97-F-33

Convergence in Law of Measurable Processes with Applications to the Prediction Process

Hideatsu Tsukahara
University of Tokyo

October 1997

Convergence in Law of Measurable Processes with Applications to the Prediction Process

Hideatsu Tsukahara*

Abstract

We study convergence in law of measurable processes with a general state space and a parameter set. The space of measurable functions are first investigated and we examine properties of probability measures on the space. A necessary and sufficient condition for convergence in law of measurable processes is obtained. These general results are applied to the prediction process, and we show that convergence of the prediction processes implies that of given processes. We also find a simple condition for convergence of the prediction processes when given processes are Markovian.

Keywords: measurable processes, convergence in measure, relative compactness, convergence in law, prediction process.

* Author's affiliation: Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

1. Introduction

The theory of weak convergence has been well known to most probabilists since the publication of Billingsley (1968), and has been effectively applied to stochastic processes. The main concept may be described in the following way. For each $n \in \overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$, let $X^n = (X^n_t)_{t\in\mathbb{T}}$ be a stochastic process defined on a probability space $(\Omega^n, \mathcal{F}^n, P^n)$ with parameter set $\mathbb{T}$ and state space $\mathbb{E}$. Suppose that the paths of X^n lie in a subset $\mathcal{X}$ of $\mathbb{E}^{\mathbb{T}}$, and that $\mathcal{X}$ is endowed with a topology, so that we can consider the Borel σ-field on $\mathcal{X}$. If the mapping $\omega \mapsto X^n_\bullet(\omega)$ from Ω into $\mathcal{X}$ is measurable, then it induces a probability measure μ_{X^n} on $\mathcal{X}$. We say that the sequence (X^n) converges in law to X^∞ if the corresponding sequence of probability measures (μ_{X^n}) converges weakly to μ_{X^∞}, i.e., $\int f(x)\,\mu_{X^n}(dx) \to \int f(x)\,\mu_{X^\infty}(dx)$ for each continuous and bounded function f on $\mathcal{X}$. The cases which have been extensively studied are when: $\mathcal{X} = \mathbb{C}_{\mathbb{E}}(\mathbb{R}_+)$, the space of $\mathbb{E}$-valued continuous functions on $\mathbb{R}_+$ with compact-open topology, and $\mathcal{X} = \mathbb{D}_{\mathbb{E}}(\mathbb{R}_+)$, the space of $\mathbb{E}$-valued càdlàg (right continuous with left limits) functions on $\mathbb{R}_+$ with Skorokhod's J_1 topology. Ethier and Kurtz (1986) and Jacod and Shiryaev (1987) are superb accounts of the recent development in these cases. Less known is the case of the space $\mathbb{L}^p(\mathbb{T}, \mathcal{T}, \nu)$ of real-valued p-th integrable functions on a σ-finite measure space $(\mathbb{T}, \mathcal{T}, \nu)$ although some papers deal with convergence of integral functionals. See Cremers and Kadelka (1986) and the references therein.

In this paper, our primary interest is in the case where the path space $\mathcal{X}$ is the space $\mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T}, \nu)$ of $\mathbb{E}$-valued measurable functions on $\mathbb{T}$. The measure space $(\mathbb{T}, \mathcal{T}, \nu)$ is assumed to be σ-finite , and $\mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T}, \nu)$ is equipped with the topology of convergence in ν-measure on each set of finite measure. To study weak convergence of probability measures on $\mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T}, \nu)$, we need to know the topological properties of the space $\mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T}, \nu)$. The purpose of Section 2 is thus to collect a number of facts on the space. It is well known that $\mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T}, \nu)$ is metrizable, and we provide conditions under which the metric space $\mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T}, \nu)$ is complete or separable. Also we shall derive a characterization of compact sets, assuming a topological group structure on $\mathbb{T}$. A special attention is paid to the particular case where $(\mathbb{T}, \mathcal{T}, \nu) = (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), m)$ and m is the Lebesgue measure. This is important for our later development of the prediction processes and their convergence.

In Section 3, we study properties of probability measures on $\mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T}, \nu)$ and convergence in law of measurable processes. If a process X is measurable, then its paths lie in $\mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T}, \nu)$. It hence induces a law on $\mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T}, \nu)$ provided that the mapping $\omega \mapsto \widetilde{X}_\bullet(\omega)$ is measurable, where $\widetilde{X}_\bullet(\omega)$ is the equivalence class containing $X_\bullet(\omega)$; this measurability is guaranteed if, for example, $\mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T}, \nu)$ is separable. It is then natural to ask when a given process has a measurable modification. The well known necessary and sufficient condition for this problem, given for example in Dellacherie and Meyer (1975), remains true for a general measurable space $(\mathbb{T}, \mathcal{T})$ and a metrizable Lusin space $\mathbb{E}$. We then go on to study the properties of probability measures on $\mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T}, \nu)$. Since $\mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T}, \nu)$ is actually a pseudo-metric space and it becomes a metric space only when we identify two functions which are ν-a.e. equal, some complications arise; for instance, it is not obvious that there exists a process having a given law on $\mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T}, \nu)$. In this connection, the notions of pseudo-path and pseudo-law are relevant. It is based on the viewpoint that we do not have access to the random variables X_t constituting the process X, but only to functions of the form $\int_a^b f(X_t)\,dt$, where f is a measurable function on $\mathbb{E}$. (see Dellacherie and Meyer (1975), IV.5 and IV.35-45). This leads us to the concept of almost equivalence, the precise definition of which will be given in Subsection 3.1. A consequence of our theory is that two measurable processes are almost equivalent if they induce the same law on $\mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T}, \nu)$. In Subsection 3.2, we show that for measurable processes, only the finite-dimensional convergence on a set of full measure is sufficient for weak convergence in $\mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T}, \nu)$. This result is a generalization of the previous work by Sadi (1988), and its converse in a sense will also be established there.

Our primary motivation to work with measurable processes is its generality. Since measurability of a process is a sort of minimum requirement and the parameter set can be any σ-finite measure space in a general setting, the class of processes to which our approach is applicable is very large. It is also interesting to see what can be done only with those minimum requirements. On the other hand, the practical importance of weak convergence comes from the fact that if X^n converges in law to X, then $h(X^n)$ converges in law to $h(X)$ for each continuous function h on the path space. And this allows us to derive many limiting distributions of continuous functionals of processes once we establish the convergence in law of the processes. The topology on $\mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T}, \nu)$ is very coarse, however, so that, as we shall see in Chapter 2, there do not seem to be many continuous functionals on this space except for integral functionals. We hope that our approach is useful in proving the existence of processes and closure properties of various classes of processes as discussed in Meyer and Zheng (1984) and Kurtz (1991).

In Section 4, we apply the results obtained in the preceding sections to the prediction process: two books by Knight (1981, 1992) are excellent expositions of his theory. His original prediction process can be defined for any measurable process X, and since our focus is on measurable processes, it suits our approach to weak convergence. After introducing the prediction process of Knight, we study their convergence in law. Conditions for tightness of conditional distributions and for finite-dimensional convergence of the prediction processes are discussed. And we show that the convergence of (Z^n) is stronger than that of (X^n). In the special case where the X^n are all homogeneous Markov processes, it will be shown that a simple condition on the resolvents of the X^n implies the convergence of their prediction processes.

2. The Space of Measurable Functions

In this section, we collect a number of results on the space of measurable functions endowed with the topology of convergence in measure. Much of the material here should be known, but we provide proofs of some statements when we are unable to locate them in the literature. They are essential in our later development.

2.1 General Case

Let $(\mathbb{T}, \mathcal{T}, \nu)$ be a σ-finite measure space and $(\mathbb{E}, d)$ a metric space. We set $\mathcal{E} = \mathcal{B}_d(\mathbb{E})$, the Borel σ-field on $\mathbb{E}$ generated by the d-open sets. Furthermore, we denote by $\mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T})$ the space of all $\mathcal{T}/\mathcal{E}$ measurable $\mathbb{E}$-valued functions on $\mathbb{T}$. We write $x \sim y$ for $x,\ y \in \mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T})$ if $\nu(t : x(t) \neq y(t)) = 0$. $\sim$ is an equivalence relation. Let $\widetilde{\mathbb{M}}_{\mathbb{E}}(\mathbb{T}, \mathcal{T}) = \mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T})/\sim$ be the space of all equivalence classes of $\mathcal{T}/\mathcal{E}$ measurable functions.

Since $(\mathbb{T}, \mathcal{T}, \nu)$ is σ-finite, there is a sequence $(A_n)_{n\in\mathbb{N}}$ with $A_n \in \mathcal{T},\ A_n \uparrow \Omega$ and $\nu(A_n) < +\infty$ for each $n \in \mathbb{N}$. Choose $\eta_n > 0$ such that $\eta_n \leq 2^{-n}$ and $\eta_n \nu(A_n) \leq 2^{-n}$. Put $\lambda(B) = \sum_{n=1}^{\infty} \eta_n \nu(A_n \cap B),\ B \in \mathcal{T}$. Then $\lambda(\Omega) \leq 1$, so λ is finite. It is easy to see that ν and λ are equivalent, i.e., $\nu \ll \lambda$ and $\lambda \ll \nu$. In fact, a stronger statement $\lambda(B) \leq \nu(B)$ is true for all $B \in \mathcal{T}$; it is trivial if $\nu(B) = \infty$. If $0 < \nu(B) < \infty$, then

$$\frac{\lambda(B)}{\nu(B)} = \sum_{n=1}^{\infty} \eta_n \frac{\nu(A_n \cap B)}{\nu(B)} \leq \sum_{n=1}^{\infty} \eta_n \leq 1.$$

To summarize, we have shown that

2.1 Lemma *If $(\mathbb{T}, \mathcal{T}, \nu)$ is σ-finite , then there exists a finite measure λ such that $\nu \ll \lambda$ and $\lambda(B) \leq \nu(B)$ for all $B \in \mathcal{T}$.*

Suppose now that $\mathbb{E}$ is separable. We say that a sequence $(w_n)_{n\in\mathbb{N}}$ of measurable $\mathbb{E}$-valued functions on $\mathbb{T}$ converges in ν-measure to a measurable $\mathbb{E}$-valued function w on $\mathbb{T}$ if for every $\epsilon > 0$, we have $\lim_{n\to\infty} \nu(t: d(w_n(t), w(t)) \geq \epsilon) = 0$. One must note that the separability of $\mathbb{E}$ ensures that the mapping $t \mapsto d(w_n(t), w(t))$ is $\mathcal{T}$ measurable, so the set $\{t: d(w_n(t), w(t)) \geq \epsilon\}$ is in $\mathcal{T}$. Since $\lambda(B) \leq \nu(B)$ for all $B \in \mathcal{T}$, convergence in ν-measure implies convergence in λ-measure. The following lemma clarifies the meaning of convergence in λ-measure in terms of ν.

2.2 Lemma *A sequence (w_n) converges in λ-measure to w if and only if for every $\epsilon > 0$ and every $A \in \mathcal{F}$ with $\nu(A) < \infty$, we have $\lim_{n\to\infty} \nu(\{d(w_n(t), w(t)) \geq \epsilon\} \cap A) = 0$.*

PROOF. (*if*) Let $\epsilon > 0$ be given. Since λ is finite and there is a sequence (A_n) with $A_n \uparrow \Omega$ and $\nu(A_n) < \infty$, for any $\eta > 0$, we can find an N such that $k \geq N$ entails $\lambda(A_k^{\complement}) < \eta$. Then

$$\begin{aligned}\lambda(d(w_n(t), w(t)) \geq \epsilon) &\leq \lambda(\{d(w_n(t), w(t)) \geq \epsilon\} \cap A_k) + \lambda(\{d(w_n(t), w(t)) \geq \epsilon\} \cap A_k^{\complement}) \\ &\leq \nu(\{d(w_n(t), w(t)) \geq \epsilon\} \cap A_k) + \eta\end{aligned}$$

for all $k \geq N$, and hence $\limsup_{n\to\infty} \lambda(d(w_n(t), w(t)) \geq \epsilon) \leq \eta$.

(*only if*) Let $\epsilon > 0$ be given, and fix $A \in \mathcal{F}$ with $\nu(A) < \infty$. The restriction ν_A of ν to A is a finite measure. Clearly $\nu_A \ll \lambda_A$, and so for any $\eta > 0$ there exists a $\delta > 0$ such that $\lambda_A(B) < \delta$ implies $\nu_A(B) < \eta$. By the assumption, there is an N such that $\lambda(d(w_n(t), w(t)) \geq \epsilon) < \delta$ for $n \geq N$, and we have $\lambda_A(d(w_n(t), w(t)) \geq \epsilon) < \delta$ for $n \geq N$. Hence for all $n \geq N$, $\nu_A(d(w_n(t), w(t)) \geq \epsilon) < \nu(\{d(w_n(t), w(t)) \geq \epsilon\} \cap A) < \eta$. ■

We shall be concerned only with the convergence in λ-measure. The above lemma shows that it really does not depend on a particular choice of λ. Any finite measure equivalent to ν will define the same convergence.

We now define three metrics which metrize the topology of convergence in λ-measure. For $v, w \in \mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T})$, let $\rho_\lambda(v, w)$ be any one of the following:

$$\inf\{\epsilon > 0: \lambda(t: d(v(t), w(t)) > \epsilon) < \epsilon\},$$

$$\int_{\mathbb{T}} \frac{d(v(t), w(t))}{1 + d(v(t), w(t))}\, \lambda(dt), \qquad \int_{\mathbb{T}} 1 \wedge d(v(t), w(t))\, \lambda(dt).$$

It is well known that all of the above metrics define the same uniformity, so using the same symbol $\rho_\lambda(v, w)$ will not be confusing. It is easy to see that the ρ_λ are pseudo-metrics on $\mathbb{M}_\mathbb{E}(\mathbb{T}, \mathcal{T})$, and ρ_λ-convergence is convergence in λ-measure. We write $\mathbb{M}_\mathbb{E}(\lambda) = (\mathbb{M}_\mathbb{E}(\mathbb{T}, \mathcal{T}), \rho_\lambda)$ (pseudo-metric space). The corresponding metric space $\widetilde{\mathbb{M}}_\mathbb{E}(\lambda)$ is defined in an obvious way; for example, for $\widetilde{v}, \widetilde{w} \in \widetilde{\mathbb{M}}_\mathbb{E}(\lambda)$, define a metric $\widetilde{\rho}_\nu(\widetilde{v}, \widetilde{w}) \triangleq \rho_\nu(v, w)$, where v and w are any representatives from the classes $\widetilde{v}$ and $\widetilde{w}$ respectively. Denote by τ_λ the topology induced by $\widetilde{\rho}_\lambda$. At first glance, this topology appears to depend on the metric d on $\mathbb{E}$. But the following proposition shows that it does not.

2.3 Proposition *Let w_n, $n \in \mathbb{N}$ and w be $\mathcal{T}/\mathcal{E}$ measurable functions on $\mathbb{T}$ into $\mathbb{E}$. Then w_n converges to w in λ-measure if and only if*

$$\int_A f(w_n(t))\,\lambda(dt) \to \int_A f(w(t))\,\lambda(dt)$$

for every $f \in C_b(\mathbb{E})$ and every $A \in \mathcal{T}$.

Proof. (*only if*) For any subsequence (n'), there exists a further subsequence (n'') for which $w_{n''}$ converges to w, λ-a.e. Since f is continuous, we have $f \circ w_{n''} \to f \circ w$, λ-a.e. and hence $f \circ w_{n''}\mathbf{1}_A \to f \circ w\mathbf{1}_A$, λ-a.e. By bounded convergence theorem, it follows that $\int f(w_{n''}(t))\mathbf{1}_A\,\lambda(dt) \to \int f(w(t))\mathbf{1}_A\,\lambda(dt)$. This being true for any subsequence, we get $\int_A f(w_n(t))\,\lambda(dt) \to \int_A f(w(t))\,\lambda(dt)$.

(*if*) The $f \circ w_n$ and $f \circ w$ are bounded, so they are in $\mathbb{L}^2(\mathbb{T}, \mathcal{T}, \lambda)$. Thus the assumption means that $(f \circ w_n)$ converges weakly in $\mathbb{L}^2(\mathbb{T}, \mathcal{T}, \lambda)$ to $f \circ w$. Also $\int [f(w_n(t))]^2\,\lambda(dt) \to \int [f(w(t))]^2\,\lambda(dt)$ since $f^2 \in C_b(\mathbb{E})$. These two facts imply that $(f \circ w_n)$ converges to $f \circ w$ strongly in $\mathbb{L}^2(\mathbb{T}, \mathcal{T}, \lambda)$. In fact,

$$\begin{aligned}
&\int [f(w_n(t)) - f(w(t))]^2\,\lambda(dt) \\
&= \int [f(w_n(t))]^2\,\lambda(dt) - 2\int f(w_n(t))f(w(t))\,\lambda(dt) + \int [f(w(t))]^2\,\lambda(dt) \\
&\to 2\int [f(w(t))]^2\,\lambda(dt) - 2\int f(w(t))f(w(t))\,\lambda(dt) = 0.
\end{aligned}$$

Thus there exists a subsequence (depending on f) (n_k) such that $f \circ w_{n_k} \to f \circ w$, λ-a.e. It is well known that we can find a countable family $(f_i)_{i \in \mathbb{N}}$ of $C_b(\mathbb{E})$ functions such that $d(x_n, x) \to 0$ if and only if $f_i(x_n) \to f_i(x)$ for all i. By the diagonal argument, we can find a single subsequence (n_k) such that $f_i \circ w_{n_k} \to f_i \circ w$ for all i, λ-a.e., i.e., $w_{n_k}(t) \to w(t)$

for λ-a.e. t. Starting with an arbitrary subsequence and applying the above argument, we easily obtain the required result. ■

We now turn to the properties of the space $\mathbb{M}_{\mathbb{E}}(\lambda)$ and $\widetilde{\mathbb{M}}_{\mathbb{E}}(\lambda)$. As seen above, $\mathbb{M}_{\mathbb{E}}(\lambda)$ is a pseudo-metric space with the pseudo-metric ρ_λ and $\widetilde{\mathbb{M}}_{\mathbb{E}}(\lambda)$ is a metric space with the metric $\widetilde{\rho}_\lambda$.

2.4 Proposition *If $(\mathbb{E}, d)$ is a separable metric space and if $\mathcal{T}$ is countably generated up to null sets, then $\mathbb{M}_{\mathbb{E}}(\lambda)$ is separable.*

PROOF. Let $\{v_1, v_2, \ldots\}$ be a countable dense set in $\mathbb{E}$. For a given $\epsilon > 0$, the open balls B_n^ϵ, $n \in \mathbb{N}$, centered at v_n and with radius ϵ cover $\mathbb{E}$. Set $C_n^\epsilon = B_n^\epsilon - \bigcup_{i=1}^{n-1} B_i^\epsilon$. Then $C_n^\epsilon \in \mathcal{E}$ and $\{C_n^\epsilon\}_{n\in\mathbb{N}}$ covers $\mathbb{E}$. Also the C_n^ϵ are mutually disjoint. We drop ϵ and write C_n since $\epsilon > 0$ is fixed throughout.

Next let $f \in \mathbb{M}_{\mathbb{E}}(\mathbb{T}, \mathcal{T})$ be given. Put $A_n = f^{-1}(C_n)$. We have $\bigcup_{n=1}^\infty A_n = \mathbb{T}$, $A_n \in \mathcal{T}$, and the A_n are mutually disjoint. Since λ is finite, for any $\eta > 0$, we can find an N such that $\lambda(\mathbb{T}) - \lambda(\bigcup_{n=1}^N A_n) < \eta$. Note now that there is a field $\mathcal{T}_0$ consisting of a countable number of sets such that $\mathcal{T} = \sigma(\mathcal{T}_0)$ because $\mathcal{T}$ is countably generated. Thus for each $n \leq N$, there exists an $H_n \in \mathcal{T}_0$ such that $\lambda(A_n \triangle H_n) < 2^{-n}\frac{\eta}{N}$ (Billingsley(1986), Theorem 11.4). Put $J_n = H_n - \bigcup_{i=1}^{n-1} H_i$. We have

$$\lambda\left(\bigcup_{n=1}^N A_n - \bigcup_{n=1}^N J_n\right) \leq \lambda\left(\bigcup_{n=1}^N (A_n \triangle H_n)\right) \leq \sum_{i=1}^N \frac{\eta}{2^i N} \leq \eta.$$

Define

$$g(t) \triangleq \begin{cases} v_n, & \text{if } t \in J_n; \\ v_0, & \text{if } t \in (\bigcup_{n=1}^N J_n)^{\mathbf{c}}, \end{cases}$$

where v_0 is any fixed point in $\mathbb{E}$. By writing out $J_n \triangle A_n$ explicitly, one finds

$$\lambda(J_n \triangle A_n) \leq \lambda(A_n \triangle H_n) + \sum_{i=1}^{n-1} \lambda(A_n \cap H_n)$$
$$\leq \frac{\eta}{2^n N} + \sum_{i=1}^{n-1} \frac{\eta}{2_i N} = \frac{\eta}{N}\left[1 - \left(\frac{1}{2}\right)^n\right],$$

and hence

$$\lambda(d(f(t),g(t))\geq\epsilon)\leq\lambda\left(\bigcup_{n=N+1}^{\infty}A_n\right)+\lambda\left(\bigcup_{n=1}^{N}A_n-\bigcup_{n=1}^{N}J_n\right)+\sum_{n=1}^{N}\lambda(J_n\triangle A_n)$$
$$\leq 2\eta+\frac{\eta}{N}\sum_{n=1}^{N}\left[1-\left(\frac{1}{2}\right)^n\right]\leq 3\eta.$$

It follows that $\inf\{\epsilon>0\colon\lambda\{d(x(t),y(t))\geq\epsilon\}\leq\epsilon\}\leq\epsilon\vee 3\eta$. We have proved that for any $f\in\mathbb{M}_{\mathbb{E}}(\mathbb{T},\mathcal{T})$ and $\epsilon>0$, we can find a $\mathcal{T}_0$-simple function g with values in $\{v_0,v_1,v_2,\ldots\}$ such that $\rho_\lambda(f,g)<\epsilon$. Thus the family of all $\mathcal{T}_0$-simple functions with values in $\{v_0,v_1,v_2,\ldots\}$ is dense in $\mathbb{M}_{\mathbb{E}}(\lambda)$, and it is obviously countable. Therefore $\mathbb{M}_{\mathbb{E}}(\lambda)$ is separable. ■

The proposition implies that $\widetilde{\mathbb{M}}_{\mathbb{E}}(\lambda)$ is a separable metric space. If in addition $(\mathbb{E},d)$ is complete, then we have

2.5 Proposition *If $(\mathbb{E},d)$ is complete and separable and if $\mathcal{T}$ is countably generated up to null sets, then $\mathbb{M}_{\mathbb{E}}(\lambda)$ is complete and separable.*

A proof of this proposition is given in Kurtz (1991), pp.1022-3 in the case where $(\mathbb{T},\mathcal{T},\lambda)=(\mathbb{R}_+,\mathcal{B}_+,e^{-t}dt)$, and it carries over to our case word for word.

When we discuss weak convergence of probability measures on $\widetilde{\mathbb{M}}_{\mathbb{E}}(\lambda)$, it is important, because of Prohorov theorem, to find a compactness criterion in $\widetilde{\mathbb{M}}_{\mathbb{E}}(\lambda)$. For a general measure space $(\mathbb{T},\mathcal{T},\lambda)$, the following result is known.

2.6 Proposition *Let $(\mathbb{E},d)$ be a complete and separable metric space. A subset Γ of $\widetilde{\mathbb{M}}_{\mathbb{E}}(\lambda)$ is relatively compact if and only if for every $\epsilon>0$, there exist a finite partition $\{A_1,A_2,\ldots,A_N\}$ of $\mathbb{T}$, a compact subset K of $\mathbb{E}$ and, corresponding to each $w\in\Gamma$, a set A_w with $\lambda(A_w)<\epsilon$ such that*

(1) $w(s)\in K$ *for all* $w\in\Gamma$ *and* $s\notin A_w$;

(2) $\sup_{s,t\in A_k-A_w} d(w(s),w(t))<\epsilon,\quad k=1,2,\ldots,N.$

See Dunford and Schwartz (1958), IV.11.1 for a proof. This criterion, however, is not useful for our purpose because the resulting tightness criterion for processes would be hard to verify. And it does not seem likely that a useful condition can be obtained for a general $(\mathbb{T}, \mathcal{T}, \lambda)$. Kurtz (1991) obtained a useful criterion for $(\mathbb{T}, \mathcal{T}, \nu) = (\mathbb{R}_+, \mathcal{B}_+, m)$, where m is the Lebesgue measure (see Proposition 2.12), and here we prove its generalization to the case where $\mathbb{T}$ has a group structure.

Let $\mathbb{T}$ be a σ-compact locally compact Hausdorff topological group. We use the multiplication as the group operation on $\mathbb{T}$. We denote by $\mathcal{T}$ the Borel σ-field $\mathcal{B}(\mathbb{T})$ on $\mathbb{T}$. It is known that a left invariant Haar measure ν exists in this case, and by the σ-compactness assumption, it is σ-finite (recall that a left invariant Haar measure ν is a Radon measure which is not identically zero and satisfies $\nu(tA) = \nu(A)$ for each $A \in \mathcal{T}$, and that a Radon measure is a regular measure on the Borel σ-field which assigns a finite measure to each compact set). We denote by λ a finite measure which satisfies the conditions in Lemma 2.1. We know that $\mathbb{M}_{\mathbb{E}}(\lambda)$ is complete, but it may not be separable as $\mathcal{T}$ here is not assumed to be countably generated up to ν-null sets. To obtain a compactness criterion, separability is not necessary although we certainly need it when we study weak convergence in $\mathbb{M}_{\mathbb{E}}(\lambda)$.

The result we shall prove is the following.

2.7 Theorem *Let $(\mathbb{E}, d)$ be a complete and separable metric space. A subset Γ of $\widetilde{\mathbb{M}}_{\mathbb{E}}(\lambda)$ is relatively compact if and only if the following conditions hold:*

(1) For every $\epsilon > 0$ and every compact set $C \subset \mathbb{T}$, there exists a compact $K \subset \mathbb{E}$ such that

$$\sup_{w \in \Gamma} \nu\{t \in C : w(t) \notin K\} \leq \epsilon.$$

(2) For every $\epsilon > 0$ and every compact set $C \subset \mathbb{T}$, there exists a neighborhood U of the identity of $\mathbb{T}$ such that $u \in U$ implies

$$\sup_{w \in \Gamma} \int_C 1 \wedge d(w(ut), w(t))\, \nu(dt) \leq \epsilon.$$

REMARK. It is easy to show that (1) is equivalent to the following condition in terms of λ:

(1') For every $\epsilon > 0$, there exists a compact $K \subset \mathbb{E}$ such that

$$\sup_{w \in \Gamma} \lambda\{t \in \mathbb{T} : w(t) \notin K\} \leq \epsilon.$$

Similarly (2) is equivalent to

(2') For every $\epsilon > 0$, there exists a neighborhood U of the identity of $\mathbb{T}$ such that $u \in U$ implies

$$\sup_{w \in \Gamma} \int_{\mathbb{T}} 1 \wedge d(w(ut), w(t))\, \lambda(dt) \leq \epsilon.$$

For the proof of Theorem 2.7, we need the following lemma due to Kurtz (1991).

2.8 Lemma *Suppose that* Γ *satisfies (1). Then* Γ *is relatively compact if and only if* $\{f \circ w : w \in \Gamma\}$ *is relatively compact in* $\mathbb{M}_{\mathbb{R}}(\lambda)$ *for every* $f \in C_b(\mathbb{E})$.

He proved the lemma for $(\mathbb{T}, \mathcal{T}, \nu) = (\mathbb{R}_+, \mathcal{B}_+, m)$, but a quick investigation of his proof shows that it is true for any σ-finite $(\mathbb{T}, \mathcal{T}, \nu)$.

Proof of Theorem 2.7. (*if*) By Lemma 2.8, it suffices to show that $B_f \triangleq \{x \in \mathbb{M}_{\mathbb{R}}(\lambda) : x = f \circ w,\ w \in \Gamma\}$ is relatively compact in $\mathbb{M}_{\mathbb{R}}(\lambda)$ for each $f \in C_b(\mathbb{E})$ with $0 \leq f \leq 1$. Since $\mathbb{M}_{\mathbb{R}}(\lambda)$ is complete, we only need to show that B_f is totally bounded.

Let $\epsilon > 0$ be given. Choose a compact $C \subset \mathbb{T}$ such that $\lambda(C^{\complement}) < \epsilon$. Also choose any compact neighborhood W of the identity of $\mathbb{T}$. Since WC is compact, by (1), there is a compact $K \subset \mathbb{E}$ such that

$$\sup_{w \in \Gamma} \nu\{t \in WC : w(t) \notin K\} \leq \epsilon. \tag{a}$$

Then select a neighborhood U of the identity of $\mathbb{T}$ with $\overline{U} \subset W$ such that $u \in U$ implies

$$\sup_{w \in \Gamma} \int_C \psi_K(d(w(s^{-1}t), w(t)))\, \nu(dt) < \epsilon, \tag{b}$$

where

$$\psi_K(r) = \sup_{x, y \in K,\ d(x,y) \leq r} |f(x) - f(y)|.$$

This is possible because $\psi_K(r)$ is continuous at 0 with $\psi_K(0) = 0$ (f is uniformly continuous on K) and by (2). Finally, pick up an $h \in C_c(\mathbb{T})$ (the space of continuous functions on $\mathbb{T}$ with compact support) satisfying $h \geq 0$, $\int h\, d\nu = 1$ and $\operatorname{supp}(h) \subset U$. Put

$$h * x(t) = \int x(s^{-1}t) h(s)\, \nu(ds),$$

for $x \in B_f$. Recalling that $\lambda(A) \le \nu(A)$ for $A \in \mathcal{T}$, we have

$$\int_{\mathbb{T}} 1 \wedge |h * x(t) - x(t)| \, \lambda(dt) \le \int_C 1 \wedge |h * x(t) - x(t)| \, \lambda(dt) + \epsilon$$
$$\le \int_C 1 \wedge |h * x(t) - x(t)| \, \nu(dt) + \epsilon,$$

and

$$\int_C 1 \wedge |h * x(t) - x(t)| \, \nu(dt) \le \int_U \int_C |x(s^{-1}t) - x(t)| \, \nu(dt) h(s) \, \nu(ds)$$
$$\le \int_U h(s) \Big[\int_{C \cap \{t: w(s^{-1}t) \in K, \ w(t) \in K\}} |f \circ w(s^{-1}t) - f \circ w(t)| \, \nu(dt)$$
$$+ \nu(C \cap \{t: w(s^{-1}t) \notin K \text{ or } w(t) \notin K\}) \Big] \, \nu(ds)$$
$$\le \int_U h(s) \left[\int_C \psi_K(d(w(s^{-1}t), w(t))) \, \nu(dt) + 2\nu\{t \in UC: w(t) \notin K\} \right] \nu(ds).$$

It hence follows from (a) and (b) that

$$\sup_{x \in B_f} \int 1 \wedge |h * x(t) - x(t)| \, \lambda(dt) < 4\epsilon. \tag{c}$$

Furthermore it is easy to see that the family $\{h * x: x \in B_f\}$ is equicontinuous; we have

$$\left|h * x(u^{-1}t) - h * x(t)\right| = \left| \int x(s^{-1}) h(u^{-1}ts) \, \nu(ds) - \int x(s^{-1}) h(ts) \, \nu(ds) \right|$$
$$\le \int |h(u^{-1}ts) - h(ts)| \, \nu(ds) = \int |h(u^{-1}s) - h(s)| \, \nu(ds),$$

and this is small uniformly in $x \in \Gamma$ if u is in a suitable neighborhood of the identity of $\mathbb{T}$ (recall that $h \in C_c(\mathbb{T})$). Obviously, $|h * x| \le 1$. It thus follows from a topological version of Ascoli-Arzela theorem (Bourbaki (1974), X.2.5) that for each $h \in C_c(\mathbb{T})$, $\{h * x: x \in B_f\}$ is relatively compact in $\mathbb{C}_{\mathbb{R}}(\mathbb{T})$ with the topology of compact convergence. Since the topology of $\mathbb{M}_{\mathbb{R}}(\lambda)$ is coarser than that topology, $\{h * x: x \in B_f\}$ is relatively compact in $\mathbb{M}_{\mathbb{R}}(\lambda)$ as well. This, together with (c), implies that B_f is relatively compact in $\mathbb{M}_{\mathbb{R}}(\lambda)$, as required.

(*only if*) Suppose that Γ is relatively compact in $\mathbb{M}_{\mathbb{E}}(\lambda)$. We shall first prove (1'), which is equivalent to (1). For any sequence in Γ, there exists a subsequence (w_n) converging in λ-measure. Then the sequence of probability measures $(w_n(\lambda))$ on $\mathbb{E}$ converges weakly (note that we have adopted the notation $w_n(\lambda)(A) = \lambda(w_n^{-1}(A))$). This means that the family $\{w(\lambda): w \in \Gamma\}$ is relatively compact. Since $\mathbb{E}$ is complete and separable, the converse half of Prohorov theorem shows that $\{w(\lambda): w \in \Gamma\}$ is tight. Namely, for any $\epsilon > 0$, there is a compact $K \subset \mathbb{E}$ such that $\sup_{w \in \Gamma} w(\lambda)(K^{\mathbf{c}}) \leq \epsilon$, which is exactly (1').

Next, we show the following assertion:

(#) For any $w \in \mathbb{M}_{\mathbb{E}}(\lambda)$, $\epsilon > 0$ and compact $C \subset \mathbb{T}$, there exists a neighborhood U of the identity of $\mathbb{T}$ such that if $u \in U$, then $\int_C 1 \wedge d(w(ut), w(t))\, \nu(dt) < \epsilon$.

Let f be an imbedding of $\mathbb{E}$ into $[0,1]^{\mathbb{N}}$ and write $f(x) = (f_1(x), f_2(x), \ldots)$. Since $f_i \circ w \mathbf{1}_C \in \mathbb{L}^1(\mathbb{T}, \mathcal{T}, \nu)$, by the standard result in Haar measure theory (Hewitt and Ross (1979), Theorem (20.4)), there is a neighborhood U_i of the identity of $\mathbb{T}$ such that $\int_C |f_i \circ w(ut) - f_i \circ w(t)|\, \nu(dt) < \epsilon/2$ for $u \in U_i$. We may assume that $\nu(C) > 0$ because the claim is trivial if $\nu(C) = 0$. Choose $N \in \mathbb{N}$ satisfying $\sum_{i=N+1}^{\infty} 2^{-i} < \epsilon/2\nu(C)$ and put $U = \bigcap_{i=1}^{N} U_i$. Then for $u \in U$, we have

$$\begin{aligned} &\sum_{i=1}^{\infty} 2^{-i} \int_C |f_i \circ w(ut) - f_i \circ w(t)|\, \nu(dt) \\ &\leq \sum_{i=1}^{N} 2^{-i} \int_C |f_i \circ w(ut) - f_i \circ w(t)|\, \nu(dt) + \sum_{i=N+1}^{\infty} 2^{-i} \nu(C) \leq \epsilon. \end{aligned}$$

Now choose any compact neighborhood W of the identity of $\mathbb{T}$. It follows from Lusin theorem (or (1)) that there is a compact $K \subset \mathbb{E}$ such that $\nu\{t \in WC: w(t) \notin K\} \leq \epsilon$. Moreover we can find an increasing functions φ_K on $\mathbb{R}_+$ into $\mathbb{R}_+$ which is continuous at 0 and satisfies $\varphi_K(0) = 0$ and

$$d(x,y) \leq \varphi_K\left(\sum_{i=1}^{\infty} 2^{-i} |f_i(x) - f_i(y)|\right), \qquad x, y \in K.$$

Then we have

$$\begin{aligned} &\int_C 1 \wedge d(w(ut), w(t))\, \nu(dt) \\ &\leq \int_C \varphi_K\left(\sum_{i=1}^{\infty} 2^{-i} |f_i \circ w(ut) - f_i \circ w(t)|\right) \nu(dt) + \nu\left(C \cap \{t: w(ut) \notin K \text{ or } w(t) \notin K\}\right). \end{aligned}$$

If we choose a suitable neighborhood $U \subset W$, we can make the first term less than ϵ. The second term is clearly less than 2ϵ. Thus (#) is proved.

Finally, we prove (2). Let $\epsilon > 0$ and a compact set $C \subset \mathbb{T}$ be given. Also let W be any compact neighborhood of the identity of $\mathbb{T}$. Then, by the total boundedness of Γ, we can find a finite subset $\{w_1, \dots, w_k\}$ of Γ such that for any $w \in \Gamma$, there exists a $w_i \in \{w_1, \dots, w_k\}$ satisfying

$$\int_{WC} 1 \wedge d(w_i(t), w(t))\, \nu(dt) < \frac{\epsilon}{3}.$$

Hence for $u \in W$,

$$\begin{aligned}
&\int_C 1 \wedge d(w_i(t), w(t))\, \nu(dt) \\
&\leq \int_C 1 \wedge d(w(ut), w_i(ut))\, \nu(dt) + \int_C 1 \wedge d(w_i(ut), w_i(t))\, \nu(dt) + \int_C 1 \wedge d(w_i(t), w(t))\, \nu(dt) \\
&< 2\epsilon + \int_C 1 \wedge d(w_i(ut), w_i(t))\, \nu(dt).
\end{aligned}$$

By (#), we can find a neighborhood $U \subset W$ such that $u \in U$ implies

$$\int_C 1 \wedge d(w_i(ut), w_i(t))\, \nu(dt) < \epsilon \quad \text{for all } i = 1, \dots, k.$$

Consequently, $\int_C 1 \wedge d(w(ut), w(t))\, \nu(dt) < 3\epsilon$ uniformly in $w \in \Gamma$, and (2) is proved. ■

NOTE. This is an $\mathbb{M}_{\mathbb{E}}(\lambda)$ analogue of Weil's compactness criterion in $\mathbb{L}^p(\mathbb{T}, \mathcal{T}, \nu)$. For this result, see Weil (1951), p.53.

2.2 A special case where $(\mathbb{T}, \mathcal{T}, \nu) = (\mathbb{R}_+, \mathcal{B}_+, m)$, *m: Lebesgue measure*

In this section, we shall investigate a special case of the preceding section where $(\mathbb{T}, \mathcal{T}, \nu) = (\mathbb{R}_+, \mathcal{B}_+, m)$. Most of the results here are taken from Knight (1992), sometimes with extended proofs. Let $\mathbb{E}$ be a metrizable Lusin space and $\mathcal{E} = \mathcal{B}(\mathbb{E})$. Also let $\mathbb{M}_{\mathbb{E}} = \mathbb{M}_{\mathbb{E}}(\mathbb{R}_+, \mathcal{B}_+, \lambda)$, where $\mathcal{B}_+ = \mathcal{B}(\mathbb{R}_+)$ and $\lambda(dt) = e^{-t}dt$. This space is of our main interest in Section 4.

We first consider $\mathbb{M}_{[0,1]}$ and give it a topology which makes $\widetilde{\mathbb{M}}_{[0,1]}$ a compact metrizable space. Define

$$d'(w_1, w_2) \triangleq \sup_{t>0} \left| \int_0^t e^{-s}(w_1(s) - w_2(s))\, ds \right|, \quad w_1, w_2 \in \mathbb{M}_{[0,1]}.$$

It is easy to see that d' is a metric on $\widetilde{\mathbb{M}}_{[0,1]}$ and $d' \leq 1$. The convergence for this metric is the same as $\int_0^t e^{-s} w_n(s)\, ds \to \int_0^t e^{-s} w(s)\, ds$ uniformly in t, but in fact it is equivalent to pointwise convergence since $t \mapsto \int_0^t e^{-s} w(s)\, ds$ is uniformly continuous, absolutely continuous and increasing. To show that $\widetilde{\mathbb{M}}_{[0,1]}$ is compact under d', it is enough to prove that it is sequentially compact. By Ascoli-Arzela theorem, any sequence (w_n) has a subsequence $(w_{n'})$ for which

$$\lim_{n\to\infty} \int_0^t e^{-s} w_{n'}(s)\, ds =: F(t) \quad (\text{uniformly in } t)$$

exists because $\{\int_0^{\bullet} e^{-s} w_n(s)\, ds\}_{n\in\mathbb{N}}$ is uniformly bounded by 1 and uniformly equicontinuous. Indeed,

$$0 \leq \int_{t_1}^{t_2} e^{-s} w_n(s)\, ds \leq e^{-t_1}(t_2 - t_1),\ t_1 < t_2,$$

and so $0 \leq F(t_2) - F(t_1) \leq e^{-t_1}(t_2 - t_1)$. Hence F is Lipschitz, and $0 \leq F'(t) \leq e^{-t}$ for a.e. t. Set

$$w_\infty(t) \triangleq \begin{cases} e^t F'(t), & \text{if } F'(t) \text{ exists;} \\ 0, & \text{otherwise.} \end{cases}$$

Then $w_\infty \in \mathbb{M}_{[0,1]}$, and clearly $F(t) = \int_0^t e^{-s} w_\infty(s)\, ds$, so $d'(w_{n'}, w_\infty) \to 0$. Thus $\widetilde{\mathbb{M}}_{[0,1]}$ is compact under d'. We also note that the d'-topology is generated by $\{w \mapsto \int_0^t e^{-s} w(s)\, ds : t > 0\}$, or by $\{w \mapsto \int_0^t w(s)\, ds : t > 0\}$ ($t > 0$ may be replaced by $t \in \mathbb{Q}_+$).

Next, let $\widehat{\mathbb{E}}$ be a compact metrizable space containing $\mathbb{E}$ as a Borel subspace (Lusin property of $\mathbb{E}$). We fix a metric d_0 which is compatible with the topology of $\widehat{\mathbb{E}}$. Let $C_u(\mathbb{E}, d_0)$ be the space of bounded, uniformly continuous real-valued functions. $C_u(\mathbb{E}, d_0)$ consists precisely of the restrictions to $\mathbb{E}$ of the functions in $C(\widehat{\mathbb{E}})$; if $f \in C(\widehat{\mathbb{E}})$, then it is uniformly continuous on $\widehat{\mathbb{E}}$ and hence $f|_{\mathbb{E}}$ is uniformly continuous. Conversely, if $f \in C_u(\mathbb{E}, d_0)$, it can be extended to the completion of $\mathbb{E}$ (which is the closure of $\mathbb{E}$ in $\widehat{\mathbb{E}}$), and then, by Tietze's theorem, further extended to $\widehat{\mathbb{E}}$ preserving continuity. $C(\widehat{\mathbb{E}})$ is separable with respect to the sup-metric, so that there exists a sequence (f_j) which is dense in $C(\widehat{\mathbb{E}}) \cap \{f : \widehat{\mathbb{E}} \to [0,1]\}$. Thus $(f_j|_{\mathbb{E}})$ is dense in $C_u(\mathbb{E}, d_0)$ with respect to the sup-metric. Define a mapping $\mathbf{f} : \widehat{\mathbb{E}} \to [0,1]^{\mathbb{N}}$ by

$$\mathbf{f}(e) \triangleq (f_1(e), f_2(e), \ldots).$$

Clearly $\mathbf{f}$ is continuous and 1-1 since the f_j separate points of $\widehat{\mathbb{E}}$. Also if $f_j(e_n) \to f_j(e)$ for each j, then $f(e_n) \to f(e)$ for all $f \in C(\widehat{\mathbb{E}})$, which in turn implies that $e_n \to e$. Hence $\mathbf{f}|_{\mathbb{E}}$ is an imbedding of $\mathbb{E}$ into $[0,1]^{\mathbb{N}}$, i.e., $\mathbf{f}|_{\mathbb{E}}$ as a mapping from $\mathbb{E}$ onto $\mathbf{f}(\mathbb{E})$ is a homeomorphism. It also follows from Lusin's theorem that $\mathbf{f}(\mathbb{E}) \in \mathcal{B}[0,1]^{\mathbb{N}}$. If we define

$$d(e_1, e_2) \triangleq \sum_{j=1}^{\infty} 2^{-j} |f_j(e_1) - f_j(e_2)|, \quad e_1, e_2 \in \mathbb{E} \text{ (or } \widehat{\mathbb{E}}),$$

then obviously d is a metric compatible with the topology of $\mathbb{E}$ (and $\widehat{\mathbb{E}}$).

We now turn to the space $\mathbb{M}_{\mathbb{E}}$ endowed with the topology of convergence in λ-measure. We can characterize the convergence in λ-measure as follows (Knight (1992), Theorem 2.3):

2.9 Proposition *$w_n \to w$ in λ-measure if and only if we have*

$$\int_0^r f_j(w_n(s))\, ds \to \int_0^r f_j(w(s))\, ds,$$

for every $r \in \mathbb{Q}_+$ and f_j.

Proof. The only if part is immediate: $w_n \to w$ in λ-measure implies $f_j \circ w_n \to f_j \circ w$ in λ-measure by the continuity of f_j, and then apply the bounded convergence theorem. To prove the if part, note that $\int_0^\infty g(s) f_j(w_n(s))\, ds \to \int_0^\infty g(s) f_j(w(s))\, ds$ for each continuous g with compact support since such a g can be uniformly approximated by a sequence of step functions. This implies $\int_0^\infty g(s) f_j(w_n(s))\, \lambda(ds) \to \int_0^\infty g(s) f_j(w(s))\, \lambda(ds)$ for such functions g. Take $T > 0$ such that $\lambda[T, \infty) < \epsilon$. For $g \in C_b(\mathbb{R}_+)$, let M be a number satisfying $|g| \le M$. For $\epsilon > 0$, define a function g_ϵ with compact support by

$$g_\epsilon(t) \triangleq \begin{cases} g(t), & t \in [0,T]\ ; \\ g(T)(T+1-t), & t \in (T, T+1]\ ; \\ 0, & t \in (T+1, \infty). \end{cases}$$

Then $|g - g_\epsilon| \le 2M$ and we have

$$\int_0^\infty |g(s) - g_\epsilon(s)| f_j(w_n(s))\, \lambda(ds) \le \int_T^\infty 2M\, \lambda(ds) \le 2M\epsilon.$$

The same inequality holds with $w(s)$ in place of $w_n(s)$. Hence

$$\begin{aligned}
&\left|\int_0^\infty g(s)f_j(w_n(s))\,\lambda(ds) - \int_0^\infty g(s)f_j(w(s))\,\lambda(ds)\right| \\
&\le \int_0^\infty |g(s)-g_\epsilon(s)|f_j(w_n(s))\,\lambda(ds) + \int_0^\infty |g(s)-g_\epsilon(s)|f_j(w(s))\,\lambda(ds) \\
&\qquad + \left|\int_0^\infty g_\epsilon(s)f_j(w_n(s))\,\lambda(ds) - \int_0^\infty g_\epsilon(s)f_j(w(s))\,\lambda(ds)\right| \\
&\le 4M\epsilon + \left|\int_0^\infty g_\epsilon(s)f_j(w_n(s))\,\lambda(ds) - \int_0^\infty g_\epsilon(s)f_j(w(s))\,\lambda(ds)\right|.
\end{aligned}$$

Letting $n \to \infty$, it follows that

$$\limsup_{n\to\infty}\left|\int_0^\infty g(s)f_j(w_n(s))\,\lambda(ds) - \int_0^\infty g(s)f_j(w(s))\,\lambda(ds)\right| \le 4M\epsilon.$$

Since $\epsilon > 0$ was arbitrary, we get $\int_0^\infty g(s)f_j(w_n(s))\,\lambda(ds) \to \int_0^\infty g(s)f_j(w(s))\,\lambda(ds)$ for every $g \in C_b(\mathbb{R}_+)$. Noting that $C_b(\mathbb{R}_+)$ is dense in $\mathbb{L}^2(\lambda)$, we conclude that $f_j \circ w_n \to f_j \circ w$ weakly in $\mathbb{L}^2(\lambda)$. However we also have $\int_0^\infty f_j^2(w_n(s))\,\lambda(ds) \to \int_0^\infty f_j^2(w(s))\,\lambda(ds)$ since $f_j^2 \in C(\widehat{\mathbb{E}})$, and so $\|f_j \circ w_n\|_2 \to \|f_j \circ w\|_2$. Hence letting $n \to \infty$ in

$$\|f_j \circ w_n - f_j \circ w\|_2 = \|f_j \circ w_n\|_2^2 - 2\langle f_j \circ w_n, f_j \circ w\rangle + \|f_j \circ w\|_2^2,$$

we obtain the strong convergence of $f_j \circ w_n$ to $f_j \circ w$ in $\mathbb{L}^2(\lambda)$, which implies $f_j \circ w_n \to f_j \circ w$ in λ-measure. A distance between w_n and w under the topology of convergence in λ-measure is given by

$$\int_0^\infty \sum_{j=1}^\infty 2^{-j}|f_j(w_n(s)) - f_j(w(s))|\,\lambda(ds),$$

and this can be made arbitrarily small if n is large. ■

Consider the mapping $F\colon \mathbb{M}_{\mathbb{E}} \to \mathbb{M}_{[0,1]}^{\mathbb{N}}$ defined by $F(w) = (f_1 \circ w, f_2 \circ w, \ldots)$. We remark that $\mathbb{M}_{\mathbb{E}}$ is endowed with the topology of convergence in λ-measure, while $\mathbb{M}_{[0,1]}$ is equipped with the topology defined earlier, which makes it a compact metrizable space. And we give $\mathbb{M}_{[0,1]}^{\mathbb{N}}$ the product topology, so $\mathbb{M}_{[0,1]}^{\mathbb{N}}$ is also compact metrizable. We see from the above proposition that the topology on $\mathbb{M}_{\mathbb{E}}$ is generated by $\{w \mapsto \int_0^r f_j(w(s))\,ds\colon r \in \mathbb{Q}_+, j \in \mathbb{N}\}$. Also the topology on $\mathbb{M}_{[0,1]}^{\mathbb{N}}$ is generated by $\{v \mapsto \int_0^r v_i(s)\,ds\colon r \in \mathbb{Q}_+, i \in \mathbb{N}\}$, where we write $v(s) = (v_1(s), v_2(s), \ldots)$ for $v \in \mathbb{M}_{[0,1]}^{\mathbb{N}}$. It is then clear that F is continuous and 1-1, and F^{-1} restricted to the range of F is continuous. Hence F is an imbedding of $\mathbb{M}_{\mathbb{E}}$ into $\mathbb{M}_{[0,1]}^{\mathbb{N}}$. We are in fact talking about equivalence class here, but to avoid confusion, note that if $w \in \widetilde{w}$, then $f_j \circ w \in \widetilde{f_j \circ w}$ and this equivalence class does not depend on the choice of $w \in \widetilde{w}$. Now if $\widetilde{v} \in \widetilde{\mathbb{M}}_{[0,1]}^{\mathbb{N}}$ is in $F(\widetilde{\mathbb{M}}_{\mathbb{E}})$, then $\widetilde{v} = \widetilde{F(w)}$ for some w, so that $\lambda\{t\colon v(t) \in \mathbf{f}(\mathbb{E})\} = 1$ for every $v \in \widetilde{v}$. Conversely, if $v(t) \in \mathbf{f}(\mathbb{E})$, λ-a.e. t, then there exists a $v^* \in \widetilde{v}$ such that $v^*(t) \in \mathbf{f}(\mathbb{E})$ for all t. Put $w^* = \mathbf{f}^{-1} \circ v^*$. Then $v^* = \mathbf{f} \circ w^*$ and so $\widetilde{v} = \widetilde{\mathbf{f} \circ w^*} \in F(\widetilde{\mathbb{M}}_{\mathbb{E}})$. Thus we have

$$F(\widetilde{\mathbb{M}}_{\mathbb{E}}) = \left\{ \widetilde{v} \in \widetilde{\mathbb{M}}_{[0,1]}^{\mathbb{N}} \colon \int_0^\infty \mathbf{1}_{\mathbf{f}(\mathbb{E})}(v(t))\, \lambda(dt) = 1 \right\}.$$

The mapping $\widetilde{v} \mapsto \int_0^\infty f(v(t))\, \lambda(dt)$ is continuous for $f \in C_b([0,1]^{\mathbb{N}})$, thus it is $\mathcal{B}(\widetilde{\mathbb{M}}_{[0,1]}^{\mathbb{N}})$ measurable, and this measurability can be extended to the functions $f \in b\mathcal{B}([0,1]^{\mathbb{N}})$ by the monotone class argument. Noting that $\mathbf{f}(\mathbb{E}) \in \mathcal{B}([0,1]^{\mathbb{N}})$, it follows that $F(\widetilde{\mathbb{M}}_{\mathbb{E}}) \in \mathcal{B}(\widetilde{\mathbb{M}}_{[0,1]}^{\mathbb{N}})$. We have therefore shown the following proposition.

2.10 Proposition *If $\mathbb{E}$ is a metrizable Lusin space, then so is $\widetilde{\mathbb{M}}_{\mathbb{E}}$.*

This particular compact metrizable space $\widetilde{\mathbb{M}}_{[0,1]}^{\mathbb{N}}$ is useful in obtaining convergence determining classes for $\Pi \triangleq \mathcal{P}(\mathbb{M}_{\mathbb{E}})$. We have only to find a sequence of continuous functions (g_k) which is dense in $C(\mathbb{M}_{[0,1]}^{\mathbb{N}}) \cap \{g\colon \mathbb{M}_{[0,1]}^{\mathbb{N}} \to [0,1]\}$. But we have seen that the topology on $\mathbb{M}_{[0,1]}^{\mathbb{N}}$ is generated by $\{v \mapsto \int_0^r v_i(s)\, ds\colon r \in \mathbb{Q}_+, i \in \mathbb{N}\}$. Let us denote $g_{r,i}(v) = \int_0^r v_i(s)\, ds$. The class $\{g_{r,i}\}_{r \in \mathbb{Q}_+, i \in \mathbb{N}}$ consists of continuous functions and separates points of $\mathbb{M}_{[0,1]}^{\mathbb{N}}$. Thus the algebra generated by the $g_{r,i}$ and 1, which consists of the linear combinations of products $\prod_{k=1}^m g_{r_k,i_k}$, is dense in $C(\mathbb{M}_{[0,1]}^{\mathbb{N}})$. Hence translating this to $\mathbb{M}_{\mathbb{E}}$ by means of F, one finds that

$$\left\{ \prod_{k=1}^m \int_0^{r_k} f_{j_k}(w(s))\, ds\colon m \in \mathbb{N},\ r_k \in \mathbb{Q}_+,\ f_{j_k} \in \{f_j\},\ 1 \le k \le m \right\}$$

and

$$\left\{ \prod_{k=1}^m \int_0^{r_k} f_{j_k}(w(s))\, \lambda(ds)\colon m \in \mathbb{N},\ r_k \in \mathbb{Q}_+,\ f_{j_k} \in \{f_j\},\ 1 \le k \le m \right\}$$

are convergence determining. Furthermore, since $\{g_{r,i}\}$ separates points, it follows from the inversion for Laplace transforms that $\{v \mapsto \int_0^\infty e^{-\lambda s} v_i(s)ds: \lambda \in \mathbb{Q}_+,\ i \in \mathbb{N}\}$ also separates points of $\mathbb{M}_{[0,1]}^{\mathbb{N}}$. The functions are clearly continuous. Hence

$$\left\{ \prod_{k=1}^{m} \int_0^\infty e^{-\lambda_k s} f_{j_k}(w(s))\, ds : m \in \mathbb{N},\ \lambda_k \in \mathbb{Q}_+,\ f_{j_k} \in \{f_j\},\ 1 \le k \le m \right\}$$

is convergence determining.

The following lemma, due basically to Knight (1992) and Kurtz (1991), gives us a way of picking up a function from each equivalence class in a measurable fashion.

2.11 Lemma *There exists a $\mathcal{B}(\widetilde{\mathbb{M}}_{\mathbb{E}}) \otimes \mathcal{B}(\mathbb{R}_+)/\mathcal{E}$ measurable mapping G from $\widetilde{\mathbb{M}}_{\mathbb{E}} \times \mathbb{R}_+$ into $\mathbb{E}$ such that $G(\widetilde{w}, \bullet) \in \widetilde{w}$ for all $\widetilde{w} \in \widetilde{\mathbb{M}}_{\mathbb{E}}$.*

PROOF. For $f \in b\mathcal{E}$, the mapping $\widetilde{x} \mapsto \int_a^b f(x(t))\, dt$ is well-defined since the value is the same for any representative $x \in \widetilde{x}$. Also, the mapping $(\widetilde{x}, t) \mapsto \frac{1}{h} \int_t^{t+h} f(x(s))\, ds$ is continuous for $f \in C_b(\mathbb{E})$ by the bounded convergence theorem. Since $\sigma(C_b(\mathbb{E})) = \mathcal{E}$ for any metric space $\mathbb{E}$, by monotone class argument, the above mapping is $\mathcal{B}(\widetilde{\mathbb{M}}_{\mathbb{E}}) \otimes \mathcal{B}(\mathbb{R}_+)$ measurable for $f \in b\mathcal{E}$. Thus the mapping

$$(\widetilde{x}, t) \mapsto y_f(\widetilde{x}, t) \triangleq \limsup_{n \to \infty} n \int_t^{t+\frac{1}{n}} f(x(s))\, ds$$

is $\mathcal{B}(\widetilde{\mathbb{M}}_{\mathbb{E}}) \otimes \mathcal{B}(\mathbb{R}_+)$ measurable for $f \in b\mathcal{E}$. By Lebesgue's density theorem, we have $y_f(\widetilde{x}, \bullet) \in f(\widetilde{x}(\bullet))$.

Since $\mathbb{E}$ is a metrizable Lusin space, there exists a 1-1 continuous mapping $h : \mathbb{E} \to [0,1]^{\mathbb{N}}$ such that $\mathbb{E}$ and $h(\mathbb{E})$ are homeomorphic and $h(\mathbb{E})$ is a Borel subset of $[0,1]^{\mathbb{N}}$ (see Dellacherie-Meyer, III.20). Let $e_0 \in \mathbb{E}$ be arbitrary but fixed, and set

$$g(y) \triangleq \begin{cases} h^{-1}(y), & \text{if } y \in h(\mathbb{E}); \\ e_0, & \text{otherwise}, \end{cases}$$

Then $g : [0,1]^{\mathbb{N}} \to \mathbb{E}$ is $\mathcal{B}([0,1]^{\mathbb{N}})/\mathcal{E}$ measurable. Write $h(e) = (h_1(e), h_2(e), \ldots)$ and define $y_i : \widetilde{\mathbb{M}}_{\mathbb{E}} \times \mathbb{R}_+ \to [0,1]$ and $G : \widetilde{\mathbb{M}}_{\mathbb{E}} \times \mathbb{R}_+ \to \mathbb{E}$ by

$$y_i(\widetilde{x}, t) \triangleq \limsup_{n \to \infty} n \int_t^{t+\frac{1}{n}} h_i(x(s))\, ds, \quad i \in \mathbb{N},$$

$$G(\widetilde{x}, t) \triangleq g(y_1(\widetilde{x}, t), y_2(\widetilde{x}, t), \ldots).$$

It is clear from the above argument that G is $\mathcal{B}(\widetilde{\mathbb{M}}_{\mathbb{E}})\otimes\mathcal{B}(\mathbb{R}_+)/\mathcal{E}$ measurable, and $G(\widetilde{x},\bullet) \in \widetilde{x}$ for all $\widetilde{x} \in \widetilde{\mathbb{M}}_{\mathbb{E}}$. ■

We state the following compactness criteria in $\mathbb{M}_{\mathbb{E}}$, which is mentioned earlier and is due to Kurtz (1991),

2.12 Proposition *Assume that $\mathbb{E}$ is Polish with a metric d with respect to which $\mathbb{E}$ is complete. A subset $\Gamma \subset \mathbb{M}_{\mathbb{E}}$ is relatively compact if and only if the following two conditions hold.*

(i) For every $\epsilon > 0$ and $T > 0$, there exists a compact subset K of $\mathbb{E}$ such that

$$\sup_{w\in\Gamma} m(t \leq T : w(t) \notin K) \leq \epsilon.$$

(ii) For every $T > 0$,

$$\lim_{h\to 0} \sup_{w\in\Gamma} \int_0^T 1 \wedge d(w(t+h), w(t))\, dt = 0.$$

It is easy to see that the conditions (i) and (ii) are equivalent to the following in terms of the measure λ.

(i') For every $\epsilon > 0$, there exists a compact subset K of $\mathbb{E}$ such that

$$\sup_{w\in\Gamma} \lambda(t : w(t) \notin K) \leq \epsilon.$$

(ii') $$\lim_{h\to 0} \sup_{w\in\Gamma} \int_0^\infty 1 \wedge d(w(t+h), w(t))\, \lambda(dt) = 0.$$

2.13 Functionals on $\mathbb{M}_{\mathbb{E}}$. Let us give a few examples of functionals on $\mathbb{M}_{\mathbb{E}}$.

(i) $S(w) \triangleq \operatorname{ess\,sup}_t |w(t)|$, *when* $\mathbb{E} = \mathbb{R}$. The set $\{w : S(w) \leq c\}$ is closed; if (w_n) satisfies $|w_n(t)| \leq c$, a.e. for all n and $w_n \to w$, then take a subsequence (n') for which $w_{n'} \to w$, a.e. It is then clear that $|w(t)| \leq c$, a.e. Thus S is lower semicontinuous (this also shows that S is measurable). S is, however, not continuous. For example, take $w_n(t) = \mathbf{1}_{[0,1/n]}(t)$. This converges to $w(t) = 0$, a.e. and $S(w) = 0$, while $S(w_n) = 1$ for all n. Hence $S(w_n) \not\to S(w)$ although $w_n \to w$ in measure. When $\mathbb{E}$ is a Banach space, the same is true with the absolute value replaced by the norm.

(ii) $T_B(w) \triangleq \inf\{t: \int_0^t \mathbf{1}_B(w(s))\,ds > 0\}$, $B \in \mathcal{E}$. This is the first time a function w has spent strictly positive time in the set B, and if we replace w by a stochastic process X, then it is called the *essential debut*, or *essential entry time* (see Dellacherie and Meyer (1975), IV.39). Note that for every $t \geq 0$, $\{w: T_B(w) \geq t\} = \{w: \int_0^t \mathbf{1}_B(w(s))\,ds = 0\}$. $w \mapsto \int_0^t \mathbf{1}_B(w(s))\,ds$ is measurable; $w \mapsto \int_0^t f(w(s))\,ds$ is continuous for $f \in C_b(\mathbb{E})$ and hence measurable. Then use the monotone class theorem. It B is open, then the equality $\{w: T_B(w) \geq t\} = \{w: m\{s \leq t: w(s) \in B^{\mathbf{c}}\} = t\}$ shows that this set is closed (use a subsequence converging a.e.). Thus when B is open, T_B is upper semicontinuous. But it is not continuous in general (even when B is open); for example, consider

$$w_n(t) = \begin{cases} -nt+1, & \text{if } 0 \leq t \leq 1/n; \\ 0, & \text{otherwise.} \end{cases}$$

(w_n) converges in measure to $w(t) \equiv 0$. Take $B = (\frac{1}{2}, \infty)$. Obviously $T_B(w_n) = 0$ for all n, but $T_B(w) = \infty$.

(iii) $I_f^t(w) \triangleq \int_0^t f(w(s))\,ds$, $f \in b\mathcal{E}$, $t \in \mathbb{R}_+$. This is clearly continuous when $f \in C_b(\mathbb{E})$. As a matter of fact, we have seen (Proposition 2.9) that $w_n \to w$ in measure if and only if $I_f^r(w_n) \to I_f^r(w)$ for every $r \in \mathbb{Q}_+$ and $f \in C_b(\mathbb{E})$. If we define, as in Knight (1992), the sojourn measures of w by

$$\mu_t(A, w) \triangleq \int_0^t \mathbf{1}_A(w(s))\,ds, \quad A \in \mathcal{E},\ t > 0,$$

then the above means that $w_n \to w$ in measure if and only if $\mu_r(\bullet, w_n) \xrightarrow{w} \mu_r(\bullet, w)$ for all $r \in \mathbb{Q}_+$.

(iv) Let $t \in \mathbb{R}_+$, and let $\tau: 0 = t_0 < t_1 < \cdots < t_n = t$ be a finite subdivision of $[0, t]$. For two disjoint open sets G_1 and G_2 in $\mathbb{E}$, define a positive integer $N_\tau^{G_1,G_2}(w)$ by the following: $N_\tau^{G_1,G_2}(w) \geq k$ if and only if there exist $0 \leq t_{i_1}^1 < t_{i_1}^2 < t_{i_2}^1 < t_{i_2}^2 < \cdots < t_{i_k}^1 < t_{i_k}^2 < t$ such that

$$\int_{t_{i_j}^1}^{t_{i_j}^1+1} \mathbf{1}_{G_1}(w(s))\,ds > 0 \quad \text{and} \quad \int_{t_{i_j}^2}^{t_{i_j}^2+1} \mathbf{1}_{G_2}(w(s))\,ds > 0,\ j = 1, \ldots, k.$$

The set $\{w: N_\tau^{G_1,G_2}(w) \geq k\} = \{w: N_\tau^{G_1,G_2}(w) > k-1\}$ is open; suppose that w is in this set. Then there are $0 \leq t_{i_1}^1 < t_{i_1}^2 < t_{i_2}^1 < t_{i_2}^2 < \cdots < t_{i_k}^1 < t_{i_k}^2 < t$ satisfying the above condition. Take $\epsilon > 0$ satisfying

$$\epsilon < \min_{1\leq j\leq k,\ \ell=1,2} \int_{t_{i_j}^\ell}^{t_{i_j+1}^\ell} \mathbf{1}_{G_\ell}(w(s))\, ds,$$

and consider

$$\Gamma \triangleq \left\{ v: \int_{t_{i_j}^\ell}^{t_{i_j+1}^\ell} \mathbf{1}_{G_\ell}(v(s))\, ds > \int_{t_{i_j}^\ell}^{t_{i_j+1}^\ell} \mathbf{1}_{G_\ell}(w(s))\, ds - \epsilon,\ \ell = 1,2,\ 1 \leq j \leq k \right\}.$$

This set is an open neighborhood of w because the topology of $\mathbb{M}_\mathbb{E}$ is the weak topology of the sojourn measures. And by construction, every element v in Γ satisfies $N_\tau^{G_1,G_2}(v) \geq k$. Hence $\{w: N_\tau^{G_1,G_2}(w) \geq k\}$ is open, so $N_\tau^{G_1,G_2}(w)$ is lower semicontinuous. Therefore if we define

$$N_t^{G_1,G_2}(w) \triangleq \sup_\tau N_\tau^{G_1,G_2}(w),$$

where the supremum is taken over all subdivision τ of $[0,t]$, then $N_t^{G_1,G_2}(w)$ is lower semicontinuous. $N_t^{G_1,G_2}(w)$ is the number of essential passages from G_1 to G_2 by w on $[0,t]$ introduced in Rebolledo (1987).

3. Measurable Processes and Their Convergence in Law

In Section 3.1, we discuss conditions for the existence of measurable modifications, and prove some properties of probability measures on $\mathbb{M}_\mathbb{E}(\mathbb{T}, \mathcal{T}, \nu)$. Their connection with almost equivalence is also mentioned. Necessary and sufficient conditions for convergence in law of measurable processes are established in Section 3.2.

3.1 Probability Measures on the Space of Measurable Functions

Let $X = (X_t)_{t\in\mathbb{T}}$ be a stochastic process on $(\Omega, \mathcal{F}, P)$ with values in $(\mathbb{E}, \mathcal{E})$. We assume throughout this section that $\mathbb{E}$ is a metrizable Lusin space and that $\mathcal{E} = \mathcal{B}(\mathbb{E})$. Suppose also that to $\mathbb{T}$ is associated a σ-field $\mathcal{T}$, so that $(\mathbb{T}, \mathcal{T})$ is a measurable space. Recall that a process X is said to be measurable if the mapping $(t,\omega) \mapsto X_t(\omega)$ is $\mathcal{T} \otimes \mathcal{F}/\mathcal{E}$ measurable.

Our starting point is a measurable process, hence it is important to know when a given process has a measurable modification. Chung and Doob (1965) obtained a necessary and sufficient condition for the existence of measurable modification for $\mathbb{T} = \mathbb{R}$ (their method of proof is attributed to P.-A. Meyer), and Cohn (1972) generalized it slightly. The condition is in fact valid for more general parameter sets and state spaces than they assumed. To state the result we need a few definitions. Let $\widetilde{\mathbb{M}}_{\mathbb{E}}(\Omega, \mathcal{F}, P)$ be the space of equivalence classes of $\mathbb{E}$-valued random variables on $(\Omega, \mathcal{F}, P)$ with the topology of convergence in P-probability and $\dot{X}_t$ the equivalence class containing X_t. A function which takes at most a countable number of values is called an elementary function. We consider the following two conditions:

(1) The mapping $t \mapsto \dot{X}_t$ is $\mathcal{T}/\mathcal{B}(\widetilde{\mathbb{M}}_{\mathbb{E}}(\Omega, \mathcal{F}, P))$ measurable and separably-valued;

(2) The mapping $t \mapsto \dot{X}_t$ is the uniform limit of a sequence of $\widetilde{\mathbb{M}}_{\mathbb{E}}(\Omega, \mathcal{F}, P)$-valued measurable elementary functions on $\mathbb{T}$.

It is quite easy to show that (1) and (2) are equivalent. The result can now be stated as follows.

3.1 Theorem *X has a measurable modification if and only if the condition (1), or equivalently (2) holds.*

PROOF. (*if*) For each $n \in \mathbb{N}$, we can choose a collection $\{C_{n,i},\ i \in \mathbb{N}\}$ of disjoint sets in $\mathcal{B}(\widetilde{\mathbb{M}}_{\mathbb{E}}(\Omega, \mathcal{F}, P))$ with diameter less than 2^{-n} that covers the range of the mapping $t \mapsto \dot{X}_t$, where the diameter refers to the metric $\rho_\lambda(v, w) = \inf\{\epsilon > 0 \colon \lambda(t \colon d(v(t), w(t)) > \epsilon) < \epsilon\}$. Then the sets $B_{n,i} = \{t \colon \dot{X}_t \in C_{n,i}\}$, $i \in \mathbb{N}$ form a $\mathcal{T}$-partition of $\mathbb{T}$. For each $n, i \in \mathbb{N}$, select $t_{n,i} \in B_{n,i}$ and define

$$X_t^n(\omega) = X_{t_{n,i}}(\omega) \qquad \text{for } t \in B_{n,i}.$$

X^n is clearly a measurable process. Note that if $s, t \in B_{n,i}$, we have $P(d(X_s, X_t) > 2^{-n}) < 2^{-n}$, where d is a metric compatible with the topology of $\mathbb{E}$. It then follows that $P(d(X_t^n, X_t) > 2^{-n}) < 2^{-n}$ for each $t \in \mathbb{T}$ and $n \in \mathbb{N}$. By Borel-Cantelli lemma, we deduce that

$$P\left(\lim_{n \to \infty} X_t^n = X_t\right) = 1 \qquad \text{for every } t \in \mathbb{T}.$$

The set $D \triangleq \{(t, \omega) \colon X_t^n(\omega) \text{ converges as } n \to \infty\}$ belongs to $\mathcal{T} \otimes \mathcal{F}$ since $\mathbb{E}$ is metrizable Lusin (see Dellacherie and Meyer (1975), I.16). Let e_0 be an arbitrary but fixed element of $\mathbb{E}$. Then the process Y^n defined by

$$Y_t^n(\omega) \triangleq \begin{cases} X_t^n(\omega), & \text{if } (t,\omega) \in D; \\ e_0, & \text{otherwise,} \end{cases}$$

is measurable, and $Y_t^n(\omega)$ converges for every t and ω. Put $Y_t(\omega) \triangleq \lim_{n\to\infty} Y_t^n(\omega)$. Y is measurable as the pointwise limit of measurable processes Y^n, and satisfies $P(X_t = Y_t) = 1$ for every $t \in \mathbb{T}$.

(*only if*) Let $\mathcal{H}$ be the family of $\mathbb{E}$-valued processes satisfying the condition (1). One can verify without difficulty that $\mathcal{H}$ is closed under sequential pointwise convergence. Let $\mathcal{C}$ be the family of mappings from $\mathbb{T} \times \Omega$ into $\mathbb{E}$ which take only finitely many values, each value occuring on a finite union of measurable rectangles. Clearly we have $\mathcal{C} \subset \mathcal{H}$. Note that any measurable function is the uniform limit of a sequence of elementary functions (for separable and metrizable $\mathbb{E}$; see Dellacherie and Meyer (1975), I.17) and that an elementary function is the pointwise limit of a sequence of simple functions. Using these facts, it is fairly easy to show that the set of $\mathcal{T} \otimes \mathcal{F}/\mathcal{E}$ measurable functions coincides with the smallest collection of functions on $\mathbb{T} \times \Omega$ into $\mathbb{E}$ that contains $\mathcal{C}$ and is closed under sequential pointwise convergence (see Lemma 2 of Cohn (1972)). It follows that $\mathcal{H}$ contains all $\mathbb{E}$-valued measurable processes. ■

The next theorem, due to Hoffmann-Jørgensen (1973), gives another necessary and sufficient condition which depends only on two-dimensional distributions of the process.

3.2 Theorem *Let $Q(s,t)(A) = P[(X_s, X_t) \in A]$ for $s, t \in \mathbb{T}$ and $A \in \mathcal{E} \otimes \mathcal{E}$. Then X has a measurable modification if and only if the following two conditions holds.*

(1) For every $G, H \in \mathcal{E}$ and every $s \in \mathbb{T}$, the mapping $t \mapsto Q(s,t,G \times H)$ from $\mathbb{T}$ into $[0,1]$ is $\mathcal{T}/\mathcal{B}[0,1]$ measurable.

(2) There exists a countable set $T_0 \subset \mathbb{T}$ such that for all $t \in \mathbb{T}$, we can find a sequence $(t_n) \subset T_0$ with $Q(t_n, t) \xrightarrow{w} Q(t,t)$, $n \to \infty$.

For a proof, see the above-mentioned paper. We also remark that Skorokhod (1980) and Engelbert (1984) discuss similar conditions for the existence of measurable, progressively measurable, optional and predictable modifications.

Now let $X = (X_t)_{t\in\mathbb{T}}$ be a measurable process on $(\Omega, \mathcal{F}, P)$ with values in $(\mathbb{E}, \mathcal{E})$. We assume that $(\mathbb{T}, \mathcal{T}, \nu)$ is a σ-finite measure space. A finite measure equivalent to ν, whose existence is guaranteed by Lemma 2.1, is denoted by λ. The notations defined in Section 2 will be used without mention. So, for example, $\mathbb{M}_{\mathbb{E}}(\lambda)$ is the space of $\mathbb{E}$-valued measurable functions on $(\mathbb{T}, \mathcal{T})$ equipped with the topology of convergence in λ-measure, and $\widetilde{\mathbb{M}}_{\mathbb{E}}(\lambda)$ is the corresponding space of equivalence classes. Since X is measurable, the paths $X_\bullet(\omega)$ belong to $\mathbb{M}_{\mathbb{E}}(\lambda)$ for each $\omega \in \Omega$. Let us denote by $\widetilde{X}(\omega)$ the equivalence class containing $X_\bullet(\omega)$. We have

3.3 Lemma *If $\mathcal{T}$ is countably generated up to ν-null sets, then the mapping $\omega \mapsto \widetilde{X}(\omega)$ is $\mathcal{F}/\mathcal{B}(\widetilde{\mathbb{M}}_{\mathbb{E}}(\lambda))$ measurable.*

PROOF. By Proposition 2.4, $\widetilde{\mathbb{M}}_{\mathbb{E}}(\lambda)$ is separable. Thus we have only to show that the inverse images of the balls under the mapping is in $\mathcal{F}$. But we have

$$\left\{\omega\colon \int_{\mathbb{T}} 1 \wedge d(X_t(\omega), w(t))\, \lambda(dt) < c\right\} \in \mathcal{F}$$

by Fubini's theorem. ∎

REMARK. The lemma remains true if we only assume that $\mathbb{E}$ is separable and metrizable.

Thus the mapping $\omega \mapsto \widetilde{X}(\omega)$ induces a probability law on $\widetilde{\mathbb{M}}_{\mathbb{E}}(\lambda)$. Conversely, we have the following.

3.4 Lemma *Suppose that $\mathcal{T}$ is countably generated up to ν-null sets. Then for any $\widetilde{\mathbb{M}}_{\mathbb{E}}(\lambda)$-valued random variable $\widetilde{X}$, there is a measurable process X such that $X_\bullet(\omega) \in \widetilde{X}(\omega)$.*

PROOF. We construct a mapping which selects a function (representative) from each equivalence class in a measurable way. First let us assume $\mathbb{E} = [0, 1]$. Consider $\mathbb{M}_{[0,1]} = \mathbb{M}_{[0,1]}(\lambda)$ as a subset of $\mathbb{L}^2(\mathbb{T}, \mathcal{T}, \lambda)$. Since $\mathcal{T}$ is countably generated, $\mathbb{L}^2(\mathbb{T}, \mathcal{T}, \lambda)$ is separable. Thus there exists a countable orthonormal basis $(\widetilde{\phi}_j)_{j\in\mathbb{N}}$ for $\mathbb{L}^2(\mathbb{T}, \mathcal{T}, \lambda)$. Pick any representative ϕ_j for each $\widetilde{\phi}_j$ and fix them. If $w \in \mathbb{M}_{[0,1]}$, then we have a representation

$$w(t) = \sum_{j=1}^{\infty} \langle w, \phi_j\rangle \phi_j(t),$$

where the limit is in $\mathbb{L}^2(\mathbb{T}, \mathcal{T}, \lambda)$, the equality is λ-a.e. and $\langle w, \phi_j \rangle = \int_{\mathbb{T}} w(t)\phi_j(t)\lambda(dt)$. Note that the value of $\langle w, \phi_j \rangle$ will remain the same when w is replaced by any $w' \in \widetilde{w}$. So it makes sense to write $\langle \widetilde{w}, \phi_j \rangle$. Put

$$\gamma_n(\widetilde{w}, t) \triangleq \sum_{j=1}^{n} \langle \widetilde{w}, \phi_j \rangle \phi_j(t).$$

$\widetilde{w} \mapsto \langle \widetilde{w}, \phi_j \rangle$ is $\mathcal{B}(\mathbb{M}_{[0,1]})$ measurable since it is continuous. Hence $(\widetilde{w}, t) \mapsto \gamma_n(\widetilde{w}, t)$ is $\mathcal{B}(\mathbb{M}_{[0,1]}) \otimes \mathcal{T}$ measurable. Define $n_k(\widetilde{w})$ to be the smallest positive integer n for which

$$\sup_{m>n} \lambda \left\{ t \in \mathbb{T} \colon |\gamma_n(\widetilde{w}, t) - \gamma_m(\widetilde{w}, t)| > \frac{1}{k^2} \right\} \leq \frac{1}{k^2}$$

holds. Then

$$\{\widetilde{w} \colon n_k(\widetilde{w}) \leq n\} = \bigcap_{m=n+1}^{\infty} \left\{ w \colon \lambda \left\{ t \in \mathbb{T} \colon |\gamma_n(\widetilde{w}, t) - \gamma_m(\widetilde{w}, t)| > \frac{1}{k^2} \right\} \leq \frac{1}{k^2} \right\},$$

and note that $\widetilde{w} \mapsto \lambda\{t \in \mathbb{T} \colon |\gamma_n(\widetilde{w}, t) - \gamma_m(\widetilde{w}, t)| > 1/k^2\}$ is $\mathcal{B}(\mathbb{M}_{[0,1]})$ measurable by Fubini's theorem. This shows that $\widetilde{w} \mapsto n_k(\widetilde{w})$ is $\mathcal{B}(\mathbb{M}_{[0,1]})$ measurable. Now set

$$g(\widetilde{w}, t) \triangleq \begin{cases} \limsup_{k \to \infty} \gamma_{n_k(\widetilde{w})}(\widetilde{w}, t), & \text{if } \limsup_{k \to \infty} \gamma_{n_k(\widetilde{w})}(\widetilde{w}, t) \in [0, 1]; \\ 0, & \text{otherwise.} \end{cases}$$

It is evident from the construction that $g(\widetilde{w}, \bullet) \in \widetilde{w}$ and $(\widetilde{w}, t) \mapsto g(\widetilde{w}, t)$ is $\mathcal{B}(\mathbb{M}_{[0,1]}) \otimes \mathcal{T}$ measurable.

As a measurable space, $(\mathbb{E}, \mathcal{E})$ is of course a measurable Lusin space. Thus, by Kuratowski's theorem (see Dellacherie and Meyer (1975), III.20), it is measurably isomorphic to a Borel subset of $[0, 1]$ endowed with its Borel σ-field . So let h be an isomorphism of $(\mathbb{E}, \mathcal{E})$ into $[0, 1]$ with $h(\mathbb{E}) \in \mathcal{B}[0, 1]$. Let $e_0 \in \mathbb{E}$ be arbitrary but fixed. Define

$$G(\widetilde{w}, t) \triangleq \begin{cases} h^{-1}(g(\widetilde{h \circ w}, t)), & \text{if } g(\widetilde{h \circ w}, t) \in h(\mathbb{E}); \\ e_0 & \text{otherwise.} \end{cases}$$

Then one can easily check that $G(\widetilde{w}, \bullet) \in \widetilde{w}$ and $(\widetilde{w}, t) \mapsto G(\widetilde{w}, t)$ is $\mathcal{B}(\mathbb{M}_{\mathbb{E}}(\lambda)) \otimes \mathcal{T}$ measurable. It then follows that for any $\widetilde{\mathbb{M}}_{\mathbb{E}}$-valued random variable $\widetilde{X}$, $X_t(\omega) \triangleq G(\widetilde{X}(\omega), t)$ defines a measurable process and $X_\bullet(\omega) \in \widetilde{X}(\omega)$. ■

The above proof is inspired by Skorokhod (1980) and Engelbert (1984).

The notion of probability law on $\widetilde{\mathbb{M}}_{\mathbb{E}}(\lambda)$ induced by X may not be easily understood, but the following theorem clarifies its meaning in terms of the finite-dimensional distributions of the process. For processes X and Y and $S \subset \mathbb{T}$, we write $X \overset{\mathrm{fd}(S)}{=} Y$ if $(X_{t_1}, \ldots, X_{t_k})$ and $(Y_{t_1}, \ldots, Y_{t_k})$ have the same law in $\mathbb{E}^k$ for all $t_i \in S$, $1 \le i \le k$, $k \in \mathbb{N}$.

3.5 Theorem *Let X and Y be $\mathbb{E}$-valued measurable processes, and suppose that $\mathcal{T}$ is countably generated up to ν-null sets. Then X and Y induce the same laws on $\widetilde{\mathbb{M}}_{\mathbb{E}}(\lambda)$ if and only if there exists an $S \in \mathcal{T}$ with $\lambda(S^{\mathbf{c}}) = 0$ such that $X \overset{\mathrm{fd}(S)}{=} Y$.*

PROOF. (*if*) Let $\mathcal{G} = \{g \in b(\mathcal{T} \otimes \mathcal{E}) \colon \{g(t,\cdot) \colon t \in \mathbb{T}\}$ is uniformly equicontinuous on $\mathbb{E}\}$. Also set

$$\Phi_{B,g}(w) \triangleq \int_B g(t, w(t))\, \lambda(dt), \quad w \in \mathbb{M}_{\mathbb{E}}, \; B \in \mathcal{T} \text{ and } g \in \mathcal{G}.$$

Suppose that $w_n \to w$ in λ-measure. For any $\epsilon, \eta > 0$, choose $\delta > 0$ such that $|g(t,x) - g(t,y)| < \epsilon$ whenever $d(x,y) < \delta$ and $t \in \mathbb{T}$. We can find n_0 such that $\lambda(t \colon d(w_n(t), w(t)) \ge \delta) < \eta$ for all $n \ge n_0$. Then for all $n \ge n_0$,

$$\begin{aligned}&\lambda(t \colon |g(t, w_n(t)) - g(t, w(t))| > \epsilon) \\ &\le \lambda(t \colon d(w_n(t), w(t)) \ge \delta) + \lambda(t \colon |g(t, w_n(t)) - g(t, w(t))| > \epsilon,\; d(w_n(t), w(t)) < \delta) < \eta.\end{aligned}$$

Hence $g(t, w_n(t))$ converges in λ-measure to $g(t, w(t))$, and it follows from the dominated convergence theorem that $\Phi_{B,g}(w_n) \to \Phi_{B,g}(w)$. Since $\mathbb{M}_{\mathbb{E}}(\lambda)$ is a metric space, this implies that $\Phi_{B,g}(w)$ is continuous. It is obviously bounded, so that we have $\Phi_{B,g}(w) \in C_b(\widetilde{\mathbb{M}}_{\mathbb{E}}))$.

Let $\mathcal{A}$ be the subalgebra in $C_b(\widetilde{\mathbb{M}}_{\mathbb{E}})$ generated by 1 and the $\Phi_{B,g}$, $B \in \mathcal{T}$, $g \in \mathcal{G}$. Suppose $\widetilde{w} \ne \widetilde{v}$ be in $\widetilde{\mathbb{M}}_{\mathbb{E}}$. Then $\lambda(t \colon d(w(t), v(t)) > 0) > 0$ for any $w \in \widetilde{w}$ and $v \in \widetilde{v}$. Put $B = \{t \colon d(w(t), v(t)) > 0\}$ and $g(t,x) = d(w(t), x) \wedge 1$. Clearly $B \in \mathcal{T}$. Since $|d(w(t), x) - d(w(t), y)| \le d(x,y)$, $\{g(t,\cdot) \colon t \in \mathbb{T}\}$ is uniformly equicontinuous. It is obvious that g is bounded and $\mathcal{T} \otimes \mathcal{E}$ measurable, so $g \in \mathcal{G}$. Moreover,

$$\Phi_{B,g}(w) = \int_B d(w(t), w(t))\, \lambda(dt) = 0, \qquad \Phi_{B,g}(v) = \int_B d(w(t), v(t))\, \lambda(dt) > 0.$$

Thus $\Phi_{B,g}(\widetilde{w}) \ne \Phi_{B,g}(\widetilde{v})$, that is, $\mathcal{A}$ separates points. By a generalized version of Stone-Weierstrass theorem, $\mathcal{A}$ is dense in $C_b(\widetilde{\mathbb{M}}_{\mathbb{E}})$ with respect to the strict topology (see Fremlin et al.(1972) for this result). It then follows that $\mathcal{A}$ separates measures on $\widetilde{\mathbb{M}}_{\mathbb{E}}$. Thus we need to show that for each $\Phi \in \mathcal{A}$, it holds that $\int \Phi(X)dP = \int \Phi(Y)dP$. But by Fubini's theorem and $X \stackrel{\mathrm{fd}(S)}{=} Y$,

$$E\left[\prod_{i=1}^{k} \Phi_{B_i,g_i}(X)\right] = \int_{B_1} \cdots \int_{B_k} E[g_1(t_1, X_{t_1}) \cdots g_k(t_k, X_{t_k})] \,\lambda(dt_1) \cdots \lambda(dt_k)$$
$$= \int_{B_1} \cdots \int_{B_k} E[g_1(t_1, Y_{t_1}) \cdots g_k(t_k, Y_{t_k})] \,\lambda(dt_1) \cdots \lambda(dt_k) = E\left[\prod_{i=1}^{k} \Phi_{B_i,g_i}(Y)\right]$$

Since each element of $\mathcal{A}$ is a finite linear combination of finite products of some $\Phi_{B,g}$'s, the desired conclusion follows.

(*only if*) We will use the function G constructed in the proof of Lemma 3.4. Denote as earlier by $\widetilde{X}(\omega)$ and $\widetilde{Y}(\omega)$ the equivalence classes containing $X_\bullet(\omega)$ and $Y_\bullet(\omega)$ respectively. Then for each ω, $G(\widetilde{X}(\omega), \bullet) = X_\bullet(\omega)$, λ-a.e., so by an application of Fubini's theorem, there is an $S_1 \in \mathcal{T}$ with $\lambda(S_1^{\mathbf{c}}) = 0$ such that $X_t(\omega) = G(\widetilde{X}(\omega), t)$, ω-a.s. for all $t \in S_1$. Similarly we can find such a set S_2 for Y. Put $S = S_1 \cap S_2$, so we have $\lambda(S^{\mathbf{c}}) = 0$. And for all $t \in S$, $X_t(\omega) = G(\widetilde{X}(\omega), t)$, ω-a.s. and $Y_t(\omega) = G(\widetilde{Y}(\omega), t)$, ω-a.s. By the assumption, $\widetilde{X}$ and $\widetilde{Y}$ have the same law on $\mathbb{M}_{\mathbb{E}}$, so for any $t_1, \ldots, t_m \in S$,

$$\left(G(\widetilde{X}(\omega), t_1), \ldots, G(\widetilde{X}(\omega), t_m)\right) \stackrel{\mathcal{L}}{=} \left(G(\widetilde{Y}(\omega), t_1), \ldots, G(\widetilde{Y}(\omega), t_m)\right).$$

It therefore follows that $(X_{t_1}, \ldots, X_{t_m})$ and $(Y_{t_1}, \ldots, Y_{t_m})$ have the same law on $\mathbb{E}^m$ for any $t_1, \ldots, t_m \in S$. ■

The above theorem connects the notion of probability measure on $\mathbb{M}_{\mathbb{E}}(\lambda)$ with that of almost equivalence. Let X and Y be two $\mathbb{E}$-valued measurable processes defined on possibly different probability spaces. Then X and Y are said to be ν-almost equivalent if for every finite system of pairs (ϕ_i, g_i), $1 \le i \le n$, where ϕ_i is positive and integrable function on $\mathbb{T}$ and g_i is bounded $\mathcal{E}/\mathcal{B}(\mathbb{R})$ measurable function on $\mathbb{E}$, the random vectors

$$\left(\int_{\mathbb{T}} g_1(X_t)\phi_1(t)\,\nu(dt), \ldots, \int_{\mathbb{T}} g_n(X_t)\phi_n(t)\,\nu(dt)\right)$$

and

$$\left(\int_{\mathbb{T}} g_1(Y_t)\phi_1(t)\,\nu(dt), \ldots, \int_{\mathbb{T}} g_n(Y_t)\phi_n(t)\,\nu(dt)\right)$$

have the same law on $\mathbb{R}^n$. It is then easy to see from the proof of Theorem 3.5 that X and Y induce the same law on $\mathbb{M}_{\mathbb{E}}(\lambda)$ if and only if X and Y are ν-almost equivalent.

The concepts of pseudo-path and pseudo-law have been introduced in Dellacherie and Meyer (1975), IV.35-45 for the case $\mathbb{T} = \mathbb{R}_+$, and they are closely related to the notions of equivalence classes and probability laws on $\mathbb{M}_{\mathbb{E}}(\lambda)$. They could be extended to a general parameter set $\mathbb{T}$ with a σ-field $\mathcal{T}$ and a σ-finite measure ν although the choice of ν, which should play the canonical role, may not be evident. We do not go into details here, and the interested reader is referred to the book mentioned above.

3.2 Convergence in Law

We continue to use the notations introduced in the preceding sections. Let again $X = (X_t)_{t\in\mathbb{T}}$ and $X^n = (X_t^n)_{t\in\mathbb{T}}$, $n \in \mathbb{N}$ be measurable processes on $(\Omega, \mathcal{F}, P)$ with state space $(\mathbb{E}, \mathcal{E})$. It appears that we would need to show $X^n \xrightarrow{\text{fd}(S)} X$ for some $S \in \mathcal{T}$ with $\lambda(S^{\complement}) = 0$ and the tightness of $(X^n)_{n\in\mathbb{N}}$ in order to get $X^n \xrightarrow{\mathcal{L}} X$ in $\mathbb{M}_{\mathbb{E}}(\lambda)$. But the next theorem shows that in fact it suffices to prove $X^n \xrightarrow{\text{fd}(S)} X$ for some $S \in \mathcal{T}$ with $\lambda(S^{\complement}) = 0$. That is, tightness is unnecessary although we must know that the limiting process X is measurable. Our method of proof is quite similar to that of Cremers and Kadelka (1986).

3.6 Theorem *Let $\mathbb{E}$ be a separable metrizable space and $\mathcal{T}$ be countably generated. Suppose that $(X^n)_{n\in\mathbb{N}}$ and X are $\mathbb{E}$-valued measurable processes on $(\Omega, \mathcal{F}, P)$. If $X^n \xrightarrow{\text{fd}(S)} X$ for some $S \in \mathcal{T}$ with $\lambda(S^{\complement}) = 0$, then $X^n \xrightarrow{\mathcal{L}} X$ in $\mathbb{M}_{\mathbb{E}}(\lambda)$.*

Proof. Let $\phi_y(t, v) = d(v, y(t)) \wedge 1$ for $t \in \mathbb{T}$, $v \in \mathbb{E}$ and $y \in \mathbb{M}_{\mathbb{E}}(\lambda)$, and define $f_{\phi_y} : \widetilde{\mathbb{M}}_{\mathbb{E}}(\lambda) \to \mathbb{R}$ by

$$f_{\phi_y}(x) \triangleq \int_{\mathbb{T}} \phi_y(t, x(t))\, \lambda(dt) = \int_{\mathbb{T}} d(x(t), y(t)) \wedge 1\, \lambda(dt).$$

Put $\Phi_M = \{f_{\phi_y} : y \in \widetilde{\mathbb{M}}_{\mathbb{E}}(\lambda)\}$. Since $\widetilde{\mathbb{M}}_{\mathbb{E}}(\lambda)$ is separable, its topology is generated by the balls, i.e. the balls constitute a base for the topology. Thus the family Φ_M generates the topology of $\widetilde{\mathbb{M}}_{\mathbb{E}}(\lambda)$. It is clear that for each fixed $y \in \mathbb{M}_{\mathbb{E}}(\lambda)$, f_{ϕ_y} is continuous. By Pollard's theorem (see Pollard (1977)), it hence suffices to prove that for any $y_1, \ldots, y_m \in \mathbb{M}_{\mathbb{E}}(\lambda)$, $k \geq 1$,

$$(f_{\phi_{y_1}}(X^n), \ldots, f_{\phi_{y_m}}(X^n)) \xrightarrow{\mathcal{L}} (f_{\phi_{y_1}}(X), \ldots, f_{\phi_{y_m}}(X)) \quad \text{in } \mathbb{R}^m.$$

We have $0 \le f_{\phi_y}(X^n_\bullet(\omega)) \le \lambda(\mathbb{T})$. Let $a_1, \ldots, a_m \in \mathbb{R}^m$. By Cramer-Wold device (Billingsley (1968)), it is enough to show that $\sum_{i=1}^m a_i f_{\phi_{y_i}}(X^n) \xrightarrow{\mathcal{L}} \sum_{i=1}^m a_i f_{\phi_{y_i}}(X)$ in $\mathbb{R}$, i.e., for $g \in C_b(\mathbb{R})$,

$$\int g\left(\sum_{i=1}^m a_i f_{\phi_{y_i}}(X^n)\right) dP \to \int g\left(\sum_{i=1}^m a_i f_{\phi_{y_i}}(X)\right) dP$$

Note that $\left|\sum_{i=1}^m a_i f_{\phi_{y_i}}(X^n)\right| \le \lambda(\mathbb{T}) \sum_{i=1}^m |a_i| =: K$. Since the polynomials are uniformly dense in $C[-K, K]$ by Weierstrass theorem, we have only to consider $g(u) = u^l$ for $l \in \mathbb{N}$. Moreover, considering a general form of $\left(\sum_{i=1}^m a_i f_{\phi_{y_i}}(X^n)\right)^l$, it is enough to prove

$$\int [f_{\phi_{y_1}}(X^n)]^{k_1} \cdots [f_{\phi_{y_m}}(X^n)]^{k_m}\, dP \to \int [f_{\phi_{y_1}}(X)]^{k_1} \cdots [f_{\phi_{y_m}}(X)]^{k_m}\, dP.$$

We have, by Fubini's theorem,

$$\int \prod_{j=1}^m [f_{\phi_{y_j}}(X)]^{k_j}\, dP = \int \prod_{j=1}^m \left[\int \phi_{y_j}(t_1^j, X^n_{t_1^j})\lambda(dt_1^j) \cdots \int \phi_{y_j}(t_{k_j}^j, X^n_{t_{k_j}^j})\lambda(dt_{k_j}^j)\right] dP$$
$$= \int \cdots \int F_n(t_1^1, \ldots, t_{k_1}^1; \cdots; t_1^m, \ldots, t_{k_m}^m)\, \lambda(dt_1^1) \cdots \lambda(dt_{k_m}^m),$$

where

$$F_n(t_1^1, \ldots, t_{k_1}^1; \ldots; t_1^m, \ldots, t_{k_m}^m) = \int_\Omega \prod_{j=1}^m \prod_{i=1}^{k_j} \phi_{y_j}(t_i^j, X^n_{t_i^j})\, dP.$$

The mapping $(v_1^1, \ldots, v_{k_1}^1; \ldots; v_1^m, \ldots, v_{k_m}^m) \mapsto \prod_{j=1}^m \prod_{i=1}^{k_j} \phi_{y_j}(t_i^j, v_i^j)$ belongs to $C_b(\mathbb{E}^{\Sigma k_j})$, and we assumed that

$$(X^n_{t_1^1}, \ldots, X^n_{t_{k_1}^1}; \ldots; X^n_{t_1^m}, \ldots, X^n_{t_{k_m}^m}) \xrightarrow{\mathcal{L}} (X_{t_1^1}, \ldots, X_{t_{k_1}^1}; \ldots; X_{t_1^m}, \ldots, X_{t_{k_m}^m})$$

for $t_1^1, \ldots, t_{k_1}^1; \ldots; t_1^m, \ldots, t_{k_m}^m \in S$. Thus

$$F_n(t_1^1, \ldots, t_{k_1}^1; \ldots; t_1^m, \ldots, t_{k_m}^m)$$
$$\to F(t_1^1, \ldots, t_{k_1}^1; \ldots; t_1^m, \ldots, t_{k_m}^m) =: \int_\Omega \prod_{j=1}^m \prod_{i=1}^{k_j} \phi_{y_j}(t_i^j, X_{t_i^j})\, dP$$

for $t_1^1, \ldots, t_{k_1}^1; \ldots; t_1^m, \ldots, t_{k_m}^m \in S$. Finally, by Lebesgue's dominated convergence theorem, we obtain

$$\int \cdots \int F_n(t_1^1, \ldots, t_{k_1}^1; \ldots; t_1^m, \ldots, t_{k_m}^m)\, \lambda(dt_1^1) \cdots \lambda(dt_{k_m}^m)$$
$$\to \int \cdots \int F(t_1^1, \ldots, t_{k_1}^1; \ldots; t_1^m, \ldots, t_{k_m}^m)\, \lambda(dt_1^1) \cdots \lambda(dt_{k_m}^m),$$

which equals $\int \prod_{j=1}^m [f_{\phi_{y_j}}(X)]^{k_j}\, dP$ by Fubini's theorem as above. ■

3.7 Remark Recalling the fact that $X^n \xrightarrow{\mathfrak{L}} X$ if and only if each subsequence (n') contains a further subsequence (n'') such that $X^{n''} \xrightarrow{\mathfrak{L}} X$, we easily see that the condition of the theorem can be replaced by:

for any subsequence (n'), there exists a further subsequence (n'') and an $S \in \mathfrak{T}$ with $\lambda(S^{\complement}) = 0$ such that $X^{n''} \xrightarrow{\mathrm{fd}(S)} X$.

We now turn to the converse of Theorem 3.6 in a sense. The following result is an extension of Meyer and Zheng (1984) and Sadi (1988).

3.8 Theorem *Suppose that $\mathfrak{T}$ is countably generated up to ν-null sets and that $\mathbb{E}$ is a metrizable Lusin space. Let $(X^n)_{n\in\mathbb{N}}$ and X be $\mathbb{E}$-valued measurable processes with parameter set $\mathbb{T}$ on some $(\Omega, \mathfrak{F}, P)$. If $X^n \xrightarrow{\mathfrak{L}} X$ in $\mathbb{M}_{\mathbb{E}}$, then there exist an $S \in \mathfrak{T}$ with $\lambda(S^{\complement}) = 0$ and a subsequence (n') such that $X^{n'} \xrightarrow{\mathrm{fd}(S)} X$.*

Proof. We have $\mathfrak{L}(X^n) \xrightarrow{w} \mathfrak{L}(X)$ in $\widetilde{\mathbb{M}}_{\mathbb{E}}$, so by Skorohod-Dudley representation theorem (see Ethier and Kurtz (1986), Theorem 3.1.8), we can find, on some probability space, $\widetilde{\mathbb{M}}_{\mathbb{E}}$-valued random variables $\widetilde{Y}^n$ and $\widetilde{Y}$ such that $\mathfrak{L}(X^n) = \mathfrak{L}(\widetilde{Y}^n)$, $\mathfrak{L}(X) = \mathfrak{L}(\widetilde{Y})$ in $\widetilde{\mathbb{M}}_{\mathbb{E}}$ and $\widetilde{Y}^n \to \widetilde{Y}$ a.s. in $\widetilde{\mathbb{M}}_{\mathbb{E}}$. By Lemma 3.4, there exist measurable processes Y^n and Y for which we have $Y^n_\bullet(\omega) \in \widetilde{Y}^n(\omega)$ and $Y_\bullet(\omega) \in \widetilde{Y}(\omega)$ for all ω. Then $Y^n_\bullet(\omega) \to Y_\bullet(\omega)$ in λ-measure for almost all ω.

Since $\mathbb{E}$ is separable, there is a family $\{g_i\}_{i\in\mathbb{N}}$ of continuous bounded functions on $\mathbb{E}$ such that $e_k \to e$ in $\mathbb{E}$ if and only if $g_i(e_k) \to g_i(e)$ for every $i \in \mathbb{N}$. For example, we can take $g_i(e) = d(e_i, e)$, where d is a metric compatible with the topology of $\mathbb{E}$ and bounded by 1, and (e_i) is dense in $\mathbb{E}$ (this is just an imbedding of $\mathbb{E}$ in $[0,1]^{\mathbb{N}}$). Clearly $g_i(Y^n_\bullet(\omega)) \to g_i(Y_\bullet(\omega))$ in λ-measure for almost all ω, and since the g_i are bounded, we get

$$\int |g_i(Y_t^n(\omega) - g_i(Y_t(\omega))| \, \lambda(dt) \to 0$$

boundedly for almost all ω. Thus integrating with respect to P yields

$$\iint |g_i(Y_t^n(\omega)) - g_i(Y_t(\omega))| \, \lambda \otimes P(dt, d\omega) \to 0.$$

Then we can find a subsequence (n_k^i) such that $g_i(Y_t^{n_k^i}(\omega)) \to g_i(Y_t(\omega))$, $\lambda \otimes P$-a.e., and by the diagonal method, there exists a single subsequence (n') so that we have $g_i(Y_t^{n'}(\omega)) \to g_i(Y_t(\omega))$, $\lambda \otimes P$-a.e. for all i. It follows from Fubini's theorem that there is an $S_1 \in \mathcal{T}$ with $\lambda(S_1^{\mathbf{C}}) = 0$ such that for every $t \in S_1$, $g_i(Y_t^{n'}(\omega)) \to g_i(Y_t(\omega))$, P-a.s. for all i. Hence for every $t \in S_1$, $Y_t^{n'}(\omega) \to Y_t(\omega)$, P-a.s., and so $Y^{n'} \overset{\mathrm{fd}(S_1)}{\longrightarrow} Y$.

Since Y^n and Y induce the same laws on $\widetilde{\mathbb{M}}_{\mathbb{E}}$ as X^n and X respectively, by Theorem 3.5, we can find an $S_2 \in \mathcal{T}$ with $\lambda(S_2^{\mathbf{C}}) = 0$ such that $X^n \overset{\mathrm{fd}(S_2)}{=} Y^n$ for all $n \in \mathbb{N}$ and $X \overset{\mathrm{fd}(S_2)}{=} Y$. But then, setting $S = S_1 \cap S_2$, we obtain $X^{n'} \overset{\mathrm{fd}(S)}{\longrightarrow} X$. ■

3.9 Corollary *Let $(X^n)_{n \in \mathbb{N}}$ and X be $\mathbb{E}$-valued measurable processes with parameter set $\mathbb{R}_+$ on some $(\Omega, \mathcal{F}, P)$. For $X^n \overset{\mathcal{L}}{\to} X$ in $\mathbb{M}_{\mathbb{E}}$, it is necessary and sufficient that for any subsequence (n'), there exist a further subsequence (n'') and an $S \in \mathcal{T}$ with $\lambda(S^{\mathbf{C}}) = 0$ such that $X^{n''} \overset{\mathrm{fd}(S)}{\longrightarrow} X$.*

For a proof, combine Theorem 3.8 and Remark 3.7. This corollary generalizes the result of Sadi (1988).

4. Convergence of the Prediction Process

In Section 4.1, we shall present a short summary of the theory of prediction process following Knight (1992). Convergence in law of the prediction processes is the main topic of Section 4.2. We first consider conditions for tightness and convergence in law of general conditional distributions. A condition for finite-dimensional convergence of the prediction processes is then derived, and we show that the convergence of the prediction processes implies that of the given processes. A special case where the given processes are Markov is discussed in detail.

Let $X = (X_t)_{t\in\mathbb{R}_+}$ be a measurable process with values in $\mathbb{E}$. We assume that $\mathbb{E}$ is a metrizable Lusin space with $\mathcal{E} = \mathcal{B}(\mathbb{E})$, and let $\mathbb{M}_\mathbb{E} = \mathbb{M}_\mathbb{E}(\mathbb{R}_+, \mathcal{B}_+, m)$ where $\mathcal{B}_+ = \mathcal{B}(\mathbb{R}_+)$ and m is the Lebesgue measure. Setting $\lambda(dt) = e^{-t}dt$, we give $\mathbb{M}_\mathbb{E}$ the topology of convergence in λ-measure. The space $\mathbb{M}_\mathbb{E}$ plays the role of canonical space for the process X, on which the prediction process of X will be defined.

The prediction process is the process consisting of the conditional distributions of the future of X given the past at each time $t \in \mathbb{R}_+$. Our first task is then to define the past and the future precisely. The problem here is that the natural filtration $\sigma(w(s), s \le t)$ is not meaningful since the coordinate projection $w \mapsto w(t)$ is not $\mathcal{B}(\mathbb{M}_\mathbb{E})/\mathcal{E}$ measurable; the atoms of $\mathcal{B}(\mathbb{M}_\mathbb{E})$ are the equivalence classes. Instead, we define the *pseudo-path* filtration $\mathcal{F}'_t$ by

$$\mathcal{F}'_t = \sigma\left(\int_0^s f(w(u))\,du; s < t,\ f \in b\mathcal{E}\right),$$

and set $\mathcal{F}' \triangleq \mathcal{F}'_\infty = \bigvee_{t>0} \mathcal{F}'_t$. Proposition 2.10 shows that $\mathcal{F}' = \mathcal{B}(\mathbb{M}_\mathbb{E})$. Moreover, the shift operator θ_t on $\mathbb{M}_\mathbb{E}$ is defined by $\theta_t w(s) = w(t+s)$ for $s, t \in \mathbb{R}_+$ and is $\mathcal{F}'_{t+s}/\mathcal{F}'_s$ measurable. The past is now defined to be $\mathcal{F}'_t$ and the future is given by $\theta_t^{-1}\mathcal{F}'$, which is isomorphic to $\mathcal{F}'$.

For the state space of the prediction process, let $\Pi \triangleq \mathcal{P}(\mathbb{M}_\mathbb{E})$, the set of probability measures on $(\mathbb{M}_\mathbb{E}, \mathcal{F}')$ endowed with the topology of weak convergence. This topology is called the *prediction topology* in Knight (1992), with which Π becomes a metrizable Lusin space. We set $\mathcal{G} = \mathcal{B}(\Pi)$. A generic element of Π is usually denoted by z, and we sometimes write P^z for z; it is actually redundant but intuitively helpful.

We are now ready to state the precise definition of the prediction process. According to Corollary 2.5 of Knight (1992), the prediction process $Z^z = (Z^z_t)_{t\in\mathbb{R}_+}$ for $z \in \Pi$ is the process with values in $(\Pi, \mathcal{G})$ that is P^z-a.s. uniquely determined by the following two requirements:

(1) $Z^z_r(A) = P^z(\,\theta_r^{-1}A \mid \mathcal{F}'_{r+}\,)$, $r \in \mathbb{Q}_+$, $A \in \mathcal{F}'$;

(2) Z^z_t is càdlàg for the prediction topology on Π defined above.

Thus the prediction process is defined for the law $z \in \Pi$ induced by X rather than for the process X itself. A remarkable result of Knight (1975) states that for any $z \in \Pi$, Z^z is a homogeneous strong Markov process (see also Knight (1981, 1992)). We are mainly concerned with the conditional probability aspect of the prediction process, so we do not have occasion to use the Markov property of Z^z.

To express (1) in terms of processes and to get around the problem of nonmeasurability of the coordinate projections mentioned above, we define a generalized coordinate process $\widetilde{X}$ on $\mathbb{M}_{\mathbb{E}}$ by

$$\widetilde{X}_t(w) \triangleq G(\widetilde{w}, t),$$

where $\widetilde{w}$ is the equivalence class containing w and G is the function constructed in Lemma 2.12. From the lemma, it is clear that $\widetilde{X}_t(w)$ is a measurable process with values in $(\mathbb{E}, \mathcal{E})$, that is, the mapping $(t, w) \mapsto \widetilde{X}_t(w)$ is $\mathcal{F}' \otimes \mathcal{B}(\mathbb{R}_+)/\mathcal{E}$ measurable and has the law z on $\mathbb{M}_{\mathbb{E}}$. One can show further that

(i) $\widetilde{X}_s(w)$ is $\mathcal{B}[0,t) \otimes \mathcal{F}'_t$ measurable on $[0,t) \times \mathbb{M}_{\mathbb{E}}$ and $\mathcal{B}[0,t] \otimes \mathcal{F}'_{t+}$ measurable on $[0,t] \times \mathbb{M}_{\mathbb{E}}$ for each $t \in \mathbb{R}_+$.

(ii) For each $w \in \mathbb{M}_{\mathbb{E}}$, $\widetilde{X}_s(w) = w(s)$ for a.e s, and $\sigma(\widetilde{X}_s, s < t) = \mathcal{F}'_t$ up to P^z-null sets.

See Knight (1992), Theorem 1.2 for a proof. (1) may then be written

$$Z^z_r(A) = P^z(\, \widetilde{X}_{t+\bullet} \in A \mid \mathcal{F}'_{r+}\,).$$

There are some situations in which one wishes to have the prediction process defined on a given probability space. If $X = (X_t)_{t \geq 0}$ is a measurable process on a probability space $(\Omega, \mathcal{F}, P)$ with values in $(\mathbb{E}, \mathcal{E})$ and if it induces the law z on $\mathbb{M}_{\mathbb{E}}$, then the prediction process Z^X of the given process X on $(\Omega, \mathcal{F}, P)$ is defined by $Z^X_t(A, \omega) = Z^z_t(A, X_\bullet(\omega))$, $t \in \mathbb{R}_+$, $A \in \mathcal{F}'$, and for $\mathcal{F}^X_{t+}$-stopping times $T < \infty$, we have

$$P(\, X_{T+\bullet} \in A \mid \mathcal{F}^X_{T+}\,) = Z^X_T(A), \quad A \in \mathcal{F}',$$

where $\mathcal{F}^X_t = X^{-1}(\mathcal{F}'_t)$.

It is often useful to consider, together with Z^z, the processes $K^z = (K^z_t)_{t \in \mathbb{R}_+}$ defined by $K^z_t(f \circ \theta_t) = Z^z_t(f)$, for $f \in b\mathcal{F}'$. Hence K^z satisfies,

$$K^z_T(A) = P^z(\, A \mid \mathcal{F}'_{T+}\,) \quad \text{on } \{T < \infty\},$$

for any $(\mathcal{F}'_{t+})$-optional T and $A \in \mathcal{F}'$. In terms of process, it means $K^z_t(A) = P^z(\, \widetilde{X}_\bullet \in A \mid \mathcal{F}'_{t+}\,)$. Hence one may say that K^z is the "prediction" process of the whole path.

Since we would like to discuss weak convergence of processes, it is necessary to assume some topological properties on $\mathbb{E}$. And we have defined the prediction process in a topological setting, so that the topology for the state space of the prediction process can be introduced in a natural way. The prediction process could, however, be defined only in a purely measure-theoretic framework (see Knight (1992)).

4.2 Tightness

First we discuss tightness of conditional distributions. Let S be a metrizable Lusin space, $\mathcal{S} = \mathcal{B}(S)$, and $\mathcal{P}(S)$ the set of all probability measures on $(S, \mathcal{S})$. We shall later apply it to the case where $S = \mathbb{M}_{\mathbb{E}}$. Suppose that $\{P^\alpha\}_{\alpha\in\Lambda}$ is a subset of $\mathcal{P}(S)$ and $\mathcal{C}$ is a sub-σ-field of $\mathcal{S}$. Let K^α be a conditional distribution given $\mathcal{C}$, i.e., $K^\alpha(A, s) = P^\alpha(\,A \mid \mathcal{C}\,)(s)$. We look at K^α as a $\mathcal{P}(S)$-valued random variable on $(S, \mathcal{S})$.

4.1 Proposition *If $\{P^\alpha\}$ is tight, so is $\{K^\alpha\}$.*

Proof. Let $\epsilon > 0$ be given. For each $j \in \mathbb{N}$, find a compact $C_j \subset S$ such that $P^\alpha(C_j) \geq 1 - \frac{\epsilon}{j2^j}$ for all $\alpha \in \Lambda$. We have

$$\begin{aligned} 1 - \frac{\epsilon}{j2^j} &\leq P^\alpha(C_j) = E^\alpha[K^\alpha(C_j)] = \int_0^1 P^\alpha[K^\alpha(C_j) \geq x]\,dx \\ &= \int_{1-1/j}^1 P^\alpha[K^\alpha(C_j) \geq x]\,dx + \int_0^{1-1/j} P^\alpha[K^\alpha(C_j) \geq x]\,dx \\ &\leq \frac{1}{j} P^\alpha\left[K^\alpha(C_j) \geq 1 - \frac{1}{j}\right] + 1 - \frac{1}{j}. \end{aligned}$$

Hence $\frac{1}{j} - \frac{\epsilon}{j2^j} \leq \frac{1}{j} P^\alpha\left[K^\alpha(C_j) \geq 1 - \frac{1}{j}\right]$, and so $P^\alpha\left[K^\alpha(C_j) \geq 1 - \frac{1}{j}\right] \geq 1 - \frac{\epsilon}{2^j}$ for all α. It follows that

$$\begin{aligned} P^\alpha\left[K^\alpha(C_j) \geq 1 - \frac{1}{j} \text{ for all } j \in \mathbb{N}\right] &= P^\alpha\left[\bigcap_{j=1}^\infty \left\{K^\alpha(C_j) \geq 1 - \frac{1}{j}\right\}\right] \\ \geq 1 - \sum_{j=1}^\infty \left[1 - P^\alpha\left(K^\alpha(C_j) \geq 1 - \frac{1}{j}\right)\right] &\geq 1 - \sum_{j=1}^\infty \frac{\epsilon}{2^j} = 1 - \epsilon. \end{aligned}$$

Set $H = \{\mu \in \mathcal{P}(S) \colon \mu(C_j) \geq 1 - 1/j \text{ for all } j \in \mathbb{N}\}$. If (μ_n) is a sequence in H, then for any $\eta > 0$, there exists a j_0 with $1/j_0 < \eta$ and $\mu_n(C_{j_0}) \geq 1 - \eta$ for all n. Thus (μ_n) is tight and by Prohorov's theorem, it is relatively compact. Hence we can find a subsequence (μ_{n_k}) converging to some $\mu \in \mathcal{P}(S)$. But since C_j is closed, $\mu(C_j) \geq \limsup_k \mu_{n_k}(C_j) \geq 1 - 1/j$ for all j and so μ is in fact in H. This shows that H is sequentially compact and, being a metric space, it is compact. Noting that

$$P^\alpha(K^\alpha \in H) = P^\alpha\left[K^\alpha(C_j) \geq 1 - \frac{1}{j} \text{ for all } j \in \mathbb{N}\right] \geq 1 - \epsilon$$

for all $\alpha \in \Lambda$, it follows that $\{K^\alpha\}$ is tight. ■

NOTE. If S is Polish, then the above proposition is true with tightness replaced by relative compactness because $\mathcal{P}(S)$ is also Polish if S is.

Now suppose that $P^n \xrightarrow{w} P$ on S. Then (P^n) is tight by Theorem 8 in Appendix III of Billingsley (1968). Also, by the above proposition, (K^n) is tight and hence relatively compact by Prohorov's theorem. Pick a convergent subsequence $(K^{n'})$ so that $K^{n'}(f) \xrightarrow{\mathcal{L}} \widetilde{K}(f)$ for some kernel $\widetilde{K}$ on S for every $f \in C_b(S)$. We want to find a condition under which $\widetilde{K} \stackrel{\mathcal{L}}{=} K$, where $K(A, s) = P(\,A \mid \mathcal{C}\,)$, or equivalently, $\widetilde{K}(f) \stackrel{\mathcal{L}}{=} K(f)$, $f \in C_b(S)$. Note that $\widetilde{K}(f)$, $K(f)$, and $K^n(f)$ are bounded random variables. So if $E^n[g(K^n(f))] \to E^n[g(K(f))]$ for every g in a family of continuous bounded functions which separates measures on $\mathbb{R}$, or a bounded interval, then we have $E[g(\widetilde{K}(f))] = E[g(K(f))]$ for all $f \in C_b(S)$, which implies $\widetilde{K}(f) \stackrel{\mathcal{L}}{=} K(f)$. In summary, we obtained

4.2 Proposition *If $P^n \xrightarrow{w} P$ on S, then it is sufficient for $K^n \xrightarrow{\mathcal{L}} K$ that $E^n[g(K^n(f))] \to E^n[g(K(f))]$ for every $f \in C_b(S)$ and g in a family of continuous bounded functions which separates measures on bounded intervals.*

NOTE. If we do not assume $P^n \xrightarrow{w} P$, then g must be in a convergence determining class.

Using Proposition 4.1, it follows that if a sequence (z_n) in $\mathcal{P}(\mathbb{M}_\mathbb{E})$ is tight, then for a fixed $t \in \mathbb{R}_+$, $(K_t^{z_n})_{n \in \mathbb{N}}$ is tight in Π. Since $Z_t^{z_n}$ is the image measure of $K_t^{z_n}$ under θ_t which is continuous, one can easily see that for a fixed $t \in \mathbb{R}_+$, $(Z_t^{z_n})_{n \in \mathbb{N}}$ is also tight in this case. The consequence of Proposition 4.2 applied to the prediction process may not be very useful, but can be stated as follows: if $z_n \xrightarrow{w} z$ in $\mathbb{M}_\mathbb{E}$ and if

$$E^{z_n}[g(E^{z_n}(\,f(\widetilde{X}_{t+\bullet}) \mid \mathcal{F}'_{t+}\,))] \to E^{z_n}[g(E^{z_n}(\,f(\widetilde{X}_{t+\bullet}) \mid \mathcal{F}'_{t+}\,))]$$

for every $f \in C_b(\mathbb{M}_{\mathbb{E}})$ and g in a family of continuous bounded functions which separates measures on bounded intervals, then $Z_t^{z_n} \xrightarrow{\mathcal{L}} Z_t^z$ in $\mathbb{M}_\Pi$.

4.3 Finite-dimensional convergence

Next we shall find a condition for finite-dimensional convergence of the prediction process. Suppose that S and $\mathcal{P}(S)$ are as above, and for $f \in C_b(S)$, define $f^*: \mathcal{P}(S) \to \mathbb{R}$ by $f^*(\mu) = \mu(f) = \int f\, d\mu$. It is clear that the family $\{f^*: f \in C_b(S)\}$ generates the weak topology on $\mathcal{P}(S)$. Let (Z^n) be a sequence of $\mathcal{P}(S)$-valued random variables on some probability spaces. By Pollard's theorem, $Z^n \xrightarrow{w} Z$ is equivalent to

$$(Z^n(f_1), \ldots, Z^n(f_k)) \xrightarrow{\mathcal{L}} (Z(f_1), \ldots, Z(f_k)) \quad \text{in } \mathbb{R}^k$$

for any $f_1, \ldots, f_k \in C_b(S)$ and $k \in \mathbb{N}$. By the Cramer-Wold device (see Billingsley (1968), theorem 7.7), the above is in turn equivalent to

$$a_1 Z^n(f_1) + \cdots + a_k Z^n(f_k) \xrightarrow{\mathcal{L}} a_1 Z(f_1) + \cdots + a_k Z(f_k)$$

for any $a_1, \ldots, a_k \in \mathbb{R}$. But obviously $a_1 Z^n(f_1) + \cdots + a_k Z^n(f_k) = Z^n(a_1 f_1 + \cdots + a_k f_k)$ and $a_1 f_1 + \cdots + a_k f_k \in C_b(S)$. Thus $Z^n \xrightarrow{\mathcal{L}} Z$ if and only if $Z^n(f) \xrightarrow{\mathcal{L}} Z(f)$ for each $f \in C_b(S)$. Clearly we can replace $C_b(S)$ by any family (f_j) which is dense in $C(\widehat{S})$ where $\widehat{S}$ is a compact metric space containing S as a Borel subset.

If we have a sequence of vectors of $\mathcal{P}(S)$-valued random variables $(Z_1^n, \ldots, Z_k^n)$, the same reasoning shows that

$$(Z_1^n, \ldots, Z_k^n) \xrightarrow{\mathcal{L}} (Z_1, \ldots, Z_k)$$

if and only if

$$(Z_1^n(f_1), \ldots, Z_k^n(f_k)) \xrightarrow{\mathcal{L}} (Z_1(f_1), \ldots, Z_k(f_k))$$

for any $f_1, \ldots, f_k \in C_b(S)$. We can use this to obtain a condition for convergence of finite-dimensional distributions of the prediction process. Suppose that we have a sequence (z_n) in Π, and let $Z_t^{z_n}$ be their prediction processes. According to what we have proved above,

$$(Z_{t_1}^{z_n}, \ldots, Z_{t_k}^{z_n}) \xrightarrow{\mathcal{L}} (Z_{t_1}^{z}, \ldots, Z_{t_k}^{z})$$

if and only if

$$(Z^{z_n}_{t_1}(g_1),\dots,Z^{z_n}_{t_k}(g_k)) \xrightarrow{\mathcal{L}} (Z^{z}_{t_1}(g_1),\dots,Z^{z}_{t_k}(g_k))$$

for every $g_1,\dots,g_k \in \{g_j\}$, where $\{g_j\}$ is dense in $C(\widehat{\mathbb{M}}_{\mathbb{E}})$. We know from the results in Section 2 that the $g_j(w)$ of the form

$$\prod_{i=1}^{m}\int_0^{t_i} f_i(w(t))\,dt \qquad \text{or} \qquad \prod_{i=1}^{m}\int_0^{\infty} e^{-\lambda_i t} f_i(w(t))\,dt,$$

where $\{f_j\}$ is dense in $C(\widehat{\mathbb{E}})$, may be used. So for the latter choice of $\{g_j\}$, the condition becomes

$$\left(E^{z_n}\left[\prod_{i=1}^{m_j}\int_0^{\infty} e^{-\lambda_i^j s} f_i^j(w(s))\,ds\circ\theta_{t_j}\middle|\mathcal{F}'_{t_j+}\right]\right)_{j=1,\dots,k}$$
$$\to\left(E^{z}\left[\prod_{i=1}^{m_j}\int_0^{\infty} e^{-\lambda_i^j s} f_i^j(w(s))\,ds\circ\theta_{t_j}\middle|\mathcal{F}'_{t_j+}\right]\right)_{j=1,\dots,k}.$$

Note that, in view of Theorem 3.6, the above condition implies the convergence in law of (Z^{z_n}) to Z^z in $\mathbb{M}_\Pi$.

4.4 Convergence of given processes

We shall now go on to show that the convergence of the prediction processes implies that of the given processes. Let X^n, $n \in \mathbb{N}$ and X be measurable processes with values in $\mathbb{E}$, and let Z^n and Z be the prediction processes of X^n and X respectively. Since $\mathbb{M}_{\mathbb{E}}$ is a metrizable Lusin space, so is Π (Dellacherie and Meyer (1975), III.60). Then the path space $\mathbb{M}_\Pi = \mathbb{M}_\Pi(\mathbb{R}_+,\mathcal{B}_+,\lambda)$ of the prediction process is also a metrizable Lusin space by Proposition 2.10.

4.3 Theorem *If $(Z^n)_{n\in\mathbb{N}}$ converges in law to Z in $\mathbb{M}_\Pi$, then $(X^n)_{n\in\mathbb{N}}$ converges in law to X in $\mathbb{M}_{\mathbb{E}}$.*

PROOF. By Corollary 3.9, for any subsequence (n'), there is a further subsequence (n'') and an $S \subset \mathbb{R}_+$ with $\lambda(S^{\complement}) = 0$ such that $Z^{n''} \xrightarrow{\mathrm{fd}(S)} Z$. Also for any $\phi \in C_b(\mathbb{M}_{\mathbb{E}})$, the mapping $\mu \mapsto \mu(\phi) = \int \phi \, d\mu$ from Π into $\mathbb{R}$ is continuous. It thus follows that for any $t \in S$ and any $\phi \in C_b(\mathbb{M}_\mathbb{E})$,

$$E[\,\phi(X^{n''}_{t+\bullet}) \mid \mathcal{F}^{n''}_{t+}\,] \xrightarrow{\mathcal{L}} E[\,\phi(X_{t+\bullet}) \mid \mathcal{F}_{t+}\,] \quad \text{in } \mathbb{R},$$

where $\mathcal{F}^{n''}_{t+} = \mathcal{F}^{X^{n''}}_{t+}$ and $\mathcal{F}_{t+} = \mathcal{F}^{X}_{t+}$. The sequence $\{E[\,\phi(X^{n''}_{t+\bullet}) \mid \mathcal{F}^{n''}_{t+}\,]\}$ is bounded, so it is uniformly integrable. Hence we get $E[\phi(X^{n''}_{t+\bullet})] \to E[\phi(X_{t+\bullet})]$ for any $t \in S$ and any $\phi \in C_b(\mathbb{M}_\mathbb{E})$. In other words, $X^{n''}_{t+\bullet} \xrightarrow{\mathcal{L}} X_{t+\bullet}$ in $\mathbb{M}_\mathbb{E}$ for each $t \in S$. Using the translation operator θ_t, we may write it as $\theta_t X^{n''} \xrightarrow{\mathcal{L}} \theta_t X$ in $\mathbb{M}_\mathbb{E}$ for each $t \in S$.

It is clear that θ_t is continuous since if $w_n \to w$ in λ-measure, then $w_n(t+\bullet) \to w(t+\bullet)$ in λ-measure. Set $\mathcal{A} = \{\theta_t^{-1} G : G \text{ open in } \mathbb{M}_\mathbb{E},\ t \in S\}$.

(i) *$\mathcal{A}$ is a base for the topology for $\mathbb{M}_\mathbb{E}$.*

Let H be any open set in $\mathbb{M}_\mathbb{E}$ and $w_0 \in H$. Then there is a ball U centered at w_0 and contained in H. Write $U = \{w : \rho(w, w_0) < \epsilon\}$, where $\rho(w, w_0) = \int_0^\infty 1 \wedge d(w(s), w_0(s))\, \lambda(ds)$ and d is a metric compatible with the topology of $\mathbb{E}$. Let $V = \{w : \rho(w, \theta_t w_0) < \epsilon/2\}$ with $t < \epsilon/2$ and $t \in S$. The we have

$$\begin{aligned}
\theta_t^{-1} V &= \left\{ w : \rho(\theta_t w, \theta_t w_0) < \frac{\epsilon}{2} \right\} = \left\{ w : \int_0^\infty 1 \wedge d(w(t+s), w_0(t+s))\, \lambda(ds) < \frac{\epsilon}{2} \right\} \\
&= \left\{ w : \int_t^\infty 1 \wedge d(w(s), w_0(s))\, \lambda(ds) < e^{-t} \frac{\epsilon}{2} \right\} \\
&\subset \left\{ w : \int_t^\infty 1 \wedge d(w(s), w_0(s))\, \lambda(ds) < \frac{\epsilon}{2} \right\}.
\end{aligned}$$

Since $d(w(s), w_0(s)) \wedge 1 \le 1$, $t < \epsilon/2$ and $\lambda[0, t] < \epsilon/2$, we see that $\theta_t^{-1} V \subset U$.

(ii) *$\mathcal{A}$ is closed under finite unions.*

It is enough to consider unions of two sets in $\mathcal{A}$. Let G_1 and G_2 be open sets in $\mathbb{M}_\mathbb{E}$ and $t_1, t_2 \in S$ with $t_1 < t_2$. We have $\theta_{t_1}^{-1} G_1 \cup \theta_{t_2}^{-1} G_2 = \theta_{t_1}^{-1} G_1 \cup \theta_{t_1}^{-1} \theta_{t_2 - t_1}^{-1} G_2 = \theta_{t_1}^{-1}(G_1 \cup \theta_{t_2 - t_1}^{-1} G_2)$. $\theta_{t_2 - t_1}^{-1} G_2$ is open because θ_t is continuous for $t \in \mathbb{R}_+$.

It follows from (i) and (ii) that the family $\{\theta_t\}_{t \in S}$ satisfies the conditions of Pollard's theorem (see Corollary A2.3). Consequently we obtain $X^{n''} \xrightarrow{\mathcal{L}} X$ in $\mathbb{M}_\mathbb{E}$.

We have shown that for any subsequence (n'), there exists a further subsequence (n'') for which $X^{n''} \xrightarrow{\mathcal{L}} X$ in $\mathbb{M}_\mathbb{E}$, which obviously implies that the sequence (X^n) of the given processes converges in law to X in $\mathbb{M}_\mathbb{E}$. ■

REMARK. The assertion of Theorem 4.3 is in fact equivalent to that of Lemma 2.21 (1) of Knight (1992). Our proof here is different from his, and the point of our proof is that the result can be shown without using the Markov property of the prediction process; only the defining property of the prediction process is necessary.

4.5 Case of Markov process

In this subsection, we assume that $\mathbb{E}$ is Polish. Let $p(t, x, B)$, $t \in \mathbb{R}_+$, $x \in \mathbb{E}$, $B \in \mathcal{E}$ be a Markov transition function which satisfies

(4.1) $(t, x) \mapsto p(t, x, B)$ is $\mathcal{B}(0, \infty) \otimes \mathcal{E}$ measurable for each $B \in \mathcal{E}$;

(4.2) $\{x \mapsto p(t, x, B): t > 0,\ B \in \mathcal{E}\}$ separates points of $\mathbb{E}$.

From Lemma 2.8 in Knight (1992), for each $x \in \mathbb{E}$, there is a measurable Markov process $X = (X_t)_{t \in \mathbb{R}_+}$ with finite-dimensional distributions determined by $p(t, x, B)$. Namely, for $0 \le t_1 < \cdots < t_k$, we have

$$(4.3) \quad P^x(X_{t_1} \in B_1, \ldots, X_{t_k} \in B_k) \\ = \int_{B_{k-1}} \cdots \int_{B_1} p(t_1, x, dx_1) \cdots p(t_{k-1} - t_{k-2}, x_{k-2}, dx_{k-1}) p(t_k - t_{k-1}, x_{k-1}, B_k).$$

Note that we do not assume $p(0, x, \bullet) = \delta_x(\bullet)$. Thus the process X may not start at x under P^x. We call the above process X the measurable Markov process having the P^x-law with transition function $p(t, x, B)$. This process induces a law on $\mathbb{M}_{\mathbb{E}}$, which we denote by $\varphi(x)$. This is uniquely determined by $p(t, x, B)$ (see Knight (1992), p. 53). We look at φ as a mapping from $\mathbb{E}$ into Π. Lemma 2.9 of Knight (1992) shows that φ is $\mathcal{E}/\mathcal{B}(\Pi)$ measurable. Furthermore, Theorem 2.36 of Knight (1992) states that $\sigma(x \mapsto p(t, x, B): t > 0,\ B \in \mathcal{E}) = \sigma(x \mapsto R_\lambda f(x): \lambda > 0,\ f \in b\mathcal{E})$, where R_λ is the resolvent defined as

$$R_\lambda f(x) \triangleq \int_0^\infty e^{-\lambda t} T_t f(x)\, dt$$

and $T_t f(x) \triangleq \int p(t, x, dy) f(y)$ is the semigroup associated with $p(t, x, B)$. Thus (4.2) amounts to assuming that $\{x \mapsto R_\lambda f(x): \lambda > 0,\ f \in b\mathcal{E}\}$ separates points of $\mathbb{E}$. Noting that $R_\lambda f(x) = E^x\left[\int_0^\infty e^{\lambda t} f(X_t)\, dt\right]$ by Fubini's theorem, we see that φ is 1-1. For, if $x \neq y$, then clearly

$$E^x\left[\int_0^\infty e^{\lambda t} f(X_t)\, dt\right] \neq E^y\left[\int_0^\infty e^{\lambda t} f(X_t)\, dt\right]$$

for some $\lambda > 0$ because those functionals separates measures on $\mathbb{M}_{\mathbb{E}}$ (see Chapter 2). This simply means that $\varphi(x) \neq \varphi(y)$. The key result is Theorem 2.10 of Knight (1992), which says that for each $x \in \mathbb{E}$, we have

$$P^{\varphi(x)}\left[\varphi(\widetilde{X}_t) = Z_t^{\varphi(x)} \text{ for a.e. } t\right] = 1.$$

This tells us that the process $\varphi(X) = (\varphi(X_t))_{t\in\mathbb{R}_+}$ and $Z^{\varphi(x)}$ induce the same law on $\mathbb{M}_{\Pi}$.

Now consider a sequence of Markov transition functions $(p_n(t,x,B))_{n\in\mathbb{N}}$ and $(p(t,x,B)$ satisfying (4.1) and (4.2) above, and denote by (X^n) and X the measurable Markov processes having the P^x-law with $p_n(t,x,B)$. Our problem is to find under which conditions on $(p_n(t,x,B))$, $X^n \xrightarrow{\mathcal{L}} X$ in $\mathbb{M}_{\mathbb{E}}$ implies $Z^{X^n} \xrightarrow{\mathcal{L}} Z^X$ in $\mathbb{M}_{\Pi}$ (note that the dependence on x is suppressed here). More precisely, if $X^n \xrightarrow{\mathcal{L}} X$ in $\mathbb{M}_{\mathbb{E}}$ for each $x \in \mathbb{E}$, then what additional conditions are necessary for Z^{X^n} to converge in law to Z^X in $\mathbb{M}_{\Pi}$ for each x ? The assumption amounts to $\varphi^n(x) \to \varphi(x)$ for each $x \in \mathbb{E}$, where $\varphi^n(x)$ and $\varphi(x)$ are the laws on $\mathbb{M}_{\mathbb{E}}$ induced by X^n and X with (4.3), as defined above. From the above observation, we know that $\varphi^n(X^n)$ and $Z^{X^n} = Z^{\varphi(x)}$ induce the same distribution on $\mathbb{M}_{\Pi}$, so what we need is $\varphi^n(X^n) \xrightarrow{\mathcal{L}} \varphi(X)$ in $\mathbb{M}_{\Pi}$ for each $x \in \mathbb{E}$. Thus the problem is reduced to a familiar one of preservation of convergence in law under mappings. This is discussed in Section 5 of Billingsley (1968) and a necessary and sufficient condition was obtained in Topsøe (1967). Here we use a simple condition, which is an easy consequence of H. Rubin's theorem (Billingsley (1968), Theorem 5.5).

4.4 Proposition *Let X^n and X be S-valued random variables with S separable metric, and let h_n and h be measurable mappings from S into a metric space S'. If h is continuous and if h_n converges to h uniformly on compact sets, then $X^n \xrightarrow{\mathcal{L}} X$ implies $h_n(X^n) \xrightarrow{\mathcal{L}} h(X)$.*

To apply this proposition to our problem, let us define $\Phi\colon \mathbb{M}_{\mathbb{E}} \to \mathbb{M}_{\Pi}$ by $\Phi(w)(t) = \varphi(w(t))$, $w \in \mathbb{M}_{\mathbb{E}}$, and similarly define Φ^n. The processes $(\varphi^n(X_t^n))_{t\in\mathbb{R}_+}$, $n \in \mathbb{N}$ are then written as $\Phi^n(X^n)$, $n \in \mathbb{N}$. It is clear that if φ is continuous, so is Φ. Denote the metrics on Π and $\mathbb{M}_{\Pi}$ by d' and ρ' so that we have

$$\rho'\left(\Phi^n(x), \Phi(x)\right) = \int_0^\infty 1 \wedge d'\left(\varphi^n(w(t)), \varphi(w(t))\right) \lambda(dt).$$

Let Γ be a compact subset on $\mathbb{M}_{\mathbb{E}}$ and $\epsilon > 0$ be given. Choose $T > 0$ satisfying $\lambda(T,\infty) < \epsilon$. It follows from Proposition 2.12 (this is the only place where the Polish assumption is used) that we can find a compact $K \subset \mathbb{E}$ such that $\sup_{w \in \Gamma} \lambda(t \leq T : w(t) \notin K) \leq \epsilon$. Now assume for the moment that $\varphi^n \to \varphi$ uniformly on compact sets. Then

$$\begin{aligned}\rho'(\Phi^n(w), \Phi(w)) &\leq \int_{[0,T]\cap\{w(t)\in K\}} 1 \wedge d'(\varphi^n(w(t)), \varphi(w(t)))\, \lambda(dt) \\ &\quad + \int_{[0,T]\cap\{w(t)\notin K\}} 1 \wedge d'(\varphi^n(w(t)), \varphi(w(t)))\, \lambda(dt) + \epsilon \\ &\leq \int_{[0,T]\cap\{w(t)\in K\}} 1 \wedge d'(\varphi^n(w(t)), \varphi(w(t)))\, \lambda(dt) + 2\epsilon.\end{aligned}$$

The integral converges to 0 uniformly in $w \in \Gamma$ by the bounded convergence theorem. Hence, as $n \to \infty$,

$$\sup_{w \in \Gamma} \rho'(\Phi^n(w), \Phi(w)) \to 0.$$

In view of Proposition 4.4, $Z^{X^n} \xrightarrow{\mathcal{L}} Z^X$ will follow.

We would like to express the assumed compact convergence of φ^n to φ in terms of resolvents R^n_λ and R_λ of X^n and X respectively. First, by Lemma 2.15 of Knight (1992), φ is continuous if and only if $R_\lambda f$ is continuous on $\mathbb{E}$ for $f \in C_b(\mathbb{E})$. Compact convergence of φ^n to φ means

$$\sup_{x \in K} \left| E^{\varphi^n(x)}(g) - E^{\varphi(x)}(g) \right| \to 0$$

for each compact $K \subset \mathbb{E}$ and each $g \in C_b(\mathbb{M}_\mathbb{E})$. We may replace g by a member of a convergence determining class. We use the following class obtained in Section 2:

$$\left\{ \prod_{k=1}^{m} \int_0^\infty e^{-\lambda_k s} f_{j_k}(w(s))\, ds : m \in \mathbb{N},\ \lambda_k \in \mathbb{Q}_+,\ f_{j_k} \in \{f_j\},\ 1 \leq k \leq m \right\},$$

where (f_j) is dense in $C(\widehat{\mathbb{E}}) \cap \{f : \widehat{\mathbb{E}} \to [0,1]\}$. So we need to find a condition for

$$E^{\varphi^n(x)}\left[\prod_{k=1}^{m} \int_0^\infty e^{-\lambda_k t} f_{j_k}(\widetilde{X}_t)\, dt\right] \to E^{\varphi(x)}\left[\prod_{k=1}^{m} \int_0^\infty e^{-\lambda_k t} f_{j_k}(\widetilde{X}_t)\, dt\right] \tag{4.4}$$

uniformly in $x \in K$ for a compact K and $m \in \mathbb{N}$, $\lambda_k \in \mathbb{Q}_+$ and $f_{j_k} \in (f_j)$. Let us look at the case $m = 1$. The left-hand side is

$$E^{\varphi^n(x)}\left[\int_0^\infty e^{-\lambda t} f_{j_k}(\widetilde{X}_t)\, dt\right] = R^n_\lambda f_j(x).$$

Hence we need the uniform convergence on compact sets of the resolvents, that is,

$$R^n_\lambda f(x) \to R_\lambda f(x) \quad \text{uniformly in } x \in K \tag{4.5}$$

for each λ and $f \in C(\widehat{\mathbb{E}})$. For a general m, we use the argument given in the proof of Lemma 2.15 of Knight (1992). Write the left-hand side of (4.4) as

$$E^{\varphi^n(x)}\left[\int_0^\infty \cdots \int_0^\infty e^{-\sum_{k=1}^m \lambda_k s_k} f_{j_1}(\widetilde{X}_{s_1}) \cdots f_{j_m}(\widetilde{X}_{s_m})\, ds_1 \cdots ds_m\right].$$

Express this multiple integral as a sum of $m!$ integrals according to $m!$ possible orderings of $s_1, \ldots, s_m$. Then it is enough to look at, for instance, the case $s_1 < \cdots < s_m$:

$$E^{\varphi^n(x)}\left[\int_0^\infty \int_{s_1}^\infty \cdots \int_{s_{m-1}}^\infty e^{-\sum_{k=1}^m \lambda_k s_k} f_{j_1}(\widetilde{X}_{s_1}) \cdots f_{j_m}(\widetilde{X}_{s_m})\, ds_1 \cdots ds_m\right]. \tag{4.6}$$

Using the Markov property, this is equal to

$$\begin{aligned}
&\int_0^\infty e^{-\lambda_1 s_1} E^{\varphi^n(x)}\Bigg[f_{j_1}(\widetilde{X}_{s_1}) E^{\varphi^n(x)}\Bigg(\int_{s_1}^\infty e^{-\lambda_2 s_2} f_{j_2}(\widetilde{X}_{s_2}) \\
&\qquad \int_{s_2}^\infty \cdots \int_{s_{m-1}}^\infty e^{-\lambda_m s_m} f_{j_m}(\widetilde{X}_{s_m})\, ds_m \cdots ds_2 \Bigg| \mathcal{F}'_{s_1+}\Bigg)\Bigg] ds_1 \\
=&\int_0^\infty e^{-\lambda_1 s_1} E^{\varphi^n(x)}\Bigg[f_{j_1}(\widetilde{X}_{s_1}) E^{\varphi^n(x)}\Bigg(\int_0^\infty e^{-\lambda_2 (t_2+s_1)} f_{j_2}(\widetilde{X}_{t_2+s_1}) \\
&\qquad \int_{t_2}^\infty \cdots \int_{t_{m-1}}^\infty e^{-\lambda_m (t_m+s_1)} f_{j_m}(\widetilde{X}_{t_m+s_1})\, dt_m \cdots dt_2 \Bigg| \mathcal{F}'_{s_1+}\Bigg)\Bigg] ds_1 \\
=&\int_0^\infty e^{-\lambda_1 s_1} E^{\varphi^n(x)}\Bigg[f_{j_1}(\widetilde{X}_{s_1}) e^{-(\lambda_2+\cdots+\lambda_m)s_1} \\
&\qquad E^{\varphi^n(\widetilde{X}_{s_1})}\Bigg(\int_0^\infty e^{-\lambda_2 t_2} f_{j_2}(\widetilde{X}_{t_2}) \int_{t_2}^\infty \cdots \int_{t_{m-1}}^\infty e^{-\lambda_m t_m} f_{j_m}(\widetilde{X}_{t_m})\, dt_m \cdots dt_2 \Bigg)\Bigg] ds_1 \\
=&E^{\varphi^n(x)}\left[\int_0^\infty e^{-\overline{\lambda} s_1} f_{j_1}(\widetilde{X}_{s_1}) g^n(\widetilde{X}_{s_1})\, ds_1\right],
\end{aligned}$$

where $\overline{\lambda} \triangleq \lambda_1 + \cdots + \lambda_m$ and

$$g^n(x) \triangleq E^{\varphi^n(x)}\left[\int_0^\infty e^{-\lambda_2 t_2} f_{j_2}(\widetilde{X}_{t_2}) \int_{t_2}^\infty \cdots \int_{t_{m-1}}^\infty e^{-\lambda_m t_m} f_{j_m}(\widetilde{X}_{t_m})\, dt_m \cdots dt_2\right].$$

Note that g^n is of the form (4.6) with $m-1$ in place of m. Thus if we assume that (4.4) holds for $m-1$ as the induction hypothesis, $g^n(s)$ will converge to $g(x)$, defined similarly, uniformly in $x \in K$. Writing $h^n(x) = f_{j_1}(x)g^n(x)$, the above expectation is equal to $R^n_\lambda h^n(x)$. Assuming the induction hypothesis, $h^n(x)$ converges to $h(x) = f_{j_1}(x)g(x)$ uniformly in $x \in K$. So the condition we need is the following:

$R^n_\lambda h^n(x) \to R_\lambda h(x)$ uniformly in $x \in K$ for each $\lambda > 0$, whenever $h^n \to h$ uniformly on compact sets.

As is seen by the above argument, the sequence (h^n) may be restricted to be uniformly bounded and we may assume that h is continuous and bounded.

We have therefore obtained the following theorem.

4.5 Theorem *Let $(p_n(t,x,B))_{n\in\mathbb{N}}$ and $p(t,x,B)$ be Markov transition functions satisfying (4.1) and (4.2) above, and let (X^n) and X be the measurable Markov processes having the P^x-laws with $p_n(t,x,B)$ and $p(t,x,B)$ respectively. Suppose that $R^n_\lambda h^n$ converges to $R_\lambda h$ uniformly on compact sets for each $\lambda > 0$ whenever a uniformly bounded sequence (h^n) converges to a continuous bounded h uniformly on compact sets, and that $R_\lambda f$ is continuous for $f \in C_b(\mathbb{E})$. Then Z^{X^n} converges in law to Z^X in $\mathbb{M}_\Pi$ for each $x \in \mathbb{E}$.*

Acknowledgements

The author would like to thank Ditlev Monrad, Frank Knight and Catherine Doléans-Dade for their helpful comments.

References

Aldous, D.J. (1981). *Weak Convergence and the General Theory of Processes*, Incomplete draft of monograph.

Bauer, H. (1981). *Probability Theory and Elements of Measure Theory*, 2nd English ed., Academic Press, London.

Billingsley, P. (1968). *Convergence of Probability Measures*, John Wiley and Sons, New York.

Billingsley, P. (1986). *Probability and Measure*, 2nd ed., John Wiley and Sons, New York.

Bourbaki, N. (1974). *Eléments de Mathématique, Topologie Générale*, Chapitres 5 à 10, Diffusion C.C.L.S., Paris.

Bourbaki, N. (1981). *Eléments de Mathématique, Espaces Vectoriels Topologiques*, Chapitres 1 à 5, Masson, Paris.

Chung, K.L. and Doob, J.L. (1965). Fields, optionality and measurability, *Amer. J. Math.* **87**, 397-424.

Cohn, D.L. (1972). Measurable choice of limit points and the existence of separable and measurable processes, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **22**, 161-165.

Cremers, H. and Kadelka, D. (1986). On weak convergence of integral functionals of stochastic processes with applications to processes taking paths in L_p^E, *Stochastic Processes Appl.* **21**, 305-317.

Dellacherie, C and Meyer, P.-A. (1975). *Probabilités et Potentiel*, Chapitres I à IV, Hermann, Paris.

Dunford, N. and Schwartz, J.T. (1958). *Linear Operators, Part I: General Theory*, John Wiley and Sons, New York.

Engelbert, H.J. (1984). On modifications of stochastic processes, *Math. Nachr.* **116**, 251-267.

Ethier, S.N. and Kurtz, T.G. (1986). *Markov Processes: Characterization and Convergence*, John Wiley and Sons, New York.

Fremlin, D.H., Garling, D.J.H. and Haydon, R.G. (1972). Bounded measures on topological spaces, *Proc. London Math. Soc.* **25**, 115-136.

Hewitt, E. and Ross, K.A. (1979). *Abstract Harmonic Analysis I*, 2nd ed, Springer-Verlag, Berlin Heidelberg New York.

Hoffmann-Jørgensen, J. (1973). Existence of measurable modifications of stochastic processes, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **25**, 205-207.

Jacod, J. and Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes*, Springer-Verlag, Berlin Heidelberg New York.

Knight, F.B. (1975). A predictive view of continuous time processes, *Ann. Probab.* **3**, 573-596.

Knight, F.B. (1981). *Essays on the Prediction Process*, Lect. Notes Series, 1, Institute of Mathematical Statistics, Hayward, CA.

Knight, F.B. (1992). *Foundations of the Prediction Process*, Oxford University Press, Oxford.

Kurtz, T.G. (1991). Random time changes and convergence in distribution under the Meyer-Zheng conditions, *Ann. Probab.* **19**, 1010-1034.

Meyer, P.-A. and Zheng, W.A. (1984). Tightness criteria for laws of semimartingales, *Ann. Inst. Henri Poincaré Probab. Statist.* **20**, 353-372.

Pollard, D. (1977). Induced weak convergence and random measures, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **37**, 321-328.

Rabolledo, R. (1987). Topologie faible et meta-stabilité, *Séminaire de Probabilités XXI, Lect. Notes in Math.* **1247**, Springer-Verlag, Berlin Heidelberg New York, 544-562.

Sadi, H. (1988). Une condition nécessaire et suffisante pour la convergence en pseudo-loi des processus, *Séminaire de Probabilités XXII, Lect. Notes in Math.* **1321**, Springer-Verlag, Berlin Heidelberg New York, 434-437.

Schwartz, L. (1973). *Radon Measures on Arbitrary Topological Spaces and Cylindrical Measures*, Oxford University Press, London.

Skorokhod, A.V. (1980). On measurability of stochastic processes, *Theory Probab. Appl.* **25**, 139-141.

Topsøe, F. (1967). Preservation of weak convergence under mappings, *Ann. Math. Statist.* **38**, 1661-1665.

Weil, A. (1951). *L'intégration dans les groupes topologiques et ses applications*, 2me édition, Hermann, Paris.