $T_C : \mathcal{K} \to \mathbb{R}$ defined by

$$T_C(K) := \mathbb{P}(C \cap K \neq \emptyset),\ K \in \mathcal{K},$$

is called the *capacity functional of* C. The properties of $\mathbb{P}$ lead to the following characteristics of T_C:

*Correspondence: dietmar.ferger@tu-dresden.de
2010 *AMS Mathematics Subject Classification:* Primary 60B05, 60B10; Secondary 26E25, 54B20

(T1) $T_C(\emptyset) = 0;\ \ 0 \leq T_C \leq 1;$

(T2) T_C is continuous from above on $\mathcal{K}$, i.e. $K_n \downarrow K$ in $\mathcal{K} \quad \Rightarrow \quad T_C(K_n) \downarrow T_C(K)$;

(T3) T_C is monotone increasing on $\mathcal{K}$ and for $K_1, K_2, \ldots, K_n \in \mathcal{K}, n \geq 2$,

$$T_C\Big(\bigcap_{i=1}^n K_i\Big) \leq \sum_{\emptyset \neq I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} T_C\Big(\bigcup_{i \in I} K_i\Big).$$

Every functional $T : \mathcal{K} \to \mathbb{R}$ satisfying (T1)–(T3) is called *Choquet capacity (functional)*. The following two results on Choquet capacities are well known in the theory of random closed sets; confer [10–12].

Theorem 1.1 ***(Choquet)*** *Every probability measure Q on $(\mathcal{F}, \mathcal{B}_{Fell})$ determines a Choquet capacity functional T on $\mathcal{K}$ through the correspondence*

$$T(K) = Q(\mathcal{H}(K)) \quad \forall\, K \in \mathcal{K}. \tag{1.2}$$

Conversely, every Choquet capacity functional T on $\mathcal{K}$ determines a unique probability measure Q on $(\mathcal{F}, \mathcal{B}_{Fell})$ that satisfies the relation (1.2).

We see that Choquet's Theorem is the counterpart of the well-known one-to-one correspondence between a distribution function on $\mathbb{R}^d$ and a probability measure on the usual Borel-σ algebra $\mathcal{B}(\mathbb{R}^d)$ on $\mathbb{R}^d, d \in \mathbb{N}$. Given a Choquet capacity T there exists a random closed set C in E such that T is equal to the capacity functional T_C of C. Let us agree to say in this case that C *is associated with* T. As to the existence, simply put $(\Omega, \mathcal{A}, \mathbb{P}) := (\mathcal{F}, \mathcal{B}_{Fell}, Q)$, where Q is the probability measure (uniquely) determined by T according to Choquet's Theorem. Then $C :=$ identity map is a random closed set with $T_C = T$ by construction. For any other random closed set D in E with $T_D = T$ it follows from Choquet's Theorem that D and C are equal in distribution. Each Choquet capacity T can be extended onto the Borel-σ algebra $\mathcal{B}$ by

$$T(B) := \sup\{T(K) : K \in \mathcal{K}, K \subseteq B\}, \quad B \in \mathcal{B}.$$

Theorem 1.2 ***(Matheron)*** *The extension $T : \mathcal{B} \to [0,1]$ is consistent in the sense that $T(B) = Q(\mathcal{H}(B)) \quad \forall\, B \in \mathcal{B}$, where the hitting sets $\mathcal{H}(B) = \{F \in \mathcal{F} : F \cap B \neq \emptyset\}$ in fact belong to $\mathcal{B}_{Fell}$.*

Therefore, from now on a Choquet capacity is a set function on the entire Borel-σ algebra $\mathcal{B}$. In general, it is not a probability measure on $\mathcal{B}$, because it is not σ-additive, but merely sub-σ-additive. Obviously, every probability measure satisfies (T1)–(T3) and thus is a Choquet capacity. Conversely, a Choquet capacity is in fact a probability measure if and only if the associated random set is a singleton with probability one; confer Proposition 2.7 below.

The definition of weak convergence of probability measures defined on the Borel-σ algebra of general topological spaces, confer Topsøe [22], includes that the limit is a probability measure as well. We extend this by allowing the limit to be a Choquet capacity. In view of the "limsup characterization" of weak convergence as given in the portmanteau theorem, confer Theorem 8.1 of Topsøe [22], we arrive at:

Definition 1.3 *Let $P_n, n \in \mathbb{N}$, be probability measures on $(E, \mathcal{B})$ and let T be a Choquet capacity. If*

$$\limsup_{n\to\infty} P_n(F) \leq T(F) \quad \forall\, F \in \mathcal{F}, \tag{1.3}$$

then we say that P_n converges weakly *to T and denote this by $P_n \xrightarrow{w} T$.*

Definition 1.4 *For every $n \in \mathbb{N}$ let $\xi_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \to (E, \mathcal{B})$ be a random variable with values in E and with distribution $\mathbb{P}_n \circ \xi_n^{-1}$. If $\mathbb{P}_n \circ \xi_n^{-1} \xrightarrow{w} T$, then we say that ξ_n* converges in distribution *to C, where C is the random closed set associated with T. In short we write:*

$$\xi_n \xrightarrow{\mathcal{L}} C. \tag{1.4}$$

Remark 1.5 *Notice that E is closed, whence from (1.3) and (T1) it follows that $T(E) = 1$, which in turn yields that $C \neq \emptyset$ a.s. Observe that the ξ_n are random* points *in the space E, whereas the limit C is a random (closed)* subset *of E. By definition the convergence in (1.4) is equivalent to*

$$\limsup_{n\to\infty} \mathbb{P}_n(\xi_n \in F) \leq \mathbb{P}(C \cap F \neq \emptyset) \quad \forall\, F \in \mathcal{F}, \tag{1.5}$$

where $(\Omega, \mathcal{A}, \mathbb{P})$ is the underlying probability space of C. If $C = \{\xi\}$ is a singleton a.s. then $\mathbb{P}(C \cap F \neq \emptyset) = \mathbb{P}(\xi \in F)$. Thus (1.5) and the (traditional) portmanteau theorem yield classical distributional convergence $\xi_n \xrightarrow{\mathcal{L}} \xi$.

Remark 1.6 *The limit T in (1.3) is not uniquely determined, since whenever T is dominated by some other Choquet capacity $\tilde{T}$, i.e. $T \leq \tilde{T}$, then $\tilde{T}$ fulfills (1.3) and therefore is a limit as well. In particular, $\tilde{T} = T_E$ is the degenerated limit, which always exists. In terms of random closed sets this means that $\xi_n \xrightarrow{\mathcal{L}} C$ implies that $\xi_n \xrightarrow{\mathcal{L}} D$ for every random closed set D such that $C \subseteq D$ a.s. and so $\xi_n \xrightarrow{\mathcal{L}} E$ in any case.*

Remark 1.7 *If T actually is a probability measure, then we are back in the classical situation (as in Remark 1.5 above). To stress this we write $P_n \overset{w}{\Rightarrow} T$.*

The following example shows an applicability of our convergence concept (1.4) in statistics for the construction of confidence regions based on the well-established principle of M-estimation.

Example 1.8 *In a statistical model let $\theta \in \mathbb{R}^d$ be the parameter of interest and let $\hat{\theta}_n$ be an* M-estimator, *i.e. $\hat{\theta}_n$ is any infimizing point of some random cadlag (rcll) criterion function $M_n(t), t \in \mathbb{R}^d$. As the usual starting point one has to investigate the asymptotic distributional behavior of $\alpha_n(\hat{\theta}_n - \theta), n \in \mathbb{N}$, with suitable sequence $\alpha_n \to \infty$. The basic idea here is to introduce the* rescaled *M_n-process X_n defined as*

$$X_n(t) := \gamma_n\{M_n(\theta + \alpha_n^{-1}t) - M_n(\theta)\},\ t \in \mathbb{R}^d,$$

where γ_n are appropriate positive constants. In fact, by Lemma 2.2 in Ferger [7] the rescaling $t \to \theta + \alpha_n^{-1}t$ yields that $\alpha_n(\hat{\theta}_n - \theta)$ is an infimizing point of the cadlag process X_n. Now, if $X_n \xrightarrow{\mathcal{L}} X$ in the multivariate Skorokhod-space $(D(\mathbb{R}^d), s)$ and if the sequence $\alpha_n(\hat{\theta}_n - \theta), n \in \mathbb{N}$, is stochastically bounded, then Theorem 3.11 *in [7] guarantees that* $\alpha_n(\hat{\theta}_n - \theta) \stackrel{\mathcal{L}}{\to} C$*, where* C *is the random closed set consisting of all infimizing points of the limit process* X*. Next, put*

$$R_n := \hat{\theta}_n - \alpha_n^{-1} G := \{\hat{\theta}_n - \alpha_n^{-1} x : x \in G\}$$

with open $G \subseteq \mathbb{R}^d$*. Then* $\{\theta \in R_n\} = \{\alpha_n(\hat{\theta}_n - \theta) \in G\}$*, whence*

$$\begin{aligned} \liminf_{n\to\infty} \mathbb{P}(\theta \in R_n) &= \liminf_{n\to\infty} \mathbb{P}(\alpha_n(\hat{\theta}_n - \theta) \in G) = 1 - \limsup_{n\to\infty} \mathbb{P}(\alpha_n(\hat{\theta}_n - \theta) \in G^c) \\ &\geq 1 - \mathbb{P}(C \cap G^c \neq \emptyset) = \mathbb{P}(C \cap G^c = \emptyset) = \mathbb{P}(C \subseteq G), \end{aligned} \tag{1.6}$$

where the inequality in (1.6) holds by (1.5), because the complement G^c *of* G *is closed. Thus, if for a given level* $\alpha \in (0,1)$ *of significance one chooses in the definition of* R_n *the open set* $G = G_\alpha$ *such that* $\mathbb{P}(C \subseteq G) \geq 1-\alpha$*, then*

$$\liminf_{n\to\infty} \mathbb{P}(\theta \in R_n) \geq 1 - \alpha$$

and thereby R_n *is an asymptotic confidence region for* θ *at level* $1-\alpha$*. For instance, let* G *be the open rectangle* $(-r,r)^d$ *with positive* r*. Then the pertaining confidence region is given by*

$$R_n = (\hat{\theta}_{n,1} - \alpha_n^{-1} r, \hat{\theta}_{n,1} + \alpha_n^{-1} r) \times \cdots \times (\hat{\theta}_{n,d} - \alpha_n^{-1} r, \hat{\theta}_{n,d} + \alpha_n^{-1} r),$$

where $\hat{\theta}_{n,i}$ *is the* $i-$*th component of* $\hat{\theta}_n = (\hat{\theta}_{n,1}, \ldots, \hat{\theta}_{n,d})$*. Notice that in the literature up to now the construction of asymptotic confidence regions required* ***classical*** *distributional convergence* $\alpha_n(\hat{\theta}_n - \theta) \stackrel{\mathcal{L}}{\to} \xi$ *to a limit* ***point*** ξ*. Thus, the essential innovation here is that distributional convergence* $\alpha_n(\hat{\theta}_n - \theta) \stackrel{\mathcal{L}}{\to} C$ *to a limit* ***set*** C *suffices for the construction, even if* C *is not a singleton. (If* $C = \{\xi\}$ *a.s. then both notions of convergence coincide.)*

This paper is organized as follows. In section 2 we derive several necessary conditions for $P_n \xrightarrow{w} T$, which are even sufficient, if the set C associated with T is compact a.s. If T actually is a measure we obtain the classical portmanteau theorem as presented in [8,22]. Recall that by Proposition 2.7 below T is a measure if and only if C is a singleton a.s. Indeed, singletons play a peculiar role in our theory. They are the most simple nonempty random closed sets. We investigate them in section 3 and show that once a sequence of singletons converges in distribution in the hyperspace $(\mathcal{F}, \tau_{Fell})$ then the limit set must be a singleton too. In this sense singletons are "closed". In section 4 we demonstrate how traditional distributional convergence in the carrier space $(E, \mathcal{G})$ and distributional convergence (of singletons) in the hyperspace are linked to each other as well as to our convergence (1.4). Moreover, if τ_{Fell} is replaced by the so-called *upper Vietoris topology* τ_{uV} it turns out that (1.4) is equivalent to distributional convergence of the singletons $\{\xi_n\}$ in the new hyperspace $(\mathcal{F}, \tau_{uV})$. Finally, in the last section the probability measures P_n in Definition 1.3 are replaced by Choquet capacities T_n. This yields a new notion of weak convergence for Choquet capacities, which corresponds to a suitable topology on the set $\mathcal{T}$ of all Choquet capacities. This topology is coarser than the narrow and the vague topology. At first sight one could expect that it coincides with the *weak convergence in the Choquet sense* as introduced by Feng and Nguyen [6]. However, the last one is strictly stronger and it coincides with the classical weak convergence, if in turn one considers the special case that $T_n = P_n$. Then the weak limit in the Choquet sense inevitably must be a probability measure as well. This does not apply to $P_n \xrightarrow{w} T$, because here T needs not to be a probability measure as demonstrated by examples. In fact this is the reason why our notion of weak convergence is more flexible in statistical application, in particular when constructing confidence regions.

2. Portmanteau theorem

We will derive our main result in this section step by step. Given a Choquet capacity $T = T_C$ with associated random set C there is the pertaining so-called *containment functional* $T^* = T^*_C$ given by $T^*(B) := \mathbb{P}(C \subseteq B)$, $B \in \mathcal{B}$. Clearly, $T^*(B) = 1 - T(B^c)$, where $B^c = E \setminus B$ denotes the complement for (any) set $B \subseteq E$. Therefore, our first lemma follows from Definition 1.3 simply by complementation.

Lemma 2.1 *P_n converges weakly to T if and only if* $\liminf_{n\to\infty} P_n(G) \geq T^*(G) \quad \forall\, G \in \mathcal{G}$.

Our next result involves the definition of the *Choquet integral* with respect to T. For any $\mathcal{B}$-measurable function $f : E \to \mathbb{R}$ one defines

$$\int f dT := \int_0^\infty T(f \geq t)\lambda(dt) + \int_{-\infty}^0 [T(f \geq t) - T(E)]\lambda(dt), \tag{2.1}$$

where λ denotes the Lebesgue measure on $\mathbb{R}$. A careful introduction and several properties of the Choquet integral are given in [12]. In fact, the integral is well-defined for any monotone set-function T on $\mathcal{B}$ and $\geq$ can be replaced by $>$ without changing the value. It coincides with the standard integral in the case T is a finite measure. In the following we will work with upper semicontinuous (usc) and lower semicontinuous (lsc) functions $f : E \to \mathbb{R}$.

Proposition 2.2 *If $P_n \xrightarrow{w} T$, then*

$$\limsup_{n\to\infty} \int f dP_n \leq \int f dT \quad \textit{for all functions } f : E \to \mathbb{R} \textit{ usc and bounded.}$$

Proof Let f be usc and bounded. Then there exist reals $a < 0 < b$ such that $a \leq f < b$. Observe that $T(f \geq t) = T(\emptyset) = 0$ for every $t \geq b$, whence $\int_0^\infty T(f \geq t)\lambda(dt) = \int_0^b T(f \geq t)\lambda(dt)$. Similarly, for $t \leq a$ one has that $T(f \geq t) - T(E) = T(E) - T(E) = 0$ and so $\int_{-\infty}^0 [T(f \geq t) - T(E)]\lambda(dt) = \int_a^0 [T(f \geq t) - T(E)]\lambda(dt)$. By definition (2.1) and linearity we arrive at

$$\int f dT = \int_a^b T(f \geq t)\lambda(dt) + aT(E) = \int_a^b T(f \geq t)\lambda(dt) + a, \tag{2.2}$$

where the second equality follows, because $T(E) = 1$ by Remark 1.5. Next, for every $k \in \mathbb{N}$ we introduce the equidistant points $t_i := a + i\,\frac{b-a}{k}$, $\quad 0 \leq i \leq k$, and the sets $F_i := \{f \geq t_i\}$, $\quad 0 \leq i \leq k$, which by upper semicontinuity of f are closed subsets of E with $E = F_0 \supseteq F_1 \supseteq \ldots \supseteq F_k = \emptyset$. Since

$$f = \sum_{i=1}^k 1_{\{t_{i-1} \leq f < t_i\}} f < \sum_{i=1}^k 1_{\{t_{i-1} \leq f < t_i\}} t_i = a + (b-a)\frac{1}{k}\sum_{i=1}^k i\,(1_{F_{i-1}} - 1_{F_i}),$$

we obtain that

$$\int f dP_n \le a + (b-a)\frac{1}{k}\sum_{i=1}^{k} i\ (P_n(F_{i-1}) - P_n(F_i)).$$

Use $P_n(F_0) = 1$ and $P_n(F_k) = 0$ when rearranging the sum in the above formula:

$$\sum_{i=1}^{k} i\ (P_n(F_{i-1}) - P_n(F_i)) = 1 + \sum_{i=1}^{k} P_n(F_i).$$

This yields that

$$\int f dP_n \le a + (b-a)\frac{1}{k}(1 + \sum_{i=1}^{k} P_n(F_i)) = a + \frac{b-a}{k} + \frac{b-a}{k}\sum_{i=1}^{k} P_n(F_i).$$

Consequently, for every $k \in \mathbb{N}$ it is

$$\limsup_{n\to\infty} \int f dP_n \le a + \frac{b-a}{k} + \frac{b-a}{k}\sum_{i=1}^{k} \limsup_{n\to\infty} P_n(F_i) \le a + \frac{b-a}{k} + \frac{b-a}{k}\sum_{i=1}^{k} T(F_i), \tag{2.3}$$

where the last equality holds by (1.3), since $F_i \in \mathcal{F}, 1 \le i \le k$. Notice that $t_i - t_{i-1} = (b-a)/k$ for every $1 \le i \le k$. Therefore,

$$S_k := \frac{b-a}{k}\sum_{i=1}^{k} T(F_i) = \sum_{i=1}^{k} T(f \ge t_i)(t_i - t_{i-1}).$$

This means that S_k is a Riemann sum of the (bounded) function $g(t) := T(f \ge t), t \in [a,b]$. Monotonicity of T ensures that g is decreasing and thus its pertaining set of all discontinuity points is at most countable and consequently has Lebesgue measure zero. From Lebesgue and Riemann integration theory, we can infer that g is Riemann-integrable and that its Riemann integral coincides with the Lebesgue integral. Conclude that

$$S_k \to \int_a^b T(f \ge t)dt = \int_a^b T(f \ge t)\lambda(dt),\ k \to \infty.$$

Thus taking the limit $k \to \infty$ in (2.3) yields the desired result upon noticing the equality. (2.2) □

For the next step on our way to the portmanteau theorem we use the following equality, which can easily be verified:

$$\int f dT^* = -\int -f dT. \tag{2.4}$$

With (2.4) the proof of the next lemma is elementary upon noticing that f is lsc if and only if $-f$ is usc.

Lemma 2.3 *The following two statements are equivalent:*

(i) $\limsup_{n\to\infty} \int f dP_n \le \int f dT$ *for all functions* $f : E \to \mathbb{R}$ *usc and bounded.*

(ii) $\liminf_{n\to\infty} \int f dP_n \ge \int f dT^*$ *for all functions* $f : E \to \mathbb{R}$ *lsc and bounded.*

The next result gives the reverse conclusion in Proposition 2.2.

Proposition 2.4 *Let T be such that its associated C is compact a.s. If*

$$\limsup_{n\to\infty} \int f dP_n \leq \int f dT \tag{2.5}$$

for all functions $f: E \to \mathbb{R}$ uniformly continuous (uc) and bounded, then $P_n \overset{w}{\to} T$.

Proof First, recall that there is a metric d on E. For a given nonempty set $F \in \mathcal{F}$ we introduce for every $k \in \mathbb{N}$ functions $f_k : E \to \mathbb{R}$ by $f_k(x) := \phi(kd(x,F)), x \in E$, where $\phi : [0,\infty) \to [0,1]$ is defined by $\phi(t) := 1_{[0,1]}(t)(1-t), t \geq 0$, and $d(x,F)$ is the distance of the point x from the set F, i.e. $d(x,F) := \inf\{d(x,y) : y \in F\}$. It is well known that

$$|d(x,F) - d(y,F)| \leq d(x,y) \quad \forall\, x,y \in E \quad \text{and that} \quad d(x,F) = 0 \Leftrightarrow x \in F. \tag{2.6}$$

Infer from this that f_k is uniformly continuous and bounded and that $f_k \downarrow 1_F, k \to \infty$. In particular, $1_F \leq f_k$ for all k, and thus

$$\limsup_{n\to\infty} P_n(F) = \limsup_{n\to\infty} \int 1_F dP_n \leq \limsup_{n\to\infty} \int f_k dP_n \leq \int f_k dT \quad \forall\, k \in \mathbb{N}, \tag{2.7}$$

where the last inequality follows from (2.5). Thus it remains to show that

$$\int f_k dT \to T(F), \quad k \to \infty.$$

To see this notice that $0 \leq f_k \leq 1$ by construction, and so the Choquet integral of f_k simplifies, confer (2.2) above, to

$$\int f_k dT = \int_0^1 T(f_k \geq t)\lambda(dt).$$

We consider the integrand $g_k(t) := T(f_k \geq t) = \mathbb{P}(A_k(t)), t \in [0,1]$, with $A_k(t) := \{C \cap \{f_k \geq t\} \neq \emptyset\}$. Since (f_k) is monotone decreasing, $A_k(t) \downarrow$ for every $t \in [0,1]$, whence (g_k) is monotone decreasing as well. Moreover, $\lim_{k\to\infty} g_k(t) = T(F)\ \forall\, t \in [0,1]$. This follows from $\bigcap_{k\geq 1} A_k(t) = \{C \cap F \neq \emptyset\}$. Indeed, assume that the event on the left-hand side occurs, that is $C \cap \{f_k \geq t\} \neq \emptyset\ \forall\ k \geq 1$. Then for every $k \geq 1$ there exists a point $x_k \in C$ with $f_k(x_k) \geq t$. W.l.o.g. we may assume that C is compact on the entire sample space Ω. Therefore, we find a subsequence (x_{k_l}) in C such that $x_{k_l} \to x \in C$ as $l \to \infty$. It is easy to check that $f_{k_l}(x_{k_l}) \geq t$ if and only if $0 \leq d(x_{k_l}, F) \leq (1-t)/k_l$. Taking the limit $l \to \infty$ yields that $d(x,F) = 0$ by continuity of $d(\cdot, F)$. Consequently, $x \in F$ by the second part in (2.6), but also $x \in C$, and so $C \cap F \neq \emptyset$. Conversely, if $C \cap F \neq \emptyset$ we find some $x \in C \cap F$. Since $x \in F$, it follows that $f_k(x) = \phi(kd(x,F)) = \phi(0) = 1 \geq t$ for all $t \in [0,1]$ and for every $k \geq 1$. This means that $x \in \{f_k \geq t\}$, but also $x \in C$, whence $C \cap \{f_k \geq t\} \neq \emptyset$ for all $k \geq 1$ as desired.

To sum up, (g_k) is monotone decreasing with (constant) limit $T(F)$. By the monotone convergence theorem we obtain

$$\int f_k dT = \int_0^1 g_k(t)\lambda(dt) \to \int_0^1 T(F)\lambda(dt) = T(F), \quad k \to \infty.$$

Hence, taking the limit $k \to \infty$ in (2.7) yields $\limsup_{n\to\infty} P_n(F) \leq T(F)$ for all nonempty closed subsets F. For $F = \emptyset$ this relation is trivially fulfilled and we can conclude that $P_n \overset{w}{\to} T$. □

Theorem 2.5 ***(Portmanteau theorem)*** *Let T be a Choquet capacity with associated random closed set C. Consider the following statements:*

(1) $P_n \overset{w}{\to} T$.

(2) $\limsup_{n\to\infty} P_n(F) \le T(F)$ *for all closed subsets* $F \subseteq E$.

(3) $\liminf_{n\to\infty} P_n(G) \ge T^*(G)$ *for all open subsets* $G \subseteq E$.

(4) $\limsup_{n\to\infty} \int f dP_n \le \int f dT$ *for all usc and bounded functions* f.

(5) $\liminf_{n\to\infty} \int f dP_n \ge \int f dT^*$ *for all lsc and bounded functions* f.

(6) $\limsup_{n\to\infty} \int f dP_n \le \int f dT$ *for all continuous and bounded functions* f.

(7) $\liminf_{n\to\infty} \int f dP_n \ge \int f dT^*$ *for all continuous and bounded functions* f.

(8) $\limsup_{n\to\infty} \int f dP_n \le \int f dT$ *for all uc and bounded functions* f.

(9) $\liminf_{n\to\infty} \int f dP_n \ge \int f dT^*$ *for all uc and bounded functions* f.

Then the following relations hold:

$$(1) \Leftrightarrow (2) \Leftrightarrow (3) \Rightarrow (4) \Leftrightarrow (5) \Rightarrow (6) \Leftrightarrow (7) \Rightarrow (8) \Leftrightarrow (9).$$

If C is compact a.s., then all statements are equivalent.

Proof (1) $\Leftrightarrow$ (2) by definition. (2) $\Leftrightarrow$ (3) by Lemma 2.1. (3) $\Rightarrow$ (4) is valid by Proposition 2.2. (4) $\Leftrightarrow$ (5) follows from Lemma 2.3. (5) $\Rightarrow$ (6) is trivial. (6) $\Leftrightarrow$ (7) and (8) $\Leftrightarrow$ (9) can be shown exactly in the same fashion as Lemma 2.3. (7) $\Rightarrow$ (8) is trivial. If C is compact, then (9) $\Rightarrow$ (1) by Proposition 2.4, which finally gives the assertion. □

Since Theorem 2.5 extends Theorem 8.1 of Topsøe [22] we adopt his denomination "portmanteau theorem". This type of results giving various equivalent conditions for convergence of measures, is also known as "Alexandrov's theorem". We refer the reader to Pfanzagl [16] for historical comments and for different naming conventions.

Remark 2.6 *Salinetti and Wets [21] consider the space $SC^u(\mathcal{F};[0,1])$ of all usc functions on $(\mathcal{F},\tau_{Fell})$ with values in $[0,1]$ equipped with the* hypo-topology $\mathcal{T}_{hypo}$. *(Actually, they use the* Wijsman topology τ_W, *but for our E lcscH it is well known that $\tau_{Fell}=\tau_W$.) They show that for every probability measure P on $\mathcal{B}$ there exists exactly one* probability sc-measure λ *and vice versa (notation: $P \leftrightarrow \lambda$). Moreover, let P and $P_n, n\in\mathbb{N}$, be probability measures on $\mathcal{B}$ with corresponding probability sc-measures λ and $\lambda_n, n \in \mathbb{N}$, that is $P\leftrightarrow\lambda$ and $P_n \leftrightarrow \lambda_n$ for all $n\in\mathbb{N}$. Then it is proved that*

$$P_n \overset{w}{\to} P \quad \Leftrightarrow \quad \lambda_n \to \lambda \ \text{ in } (SC^u(\mathcal{F};[0,1]),\mathcal{T}_{hypo}). \tag{2.8}$$

Now what can be said if P in (2.8) is replaced by a Choquet capacity T? To answer this question, assume that the restriction $T:(\mathcal{F},\tau_{Fell})\to[0,1]$ is usc, then it follows with the same arguments of [21] for the validity of $\Leftarrow$ in (2.8) that

$$P_n \overset{w}{\to} T \quad \Leftarrow \quad \lambda_n \to T \ \text{ in } (SC^u(\mathcal{F};[0,1]),\mathcal{T}_{hypo}). \tag{2.9}$$

The other direction $\Rightarrow$ in (2.9) cannot hold, because the hypo-limit is uniquely determined in contrast to our weak limit T; confer Remark 1.6. As to the assumption on T notice that according to Wei et al. [25] $T : (\mathcal{F}, \tau_{Fell}) \to [0,1]$ is usc if and only if T is continuous from above on $\mathcal{F}$, which in general is stronger than continuity from above on $\mathcal{K}$ as required by definition in (T2). In particular, any T is usc as long as E is compact, and for noncompact E this is in general not true; confer Example 3.1 of Wei et al. [25]. Therefore, the approach reported by Salinetti and Wets [21] works very well for traditional weak convergence via the *correspondence theorem* (2.8), but for our extended notion of weak convergence it is of little use. However, in Theorem 4.2 below we will give a suitable corresponding relation for $P_n \stackrel{w}{\to} T$ in terms of the associated random closed sets.

We end this section with a necessary and sufficient condition via the Choquet capacity functional of a random closed set C, which in particular guarantees that C is a singleton. It will be a very useful tool in the next sections. A proof is given in [7].

Proposition 2.7 *Let T_C be the capacity functional of a random closed set C defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then the following statements are equivalent:*

(i) T_C is a probability measure.

(ii) $C = \{\xi\}$ $\mathbb{P}$ - a.s. for some random variable $\xi : (\Omega, \mathcal{A}) \to (E, \mathcal{B})$.

We would like to mention that there is a weaker version of the above proposition, where the almost sure equality in (ii) is replaced by equality in distribution. This follows easily from Theorem 1.1 of Choquet.

3. Distributional convergence of singletons in hyperspaces

As before let $\xi_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \to (E, \mathcal{B})$ be random variables with values in E, $n \in \mathbb{N}$. Then the singletons $\{\xi_n\}, n \in \mathbb{N}$, are random closed sets. We wish to relate $\xi_n \stackrel{\mathcal{L}}{\to} C$ to distributional convergence of the singletons $\{\xi_n\}$ in appropriate hyperspaces and also to distributional convergence of ξ_n in the basic space $(E, \mathcal{G})$. The following proposition will help to give an answer. In short it says that the distributional limit of singletons necessarily has to be a singleton as well. Here we use the notation $C_n \stackrel{\mathcal{L}}{\to} C$ in $(\mathcal{F}, \tau_{Fell})$ to indicate that the random closed sets C_n as random variables in the measurable space $(\mathcal{F}, \mathcal{B}_{Fell})$ converge in distribution to the random closed set C. More precisely, the distributions of C_n as probability measures on $\mathcal{B}_{Fell}$ converge weakly in the sense of Topsøe [22] to the distribution of C.

Proposition 3.1 *If $\{\xi_n\} \stackrel{\mathcal{L}}{\to} C$ in $(\mathcal{F}, \tau_{Fell})$, where C is a random closed set on $(\Omega, \mathcal{A}, \mathbb{P})$ with $C \neq \emptyset$ $\mathbb{P}$-a.s., then there exists a random variable $\xi : (\Omega, \mathcal{A}) \to (E, \mathcal{B})$ such that $C = \{\xi\}$ $\mathbb{P}$-a.s.*

Proof Let $T := T_C$ be the capacity functional of C with corresponding probability measure Q. By Proposition 2.7 we have to show that T_C is a probability measure. To begin with consider $\mathcal{K}_T := \{K \in \mathcal{K} : T(K) = T(K^o)\} = \{K \in \mathcal{K} : Q(\partial\mathcal{H}(K) = 0\}$, where the second equality holds according to Lemma 7.2 in [12]. (This lemma is formulated only for $E = \mathbb{R}^d$, but its proof carries over analogously to general E lcscH.) From Theorem 2.1 of Norberg [13] we can conclude that $\mathbb{P}_n(\xi_n \in K) = \mathbb{P}_n(\{\xi_n\} \cap K \neq \emptyset\} \to T(K) \quad \forall\, K \in \mathcal{K}_T$, whence

$$\lim_{n\to\infty} \mathbb{P}_n(\xi_n \notin K) = 1 - T(K) \quad \forall\, K \in \mathcal{K}_T. \tag{3.1}$$

With a look at (3.1) we will construct a sequence $(K_i) \subseteq \mathcal{K}_T$ such that $T(K_i) \uparrow 1, i \to \infty$. For that purpose fix some $x_0 \in E$ and let $B_r := B(x_0, r) := \{x \in E : d(x, x_0) \le r\}, r > 0$. These closed balls are compact. By (1.10) of [20],

$$\partial\mathcal{H}(K) = \{F \in \mathcal{F} : \emptyset \neq F \cap K \subseteq \partial K\} = \{F \in \mathcal{F} : F \cap K \neq \emptyset, F \cap K^o = \emptyset\} \tag{3.2}$$

for every compact K. Therefore, the sets

$$\mathcal{D}_r := \partial\mathcal{H}(B_r),\ r > 0,\ \text{are pairwise disjoint.} \tag{3.3}$$

To see this, assume there exist $0 < r < s$ such that $\partial\mathcal{H}(B_r) \cap \partial\mathcal{H}(B_s) \neq \emptyset$. Consequently, there exists a closed set $F \in \partial\mathcal{H}(B_r) \cap \partial\mathcal{H}(B_s)$, which by (3.2) satisfies:

(i) $\emptyset \neq F \cap B_r \subseteq \partial B_r \subseteq \{x \in E : d(x, x_0) = r\}$ and (ii) $\emptyset \neq F \cap B_s \subseteq \partial B_s \subseteq \{x \in E : d(x, x_0) = s\}$. According to (i) we find some $x \in F \cap B_r$ such that $d(x, x_0) = r$. Since $B_r \subseteq B_s$ it is $x \in F \cap B_s$, which by (ii) means that $d(x, x_0) = s$. However, this is a contradiction to $r \neq s$.

From (3.3) we can conclude with a standard argument from measure theory that $R_+ := \{r > 0 : Q(\mathcal{D}_r) > 0\}$ is denumerable. Therefore, $R_0 := \{r > 0 : Q(\mathcal{D}_r) = 0\}$ lies dense in $[0, \infty)$. In particular, we can find a sequence $(r_i)_{i \ge 1} \subseteq R_0$ with $r_i \uparrow \infty, i \to \infty$. Define $K_i := B_{r_i}, i \ge 1$. Then $(K_i) \subseteq \mathcal{K}_T$ by construction, so that by (3.1) it follows that

$$\lim_{n\to\infty} \mathbb{P}_n(\xi_n \notin K_i) = 1 - T(K_i) \quad \forall\, i \ge 1. \tag{3.4}$$

Notice that $T(K_i) = \mathbb{P}(A_i)$, where $A_i = \{C \cap K_i \neq \emptyset\}$. It is easy to see that $A_i \uparrow \{C \neq \emptyset\}$, whence $T(K_i) \uparrow 1$, because $\mathbb{P}(C \neq \emptyset) = 1$ by our assumption on C. Thus by taking the limit $i \to \infty$ in (3.4) we obtain that $\lim_{i\to\infty} \lim_{n\to\infty} \mathbb{P}_n(\xi_n \notin K_i) = 0$. If $P_n := \mathbb{P}_n \circ \xi_n^{-1}, n \ge 1$, this implies that the sequence $(P_n)_{n\ge1}$ is tight. According to Prohorov's theorem, $(P_n)_{n\ge1}$ is relatively compact, so that there exists a subsequence $(P_{n_k})_{k\ge1}$ of $(P_n)_{n\ge1}$ and a probability measure P' on $(E, \mathcal{B})$ such that $P_{n_k} \stackrel{w}{\Rightarrow} P', k \to \infty$. By, e.g., the canonical construction we find a probability space $(\Omega', \mathcal{A}', \mathbb{P}')$ and a random variable $\xi' : (\Omega', \mathcal{A}') \to (E, \mathcal{B})$ with $P' = \mathbb{P}' \circ {\xi'}^{-1}$. It follows that $\xi_{n_k} \stackrel{\mathcal{L}}{\to} \xi', k \to \infty$. Let $i : (E, \mathcal{G}) \to (\mathcal{F}, \tau_{Fell})$ be the map defined by $i(x) := \{x\}, x \in E$. The map i is continuous on its entire domain, because for every $x \in E$ and every sequence $x_n \to x$ one easily shows that $K - \lim_{n\to\infty}\{x_n\} = \{x\}$ and the assertion follows from (1.1). Thus an application of the continuous mapping theorem yields that

$$\{\xi_{n_k}\} \stackrel{\mathcal{L}}{\to} \{\xi'\}, \quad k \to \infty. \tag{3.5}$$

On the other hand, our assumption entails that

$$\{\xi_{n_k}\} \stackrel{\mathcal{L}}{\to} C, \quad k \to \infty. \tag{3.6}$$

From (3.5) and (3.6) we can infer that $\mathbb{P} \circ C^{-1} = \mathbb{P}' \circ \{\xi'\}^{-1}$, because $(\mathcal{F}, \tau_{Fell})$ is compact and metrizable; hence it is a polish space and according to [8], p.344, the weak limit is unique in that case. Consequently, $T_C = \mathbb{P}' \circ {\xi'}^{-1}$ is a probability measure and the assertion follows from Proposition 2.7. □

4. Relating the different types of distributional convergence

Recall our aim to establish relationships between distributional convergence to a random closed set and distributional convergence in hyperspaces and, if possible, in the basic space, respectively. It turns out that these concepts of convergence all coincide as long as the limit set is a singleton.

Theorem 4.1 *The following statements are equivalent:*

(1) $\xi_n \stackrel{\mathcal{L}}{\to} \{\xi\}$ *for some random variable* $\xi : (\Omega, \mathcal{A}) \to (E, \mathcal{B})$.

(2) $\{\xi_n\} \stackrel{\mathcal{L}}{\to} C$ *in* $(\mathcal{F}, \tau_{Fell})$ *for some random closed set* $C \neq \emptyset$ *a.s.*

(3) $\xi_n \stackrel{\mathcal{L}}{\to} \xi$ *in* $(E, \mathcal{G})$ *for some random variable* $\xi : (\Omega, \mathcal{A}) \to (E, \mathcal{B})$.

In each case, $C = \{\xi\}$ *a.s.*

Proof $(1) \Rightarrow (2)$: By definition we have that

$$\limsup_{n\to\infty} \mathbb{P}(\xi_n \in F) \leq \mathbb{P}(\{\xi\} \cap F \neq \emptyset) = \mathbb{P}(\xi \in F) \quad \forall\, F \in \mathcal{F},$$

whence $\xi_n \stackrel{\mathcal{L}}{\to} \xi$ in $(E, \mathcal{G})$ by the (classical) portmanteau theorem. Another application of the portmanteau theorem yields

$$\lim_{n\to\infty} \mathbb{P}(\xi_n \in B) \quad \forall\, B \in \mathcal{B} \text{ with } \mathbb{P}(\xi \in \partial B) = 0. \tag{4.1}$$

Put $D := \{\xi\}$, which is a random closed set, and define

$$\begin{aligned} \mathcal{S}_D &:= \{B \in \mathcal{B} : \overline{B} \in \mathcal{K}, \mathbb{P}(D \cap \overline{B} \neq \emptyset) = \mathbb{P}(D \cap B^o \neq \emptyset)\} \\ &= \{B \in \mathcal{B} : \overline{B} \in \mathcal{K}, \mathbb{P}(\xi \in \overline{B}) = \mathbb{P}(\xi \in B^o)\} \subseteq \{B \in \mathcal{B} : \mathbb{P}(\xi \in \partial B) = 0)\}, \end{aligned}$$

where the inclusion follows from $\partial B = \overline{B} \setminus B^o$. Thus by (4.1) we obtain that

$$\mathbb{P}(\{\xi_n\} \cap B \neq \emptyset) = \mathbb{P}(\xi_n \in B) \to \mathbb{P}(\xi \in B) = \mathbb{P}(\{\xi\} \cap B \neq \emptyset) = \mathbb{P}(D \cap B \neq \emptyset)$$

for all $B \in \mathcal{S}_D$. Therefore, an application of Theorem 6.5 of [11] gives (2) with $C = D = \{\xi\}$.

$(2) \Rightarrow (3)$: By Proposition 3.1 there exists a random variable $\xi : (\Omega, \mathcal{A}) \to (E, \mathcal{B})$ such that $C = \{\xi\}$ a.s. Consequently, $\{\xi_n\} \stackrel{\mathcal{L}}{\to} \{\xi\}$ in $(\mathcal{F}, \tau_{Fell})$. Let $\mathcal{F}_s := \{\{x\} : x \in E\} \subseteq \mathcal{F}$ be the subspace of all singletons. We already know that the Kuratowski metric δ generates the Fell topology τ_{Fell}; confer (1.1). By Lemma 3.26 in [9] we can conclude that $\{\xi_n\} \stackrel{\mathcal{L}}{\to} \{\xi\}$ in $(\mathcal{F}_s, \delta)$. Let $j : (\mathcal{F}_s, \delta) \to (E, \mathcal{G})$ be the map defined by $j(\{x\}) := x, x \in E$. The map j is continuous on $\mathcal{F}_s$. Indeed, observe that $x_n \to x$ in E if (and only if) $K - \lim_{n\to\infty}\{x_n\} = \{x\}$. Thus with (1.1) and the continuous mapping theorem we arrive at (3).

$(3) \Rightarrow (1)$: A further application of the (classical) portmanteau theorem results in

$$\limsup_{n\to\infty} \mathbb{P}(\xi_n \in F) \leq \mathbb{P}(\xi \in F) = \mathbb{P}(\{\xi\} \cap F \neq \emptyset) \quad \forall\, F \in \mathcal{F},$$

which by definition means that $\xi_n \stackrel{\mathcal{L}}{\to} \{\xi\}$. □

Theorem 4.1 in particular tells us that $\xi_n \stackrel{\mathcal{L}}{\to} \{\xi\}$ is equivalent to $\{\xi_n\} \stackrel{\mathcal{L}}{\to} \{\xi\}$ in $(\mathcal{F}, \tau_{Fell})$. In other words, convergence in distribution of (ξ_n) to a limit set C, which is a singleton a.s. corresponds to classical weak convergence of the singletons $(\{\xi_n\})$ to C in the hyperspace $(\mathcal{F}, \tau_{Fell})$. Thus, the question arises, how can we extend this result to general C? The answer is: By changing the underlying hyperspace topology. Let τ_{uV} be the *upper Vietoris topology*, which is generated from a subbase $\{\mathcal{M}(F) : F \in \mathcal{F}\}$. This new topology is neither coarser nor finer than the Fell topology, but the pertaining σ-algebras $\mathcal{B}_{uV} := \sigma(\tau_{uV})$ and $\mathcal{B}_{Fell}$ coincide; confer [7]. Consequently, every random closed set in E is $\mathcal{B}_{uV}$-measurable. This is necessary when talking about weak convergence in $(\mathcal{F}, \tau_{uV})$.

Theorem 4.2 ***(Correspondence theorem)*** *The following statements are equivalent:*

(1) $\xi_n \stackrel{\mathcal{L}}{\to} C$.

(2) $\{\xi_n\} \stackrel{\mathcal{L}}{\to} C$ *in* $(\mathcal{F}, \tau_{uV})$.

Proof (1) $\Rightarrow$ (2) : It follows from the portmanteau theorem, confer Theorem 8.1 of Topsøe [22], that (2) is equivalent to

$$\limsup_{n\to\infty} \mathbb{P}_n(\{\xi_n\} \in \mathcal{C}) \le \mathbb{P}(C \in \mathcal{C}) \quad \text{for all } \tau_{uV}\text{-closed } \mathcal{C}. \tag{4.2}$$

Let $\mathcal{C}$ be an arbitrary τ_{uV}-closed set. Since $\{\mathcal{M}(F), F \in \mathcal{F}\}$ is a base for τ_{uV}, there exist $F_i \in \mathcal{F}, i \in I$, where I is some index set, such that $\mathcal{C} = (\bigcup_{i\in I} \mathcal{M}(F_i))^c = \bigcap_{i\in I} \mathcal{H}(F_i)$. Thus $\{\{\xi_n\} \in \mathcal{C}\} = \{\xi_n \in \bigcap_{i\in I} F_i\}$. Now $\bigcap_{i\in I} F_i \in \mathcal{F}$ and so (1.5) gives

$$\limsup_{n\to\infty} \mathbb{P}_n(\{\xi_n\} \in \mathcal{C}) \le \mathbb{P}(C \cap \bigcap_{i\in I} F_i \neq \emptyset) = \mathbb{P}(C \in \mathcal{H}(\cap_{i\in I} F_i)).$$

Since $\mathcal{H}(\cap_{i\in I} F_i) \subseteq \bigcap_{i\in I} \mathcal{H}(F_i) = \mathcal{C}$ we obtain (4.2).

(2) $\Rightarrow$ (1) : For every $F \in \mathcal{F}$ it is $\{\xi_n \in F\} = \{\{\xi_n\} \in \mathcal{H}(F)\}$, where $\mathcal{H}(F) = \mathcal{M}(F)^c$ is τ_{uV}-closed. Thus, (4.2) yields (1.5) as desired. □

There is a third hyperspace topology on $\mathcal{F}$, namely the *upper Fell topology* τ_{uF}, which is generated by the family $\{\mathcal{M}(K) : K \in \mathcal{K}\}$. By construction τ_{uF} is coarser than τ_{uV}. Vogel [24] shows (for $E = \mathbb{R}^d$, but her proof can easily be extended to E lcscH):

$$F_n \to F \text{ in } (\mathcal{F}, \tau_{uF}) \quad \Leftrightarrow \quad K - \limsup_{n\to\infty} F_n \subseteq F, \tag{4.3}$$

where $K - \limsup_{n\to\infty} F_n$ denotes the Painlevé–Kuratowski outer limit of (F_n). To set a simple example, recall from Analysis that a sequence (x_n) in a metric space E converges to a (closed) set $A \subseteq E$ (notation $x_n \to A$), if A contains all cluster points of the sequence. It follows from the definition of the outer limit that $x_n \to A$ if and only if $K - \limsup_{n\to\infty}\{x_n\} \subseteq A$ and therefore by (4.3) we obtain that $\{x_n\} \to A$ in $(\mathcal{F}, \tau_{uF})$ is equivalent to $x_n \to A$. The equivalent characterization (4.3) also shows that every superset of F is a limit too. Consequently, the limit is not necessarily unique and thus $(\mathcal{F}, \tau_{uF})$ is not a Hausdorff space.

In Ferger [7] it is shown that $\mathcal{B}_{uF} := \sigma(\tau_{uF}) = \mathcal{B}_{Fell}$, whence every random closed set in E is $\mathcal{B}_{uF}-$measurable as well. Vogel [24] refers to distributional convergence $C_n \stackrel{\mathcal{L}}{\to} C$ in $(\mathcal{F}, \tau_{uF})$ as *inner approximation in distribution* or *semiconvergence in distribution.* Since $\tau_{uF} \subseteq \tau_{uV}$, we have that

$$C_n \stackrel{\mathcal{L}}{\to} C \text{ in } (\mathcal{F}, \tau_{uV}) \quad \Rightarrow \quad C_n \stackrel{\mathcal{L}}{\to} C \text{ in } (\mathcal{F}, \tau_{uF}) \tag{4.4}$$

Thus, weak convergence in $(\mathcal{F}, \tau_{uV})$ is stronger than that in $(\mathcal{F}, \tau_{uF})$. However, the reverse in (4.4) is not true as will be shown below. The following result is the counterpart of our correspondence theorem. It sheds more light on the connection between inner approximation in distribution (of singletons) and our convergence concept. Recall that $\xi_n \stackrel{\mathcal{L}}{\to} C$ by definition means that

$$\limsup_{n\to\infty} \mathbb{P}_n(\xi_n \in F) \leq \mathbb{P}(C \cap F \neq \emptyset) \quad \forall\, F \in \mathcal{F}. \tag{4.5}$$

Theorem 4.3 ***(Second correspondence theorem)*** *The following statements are equivalent:*

(1) $\limsup_{n\to\infty} \mathbb{P}_n(\xi_n \in K) \leq \mathbb{P}(C \cap K \neq \emptyset) \quad \forall\, K \in \mathcal{K}.$

(2) $\{\xi_n\} \stackrel{\mathcal{L}}{\to} C$ *in* $(\mathcal{F}, \tau_{uF})$.

Proof In the proof of Theorem 4.2 replace τ_{uV} by τ_{uF} and $\mathcal{F}$ by $\mathcal{K}$. □

Notice that in Theorem 4.3 the class $\mathcal{K}$ of all compact sets plays exactly the same role as $\mathcal{F}$ does in Theorem 4.2. If E is noncompact, then $\mathcal{K}$ is strictly smaller than $\mathcal{F}$, whence the reverse conclusion in (4.4) should not hold. In fact, to see this let $\xi_n := n$ for every $n \in \mathbb{N}$. Then (1) in Theorem 4.3 is fulfilled with $C = \emptyset$ and therefore $\{\xi_n\} \stackrel{\mathcal{L}}{\to} \emptyset$ in $(\mathcal{F}, \tau_{uF})$, whereas by Theorem 4.2 in combination with Remark 1.5 $\{\xi_n\} \stackrel{\mathcal{L}}{\to} \emptyset$ in $(\mathcal{F}, \tau_{uV})$ is impossible.

Finally, from a statistician's point of view it is important to note that semiconvergence in distribution does not suffice for the construction of confidence regions as explained in Example 1.8. This is so, because the complement of the open set G there (for instance the open rectangle $G = (-r, r)^d$) in general is only closed, but not compact!

5. Weak convergence of Choquet capacities

Let T and $T_n, n \in \mathbb{N}$, be Choquet capacities. We extend our Definition 1.3 by saying that T_n *converges weakly to* T $(T_n \stackrel{w}{\to} T)$, if

$$\limsup_{n\to\infty} T_n(F) \leq T(F) \quad \forall\, F \in \mathcal{F}. \tag{5.1}$$

Let $C_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \to (\mathcal{F}, \mathcal{B}_{Fell}), n \in \mathbb{N}$, and $C : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathcal{F}, \mathcal{B}_{Fell})$ be random closed sets in E with pertaining capacity functional T_n and T, respectively. If $T_n \stackrel{w}{\to} T$, then we say that C_n *converges capacitively in distribution to* C and write $C_n \stackrel{c-\mathcal{L}}{\longrightarrow} C$. Notice that this is equivalent to

$$\limsup_{n\to\infty} \mathbb{P}_n(C_n \cap F \neq \emptyset) \leq \mathbb{P}(C \cap F \neq \emptyset) \quad \forall\, F \in \mathcal{F}. \tag{5.2}$$

Remark 1.5 and 1.6 are applicable in the general framework so that the limits T and C, respectively, are not uniquely determined. Note that one has to distinguish between $C_n \stackrel{c-\mathcal{L}}{\longrightarrow} C$ and $C_n \stackrel{\mathcal{L}}{\longrightarrow} C$ in $(\mathcal{F}, \tau_{Fell})$.

Our concept of capacitive convergence in distribution leads to innovative confidence regions in the situation of Example 1.8. Here we use the entire set of all $M-$estimators.

Example 5.1 *In Example 1.8 let A_n be the set of all infimizing points of the criterion function M_n, i.e. A_n consists of all M-estimators for the parameter θ. Consider*

$$C_n := \alpha_n(A_n - \theta) = \{\alpha_n(t-\theta) : t \in A_n\}.$$

By Lemma 2.2 in Ferger [7] C_n is equal to the (random closed) set of all infimizing points of the rescaled process X_n. If $X_n \stackrel{\mathcal{L}}{\longrightarrow} X$ in $(D(\mathbb{R}^d), s)$ and if the sequence (C_n) is stochastically bounded*, that is*

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}(C_n \not\subseteq [-k,k]^d) = 0,$$

then from Theorem 3.4 in Ferger [7] we can infer that $C_n \stackrel{c-\mathcal{L}}{\longrightarrow} C$. Put

$$U_n := A_n - \alpha_n^{-1} G_\alpha = \{t - \alpha_n^{-1}x : t \in A_n, x \in G_\alpha\}$$

and observe that $\{\theta \in U_n\} \supseteq \{\alpha_n(A_n-\theta) \subseteq G_\alpha\} = \{C_n \subseteq G_\alpha\} = \{C_n \cap G_\alpha^c = \emptyset\}$, where G_α^c is closed. Therefore, by (5.2) it follows that

$$\begin{aligned}\liminf_{n\to\infty} \mathbb{P}(\theta \in U_n) &\geq \liminf_{n\to\infty} \mathbb{P}(C_n \cap G_\alpha^c = \emptyset)) = 1 - \limsup_{n\to\infty} \mathbb{P}(C_n \cap G_\alpha^c \neq \emptyset) \\ &\geq 1 - \mathbb{P}(C \cap G_\alpha^c \neq \emptyset) = \mathbb{P}(C \cap G_\alpha^c = \emptyset) = \mathbb{P}(C \subseteq G_\alpha) \geq 1-\alpha.\end{aligned}$$

Thus, U_n is an asymptotic confidence region for θ at level $1-\alpha$. According to Theorem 3.4 in [7] these statements remain valid, if A_n is replaced by any nonempty random closed set A_n^* with $A_n^* \subseteq A_n$. The special choice $A_n^* := \{\hat{\theta}_n\}$ yields $U_n = R_n$.

Feng and Nguyen [6] introduce another concept of weak convergence for capacity functionals by replacing (5.1) through the requirement

$$\lim_{n\to\infty} \int f dT_n = \int f dT \text{ for all continuous and bounded functions } f. \tag{5.3}$$

In that case they say that T_n *converges in the Choquet weak sense to* T and denote this by $T_n \stackrel{C-W}{\Longrightarrow} T$. It turns out that this concept of convergence is strictly stronger than (5.1):

Proposition 5.2 *If $T_n \stackrel{C-W}{\Longrightarrow} T$ then $T_n \stackrel{w}{\to} T$. The reverse conclusion is not true.*

Proof For the first assertion, see Lemma 3.2 in [6]. As to the second one, assume that $T_n \stackrel{w}{\to} T$ entails $T_n \stackrel{C-W}{\Rightarrow} T$. From Lemma 3.8 of [6] we can conclude that $Q_n \stackrel{w}{\Rightarrow} Q$, where Q_n and Q are the probability measures pertaining to T_n and T (Choquet theorem). Since the classical weak limit is uniquely determined, confer [8], p. 344, it follows that T is uniquely determined, which in turn is a contradiction to Remark 1.5. □

Example 5.3 *For every $n \in \mathbb{N}$ let $C_n := B(\zeta_n, \rho_n) := \{x \in E : d(x, \zeta_n) \leq \rho_n\}$ be the closed ball with random center ζ_n and random radius ρ_n both defined on some common probability space $(\Omega, \mathcal{A}, \mathbb{P})$. If $(\rho_n, \zeta_n) \xrightarrow{\mathcal{L}} (\rho, \zeta)$ in $\mathbb{R}_+ \times E$, then*

$$B(\zeta_n, \rho_n) \xrightarrow{c-\mathcal{L}} B(\zeta, \rho). \tag{5.4}$$

Proof Let F be any nonempty closed subset of E. For every $r \geq 0$ the *r-neighborhood of F* is defined by

$$F^r := \{x \in E : d(x, F) \leq r\} = \{x \in E : B(x, r) \cap F \neq \emptyset\}, \tag{5.5}$$

where the second equality can easily be verified upon noticing that, since $F \neq \emptyset$ is closed, for each $x \in E$ there exists a $y \in F$ such that $d(x, F) = d(x, y)$. It follows that

$$\begin{aligned} \{B(\zeta_n, \rho_n) \cap F \neq \emptyset\} &= \bigcup_{r \geq 0} \{\rho_n = r, B(\zeta_n, r) \cap F \neq \emptyset\} \\ &= \bigcup_{r \geq 0} \{(\rho_n, \zeta_n) \in \{r\} \times F^r\} \quad \text{by (5.5)} \\ &= \{(\rho_n, \zeta_n) \in \bigcup_{r \geq 0} (\{r\} \times F^r)\}. \end{aligned} \tag{5.6}$$

The set $A := A(F) := \bigcup_{r \geq 0} (\{r\} \times F^r)$ occurring in (5.6) is closed. To see this consider a sequence (r_n, z_n) in A with $(r_n, z_n) \to (r, z)$. For every $n \in \mathbb{N}$ we find some $s_n \geq 0$ such that $r_n = s_n$ and $z_n \in F^{s_n}$. Consequently, $d(z_n, F) \leq s_n = r_n$ for all $n \in \mathbb{N}$. Recall that the map $x \mapsto d(x, F)$ is continuous, so that by taking the limit $n \to \infty$ we arrive at $d(z, F) \leq r$, because $z_n \to z$ and $r_n \to r \geq 0$. Thus by (5.5) the limit (r, z) lies in A as desired.

Now equality (5.6) ensures that

$$\limsup_{n \to \infty} \mathbb{P}(B(\zeta_n, \rho_n) \cap F \neq \emptyset) = \limsup_{n \to \infty} \mathbb{P}((\rho_n, \zeta_n) \in A(F)) \leq \mathbb{P}((\rho, \zeta) \in A(F)),$$

where the inequality holds by the (classical) portmanteau theorem. Clearly, in the derivation of (5.6) we are free to replace $B(\zeta_n, \rho_n)$ by $B(\zeta, \rho)$ and therefore $\mathbb{P}((\rho, \zeta) \in A(F)) = \mathbb{P}(B(\zeta, \rho) \cap F \neq \emptyset)$, which according to (5.2) gives the assertion. □

Our notion of capacitive convergence in distribution immediately yields a huge class of examples for convergence in distribution of random points to a random set.

Lemma 5.4 *Let C and $C_n, n \in \mathbb{N}$, be random closed sets in E and let ξ_n be random variables with $\xi_n \in C_n$ a.s. for every $n \in \mathbb{N}$. If $C_n \xrightarrow{c-\mathcal{L}} C$, then $\xi_n \xrightarrow{\mathcal{L}} C$.*

Proof The assertion follows at once from the inequality $\mathbb{P}_n(\xi_n \in F) \leq \mathbb{P}_n(C_n \cap F \neq \emptyset)$ for all nonempty closed $F \subseteq E$ upon noticing (1.5) and (5.2). □

As to the existence of the ξ_n occurring in the above lemma we refer to the *measurable selection theorem*; confer Theorem 8.1.3 in [4]. It states that for every nonempty random closed set $D : (\Omega, \mathcal{A}) \to (\mathcal{F}, \mathcal{B}_{Fell})$ there exists a measurable map $\xi : (\Omega, \mathcal{A}) \to (E, \mathcal{B})$ such that $\xi(\omega) \in D(\omega)$ for all $\omega \in \Omega$. Here it is required that

(E, d) is a complete, separable metric space. However, it is well known that every E lcscH has this property; confer, e.g., [18], p. 260.

Finally, we consider the convergence in the Choquet weak sense in the special case that the T_n's are actually probability measures. It turns out that in this situation it is the same as classical weak convergence.

Proposition 5.5 *Let $P_n, \in \mathbb{N}$, be probability measures. Then $P_n \stackrel{C-W}{\Longrightarrow} T$ if and only if $P_n \stackrel{w}{\Rightarrow} T$. In either case T is a probability measure.*

Proof For the proof of the direct half $\Rightarrow$ recall that from Lemma 3.8 of [6] we can conclude that

$$Q_n \stackrel{w}{\to} Q, \tag{5.7}$$

where Q_n and Q are the probability measures on $(\mathcal{F}, \mathcal{B}_{Fell})$, which by Choquet's theorem are uniquely determined through P_n and T, respectively. Let C_n and C be random closed sets on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that C_n and C have distribution Q_n and Q, respectively, i.e. $\mathbb{P} \circ C_n^{-1} = Q_n$ and $\mathbb{P} \circ C^{-1} = Q$. Then by construction the capacity functional T_{C_n} of C_n is equal to P_n, because $T_{C_n}(B) = \mathbb{P}(C_n \cap B \neq \emptyset) = \mathbb{P} \circ C_n^{-1}(\mathcal{H}(B)) = Q_n(\mathcal{H}(B)) = P_n(B)$ for all $B \in \mathcal{B}$. Here notice that all involved sets are measurable by Theorem 1.2 of Matheron [10]. Thus, T_{C_n} is a probability measure and we may apply Proposition 2.7. It guarantees the existence of random variables ξ_n on $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $(E, \mathcal{B})$ such that $C_n = \{\xi_n\}$ a.s. Therefore from (5.7) we can deduce that

$$\{\xi_n\} \stackrel{\mathcal{L}}{\to} C \text{ in } (\mathcal{F}, \tau_{Fell}). \tag{5.8}$$

An application of Proposition 5.2 with $T_n = P_n$ yields that $P_n \stackrel{w}{\to} T$, whence $T(E) = 1$ by Remark 1.5. Repeating our arguments above with C instead of C_n shows that $T_C = T$ and we now know that $C \neq \emptyset$ a.s. By (5.8) and Proposition 3.1 there exists a random variable $\xi : (\Omega, \mathcal{A}) \to (E, \mathcal{B})$ such that $C = \{\xi\}$ a.s. Consequently $T = T_C$ is a probability measure and the defining relation (5.3) (with $T_n = P_n$) now coincides with $P_n \stackrel{w}{\Rightarrow} T$ as desired. This also yields the converse half $\Leftarrow$ because here T is a probability measure by definition. □

To sum up we have on the one hand $P_n \stackrel{C-W}{\Rightarrow} T$ iff $P_n \stackrel{w}{\Rightarrow} T$, but on the other hand $P_n \stackrel{w}{\to} T \not\Rightarrow P_n \stackrel{w}{\Rightarrow} T$, for otherwise we obtain a contradiction to the fact that T is not a probability measure in general; confer Proposition 2.7 in combination with Example 5.3 and Lemma 5.4.

Remark 5.6 *Our notion of weak convergence $T_n \stackrel{w}{\to} T$ corresponds to a topology on the set $\mathcal{T}$ of all Choquet capacities. The principle of its construction goes back to Topsøe [22], who in turn refers to previous research by Alexandrov [1–3]. Let $\mathcal{O}$ be the topology on $\mathcal{T}$ generated by the family $\{\{T \in \mathcal{T} : T(F) < x\}, F \in \mathcal{F}, x \in \mathbb{R}\}$. Thus, $\mathcal{O}$ is the coarsest topology on $\mathcal{T}$ such that the evaluation maps $e_F : \mathcal{T} \to \mathbb{R}$ defined by $e_F(T) := T(F), T \in \mathcal{T}$, are upper semicontinuous for every $F \in \mathcal{F}$. This construction ensures that*

$$T_n \to T \ \ \textit{in } (\mathcal{T}, \mathcal{O}) \quad \Leftrightarrow \quad \limsup_{n \to \infty} T_n(F) \leq T(F) \quad \forall\, F \in \mathcal{F}. \tag{5.9}$$

The equivalence (5.9) enables a direct comparison with the *narrow topology* $\mathcal{O}_n$ of O'Brien [14]. This topology is generated by the family

$$\{\{T \in \mathcal{T} : T(F) < x\}, F \in \mathcal{F}, x \in \mathbb{R}\} \cup \{\{T \in \mathcal{T} : T(G) > x\}, G \in \mathcal{G}, x \in \mathbb{R}\}.$$

Analogous to $\mathcal{O}$ it follows that

$$T_n \to T \text{ in } (\mathcal{T}, \mathcal{O}_n) \quad \Leftrightarrow$$

$$\limsup_{n\to\infty} T_n(F) \le T(F) \quad \forall\, F \in \mathcal{F}, \tag{5.10}$$

$$\liminf_{n\to\infty} T_n(G) \ge T(G) \quad \forall\, G \in \mathcal{G}. \tag{5.11}$$

Therefore, firstly, by definition (5.1) and relation (5.9) our concept of weak convergence $T_n \overset{w}{\to} T$ coincides with convergence in $(\mathcal{T}, \mathcal{O})$. Secondly, compared to *narrow convergence* $T_n \to T$ in $(\mathcal{T}, \mathcal{O}_n)$, it incorporates only the "upper part" (5.10), but not the "lower part" (5.11). If in the definition of the narrow topology the role of $\mathcal{F}$ is taken over by $\mathcal{K}$, one obtains the *vague topology* and convergence in that topology is equivalent to (5.10) with $\mathcal{K}$ instead of $\mathcal{F}$ and (5.11). O'Brien [14] and O'Brien and Watson [15] actually consider set $\mathcal{C}$ of *capacities*, where a capacity is a set function similar to but not the same as a Choquet capacity. They find relations between the narrow and vague topology on $\mathcal{C}$ including a necessary and sufficient condition for relative compactness in the narrow topology; confer also Vervaat [23] for previous work on that topic.

Concluding remark

In the theory of classical weak convergence there are various methods to derive $P_n \overset{w}{\Rightarrow} T$. For example, if $E = \mathbb{R}^d$ it suffices to show pointwise convergence of the pertaining distribution functions or characteristic functions, respectively. Further, if E is equal to the function space $C[0,1]$ or $D[0,1]$ then convergence of the finite dimensional distributions plus tightness ensure weak convergence; confer Billingsley [5], Theorem 8.1 or Theorem 15.1, respectively. It turns out that a comparable result holds for capacitive convergence in distribution $C_n \overset{c-\mathcal{L}}{\longrightarrow} C$. In fact, assume that

$$(1) \quad C_n \overset{\mathcal{L}}{\to} C \text{ in } (\mathcal{F}, \tau_{uF}) \quad \text{and} \quad (2) \quad \forall \epsilon > 0\, \exists K \in \mathcal{K} : \limsup_{n\to\infty} \mathbb{P}(C_n \not\subseteq K) \le \epsilon.$$

Then we prove (unpublished future work) that $C_n \overset{\mathcal{L}}{\to} C$ in $(\mathcal{F}, \tau_{uV})$, which in turn yields $C_n \overset{c-\mathcal{L}}{\longrightarrow} C$. Notice that in the case of singletons $C_n = \{\xi_n\}$ condition (2) exactly means that the sequence (ξ_n) is tight. Occasionally, the random closed sets are given as $C_n = h(Z_n), n \in \mathbb{N}$, and $C = h(Z)$, where $Z_n \overset{\mathcal{L}}{\to} Z$ in some metric space (S, σ). Then by using the characterization (4.3) one shows that the map $h : (S, \sigma) \to (\mathcal{F}, \tau_{uF})$ is continuous and condition (1) follows immediately.

References

[1] Alexandrov AD. Additive set functions in abstract spaces. Mat Sb 1940; 8: 307-348.

[2] Alexandrov AD. Additive set functions in abstract spaces. Mat Sb 1941; 9: 563-628.

[3] Alexandrov AD. Additive set functions in abstract spaces. Mat Sb 1943; 13: 169-238.

[4] Aubin JP, Frankowska H. Set-valued Analysis. Boston, MA, USA: Birkhäuser, 1990.

[5] Billingsley P. Convergence of Probability Measures, Second Edition. New York, NY, USA: John Wiley & Sons, 1999.

[6] Feng D, Nguyen HT. Choquet weak convergence of capacity functionals of random sets. Inform Sci 2007; 177: 3239-3250.

[7] Ferger D. Arginf-sets of multivariate cadlag processes and their convergence in hyperspace topologies. Theory Stoch Process 2015; 20: 13-41.

[8] Gaenssler P, Stute W. Wahrscheinlichkeitstheorie. New York, NY, USA: Springer-Verlag, 1977.

[9] Kallenberg O. Foundations of Modern Probability. New York, NY, USA: Springer-Verlag, 1997.

[10] Matheron G. Random Sets and Integral Geometry. New York, NY, USA: John Wiley & Sons, 1975.

[11] Molchanov I. Theory of Random Sets. London, UK: Springer-Verlag, 2005.

[12] Nguyen HT. An Introduction to Random Sets. Boca Raton, FL, USA: Chapman & Hall/CRC, 2006.

[13] Norberg T. Convergence and existence of random set distributions. Ann Probab 1984; 12: 726-732.

[14] O'Brien GL. Sequences of capacities, with connections to large-deviation theory. J Theoretical Probab 1996; 9: 19-35.

[15] O'Brien GL, Watson S. Relative compactness for capacities, measures, upper semicontinuous functions and closed sets. J Theoretical Probab 1998; 11: 577-588.

[16] Pfanzagl J. Mathematical Statistics. Essays on History and Methodology. Berlin, Germany: Springer-Verlag, 2017.

[17] Pflug G. Asymptotic dominance and confidence for solutions of stochastic programs. Czechoslovak J Oper Res 1992; 1: 21-30.

[18] Ramsay AB. Polish Groupoids. In: Louveau A, Foreman M, Kechris AS, Weiss B, editors. Descriptive Set-Theory and Dynamical Systems. London Mathematical Society Lecture Notes Series, 277. Cambridge University Press, 2000, pp. 259-272.

[19] Rockafellar RT, Wets RJB. Variational Analysis. Berlin, Germany: Springer-Verlag, 1998.

[20] Salinetti G, Wets RJB. On the convergence in distribution of measurable multi-functions (random sets), normal integrands, stochastic processes and stochastic infima. Math Oper Res 1986; 11: 385-419.

[21] Salinetti G, Wets RJB. Weak convergence of probability measures revisited. Working Paper WP-87-030, International Institute for Applied System Analysis, Laxenburg, Austria, 1987.

[22] Topsøe F. Topology and Measure. Lecture Notes in Mathematics Vol. 133. Berlin, Heidelberg, New York: Springer-Verlag, 1970.

[23] Vervaat W. Narrow and vague convergence of set functions. Statist Probab Lett 1988; 6: 295-298.

[24] Vogel S. Semiconvergence in distribution of random closed sets with application to random optimization problems. Ann Oper Res 2006; 142: 269-282.

[25] Wei G, Wang Y, Nguyen HT, Beken DE. On the upper semicontinuity of Choquet capacities. Int J Approx Reason 2010; 51: 429-440.