A CRITERION FOR WEAK CONVERGENCE
OF MEASURES WITH AN APPLICATION
TO CONVERGENCE OF MEASURES ON $D[0, 1]$

FLEMMING TOPSØE

Introduction.

The usual way to establish weak convergence of a sequence of probability measures on $C[0, 1]$ is to prove that all the finite dimensional distributions converge weakly to the "right" limits and that the sequence is tight; then a theorem, which is quite easy to establish, tells us that we have weak convergence. If we turn our attention to the Skorohod space $D[0, 1]$ we find that the analogous theorem is much harder to obtain. Recently, P. Billingsley has obtained a suitable result published in [1, p. 124]. My aim has been to find a general theorem, valid in any Polish space, which implies the desired result in $D[0, 1]$.

All measures below are supposed to be defined on the Borel $\sigma$-field.

A reasonable problem inspired by the concrete question about $D[0, 1]$ is the following: Let $X$ be a Polish space (that is, separable and metrizable in such a way that it becomes complete), and let $\mathcal{A}$ be a field of Borel sets generating the entire Borel $\sigma$-field $\mathcal{B}(X)$. If $P, (P_n)_{n \geq 1}$ are probability measures on $X$ such that $(P_n)$ is tight and such that $P_n A \rightarrow PA$ for all sets $A$ in $\mathcal{A}$, is it then true, or under what additional assumptions is it true, that $P_n$ converges weakly to $P$?

Equivalently, we could ask if, under the just mentioned hypotheses on $X$ and $\mathcal{A}$, the facts that $P_n \rightarrow Q$ for some probability measure $Q$ and $P_n A \rightarrow PA$ for all $A$ in $\mathcal{A}$, imply $P=Q$. Here, as below, the sign $\rightarrow$ is used to indicate weak convergence of probability measures as well as ordinary convergence of real numbers.

Since $P_n \rightarrow Q$ and $P_n A \rightarrow PA$ imply

$$Q A \leq \liminf P_n A \leq PA \leq \limsup P_n A \leq Q A,$$

a problem related to the problem above arises: Let again $X$ be Polish and $\mathcal{A}$ a field generating $\mathcal{B}(X)$. If $P$ and $Q$ are two probability measures

Received November 13, 1967; in revised form January 28, 1969.

Math. Scand. 25 — 7
such that $Q\hat{A} \leq PA \leq Q\overline{A}$ holds for all $A$ in $\mathscr{A}$, is it then true, or under what additional assumptions is it true, that $P \equiv Q$?

Without additional assumptions the answers to both problems are negative. Indeed, O. Bjørnson observed that there is a counterexample in which $P$ and $Q$ are as simple as possible, namely point masses.

1. A convergence criterion.

A class $\mathscr{A}$ of subsets of the topological space $X$ is said to separate points $T_2$ if, for every pair of distinct points $x$ and $y$ in $X$ there exists a set $A$ in $\mathscr{A}$ such that $x \in \hat{A}$ and $y \notin \overline{A}$.

**Theorem 1.** Let $\mathscr{A}$ be a lattice of subsets of the Polish space $X$ and suppose that $\mathscr{A}$ separates points $T_2$.

(i) If $P$ and $Q$ are probability measures on $X$ such that $Q\hat{A} \leq PA \leq Q\overline{A}$ for all $A \in \mathscr{A}$, then $P \equiv Q$.

(ii) If $P$, $(P_n)_{n \geq 1}$ are probability measures on $X$ such that $(P_n)_{n \geq 1}$ is tight and such that $\limsup P_n\hat{A} \leq PA \leq \limsup P_n\overline{A}$ for all $A \in \mathscr{A}$, then $P_n \rightarrow P$.

Part (ii) is the hoped for convergence criterion.

**Proof.** (i) If $K_1$ and $K_2$ are disjoint compact subsets of $X$, we can find $A \in \mathscr{A}$ such that $K_1 \subset \hat{A}$ and $\overline{A} \cap K_2 = \emptyset$. It follows that $QK_1 \leq 1 - PK_2$. Employing the tightness of $P$ and $Q$, one deduces from this that $P \equiv Q$.

(ii) Let $(P_{n_k})_{k \geq 1}$ be a convergent subsequence of $(P_n)_{n \geq 1}$, say $P_{n_k} \rightarrow Q$. Then

$$Q\hat{A} \leq \liminf_{k \to \infty} P_{n_k}\hat{A} \leq \limsup_{n \to \infty} P_n\hat{A} \leq PA$$

for all $A \in \mathscr{A}$. By (i), $Q \equiv P$ follows.

Part (i) in the theorem can of course be considered as a special case of part (ii). It is easy to see, by means of a simple counterexample, that we can not drop the assumption that $\mathscr{A}$ be a lattice. We have been unable to decide whether one can relax this assumption assuming only that $\mathscr{A}$ is closed under finite intersections. The condition in Theorem 1 that $\mathscr{A}$ separates points $T_2$ is necessary, as is easily seen.

It is not difficult to extend the convergence criterion from measures on Polish spaces to tight measures (Radon measures) on arbitrary Hausdorff spaces; also, one may consider a lattice of functions in stead of a lattice of sets. It is our intention to publish elsewhere some results on weak convergence of tight measures on arbitrary Hausdorff spaces.
If $\mathcal{A}$ in the convergence criterion is a subclass of $\mathcal{B}(X)$, then $\mathcal{A}$ generates $\mathcal{B}(X)$; to see this, note that $\mathcal{A}$ contains a countable subclass separating points and apply Theorem 3.3 of [3]. This argument was pointed out to us by E. T. Kehlet.

2. An application to measures on the Skorohod space $D[0,1]$.

$D[0,1]$ consists of those real-valued functions on $[0,1]$ which are continuous from the right for $0 \leq t < 1$ and have limits from the left for $0 < t \leq 1$. The distance $d(x,y)$ between two functions in $D[0,1]$ is the infimum of those $\epsilon \geq 0$ for which there exists an increasing homeomorphism $\lambda$ from $[0,1]$ onto itself such that $\|\lambda - i\| < \epsilon$ and $\|x - y \circ \lambda\| < \epsilon$; here $i$ denotes the identity map and $\|\cdot\|$ the uniform norm. The metric space $D[0,1]$ is known to be Polish. For a finite (ordered) subset $\mathbf{t} = \{t_1, \ldots, t_k\}$ of $[0,1]$ we denote by $\pi_{\mathbf{t}}$ the projection from $D[0,1]$ onto $\mathbb{R}^k$. The projections are all measurable. For a probability measure $P$ on $D[0,1]$ we denote by $T_P$ the set of $t \in [0,1]$ such that $\pi_t$ is continuous a.e. $P$. There are at most countably many points in $[0,1] \setminus T_P$; these points are called fixed points of discontinuity for $P$.

The purpose of this section is to prove the following result:

**Theorem 2.** Let $(P_n)_{n \geq 1}$ be a tight sequence of probability measures on $D[0,1]$ and let $T$ be a dense subset of $[0,1]$ containing the point 1. Suppose that for each finite subset $\mathbf{t}$ of $T$ there is a probability measure $P_\mathbf{t}$ on the proper Euclidean space such that $P_n\pi_{\mathbf{t}}^{-1}$ converges weakly to $P_\mathbf{t}$. Then the sequence $(P_n)_{n \geq 1}$ converges weakly. Furthermore, the limit measure $P$ can be identified by the formula

$$P_{\pi_{\mathbf{t}}^{-1}} = P_{\mathbf{t}^+},$$

which holds for any finite subset $\mathbf{t}$ of $[0,1]$.

The formula $P_{\pi_{\mathbf{t}}^{-1}} = P_{\mathbf{t}^+}$ means, first of all, that the limit from the right, in the sense of weak convergence, exists at $\mathbf{t}$ and, secondly, that this limit is the finite dimensional distribution of $P$ at $\mathbf{t}$. In more detail, what we claim is the following: Let $\mathbf{t} = \{t_1, \ldots, t_k\}$ be any finite subset of $[0,1]$; let, for each $\nu = 1, 2, \ldots$,

$$t_\nu = \{t_{\nu 1}, \ldots, t_{\nu k}\}$$

be a finite subset of $T$ with $t_{\nu i} > t_i$ for each $i = 1, \ldots, k$ (unless $t_i = 1$ in which case we demand $t_{\nu i} = 1$); suppose further that

$$t_{\nu i} \downarrow t_i \quad \text{as} \quad \nu \to \infty \quad \text{for each} \quad i = 1, \ldots, k$$
(shortly: \( t_v \downarrow t \)). Then we claim that the measures \( P_{t_v} \) converge weakly in \( R^k \) to \( P_{t_0} \) as \( v \to \infty \).

In Theorem 2, the limit measure is not supposed to be known in advance; thus the result can be used to construct various measures.

A special case of Theorem 2 (with known limit measure and special \( T \)) has been established by Billingsley (Theorem 15.1 of [1]). Our method is completely different from Billingsley’s.

To prove that \( (P_n) \) converges weakly in Theorem 2 we need some simple lemmas.

**Lemma 1.** For any finite subset \( t = \{t_1, \ldots, t_k \} \) of \([0, 1]\) and any subset \( E \) of \( R^k \) we have

\[
\pi_t^{-1}(\overline{E}) \subset \pi_t^{-1}(\overline{E}) \quad \text{and} \quad \pi_t^{-1}(\overline{E}) \supset \pi_t^{-1}(\overline{E}).
\]

In other words, all projections are open mappings.

**Proof.** Assume that \( x \in \pi_t^{-1}(\overline{E}) \). Then \((x(t_1), \ldots, x(t_k)) \) lies in \( \overline{E} \). Thus, to any \( \delta > 0 \) we can find real numbers \( r_1, \ldots, r_k \) with \( |r_i| < \delta \) for all \( i \), and such that \((x(t_1) + r_1, \ldots, x(t_k) + r_k) \) lies in \( E \). Clearly then, there exists a function \( y \) in \( D \) with \( y(t_i) = x(t_i) + r_i \) for all \( i \), and such that the uniform distance from \( x \) to \( y \) is less than \( \delta \). Then the Skorohod distance from \( x \) to \( y \) is also less than \( \delta \). Since \( y \in \pi_t^{-1}(E) \) and \( \delta \) is arbitrary the first inclusion follows. The second inclusion is a consequence of the first.

**Lemma 2.** Let \( s \) be a point in \([0, 1]\) and \( E \) a subset of \( R \). Denote by \( A \) the cylinder set \( \pi_s^{-1}(E) = \{x \in D: x(s) \in E\} \). If \( 0 < s < 1 \), then we have

\[
\overline{A} = \{x \in D: x(s-) \in \overline{E} \text{ or } x(s) \in \overline{E}\},
\]

\[
\hat{A} = \{x \in D: x(s-) \in \hat{E} \text{ and } x(s) \in \hat{E}\}.
\]

If \( s \) is either 0 or 1, then

\[
\overline{A} = \{x \in D: x(s) \in \overline{E}\},
\]

\[
\hat{A} = \{x \in D: x(s) \in \hat{E}\}.
\]

**Proof.** The case \( s = 0 \) or 1 is easily treated. Now assume that \( s \in (0, 1) \). If \( x \in \overline{A} \) then there exists a sequence \((x_n)\) of functions in \( A \) and a sequence \((\lambda_n)\) of increasing homeomorphisms of \([0, 1]\) onto \([0, 1]\) such that

\[
||\lambda_n - \delta|| \to 0 \quad \text{and} \quad ||x_n - x \circ \lambda_n|| \to 0.
\]
Put $s_n = \lambda_n(s)$. We may assume that either $s_n < s$ holds for all $n$ or else $s_n \geq s$ holds for all $n$. If the first alternative takes place, then $x(s_n) \to x(s-)$ and one finds that $x_n(s) \to x(s-)$ so that $x(s-) \in \overline{E}$. The second alternative leads to $x(s) \in \overline{E}$.

To prove the reverse inclusion, assume first that $x(s) \in \overline{E}$; it follows from Lemma 2 that $x \in \overline{A}$. Now assume that $x(s-) \in \overline{E}$. By moving the function a little to the right and then adding a small constant function, one arrives at a function in $A$. Intuitively, it is thus clear that $x \in \overline{A}$. It is left to the reader to make this argument rigorous.

**Lemma 3.** Let $T$ be a dense subset of $[0, 1]$ containing the point 1. Denote by $\mathcal{A}$ the class of cylinder sets based on time-points in $T$, that is, $\mathcal{A}$ is the class of all sets $\pi_t^{-1}(E)$, where $t$ ranges over all finite subsets of $T$ and $E$ ranges over all Borel subsets of the proper Euclidean spaces. Then $\mathcal{A}$ is a field separating points $T_2$.

**Proof.** Clearly, $\mathcal{A}$ is a field. We shall prove that $\mathcal{A}$ separates points $T_2$. Let $x$ and $y$ be distinct functions in $D$. This means that, for some $s$ in $[0, 1]$, $x(s)$ is distinct from $y(s)$. We shall assume that $x(s) < y(s)$ holds.

If $s = 1$, then the set $A = \pi_1^{-1}((\infty, m))$, where $m$ is the midpoint of $[x(1), y(1)]$ lies in $\mathcal{A}$ and by Lemma 2 we also find that $x \in \overline{A}$ and $y \notin \overline{A}$.

If $s < 1$, we argue as follows. First choose three real numbers $m_1$, $m$ and $m_2$ such that $x(s) < m_1 < m < m_2 < y(s)$. Then, by the right continuity, we can find a positive $\delta$ with $s + \delta < 1$ such that $x(t) \leq m_1$ and $y(t) \geq m_2$ hold for any $t$ in $(s, s + \delta)$. Since $T$ is dense in $[0, 1]$, we can find a $t$ from $T$ in $(s, s + \delta)$. Now put $A = \pi_t^{-1}((\infty, m))$. $A$ lies in $\mathcal{A}$ and by Lemma 1 we also find that $x \in \overline{A}$ and $y \notin \overline{A}$. Thus $\mathcal{A}$ separates points $T_2$.

**Proof of the weak convergence in Theorem 2.** We begin by remarking that the family of measures $(P_t)$ where $t$ ranges over all finite subsets of $T$ is consistent. Now, let $Q_1$ be any limit measure for $(P_n)$, say $P_n \to Q_1$. Consider a finite subset $t = \{t_1, \ldots, t_k\}$ of $T$ and a $k$-dimensional Borel set $E$. Then

$$Q_1(\pi_t^{-1}E) \leq \liminf P_n(\pi_t^{-1}E) \leq \liminf P_n(\pi_t^{-1}\overline{E}) \leq \limsup P_n(\pi_t^{-1}\overline{E}) \leq P_t(\overline{E}),$$

that is, we have

$$Q_1(\pi_t^{-1}E) \leq P_t(\overline{E}).$$

(2)
If $Q_2$ is another limit measure for $(P_n)$ then we find in an analogous manner

\begin{equation}
\tag{3}
P_t(\bar{E}) \leq Q_2(\pi_t^{-1}E).
\end{equation}

Consider the class $\mathcal{A}$ of those subsets $A$ of $D[0,1]$ for which there exist a finite subset $t=\{t_1, \ldots, t_k\}$ of $T$ and a $k$-dimensional Borel set $E$ such that $A=\pi_t^{-1}E$ and $P_t(\partial E)=0$. Here $\partial E$ is the boundary in $\mathbb{R}^k$ of $E$. In checking which sets $A$ belong to $\mathcal{A}$ it does not matter which representation we use for $A$. Clearly, $\mathcal{A}$ is a field. The proof of Lemma 3 shows that $\mathcal{A}$ separates points $T_2$. By (2) and (3), the inequality $Q_1(\bar{A}) \leq Q_2(\bar{A})$ holds for any set $A$ in $\mathcal{A}$. By Theorem 1, $Q_1$ and $Q_2$ are identical.

We shall now prove (1). In case the limit measure $P$ has only finitely many fixed points of discontinuity, this formula obviously holds. In the general case we use a rather elaborate argumentation and we begin with a lemma.

\textbf{Lemma 4.} Let $x$ be a function in $D[0,1]$ and $t$ a point in $[0,1]$ such that $|x(t)-x(t^-)|<\varepsilon$. Then there exists a positive $\delta$ and a positive $h$ such that any element in the open sphere $S(x,\delta)$ with center $x$ and radius $\delta$ oscillates by less than $\varepsilon$ in the interval $[t-h, t+h]$.

This result is obvious.

\textbf{Proof of (1) of Theorem 2.} To ease the notation, we shall assume that the finite subset $t$ that we consider in fact only contains one point $t_0$. We may also assume that $t_0<1$. What we have given is a sequence $(s_k)$ of points in $T$ with $s_k > t_0$ for all $k$ and $s_k \downarrow t_0$ as $k \to \infty$. We want to prove that $P_{s_k} \to P_{\pi_{t_0}^{-1}}$. Fix, for some time, two positive numbers $\varepsilon$ and $\eta$. Since there are at most finitely many points $t$ for which

$$P\{x : |x(t)-x(t^-)| \geq \varepsilon\} \geq \eta$$

holds, we can find an integer $k_0$ such that

$$PA_k > 1-\eta$$

for all $k \geq k_0$,

where we have put

$$A_k = \{x : |x(s_k)-x(s_k^-)| < \varepsilon\}.$$ 

Choose a compact set $K_k$ with $K_k \subset A_k$ and $PK_k > 1-\eta$. To any function $x$ in $K_k$ we choose two positive numbers $\delta_x$ and $h_x$ such that the oscillation of any function from $S(x,\delta_x)$ over the interval $[s_k-h_x, s_k+h_x]$ is less than $\varepsilon$. Finitely many of the spheres, say $S(x_1, \delta_{x_1}), \ldots, S(x_r, \delta_{x_r})$, cover $K_k$. Put
\[ G_k = S(x_1, \delta_{x_1}) \cup \ldots \cup S(x_r, \delta_{x_r}) \quad \text{and} \quad h_k = \min \{h_{x_1}, \ldots, h_{x_r}\}. \]

Then \( G_k \) is open, \( G_k \) contains \( K_k \) and any function in \( G_k \) oscillates by less than \( \epsilon \) in the interval \([s_k - h_k, s_k + h_k]\). Now choose a point \( t_k \) in the interval \([s_k - h_k, s_k + h_k]\) such that \( t_k > t_0 \), \( |t_k - s_k| < 1/k \) and such that \( t_k \in T_P \). Let \( E \) be any Borel subset of \( \mathbb{R} \). Then, for \( k \geq k_0 \) the following inclusion holds:

\[ \{x : x(s_k) \in E\} \subseteq \{x : x(t_k) \in E^c\} \cup G_k^c. \]

Here \( E^c \) denotes the \( \epsilon \)-neighbourhood of \( E \), that is, the set of points within distance less than \( \epsilon \) from \( E \), and \( ^c \) indicates complementation. Considering a \( k \geq k_0 \) and using the weak convergence \( P_n \pi_{t_k}^{-1} \rightarrow P \pi_{t_k}^{-1} \) (which follows since \( t_k \in T_P \)), we now find that

\[
P_{s_k}^c(E) \leq \liminf_{n \to \infty} P_n \pi_{s_k}^{-1}(E) \leq \limsup_{n \to \infty} P_n \pi_{t_k}^{-1}(E^c) + \limsup_{n \to \infty} P_n(G_k^c) \leq P \pi_{t_k}^{-1}(E^c) + \eta.
\]

Since \( t_k \downarrow t_0 \), the projections \( \pi_{t_k} \) converge everywhere to \( \pi_{t_0} \) as \( k \to \infty \), hence

\[ P \pi_{t_k}^{-1} \rightarrow P \pi_{t_0}^{-1} \quad \text{as} \quad k \to \infty. \]

From the above inequality we thus find

\[
\limsup_{k \to \infty} P_{s_k}^c(E) \leq \limsup_{k \to \infty} P \pi_{t_k}^{-1}(E^c) + \eta \leq P \pi_{t_0}^{-1}(E^c) + \eta \leq P \pi_{t_0}^{-1}(E^{2\epsilon}) + \eta.
\]

Since this holds for all positive \( \epsilon \) and \( \eta \), we finally find that

\[
\limsup_{k \to \infty} P_{s_k}^c(E) \leq P \pi_{t_0}^{-1}(E).
\]

This being so for any \( E \), we conclude by Theorem 1 that \( P_{s_k} \rightarrow P \pi_{t_0}^{-1} \).

**Acknowledgement.**

I have had helpful discussions with K. R. Parthasarathy who suggested to me the two problems stated in the introduction; he was led to these problems by observing an error in the book by Gihman and Skorohod (p. 578 in [2]), where these authors seem to rely on a result like Theorem 2 without observing the difficulties involved.

**References**


UNIVERSITY OF COPENHAGEN, DENMARK