ON STRONG FORMS OF WEAK CONVERGENCE

S D JACKA*, *University of Warwick*

G O ROBERTS**, *University of Cambridge*

Abstract

We discuss three forms of convergence in distribution which are stronger than the normal weak convergence. They have the advantage that they are non-topological in nature and are inherited by discontinuous functions of the original random variables—clearly an improvement on 'normal' weak convergence. We give necessary and sufficient conditions for the three types of convergence and go on to give some applications which are very hard to prove in a more restricted setting.

CONVERGENCE IN DISTRIBUTION; SKOROKHOD REPRESENTATION; TOTAL VARIATION METRIC.

AMS 1991 SUBJECT CLASSIFICATION: PRIMARY 60B10
SECONDARY 60J25, 60J27
60H10, 62B05

§1. Introduction

Any introduction to weak convergence includes several examples of probability measures which 'ought' to converge but don't unless you adopt the weak formulation of the Portmanteau Theorem (see Billingsley (1968)) —for example the probability measures on $\mathbb{R}$: $\mathbb{P}_n(A) = 1_{({}^1\!/_n \in A)}$; $\mathbb{P}(A) = 1_{(0\in A)}$.

Concentration on this form of convergence of (probability) measures has, we believe, led to a failure to notice when stronger forms of convergence of measures may hold (with the exception of convergence with respect to the total variation metric). In this article we discuss three 'strong' forms of convergence of measure which are well-known in the function-analytic literature (see, for example, Dunford and Schwartz (1958)) but whose probabilistic scope is not widely appreciated. These forms of convergence are not topological in nature and therefore have significant advantages over the usual weak convergence. Firstly we demonstrate that two of these forms of convergence allow types of Skorokhod representation with respect to the discrete metric. Secondly we prove a result relating these types of convergence to convergence of sufficient statistics — we have found extensive applications of these results in problems concerning convergence

* Postal Address: Department of Statistics, University of Warwick, Coventry CV4 7AL, UK.

** Postal Address: Statistical Laboratory, 16 Mill Lane, Cambridge CB2 1SB, UK

of conditioned Markov processes. We go on to point out situations in which these forms of convergence apply and note some deductions which are very hard to prove in a more restricted setting.

The article is organised as follows—in chapter 2 we introduce the definitions. Chapter 3 offers alternative characterisations for each type of convergence, including the Skorokhod representation results. Chapter 4 gives applications to sufficient statistics and conditioned Markov processes. Chapter 5 gives applications to the convergence of more general time-inhomogeneous Markov processes.

§2. Notation, preliminaries and definitions

We shall work throughout with a fixed measurable space $(\Omega, \mathcal{F})$. All measures (unless otherwise stated) will live on $(\Omega, \mathcal{F})$—we make no topological assumptions about Ω (or $\mathcal{F}$). For ease of notation we shall index any sequence of probability measures on $(\Omega, \mathcal{F})$ by $\bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$. $\mathbb{P}_\infty$ will be our candidate limit law and any statement about the sequence $\mathbb{P}_n$ will be meant to include $\mathbb{P}_\infty$; i.e. $\{\mathbb{P}_n\}$ is shorthand for $\{\mathbb{P}_n; n \in \bar{\mathbb{N}}\}$ and $(\mathbb{P}_n)$ is shorthand for $(\mathbb{P}_n)_{n\in\bar{\mathbb{N}}}$. As is customary, we shall denote the distribution of a random object X under a probability measure $\mathbb{P}$ by $\mathbb{P}_X \equiv \mathbb{P} \circ X^{-1}$.

The first of our three forms of convergence is the standard undergraduate guess at a definition of convergence of measures:

Definition 2.1 Given a sequence $(\mathbb{P}_n)$ of probability measures on $(\Omega, \mathcal{F})$ we say the $\mathbb{P}_n$ converge strongly to $\mathbb{P}_\infty$, written

$$\mathbb{P}_n \overset{s}{\Rightarrow} \mathbb{P}_\infty \,,$$

if *for all* $A \in \mathcal{F}$

$$\mathbb{P}_n(A) \to \mathbb{P}_\infty(A) \text{ as } n \to \infty \,.$$

Before we introduce the other two types of convergence a reminder is in order.

Definition 2.2 A family of measures $\{\mu^\theta; \theta \in \Theta\}$ is said to be dominated by a measure μ if

$$\mu^\theta \ll \mu \text{ for all } \theta \in \Theta,$$

and such a μ is said to be a dominating measure for the family.

Note that for any countable collection $\{\mathbb{P}_n\}$ of probability measures on $(\Omega, \mathcal{F})$ there is a dominating measure (call it $\mathcal{R}$) such that each $\mathbb{P}_n$ is absolutely continuous with respect to $\mathcal{R}$ (and thus has a density by the Radon-Nikodym theorem). To see this simply set

$$\mathcal{R} = \tfrac{1}{2}(\mathbb{P}_\infty + \sum_{n=1}^{\infty} 2^{-n}\mathbb{P}_n) \,.$$

Note that, in fact, $\mathcal{R}$ is a probability measure

Now for the two other types of convergence:

Definition 2.3 Given a sequence of probability measures $(\mathbb{P}_n)$ we say the $\mathbb{P}_n$ converge *Skorokhod weakly* to $\mathbb{P}_\infty$, written

$$\mathbb{P}_n \overset{SW}{\Rightarrow} \mathbb{P}_\infty\,,$$

if there is a dominating (probability) measure $\mathbb{Q}$ such that:

$$f_n^{\mathbb{Q}} \overset{prob(\mathbb{Q})}{\longrightarrow} f_\infty^{\mathbb{Q}} \text{ as } n \to \infty,$$

where $f_n^{\mathbb{Q}}$ is a version of $\frac{d\mathbb{P}_n}{d\mathbb{Q}}$.

Definition 2.4 Given the $(\mathbb{P}_n)$, we say the $\mathbb{P}_n$ converge *Skorokhod strongly* to $\mathbb{P}_\infty$, written

$$\mathbb{P}_n \overset{SS}{\Rightarrow} \mathbb{P}_\infty\,,$$

if there exists a dominating probability measure $\mathbb{Q}$ such that:

$$f_\infty^{\mathbb{Q}} \wedge f_n^{\mathbb{Q}} \overset{\mathbb{Q}a.s.}{\longrightarrow} f_\infty^{\mathbb{Q}} \text{ as } n \to \infty\,.$$

Remark 2.5 The reason for the nomenclature will become apparent in the next chapter.

Remark 2.6 There is no need to restrict the choice of $\mathbb{Q}$ to probability measures— any σ-finite measure will do (with convergence in probability replaced by convergence in measure and convergence a.s. replaced by convergence a.e.). Conversely, we gain nothing by allowing more general σ-finite measures, since if $\mathcal{R}$ is a σ-finite dominating measure with $T_n \nearrow \Omega$ and $\mathcal{R}(T_n) < \infty$ for each n, then there exists a sequence $(a_n) \subseteq (0,\infty)$ such that $\sum_n a_n R(T_n \backslash T_{n-1}) = 1$ and, defining $\mathbb{Q}$ by

$$\frac{d\mathbb{Q}}{dR} = \sum_n a_n 1_{(T_n \backslash T_{n-1})},$$

we see that $\mathbb{Q}$ is a probability measure on $(\Omega, \mathcal{F})$ and $\mathbb{Q} \sim R$, so we may substitute $\mathbb{Q}$ for $\mathcal{R}$ and $f_n^{\mathbb{Q}}$ $(\equiv f_n^R \frac{dR}{d\mathbb{Q}})$ for f_n^R.

Remark 2.7 The use of weak in the phrase 'Skorokhod weak convergence' is in the same style as in 'the weak law of large numbers', denoting convergence in probability.

Before we continue with the body of the paper we recall the definition of the total variation metric and the fundamental inequality of coupling: given two probability measures $\mathbb{P}$ and $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ we define the total variation metric d by

$$d(\mathbb{P}, \mathbb{Q}) = \sup_{A \in \mathcal{F}} |\mathbb{P}(A) - \mathbb{Q}(A)|\,.$$

Note that d is a metric on the space $\mathcal{P}$ of probability measures on $(\Omega, \mathcal{F})$ and $(\mathcal{P}, d)$ is a complete metric space. The fundamental inequality of coupling states that if $\mathbb{P}$ and $\mathbb{Q}$ are two probability measures on $(\Omega, \mathcal{F})$ then

(a) if X and Y are random objects: $X, Y : (\Omega', \mathcal{F}', \mathbb{P}') \to (\Omega, \mathcal{F})$ with distributions $\mathbb{P}'_X = \mathbb{P}$ and $\mathbb{P}'_Y = \mathbb{Q}$ then

$$\mathbb{P}'(X \neq Y) \geq d(\mathbb{P}, \mathbb{Q})\,;$$

(b) there exists a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ and random objects $X, Y : (\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}) \to (\Omega, \mathcal{F})$ such that

(i) $\tilde{\mathbb{P}}_X = \mathbb{P}$ and $\tilde{\mathbb{P}}_Y = \mathbb{Q}$

and

(ii) $\tilde{\mathbb{P}}(X \neq Y) = d(\mathbb{P}, \mathbb{Q})$.

The proof of this result relies on the following observation: suppose $\mathbb{P}$ and $\mathbb{Q}$ are as above then, taking any dominating measure $\mathcal{R}$ (i.e. an $\mathcal{R}$ s.t. $\mathbb{P}, \mathbb{Q} \ll R$), and, defining $f_{\mathbb{P}} = \frac{d\mathbb{P}}{dR}, f_{\mathbb{Q}} = \frac{d\mathbb{Q}}{dR}$,

$$\begin{aligned} d(\mathbb{P}, \mathbb{Q}) &= \int (f_{\mathbb{P}} - f_{\mathbb{Q}})^+ dR = \int (f_{\mathbb{P}} - f_{\mathbb{Q}})^- dR \\ &= \int_K (f_{\mathbb{P}} - f_{\mathbb{Q}}) dR = \int_{K^c} (f_{\mathbb{Q}} - f_{\mathbb{P}}) dR = \mathbb{P}(K) - \mathbb{Q}(K) = \mathbb{Q}(K^c) - \mathbb{P}(K^c) \end{aligned}$$

where $K = \{\omega : f_{\mathbb{P}}(\omega) > f_{\mathbb{Q}}(\omega)\}$ †

The proof of (a) then follows by defining the measure S by $S(A) = \mathbb{P}'(X = Y \in A)$ and observing that $S \ll R$ with density dominated by $f_{\mathbb{P}} \wedge f_{\mathbb{Q}}$ so that

$$S(\Omega) = \mathbb{P}'(X = Y) \leq \int_\Omega f_{\mathbb{P}} \wedge f_{\mathbb{Q}} dR.$$

To prove (b) we construct the measurable space $(\tilde{\Omega}, \tilde{\mathcal{F}}) \stackrel{def}{=} (\Omega \times \Omega \times [0,1], \mathcal{F} \otimes \mathcal{F} \otimes \mathcal{B}[0,1])$, and the probability measure T on $(\Omega, \mathcal{F})$ by setting $\frac{dT}{dR} = \frac{(f_{\mathbb{Q}} - f_{\mathbb{P}}) 1_{K^c}}{d(\mathbb{P}, \mathbb{Q})}$, then define (for each $\tilde{\omega} = (\omega_1, \omega_2, t) \in \Omega \times \Omega \times [0,1]$)

$$X(\tilde{\omega}) = \omega_1, Y(\tilde{\omega}) = \omega_1 1_{\left(t \leq \frac{f_{\mathbb{P}} \wedge f_{\mathbb{Q}}(\omega_1)}{f_{\mathbb{P}}(\omega_1)}\right)} + \omega_2 1_{\left(t > \frac{f_{\mathbb{P}} \wedge f_{\mathbb{Q}}(\omega_1)}{f_{\mathbb{P}}(\omega_1)}\right)}$$

and then set $\tilde{\mathbb{P}} = \mathbb{P} \otimes T \otimes \Lambda$, where Λ is Lebesgue measure (on $\mathcal{B}[0,1]$).

---

† We need to make a technical assumption here—that $(X \neq Y)$ is measurable: if we don't assume this then the result remains valid if we reinterpret $\mathbb{P}'(X \neq Y)$ as $\mathbb{P}'_-(X \neq Y)$, where $\mathbb{P}'_-$ is the inner measure generated by $\mathbb{P}'$ (= 1− outer measure).

§3. Equivalent formulations and counter examples

3.1. Equivalent formulations

Recall that for an arbitrary space $\mathbb{S}$, the discrete metric ρ_{dis} is defined by

$$\rho_{dis}(x,y) = I_{(x \neq y)}.$$

Hidden just underneath the surface (despite our protestations about the non-topological nature of our definitions) is a metric—the discrete metric. To see this with Definition 2.1 is the work of a moment—if you accept that 'measurable' is the right substitute for 'continuous' when working with the discrete metric. Recall part of the Portmanteau Theorem:

$$\mathbb{P}_n \overset{w}{\Rightarrow} \mathbb{P}_\infty$$

if and only if

$$\int g\, d\mathbb{P}_n \to \int g\, d\mathbb{P}_\infty \text{ for all bounded continuous } g : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}).$$

Theorem 3.1.1 Given $(\mathbb{P}_n)$ on $(\Omega, \mathcal{F})$ the following are equivalent:

(i) $\mathbb{P}_n \overset{s}{\Rightarrow} \mathbb{P}_\infty$:

(ii) $\int g\, d\mathbb{P}_n \to \int g\, d\mathbb{P}_\infty$ $\forall$ bounded measurable $g : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$:

(iii) $\exists$ a probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ s.t.

$$\int_A f_n^{\mathbb{Q}} d\mathbb{Q} \to \int_A f_\infty^{\mathbb{Q}} d\mathbb{Q} \text{ for all } A \in \mathcal{F}.$$

Remark 3.1.2 To put Definition 2.1 in line with the other two we might have adopted (iii) above as the definition of strong convergence.

Proof of Theorem 3.1.1 The equivalence of (i) and (iii) is immediate (using the comment after Definition 2.2).

Moreover setting $g = 1_A$ shows that (ii) $\Rightarrow$ (i).

To prove (i) $\Rightarrow$ (ii) just mimic the proof of the equivalent implication in the Portmanteau Theorem—approximating g by simple functions □

To see the underlying importance of the discrete metric for Skorokhod weak and strong convergence first recall Skorokhod's representation theorem: one version of it states that if $(\Omega, \mathcal{F})$ is a separable metric measurable space then $\mathbb{P}_n \overset{w}{\Rightarrow} \mathbb{P}_\infty$ if and only if there is a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ and random objects $(X_n) : (\Omega', \mathcal{F}', \mathbb{P}') \to (\Omega, \mathcal{F})$ such that

(i) $\mathbb{P}'_{X_n} = \mathbb{P}_n$

and

(ii) $X_n \stackrel{\mathbb{P}' a.s.}{\longrightarrow} X_\infty$

Now we can ask ' what about Skorokhod representation for the discrete metric?'

Theorem 3.1.3 If $(\mathbb{P}_n)$ are probability measures on $(\Omega, \mathcal{F})$ then the following are equivalent

(i) $\mathbb{P}_n \stackrel{SW}{\Rightarrow} \mathbb{P}_\infty$

(ii) $\exists$ a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ and random objects $(X_n) : (\Omega', \mathcal{F}', \mathbb{P}') \to (\Omega, \mathcal{F})$ such that

(a) $\mathbb{P}'_{X_n} = \mathbb{P}_n$

and

(b) $X_n \stackrel{prob(\mathbb{P}')}{\longrightarrow} X_\infty$ with respect to the discrete metric

i.e.

$$\mathbb{P}'(X_n \neq X_\infty) \to 0 \text{ as } n \to \infty$$

(iii) $\mathbb{P}_n \to \mathbb{P}_\infty$ with respect to the total variation metric

i.e.

$$d(\mathbb{P}_n, \mathbb{P}_\infty) \to 0 \text{ as } n \to \infty$$

(iv) $\exists$ a dominating probability measure $\mathbb{Q}$ s.t.

$$f_n^{\mathbb{Q}} \stackrel{L^1(\mathbb{Q})}{\longrightarrow} f_\infty^{\mathbb{Q}}$$

(v) $\mathbb{P}_n(A) \to \mathbb{P}_\infty(A)$ uniformly in $A \in \mathcal{F}$.

Remark 3.1.4 We've given you more than we promised here but trust the relative simplicity of the proofs will justify a (relatively) long list of equivalent conditions.

Remark 3.1.5 The equivalence of (i) and (iii) is Scheffé's lemma (see Billingsley (1968)). The equivalence of (ii) and (iii) is an immediate consequence of the fundamental theorem of coupling (see, for example, Rogers and Williams (1987)). We prove all the equivalences for the sake of completeness.

Proof of Theorem 3.1.3 Throughout the proof $\mathcal{R}$ is as defined in Remark 2.2.

(iv) $\Rightarrow$ (ii) This mimics part of the proof of the fundamental inequality for coupling. Given $\mathbb{Q}$ and the densities $(f_n^{\mathbb{Q}})$, define

$$\Omega' = \Omega \times \Omega^\infty \times [0, 1],$$
$$\mathcal{F}' = \mathcal{F} \otimes \mathcal{F}^{*\infty} \otimes \mathcal{B}([0, 1]),$$

and the probability measures T_n by

$$\frac{dT_n}{d\mathbb{Q}} = \frac{(f_n^{\mathbb{Q}} - f_\infty^{\mathbb{Q}})^+}{d(\mathbb{P}_n, \mathbb{P}_\infty)}.$$

Then define

$$\mathbb{P}' = \mathbb{P}_\infty \otimes \bigotimes_{n=1}^{\infty} T_n \otimes \Lambda$$

and define, for each $\omega' = (\omega_\infty, \omega_1, \ldots; t) \in \Omega'$,

$$X_\infty(\omega') = \omega_\infty,$$
$$X_n(\omega') = \omega_\infty 1_{\left(t \le \frac{f_\infty^{\mathbb{Q}} \wedge f_n^{\mathbb{Q}}}{f_\infty^{\mathbb{Q}}}(\omega_1)\right)} + \omega_n 1_{\left(t > \frac{f_\infty^{\mathbb{Q}} \wedge f_n^{\mathbb{Q}}}{f_\infty^{\mathbb{Q}}}(\omega_1)\right)},$$

and

$$Y(\omega') = t\,.$$

What we're doing is constructing X_∞ to have the right law under $\mathbb{P}'$; then, taking an independent $U[0,1]$ r.v. (called Y), setting $X_n = X_\infty$ if (and only if) $Y \le \frac{f_\infty^{\mathbb{Q}} \wedge f_n^{\mathbb{Q}}}{f_n^{\mathbb{Q}}}(X_\infty)$ and otherwise giving X_n a conditional distribution which gives it the right (unconditional) distribution. It's not hard to check that $\mathbb{P}'_{X_n} = \mathbb{P}_n$ for all n, whilst

$$\begin{aligned}
\mathbb{P}'(X_n \neq X_\infty) \le (=) \mathbb{P}'\big(Y > \frac{f_\infty^{\mathbb{Q}} \wedge f_n^{\mathbb{Q}}}{f_\infty^{\mathbb{Q}}}(X_\infty)\big) \\
= \int_\Omega \frac{(f_n^{\mathbb{Q}} - f_\infty^{\mathbb{Q}})^+}{f_\infty^{\mathbb{Q}}}\, d\mathbb{P}_\infty \\
= \int_\Omega (f_n^{\mathbb{Q}} - f_\infty^{\mathbb{Q}})^+ d\mathbb{Q},
\end{aligned} \tag{3.1.1}$$

and by (iv) the last term in (3.1.1) tends to 0.

(i) $\Leftrightarrow$ (iv) The reverse implication is obvious (since convergence in L^1 is equivalent to {convergence in probability *and* uniform integrability}). The forward implication is well-known since (by virtue of the fact that $f_\infty^{\mathbb{Q}}$ and $f_n^{\mathbb{Q}}$ are densities):

$$\int_\Omega |f_\infty^{\mathbb{Q}} - f_n^{\mathbb{Q}}| d\mathbb{Q} = 2\int_\Omega (f_\infty^{\mathbb{Q}} - f_n^{\mathbb{Q}})^+ d\mathbb{Q} \tag{3.1.2}$$

and the integrand on the right of (3.1.2) is uniformly bounded by $f_\infty^{\mathbb{Q}}$ (which is, by definition, in $L^1(\mathbb{Q})$).

(ii) $\Rightarrow$ (iii) This follows immediately from the coupling inequality.

(iii) $\Rightarrow$ (iv) This follows on taking the dominating measure $\mathcal{R}$:

$$d(\mathbb{P}_n, \mathbb{P}_\infty) \to 0$$

tells us that

$$\int_\Omega (f_\infty^R - f_n^R)^+ dR \to 0 \tag{3.1.3}$$

and (as before) $\int_\Omega |f_\infty^R - f_n^R| dR = 2\int_\Omega (f_\infty^R - f_n^R)^+ dR$ establishing (iv).

(iii) $\Leftrightarrow$ (v) This is obvious □

Now for Skorokhod strong convergence.

Theorem 3.1.6 Suppose $(\mathbb{P}_n)$ are probability measures on $(\Omega, \mathcal{F})$, then the following are equivalent

(i) $\mathbb{P}_n \overset{SS}{\Rightarrow} \mathbb{P}_\infty$

(ii) There exists a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ and random objects

$$(X_n) : (\Omega', \mathcal{F}', \mathbb{P}') \to (\Omega, \mathcal{F})$$

such that

(a) $\mathbb{P}'_{X_n} = \mathbb{P}_n$

and

(b) $\mathbb{P}'(X_n \neq X_\infty \text{ i.o.}) = 0$.

Proof (i) $\Rightarrow$ (ii) Take the representation given in the proof of Theorem 3.1.3, then

$$\mathbb{P}'(\exists n \geq N : X_n \neq X_\infty) = \mathbb{P}'(Y > \inf_{n \geq N} \frac{f_\infty^{\mathbb{Q}} \wedge f_n^{\mathbb{Q}}}{f_\infty^{\mathbb{Q}}}(X_\infty))$$

$$= \int_\Omega (1 - \frac{f_\infty^{\mathbb{Q}} \wedge \inf_{n \geq N} f_n^{\mathbb{Q}}}{f_\infty^{\mathbb{Q}}}(\omega))\, d\mathbb{P}_\infty(\omega)$$

$$= \int_\Omega (f_\infty^{\mathbb{Q}}(\omega) - \inf_{n \geq N} f_n^{\mathbb{Q}}(\omega))^+ d\mathbb{Q}(\omega)$$

and by monotone convergence this expression converges to

$$\int_\Omega (f_\infty^{\mathbb{Q}} - \liminf f_n^{\mathbb{Q}})^+\, d\mathbb{P}_\infty$$

$$= 0 \text{ (by (i))}.$$

(ii) $\Rightarrow$ (i) Given $\mathbb{P}'$ and (X_n) as in (ii), define $\mathbb{Q}$ as in Remark 2.2, and define, for each $m \geq 1$, the measure $\mathbb{Q}_m$ on $(\Omega, \mathcal{F})$ by

$$\mathbb{Q}_m(A) = \mathbb{P}'(\sup_{n \geq m} \delta(X_n, X_\infty) = 1, X_\infty \in A),$$

where δ is the discrete metric on Ω. Note that (since $\mathbb{Q}_m(A) \leq \mathbb{P}'(X_\infty \in A) = \mathbb{P}_\infty(A)$ by hypothesis) $\mathbb{Q}_m \ll \mathbb{P}_\infty$, whilst

$$\mathbb{Q}_m(\Omega) = \mathbb{P}'(\sup_{n \geq m} \delta(X_n, X_\infty) = 1),$$

so that

$$\lim \mathbb{Q}_m(\Omega) = \mathbb{P}'(X_n \neq X_\infty \text{ i.o.}).$$

Now

$$\begin{aligned}\mathbb{Q}_m(A) &\geq \mathbb{P}'(X_n \neq X_\infty, X_\infty \in A)\\ &\geq \mathbb{P}'(X_n \in A^c, X_\infty \in A)\\ &\geq \mathbb{P}'(X_\infty \in A) - \mathbb{P}'(X_n \in A)\\ &= \mathbb{P}_\infty(A) - \mathbb{P}_n(A) \qquad \text{(for any } n \geq m),\end{aligned}$$

so that, for any $n \geq m$,

$$g_m \stackrel{def}{=} \frac{d\mathbb{Q}_m}{d\mathbb{Q}} \geq f_\infty^{\mathbb{Q}} - f_n^{\mathbb{Q}} \quad (\mathbb{Q}\text{ a.s.}),$$

so

$$g_m \geq (f_\infty^{\mathbb{Q}} - f_n^{\mathbb{Q}})^+ \quad (\mathbb{Q}\text{ a.s.) for any } n \geq m\,.$$

It follows that $g_m \geq (f_\infty^{\mathbb{Q}} - \inf_{n\geq m} f_n^{\mathbb{Q}})^+$ ($\mathbb{Q}$ a.s.) and hence

$$\begin{aligned}0 = \lim \mathbb{Q}_m(\Omega) &= \lim \int_\Omega g_m d\mathbb{Q}\\ &\geq \lim \int_\Omega (f_\infty^{\mathbb{Q}} - \inf_{n\geq m} f_n^{\mathbb{Q}})^+ d\mathbb{Q}.\end{aligned}$$

It follows (by monotone convergence) that lim inf $f_n^{\mathbb{Q}} \geq f_\infty^{\mathbb{Q}}$ ($\mathbb{Q}$ a.s.) from which we may easily deduce (using Fatou's lemma) that lim inf $f_n^{\mathbb{Q}} = f_\infty^{\mathbb{Q}}$ ($\mathbb{Q}$ a.s.) and hence

$$f_\infty^{\mathbb{Q}} \wedge f_n^{\mathbb{Q}} \stackrel{\mathbb{Q}a.s.}{\longrightarrow} f_\infty^{\mathbb{Q}}$$

□

Remark 3.1.7 We can see from the proof of Theorem 3.1.6 that we could have adopted two slightly different definitions of Skorokhod strong convergence:

(a) lim inf $f_n^{\mathbb{Q}} = f_\infty^{\mathbb{Q}}$ ($\mathbb{Q}$ a.s.)

(b) lim inf $f_n^{\mathbb{Q}} \geq f_\infty^{\mathbb{Q}}$ ($\mathbb{Q}$ a.s.).

Remark 3.1.8 For an example of these constructions see Roberts and Jacka (1993) where we exhibit an explicit construction to demonstrate the convergence of time-inhomogeneous birth and death processes.

Remark 3.1.9 There is, of course, a still stronger form of convergence of measures:

Definition 3.1.10 If $(\mathbb{P}_n)$ are probability measures on $(\Omega, \mathcal{F})$ we say that $\mathbb{P}_n$ tends strictly to $\mathbb{P}_\infty$, written

$$\mathbb{P}_n \stackrel{strict}{\Rightarrow} \mathbb{P}_\infty\,,$$

if there is a dominating probability measure $\mathbb{Q}$ such that

$$\mathbb{P}_n \ll \mathbb{Q}, \text{ with densities } f_n^{\mathbb{Q}}$$

such that

$$f_n^{\mathbb{Q}} \stackrel{\mathbb{Q}a.s.}{\longrightarrow} f_\infty^{\mathbb{Q}}.$$

Remark 3.1.11 Theorem 3.1.6 allows us to deduce another equivalent formulation for Skorokhod weak convergence as we see in the corollary below.

Corollary 3.1.12 Suppose $(\mathbb{P}_n)$ are probability measures on $(\Omega, \mathcal{F})$ then the following are equivalent:

(i) $\mathbb{P}_n \stackrel{SW}{\Rightarrow} \mathbb{P}_\infty$

(ii) There exists a dominating probability measure $\mathbb{Q}$ such that

$$f_\infty^{\mathbb{Q}} \wedge f_n^{\mathbb{Q}} \stackrel{prob(\mathbb{Q})}{\longrightarrow} f_\infty^{\mathbb{Q}}$$

Proof This follows by looking at a.s. convergent subsequences:

(i) $\Rightarrow$ (ii) If (i) holds then by Theorem 3.1.3(ii) $\exists \mathbb{P}', (X_n)$ s.t.

(a) $\mathbb{P}'_{X_n} = \mathbb{P}_n$

and

(b) $\delta(X_n, X_\infty) \stackrel{prob(\mathbb{P}')}{\longrightarrow} 0$

so, given any subsequence (n_k) there is a sub-sub-sequence (n_{k_j}) such that

$$\delta(X_{n_{k_j}}, X_\infty) \stackrel{\mathbb{P}' a.s.}{\longrightarrow} 0$$

so, by Theorem 3.1.6,

$$f_\infty^{\mathbb{Q}} \wedge f_{n_{k_j}}^{\mathbb{Q}} \stackrel{\mathbb{Q} a.s.}{\longrightarrow} f_\infty^{\mathbb{Q}}.$$

Since the subsequence n_k is arbitrary it follows that

$$f_\infty^{\mathbb{Q}} \wedge f_n^{\mathbb{Q}} \stackrel{prob(\mathbb{Q})}{\longrightarrow} f_\infty^{\mathbb{Q}}.$$

(ii) $\Rightarrow$ (i) This is essentially the same argument: for any (n_k) there is a sub-subsequence (n_{k_j}), s.t.

$$f_\infty^{\mathbb{Q}} \wedge f_{n_{k_j}}^{\mathbb{Q}} \stackrel{\mathbb{Q} a.s.}{\longrightarrow} f_\infty^{\mathbb{Q}}$$

and hence (essentially) by Theorem 3.1.6(ii), $\exists \mathbb{P}', (X_{n_{k_j}})$ s.t.

(a) $\mathbb{P}'_{X_{n_{k_j}}} = \mathbb{P}_{n_{k_j}}$

and

(b) $\mathbb{P}'(\delta(X_{n_{k_j}}, X_\infty) \to 0) = 1$.

Since (n_k) is arbitrary, $\delta(X_n, X_\infty) \stackrel{prob(\mathbb{P}')}{\longrightarrow} 0$ and thus, by Theorem 3.1.3,

$$\mathbb{P}_n \stackrel{SW}{\Rightarrow} \mathbb{P}_\infty$$

□

Remark 3.1.13 If $(\Omega, \mathcal{F})$ is a *countable* metric measurable space, and $(\mathbb{P}^n)$ are probability measures on $(\Omega, \mathcal{F})$ then

$$\mathbb{P}^n \Rightarrow \mathbb{P}^\infty \text{ iff } \mathbb{P}^n \overset{strict}{\Rightarrow} \mathbb{P}^\infty.$$

This is because $\{\omega\}$ must be a $\mathbb{P}^\infty$-continuity set for each $\omega \in \Omega$, so that, enumerating Ω as $\{\omega_k : k \geq 1\}$, and defining $\mathbb{Q}$ by

$$\mathbb{Q}(A) = \sum_k 2^{-k} 1_{(\omega_k \in A)},$$

we have

$$\frac{d\mathbb{P}^n}{d\mathbb{Q}}(\omega_k) \equiv 2^k \mathbb{P}^n(\{\omega_k\}) \to 2^k \mathbb{P}^\infty(\{\omega_k\}) \equiv \frac{d\mathbb{P}^\infty}{d\mathbb{Q}}(\omega_k)$$

□

Remark 3.1.14 It is easy to show that (in general) these forms of convergence are all distinct (see, for example, Jacka and Roberts (1992)).

3.2 Counterexamples

We want now to show that none of our four definitions of convergence are equivalent. It is obvious that 'strict ⇒ Skorokhod strong ⇒ Skorokhod weak ⇒ strong convergence' so the following three counterexamples will do, but before we do this we'll introduce a general set-up.

In the examples (B_n) are a sequence of Bernoulli random variables:

$$(B_n) : (\Omega, \mathcal{F}) \to (\{0, 1\}, 2^{\{0,1\}}),$$

and $\mathbf{B}$ is the random vector $(B_1, B_2, \ldots)$. Note that setting $Y = \cdot B_1 B_2 \ldots$ [it being understood that a dyadic representation is being given] it follows from the fact that the Borel sets of $[0, 1]$ are generated by the intervals with dyadic rational endpoints that Y is a random variable:

$$Y : (\Omega, \mathcal{F}) \to ([0, 1], \mathcal{B}([0, 1])).$$

Example 3.2.1 (This example is due to Dudley). *Strong convergence does not imply Skorokhod weak convergence* Define $\mathbb{P}_k$ by:

(i) the (B_n) are independent under $\mathbb{P}_k$;

(ii) $\mathbb{P}_k(B_n = 1) = 1/2 + 1/2\delta_{k,n}$,

$\mathbb{P}_k(B_n = 0) = 1/2 - 1/2\delta_{k,n}$;

(iii) $\mathbb{P}_\infty$ is, of course, the 'uniform measure', ie under $\mathbb{P}_\infty$ the (B_n) are iid Bernoulli with parameter $1/2$.

In other words, under $\mathbb{P}_k$ the (B_n) are independent and for all $n \neq k$, B_n is equally likely to be 0 or 1 but (under $\mathbb{P}_k$) B_k is 1. Taking our reference measure as Lebesgue measure, Λ, on $[0,1]$ it is clear that the density of Y under $\mathbb{P}_n$ is given by f_n where:

$$f_n(x) = \begin{cases} 2: & \frac{2k-1}{2^n} \leq x < \frac{2k}{2^n};\ k = 1, \ldots, 2^{n-1} \\ 0: & \text{otherwise} \end{cases}$$

so $f_n \overset{prob}{\not\to} f_\infty \equiv 1$.

It remains to prove that $\mathbb{P}_n \overset{s}{\Rightarrow} \mathbb{P}_\infty$ or equivalently, that $\mathbb{P}_n(Y \in A) \to \Lambda(A)$, for all $A \in \mathcal{B}([0,1])$.

Now given $A \in \mathcal{B}[0,1]$ and using the equivalence mentioned above we see that

$$(Y \in A) = (\mathbf{B} \in D) = (\omega \in C)$$

for some $C \in \sigma(\mathbf{B})$.

But, using the approximation lemma, for any $C \in \sigma(\mathbf{B})$ and any $\varepsilon > 0$ there exists an $n \equiv n(\varepsilon, C)$ and a $C_n \in \sigma(B_1, \ldots, B_n)$ such that

$$\mathbb{P}_\infty(C \triangle C_n) \leq \varepsilon\,.$$

Now, for any $k > n$, $\mathbb{P}_k(C_n) = \mathbb{P}_\infty(C_n)$ (since under $\mathbb{P}_k$ the first n of the B's are iid Bernoulli) so

$$|\mathbb{P}_k(C) - \mathbb{P}_\infty(C)| \leq \mathbb{P}_k(C \triangle C_n) + \mathbb{P}_\infty(C \triangle C_n),$$

but $\frac{d\mathbb{P}_k}{d\mathbb{P}_\infty} \leq 2$ so

$$\mathbb{P}_k(C \triangle C_n) \leq 2\mathbb{P}_\infty(C \triangle C_n) \leq 2\varepsilon$$

and hence, for all $k > n(\varepsilon, C)$,

$$|\mathbb{P}_k(C) - \mathbb{P}_\infty(C)| \leq 3\varepsilon$$

establishing that $\mathbb{P}_n \overset{s}{\Rightarrow} \mathbb{P}_\infty$ □

Example 3.2.2 *Skorokhod weak does not imply Skorokhod strong convergence* Essentially we just want an example of a sequence of densities which converge in probability, but not almost surely. Given the (B_n), define $\mathbb{P}_k$ as follows: express $k = 2^n + r$ $(0 \leq r \leq 2^n - 1)$, then

(i) under $\mathbb{P}_k$, $(B_1, \ldots, B_n, B_{n+2}, \ldots)$ are iid Bernoulli (parameter $1/2$);

(ii) if $\cdot B_1 \ldots B_n$ is *not* the dyadic representation of $\frac{r}{2^n}$ then make B_{n+1} conditionally independent Bernoulli ($1/2$);

(iii) if $\cdot B_1 \ldots B_n$ *is* the representation of $\frac{r}{2^n}$, then set $B_{n+1} = 1$.

It follows, setting $Y = \cdot B_1 \ldots$ as before, that

$$f_k(x) = \begin{cases} 1: & x \notin [\frac{r}{2^n}, \frac{r+1}{2^n}) \\ 2: & x \in [\frac{r+1/2}{2^n}, \frac{r+1}{2^n}) \\ 0: & x \in [\frac{r}{2^n}, \frac{r+1/2}{2^n}) \end{cases}$$

where k (as before) is $2^n + r$ $(0 \leq r \leq 2^n - 1)$. Clearly $f_n \xrightarrow{prob} f_\infty (\equiv 1)$, since f_n differs from f_∞ only on a set of Lebesgue measure $O(\frac{1}{\log_2 n})$, but equally clearly

$$\liminf f_n = 0 \text{ Lebesgue a.e.}$$

□

Example 3.2.3 *Skorokhod strong convergence does not imply strict convergence.* Here we just content ourselves with giving f_k:

$$f_k(x) = \begin{cases} 1 - 2^{-n} : & x \notin [\frac{r}{2^n}, \frac{r+1}{2^n}) \\ 2 - 2^{-n} : & x \in [\frac{r}{2^n}, \frac{r+1}{2^n}) \end{cases}$$

where, as usual, $k = 2^n + r$ $(0 \leq r \leq 2^n - 1)$. Clearly,

$$\liminf f_n = 1,$$

but

$$\limsup f_n = 2 \text{ (Lebesgue a.e.)}$$

□

§4. Applications to sufficient statistics and conditioned Markov processes

For any unexplained notation or terminology in this and the subsequent chapter, the reader is referred to Ethier and Kurtz (1986). For a more expansive account of the formalism of sufficient statistics see Le Cam (1986).

4.1. Sufficient statistics

We need to introduce a number of definitions in order to state and prove the main result of this section. However, we trust that the benefits are reaped in the subsequent section.

Suppose $(\mathbb{P}^\theta; \theta \in \Theta)$ are a collection of probability measures on $(\Omega, \mathcal{F})$, and S is a measurable function $S : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$. Recall that S is a sufficient statistic for $(\mathbb{P}^\theta; \theta \in \Theta)$ (S is sufficient for $(\mathbb{P}_\theta)$) if $\mathbb{P}^\theta(\cdot|S = \cdot)$ (the conditional probability under θ given S) satisfies:

(i) for each $\theta \in \Theta$, $\mathbb{P}^\theta(\cdot|S = \cdot)$ is a regular conditional probability (i.e. $\mathbb{P}^\theta(\cdot|S = s)$ is a probability measure on $(\Omega, \mathcal{F})$ for each $s \in \Omega'$ *and* $\mathbb{P}^\theta(A|S = \cdot)$ is a measurable function from $(\Omega', \mathcal{F}')$ to $([0,1], \mathcal{B}([0,1]))$ for each $A \in \mathcal{F}$);

and

(ii) $\mathbb{P}^\theta(\cdot|S = \cdot)$ is independent of θ.

It's worth quoting the Factorisation Theorem carefully: we do so below.

The Factorisation Theorem Suppose that S is sufficient for $(\mathbb{P}^\theta)$; denote (the common value of) $\mathbb{P}^\theta(\cdot|S = \cdot)$ by $\tilde{\mathbb{P}}(\cdot|S = \cdot)$, and the distribution of S under $\mathbb{P}^\theta$ by $\mathbb{P}^\theta_S$.

Then, if P^S is a dominating (probability) measure for the $(\mathbb{P}^\theta_S; \theta \in \Theta)$ and $\frac{d\mathbb{P}^\theta_S}{dP^S} = f^{\theta,S}$, and, if we define $\mathbb{Q}$ by

$$\mathbb{Q}(A) = \int_{\Omega'} \tilde{\mathbb{P}}(A|S = s) dP^S(s), \tag{4.1}$$

then:

(i) $\mathbb{Q}$ is a probability measure on $(\Omega, \mathcal{F})$ and $\mathbb{P}^\theta \ll Q$ for each $\theta \in \Theta$;

(ii) the law of S under $\mathbb{Q}$ is given by

$$\mathbb{Q}_S \equiv P^S;$$

and

(iii) $f^\theta(\omega) \stackrel{def}{=} \frac{d\mathbb{P}^\theta}{d\mathbb{Q}} = f^{\theta,S}(S(\omega))$, i.e. the density at ω of $\mathbb{P}^\theta$ with respect to $\mathbb{Q}$ is the density at $S(\omega)$ of $\mathbb{P}^\theta_S$ wrt P^S.

Conversely, suppose $(\mathbb{P}^\theta; \theta \in \Theta)$ are dominated by the probability measure $\mathbb{Q}$, with densities $\frac{d\mathbb{P}^\theta}{d\mathbb{Q}}$ given by

$$\frac{d\mathbb{P}^\theta}{d\mathbb{Q}}(\omega) = f^\theta(S(\omega))\,.$$

Then, defining the conditional $\mathbb{Q}$-probabilities given S as $\mathbb{Q}(\cdot|S = \cdot)$, we have:

$$\mathbb{P}^\theta(A|S = s) = \mathbb{Q}(A|S = s),$$

so that if $\mathbb{Q}(\cdot|S = \cdot)$ is a collection of regular conditional probabilities, then:

(iv) S is sufficient for the $\mathbb{P}^\theta$;

(v) $\mathbb{Q}_S$ is a dominating probability measure for the $(\mathbb{P}^\theta_S)$

and

(vi) $\frac{d\mathbb{P}^\theta_S}{d\mathbb{Q}_S}(s) = f^\theta(s)$. (4.2)

It is this converse which is usually used to identify sufficient statistics.

Remark 4.1.1 The existence of regular conditional probabilities is in general very hard to establish but the following theorem is generally applicable.

Theorem 4.1.2 (see Shiryayev (1984) Theorem II.7.5) If $(\Omega', \mathcal{F}')$ is a Borel space, and X is a random object taking values in $(\Omega', \mathcal{F}')$, i.e.

$$X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}'),$$

then for any probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ and any sub σ-field $\mathcal{G}$ of $\mathcal{F}$, there is a (unique) regular conditional distribution for X given $\mathcal{G}$).

In particular, if $(\Omega, \mathcal{F})$ is a Borel space, then, taking X to be the identity, for any probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ and any sub σ-field $\mathcal{G}$ of $\mathcal{F}$, regular conditional probabilities $\mathbb{Q}(\cdot|\mathcal{G})$ exist.

Remark 4.1.3 The definition of a Borel space may also be found in Shiryayev (1984)—who points out that any complete, separable metric space (equipped with its Borel sets) is a Borel space. In particular if Ω is $D([0,\infty),\mathbb{Z}^+)$, $C^n([0,\infty])$ or $D^n([0,\infty])$ and $\mathcal{F}$ is the corresponding Borel σ-field (under the usual topology) then $(\Omega,\mathcal{F})$ is a Borel space.

Remark 4.1.4 See also Theorem II.69 of Williams (1979)—this tells us that if $(\Omega,\mathcal{F})$ is a Polish space equipped with its Borel sets then any countably generated sub-σ-field $\mathcal{G}$ gives rise to regular conditional probability measures under any probability measure on $(\Omega,\mathcal{F})$.

We are now nearly in a position to give our general result on convergence in the presence of a sufficient statistic but firstly, to simplify reference to them, we shall use the following nomenclature for types of convergence of measure: weak, strong, Skorokhod weak, Skorokhod strong, and strict convergence will henceforth be referred to as convergence type 0 to 4 respectively. Secondly, we shall extend the definitions of these types of convergence in the following way:

if Θ is any subset of $(-\infty,\infty]$ which is unbounded from above, and $(\mathbb{P}^\theta;\theta\in\Theta)$ are a collection of probability measures on $(\Omega,\mathcal{F})$, we say

$$\mathbb{P}^\theta \Rightarrow \mathbb{P}^\infty \text{ (type } i\text{) or } \mathbb{P}^\theta \overset{\boxed{type\ i}}{\Longrightarrow} \mathbb{P}^\infty\,,$$

for $i=0,1,\ldots,4$, if for every sequence $(\theta_k)\subseteq\Theta$ with $\theta_k\to\infty$ as $k\to\infty$ † ,

$$\mathbb{P}^{\theta_k} \Rightarrow \mathbb{P}^\infty \text{ (type } i\text{) for the same value of } i\,.$$

Now we can state our next theorem.

Theorem 4.1.5 Suppose Θ is such an index set, $(\mathbb{P}^\theta;\theta\in\Theta)$ are a collection of probability measures on $(\Omega,\mathcal{F})$ and S is sufficient for $(\mathbb{P}^\theta)$, then, for each $i=1,\ldots,4$,

$$\mathbb{P}^\theta \overset{\boxed{type\ i}}{\Longrightarrow} \mathbb{P}^\infty$$

if and only if

$$\mathbb{P}^\theta_S \overset{\boxed{type\ i}}{\Longrightarrow} \mathbb{P}^\infty_S$$

Proof Since S is sufficient for $(\mathbb{P}^{\theta_k})$ for any sequence (θ_k) with $\theta_k\to\infty$ we may restrict attention to the countable case. The result now follows (essentially) from the fact that each type of convergence may be defined in terms of the densities $\frac{d\mathbb{P}^{\theta_k}}{d\mathbb{Q}}$ (or equivalently $\frac{d\mathbb{P}^{\theta_k}_S}{d\mathbb{Q}_S}$).

† We may make suitable generalisations for $\Theta\subseteq\mathbb{S}$, where $\mathbb{S}$ is a complete metric space.

Perhaps the only equivalence which is not clear is type 1; to prove this notice that

$$\begin{aligned}\mathbb{P}^{\theta_k}(A) &= \int_{\Omega'} \mathbb{P}^{\theta_k}(A|S=s)d\mathbb{P}_S^{\theta_k}(s)\\ &= \int_{\Omega'} \tilde{\mathbb{P}}(A|S=\cdot)d\mathbb{P}_S^{\theta_k}\end{aligned}$$

and since $\tilde{\mathbb{P}}(A|S=s)$ is a bounded measurable function of s it follows from Theorem 3.1.1 (ii) that $\mathbb{P}^{\theta_k} \overset{\boxed{type\ 1}}{\Longrightarrow} \mathbb{P}^\infty$. □

4.2. Convergence of conditioned Markov processes

We consider now an application to conditioned Markov processes. We suppose that $\mathbb{S}$ is a statespace, and for concreteness we suppose that $\mathbb{S}$ is either $\mathbb{R}^n$ or $\mathbb{Z}^+$. We suppose that $(\mathbb{P}_{x,t}; x \in \mathbb{S}, t \in \mathbb{R}_+)$ constitute a collection of time-inhomogeneous strong Markov probability measures on $D([0,\infty);\mathbb{S})$ equipped with its Borel sets (henceforth denoted by $(\Omega, \mathcal{F})$), and the usual filtration. X is the process given by

$$X_t(\omega) = \omega_t$$

and X is, of course, a time-inhomogeneous strong Markov process under each $\mathbb{P}_{x,t}$. Finally $(\vartheta_t; t \in \mathbb{R}_+)$ is the usual collection of shift operators:

$$\vartheta_t : \omega \mapsto \omega^t$$

where $\omega_s^t = \omega_{t+s}$.

We define $\mathcal{G}_t = \sigma(\{X_s : s \geq t\})$.

We shall now define the sort of events on which we want to condition.

Definition 4.2.1 An event A in $\mathcal{F}$ is said to be *$t-$decomposable* if it can be written as

$$A = B_t \cap C_t,$$

for some $B \in \mathcal{F}_t$ and $C \in \mathcal{G}_t$.

Definition 4.2.2 An event A in $\mathcal{F}$ is said to be *uniformly $t-$decomposable* if it is s-decomposable for each $s \leq t$.

Definition 4.2.3 Suppose that $\mathbb{T} \subseteq \mathbb{R}_+$. A collection of events $\{A_T : T \in \mathbb{T}\}$ in $\mathcal{F}$ is said to be *uniformly decomposable* if, for each $T \in \mathbb{T}$, A_T can be written as

$$A_T = B_s \cap C_s^T,$$

for some $B_s \in \mathcal{F}_s$ and $C_s^T \in \mathcal{G}_s$, for each $s \leq T$.

Remark 4.2.4 Note that in Definition 4.2.3 we are asking slightly more than that each A_T is uniformly T-decomposable; we're asking that the 'B sets' are the same in each case.

Remark 4.2.5 The sort of situation we're envisaging is where the (A_T) are all of the form

$$A_T = (T < \tau < \infty),$$

where τ is a, possibly infinite, Markov time (i.e a stopping time satisfying $(\tau > \ t+s) = (\tau > t) \cap (\tau \circ \vartheta_t > \ s)$ for all $s, t > 0$). In this case we can take

$$B_t = (\tau > t) \text{ and } C_t^T = (\infty > \tau \circ \vartheta_t > T - t).$$

Remark 4.2.6 Notice that if, for each i, (A_T^i) is a uniformly decomposable collection with index set $\mathbb{T}$, then, defining $A_T^\infty = \bigcap_{i=1}^\infty A_T^i$,

$$\{A_T^\infty : T \in \mathbb{T}\} \text{ is a uniformly decomposable collection.}$$

We suppose that we are given a uniformly decomposable collection of events $\{A_T : T \in \mathbb{R}\}$, satisfying

$$\mathbb{P}_{x,t}(A_T) > 0$$

for some non-trivial collection $\mathcal{X}$ of (x, t) pairs. We define, for each $(x, t) \in \mathcal{X}$:

$$\mathbb{P}_{x,t}^T(\cdot) = \mathbb{P}_{x,t}(\cdot | A_T)$$

and we wish to consider the convergence of $\mathbb{P}_{x,t}^T$ as $T \to \infty$. In fact we shall only consider convergence of $\mathbb{P}_{x,t}^T|_{\mathcal{F}_R}$ for arbitrary but finite R. We denote such probability measures by $\mathbb{P}^{T,R}$ (suppressing the dependence on x and t).

Lemma 4.2.7 Setting $\mathbb{R}_R = [R, \infty)$, the value of X at time R, X_R, is sufficient for $(\mathbb{P}_{x,t}^{T,R}; T \in \mathbb{R}_R)$ for each fixed x and t.

Proof Denote $(X_t; 0 \leq t \leq R)$ by X^R, then

$$\begin{aligned}
&\mathbb{P}_{x,t}^T(X^R \in A | X_R = s) \\
&= \mathbb{P}_{x,t}(X^R \in A | A_t \cap X_R = s) \\
&= \mathbb{P}_{x,t}(X^R \in A | B_R \cap C_R^T \cap (X_R = s)) \\
&= \frac{\mathbb{P}_{x,t}((X^R \in A) \cap B_R | C_R^T \cap (X_R = s))}{\mathbb{P}_{x,t}(B_R | C_R^T \cap (X_R = s))}
\end{aligned}$$

and by the Markov property this is

$$\begin{aligned}
&\mathbb{P}_{x,t}((X^R \in A) \cap B_R | X_R = s) / \mathbb{P}_{x,t}(B_R | X_R = s) \\
&= \mathbb{P}_{x,t}(X^R \in A | B_R \cap (X_R = s))
\end{aligned} \tag{4.3}$$

which is independent of T. Moreover it is clear from (4.3) that $\mathbb{P}_{x,t}^T(\cdot | X_R = s)$ constitutes a regular conditional probability measure. □

Corollary 4.2.8 If the law of X_R under $\mathbb{P}_{x,t}^T$ converges as $T \to \infty$ (type 1 to 4) then the law of X^R under $\mathbb{P}_{x,t}^T$ converges (type 1 to 4) and, in particular, if $\mathbb{S}$ is countable, and the law of X_R under $\mathbb{P}^T_{x,t}$ converges weakly (as $T \to \infty$) then the law of X^R under $\mathbb{P}^T_{x,t}$ converges type 4 as $T \to \infty$.

§5. Convergence of time-inhomogeneous Markov processes

5.1. Time-inhomogeneous Markov chains

We assume that $(\mathbb{P}^n)$ are a collection of probability measures on $D([0,\infty);\mathbb{Z}^+)$: under $\mathbb{P}^n$, X (given by $X_t(\omega) = \omega_t$) is a time-inhomogeneous *non-explosive* Markov chain with initial distribution (p^n_i). We assume the existence of a dominating measure μ (finite on compact sets) with respect to which each probability measure has transition rates $q^n_{i,j}(t)$ $(t \geq 0, i, j \in \mathbb{Z})$ and, as usual we write $q^n_i(t) = -q^n_{i,i}(t)$.

Recall the following notation: we write

$$f_n \overset{L^1_{\text{loc}}(\mu)}{\longrightarrow} f$$

if, for all $T \geq 0$, $\int_{[0,T]} |f_n - f|\, d\mu \to 0$ as $n \to \infty$; and we write

$$f_n \overset{(\mu \text{ loc})}{\longrightarrow} f$$

if, for all $T \geq 0$, and for each $\varepsilon > 0$, $\mu\{t \in [0,T] : |f^n(t) - f(t)| > \varepsilon\} \to 0$.

Theorem 5.1.1 (a) If

$$p^n_i \to p^\infty_i \text{ as } n \to \infty \text{ for each } i; \tag{5.1}$$

$$q^n_i \overset{L^1_{\text{loc}}(\mu)}{\longrightarrow} q^\infty_i \text{ as } n \to \infty \text{ for each } i; \tag{5.2}$$

and

$$q^n_{i,j} \overset{\mu a.e.}{\longrightarrow} q^\infty_{i,j} \text{ for each } i \text{ and } j \text{ in } \mathbb{Z}^+; \tag{5.3}$$

then, for each $T > 0$,

$$\mathbb{P}^n|_{[0,T]} \overset{\boxed{type\ 4}}{\Longrightarrow} \mathbb{P}^\infty|_{[0,T]}$$

(b) If (5.1) and (5.2) hold and

$$q^n_{i,j} \overset{(\mu \text{ loc})}{\longrightarrow} q^\infty_{i,j} \text{ for each } i, j \text{ in } \mathbb{Z}^+ \tag{5.4}$$

then for, each $T > 0$,

$$\mathbb{P}^n|_{[0,T]} \overset{\boxed{type\ 2}}{\Longrightarrow} \mathbb{P}^\infty|_{[0,T]}$$

Remark 5.1.2 We stress that we are assuming that $\mathbb{P}^\infty$ is non-explosive.

Proof We give first a dominating (probability) measure $\mathbb{Q}$: it is specified by having waiting time distribution " exponential (μ)" in each state, i.e. $q_i(t) \equiv 1$ for each i. Under $\mathbb{Q}$, the jump chain forms a sequence of iid geometric(½) r.v.s so that $q_{i,j}(t) = 2^{-(j+1)}$

and $\mathbb{Q}(X_0 = i) = 2^{-(i+1)}$. We assume that μ is continuous i.e. non-atomic. It is then obvious that the density of $\mathbb{P}^k|_{[0,T]}$ wrt $\mathbb{Q}|_{[0,T]}$ is $f^k \equiv f_T^k$ given by

$$\begin{aligned} f^k(\omega) =& e^{\mu([0,T])} p^k_{\omega_0} 2^{\left(\sum_{n=0}^N (\omega_{T_n}+1)\right)} \exp(-\int_{T_N}^T q^k_{\omega_{T_N}}(t) d\mu(t)) \\ & \times \prod_{n=1}^N q^k_{\omega_{T_{n-1}},\omega_{T_n}}(T_n) \exp(-\int_{T_{n-1}}^{T_n} q^k_{\omega_{T_{n-1}}}(t) d\mu(t)), \end{aligned} \tag{5.5}$$

where $N = N^T(\omega) = \#$ {jumps of X on $[0,T]$}, $T_0 = 0$, and T_n $(1 \leq n \leq N)$ are the successive jump times of X (on $[0,T]$). Finally, since $\mathbb{Q}$ is non-explosive, notice that for any $\varepsilon > 0$, there is an $n(\varepsilon)$ s.t. $\mathbb{Q}(N > n) \leq \varepsilon/2$ and then $\exists\ m(n(\varepsilon),\varepsilon)$ s.t.

$$\mathbb{Q}(X \text{ leaves } \{0,\ldots,m\} \text{ before } T) \leq \varepsilon/2.$$

Denote the union of the two sets involved in these statements by A_ε. We are now ready to prove (a). Under the assumption (5.2)

$$e^{-\int_u^v q_i^k(t)d\mu(t)} \to e^{-\int_u^v q_i^\infty(t)d\mu(t)},$$

for any $0 \leq u \leq v \leq T$. Hence, off A_ε, there are only finitely many terms in (5.5) and (by (5.1) and (5.3)) each converges $\mathbb{Q}$ a.s. to the corresponding term in f^∞. Thus $\mathbb{Q}(f^k \not\to f^\infty) \leq \mathbb{Q}(A_\varepsilon) \leq \varepsilon$ and since ε is arbitrary we have established (a).

To prove (b) we need only take subsequences: given a subsequence (n_k) take a sub-subsequence (n_{k_j}) (by diagonalisation), along which (5.3) holds (at least for $t \in [0,T]$) then $f^{n_{k_j}} \overset{\mathbb{Q}a.s.}{\longrightarrow} f^\infty$ as $j \to \infty$ by (a). The subsequence is arbitrary so $f^n \overset{prob(\mathbb{Q})}{\longrightarrow} f^\infty$ □

Remark 5.1.3 The proof of Theorem 5.1.1 only deals with the case where μ is non-atomic; if μ has atoms there is no great additional difficulty: we simply need to replace $\exp(-\int f d\mu)$ by $\exp(-\int f d\mu^c)\prod(1 - f\Delta\mu)$ wherever such terms appear in (5.5).

Remark 5.1.4 If some of the $\mathbb{P}^k$ are explosive we can restrict attention to the time-interval $[0, T \wedge \tau_n]$, where $\tau_n \overset{def}{=} \inf\{t : X_t \geq n\}$ and retain (in this more restricted setting) the results of Theorem 5.1.1.

Remark 5.1.5 If we retain the hypothesis that $\mathbb{P}^\infty$ is non-explosive but allow some of the $\mathbb{P}^k$ $(k < \infty)$ to be explosive then (the RHS of) (5.5) gives a lower bound for f^k†. It follows that, under these circumstances:

† Of course, an explosive $\mathbb{P}^k$ is not a probability measure on $D([0,\infty);\mathbb{Z}^+)$, but its restriction to $D([0,T);\mathbb{Z}^+)$ is a sub-probability measure and f^k is the density of this restriction.

(a) if (5.1) to (5.3) hold then, by Remark 3.1.7 (b),

$$\mathbb{P}^k|_{[0,T]} \overset{\boxed{type\ 3}}{\Longrightarrow} \mathbb{P}^\infty|_{[0,T]};$$

whilst

(b) if (5.1), (5.2), and (5.4) hold then, as before,

$$\mathbb{P}^k|_{[0,T]} \overset{\boxed{type\ 2}}{\Longrightarrow} \mathbb{P}^\infty|_{[0,T]}.$$

5.2. Time-inhomogeneous Itô diffusions

By a (time-inhomogeneous) Itô diffusion we mean a solution to a stochastic integral equation of the form

$$X_t = x_0 + \int_0^t \sigma_s(X_s)dB_s + \int_0^t \mu_s(X_s)ds \tag{5.6}$$

where B is a d-dimensional Brownian motion, σ_s is $n \times d$ and μ_s is $n \times 1$, and σ and μ are such that the solution is *strict* (i.e. adapted to the filtration of B—see Rogers and Williams (1987)).

We use the following version of the Cameron-Martin-Girsanov formula (a straightforward adaptation of Rogers and Williams (1987), Theorems IV.38.5 and IV.38.9).

Theorem 5.2.1 Suppose $\mathbb{P}$ and $\mathbb{Q}$ are probability measures on $C([0,\infty),\mathbb{R}^d)$ (equipped with its natural filtration), τ is some stopping time, b is a previsible (predictable) process, and Z^τ is given by

$$Z_t^\tau = \exp(\int_0^{t\wedge\tau} b_s d\omega_s - {}^1\!/_2 \int_0^{t\wedge\tau} |b_s|^2 ds).$$

Then if, under $\mathbb{P}$, ω is a Brownian motion (BM) started at 0, whilst under $\mathbb{Q}$, ω' given by

$$\omega'_t = \omega_t - \int_0^t b_s ds\,,$$

is a BM on $[0,\tau]$, then, provided Z^τ is uniformly integrable, (Z_t^τ) is a ui martingale and $\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_\tau} = Z_\tau^\tau$.

We consider the case where we have a collection of solutions to stochastic integral equations with the same diffusion co-efficient.

Corollary 5.2.2 Suppose $(X^k(x;B))$ are strict solutions to the SDEs

$$X_t^k = x_0 + \int_0^t \sigma_s(X_s^k)dB_s + \int_0^t \mu_s^k(X_s^k)ds, \tag*{$(5.7)_k$}$$

where B is a $\mathbb{P}$-BM, the initial distribution of X^k is to be ν^k, and $\nu^k \ll \nu$, for each k, where ν is a dominating probability measure

Define $\mathbb{Q}$ (a probability measure on $\mathbb{R}^n \times C([0,\infty),\mathbb{R}^d)$) by

$$\mathbb{Q} = \nu \otimes \mathbb{P}.$$

Set

$$\rho^k_t(x) = \sigma'_t(x)(\sigma_t(x)\sigma'_t(x))^{-1}(\mu^k_t(x) - \mu^\infty_t(x)),$$

and define

$$\tilde{Z}^k_t(x;B) \stackrel{def}{=} \frac{d\nu^k}{d\nu}(x) Z^k_t(x;B),$$

where

$$Z^k_t(x;B) \stackrel{def}{=} \exp(\int_0^t \rho^k_s(X^\infty_s(x;B))dB_s - 1/2 \int_0^t |\rho^k_s(X^\infty_s(x;B))|^2 ds).$$

Define $\mathbb{P}^k(x;\cdot)$ by

$$\frac{d\mathbb{P}^k}{d\mathbb{P}}(x;B) \stackrel{def}{=} Z^k_\tau(x;B),$$

and $\tilde{\mathbb{P}}^k$ by

$$\frac{d\tilde{\mathbb{P}}^k}{d\mathbb{Q}} \stackrel{def}{=} \tilde{Z}^k_\tau(x;B). \tag{5.8}$$

Then, if $(Z^k_{t\wedge\tau})$ is ui (for each k): $\mathbb{P}^k$ and $\tilde{\mathbb{P}}^k$ are probability measures and under $\tilde{\mathbb{P}}^k$, X^∞ is the solution to $(5.7)_k$ [with B replaced by $B^k \stackrel{def}{=} B - \int_0^\cdot \rho^k_s(X^\infty_s)ds$—a $\tilde{\mathbb{P}}^k$-BM (at least on $[0,\tau]$)], and has initial distribution ν^k.

Thus, if $i, j \geq 2$, and

$$\nu^k \xrightarrow{\boxed{type\ i}} \nu^\infty$$

and

$$\mathbb{P}^k|_{[0,\tau]}(x;\cdot) \xrightarrow{\boxed{type\ j}} \mathbb{P}^\infty|_{[0,\tau]}(x;\cdot) \text{ for each } x \in \mathbb{R}^n, \tag{5.9}$$

then

$$\tilde{\mathbb{P}}^k|_{[0,\tau]} \xrightarrow{\boxed{type\ i\wedge j}} \tilde{\mathbb{P}}^\infty|_{[0,\tau]},$$

which implies that

$$X^k|_{[0,\tau]} \xrightarrow{\boxed{type\ i\wedge j}} X^\infty|_{[0,\tau]}.$$

Proof All we've done is to expand the sample space to include the initial distribution of the X's. Thus the first paragraph of the corollary follows immediately from Theorem 5.2.1 on substituting $B = B^k + \int_0^\cdot \rho^k_s(X^\infty(x;B))ds$.

The second paragraph follows from the density characterisation of type i convergence and the representation (5.8) □

Remark 5.2.3 Notice that if the ν^k are all equal to δ_{x_0}, the point mass at x_0, then we need only require convergence of the $\mathbb{P}^k(x_0;\cdot)$ in the corollary.

The corollary above is a little bit ethereal so let's give a slightly more concrete application.

Theorem 5.2.4 Suppose that σ and μ^∞ are such that X^∞ is a strict solution to (5.7). Suppose that

$$\rho^k_\cdot(\cdot) \to 0 \text{ uniformly on compact subsets of } \mathbb{R}^+ \times \mathbb{R}^n, \tag{5.10}$$

then, defining

$$\tau_N = \inf\{t : |X^\infty_t| \geq N\},$$

the following are equivalent for each $i \geq 2$:

(i) $\nu^k \overset{\boxed{type\ i}}{\Longrightarrow} \nu^\infty$;

and

(ii) $X^k|_{[0,\tau_N]} \overset{\boxed{type\ i}}{\Longrightarrow} X^\infty|_{[0,\tau_N]}$ for any (and then for all) $N > 0$.

Proof (i) $\Rightarrow$ (ii): it is a fairly trivial application of stochastic calculus to show that (5.10) implies that

$$\mathbb{E}_{\mathbb{P}} \sup_{0\leq t\leq \tau_N} (Z^j_t(x;\cdot)-1)^2 \leq c^N_k(t) \text{ for all } j \geq k \text{ and } x \in \mathbb{R}^n,$$

for a suitable sequence of functions $c^N_k(\cdot)$, and that $c^N_k(\cdot) \to 0$ as $k \to \infty$. It follows that condition (5.9) holds with $j = 4$.

(ii) $\Rightarrow$ (i): this is immediate on observing that X^k_0 is a measurable function of $X^k|_{[0,\tau_N]}$ (for any N) □

Remark 5.2.5 In the case where the diffusion coefficient σ varies with k, these results break down—essentially because different σs give rise to mutually singular measures. In the one-dimensional case some results may be obtained by time-changing—the details are left to the reader.

Acknowledgments We are grateful to Wilfrid Kendall, and Robin Reed for several stimulating discussions on the topics in this article.

References:

Bauer, H. (1981): “Probability Theory and Elements of Measure Theory” (2nd English edition). Academic Press, London.

Billingsley, P. (1965): “Convergence of Probability Measures”. Wiley, New York,

Dunford, N. and Schwartz, J.T. (1958): Linear Operators; Part I: General Theory. Interscience, New York.

Ethier, S.N. and Kurtz, T.G. (1986): “Markov Processes: Characterisation and Convergence”. Wiley, New York.

Jacka, S.D. and Roberts, G.O. (1992): Strong Forms of Weak Convergence. University of Warwick, Department of Statistics Research Report no. 243.

Le Cam, L. (1986): Asymptotic Methods in Statistical Decision Theory. Springer-Verlag, New York.

Roberts, G.O. and Jacka, S.D. (1993): Weak Convergence of Conditioned Birth and Death Processes. *To appear in* JAP.

Rogers, L.C.G. and Williams, D. (1987): “Diffusions, Markov Processes, and Martingales Vol. 2”. Wiley, New York.

Shiryayev, A.N. (1984): “Probability”. Springer, Berlin.

Williams, D. (1979): “Diffusions, Markov Processes, and Martingales Vol. 1”. Wiley, New York.