WEAK CONVERGENCE AND THE GENERAL THEORY OF PROCESSES

(Incomplete draft of monograph / July 1981)

David J. Aldous

Department of Statistics
University of California
Berkeley, CA 94720

PREFACE

Billingsley's Convergence of Probability Measures (1968) was perhaps the major influence in making the theory of weak convergence into a standard part of both theoretical and applied probability. The purpose of this book is to develop one or two further topics in weak convergence. We assume the reader is familiar with the material presented by Billingsley; the only other prerequisite is knowledge of stochastic processes up to the strong Markov property of Brownian motion and the optional sampling theorem for submartingales - Breiman (1968) is recommended.

One topic is to connect weak convergence with two post-1968 developments, the "general theory of processes" of the Strasbourg school and the martingale approach to stochastic processes (these developments are surveyed in Chapter 2). In Chapter 3 we describe the systematic technique for establishing weak convergence to a limit process with a martingale characterisation; this is based upon a stopping time criterion for tightness (Section 4).

The major topic is the development (Chapter 5 onwards) of a new variant of weak convergence, extended weak convergence. This is motivated by both theoretical and practical considerations. Theoretically, extended weak convergence is designed to be compatible with the techniques of the Strasbourg school (e.g. the Doob-Meyer decomposition of submartingales, Section 19), with which classical weak convergence is incompatible. For the practical motivation, consider the prototype weak convergence result: approximating normalised partial sums $(n^{-\frac{1}{2}}S_{[nt]})$ by Brownian motion (W_t). The purpose of the weak convergence formulation is to be able to deduce convergence of functionals $\Lambda(n^{-\frac{1}{2}}S_{nt})$ to $\Lambda(W_t)$, for functionals Λ of sample paths:

typical functionals of interest being $\Lambda(f) = \sup_{t \leq 1} f(t)$ and $\Lambda(f) = \text{measure}\{t \leq 1 : f(t) > 0\}$. However, when approximating one process (X_t) by another we may be interested in something more than the distributions of functionals of sample paths. For example, the optimally-stopped value $\sup\{EX_T : T \text{ a stopping time}\}$ of a process cannot be represented by a functional of sample paths, so weak convergence cannot handle questions of convergence of optimally-stopped values (Section 11), whereas extended weak convergence can (Section 17).

This is unashamedly a "theoretical" book, but it is not merely theory for its own sake. Rather, we are trying to present theoretical tools which researchers interested in specific applications may find useful. The emphasis is on techniques and concepts, not theorems and proofs. In the first three chapters, where the material will be partly familiar to the reader and is accessible elsewhere, proofs will sometimes be sketchy. But we have tried to be more scrupulous later.

Finally, it should be emphasised that this book represents the author's personal viewpoint. Since a preliminary version was circulated in 1978, interest in connections between weak convergence and the general theory of processes has started to grow; but this book does not purport to be a definitive account of this growing field. For instance, it seemed unreasonable to expect the reader interested in applications to master technical aspects of the general theory. So we make no mention of semimartingales, though theoreticians would regard semimartingales as the natural setting for the results of Chapter 3.

CONTENTS

CHAPTER 1 - WEAK CONVERGENCE

1 BACKGROUND

As mentioned in the preface, the reader is assumed to be familiar with the basics of measure-theoretic probability and stochastic processes, and with the theory of weak convergence as expounded in Billingsley (1968): references of the form (B.5.1) refer to Theorem 5.1 of that book. This first section is intended to establish some notation, refresh the reader's memory of some fundamental results and techniques, and mention some variations of these fundamental results.

WEAK CONVERGENCE ON ABSTRACT SPACES

We consider only Polish (i.e. complete separable metric) spaces (S,d). $\mathcal{L}(X)$ denotes the distribution of a random element X of S. $\mathcal{P}(S)$ is the space of probability measures μ on S, equipped with the topology of weak convergence. $\mu_n \Rightarrow \mu$ denotes weak convergence of probability measures. For random elements (X_n), $X_n \xrightarrow[p]{} X$ denotes convergence in probability, i.e. $d(X_n,X) \xrightarrow[p]{} 0$ on R. And $X_n \Rightarrow X$ denotes convergence in distribution, i.e. $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$. We shall say $X_n \Rightarrow X$ <u>on S</u> if there is any possible doubt about the range space S.

Because S is Polish, tightness of a sequence (X_n) of random elements is equivalent to precompactness (i.e. relative compactness) of the sequence of distributions $\mathcal{L}(X_n)$ (B.6.1,6.2).

Consider a map $h:S \rightarrow S'$, and suppose $X_n \Rightarrow X$ on S. The continuous mapping theorem (B.5.1) says: if h is continuous then $h(X_n) \Rightarrow h(X)$ on S'. It is sometimes helpful to view this more abstractly. Define the <u>induced</u> map $\tilde{h}:\mathcal{P}(S) \rightarrow \mathcal{P}(S')$ by

$$(1.1) \qquad \tilde{h}(\mathcal{L}(X)) = \mathcal{L}(h(X)).$$

The continuous mapping theorem may be paraphrased as: if h is continuous then $\tilde{h}$ is continuous.

Recall the following fundamental technique for proving weak convergence $X_n \Rightarrow X$.

(i) Prove (X_n) is tight.

(ii) Consider an arbitrary subsequential weak limit Y, and prove $\mathcal{L}(Y) = \mathcal{L}(X)$.

In the proofs in (B) of weak convergence on the function spaces $C[0,1]$ and $D[0,1]$, step (ii) is achieved by proving convergence of finite-dimensional distributions. Our proofs (Chapter 3) use martingale characterisations, and thus avoid the necessity to prove convergence of finite-dimensional distributions.

The next theorem gives an extremely useful technique for establishing consequences of weak convergence.

SKOROHOD REPRESENTATION THEOREM. Suppose $X_n \Rightarrow X_\infty$ on S. Then there exist, on some probability triple, random elements $X_1', X_2', \ldots; X_\infty'$ such that

(i) $\mathcal{L}(X_n') = \mathcal{L}(X_n)$, $n = 1,2,\ldots$;

(ii) $X_n' \to X'$ a.s.

See Billingsley (1971) for a proof. An example of its use will be given shortly.

APPROXIMATING MEASURABLE FUNCTIONS

In measure theory, one often verifies identities involving measurable functions by the following technique:

(i) verify the identity for simple functions

(ii) prove thst the class of functions satisfying the identity is closed under pointwise convergence.

In the context of weak convergence, it is more natural to start with continuous functions. Then we need the following result (Halmos (1950) p. 241).

(1.2) LEMMA. Let $\phi : S \to R$ be measurable, $|\phi| \leq 1$, and let X be a random element of S. Then there exist continuous $\phi_n : S \to R$, $|\phi_n| \leq 1$, such that

$$E\,|\phi(X) - \phi_n(X)| \to 0.$$

UNIFORM INTEGRABILITY

Call a sequence (X_n) of real-valued random variables uniformly integrable if

(1.3) $$\lim_{\lambda\to\infty} \limsup_{n\to\infty} E|X_n| 1_{(|X_n|\geq\lambda)} = 0$$

Note that (1.3) implies

(1.4) $$E|X_n| < \infty \quad , n \geq n_0$$

for some n_0. The usual definition of uniform untegrability requires that (1.4) hold for all $n \geq 1$. In the context of limit theorems, the distinction is plainly irrelevant. Here are some of the basic facts about uniform integrability

(1.5) LEMMA. *If* (X_n) *is uniformly integrable and if* $P(A_n) \to 0$ *then* $E|X_n| 1_{A_n} \to 0$.

(1.6) LEMMA. *Suppose* $X_n \to X$ a.s. *Then the following are equivalent*.

(i) (X_n) *is uniformly integrable*;

(ii) $E|X| < \infty$ *and* $E|X_n - X| \to 0$;

(iii) $E|X| < \infty$ *and* $E|X_n| \to E|X|$.

We now take the opportunity to illustrate the use of the Skorohod representation theorem.

(1.7) COROLLARY. *Suppose* $X_n \Rightarrow X$. *Then* (i) *and* (iii) *of Lemma 1.6 are equivalent*.

Proof, Let $X_1', X_2', \ldots, X'$ be as in the Skorohod representation theorem. Observe that condition (i) (resp. (iii)) of Lemma 1.6 depends only on the individual distributions $\mathcal{L}(X_1)$, $\mathcal{L}(X_2), \ldots, \mathcal{L}(X)$. So if condition (i) (resp. (iii)) is satisfied by (X_n) and X then it is also satisfied by (X_n') and X'. So by Lemma 1.6, condition (iii) (resp. (i)) also is satisfied by (X_n') and X'. Using the initial observation again, condition (iii) (resp. (i)) is satisfied by (X_n) and X.

Note carefully why we cannot prove (i) implies (ii): it is because condition (ii) involves the bivariate distributions $\mathcal{L}(X_n, X)$, and the Skorohod representation theorem does not guarantee that any joint distributions of (X'_n) should coincide with those of (X_n).

Readers unfamiliar with this technique practise by proving (B.5.1 and 5.5) using the Skorohod representation theorem. Once understood, it is not necessary to be as pedantic as we were in the proof of Corollary 1.7: instead, we could simply write

Proof We may assume (Skorohod representation) that $X_n \to X$ a.s.. Apply Lemma 1.6.

Let us record a useful variant of the definition (1.3).

LEMMA. (X_n) *is uniformly integrable if and only if*

(1.8) $$E|Y_n| 1_{(|Y_n| \geq n)} \to 0$$

for every subsequence $Y_n = X_{j_n}$.

In Chapter 3 we shall need to prove uniform integrability of sequences of stochastic processes. We shall see that natural stopping time arguments lead to estimates of the type

(i) $E(|X_n|-\lambda) 1_{(|X_n| > \lambda)} \leq C\, EV_n 1_{A_{n,\lambda}}$; $\lambda \geq \lambda_0$, $n \geq 1$; where

(1.9)(ii) (V_n) is uniformly integrable,

(iii) $P(A_{n,\lambda_n}) \to 0$ for any $\lambda_n \to \infty$.

And we shall need to prove this implies (X_n) is uniformly integrable. It is generally easier to use (1.8) rather than (1.3) in such cases. For example, from (1.9) we argue

$$\limsup_{n\to\infty} E|X_{j_n}| 1_{(|X_{j_n}| \geq n)}$$

$$\leq \limsup E(|X_{j_n}| - \lambda) 1_{(|X_{j_n}| \geq n)} + \limsup \lambda P(|X_{j_n}| \geq n)$$

$$\leq \limsup C.EV_{j_n} 1_{A_{j_n,n}}$$ by (i); the second term vanishes because

$\limsup E|X_n| < \infty$ by (i)

$= 0$ by (ii) and (iii).

METRISATIONS OF CONVERGENCE

The space $\mathcal{P}(S)$ of probability measures is itself a Polish space (B.Appendix III), and so we may consider random elements ξ of $\mathcal{P}(S)$. For such random elements, one must think carefully about modes of convergence: thus

$\xi_n \to \xi$ a.s. means $\{\omega : \xi_n(\omega) \to \xi(\omega)$ in $\mathcal{P}(S)\}$ has probability one;

$\xi_n \Rightarrow \xi$ means $\mathcal{L}(\xi_n) \to \mathcal{L}(\xi)$ in $\mathcal{P}(\mathcal{P}(S))$.

It is occasionally useful to have explicit metrics for convergence in probability and weak convergence of S-valued random elements. Many metrics are known: the following suffice for our purposes.

$$(1.10) \quad \begin{aligned} d_0(X,Y) &= \inf\{\varepsilon > 0 : P(d(X,Y) > \varepsilon) \le \varepsilon\}. \\ \rho(\mu,\nu) &= \inf\{d_0(X,Y) : \mathcal{L}(X) = \mu, \mathcal{L}(Y) = \nu\}. \end{aligned}$$

(Recall that d is the metric on S). The proofs of the following facts are left as an exercise for the reader interested in abstract theory.

(i) The infima are attained.

(1.11) (ii) d_0 is a complete metrisation of convergence in probability.

(iii) ρ is a complete separable metrisation of $\mathcal{P}(S)$.

2 THE SPACE D = D($[0,\infty)$,S).

Adopting a name suggested by Williams (1979), we shall call a function <u>Skorohod</u> if it is right continuous with limits from the left. [B] discusses the space $D([0,1],R)$ of real-valued Skorohod functions defined on $[0,1]$. For many stochastic processes it seems more natural to use $[0,\infty)$ as the time interval; and later we shall need to consider functions taking values in a Polish space S; we shall therefore consider the space $D = D([0,\infty),S)$ of Skorohod functions $f:[0,\infty) \to S$. Fortunately the results developed in [B] for $D([0,1],R)$ pass over with only the obvious modifications to our setting. This section is mostly devoted to the statements of these modified results.

Here is some useful notation from [DM].

$u \downarrow t$ means $u \to t$, $u \geq t$;

$u \uparrow\uparrow t$ means $u \to t$, $u < t$;

and analogously for $u \uparrow t$ and $u \downarrow\downarrow t$. In this notation, a function f is Skorohod iff

$f(t) = \lim(\ f(u): u \downarrow t)\ , \quad 0 \leq t < \infty$

$f(t-) = \lim(\ f(u): u \uparrow\uparrow t)$ exists, $\quad 0 < t < \infty$.

The set of Skorohod functions has a natural σ-field, i.e. the σ-field generated by the evaluation maps

(2.1) $\qquad f \to f(t_0), \qquad 0 \leq t_0 < \infty$.

To define a topology, let Λ be the set of continuous strictly increasing functions $\lambda:[0,\infty) \to [0,\infty)$ such that $\lambda(0) = 0$ and $\lim_{t\to\infty} \lambda(t) = \infty$. Call a sequence (λ_n) in Λ a <u>scaling sequence</u> if

$$\sup_{t \leq L} |\lambda_n(t) - t| \to 0, \quad \text{each } L < \infty .$$

Here is the topology we use on D.

THEOREM ([B,14]; Lindvaal (1973)). *There exists a Polish topology on the set of Skorohod functions such that* $f_n \to f$ *if and only if*

$$(2.2) \qquad \sup_{t \le L} d(f_n(\lambda_n(t)), f(t)) \to 0, \quad \textit{each } L < \infty,$$

for some scaling sequence (λ_n). *The Borel* σ*-field of this topology coincides with the natural* σ*-field generated by the evaluation maps* (2.1).

When (2.2) holds, call (λ_n) a *scaling sequence for* (f_n).

The relation between $D[0,1]$ and $D[0,\infty)$ is the same as the relation between $C[0,1]$ and the space $C[0,\infty)$ equipped with the topology of uniform convergence on bounded intervals. For brevity we shall often omit the quantifier "for each $L < \infty$" from assertions like (2.2): just think of L as an arbitrarily large endpoint.

The first picture below shows two functions which are close; in the other pictures the functions are not close.

Call a stochastic process X *Skorohod* if its sample paths are Skorohod. Throughout this book we assume all processes to be Skorohod, except where otherwise stated. Thus a S-valued process may be regarded as a random element of D(S).

Call t a *continuity point* of f (resp. X) if $f(t) = f(t-)$ (resp. $X_t = X_{t-}$ a.s.); a *discontinuity point* if not. By convention 0 is a continuity point.

In contrast to the situation with continuous function space, the evaluation maps on D are not continuous. Writing out proofs of the next two lemmas (draw pictures first) will help the reader understand the topology on D.

We consider the map

(2.3) $(f,t) \to f(t)$

from $D(S) \times [0,\infty)$ to S.

(2.4) LEMMA. The map (2.3) is continuous at (f_0,t_0) if t_0 is a continuity point of f_0.

By the Skorohod Representation Theorem any continuity result about D immediately gives a continuity result for processes. In this case, Lemma 2.4 gives

(2.5) COROLLARY. If $(X^n,T_n) \Rightarrow (X,T)$ and if $T(\omega)$ is a.s. a continuity point of the sample path $X(\omega)$ then $X^n_{T_n} \Rightarrow X_T$.

Here is what happens at discontinuity points.

(2.6) LEMMA. Suppose $(f_n,t_n) \to (f_0,t_0)$ where t_0 is a discontinuity point of f_0.

(i) $\{f_n(t_n)\}$ is precompact, and $f_0(t_0)$ and $f_0(t_0-)$ are the only possible limit points.

(ii) There exist $u_n \to t_0$ such that $f_n(u_n) \to f_0(t_0)$ and $f_n(u_n-) \to f_0(t_0-)$.

(iii) If $t_n-u_n \downarrow 0$ then $f_n(t_n) \to f_0(t_0)$,

if $t_n-u_n \uparrow\uparrow 0$ then $f_n(t_n) \to f_0(t_0-)$.

As a convenient shorthand, we sometimes write $t_n \to t\pm$ to mean " either $t_n \downarrow t$ or $t_n \uparrow\uparrow t$". Thus we could express (iii) more succinctly as

if $t_n-u_n \to 0\pm$ then $f_n(t_n) \to f_0(t_0\pm)$.

Here is a useful reformulation of the definition (2.2) of convergence in D.

(2.7) LEMMA. Let (λ_n) be a scaling sequence. Let (f_n), f be elements of D. The following are equivalent.

(i) $f_n \to f$ and (λ_n) is a scaling sequence for (f_n).

(ii) $t_n \to t\pm$ implies $f_n(\lambda_n(t_n)) \to f(t\pm)$.

This shows that the topology on D depends only on the topology of S, not on the particular metrisation used in (2.2). Here is a very simple illustration of the use of Lemma 2.7.

(2.8) LEMMA. *Let* $\Phi : S \to S'$ *be continuous. Then for* $f \in D(S)$ *the map* $t \to \Phi(f(t))$ *is Skorohod and so defines an element* $\Phi \circ f$ of $D(S')$. *The map* $f \to \Phi \circ f$ *from* $D(S)$ *to* $D(S')$ *is continuous.*

Proof. Given $f_n \to f$, choose a scaling sequence for (f_n) and use (2.7)(ii) to see that it is also a scaling sequence for $(\Phi \circ f_n)$.

The next three lemmas state the fundamental convergence and compactness properties of D, proved in [B] and modified here for our setting.

(2.9) LEMMA. *Let* (f_n), f *be elements of* $D(S)$, *and let* (X^n), X *be* S-*valued processes. Let* Δ *be some dense subset of* $[0,\infty)$.

(i) f (*resp.* X) *has only countably many discontinuity points.*

(ii) *If* $f_1(t) = f_2(t)$ *for all* $t \in \Delta$ *then* $f_1 = f_2$. *If* $\mathcal{L}(X^1_{t_1},\ldots,X^1_{t_k}) = \mathcal{L}(X^2_{t_1},\ldots,X^2_{t_k})$ *for all* $(t_1,\ldots,t_k) \in \Delta$ *then* $\mathcal{L}(X^1) = \mathcal{L}(X^2)$.

(iii) *If* (f_n) *is precompact and if* $f_n(t) \to f(t)$ *for all* $t \in \Delta$ *then* $f_n \to f$. *If* (X^n) *is tight and if* $(X^n_{t_1},\ldots,X^n_{t_k}) \Rightarrow (X_{t_1},\ldots,X_{t_k})$ *for all* $(t_1,\ldots,t_k) \in \Delta$ *then* $X^n \Rightarrow X$.

(2.10) *Definition*. For $f \in D(S)$, $\delta > 0$, $L < \infty$ define

$$w(f,\delta,L) = \sup\{d(f(t_1),f(t_2)) : t_1 < t_2 \le L,\ t_2 \le t_1+\delta\}.$$

$$w'(f,\delta,L) = \inf\{\max_i \sup_{t_i<t<t_{i+1}} d(f(t),f(t_i)) : 0 = t_0 < t_1 < \ldots < t_I = L,\ t_{i+1} \le t_i+\delta\}$$

$$w''(f,\delta,L) = \sup\{d(f(t_1),f(t_2)) \wedge d(f(t_2),f(t_3)) : t_1 < t_2 < t_3 \le L,\ t_3 \le t_1+\delta\}$$

Here $a \wedge b$ denotes $\min(a,b)$.

(2.11) LEMMA. $F \subset D(S)$ is precompact if and only if

(C1) $\{f(t): t \leq L, f \in F\}$ is precompact in S;

(C2) $$\lim_{\delta \downarrow 0} \sup_{f \in F} w''(f,\delta,L) = 0;$$

(C3) $$\lim_{\delta \downarrow 0} \sup_{f \in F} \sup_{t \leq \delta} d(f(t),f(0)) = 0.$$

As mentioned earlier, conditions like these are implicitly required to hold for each $L < \infty$.

(2.12) LEMMA. (X^n) is tight on $D(S)$ if and only if for each $\varepsilon > 0$ and $L < \infty$,

(a) there exists compact $K \subset S$ such that $P(X_t \in K$ for all $t \leq L) \geq 1-\varepsilon$;

(b") $$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} P(w''(X^n,\delta,L) > \varepsilon) = 0;$$

(c) $$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} P(\sup_t d(X_t^n, X_0^n) > \varepsilon) = 0.$$

In this result, conditions (b") and (c) may be replaced by

(b') $$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} P(w'(X^n,\delta,L) > \varepsilon) = 0.$$

If (X^n) satisfies

(b) $$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} P(w(X^n,\delta,L) > \varepsilon) = 0$$

(d) $\{X_0^n\}$ is precompact in S

then (X^n) is tight and every limit process has continuous sample paths.

As the reader probably knows, these compactness conditions are hard to verify directly. In Section 4 we shall give a more usable sufficient condition for tightness. In the Polish space setting, these conditions are unsatisfactory because they involve a metric. Often, for example if S is the space of probability measures on $D(S')$, we do not have any very tractable metric on S. It is therefore useful to have conditions for compactness which do not involve the metric explicitly. Recall conditions (C1)-(C3) of Lemma 2.11. Here are two more conditions on a subset F of D(S).

(C4) If $(f_j) \in F$ and $u_j^1 < u_j^2 < u_j^3$ are such that $u_j^i \to u$ and $f_j(u_j^i) \to s_i$ for each i, then either $s_1 = s_2$ or $s_2 = s_3$.

(C5) If $(f_j) \in F$ and $\delta_j \downarrow 0$ are such that $f_j(0) \to s_0$ and $f_j(\delta_j) \to s_1$, then $s_1 = s_0$.

(2.13) LEMMA. Suppose $F \subset D(S)$ satisfies (C1). Then (C2) is equivalent to (C4), and (C3) is equivalent to (C5). Hence F is precompact if and only if it satisfies (C1), (C4) and (C5).

Proof. (C2) implies (C4) because

$$d(f_j(u_j^1),f_j(u_j^2)) \wedge d(f_j(u_j^2),f_j(u_j^3)) \le w''(f_j,u_j^3-u_j^1,L)$$

for $L = \sup_{i,j} u_j^i$. Conversely, suppose (C2) fails. Then there exist $L<\infty$, $\varepsilon > 0$, $(f_j) \in F$ and $u_j^1 < u_j^2 < u_j^3 \le L$ such that $u_j^3 - u_j^1 \to 0$ but $d(f_j(u_j^1),f_j(u_j^2)) \wedge d(f_j(u_j^2),f_j(u_j^3)) \ge \varepsilon$. By (C1) we can pass to a subsequence in which $u_j^i \to u$, say, and $f_j(u_j^i) \to s_i$, say. Then $d(s_1,s_2) \wedge d(s_2,s_3) \ge \varepsilon$, so (C4) fails. The argument that (C3) is equivalent to (C5) is similar but easier.

3 MORE ABOUT D

In this section we discuss some technical properties of the space D which are not covered in B .

INCREASING PROCESSES

For increasing real-valued processes, weak convergence can be deduced from convergence of finite-dimensional distributions with only mild extra conditions, avoiding the usual need to establish tightness.

Let $D_+(R)$ be the set of "counting fuctions", i.e. functions of the form

$$f(t) = \sum_{i\geq 1} 1_{(t \geq t_i)}$$

where $0 < t_1 < t_2 < \ldots\ldots$. Let Δ be some countable dense set in $[0,\infty)$ containing 0.

(3.1) LEMMA. <u>Suppose</u> $X, X^1, X^2, \ldots.$ <u>are real-valued processes such that</u> $(X^n_{t_1}, X^n_{t_2}, \ldots, X^n_{t_k}) \Rightarrow (X_{t_1}, \ldots\ldots, X_{t_k})$ <u>for all</u> $(t_1, \ldots, t_k) \in \Delta$.

<u>Suppose either</u>

(i) <u>each</u> X^n <u>is increasing, and</u> X <u>is continuous; or</u>

(ii) <u>each</u> X^n <u>and</u> X <u>has sample paths in</u> $D_+(R)$.

<u>Then</u> $X^n \Rightarrow X$.

<u>Proof</u>. Consider first the deterministic case. Suppose $f_n(t) \to f(t)$, $t \in \Delta$, and suppose either

(i) each f_n is increasing, and f is continuous; or

(ii) each f_n and f are in $D_+(R)$.

We assert that $f_n \to f$ in either case. In case (i) elementary analysis gives uniform convergence on bounded intervals. In case (ii) consider a scaling sequence (λ_n) such that $\lambda_n(t^n_i) = t_i$.

To prove the lemma, let $\Delta = \{t_1, t_2, \ldots\ldots\}$. Then $(X^n_{t_1}, X^n_{t_2}, \ldots\ldots\ldots) \Rightarrow (X_{t_1}, X_{t_2}, \ldots\ldots)$ on R^∞. By the Skorohod Representation Theorem we may suppose $(X^n_{t_1}, X^n_{t_2}, \ldots\ldots) \to (X_{t_1}, X_{t_2}, \ldots\ldots)$ a.s. on R^∞. Apply the deterministic result.

JOINT CONVERGENCE

If U, V are random elements of spaces S_1, S_2 respectively, then the pair (U,V) defines a random element of the product space $S_1 x S_2$. Weak convergence on this product space

$(U_n, V_n) \Rightarrow (U,V)$

is the usual concept of joint convergence. So if X_t (resp. Y_t) is a S_1 (resp. S_2)-valued process, then X (resp. Y) is a random element of $D(S_1)$ (resp. $D(S_2)$), so the pair (X,Y) is a random element of $D(S_1) x D(S_2)$, and so the usual concept of joint convergence is

(3.2) $(X^n, Y^n) \Rightarrow (X,Y)$ on $D(S_1) \times D(S_2)$.

In the real-valued case we would naturally like to deduce, for example,

(3.3) $X^n + Y^n \Rightarrow X + Y$.

But (3.2) does not imply (3.3). For consider $f_n(t) = 1_{(t \geq 1-1/n)}$, $g_n(t) = 1_{(t \geq 1+1/n)}$, $f(t) = 1_{(t \geq 1)}$. Here $(f_n, g_n) \to (f,f)$ in $D(S_1) \times D(S_2)$ but $f_n + g_n \not\to f + f$.

Fortunately there is another concept of joint convergence which avoids this type of problem. Given a pair (X,Y) of processes, the sample paths $t \to (X_t(\omega), Y_t(\omega))$ are $S_1 x S_2$-valued Skorohod functions, i.e. elements of the space $D(S_1 x S_2)$. So we can regard (X,Y) are a random element of $D(S_1 x S_2)$ and consider weak convergence on this space:

(3.4) $(X^n, Y^n) \Rightarrow (X,Y)$ on $D(S_1 x S_2)$.

In the real-valued case, (3.4) is sufficient to imply (3.3), by Lemma 2.8. The next lemma describes the relationship between $D(S_1)xD(S_2)$ and $D(S_1xS_2)$.

(3.5) LEMMA. $(f_n,g_n) \to (f,g)$ in $D(S_1xS_2)$ <u>if and only if</u>

(i) $f_n \to f$ in $D(S_1)$, $g_n \to g$ in $D(S_2)$; that is, $(f_n,g_n) \to (f,g)$ in $D(S_1)xD(S_2)$.

(ii) whenever t is a discontinuity point of f, and $t_n \to t$ is such that

$f_n(t_n) \to f(t)$ and $f_n(t_n-) \to f(t-)$, then $g_n(t_n) \to g(t)$ and $g_n(t_n-) \to g(t-)$.

Note that (ii) holds if either f or g is continuous, or if their discontinuity points are disjoint (by Lemma 2.4).

<u>Proof</u>. Suppose $(f_n,g_n) \to (f,g)$ in $D(S_1xS_2)$ with scaling sequence (λ_n). Plainly (λ_n) is a scaling sequence for (f_n) and for (g_n), so (i) holds. For t, (t_n) as in (ii) we must have $\lambda_n(t) = t_n$ for all sufficiently large n, so (ii) holds.

Conversely, suppose (i) and (ii) hold. It suffices to show the sequence (f_n,g_n) is precompact in $D(S_1xS_2)$, since (i) identifies the limit. We shall verify the conditions of Lemma 2.13. Conditions (C1) and (C5) follow from the same conditions on (f_n) and (g_n). To verify (C4) we need the following observation.

SUBLEMMA. <u>Suppose</u> $u \geq 0$. <u>Then there exist</u> $t_n \to u$ <u>such that</u>

(a) $f_n(t_n) \to f(u)$ and $f_n(t_n-) \to f(u-)$;

(b) $g_n(t_n) \to g(u)$ and $g_n(t_n-) \to g(u-)$.

<u>Proof</u>. Suppose u is a continuity point of f. By Lemma 2.4 or 2.6, there exist (t_n) satisfying (b), and then by Lemma 2.4 they satisfy (a). Alternatively, suppose u is a discontinuity point of f. By Lemma 2.6 there exist (t_n) satisfying (a), and then by hypothesis (ii) they satisfy (b).

To verify (C4) consider $u_j \to u$, and let (t_j) be as in the sublemma.

Lemma 2.6 says:

if $u_j \geq t_j$ for all j then $(f_j(u_j), g_j(u_j)) \to (f(u), g(u))$;

if $u_j < t_j$ for all j then $(f_j(u_j), g_j(u_j)) \to (f(u-), g(u-))$.

Given (f_j) and (u_j^i) as in (C4), pass to a subsequence in which (u_j^2) satisfies one of the requirements above: in the first case $s_2 = s_3 = (f(u), g(u))$, while in the second case $s_1 = s_2 = (f(u-), g(u-))$.

(3.6) COROLLARY. Suppose $(X^n, Y^n) \Rightarrow (X, Y)$ on $D(S_1) \times D(S_2)$. If either X or Y is continuous then $(X^n, Y^n) \Rightarrow (X, Y)$ on $D(S_1 \times S_2)$.

Proof. Use the Skorohod Representation Theorem and the remark after Lemma 3.5.

APPROXIMATING MEASURABLE FUNCTIONALS

Lemma 1. says that a measurable map $\phi: D(S) \to R$ can be approximated by continuous maps ϕ_n. Sometimes we need to know that when ϕ depends only on $0 \leq t \leq v$ then we can choose ϕ_n depending only on $0 \leq t \leq v$. Lemma 3.8 makes this idea precise.

(3.7) Definition. Let $\mathcal{F}_t^D$ be the σ-field on D(S) generated by the evaluation maps $f \to f(u)$, $u \leq t$.

(3.8) LEMMA. Let $v > 0$. Let $\phi: D(S) \to R$ be $\mathcal{F}_v^D$-measurable, $|\phi| \leq 1$. Let X be a process such that $X_v = X_{v-}$ a.s. Then there exist $\mathcal{F}_v^D$-measurable continuous maps ϕ_n, $|\phi_n| \leq 1$, such that $E|\phi(X) - \phi_n(X)| \to 0$.

Proof. Define $a_v: D(S) \to D(S)$ by

$$a_v \circ f(t) = f(t \wedge v).$$

Then a_v is $\mathcal{F}_v^D / \mathcal{F}_\infty^D$-measurable. And a_v is continuous at f satisfying

(3.9) v is a continuity point of f.

Since ϕ is $\mathcal{F}_v^D$-measurable, $\phi = \phi \circ a_v$. Applying Lemma 1. to $a_v(X)$, there exist continuous ϕ_n, $|\phi_n| \leq 1$, such that

$$E\left|\phi(X) - \phi_n \circ a_v(X)\right| \to 0 \quad \text{as } n \to \infty.$$

It now suffices to prove the following assertion, for fixed n.

<u>Assertion</u>. There exist continuous $\mathcal{F}_v^D$-measurable maps ψ_m, $|\psi_m| \leq 1$, such that $\psi_m(f) \to \phi_n \circ a_v(f)$ as $m \to \infty$ for all f satisfying (3.9).

To prove this, for $0 < u \leq 1$ define $b_u: D(S) \to D(S)$ by

$$b_u \circ f(t) = f(ut).$$

Then b_u is continuous. Also b_u is $\mathcal{F}_v^D/\mathcal{F}_v^D$-measurable, and so $a_v \circ b_u$ is $\mathcal{F}_v^D/\mathcal{F}_\infty^D$-measurable. So the definition

$$\psi_m(f) = m\int_{1-1/m}^{1} (\phi_n \circ a_v \circ b_u \circ f)\, du$$

produces a $\mathcal{F}_v^D$-measurable map $\psi_m: D(S) \to R$. If $f_n \to f$ in D(S) then $b_u \circ f_n \to b_u \circ f$ for each u, so $a_v \circ b_u \circ f \to a_v \circ b_u \circ f$ for all u except perhaps the countable set where $b_u \circ f$ fails (3.9): now since ϕ_n is continuous it follows that ψ_m is continuous. Finally, suppose f satisfies (3.9). As $u \uparrow 1$ we have $b_u \circ f \to f$, so $a_v \circ b_u \circ f \to a_v \circ f$: now the continuity of ϕ_n gives the convergence part of the assertion.

CONTINUITY OF OPERATIONS ON D

The compactness conditions (C1)-(C5) of section 2 may be used to prove continuity of operations involving elements of D. We illustrate this technique by proving two technical lemmas needed later.

(3.10) LEMMA. Suppose $f_n \to f_\infty$ in $D(S)$, $t_n \to t_\infty$ in $[0,\infty)$, and $f_n(t_n) \to f_\infty(t_\infty)$. Define $h_n(u) = f_n(u + t_n)$. Then $h_n \to h_\infty$ in $D(S)$.

Proof. If u is a continuity point of h_∞ then $h_n(u) \to h_\infty(u)$ by Lemma 2.4. So it is sufficient to prove (h_n) is precompact. Conditions (C1) and (C4) follow from the same conditions for (f_n). To verify (C5), suppose $\delta_j \downarrow 0$, $h_j(\delta_j) \to s_1$, $h_j(0) \to s_0$. That is, $f_j(t_j+\delta_j) \to s_1$, $f_j(t_j) \to s_0$. Now $s_0 = f_\infty(t_\infty)$ by hypothesis, and $s_1 = f_\infty(t_\infty)$ by Lemma 2.6(iii).

(3.11) LEMMA. Suppose $f_n \to f_\infty$ and $g_n \to g_\infty$ in $D(S)$, and $t_n \to t_\infty$ in $[0,\infty)$. Define

$$h_n(u) = f_n(u), \quad u < t_n,$$
$$= g_n(u-t_n), \quad u \geq t_n.$$

If either

(a) $f_\infty(t_\infty) = g_\infty(0)$; or

(b) $t_\infty > 0$ and $f_n(t_n-) \to f_\infty(t_\infty-)$

then $h_n \to h_\infty$ in $D(S)$.

Proof. If u is a continuity point of h_∞ and $u \neq t_\infty$ then $h_n(u) \to h_\infty(u)$ by Lemma 2.4. So it suffices to prove (h_n) is precompact. Condition (C1) holds for (h_n) because it holds for (f_n) and for (g_n).

If $t_\infty > 0$ then (C5) holds for (h_n) because it holds for (f_n). If $t_\infty = 0$ then

$$0 \le \delta_j < t_j \quad \text{implies } h_j(\delta_j) = f_j(\delta_j) \to f_\infty(0)$$
$$t_j \le \delta_j \downarrow 0 \quad \text{implies } h_j(\delta_j) = g_j(\delta_j - t_j) \to g_\infty(0)$$

and now hypothesis (a) establishes (C5).

It remains to prove (C4). Suppose (h_j) is a subsequence of (h_n), $\lim_j u_j^i = u$, $\lim_j h_j(u_j^i) = s_i$. We must prove $s_1 = s_2$ or $s_2 = s_3$. If $u \neq t_\infty$ this follows from (C4) applied to (f_n) and (g_n). So suppose $u = t_\infty$. By passing to a further subsequence, we reduce to three cases.

(i) $t_j < u_j^2$

(ii) $u_j^2 \le t_j < u_j^3$

(iii) $u_j^3 \le t_j$.

In case (i), $s_2 = s_3 = g_\infty(0)$.

In case (iii), the condition holds because it holds for (f_n).

In case (ii), we must treat hypotheses (a) and (b) separately. Under (b), $s_1 = s_2 = f_\infty(t_\infty-)$ by Lemma 2.6(iii). Under (a), $s_3 = g_\infty(0) = f_\infty(t\ )$. By Lemma 2.6(i),(iii), the only way to have $s_1 = s_2$ is to have $s_1 = f_\infty(t_\infty-)$, $s_2 = f_\infty(t_\infty)$: but then $s_2 = s_3$.

4 STOPPING TIMES AND TIGHTNESS

Because the basic criteria for tightness are in practice hard to check, much effort has gone into finding more tractable sufficient conditions. There seem now to be four classes of conditions known. Two of these (based on maximal inequalities or moment inequalities) are discussed in (B). Strook and Varadhan (1978) give conditions specially designed for diffusions. We shall present a condition specially suitable for martingales.

Recall the classical definition of a stopping time T on a process X:

(4.1) $\{T \leq t\} \in \sigma(X_r : r \leq t) \qquad t \geq 0.$

Write $\mathcal{T}$ for the set of stopping times (on a specified process), and $\mathcal{T}_L$ for the subset bounded by L.

We shall be concerned with the following condition on a sequence of processes (X^n):

(4.2) $X^n_{T_n+\delta_n} - X^n_{T_n} \underset{p}{\to} 0$ for each sequence $T_n \in \mathcal{T}^n_L$ and each sequence of constants $\delta_n \downarrow 0$.

Though the form above is more useful for applications, note that (4.2) can be rewritten as

(4.3) $$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \sup_{\substack{T \in \mathcal{T}^n_L \\ 0 \leq u \leq \delta}} P(|X^n_{T+u} - X^n_T| > \varepsilon) = 0 \ ; \quad \text{each} \quad \varepsilon > 0.$$

Now if T were not required to be a stopping time, this condition would just be condition (.) for the processes to be tight and subsequential limits to be continuous. The significance of this observation will become clear later (

Condition (4.2) is not quite sufficient for tightness: consider

$X_t^n = n.1_{(t \geq U)}$, where $P(U > u) = e^{-u}$. But with mild supplementary conditions it becomes sufficient.

(4.4) THEOREM <u>Suppose</u> (X^n) <u>satisfies</u> (4.2), <u>and suppose</u> (X_0^n) <u>is tight on</u> R. <u>Suppose either</u>

(i) (4.2) <u>holds with</u> L^1 <u>convergence</u>;

<u>or</u> (ii) (X_t^n) <u>is tight on</u> R, <u>for each</u> t.

<u>Then</u> (X^n) <u>is tight on</u> D(R).

The result is stated for real-valued processes, but the proof (under (ii)) extends to the general Polish space S (to make sense of (i) we need S to be a Banach space).

Since Theorem 4.4 plays a major role later we feel obliged to present the proof, although it is rather technical. The reader may well omit the rest of this section.

A few remarks about stopping times are needed. That $T = \inf\{t : X_t \geq a\}$ is a stopping time is obvious - until you try to prove it! (see Williams (1979)). On the other hand the time

(4.5) $T = \inf\{t : X_t > a\}$

is easily seen to be at least a <u>weak</u> stopping time: that is,

(4.6) $\{T < t\} \in \sigma\{X_r : r \leq t\}$; $t \geq 0$.

Now a weak stopping time T is a limit of ordinary stopping times $T+\delta$, and so condition (4.2) extends automatically to weak stopping times. In the proof below we shall use weak stopping times like (4.5). Chapter 2 has a more definitive discussion of stopping times.

<u>Proof of Theorem 4.4</u>. We first show that (i) implies (ii), by showing that (i) implies $\limsup E|X_t^n - X_0^n| < \infty$ for each t. For if the latter condition

fails, then for any $0 = t_0 < \dots < t_k = t$ we have $\limsup_{n\to\infty} E|X^n_{t_i} - X^n_{t_{i-1}}| = \infty$ for some i, and follows that (i) fails for certain (non-random) stopping times.

Our strategy now is to verify, using (ii), the basic tightness conditions of (B.15.2). Fix L and $\varepsilon > 0$. By hypothesis (4.2) there exist $\delta > 0$ and n_0 such that

(4.7) $P(|X^n_{T+u} - X^n_T| \geq \varepsilon) < \varepsilon \quad ; T \in \mathcal{T}^n_L,\ u \leq 2\delta,\ n \geq n_0.$

Fix an integer $q > 2L/\delta$. Then similarly there exist $\delta_2 > 0$, $n_1 \geq n_0$ such that

(4.8) $P(|X^n_{T+u} - X^n_T| \geq \varepsilon) < \varepsilon/q \ ; \ T \in \mathcal{T}^n_L,\ u \leq 2\delta_2,\ n \geq n_1.$

Next let B_c be the set of t such that

$$\sup_n P(|X^n_t| > c) \leq \varepsilon\delta/L .$$

By hypothesis (ii), $\bigcup B_c = [0,\infty)$. So we can choose c such that $[0,2L]$ has Lebesgue measure less than $\varepsilon\delta/L$. We shall prove later that

(4.9) $P(\sup_{t \leq L} |X^n_T| > c+\varepsilon) \leq 3\varepsilon, \quad n \geq n_1.$

Consider now the process X^n for some $n \geq n_1$, and drop superscripts "n". Define stopping times T_i by: $T_0 = 0$,

$$T_{i+1} = \inf\{t > T_i : |X_t - X_{T_i}| > 2\varepsilon\} \wedge L \ ; \ i = 1,\dots,q.$$

We shall prove later that

(4.10) $P(T_q < L) \leq 16\varepsilon$

(4.11) $P(T_i < (T_{i-1} + \delta_2) \wedge L$ for some i$) \leq 8\varepsilon$.

These inequalities are sufficient to establish tightness on $D[0,L]$, from which the Theorem follows. For let w' be the modulus of (B.equation 14.6), modified for $D[0,L]$ by taking the infimum in (B.equation 14.6) over $0 = t_0 < \dots < t_r = L$ to give a modulus $w'(f,\delta,L)$. From the definition of w' and (4.10), (4.11),

$$P(w'(X,\delta_2,L) > 4\varepsilon) \leq 24\varepsilon.$$

This and (4.9) are the hypotheses of (B.15.2), which establishes tightness.

Proof of (4.10) and (4.11). Let θ be independent of X, distributed uniformly on $[0,2\delta]$. Consider $f \in D(R)$ and $0 \leq t_1 \leq t_2 \leq L$. If

$$t_2 - t_1 < \delta,$$
$$P(|f(t_j + \theta) - f(t_j)| < \varepsilon) > 3/4, \quad j = 1,2,$$

then there must exist some constant θ_0 in $[t_1+\delta, t_1+2\delta]$ such that

$$|f(\theta_0) - f(t_j)| < \varepsilon \quad , \quad j = 1,2,$$

and so $|f(t_2) - f(t_1)| < 2\varepsilon$. In other words, the set

$$\{(f,t_1,t_2) : |f(t_2) - f(t_1)| \geq 2\varepsilon \text{ and } t_2 < t_1 + \delta\}$$

is contained in the set

$$\{(f,t_1,t_2) : P(|f(t_j+\theta) - f(t_j)| \geq \varepsilon) \geq 1/4 \text{ for } j = 1 \text{ or } 2\}.$$

Applying this to the process X, we have for each $i = 1,\dots,q$

(4.12) $\quad P(|X(T_i) - X(T_{i-1})| \geq 2\varepsilon \text{ and } T_i < T_{i-1} + \delta) \leq P(A_i) + P(A_{i-1}),$

where $A_i = \{P(|X(T_i+\theta) - X(T_i)| \geq \varepsilon \,|X) \geq 1/4\}$.

But $P(A_i) \leq 4.P(|X(T_i+\theta) - X(T_i)| \geq \varepsilon)$, and by averaging over u in (4.7) we find

(4.13) $\quad P(|X(T+\theta) - X(T)| \geq \varepsilon) < \varepsilon \quad ; \quad T \in \mathcal{T}_L.$

And by definition of T_i we have $|X(T_i) - X(T_{i-1})| \geq 2\varepsilon$ on $\{T_i < L\}$.

So (4.12) implies

(4.13) $\quad P(T_i < L, T_i < T_{i-1} + \delta) \leq 8\varepsilon; \quad i = 1,\dots,q.$

The same argument using (4.8) in place of (4.7) shows

$$P(T_i < L,\ T_i < T_{i-1} + \delta_2) \leq 8\varepsilon/q \ ; \quad i = 1,\dots,q,$$

which gives (4.11). To get (4.10) we compute

$$\begin{aligned} L &> E(T_q \mid T_q < L) \\ &= \sum E(T_i - T_{i-1} \mid T_q < L) \\ &\geq \delta \sum P(T_i - T_{i-1} \geq \delta \mid T_q < L) \\ &= \delta \sum (1 - P(T_i - T_{i-1} < \delta \mid T_q < L)) \\ &\geq \delta q\, (1 - 8\varepsilon/P(T_q < L)) \quad \text{, using (4.13).} \end{aligned}$$

Rearranging, this gives (4.10) since we chose $q > 2L/\delta$.

<u>Proof of</u> (4.9). Take U uniform on $[0,2L]$, independent of X. Then for f in D(R) and $0 \leq t \leq L$ we have, by considering densities,

(4.14) $\quad P(|f(t+\theta)| \geq c) \leq L/\delta .\ P(|f(U)| \geq c).$

So for the stopping time $T = \inf\{t : |X(t)| > c+\varepsilon\} \wedge L$,

$$\begin{aligned} P(|X(T)| \geq c+\varepsilon) &\leq \varepsilon + P(|X(T+\theta)| \geq c) \quad \text{by (4.12)} \\ &\leq \varepsilon + L/\delta .\ P(|X(U)| \geq c) \quad \text{by (4.14)} \\ &\leq \varepsilon + L/\delta .\ (\varepsilon\delta/L + P(U \notin B_c)) \leq 3\varepsilon \end{aligned}$$

by definition of B_c.

5 WEAK CONVERGENCE OF MARTINGALES

Here we are concerned with theoretical properties of weak convergence of martingales, not with conditions for convergence to occur.

We first want to show that a weak limit of (sub)martingales is itself a (sub)martingale. The alert reader will instantly see that this isn't quite true. For if $P(V_n = n) = 1/(n+1)$, $P(V_n = -1) = n/(n+1)$, then $X_t^n = V_n \cdot 1_{(t \geq 1)}$ is a sequence of martingales whose limit is not a martingale. Pursuing this line of thought, it is not hard to see that any process is a limit of martingales. The problem, of course, is that the martingale property involves integrability: the natural solution is to impose uniform integrability.

Some care is needed with definitions. We have in mind the classical definition of submartingale:

$$E(S_u | S_r : r \leq t) \geq S_t \quad \text{a.s.,} \quad t < u.$$

But sometimes we will have a bivariate process (X_t, S_t), and then we want the submartingale property for S_t to be:

$$E(S_u | X_r, S_r : r \leq t) \geq S_t \quad \text{a.s.,} \quad t < u.$$

Similarly, we want $\mathcal{T}$ to denote the stopping times for the bivariate process. The result is stated for submartingales – the martingale form is an immediate consequence.

(5.1) PROPOSITION *Suppose* $(X^n, S^n) \Rightarrow (X,S)$ *on* $D(S) \times D(R)$, *and suppose* (S_t^n) *is uniformly integrable for each* t. *If each* S^n *is a submartingale, or alternatively if*

$$C_L^n = \inf \{ ES_U^n - ES_T^n : T, U \in \mathcal{T}_L, T \leq U \} \rightarrow 0,$$

then S_t *is a submartingale.*

Proof. Let $t < u$ and $G \in \sigma(X^n_r, S^n_r : r \le t)$. By considering

$T = t$ on B; $T = u$ outside B;

$U = u$,

we find that $E(S^n_u - S^n_t)1_G \ge C^n_L$ (where $C^n_L = 0$ in the submartingale case). It follows that for any $\sigma(X^n_r, S^n_r)$-measurable random variable γ^n with $0 \le \gamma^n \le 1$,

$$E(S^n_u - S^n_t)\gamma^n \ge C^n_L .$$

Let $t_1 < t_2 \; .... < t < t_k < u$, where (t_i) are continuity points of (X,S). For $g \in C(S^k x R^k)$ with $0 \le g \le 1$ define $\gamma^n = g(X^n_{t_1},..,X^n_{t_k}, S^n_{t_1},....,S^n_{t_k})$. By weak convergence and Corollary 1.4,

$$E(S_u - S_{t_k})\gamma \ge 0.$$

By Lemma 5.2 below (for $R \times (S^k x R^k)$), this inequality extends to all bounded positive $\sigma(X_{t_1},..,X_{t_k}, S_{t_1},...,S_{t_k})$-measurable random variables γ, and thence all $(X_r, S_r : r \le t_k)$-measurable variables, proving

$$E(S_u \mid X_r, S_r : r \le t_k) \ge S_{t_k} \text{ a.s.}$$

Letting t_k decrease to t establishes the submartingale property.

In the course of the proof we appealed to the following variation of a standard fact (Halmos (1950)).

(5.2) LEMMA <u>Let μ be a probability measure on a product</u> $(S \times S', \mathcal{S} \times \mathcal{S}')$ <u>of Polish spaces. Let</u> f <u>be positive bounded and</u> $\mathcal{S}'$<u>-measurable. Then there exist positive</u> $f_m \in C(S')$ <u>such that</u> $\int |f_m - f| \, d\mu \to 0$.

Our second project is to show that weak convergence of martingales to a continuous martingale follows automatically from convergence of finite-dimensional distributions. Again it is plainly necessary to require uniform integrability. This is in fact sufficient, but there seems no easy proof of this (see remark later). Fortunately, in the application we need ( ) the martingales are bounded, and for this case the proof is easy.

(5.3) PROPOSITION <u>Let</u> (M^n) <u>be a sequence of martingales such that</u> $|M^\infty| \leq 1$, <u>and suppose the finite-dimensional distributions converge to those of</u> M^∞. <u>If</u> M^∞ <u>is continuous then</u> $M^n \Rightarrow M^\infty$.

<u>Proof</u>. First verify the inequality

if $0 \leq V \leq 2$ and $0 < \epsilon < 1$ then $P(V \geq EV - \epsilon) \geq \epsilon/2$.

Replacing V by V-1 and conditioning, we get the inequality below.

(5.4) If $|V| \leq 1$, $0 < \epsilon < 1$ and $\mathcal{F}$ is any σ-field, then

$$P(V \geq E(V|\mathcal{F}) - \epsilon \mid \mathcal{F}) \geq \epsilon/2 \quad \text{a.s.}$$

Now fix $0 = t_0 < t_1 \; \dots < t_k = L$. Let M be any martingale with $|M| \leq 1$. We assert

(5.5) $$P(\max_i M_{t_{i+1}} - M_{t_i} > \epsilon) \geq \epsilon/2 \; P(\max_i \sup_{t_i < t \leq t_{i+1}} M_t - M_{t_i} > 2\epsilon).$$

To prove (5.5) consider the stopping times

$$S_0 = \inf\{t : M_t - M_{t_i} > 2\epsilon,\ t_i < t \leq t_{i+1}\}$$

$$S = S_0 \wedge L$$

$$T = t_{i+1} \quad \text{on } \{t_i < S \leq t_{i+1}\}.$$

Applying (5.4) to M_T and $\mathcal{F}_S = \sigma(M_r : r \leq S)$,

$$P(M_T \geq E(M_T|\mathcal{F}_S) - \epsilon \mid \mathcal{F}_S) \geq \epsilon/2 \quad \text{a.s.}$$

Using the optional sampling theorem we find

$$P(M_T \geq M_S - \varepsilon,\ S_0 \leq L) \geq \varepsilon/2\ P(S_0 \leq L)$$

and (5.5) follows.

By convergence of finite-dimensional distributions

$$\limsup_{n\to\infty} P(\max_i \sup_{t_i < t \leq t_{i+1}} M^n_t - M^n_{t_i} > 2\varepsilon) \leq 2/\varepsilon\ P(\max_i M^\infty_{t_{i+1}} - M^\infty_{t_i} \geq \varepsilon).$$

But M^∞ is continuous, so the right side tends to zero as $\max |t_{i+1}\ t_i| \to 0$. This establishes tightness

Remarks. The result is not true for discontinuous M^∞ - consider $M^n_t = V_1 \cdot 1_{(t\geq 1)} + V_2 \cdot 1_{(t \geq 1-1/n)}$, where $P(V_i = \pm \frac{1}{2}) = \frac{1}{2}$.

For the record, we state a sharper result, which implies that Proposition 5.3 remains true when only uniform integrability is imposed: conceivably this may be useful elsewhere.

(5.6) PROPOSITION Let (X^n) be a sequence of processes whose finite-dimensional distributions converge to those of a continuous process X. Suppose (i) $\{X^n_T : T \in \mathcal{J}^n_L,\ n \geq 1\}$ is uniformly integrable;

(ii) $EX^n_{T_n+\delta_n} - EX^n_{T_n} \to 0$ for all $(T_n) \in \mathcal{J}^n_L$, $\delta_n \downarrow 0$.

Then $X^n \Rightarrow X$.

The proof occupies six pages of Aldous (1977). Note that (ii) is essentially a weaker form of (.). Here is an instructive example which shows that some condition like (ii) is necessary. It also shows that Proposition 5.3 is not true for submartingales (where does the proof break down?)

(5.7) EXAMPLE Let $X_t = t \wedge 1$, $X^n_t = (t - 1_{(U \leq t < U+1/n)}) \wedge 1$, where $P(U > u) = e^{-u}$.

fig here

Here X^n and X are submartingales, the finite-dimensional distributions converge, but weak convergence fails.

CHAPTER 2 - THE GENERAL THEORY OF PROCESSES

This is intended as a gentle introduction to the concepts in the general theory needed later. For the full story see Dellacherie and Meyer (1975,1980) - abbreviated (DM).

6 BASICS OF THE GENERAL THEORY

Given a process X, write

$$\mathcal{F}_t^\circ = \sigma(X_r : r \leq t) \qquad t \geq 0.$$

Think of $\mathcal{F}_t^\circ$ as representing what can be learned by observing the process over time $[0,t]$, and call $F^\circ = (\mathcal{F}_t^\circ)_{t\geq 0}$ the <u>natural filtration</u> of X. The classical definitions of "martingale", "Markov process", "stopping time on a process" refer to this natural filtration. In the general theory we adopt a different view. We suppose there is one filtration F - that is, an increasing family $(\mathcal{F}_t)$ of σ-fields - which is fixed once and for all. Think of $\mathcal{F}_t$ as representing what can be learned by observing the entire universe over time $[0,t]$. When we consider a process X we assume (unless otherwise stated) that the value X_t is knowable at time t, i.e. X is <u>adapted</u> in the sense

$$X_t \text{ is } \mathcal{F}_t\text{-measurable}, \quad t \geq 0.$$

We now define concepts like "submartingale", "stopping time" in terms of the fixed filtration F :

$$E(X_t \mid \mathcal{F}_s) \geq X_s \quad \text{a.s.,} \qquad s \leq t$$

$$\{T \leq t\} \in \mathcal{F}_t, \qquad t \geq 0.$$

For technical reasons we assume F satisfies certain regularity conditions (the <u>usual conditions</u>):

$$\mathcal{F}_t = \bigcap_{u>t} \mathcal{F}_u ;$$

the σ-field $\mathcal{B}$ of the underlying probability triple is complete;

$\mathcal{F}_t$ contains the null sets of $\mathcal{B}$.

We shall not discuss the significance of these conditions, save to observe that a weak stopping time

$\{T<t\} \in \mathcal{F}_t$ c.f. (4.6)

becomes a genuine stopping time. Write $\mathcal{T}$ for the class of stopping times; $\mathcal{T}_L$ for the subclass bounded by L.

Given a process X, there is a minimal filtration which satisfies the usual conditions and to which X is adapted - call this the usual filtration. When we construct a process X as a weak limit, we give it the usual filtration. The reader may verify that a Skorohod process which is a submartingale with respect to its natural filtration remains a submartingale with respect to its usual filtration.

A simple yet elegant concept which arises from the general theory is predictability. The times of accidents or radioactive disintegrations are inherently unpredictable, so when we model such events by a Poisson process (say N_t is the number of events occurring before time t), the time of the k'th event

(6.1) $T = \inf\{t : N_t = k\}$

should be "unpredictable". On the other hand a hitting time for Brownian motion

(6.2) $T = \inf\{t : W_t = 1\}$

should be "predictable". The intuitive idea here is: can we tell whether $T = t$ by looking at what happens strictly before t? Unfortunately the obvious formalisation

$\{T = t\} \in \sigma(\mathcal{F}_s : s < t)$

turns out to be unsatisfactory. Here is the right formalisation.

(6.3) <u>Definition</u>. A stopping time T is <u>predictable</u> if there exist stopping times $T_n \leq T$ such that $T_n \uparrow\uparrow T$ on $\{T > 0\}$. Say (T_n) <u>announces</u> T.

How do we tell if this is so? The time in (6.2) is predictable, being announced by $T_n = \inf\{t : W_t = 1 - n^{-1}\}$. What about (6.1)? Let us take this opportunity to define a <u>Poisson process of rate</u> λ as a simple point process N_t such that $N_{s+t} - N_t$ is distributed as Poisson (λs) independent of $\mathcal{F}_t$. Then the strong Markov property implies that $N_{T+s} - N_T$ is also Poisson (λs) for any stopping time T. So if T is predictable then by considering an announcing sequence we find that $N_{(T+s)-} - N_{T-}$ is Poisson (λs). So letting $s \to 0$ we see $N_T = N_{T-}$ a.s.. So the time in (6.1) is not predictable. In fact, a little more is true.

(6.4) <u>Definition</u>. A Skorohod process X is <u>quasi left continuous</u> if $X_T = X_{T-}$ a.s. for each predictable stopping time T.

The argument above shows the Poisson process is quasi left continuous, and suggests that strong Markov processes might be quasi left continuous in general. This is true under wide conditions - see Blumenthal and Getoor (196).

<u>Remark</u>. Of course in conditions like (6.4) we really mean "$X_T = X_{T-}$ a.s. on $\{0 < T < \infty\}$", since X_∞ and X_{0-} are not defined. But we shall reserve such meticulousness for the occasions it is really important.

We omit various elementary facts about stopping times where the classical theory goes over unchanged (e.g. closure under limits and lattice operations; the pre-T σ-field $\mathcal{F}_T$). Liptser and Shiryayev (1977) Section 1.3 has a concise account. Here is a non-elementary fact.

(6.5) LEMMA <u>Suppose</u> X <u>is Skorohod and</u> B <u>measurable. Then</u> $T = \inf\{t : X_t \in B\}$ <u>is a stopping time</u>.

That follows from (DM.III.44). One elementary fact deserves explicit mention.

(6.6) LEMMA <u>Let</u> T <u>be a stopping time. Then there exist discrete-valued stopping times</u> (T_n) <u>such that</u> $T_n \downarrow T$.

<u>Proof</u>. Put $T_n = k/2^{-n}$ on $\{k-1/2^{-n} \leq T < k/2^n\}$.

For Skorohod processes we have $X_{T_n} \to X_T$, and this gives the fundamental technique for extending results from discrete to continuous time (e.g. the proof of the strong Markov property of Brownian motion in Breiman (1968)). As another example, consider the optional sampling theorem for Skorohod submartingales.

(6.7) $E(X_T \mid \mathcal{F}_S) \geq X_S \quad ; \quad 0 \leq S \leq T, \quad S,T \in \mathcal{T}_L.$

To prove this, approximate S,T by discrete S_n, T_n as in Lemma 6.6. The discrete time theorem says $E(X_{T_m} \mid \mathcal{F}_{S_n}) \geq X_{S_n}$ for $m \leq n$. Letting $n \to \infty$, then $m \to \infty$, gives (6.7) - see (DM.VI.10) for details.

Though primarily interested in Skorohod processes, we must occasionally leave this class. Temporarily think of a process X as a map $(t,\omega) \to X_t(\omega)$ defined on $[0,\infty) \times \Omega$.

(6.8) <u>Definition</u>. A process X is <u>optional</u> if it is measurable with respect to the σ-field on $[0,\infty) \times \Omega$ generated by the class of Skorohod processes.

For example, measurable functions $h(X_t)$ or limits $\limsup_{n\to\infty} X^n_t$ of Skorohod processes are optional, but in general are not Skorohod. This "measure-theoretic" view of processes leads to a concept of <u>predictable process</u>. The intuitive idea is: can we tell the value of X_t by looking at what happens strictly before t? Again the obvious formalisation

X_t is $\sigma(\mathcal{F}_s : s < t)$-measurable

turns out to be unsatisfactory. But a left-continuous process should be predictable, hence

(6.9) Definition. A process X is predictable if it is measurable with respect to the σ-field on $[0,\infty) \times \Omega$ generated by the class of left-continuous processes. In particular, a continuous process is predictable.

(6.10) LEMMA (DM.IV.67) A predictable process is optional.

Warning. Since any Skorohod process X is the pointwise limit of the left-continuous processes $X_{(t+\delta)-}$, it is tempting to believe that any Skorohod process is predictable. This is wrong. For by "process" we implicitly mean "adapted process", and in general $X_{(t+\delta)-}$ is not adapted.

So what does it mean for a Skorohod process to be predictable? (DM) is not explicit here. Informally, it means that any stopping time T picking out a discontinuity ($X_T \neq X_{T-}$ a.s.) is predictable and the value X_T is determined by what happens strictly before T (i.e. X_T is $\mathcal{F}_{T-}$-measurable).
See
Again informally, predictability is the opposite of quasi left continuity, where discontinuity times are unpredictable.

We conclude by quoting several technical lemmas.

(6.11) LEMMA (follows from DM.IV.85). If X is a predictable Skorohod process then

$$T_1 = \inf\{t : X_t \geq a\},$$
$$T_2 = \inf\{t : X_t - X_{t-} \geq \varepsilon\}$$

are predictable stopping times.

(6.12) LEMMA (DM.IV.86) Suppose X and Y are optional (resp. predictable) processes and $X_T = Y_T$ a.s. for each (predictable) stopping time T. Then X and Y are indistinguishable, i.e. $P(X_t = Y_t \text{ for all } t) = 1$.

7 MARTINGALE DECOMPOSITIONS

We assume the reader is familiar with basic results about discrete-time (sub)martingales - convergence, maximal inequalities, optional sampling. In the continuous-time setting, the fundamental fact is that submartingales may be assumed to be Skorohod, in the following sense.

PROPOSITION (DM.VI.4) _Let_ X _be a submartingale, and suppose the map_ $t \to EX_t$ _is right-continuous. Then there exists a Skorohod process_ $\hat{X}$ _such that_ $P(X_t = \hat{X}_t) = 1$ _for each_ t.

This indicates why D is the "right" function space for much of probability theory. In this section all processes are assumed to be Skorohod, and adapted to the fixed filtration F. In (6.7) we saw how the optional sampling theorem extended to continuous time, and other results extend just as easily (DM.VI). We record two maximal inequalities.

(7.1) LEMMA _For any martingale_ X,

$$P(\sup_{t \le L} |X_t| > \lambda) \le \lambda^{-1} E|X_L| ;$$
$$E(\sup_{t \le L} X_t^2) \le 4 EX_L^2 .$$

The rest of the section is devoted to the Doob-Meyer decomposition. Consider first an integrable discrete-time process X_n adapted to $(\mathcal{F}_n)$. Define a martingale M by:

$$M_0 = 0, \quad M_n - M_{n-1} = X_n - X_{n-1} - E(X_n - X_{n-1} | \mathcal{F}_{n-1}).$$

Then $X_n = M_n + A_n$, where A is given by:

$$A_0 = X_0, \quad A_n - A_{n-1} = E(X_n - X_{n-1} | \mathcal{F}_{n-1}).$$

Call $X = M + A$ the _Doob-Meyer decomposition_ of X. Note that if X is a submartingale then A is an increasing process.

Here is a mental picture the author finds helpful. Imagine you are engaged in some hazardous enterprise, and X_n represents your accumulated losses at time n. If at time n-1 you wish to insure against the prospective loss $X_n - X_{n-1}$ to be incurred over the interval $[n-1,n]$, then the fair insurance premium is $E(X_n - X_{n-1}|\mathcal{F}_{n-1}) = A_n - A_{n-1}$. So if you always insure, then A_n represents your accumulated insurance premiums at time n, and M_n represents the accumulated losses of your insurance company.

The continuous-time version of this decomposition lies much deeper. We discuss only the submartingale case.

(7.2) Definition. A process X is class (DL) if $\{X_T : T \in \mathcal{J}_L\}$ is uniformly integrable, for each $L < \infty$.

In particular, a positive submartingale is class (DL), because $0 \leq X_T \leq E(X_L|\mathcal{F}_T)$ and the family $E(X_L|\mathcal{F}_T : T \in \mathcal{J}_L$ is uniformly integrable.

(7.3) Definition. A submartingale X is regular if $EX_T = EX_{T-}$ for each bounded predictable T.

So regularity is a property somewhat weaker than quasi left continuity. Here is the continuous-time Doob-Meyer decomposition.

(7.4) THEOREM (DM.VII.8,10) *Let* X *be a class (DL) submartingale. Then there exist unique processes* M, A *such that*

$X = M + A$

M *is a martingale with* $M_0 = 0$

A *is a predictable process.*

The process A *is increasing; it is continuous if and only if* X *is regular*.

Remarks. (a) "Unique" means unique up to indistinguishability.

(b) The normalisation $M_0 = 0$ is convenient for our purposes. Other authors use $A_0 = 0$.

(c) Actually (DM) work with supermartingales. To convert, note that if X is a class (DL) submartingale then $E(X_L|\mathcal{F}_t) - X_{t \wedge L}$ is a class (D) potential in the terminology of (DM).

Call A the <u>compensator</u> of X. Applying the optional sampling theorem to M gives the next result, which is the main way of exploiting the decomposition.

(7.5) COROLLARY <u>Let</u> X, A <u>be as in Theorem 7.4, and let</u> $S \le T$ <u>be bounded stopping times. Then</u>

$$EX_T = EA_T$$
$$E(X_T - X_S|\mathcal{F}_S) = E(A_T - A_S|\mathcal{F}_S).$$

There is a partial converse.

(7.6) LEMMA <u>Let</u> X <u>be a submartingale and</u> A <u>a predictable process. If</u> $X_0 = A_0$ <u>and</u> $EX_T = EA_T$ <u>for each bounded stopping time then</u> A <u>is the compensator of</u> X.

<u>Proof</u>. If $M = X - A$ then $M_0 = 0$ and $EM_T = 0$ for each T. By considering T taking values s, t only ($s < t$), we find $EM_s 1_B = EM_t 1_B$, $B \in \mathcal{F}_s$, and hence $E(M_t|\mathcal{F}_s) = M_s$.

Point processes provide an illustration of the decomposition. Let N be an integrable simple point process. Suppose $EN_t = \int_0^t m(s)\, ds$. Then m(s) is the "intensity" of the process: informally, m(s)ds is the probability of a point in $[s, s+ds]$. Now let A be the compensator of N, and suppose $A_t = \int_0^t a(s,\omega)\, ds$. Then a(s) is the "conditional intensity": informally, $a(s,\omega)ds$ is the conditional probability of a point in $[s,s+ds]$ given $\mathcal{F}_s$. For the Poisson process of rate λ we have $E(N_{t+s} - N_t|\mathcal{F}_t) = E(N_{t+s} - N_t) = \lambda s$, and it follows that the compensator is $A_s = \lambda s$. The converse is true, but more difficult.

(7.7) THEOREM. <u>Let</u> N <u>be an integrable simple point process whose compensator is</u> $A_s = \lambda s$. <u>Then</u> N <u>is a Poisson process of rate</u> λ.

This result and systematic development of the martingale approach to point processes can be found in Bremaud and Jacod (1977) or Liptser and Shiryayev (1978) Chapter 18.

Square integrable martingales provide a second, and more important, use for the decomposition. Let M be a square integrable martingale. Then M_t^2 is a submartingale, and its compensator has acquired a special name $\langle M\rangle_t$ (or $\langle M,M\rangle_t$). As an exercise in using the tools of Chapter 6, we prove some simple facts about $\langle M\rangle$.

(7.8) LEMMA <u>Let</u> M <u>be a square integrable martingale</u>.

(a) $E(\langle M\rangle_T - \langle M\rangle_S) = E(M_T - M_S)^2$ <u>for bounded stopping times</u> $S \le T$.

(b) $E(\langle M\rangle_T - \langle M\rangle_{T-}) = E(M_T - M_{T-})^2$ <u>for bounded predictable</u> $T > 0$.

(c) <u>Outside a null set</u> Ω_0, <u>the sample path</u> $M_t(\omega)$ <u>is constant on each interval on which the sample path</u> $\langle M\rangle_t$ <u>is constant.</u>

<u>Proof</u>.
$$E(\langle M\rangle_T - \langle M\rangle_S) = E(M_T^2 - M_S^2) \quad \text{by (7.6)}$$
$$= E(M_T - M_S)^2 \quad \text{since } E(M_T - M_S \mid \mathcal{F}_S) = 0.$$

This is (a), and (b) follows by considering a sequence announcing T. To prove (c), fix a rational $r < L$. Define

$$S_n = \inf\{t > r : |M_t - M_r| \ge n^{-1}\}$$
$$S = \inf\{t > r : M_t \ne M_r\}$$
$$U_m = \inf\{t > r : \langle M\rangle_t \ge \langle M\rangle_r + m^{-1}\} \wedge L$$
$$U = \inf\{t > r : \langle M\rangle_t > \langle M\rangle_r\} \wedge L.$$

We must prove $S \ge U$ a.s.. For any stopping time T with $r \le T \le L$,

$$E(M_{S_n \wedge T} - M_r)^2 = E(\langle M\rangle_{S_n\wedge T} - \langle M\rangle_r) \quad \text{by (a).}$$

Now U_m is predictable by (6.11). So by considering an announcing sequence,

$$E(M_{S_n\wedge(U_m-)} - M_r)^2 = E(\langle M\rangle_{S_n\wedge(U_m-)} - \langle M\rangle_r)$$
$$\le m^{-1} \quad \text{by definition of } U_m.$$

So $P(S_n < U_m) \to 0$ as $m \to \infty$, by definition of S_n. Since $U_m \downarrow U$, we have $P(S_n < U) = 0$. But $S_n \downarrow S$, so $S \ge U$ a.s..

Although Theorem 7.4 gives no method of calculating $\langle M \rangle$, one can in practice find it whenever M has an explicit definition. For Brownian motion W, simple computations show that $W_t^2 - t$ is a martingale, so $\langle W \rangle_t = t$. And this characterises Brownian motion, in the following sense.

(7.9) LEVY'S THEOREM (Liptser and Shiryayev (1977) Th.4.1). Let M be a continuous square integrable martingale such that $\langle M \rangle_t = t$. Then M is Brownian motion; more specifically, for each t the process $(M_{t+u} - M_t)_{u \geq 0}$ is distributed as Brownian motion independent of $\mathcal{F}_t$.

As another example where $\langle M \rangle$ is manifest, let $(\xi_i, \mathcal{F}_i)$ be a discrete-time square integrable martingale. Given constants $0 < t_1 < t_2 < \dots\dots$, form a continuous-time martingale by

(7.10) $$M_t = \sum_{i:t_i \leq t} \xi_i .$$

Then $$\langle M \rangle_t = \sum_{i:t_i \leq t} E(\xi_i^2 | \mathcal{F}_{i-1}) .$$

Remark. The notion $\langle M \rangle$ can be extended from square integrable martingales to a rather more general class of processes, the semimartingales (DM.VII). From the theoretical viewpoint, semimartingales provide the natural setting for the theorems of the next chapter. We stick to square integrable martingales: the generalisations require greater technical skill with the general theory, but the same weak convergence ideas can be used.

CHAPTER 3

We now have the tools for a systematic technique for establishing weak convergence theorems:

(i) find a martingale characterisation of the desired limit X;

(ii) prove tightness via Theorem 4.4;

(iii) identify subsequential limits via Proposition 5.1.

The purpose of this chapter is to illustrate this technique by proving simple weak convergence theorems for a variety of types of process.

Because we combine ideas from the weak convergence theory and from the general theory of processes, let us explain carefully the set-up we adopt for the rest of the book. By a <u>process</u> (X,F) we mean a Skorohod process X adapted to a filtration F. When we have a sequence (X^n,F^n) of processes, the filtrations F^n and indeed the probability spaces Ω^n will in general be different for different n. By the device of forming a product space, there is no loss of generality in assuming a common probability space $(\Omega,\mathfrak{F},P)$. But the filtrations remain different for different n: for example, it would be quite unnatural to try to discuss convergence of random walks to Brownian motion with reference to a single filtration. When a process X is constructed as a weak limit, it is given its usual filtration.

In practice we shall often be less formal and just write "a process X ", leaving the filtration implicit.

In this chapter we consider only real-valued processes.

8 THE MARTINGALE INVARIANCE PRINCIPLE

We shall consider martingales satisfying

(8.1) $M_0 = 0$, $E(M_t)^2 < \infty$, for each t.

For in most practical settings we can reduce a given martingale M to one satisfying (8.1) by first considering $M_t - M_0$ and then truncating.

Levy's theorem suggests that for a sequence (M^n) to converge to Brownian motion W, it should suffice to show

(i) the discontinuities become small

(ii) $\langle M^n, M^n \rangle_t$ approaches t.

Write $A^n = \langle M^n, M^n \rangle$. Note that the following formalisations of (ii) are equivalent, by Lemma 3.1.

(8.2)(a) $A^n_t \underset{p}{\rightarrow} t$, for each t.

(b) $A^n \Rightarrow t$.

In the special case of continuous martingales, this is sufficient for convergence.

(8.3) PROPOSITION. <u>Let</u> (M^n) <u>be a sequence of continuous martingales satisfying</u> (8.1) <u>and</u> (8.2). <u>Then</u> $M^n \Rightarrow W$.

The proof is based on the well-known representation of continuous martingales as time-changes of Brownian motion. The following crude version of this representation is all we need.

(8.4) LEMMA. <u>Let</u> M <u>be a continuous martingale satisfying</u> (8.1). <u>Let</u> $A = \langle M, M \rangle$ <u>and suppose there exists</u> t_0 <u>such that</u>

(8.5) $A_t \geq t - t_0$.

<u>Let</u> τ <u>be the inverse of</u> A (<u>that is,</u> $\tau_t = \inf\{s : A_s > t\}$). <u>Let</u> $W = M \circ \tau$ (<u>that is</u>, $W_t = M_{\tau_t}$), <u>and let</u> $\mathcal{G}_t = \mathcal{F}_{\tau_t}$. <u>Then</u> (W,G) <u>is Brownian motion</u>.

Proof. We verify the hypotheses of Levy's theorem. W is continuous by Lemma 7.8(c). Assumption (8.5) implies $\tau_t \leq t + t_0$. So for $s < t$ the optional sampling theorem implies

$$E(W_t \mid \mathcal{G}_s) = E(M_{\tau_t} \mid \mathcal{F}_{\tau_s}) = M_{\tau_s} = W_s$$

$$E(W_t^2 - W_s^2 \mid \mathcal{G}_s) = E(M_{\tau_t}^2 - M_{\tau_s}^2 \mid \mathcal{F}_{\tau_s}) = E(A_{\tau_t} - A_{\tau_s} \mid \mathcal{F}_{\tau_s}) = t-s.$$

Proof of Proposition 8.3. Replacing M^n by $M^n_{t\wedge n} + \hat{W}_{(t-n)\vee 0}$, where $\hat{W}$ is an independent Brownian motion, we see there is no loss of generality in assuming M^n satisfies (8.5) for $t_0 = n$. Let $W^n = M^n \circ \tau^n$, in the notation of Lemma 8.4. By (8.2), $(W^n, A^n) \Rightarrow (W, t)$. So $M^n = W^n \circ A^n \Rightarrow W \circ t = W$, using a simple result on the continuity of random time changes [B, eq. 17.9].

Remark. Discontinuous martingales also can be embedded into Brownian motion (see e.g. Drogin (1972)), and this leads to the "Skorohod embedding" technique for proving martingale invariance principles. However, our aim is to illustrate the technique outlined at the start of the chapter.

Here is a simple result (exercise) about convergence of arbitrary processes.

(8.6) LEMMA. Suppose $X^n \Rightarrow X$. The following are equivalent.

(i) X is continuous.

(ii) $\sup_{t \leq L} |X^n_t - X^n_{t-}| \underset{p}{\to} 0$.

(iii) $\sup_{\substack{s,t \leq L \\ |s-t| \leq \delta_n}} |X^n_t - X^n_s| \underset{p}{\to} 0$, for each sequence $\delta_n \downarrow 0$.

In seeking to prove the invariance principle for discontinuous martingales, (ii) is a natural necessary condition to impose. However, (ii) and (8.2) are not sufficient: we need integrability conditions also. In the L^2 setting it is natural to be interested in convergence of moments, and then we obtain a definitive result.

(8.7) THEOREM. Let (M^n) be a sequence of martingales satisfying (8.1). Then (I) and (II) are equivalent.

(I)
(a) $A^n \Rightarrow t$;
(b) $EA^n_t \to t$;
(c) $E \sup_{t \leq L} (M^n_t - M^n_{t-})^2 \to 0$.

(II)
(d) $M^n \Rightarrow W$;
(e) $E(M^n_t)^2 \to t$.

Proof. Here is a technical lemma whose proof is deferred.

(8.8) LEMMA. The following are equivalent.

(i) $\{(M_t^n)^2 : n \geq 1\}$ is uniformly integrable, for each t;

(ii) $\{A_t^n : n \geq 1\}$ is uniformly integrable, for all t;

(iii) $\{\sup_{s \leq t} (M_s^n)^2 : n \geq 1\}$ is uniformly integrable, for all t.

Hypothesis (II) implies the first assertion of the lemma; hypothesis (I) implies the second assertion. Hence all three assertions are true under either hypothesis. Observe also that $E\, A_t^n = E\, (M_t^n)^2$, so conditions (b) and (e) are equivalent. Also, (c) follows from (d), Lemma 8.6 and Lemma 8.8(iii). So we need only establish the "weak convergence" conditions (a) and (d).

Proof of (d). To prove (M^n) is tight, let $T_n \leq L$ be stopping times and let $\delta_n \downarrow 0$. Then

$$(8.9) \quad E\left(M^n_{T_n+\delta_n} - M^n_{T_n} \right)^2 = E\left(A^n_{T_n+\delta_n} - A^n_{T_n} \right).$$

But by (a) and Lemma 8.6, $A^n_{T_n+\delta_n} - A^n_{T_n} \underset{p}{\to} 0$, and uniform integrability extends this to L^1 convergence. So (M^n) is tight by Theorem 4.4. Let M be a subsequential weak limit of (M^n). We must show M = W. Passing to a subsequence

(8.10) $(M^n, A^n) \Rightarrow (M, t)$ on D(R) x D(R).

Proposition 5.1 and Lemma 8.8 show that M is a martingale. By (8.10) and Corollary 3.6,

$$(M^n)^2 - A^n \Rightarrow M^2 - t .$$

Proposition 5.1 and Lemma 8.8 show that $M_t^2 - t$ is a martingale. And M is continuous by (c) and Lemma 8.6. Levy's theorem identifies M as W.

Proof of (a). The proof that (A^n) is tight is precisely the same as the proof above that (M^n) is tight, using (8.9) in the other direction. Let A be a subsequential weak limit of (A^n). First, we assert that A is continuous. To see this, apply Lemma 7.8(b) to stopping times of the form

$$T_n = \inf\{t: A^n_t - A^n_{t-} \geq \delta\} \wedge L$$

and use (c) and Lemma 8.6. (We have already seen that (c) follows from (II)). Next, passing to a subsequence,

$$(M^n, A^n) \Rightarrow (W, A) \quad \text{on } D(R) \times D(R).$$

Using Proposition 5.1 and Lemma 8.8 as in the proof of (d), we see that W and $W^2 - A$ are martingales with respect to G, the usual filtration of (W,A). Thus $A = \langle W,W \rangle$, where $\langle \, , \rangle$ is interpreted with respect to G. But Proposition 7.11 shows that $\langle W,W \rangle$ does not depend on the filtration, so $A_t = t$ as desired.

Proof of Lemma 8.8. Let M be a martingale satisfying (8.1), $A = \langle M,M \rangle$.

(ii) implies (i). Fix $\lambda > 0$, $t < \infty$. Let $T = \inf\{s : |M_s| \geq \lambda\}$. The elementary inequality $\frac{1}{2}(a+b)^2 \leq a^2 + b^2$ gives

(8.11) $$\tfrac{1}{2}E(|M_t| - \lambda)^2 1_{(T \leq t)} \leq E(M_t - M_T)^2 1_{(T \leq t)} + E(|M_T| - \lambda)^2 1_{(T \leq t)}$$

The first term on the right equals $E(A_t - A_T)1_{(T \leq t)}$ by (7.8), and this is bounded by $EA_t 1_{(T \leq t)}$. The second term on the right is bounded by $E \sup_{s \leq t} (M_s - M_{s-})^2$. Since $\{T \leq t\} = \{|M_T| \geq \lambda\} \supset \{|M_t| \geq \lambda\}$,

(8.12) $$E(|M_t| - \lambda)^2 1_{(|M_t| \geq \lambda)} \leq 2EA_t 1_{(|M_T| \geq \lambda)} + 2E \sup_{s \leq t} (M_s - M_{s-})^2$$

and $P(M_T \geq \lambda) \leq \lambda^{-2} E M_T^2 \leq \lambda^{-2} E M_t^2 = \lambda^{-2} EA_t$.

Apply this to a sequence (M^n) satisfying (I). Using (c) and the uniform integrability of (A^n),

$$\limsup_{n \to \infty} E(|M_t^n| - \lambda)^2 1_{(|M_t^n| \geq \lambda)} \to 0 \quad \text{as} \quad \lambda \to \infty .$$

This implies that $(M_t^n)^2$ is uniformly integrable (recall discussion in Section 1).

<u>(i) implies (ii)</u>. This is very similar. This time, put $T = \inf\{s: A_s \geq \lambda\}$. So $\{T \leq t\} = \{A_t \geq \lambda\}$. Now

$$E(A_t - \lambda)1_{(A_t \geq \lambda)} = E(A_t - \lambda)1_{(T \leq t)}$$

$$\leq E(A_t - A_T)1_{(T \leq t)} + E(A_T - A_{T-})1_{(T \leq t)}. \tag{8.13}$$

By (7.5) the first term on the right equals $E(M_t^2 - M_T^2)1_{(T \leq t)}$. And since T is predictable by (6.11), by considering a sequence announcing T, and using (7.5),

$$E(A_T - A_{T-})1_{(T \leq t)} = E(M_T^2 - M_{T-}^2)1_{(T \leq t)}.$$

Now (8.13) gives

$$E(A_t - \lambda)1_{(A_t \geq \lambda)} \leq E\,M_t^2 1_{(T \leq t)}. \tag{8.14}$$

But $P(T \leq t) = P(A_t \geq \lambda) \leq \lambda^{-1} EA_t = \lambda^{-1} EM_t^2$. So uniform integrability of a sequence $(M_t^n)^2$ implies uniform integrability of (A_t^n).

<u>(i) implies (iii)</u>. Fix t, and set $S = \sup_{s \le t} |M_s|$. Let $\lambda \ge 0$. We shall prove

(8.15) $\quad E((S-\lambda)\vee 0)^2 \le 4 E M_t^2 1_{(|M_t| \ge \lambda)}$,

and the result follows. Actually, for $\lambda = 0$ (8.15) is a well-known inequality of Doob (see e.g. Williams (1979)), and our argument below is just a slight modification of the usual argument for $\lambda = 0$.

Fix $c > \lambda$. Let $T = \inf\{s : |M_s| \ge c\}$, so

(8.16) $\quad E(|M_t| \mid T \le t) \ge c.$

Putting $Z = |M_t| 1_{(|M_t| \ge \lambda)}$ we can rewrite (8.16) as

$$E(Z \mid T \le t) \ge c - E(|M_t| 1_{(|M_t| < \lambda)} \mid T \le t)$$

$$\ge c - \lambda .$$

Since $\{T \le t\} = \{S \ge c\}$, this gives

(8.17) $\quad (c-\lambda)P(S \ge c) \le EZ1_{(S \ge c)}.$

We now compute

$$E((S-\lambda)\vee 0)^2 = -\int_{c=\lambda}^{\infty} (c-\lambda)^2 \, dP(S \ge c)$$

$$= 2\int (c-\lambda)\, P(S \ge c)\, dc - \lim_{c\to\infty} c^2 P(S \ge c) \quad \text{(integration by parts)}$$

$$\le 2\int EZ1_{(S \ge c)}\, dc \quad \text{by (8.17)}$$

$$= 2\, E\, Z((S-\lambda)\vee 0)$$

$$\le 2\, \{EZ^2\}^{\frac{1}{2}} \cdot \{E((S-\lambda)\vee 0)^2\}^{\frac{1}{2}} \quad \text{(Cauchy-Schwarz)}.$$

Rearranging this, we get (8.15).

Let us see what Theorem 8.7 means in the classical case of a square integrable martingale difference sequence (ξ_i), that is where $E(\xi_i | \mathcal{F}_{i-1}) = 0$ for $\mathcal{F}_i = \sigma(\xi_1, \ldots, \xi_i)$. Set

$$S_n = \sum_1^n \xi_i \quad , \quad M_t^n = n^{-\frac{1}{2}} S_{[nt]}$$

Then $\langle M^n, M^n \rangle_t = n^{-1} \sum_1^{[nt]} E(\xi_i^2 | \mathcal{F}_{i-1})$, and the conditions of (I) become

(8.18)

(a) $n^{-1} \sum_1^n E(\xi_i^2 | \mathcal{F}_{i-1}) \underset{p}{\longrightarrow} 0$;

(b) $n^{-1} \sum_1^n E\,\xi_i^2 \to 1$;

(c) $n^{-1} E \max_{i \leq n} \xi_i^2 \to 0$.

Note that (c) is weaker than

(c') $\sup_i E|\xi_i|^{2+\delta} = K < \infty$, for some $\delta > 0$.

For (c') implies

$$n^{-1} P(\max_{i \leq n} \xi_i^2 \geq x) \leq \max_{i \leq n} P(\xi_i^2 \geq x) \leq Kx^{-(1+\delta/2)} .$$

If one is interested only in proving weak convergence $M^n \Rightarrow W$, without caring about convergence of moments, it is natural to suppose that some of the integrability hypotheses in (I) could be removed. The next example shows one cannot go far in this direction.

For each n let $\{\xi_{n,i} : i \geq 1\}$ be i.i.d. with

$P(\xi_{n,i} = n) = P(\xi_{n,i} = -n) = 1/2n^3$; $P(\xi_{n,i} = 0) = 1 - 1/n^3$.

Let $M_t^n = n^{-\frac{1}{2}} \sum_1^{[nt]} \xi_{n,i}$. Then $M^n \Rightarrow 0$. But conditions (a) and (b) of (8.18) hold, and also

$$\sup_{t \leq L} |M_t^n - M_{t-}^n| = n^{-\frac{1}{2}} \max_{i \leq nL} |\xi_i| \underset{p}{\longrightarrow} 0 .$$

Thus to deduce $M^n \Rightarrow W$ in Theorem 8.7 we cannot replace (c) by the (necessary) condition $\sup_{t \leq L} | M_t^n - M_{t-}^n | \underset{p}{\longrightarrow} 0$. The classical central limit theorem for independent variables suggests that we need Lindeberg-type conditions, but we shall not pursue this topic.

CONVERGENCE TO DIFFUSIONS

The technique used to prove convergence of martingales to Brownian motion can also be used to discuss convergence of arbitrary processes to diffusions. The ideas are exactly the same, though the details become more messy. We must first discuss the analogue of Lévy's theorem for diffusions.

For Brownian motion W and constants $a > 0$, b, call the process $Y_t = a^{\frac{1}{2}}W_t + bt$ _Brownian motion with drift_ b _and variance_ a. Levy's theorem shows that Y is characterised amongst continuous processes started at 0 by

$$(8.19) \qquad \begin{aligned} M_t &= Y_t - bt && \text{is a martingale} \\ S_t &= M_t^2 - at && \text{is a martingale.} \end{aligned}$$

Now let $a(x) > 0$, $b(x)$ be bounded continuous functions on the line. Informally, we can construct a continuous Markov process X such that, conditional on $X_{t_0} = x$, the incremental process $Y_t = X_{t_0+t} - x$ behaves like Brownian motion with drift $b(x)$ and variance $a(x)$ over an infinitesimal time interval $0 < t < \varepsilon$. Such a process - which we call a diffusion with drift $b(x)$ and variance $a(x)$ - may be defined formally in several ways, but the formal definitions need not concern us. The informal description of X and (8.19) suggest the properties

$$(8.20) \qquad \begin{aligned} M_t &= X_t - \int_0^t b(X_s)\, ds && \text{is a martingale} \\ S_t &= M_t^2 - \int_0^t a(X_s)\, ds && \text{is a martingale.} \end{aligned}$$

(we interpret "martingale" with respect to the usual filtration of X). Let us quote the following facts from diffusion theory.

(8.21) PROPOSITION. _For each_ $x \in R$ _there exists a continuous process_ X _satisfying_ (8.20) _with_ $X_0 = x$. _The distribution_ Δ_x _of this process is unique_. _The family_ $(\Delta_x)_{x \in R}$ _constitute a strong Markov process._

This result can be found in Stroock and Varadhan (1979) Exercise 4.6.6.; there does not seem to be any simple proof.

For our purposes, the important fact is that (8.20) characterises the diffusion amongst all continuous processes. We can now use our technique to show that a sequence of processes whose discontinuities become small and such that (8.20) is more and more nearly satisfied will converge to the diffusion.

(8.22) THEOREM. Let $a(x) > 0$, $b(x)$ be bounded continuous functions, and let $x_0 \in R$. Let X be the diffusion with drift $b(x)$ and variance $a(x)$, and $X_0 = x_0$. Let (X^n, F^n) be a sequ nce of processes. Suppose

(a) $X_0^n \Rightarrow x_0$

(b) $E \sup_{t \leq L} (X_t^n - X_{t-}^n)^2 \to 0$ as $n \to \infty$.

Suppose also that for each n there exist N^n and N^n adapted to F^n such that

(c) (M^n, F^n) is a martingale, where $M_t^n = X_t^n - \int_0^t b(X_s^n)\,ds - N_t^n$

(d) (S^n, F^n) is a martingale, where $S_t^n = (M_t^n)^2 - \int_0^t a(X_s^n)\,ds - N_t^n$

(e) $\sup_{T \in \mathcal{J}_L^n} E\,(N_T^n)^2 \to 0$ as $n \to \infty$.

(f) $\sup_{T \in \mathcal{J}_L^n} E|N_T^n| \to 0$ as $n \to \infty$.

Then $X^n \Rightarrow X$.

Remarks. (i) Note that there is no assumption that M^n, S^n be similar to M, S in (8.20).

(ii) The case $a(x) \equiv 1$, $b(x) \equiv 0$, $x_0 = 0$, $N^n \equiv 0$ is essentially the main implication of Theorem 8.7.

(iii) These conditions are not necessary for weak convergence. But they turn out to be essentially necessary and sufficient for extended weak convergence

( )

Proof of Theorem 8.22. Suppose we can prove

(8.23) LEMMA. $\{(X_t^n)^2 : n \geq 1\}$ is uniformly integrable, for each t.

(8.24) LEMMA. (X^n) is tight.

Then passing to a subsequence suppose $X^n \Rightarrow Y$, say. Then $Y_0 = x_0$ a.s. by (a), and Y is continuous by (b). By the continuous mapping theorem and (c)-(f),

$$(X^n, M^n) \Rightarrow (Y, Q), \quad \text{where } Q_t = Y_t - \int_0^t b(Y_s)\, ds$$

$$(X^n, S^n) \Rightarrow (Y, Q'), \text{ where } Q'_t = Q_t^2 - \int_0^t a(Y_s)\, ds.$$

By Lemma 8.23 we can apply Proposition 5.1 and deduce that Q and Q' are martingales (with respect to the usual filtration of Y). So $\mathcal{L}(Y) = \mathcal{L}(X)$ by the uniqueness assertion of Proposition 8.21.

Proof of Lemma 8.24. For fixed $L < \infty$ define

$$\bar{b} = \sup |b(x)|$$
$$\bar{a} = \sup \; a(x)$$
$$c_n = \sup_{T \in \mathcal{T}_L^n} E\,(N_T^n)^2$$
$$d_n = \sup_{T \in \mathcal{T}_L^n} E|\tilde{N}_T^n|.$$

We shall prove that, for $T \in \mathcal{T}_{L-1}^n$ and $0 < \delta < 1$,

$$(8.25) \quad E(X_{T+\delta}^n - X_T^n)^2 \leq 4(\bar{a}\,\delta + 2d_n + \bar{b}^2\delta^2 + 2c_n).$$

Then for a sequence $T_n \in \mathcal{T}_{L-1}^n$, $\delta_n \downarrow 0$ we have $E(X_{T_n+\delta_n}^n - X_{T_n}^n)^2 \to 0$ by (e) and (f), and so (X^n) is tight by Theorem 4.4.

To prove (8.25), fix n and omit it. By (c)

Proof of Lheorem 8.22. Replacing X^n by $X_t^n 1_{(|X_0^n - x_0| \le \varepsilon_n)} + X_t 1_{(|X_0^n - x_0| > \varepsilon_n)}$ for $\varepsilon_n \downarrow 0$ suffiviently slowly, we may suppose (a) is strengthened to

(a') $E(X_0^n - x_0)^2 \to 0.$

We shall prove

(8.26) $|X_{T+\delta} - X_T| \le |M_{T+\delta} - M_T| + \bar{b}\delta + |N_{T+\delta}| + |N_T|.$

From the elementary inequality

(8.27) $(\sum_1^m y_i)^2 \le m(\sum_1^m y_i^2)$

we deduce

~~From the elementary inequality $(\sum_1^4 y_i)^2 \leq 4(\sum_1^4 y_i^2)$ we deduce~~

$$E(X_{T+\delta} - X_T)^2 \leq 4(E(M_{T+\delta} - M_T)^2 + \bar{b}^2\delta^2 + 2c).$$

But M is a martingale, so

$$E(M_{T+\delta} - M_T)^2 = EM^2_{T+\delta} - EM^2_T \leq \bar{a}\,\delta + 2d \quad \text{by (d)},$$

and this gives (8.25).

Proof of Lemma 8.24. We keep the notation of the lemma above, and introduce

$$e_n = E \sup_{t \leq L} (X^n_t - X^n_{t-})^2.$$

Fix $\lambda < \infty$. We shall derive the following estimate.

(8.28) $E(|X^n_L| - \lambda)^2 1_{(|X^n_L| > \lambda)} \leq 16(c_n + d_n + e_n) + 8\lambda^{-1}(\bar{b}L^2 + L\bar{a})\eta$

where $\eta = \bar{b}L + c_n^{\frac{1}{2}} + (2d_n + \bar{a}^{\frac{1}{2}}L + 2(c_n + E(X^n_0)^2))^{\frac{1}{2}}$.

By hypothesis c_n, d_n and e_n converge to zero, so

$$\limsup_{n \to \infty} E(|X^n_L| - \lambda)^2 1_{(|X^n_L| > \lambda)} \to 0 \quad \text{as} \quad \lambda \to \infty$$

establishing the lemma.

To prove (8.28), fix n and omit it. Set $T = \inf\{t : |X_t| \geq \lambda\}$, so $\{|X_L| > \lambda\} \subset \{T \leq L\}$. Consider

$$E(|X_L| - \lambda)^2 1_{(T \leq L)} \leq E(|X_L - X_T| + (|X_T| - \lambda))^2 1_{(T \leq L)}$$

$$2(E(X_L - X_T)^2 1_{(T \leq L)} + e),$$

using (8.27) and the fact $|X_T| - \lambda \leq |X_T - X_{T-}|$. As in (8.26) we can write

$$|X_L - X_T| \leq |M_L - M_T| + \bar{b}L + |N_T| + |N_L| \qquad \text{on } \{T \leq L\}.$$

So using (8.27)

$$E(X_L - X_T)^2 1_{(T \leq L)} \quad 4\{E(M_L - M_T)^2 1_{(T \leq L)} + \bar{b}^2 L^2 P(T \leq L) + 2c\}.$$

But M is a martingale, so

$$E(M_L - M_T)^2 1_{(T \leq L)} = E(M_L^2 - M_T^2) 1_{(T \leq L)}$$

$$\leq L\bar{a}P(T \leq L) + 2d \qquad \text{by (d).}$$

It remains only to estimate $P(T \leq L)$. Observe that $\{T \leq L\} \subset \{|X_{T\wedge L}| \geq \lambda\}$ and so $P(T \leq L) \leq \lambda^{-1}\eta$, where $\eta = E\, X_{T\wedge L}$. Now

$$E|X_{T\wedge L}| - \bar{b}L - c^{\frac{1}{2}} \leq E|M_{T\wedge L}| \qquad \text{by (c)}$$

$$E|M_L| \leq (E\, M_L^2)^{\frac{1}{2}}.$$

And $E\, M_L^2 \leq ES_L + \bar{a}L + d$ by (d)

$$ES_L = ES_0$$

$$ES_0 - d \leq E(X_0 + N_0)^2 \qquad \text{by (d)}$$

$$\leq 2(EX_0^2 + c) \qquad \text{by (8.27).}$$

Putting together our estimates gives (8.28).

Notes on Chapter 3. The literature on weak convergence is so vast that we dare not attempt tl give complete references. Here is a brief history of the martingale technique. Stroock and Varadhan (1969) used it for proving convergence of Markov processes to diffusions; Grigelionis (197) observed that the technique could be used more generally; Rebolledo (19) gave a systematic treatment from the general theory of processes viewpoint. But the latter two authors, not having Theorem 4.4, were forced to use more complicated tightness arguments.

Notes on Section 8. Hall and Heyde (198) discuss the central limit theorem and invariance principle for discrete martingales. For the continuous case see Rebolledo (198), Helland (198)

For recent work on convergence to diffusions see Helland (
Rebolledo (19),

9 POINT PROCESSES

Our technique readily applies to convergence of point processes to the Poisson process.

(9.1) PROPOSITION. Let (N^n) be a sequence of point processes such that $E\,N^n_t < \infty$ for all n,t. Let (A^n) be their compensators. Let $0 < \lambda < \infty$ and let N denote the Poisson process with rate λ. Suppose

(9.2) $$A^n_t \underset{p}{\longrightarrow} \lambda t, \quad \text{for each } t.$$

Then $N^n \Rightarrow N$.

Proof. By Lemma 3.1, (9.2) implies

(9.3) $$A^n \Rightarrow \lambda t.$$

Let us first assume the extra hypothesis

(9.4) $\{N^n_t, A^n_t : n \geq 1\}$ is uniformly integrable, for each t.

Let $T_n \leq L$ be a sequence of stopping times, and let $\delta_n \downarrow 0$. Then

$$E(N^n_{T_n+\delta_n} - N^n_{T_n}) = E(A^n_{T_n+\delta_n} - A^n_{T_n})$$

$$\to 0 \quad \text{by (9.3) and Lemma 8.6.}$$

So (c.f. Remark 4,) we can apply Theorem 4.4 to (N^n,A^n) and deduce

(N^n,A^n) is tight on D(RxR).

Passing to a subsequence we may suppose

(9.5) $(N^n,A^n) \Rightarrow (N^0,A^0)$, say, on D(RxR).

By (9.3), $A^0_t = \lambda t$. So by the continuous mapping theorem

$$(N^n, N^n - A^n) \Rightarrow (N^0, N^0 - \lambda t).$$

Now $N^n - A^n$ is a martingale, so Proposition 5.1 and (9.4) imply that $N^0_t - \lambda t$ is a martingale. But N^0 is a point process, since the set $D_+(R)$ of counting functions is closed in D(R). Theorem 7.7 identifies N^0 as the Poisson process of rate λ.

This proves the Proposition under hypothesis (9.4). In the general case we use a truncation argument. Let

$$T_n = \inf\{t: A_t^n \geq \lambda t + 1\}.$$

By (9.3), $T_n \underset{p}{\to} \infty$. Let $\tilde{N}_t^n = N_{t \wedge T_n}^n$, which has compensator $\tilde{A}_t^n = A_{t\wedge T_n}^n$. It suffices to show that $\tilde{N}^n$, $\tilde{A}^n$ satisfy (9.4), for then $\tilde{N}^n \Rightarrow N$ and so $N^n \Rightarrow N$.

Fix t, and note

$$A_{t\wedge T_n}^n \leq \lambda t + 1 + \sup_s (A_s^n - A_{s-}^n).$$

From (9.8) below we see

(9.6) $\{ \tilde{A}_t^n : n \geq 1 \}$ is uniformly integrable.

Now for fixed c define $U_n = \inf\{s: \tilde{N}_s^n \geq c\} \wedge t$. Then

$$\begin{aligned} E(\tilde{N}_t^n - c - 1)^+ 1_{(\tilde{N}_t^n \geq c)} &\leq E(\tilde{N}_t^n - \tilde{N}_{U_n}^n) \\ &= E(\tilde{A}_t^n - \tilde{A}_{U_n}^n) \\ (9.7) \qquad &\leq E\, \tilde{A}_t^n 1_{(\tilde{N}_t^n \geq c)} . \end{aligned}$$

But $P(N_t^n \geq c) \leq c^{-1} E\tilde{N}_t^n = c^{-1} E\tilde{A}_t^n \to c^{-1}\lambda t$ by (9.6) and (9.2). So (9.7) and (9.6) establish the uniform integrability of $\{\tilde{N}_t^n\}$.

Finally, we used the "obvious" fact that the compensator A of any point process N satisfies

(9.8) $\sup_s (A_s - A_{s-}) \leq 1$ a.s.

To prove this, let $V = \inf\{s: A_s - A_{s-} \geq 1+\varepsilon\}$. By (6.11) V is predictable, and by considering an announcing sequence

$$E(A_V - A_{V-}) 1_{(V<\infty)} = E(N_V - N_{V-}) 1_{(V<\infty)}.$$

But $A_V - A_{V-} > 1 \geq N_V - N_{V-}$ on $\{V<\infty\}$, and so $P(V<\infty) = 0$.

By analogy with Theorem 8.7, one might expect condition (9.2) to be necessary as well as sufficient. The next example shows it is not: however, we shall see in section 21 that (9.2) is essentially necessary and sufficient for extended weak convergence.

(9.9) <u>Example</u>. Let U_1 be uniform on (0,1), and let $U_2^n = \{nU_1\}$, where $\{x\} = x - [x]$ denotes the fractional part of x. Then U_2^n is uniform on (0,1) and $(U_1, U_2^n) \Rightarrow (U_1, U^*)$ where U^* is independent of U_1. Now let

$T_1 = -\log U_1$,

$T_2^n = -\log U_2^n$,

$T_3, T_4, \ldots\ldots$ be i.i.d. independent of T_1 with exponential (mean 1) distributions,

$N_t^n = \max\{i: T_1 + T_2^n + T_3 + \ldots + T_i \leq t\}$.

Then N^n converges weakly to the Poisson process of rate 1, but (9.2) fails because T_2^n is a function of T_1 and so

$$A_t^n = A_{T_1}^n , \quad T_1 \leq t < T_2^n.$$

Proposition 9.1 can be generalised in several ways. One can remove the integrability condition; treat convergence to time-dependent Poisson processes or to doubly stochastic Poisson processes (i.e. mixtures of time-dependent Poisson processes). Alternatively, for processes with conditional intensities (i.e. if $A_t^n = \int_0^t a_n(s)\, ds$) we can seek conditions under which weak convergence is strengthened to convergence in total variation. References may be found in the Notes.

We should, however, point out an inherent limitation of our technique. To use the technique for convergence to a general point process N, we would need to know that the distribution $\mathcal{L}(N)$ was determined by the distribution $\mathcal{L}(A)$ of its compensator. But this is false in general.

(9.10) Example. Let N be the Poisson process of rate 1, so $P(N_1 = 0) = P(N_1 = 1) = e^{-1}$. For $j = 0,1$ let $T_j = 1$ if $N_1 = j$; $T_j = \infty$ otherwise. Let $N^j_t = N_{t\wedge T_j}$. Then N^j has compensator $A^j_t = t\wedge T_j$. So here $\mathcal{L}(A^0) = \mathcal{L}(A^1)$ but $\mathcal{L}(N^0) \neq \mathcal{L}(N^1)$.

Actually, it is not hard to show that doubly stochastic Poisson processes are the only point processes which are characterised by the distribution of their compensator.

Notes on section 9. Brown (1978) gave a rather complicated version of Proposition 9.1. Brown (1981) and Kabanov, Liptser and Shiryayev (1980) give generalisations.

Since processes with independent increments are sums of Brownian motion and Poisson processes, it is natural to suppose that Theorem 8.7 and Proposition 9.1 can be combined to give conditions for convergence to such processes: this project has been carried out by Jacod and Memin (1980).

10 WORKED EXAMPLES

The results given so far in this chapter have been rather abstract, and have involved processes plainly related to martingales. To demonstrate that our technique is more widely applicable, we shall in this section re-prove two results in [B], as "worked examples" of the technique (We do not claim our proofs are substantially simpler than those of [B]).

SAMPLING

Let $x_1, \dots, x_k$ be real numbers such that

(10.1) $$\sum x_i = 0 , \quad \sum x_i^2 = 1.$$

Let $\xi_1, \dots, \xi_k$ be a uniform random permutation of $\{1, \dots, k\}$. Define a random element Y of $D[0,1]$ by

(10.2) $$Y_t = \sum_{i=1}^{[kt]} x_{\xi_i}$$

(10.3) PROPOSITION [B,24.1] _Suppose for each_ n _the sequence_ $x_{n,1}, \dots, x_{n,k_n}$ _satisfies (10.1), and define_ Y^n _by_ (10.2). _If_

(10.4) $$\max_i |x_{n,i}| \to 0 \text{ as } n \to \infty$$

then $Y^n \Rightarrow W^o$, _where_ W^o _is the Brownian bridge._

Proof. To simplify computations, we shall replace the Y defined in (10.2) by an asymptotically equivalent process X defined below. Let $U_1, U_2, \dots, U_k$ be independent, uniformly distributed on $(0,1)$, and let

(10.5) $$X_t = \sum_1^k x_i 1(U_i \le t) .$$

Now writing $\tilde{U}_1, \dots, \tilde{U}_k$ for the order statistics, and putting $\tau_t = U_{[kt]}$, then the time-changed process $X \circ \tau$ is distributed as the process Y of (10.2). Since $\tau_t^n \Rightarrow t$ by the Glivenko-Cantelli theorem, it suffices to prove

(10.6) $$X^n \Rightarrow W^o.$$

Consider a process X defined by (10.5). Let $S_t = \sum x_i^2 1_{(U_i \le t)}$.
Given a subset A of $\{1,\ldots,k\}$ and $0 < t < 1$ write

$$x = \sum_{i \in A} x_i, \quad \sigma = \sum_{i \in A} x_i^2, \quad \Omega_0 = \{\omega : \{i : U_i \le t\} = A\}.$$

Conditional on Ω_0 the random variables $\{U_i : i \notin A\}$ are independent, uniformly distributed on (t,1). Easy calculations give, for $t < u < 1$,

$$E(X_u - X_t \mid \Omega_0) = -(u-t)(1-t)^{-1}x$$
$$\mathrm{var}(X_u - X_t \mid \Omega_0) = (u-t)(1-u)(1-t)^{-2}(1-\sigma).$$

Hence

$$E(X_u - X_t \mid \mathcal{F}_t^X) = -(u-t)(1-t)^{-1}X_t \tag{10.7}$$
$$\mathrm{var}(X_u - X_t \mid \mathcal{F}_t^X) = (u-t)(1-u)(1-t)^{-2}(1-S_t).$$

Consider now a sequence X^n of processes, each of the form (10.5), and suppose

(10.8) $S_t^n \to t$ in L^2;

(10.9) $\{(X_t^n)^2 : n \ge 1\}$ is uniformly integrable, for each t;

(10.10) (X^n) is tight on $D[0,1]$.

Let Z be a weak limit of (X^n). Then $Z_0 = Z_1 = 0$, and Z is continuous by (10.4). Proposition 5.1 and (10.7)-(10.9) imply

$$\left.\begin{aligned} E(Z_u - Z_t \mid \mathcal{F}_t^Z) &= -(u-t)(1-t)^{-1}Z_t \\ \mathrm{var}(Z_u - Z_t \mid \mathcal{F}_t^Z) &= (u-t)(1-u)(1-t)^{-1} \end{aligned}\right\} \quad 0 < t < u < 1$$

These properties characterise Z as the Brownian bridge. For the process $W_t = (1+t)Z_{t/(1+t)}$ is easily seen to satisfy the hypotheses of Levy's theorem, and hence is Brownian motion: then $Z_t = (1-t)W_{t/(1-t)}$ is the Brownian bridge.

It remains to verify conditions (10.8)-(10.10). First, we compute

$$E(S_t^n) = t$$

$$\begin{aligned} \operatorname{var}(S_t^n) &= t(1-t) \sum x_{n,i}^4 \\ &\leq t(1-t) \max_i x_{n,i}^2 \quad \text{by (10.1)} \\ &\to 0 \quad \text{by (10.4).} \end{aligned}$$

This proves (10.8). To prove (10.9), we can follow [B,p.211] to get a uniform bound on the fourth moments of X_t^n ; alternatively, stopping time arguments analogous to those in Lemmas 8.8 and 8.23 may be employed. To prove (10.10), observe first that (10.7) extends to stopping times. So for a stopping time $T \leq 2/3$ on a process X satisfying (10.7),

$$\begin{aligned} E(X_{2/3}^2 | \mathcal{F}_T) &\geq (E(X_{2/3} | \mathcal{F}_T))^2 \\ &= \{ \tfrac{1}{3}(1-T)^{-1} X_T \}^2 \\ &\geq X_T^2/9 , \end{aligned}$$

and so $E X_T^2 \leq 9 EX_{2/3}^2$. Now consider $T_n \leq 2/3$, $\delta_n \downarrow 0$. From (10.7)

$$\begin{aligned} E(X_{T_n+\delta_n}^n - X_{T_n}^n)^2 &= \delta_n^2 E (1-T_n)^{-2} X_{T_n}^2 + \delta_n E(1-\delta_n - T_n)(1-T_n)^{-2}(1-S_{T_n}) \\ &\to 0, \quad \text{using (10.9).} \end{aligned}$$

So Theorem 4.4 shows that (X^n) is tight on $D[0,2/3]$. By symmetry, the process X_t^n has the same distribution as the process $X_{(1-t)-}^n$, and so (X^n) is also tight on $D[1/3,1]$. Using the description of tightness in Lemmas 2.11 and 2.12, we see that (X^n) is tight on $D[0,1]$.

MIXING PROCESSES

As our second worked example we consider ϕ-mixing processes. Let $\xi_1, \xi_2, \ldots\ldots$ be stationary, $E\xi_1 = 0$, $E\,\xi_1^2 < \infty$. Let $\mathcal{M}_a^b = \sigma(\xi_a, \xi_{a+1}, \ldots, \xi_b)$. Recall that (ξ_n) is called ϕ-mixing, for a sequence $\phi = (\phi(n))$, if

(10.11) $\left| P(E_2|E_1) - P(E_1) \right| \leq \phi(b-a)$; all $E_1 \in \mathcal{M}_1^a$, $E_2 \in \mathcal{M}_b^\infty$, $P(E_2) > 0$, $a < b$.

Suppose this holds for some ϕ satisfying

(10.12) $\sum \phi^{\frac{1}{2}}(n) < \infty$

Then [B,p.174] we can define

(10.13) $\sigma^2 = E(\xi_1^2) + 2\sum_{k=2}^{\infty} E(\xi_1 \xi_k)$

where the series is absolutely convergent. Suppose $\sigma > 0$. Let $S_n = \sum_1^n \xi_i$ and define processes X^n in $D[0,\infty)$ by $X_t^n = \sigma^{-1} n^{-\frac{1}{2}} S_{[nt]}$.

(10.14) PROPOSITION [B.20.1] Under the hypotheses above, $X^n \Rightarrow W$.

Proof. Our proof, like that in [B], falls into three parts: establishing

(10.15) $\{ n^{-1} S_n^2 \}$ is uniformly integrable; $E(n^{-1} S_n^2) \to \sigma^2$;

(10.16) (X^n) is tight;

and then deducing the convergence. We refer the reader to [B] for the proof of (10.15), but will use our martingale techniques to give proofs of the other two parts.

Suppose that (10.15) and (10.16) have been established. Fix $L < \infty$. Then

(10.17) $\sup_{t \leq L} |X_t^n - X_{t-}^n| = \sigma^{-1} n^{-\frac{1}{2}} \sup_{i \leq nL} |\xi_i|$

Now $P(n^{-\frac{1}{2}} \sup_{i \leq nL} |\xi_i| > \epsilon) \leq nL.P(|\xi_1| > \epsilon n^{\frac{1}{2}}) \to 0$ because ξ_1 has finite second moment; so

(10.18) $n^{-\frac{1}{2}} \sup_{i \leq nL} |\xi_i| \underset{p}{\to} 0.$

Consider a subsequential weak limit X of (X^n). By (10.17) and (10.18), X is continuous. We want to use to show X is Brownian motion. We need a simple consequence of (10.11) [B,p.171]

(10.19) $|E(\theta\eta) - E\theta . E\eta| \le 2C\phi(b-a)$; $\theta \in \mathcal{M}_1^a$, $|\theta| \le 1, \eta \in \mathcal{M}_b^\infty$, $|\eta| \le C$.

Let $t_1 < t_2 < t_3$ and let $G:D[0,\infty) \to R$ be $\mathcal{F}_{t_1}^D$-measurable, $|G| \le 1$. Then

$$EG(X^n)(X_{t_3}^n - X_{t_2}^n)1(|X_{t_3}^n - X_{t_2}^n| \le C) \quad - EG(X^n).\ E(X_{t_3}^n - X_{t_2}^n)1(|X_{t_3}^n - X_{t_2}^n| \le C)$$

$$\le 2C\ \phi([n(t_2-t_1)]) \quad \text{by (10.19)}$$

$$\to 0.$$

By (10.15), $\{X_{t_i}^n\}$ is uniformly integrable. Letting $C \to \infty$

It remains to prove the tightness assertion (10.16). Let $T_n \leq nL$ be stopping times on $(\xi_1, \xi_2, \ldots\ldots)$ and let d_n be positiove integers, $n^{-1}d_n \to 0$. Theorem 4.4 will give tightness if we can show

$$n^{-\frac{1}{2}}(S_{T_n+d_n} - S_{T_n}) \underset{p}{\to} 0.$$

Now by (10.18) we can find $k_n \to \infty$ so slowly that

$$k_n . n^{-\frac{1}{2}} . \sup_{i \leq n(L+1)} |\xi_i| \underset{p}{\to} 0.$$

We may now assume $d_n > k_n$, and it suffices to prove

(10.20) $$n^{-\frac{1}{2}}(S_{T_n+d_n} - S_{T_n+k_n}) \underset{p}{\to} 0.$$

Consider first $\lambda < \infty$ and integers k,d,t with $t < t+k < t+d$. Let $E \in \mathcal{M}_1^t$, $P(E) > 0$. Then

$$\begin{aligned} P(|S_{t+d} - S_{t+k}| > \lambda \mid E) &\leq \phi(k) + P(|S_{t+d} - S_{t+k}| > \lambda) \quad \text{by (10.11)} \\ &\leq \phi(k) + C\lambda^{-2}d, \end{aligned}$$

where $C = \sup_j j^{-1}ES_j^2 < \infty$ by (10.15). By putting $E = \{T_n = t\}$, $\lambda = \delta n^{\frac{1}{2}}$ we find

$$P(|S_{T_n+d_n} - S_{T_n+k_n}| > \delta n^{\frac{1}{2}}) \leq \phi(k_n) + C\delta^{-2}n^{-1}d_n \to 0$$

which gives (10.20).

Remarks. (a) It was not really necessary to use , since the weaker results concerning asymptotically independent inceements B,19.2 would suffice. We just wanted to display the machinery.

(b) Note that hypothesis (10.12) is used only for (10.13) and (10.15); the rest of the argument requires only that $\phi(n) \to 0$.

Notes on Section 10. Proposition 10.3 is the prototype for results about weak convergence of exchangeable processes, e.g. [B,24.2]. For more recent work see Kallenberg (1973), Eagleson and Weber (1978).

Proposition 10.14 is now an anachronism. Under ϕ-mixing hypotheses it is possible to obtain much better conclusions of the "strong approximation" type - see Berkes and Phillip (1979). And weak convergence can be obtained under less restrictive types of mixing - see Ibragimov (1975). However, the argument for Proposition 10.14 would probably be effective under the "mixingale" type of hypothesis considered in McLeish (1975).

We should admit that there are several other areas of weak convergence theory where our technique does not seem effective, e.g.

empirical distributions

self-similar processes - Taqqu (1979).

CHAPTER 4 - THE PREDICTION PROCESS

11 LIMITATIONS OF WEAK CONVERGENCE

Suppose a sequence of processes is converging weakly to Brownian motion

(11.1) $\quad X^n \Rightarrow W.$

By definition, if $\Phi(X) = E\phi(X)$ for some bounded continuous function ϕ on D, then

(11.2) $\quad \Phi(X^n) \to \Phi(W).$

Does (11.2) remain true for more complicated functionals of processes, such as

(11.3)
$$\Phi_1(X) = \sup\{Eg(X_T) : T \in \mathcal{T}_1\} \qquad (g \in C(R));$$
$$\Phi_2(X) = E \sup_{t \leq 1} P(X_1 > 2 | \mathcal{F}_t) \ ?$$

(Recall our convention from Chapter 3: each process X is associated with some filtration F). Here is a simple example to show that, for functionals like these, (11.2) does not follow from (11.1).

(11.4) <u>Example</u>. Let $f_1, f_2, \ldots\ldots$ be continuous functions on $[0,\infty)$ such that

(11.5) $\quad \{t : f_i(t) = f_j(t)\}$ has measure 0, $i \neq j$.

For each n define a process X^n by picking one of the first n functions at random. Formally, $X^n_t = f_{I_n}(t)$, where I_n is uniform on $\{1,\ldots,n\}$. By (11.5) each process is essentially deterministic, that is

X^n is $\mathcal{F}^n_0$-measurable, where F^n is the usual filtration.

Now it is possible to choose (f_i) such that (11.1) holds; in fact, this is achieved by picking (f_i) independently from the distribution of Brownian motion.(B.Ex.4.4). Despite (11.1), the deterministic processes X^n are really totally unlike Brownian motion so far as probabilistic structure is concerned. For the functionals in (11.3),

$$\Phi_1(X^n) = E \sup_{t\leq 1} g(X^n_t) \to E \sup_{t \leq 1} g(W_t) > \Phi_1(W) \text{ in general;}$$

$\Phi_2(X^n) = P(X_1^n > 2) \rightarrow P(W_1 > 2) < \Phi_2(W)$.

Of course this example is artificial, but that is the point. When we encounter processes satisfying (11.1) in practice we do not expect them to be so pathological, and we can hope to be able to prove convergence of functionals like (11.3). But to make a systematic theory, we must use something stronger than weak convergence.

The rest of this section provides some background and motivation for later definitions, and can be omitted.

Looking at Example (11.4) intuitively, the correct limit of X^n is deterministic Brownian motion, that is W with the filtration G where $\mathcal{G}_0$ contains $\sigma(W)$. With this limit, the functionals (11.3) do converge. So if we want a theory of "convergence in distribution" which can handle these functionals, we first need a notion of "distribution" which distinguishes between deterministic and ordinary Brownian motion. The next examples show that the need for distinctions may arise in more natural settings.

(11.6) Example. Consider the simplest queue model - single server, Poisson input, exponential service times. Let Q_t be the length of the queue at time t, and D_t the number of departures by time t. It is well known (see e.g. Kelly (1979)) that the inter-departure times are independent exponentials, and so D is a Poisson process with respect to its usual filtration. But if G is the usual filtration of (Q,D) then the process (D,G) is not a Poisson process in the sense of Chapter 6, because $D_{t+u} - D_t$ is not independent of $\mathcal{G}_t$ (consider u small and $Q_t = 0$). In studying D it does matter which filtration we use to represent the past.

(11.7) Example. Consider two types of model for automobile accidents. In the first, the accident rate varies between individuals but is constant in time; in the second, the accident rate depends on the individual's history of accidents, but in the same way for each individual. These models are logically distinct, but may produce point processes N_t with the same distribution on D(R) - see Cane (1977). The distinction is seen only when the associated filtrations are described.

Mathematically, we are faced with the following problem: when to regard two processes (X,F) and (X',F') as being essentially the same. For another view of this problem see (DM.IV.4). The examples above show that the function space distribution is too crude a characteristic. As discussed in Chapter 3, in the context of "convergence in distribution" it is unreasonable to suppose all processes are defined on the same filtration, so we cannot use the concept of indistinguishability. One could borrow the idea of isomorphism from ergodic theory, and call (X,F) isomorphic to (X',F') if there is a function θ: such that $\theta(X) = X'$ and θ, θ^{-1} are $\mathcal{F}_t/\mathcal{F}'_t$ -measurable for each t. But this seems rather strong. We propose the following answer.

(11.8) <u>Definition</u>. Call processes (X,F) and (X',F') <u>synonymous</u>, and write $(X,F) \equiv (X',F')$, if

$$\mathcal{L}(Eh_1(X_{u_1}|\mathcal{F}_{t_1}),\ldots\ldots,Eh_k(X_{u_k}|\mathcal{F}_{t_k})) = \mathcal{L}(Eh_1(X'_{u_1}|\mathcal{F}'_{t_1}),\ldots\ldots,Eh_k(X'_{u_k}|\mathcal{F}'_{t_k}))$$

for all $k \geq 1$; all t_i, $u_i \in [0,\infty)$; and all bounded measurable $h_i : S \to R$.

Informally, if an algorithm for assigning numbers to processes is built up from the basic operations of taking measurable functions and conditional expectations, then the algorithm will assign the same number to synonymous processes.

Definition (11.8) is rather unpleasant to work with, and will shortly be replaced by a more elegant and sophisticated reformulation. Meanwhile, note that $(X,F) \equiv (X',F')$ certainly implies $\mathcal{L}(X) = \mathcal{L}(X')$. The following easily proved facts may give some feeling for the definition.

(11.9) LEMMA. Let (X,F) be a process with usual filtration.

(a) Suppose X is also adapted to G. Then $(X,F) \equiv (X,G)$ if and only if for each $t < u$, X_u and $\mathcal{G}_t$ are conditionally independent given $\mathcal{F}_t$; that is,

$$E(h(X_u)|\mathcal{G}_t) = E(h(X_u)|\mathcal{F}_t) \quad \text{for bounded measurable } h.$$

(b) Let (X',F') be another process with usual filtration. Then $(X,F) \equiv (X',F')$ if and only if $\mathcal{L}(X) = \mathcal{L}(X')$.

12 REGULAR CONDITIONAL DISTRIBUTIONS

Before reaching the subject of this chapter, the prediction process, we must digress to discuss regular conditional distributions.

Let V be a random element of the Polish space $(S,\mathcal{S})$, defined on $(\Omega,\mathcal{B},P)$. Let $\mathcal{B}_0$ be a sub σ-field of $\mathcal{B}$. For fixed $A \in \mathcal{S}$ we can define conditional probabilities in terms of conditional expectations:

$$P(V\in A|\mathcal{B}_0) = E(1_A(V)|\mathcal{B}_0).$$

But some care is needed in putting these conditional probabilities together as A varies. See Freedman (1971) for proofs of (12.1) and (12.3).

(12.1) PROPOSITION. <u>There exists a function</u> $\xi(\omega,A)$ <u>such that</u>

(i) <u>for each</u> ω <u>the map</u> $A \to \xi(\omega,A)$ <u>is a probability measure on</u> $(S,\mathcal{S})$;

(ii) <u>for each</u> A <u>the random variable</u> $\omega \to \xi(\omega,A)$ <u>is a version of</u> $P(V\in A|\mathcal{B}_0)$.

Write $\xi(\omega)$ for the probability measure in (i). From (ii), the map $\omega \to \xi(\omega)$ defines a $\mathcal{B}_0$-measurable random element of $\mathcal{P}(S)$. We shall call this random element a <u>regular conditional distribution</u> (r.c.d.) for V given $\mathcal{B}_0$. When $\mathcal{B}_0$ is finite we have an elementary construction:

$$\xi(\omega,A) = P(V\in A|B), \quad \text{where B is the atom of } \mathcal{B}_0 \text{ containing } \omega.$$

The rest of this section describes technical facts about r.c.d.'s. Let h be a real-valued function such that $E|h(V)|<\infty$. If $P(B)>0$ we can compute $E(h(V)|B)$ in two ways. First, by considering the random variable h(V) and taking its conditional expectation given B. Or second, by considering the conditional distribution of V given B, and integrating h against this measure. This suggests:

(12.2) $\int h(s)\,\xi(\omega,ds)$ is a version of $E(h(V)|\mathcal{B}_0)$.

We state below a more general result. Throughout, ξ is a r.c.d. for V given $\mathcal{B}_0$.

(12.3) LEMMA. Let U be a $\mathcal{B}_0$-measurable random element of S'. Let h be a real-valued function on $S \times S'$ such that $E|h(V,U)| < \infty$. Then

$$\int h(s,U(\omega))\ \xi(\omega,ds) \quad \text{is a version of } E(h(V,U)|\mathcal{B}_0).$$

The main use of r.c.d.'s is in computing conditional expectations via (12.2) and (12.3).

When handling r.c.d.'s, one piece of abstract weak convergence theory not discussed in Billingsley (1968) is useful. Call a subset H of $C(S)$ convergence-determining if:

$$\int h\, d\mu_n \to \int h\, d\mu\ ,\ h \in H, \quad \text{implies } \mu_n \to \mu\ .$$

In particular H is separating:

$$\int h\, d\mu = \int h\, d\nu\ ,\ h \in H, \quad \text{implies } \mu = \nu.$$

(12.4) LEMMA (Parthasarathy (1967) Th.6.6) There exists a countable convergence-determining subset H_0 of $C(S)$.

Let H_0 denote such a subset. The first two lemmas below are obvious.

(12.5) LEMMA. Suppose ξ' is a $\mathcal{B}_0$-measurable random element of $\mathcal{P}(S)$. Then ξ' is a r.c.d. for V given $\mathcal{B}_0$ if and only if

$$\int h(s)\ \xi'(\omega,ds) = E(h(V)|\mathcal{B}_0) \quad \text{a.s.},\quad h \in H_0.$$

So if ξ, ξ' are both r.c.d.'s for V given $\mathcal{B}_0$ then $\xi = \xi'$ a.s.

(12.6) LEMMA. Let $g: S \to S'$ be measurable, and let $\tilde{g}: \mathcal{P}(S) \to \mathcal{P}(S')$ be the induced map in the sense of (1.1). Then $\tilde{g}(\xi)$ is a r.c.d. for $g(V)$ given $\mathcal{B}_0$.

(12.7) LEMMA. Let $(\mathcal{B}_k)$ be an increasing family of finite σ-fields whose union generates $\mathcal{B}_0$. Let ξ_k be the (elementary) conditional distribution of V given $\mathcal{B}_k$. Then $\xi_k(\omega) \to \xi(\omega)$ in $\mathcal{P}(S)$ a.s..

Proof. For fixed $h \in C(S)$ the martingale limit theorem says

$$E(h(V)|\mathcal{B}_k) \to E(h(V)|\mathcal{B}_0) \text{ a.s.}$$

So by (12.2),

$$\int h(s)\ \xi_k(\omega,ds) \rightarrow \int h(s)\ \xi(\omega,ds) \qquad \text{a.s.}$$

Apply (12.4).

13 THE PREDICTION PROCESS

Let (X,F) be a process, fixed throughout this section and the next. Consider X as a random element of D(S). For each t there exists a regular conditional distribution Z_t for X given $\mathcal{F}_t$. As discussed in the previous section, we regard Z_t as a random element of $\mathcal{P}(D(S))$. Write Π for the space $\mathcal{P}(D(S))$, and think of Π as the space of all possible distributions of processes. The important fact is that the random elements Z_t, $t \geq 0$, can be put together to form a Skorohod process.

(13.1) THEOREM. There exists a Skorohod process $Z = (Z_t)$, adapted to F, with Z_t taking values in Π, such that Z_t is a regular conditional distribution for X given $\mathcal{F}_t$, for each $t \geq 0$. This prediction process is unique up to indistinguishability. Moreover Z_T is a regular conditional distribution for X given $\mathcal{F}_T$, for each stopping time $T < \infty$.

Now Z can be regarded as a random element of $D(\Pi)$. It turns out that the distribution of the prediction process describes the underlying process up to synonymity (in the sense of (11.8)).

(13.2) LEMMA $(X',F') \equiv (X,F)$ if and only if $\mathcal{L}(Z') = \mathcal{L}(Z)$.

This is the reason for our interest in the prediction process. But we defer this topic until Section 15, after we have looked at the prediction process in its own right.

Proof of Theorem 13.1. By Lemma 12.4 there is a countable separating subset of C(D(S)); call these functions (h_{2i-1}). Since a single random element is tight, there exist compact subsets K_i of D(S) such that $Eh_{2i}(X) \leq 2^{-2i}$, where $h_{2i} = 1_{D(S)\setminus K_i}$. Now define a map $*: \Pi \to R^{\infty}$ by

$$(\mu)^* = \left(\int h_i \, d\mu\right)_{i \geq 1}.$$

Give R^{∞} the product topology.

(13.3) LEMMA. *If* $(\mu_n)^* \to v = (v_i)$ *and if* $\lim_{i\to\infty} v_{2i} = 0$ *then*

(i) $\mu_n \to \mu$ *(say) in* Π ;

(ii) $\int h_{2i-1}\, d\mu = v_{2i-1}$ *for each* i.

Proof. The set $\{\mu_n\}$ is precompact because

$$\lim_{n\to\infty} \mu_n(D(S)\setminus K_i) = \lim_{n\to\infty} \int h_{2i}\, d\mu_n = v_{2i}.$$

Now any weak limit point μ must satisfy (ii). But (h_{2i-1}) is separating, so (ii) determines some unique distribution.

Returning to the construction, define $Z_t' : \Omega \to R^\infty$ by

$$Z_t' = (E(h_i(X)|\mathcal{F}_t))_{i\geq 1},$$

using a Skorohod version of each martingale. For each rational r let Z_r be a regular conditional distribution for X given $\mathcal{F}_r$. Lemma 12.5 implies that, outside some null set Ω_1,

$$(Z_r(\omega))^* = Z_r'(\omega) \quad \text{for each rational } r.$$

Now by the maximal inequality (7.1),

$$P(\sup_t E(h_{2i}(X)|\mathcal{F}_t) > 2^{-i}) \leq 2^i\, Eh\ (X)$$
$$\leq 2^{-i}.$$

The Borel-Cantelli lemma implies that, outside some null set Ω_2,

$$\lim_{i\to\infty} \sup_t\ (Z_t'(\omega))_{2i} = 0.$$

Now let $r \to t\pm$. Then $Z_r'(\omega) \to Z_{t\pm}'(\omega)$ for each ω. Applying Lemma 13.3 we see that for ω outside $\Omega_1 \cup \Omega_2$,

(a) $Z_r(\omega) \to Z_{t\pm}(\omega)$, say;

(b) $\int h_{2i-1}\, dZ_t(\omega,\cdot) = (Z_t'(\omega))_{2i-1}$.

For fixed t, (a) and the "usual conditions" on F imply that Z_t is $\mathcal{F}_t$-measurable. By definition $(Z_t')_{2i-1}$ is a version of $E(h_{2i-1}(X)|\mathcal{F}_t)$, so (b) and Lemma 12.5

show that Z_t is a regular conditional distribution for X given $\mathcal{F}_t$. And by redefining Z on $\Omega_1 \cup \Omega_2$ we obtain a Skorohod process.

The uniqueness of the prediction process follows from the uniqueness of r.c.d.'s (Lemma 12.5) and the Skorohod property. It remains to prove that for stopping times T,

Z_T is a r.c.d. for X given $\mathcal{F}_T$.

For discrete T this follows from the case T = t. In general, approximate T from above by discrete stopping times (T_n) as in Lemma 6.6. Then, arguing as in Lemma 12.7, for $h \in C(D(S))$

$$\int h(s)\, Z_{T_n}(\omega,ds) = E(h(X)|\mathcal{F}_{T_n}) \to E(h(X)|\mathcal{F}_T) \quad \text{a.s. (martingale convergence)}$$
$$\to \int h(s)\, Z_T(\omega,ds) \quad \text{a.s. (Z is Skorohod)}$$

and we can apply Lemma 12.5 again.

Theorem 13.1 gives no insight into the form of the prediction process. The simplest case, unsurprisingly, is when X is Markov, so let us briefly discuss this case. Write ${}^tX$ for the post-t process ${}^tX_u = X_{t+u}$. The Markov property for (X,F) is equivalent to the assertion: there exists a function $\phi:[0,\infty) \times S \to \Pi$ such that for each t,

(13.4) $\quad \phi(t,X_t)$ is a regular conditional distribution for ${}^tX$ given $\mathcal{F}_t$.

Of course we think of $\phi(t,s)$ as the distribution of the Markov process started from position s at time t. So what is Z_t? Informally, given $\mathcal{F}_t$ we know the sample path $X(\omega)$ over time $[0,t]$, and we know the post-t process has conditional distribution $\phi(t,X_t(\omega))$. This suggests putting

(13.5) $\quad Z_t'(\omega) = \theta(X(\omega),t,\phi(t,X_t(\omega))),$

where $\theta: D(S) \times [0,\infty) \times \Pi \to \Pi$ is defined by:

(13.6) $\quad \theta(f,t,\mu) = \mathcal{L}(Y)$, where $Y_u = f(u), \ u < t;$

$$\mathcal{L}({}^tY) = \mu .$$

To prove Z' is indeed the prediction process we must show

(a) Z'_t is a r.c.d. for X given $\mathcal{F}_t$;

(b) Z'is Skorohod.

The proof of (a) from (13.4) is an exercise in manipulating r.v.d.'s. To prove (b) we need a regularity condition.

(13.7) Definition. Call X a Feller process if the map ϕ in (13.4) may be chosen to be jointly continuous.

(13.8) PROPOSITION If (X,F) is a Feller process then (13.5) defines the prediction process. Moreover

$$\{t: Z_t(\omega) \neq Z_{t-}(\omega)\} \subset \{t: X_t(\omega) \neq X_{t-}(\omega)\} \quad \text{a.s.}$$

The proof will be given in Section

(13.9) Remarks. The idea of the prediction process comes from Knight (1975), where it is used for quite different purposes. Knight considers a process X assumed only to satisfy measurability conditions, so X can be regarded as a random element of the space M(S) of measurable functions $[0,\infty) \to S$ with the topology of convergence in measure. Then Z_t becomes a random element of $\mathcal{P}(M(S))$. Knight shows that Z can be made Skorohod even though X is not, and this is his motivation in studying Z: "its potential utility derives from the fact that it is possible to impose a greater degree of regularity in an auxiliary space than one would be justified in assuming a priori". This is not relevant in our setting, where X is assumed Skorohod.

14 PROPERTIES OF THE PREDICTION PROCESS

Here is some useful notation. For $s \in S$ let $\delta_s \in \mathcal{P}(S)$ be the distribution degenerate at s: $\delta_s(A) = 1_{(s \in A)}$. For each $t \geq 0$ let $\pi_t : \Pi \to \mathcal{P}(S)$ be the map induced in the sense of (1.1) from the evaluation map $f \to f(t)$, so $\pi_t(\mathcal{L}(Y)) = \mathcal{L}(Y_t)$. Similarly, $\pi_{t-}(\mathcal{L}(Y)) = \mathcal{L}(Y_{t-})$.

Let (X,F) be a process with prediction process Z. Intuitively, for fixed t, ω the distribution $Z_t(\omega)$ describes a process which evolves deterministically as $X(\omega)$ over time $[0,t]$. This idea is formalised in Lemma 14.2 below.

By definition (12.1), if V is $\mathcal{B}_0$-measurable then δ_V is a regular conditional distribution for V given $\mathcal{B}_0$. In particular for $r \leq t$,

δ_{X_r} is a r.c.d. for X_r given $\mathcal{F}_t$.

But by Lemma 12.6,

$\pi_r Z_t$ is a r.c.d. for X_r given $\mathcal{F}_t$.

So for each fixed pair $r \leq t$,

$$(14.1) \quad \pi_r Z_t = \delta_{X_r} \quad \text{a.s.}$$

But more is true.

(14.2) LEMMA Outside a null set,

$$(14.3) \quad \pi_r Z_t(\omega) = \delta_{X_r(\omega)} \quad \text{for all } 0 \leq r \leq t < \infty .$$

<u>Proof</u>. Certainly (14.3) holds for all rational pairs $r \leq t$, outside a null set. Fix ω outside that set. First observe that for $\mu \in \Pi$ the map $r \to \pi_r(\mu)$ is Skorohod. So by approximating a real r from above by rationals, (14.3) holds for $\{(r,t)$: r real, t rational, $r \leq t\}$. Next recall that $\mu_n \to \mu$ implies $\pi_r(\mu_n) \to \pi_r(\mu)$ provided r is a continuity point of μ. So by approximating t from above, (14.3) extends to $\{(r,t)$: r,t real, $r \leq t$, r is a continuity point of $Z_t(\omega)\}$. But continuity points are dense, so repeating the first argument shows that (14.3) holds for all real pairs $r < t$.

To complete the proof we must show

$P(\pi_t Z_t = \delta_{X_t} \text{ for all } t) = 1.$

Both processes $\pi_t Z_t$ and δ_{X_t} are measurable functions of Skorohod processes, and so are optional; by Lemma 6.12 it suffices to prove

(14.4) $\pi_T Z_T = \delta_{X_T}$ a.s.

for each stopping time T. By Theorem 13.1

(14.5) Z_T is a r.c.d. for X given $\mathcal{F}_T$.

We could prove (14.4) in the way we proved (14.1), but that requires a more general form of Lemma 12.6. Here is a different argument. Let d be a bounded metric on S, and put

$$h(f,t,s) = d(f(t),s)$$

$$h^*(\mu,t,s) = \int h(f,t,s)\ \mu(df).$$

Since T and X_T are $\mathcal{F}_T$-measurable, (14.4) and Lemma 12.3 show

$$h^*(Z_T,T,X_T) = E(h(X,T,X_T) \mid \mathcal{F}_T) \quad \text{a.s.}$$

$$= 0 \quad \text{a.s.}$$

But $h^*(\mu,t,s) = 0$ implies $\pi_t(\mu) = \delta_s$, so (14.4) follows.

In view of Lemma 14.2 we can and will choose Z so that (14.3) holds for all ω. Recall the idea behind (14.3): that $\pi_r Z_t$ is the conditional distribution of X_r given $\mathcal{F}_t$, and for $r \le t$ X_r is determined by $\mathcal{F}_t$. Similarly we may regard (for example) $\pi_{r-}Z_{t-}$ as the conditional distribution of X_{r-} given events strictly before t. The same intuitive idea suggests

(14.6) $\pi_{r-}Z_t(\omega) = \pi_{r-}Z_{t-}(\omega) = \delta_{X_{r-}(\omega)}$ for all ω, all $r \le t$;

(14.7) $\pi_r Z_{t-}(\omega) = \delta_{X_{r-}(\omega)}$ for all ω, all $r < t$.

These identities follow from (14.3) by easy continuity arguments.

Our results about the prediction process so far have been dull technicalities. Here is a more interesting idea. Consider the process $\pi_t Z_{t-}$ and interpret this as the conditional distribution of X_t given events strictly before t. Now the intuitive idea of a predictable process is that X_t should be determined by the events strictly before t, so

$$\pi_t Z_{t-} = \delta_{X_t} .$$

And the intuitive idea of quasi left continuity is that discontinuities are unpredictable, so that given events strictly before t we predict $X_t = X_{t-}$, so

$$\pi_t Z_{t-} = \delta_{X_{t-}} .$$

It turns out that these intuitive ideas are correct. The result below will be proved in Section

(14.8) PROPOSITION (i) X <u>is predictable if and only if</u> $P(\pi_t Z_{t-} = \delta_{X_t}$ <u>for all</u> $t) = 1$.

(ii) X <u>is quasi left continuous if and only if</u> $P(\pi_t Z_{t-} = \delta_{X_{t-}}$ <u>for all</u> $t) = 1$.

In particular, predictability and quasi left continuity for a process are sample-path properties of the prediction process.

(14.9) <u>Remark</u>. We could define $Z_\infty(\omega) = \delta_{X(\omega)}$, and then Z_T is a r.c.d. for X given $\mathcal{F}_T$ for all stopping times $T \leq \infty$. This leads to an alternative construction of Z. Start with Z_∞ as above, and define Z_t as the distribution-valued martingale $E(Z_\infty | \mathcal{F}_t)$. This approach seems to be envisaged in Schwartz (197). Actually the difference in approaches is only superficial: the proof that a distribution-valued martingale has a Skorohod version is essentially the same as the proof of Theorem 13.1.

Here are some more technicalities.

(14.10) LEMMA <u>For bounded measurable</u> $\phi : D(S) \to R$ <u>define</u> $\phi^* : \Pi \to R$ <u>by</u> $\phi^*(\mu) = \int \phi \, d\mu$. <u>Then</u>

(a) $\phi^*(Z_t) = E(\phi(X) | \mathcal{F}_t)$ a.s., <u>for each</u> t.

(b) $E|\phi_1^*(Z_t) - \phi_2^*(Z_t)|^2 \le E|\phi_1(X) - \phi_2(X)|^2$.

(c) <u>if</u> $\phi \in C(D(S))$ <u>then</u> $\phi^* \in C(\Pi)$ <u>and</u> $\phi^*(Z_t)$ <u>is a Skorohod version of</u> $E(\phi(X)|\mathcal{F}_t)$.

<u>Proof</u>. (a) is a special case of (12.2), and (b) follows because for any variable $E|E(V|\mathcal{G})|^2 \le E V^2$. The continuous mapping theorem and the fact that a continuous function of a Skorohod function is Skorohod establish (c).

(14.11) <u>Remarks</u>. A more delicate argument shows that even for non-continuous ϕ the sample paths of $\phi^*(Z_t)$ are a.s. Skorohod. In particular, for measurable $A \subset D(S)$ the process $Z_t(\omega, A)$ is an a.s. Skorohod version of $P(X \in A | \mathcal{F}_t)$. However, there is a limit to the amount of "tidying up" of Z which is possible. It is <u>not</u> true that, outside a null set,

the map $t \to Z_t(\omega, A)$ is Skorohod for all A.

For example, let X be Brownian motion and let $A_y = \{f : f(y) = 0, f > 0$ over some interval $(y, y+\varepsilon)\}$. Then

$$Z_t(\omega, A_y) = 1_{(X(\omega) \in A_y)} \cdot 1_{(t > y)}$$

and the sample paths are left-continuous.

The reader may have wondered why we did not define the prediction process using only the future in the conditional distribution, i.e. why not

(14.12) $\hat{Z}_t$ is a r.c.d. for ${}^tX$ given $\mathcal{F}_t$,

where ${}^tX$ is the post-t process $X_{t+u} = ({}^tX)_u$. There are no a priori grounds for preferring Z or $\hat{Z}$, since either can be derived from the other. But it turns out that the analogue of Theorem 13.1 for $\hat{Z}$ is not true in general.

(14.13) PROPOSITION. The following are equivalent.

(i) There exists a Skorohod process $\hat{Z}$ adapted to F such that $\hat{Z}_T$ is a r.c.d. for ${}^T X$ given $\mathcal{F}_T$, for each finite stopping time T.

(ii) X is quasi left continuous.

Remark. Knight (1975) uses $\hat{Z}$ and gets a Skorohod process in general, but this is because he uses a weaker topology on function space - see Remark 13.9.

Proposition 14.13 is not used in the sequel. We give the proof because it provides a nice introduction to the techniques used in handling the prediction process. Recall from Section 3 the shift map $\sigma_t : D(S) \to D(S)$,

$$\sigma_t(f)(u) = f(0 \vee (t+u)).$$

Lemma 3. extends to the following result for processes, by the Skorohod representation theorem.

(14.14) LEMMA Suppose $t_n \to t$ and $Y^n \Rightarrow Y$ on $D(S)$. Either of the conditions below is sufficient to show $\sigma_{t_n}(Y^n) \Rightarrow \sigma_t(Y)$.

(a) $t \le 0$.

(b) t is a continuity point of Y.

(c) for each n, $t_n \ge t$ and $Y^n_t = Y_t = s_0$ a.s. for some $s_0 \in S$.

Proof of Proposition 14.13. Suppose X is quasi left continuous. Let Z be the prediction process. Put $\hat{Z}_t = \tilde{\sigma}_t(Z_t)$, where $\tilde{\sigma}_t$ is the map from Π to Π induced by σ_t. By Lemma 12.6,

(14.15) $\hat{Z}_t$ is a r.c.d. for $\sigma_t(X) = {}^t X$ given $\mathcal{F}_t$.

We shall show $\hat{Z}$ is Skorohod; the extension of (14.15) to stopping times is then straightforward. Fix ω. By Proposition 14.8 we may assume

(14.16) $\pi_t Z_{t-}(\omega) = \delta_{X_{t-}(\omega)}$ for all t.

For each t let Y^t, Y^{t-} denote processes with distributions $Z_t(\omega)$, $Z_{t-}(\omega)$.

Because Z is Skorohod,

(14.17) $\quad t_n \to t\pm$ implies $Y^{t_n} \Rightarrow Y^{t\pm}$.

It suffices to prove

(14.18) $\quad t_n \to t\pm$ implies $\sigma_{t_n}(Y^{t_n}) \Rightarrow \sigma_t(Y^{t\pm})$.

But this follows from Lemma 14.14. For when $t_n \downarrow t$ then $(Y^{t_n})_t = (Y^t)_t = X_t(\omega)$ by (14.3), so condition (c) holds. And when $t_n \uparrow\uparrow t$ then (b) holds: for t is a continuity point of Y^{t-} because

$$\pi_t Y^{t-} = \delta_{X_{t-}(\omega)} \quad \text{by (14.16)}$$

$$= \pi_{t-} Y^{t-} \quad \text{by (14.6).}$$

Conversely, suppose there is a Skorohod process $\hat{Z}$ as in part (i) of the Proposition. For $\delta > 0$ define

$$\phi_\delta(f) = \sup_{u \le \delta} d(f(u), f(0)) ;$$

where d is a bounded metric on S. It is easy to check

$\phi_\delta(f) \to 0$ as $\delta \to 0$;

if $f_n \to f$ and δ is a continuity point of f then $\phi_\delta(f_n) \to \phi_\delta(f)$.

Define ϕ^*_δ as in Lemma 14.10. Then for μ_n, μ in Π,

(14.19) $\quad \phi^*_\delta(\mu) \to 0$ as $\delta \to 0$;

(14.20) $\quad$ if $\mu_n \to \mu$ and δ is a continuity point of μ then $\phi^*_\delta(\mu_n) \to \phi^*_\delta(\mu)$.

Now consider a predictable stopping time T $(0 < T < \infty)$, and suppose (S_n) is a sequence announcing T. Then $\hat{Z}_{S_n} \to \hat{Z}_{T-}$ a.s.. So for δ outside some countable set Δ we have by (14.20)

$$\phi^*_\delta(\hat{Z}_{S_n}) \to \phi^*_\delta(\hat{Z}_{T-}) \quad \text{a.s..}$$

By hypothesis $\hat{Z}_{S_n}$ is a r.c.d. for $\sigma_{S_n}(X)$ given $\mathcal{F}_{S_n}$. So

$$\phi^*_\delta(\hat{Z}_{S_n}) = E(\phi_\delta(\sigma_{S_n}(X)) \mid \mathcal{F}_{S_n}) \qquad \text{by (12.3)}$$

$$= E(\sup_{u \le \delta} d(X_{S_n+u}, X_{S_n}) \mid \mathcal{F}_{S_n}).$$

So

$$E \sup_{u \le \delta} d(X_{S_n+u}, X_{S_n}) \to E\,\phi^*_\delta(Z_{T-}), \quad \delta \notin \Delta.$$

But $\sup_{u \le \delta} d(X_{S_n+u}, X_{S_n}) \ge \frac{1}{2} d(X_T, X_{T-})$ on the set $\{T > S_n > T-\delta\}$, and so letting $n \to \infty$ we see

$$E\,\phi^*_\delta(Z_{T-}) \ge \tfrac{1}{2} E d(X_T, X_{T-}), \quad \delta \notin \Delta.$$

Letting $\delta \to 0$, (14.19) implies $X_T = X_{T-}$ a.s., so X is quasi left continuous.

Finally we give a proof of Lemma 13.2.

(14.21) LEMMA. Let V, X be random elements of D(S), and let Y, Z be random elements of $D(\Pi)$. Suppose

(14.22) $\mathcal{L}(V, \phi_1^*(Y_{t_1}), \ldots, \phi_k^*(Y_{t_k})) = \mathcal{L}(X, \phi_1^*(Z_{t_1}), \ldots, \phi_k^*(Z_{t_k}))$

for all $(\phi_i) \in C(D(S))$ and all $(t_i) \in [0,\infty)\setminus T_0$, where T_0 is countable. Then $\mathcal{L}(V,Y) = \mathcal{L}(X,Z)$.

Proof. Fix $(t_i) \in [0,\infty)\setminus T_0$. It suffices to show that the finite-dimensional distributions $\mathcal{L}(V, Y_{t_1}, \ldots, Y_{t_k})$ and $\mathcal{L}(X, Z_{t_1}, \ldots, Z_{t_k})$ coincide. By hypothesis, these distributions agree on subsets of $D(S) \times \Pi^k$ of the form

$$G \times \bigcap_{1\le i\le j_1} \{\lambda : \phi_{1,i}^*(\lambda) < a_{1,i}\} \times \cdots \times \bigcap_{1\le i\le j_k} \{\lambda : \phi_{k,i}^*(\lambda) < a_{k,i}\},$$

where G is open in D(S). But these form a base of open sets for the topology on $D(S) \times \Pi^k$, and hence generate the Borel σ-field.

Proof of Lemma 13.2. We must check that (14.22) holds for $(V,Y) = (X',Z')$. But definition 11.8 says that (14.22) holds for ϕ_i of the form $h_i \circ \pi_{u_i}$, $h_i \in C(S)$. The extension to general ϕ_i is routine, since $\{\pi_u : u \ge 0\}$ generate the Borel σ-field on D(S).

CHAPTER 5 - EXTENDED WEAK CONVERGENCE

15 EXTENDED WEAK CONVERGENCE

Given a sequence (X^n, F^n) of processes, consider the mode of convergence defined by weak convergence of their prediction processes

(15.1) $\quad Z^n \Rightarrow Z^\infty.$

In this section we discuss why this seems a sensible mode of convergence to consider: the rest of the book develops the mathematics.

The fundamental idea underlying weak convergence theory is that a process can be essentially described by its function space distribution. For many purposes this is true. But as we saw in Section 11, there are natural examples where the function space distribution fails to distinguish between essentially different processes. When we use mode (15.1) of convergence, we are assuming only that processes are described by the distribution of their prediction processes. In other words (Lemma 13.2) we regard two processes as being essentially the same if and only if they are synonymous in the sense of (11.8). This is the viewpoint we shall adopt. Of course we do not assert that synonymous processes are similar in all respects.

We saw in Section 14 that certain structural properties of X (quasi left continuity; predictability) are sample-path properties of Z, though they are plainly not sample-path properties of X in general. This is true for other properties. For example, it is easy to check that X is a martingale if and only if outside some null set

$$e\,\pi_r Z_t(\omega) = X_t(\omega) \quad , \quad \text{all } r \geq t,$$

where $e : \mathcal{P}(R) \rightarrow R$ is the expectation map. Now when (15.1) holds, the

Skorohod representation says that the sample-paths of Z^n and Z^∞ can be nearly matched; this suggests that X^n and X^∞ will have nearly matched structural properties. This "structure-preserving" feature of (15.1) is one of its main advantages: we saw in Example 11.4 that weak convergence does not have this feature.

For technical reasons we use a slight variant of (15.1) for the formal definition.

(15.2) Definition. Let (X^n, F^n) be processes with prediction processes Z^n. Write $(X^n, F^n) \overset{e}{\Rightarrow} (X^\infty, F^\infty)$ if $(X^n, Z^n) \Rightarrow (X^\infty, Z^\infty)$ on $D(S \times \Pi)$. Call this extended weak convergence.

Here we are regardeng (X,Z) as the $S \times \Pi$ -valued process $t \to (X_t, Z_t)$. Usually we shall suppress the filtrations and just write $X^n \overset{e}{\Rightarrow} X$.

In the rest of this chapter we develop properties of extended weak convergence. The question of how in practice to establish extended weak convergence is deferred to Chapter 6.

Remark. The concept of extended weak convergence is new, although independently Helland (1980) has described a closely related form of convergence - see

16 TECHNICAL PROPERTIES

Most of these technical lemmas do not require much comment. The reader will recognise several analogues of standard weak convergence results.

(16.1) LEMMA. *Let* $(X^n,F^n) \Rrightarrow (X^\infty,F^\infty)$.

(a) *Suppose* $h:S \to S'$ *is continuous. Put* $Y^n_t = h(X^n_t)$. *Then* $(Y^n,F^n) \Rrightarrow (Y^\infty,F^\infty)$.

(b) *Suppose* $H:D(S) \to D(S')$ *is a continuous mapping such that*:

if $f(u) = g(u)$ *for* $u \leq t$ *then* $(Hf)(t) = (Hg)(t)$.

Put $Y^n = H(X^n)$. *Then* $(Y^n,F^n) \Rrightarrow (Y\ ,F\ )$.

Proof. Plainly (a) is a special case of (b). To prove (b), note first that Y^n is indeed adapted to F^n. Let $\tilde{H}:\Pi \to \Pi'$ be the map induced in the sense of (1.1). By Lemma 12.6, $\tilde{H}(Z^n_t)$ is a regular conditional distribution for Y^n given $\mathcal{F}^n_t$. Since $\tilde{H}$ is continuous, the process $\tilde{H}(Z^n_t)$ is the prediction process of Y^n. By hypothesis $(X^n,Z^n) \Rightarrow (X^\infty,Z^\infty)$ on $D(S\times\Pi)$, so by the continuous mapping theorem $(\tilde{H}(X^n),\tilde{H}(Z^n)) \Rightarrow (H(X^\infty),\tilde{H}(Z^\infty))$ on $D(S'\times\Pi')$.

(16.2) LEMMA. *Let* $X^n \Rrightarrow X$. *Let* u *be a continuity point of the prediction process* Z *of* X. *Let* $\phi:D(S) \to R$ *be a measurable function such that*

(a) $(\phi(X^n))$ *is uniformly integrable*;

(b) $P(X\in C_\phi) = 1$, *where* $C_\phi = \{f: \phi$ *is continuous at* $f\}$.

Then $E(\phi(X^n)|\mathcal{F}^n_u) \Rightarrow E(\phi(X)|\mathcal{F}_u)$.

Proof. By truncating, we may assume ϕ is bounded. Define $\phi^*:\Pi \to R$ as at (14.10). We must prove

(16.3) $\quad \phi^*(Z^n_u) \Rightarrow \phi^*(Z_u)$.

By hypothesis $Z^n \Rightarrow Z$, and since u is a continuity point we have

(16.4) $\quad Z^n_u \Rightarrow Z_u$.

Now (B.5.2) says that ϕ^* is continuous at distributions λ such that $\lambda(C_\phi) = 1$. And $Z_u(\omega,C_\phi) = P(X\in C_\phi|\mathcal{F}_u) = 1$ a.s. by (b). So

$P(\omega\colon \phi^* \text{ is continuous at } Z_u(\omega)\ ) = 1.$

Now (16.4) implies (16.3), using (B.5.2) again.

We state for later use a multidimensional version of Lemma 16.2. The proof is the same.

(16.5) LEMMA. Let $X^n \Rightarrow X$. For each $i \geq 1$ let u_i, ϕ_i be as in Lemma 16.2. Then $(E(\phi_1(X^n)|\mathcal{F}^n_{u_1}), E(\phi_2(X^n)|\mathcal{F}^n_{u_2}), \ldots\ldots\ldots) \Rightarrow$

$$(E(\phi_1(X)|\mathcal{F}_{u_1}), E(\phi_2(X)|\mathcal{F}_{u_2}), \ldots\ldots\ldots) \text{ on } R^\infty.$$

(16.6) LEMMA. For a sequence (X^n, F^n) of processes, the following are equivalent.

(a) (X^n) is tight on $D(S)$.

(b) For each $\varepsilon > 0$ there exists a compact subset $\mathcal{M}$ of Π such that

$P(Z^n_t \in \mathcal{M} \text{ for all } t) \geq 1 - \epsilon\ ;\ n \geq 1.$

(c) (Z^n_0) is tight on Π.

Proof. Suppose (a) is true. Fix ε. Choose compact subsets K_j of $D(S)$ such that $P(X^n \in K^c_j) \leq \varepsilon 2^{-2j}$ for each j,n. Put

$$\mathcal{M} = \{\lambda : \lambda(K^c_j) \leq 2^{-j} \text{ for all } j\}.$$

Then $\mathcal{M}$ is compact. So

$$\begin{aligned} P(Z^n_t \in \mathcal{M} \text{ for all } t) &= P(Z^n_t \in \mathcal{M} \text{ for rational } t) \\ &\ 1 - \sum_j P(Z^n_t(\omega, K^c_j) > 2^{-j} \text{ for some rational } t) \\ &= 1 - \sum_j P(\ P(X^n \in K^c_j | \mathcal{F}^n_t) > 2^{-j} \text{ for some rational } t) \\ &\ 1 - \sum \varepsilon 2^{-j} = 1 - \varepsilon \end{aligned}$$

by the martingale maximal inequality (7.1).

Plainly (b) implies (c). Now suppose (c) is true. Given $\varepsilon > 0$ there exists a compact subset $\mathcal{M}$ of Π such that $P(Z^n_0 \in \mathcal{M}\ ) \geq 1 - \epsilon\ ,\ n \geq 1.$

Now there exists a compact subset K of D(S) such that $\lambda(K^c) \leq \varepsilon$, $\lambda \in \mathcal{M}$. So $P(Z_0^n(\omega,K^c) > \varepsilon) \leq \varepsilon$. But $Z_0^n(\omega,K^c) = P(X^n \in K^c | \mathcal{F}_0^n)$, and so $P(X^n \in K^c) \leq 2\varepsilon$.

(16.7) LEMMA. <u>Let</u> (X^n,F^n) <u>be a sequence of processes such that</u> $Z^n \Rightarrow Z^\infty$. <u>Then</u> $(X^n,Z^n) \Rightarrow (X^\infty,Z^\infty)$ <u>on</u> $D(S) \times D(\Pi)$.

<u>Remark</u>. But not necessarily on $D(S\times\Pi)$. In other words, (15.1) is strictly weaker than extended weak convergence - see

<u>Proof</u>. Using the Skorohod representation theorem, suppose $Z^n \to Z^\infty$ a.s.; it suffices to prove

(16.8) $\quad X^n \to X^\infty$ a.s..

Let L be a continuity point of the process Z^∞. Then

(16.9) $\quad Z_L^n \to Z_L^\infty$ a.s..

Fix ω in the set where (16.9) holds. Let $t_0 < L$ be a continuity point of the distribution $Z_L(\omega)$. By (14.3) the distribution $Z_L^n(\omega)$ coincides, on the interval $[0,t_0]$, with the degenerate distribution $\delta_{X^n(\omega)}$. Now (16.9) implies $X^n(\omega) \to X^\infty(\omega)$ in $D([0,t_0])$. Since L and t_0 may be arbitrarily large, we deduce (16.8).

The next sequence of lemmas lead up to a technique (16.) for proving extended weak convergence. For a space S and a subset H of C(S), consider the property:

(16.10) if $s_1 \neq s_2$ and $s_2 \neq s_3$ then there exists $h \in H$ such that $h(s_1) \neq h(s_2)$ and $h(s_2) \neq h(s_3)$.

(16.11) LEMMA. (a) <u>There exists a countable subset</u> H <u>of</u> C(S) <u>satisfying</u> (16.10).

(b) <u>Moreover if</u> $S = \mathcal{P}(S')$ <u>then we may take</u> $H = \{\phi^* : \phi \in H'\}$ <u>for some countable subset</u> H' <u>of</u> C(S'), <u>where</u> $\phi^*(\lambda) = \int \phi(s')\, d\lambda$.

<u>Proof</u>. Since S is separable there is a countable subset H_0 of C(S) which separates points, that is:

if $s_1 \neq s_2$ then there exists $h \in H_0$ such that $h(s_1) \neq h(s_2)$.

Then $H = H_0 \cup \{h_1+h_2 : h_i \in H_0\}$ satisfies (16.10), proving (a). To prove (b) let $H_0' \subset C(S')$ be as in Lemma 12.4, and construct H' similarly.

(16.12) LEMMA. <u>Let</u> (X^n) <u>be a sequence of processes in</u> D(S). <u>Suppose</u>

(a) <u>for each</u> $L < \infty$, $\varepsilon > 0$ <u>there exists a compact subset</u> K <u>of</u> S <u>such that</u>

$$\liminf_n P(X_t^n \in K \text{ for all } t \leq L) \geq 1 - \varepsilon ;$$

(b) <u>the set of processes</u> $\{h(X_t^n) : n \geq 1\}$ <u>is tight on</u> D(R), <u>for each</u> h <u>in some countable set satisfying</u> (16.10).

<u>Then</u> (X^n) <u>is tight on</u> D(S).

<u>Proof</u>. Suppose $H = (h_i)$ satisfies (16.10). Let A be a subset of D(S) such that

(a') $\{f(u) : u \leq L, f \in A\}$ is precompact, for each L;

(b') there exist compact subsets K_i of D(R) such that $h_i \circ f \in K_i$ for each $i \geq 1$ and each $f \in A$.

We shall prove A is precompact; the lemma follows, in the same way that the more familiar tightness conditions (B.15.2) follows from conditions for compactness in D(R).

(16.13) PROPOSITION. <u>Let</u> (X^n,F^n), $1 \leq n \leq \infty$, <u>be processes, and suppose</u> Z^∞ <u>is continuous. If</u> $(Z^n_{t_1},\ldots,Z^n_{t_k}) \Rightarrow (Z^\infty_{t_1},\ldots,Z^\infty_{t_k})$ <u>for all</u> $(t_1,\ldots,t_k)$, <u>then</u> $(X^n,F^n) \Rrightarrow (X^\infty,F^\infty)$.

<u>Proof</u>. We shall prove (Z^n) is tight; then $Z^n \Rightarrow Z^\infty$, and the result follows from

We shall show that (Z^n) satisfies the hypotheses of Lemma 16.12. Hypothesis (a) follows from Lemma 16.6. To check (b) it suffices, by Lemma 16.11(b), to show that $\{\phi^*(Z^n_t): n\geq 1\}$ is tight on D(R), for each $\phi \in C(D(S))$. But Lemma 14.10 says that $\phi^*(Z^n_t)$ is the bounded martingale $E(\phi(X)|\mathcal{F}^n_t)$, and so $\phi^*(Z^n_t) \Rightarrow \phi^*(Z_t)$ on D(R) by hypothesis and Proposition 5.3.

(16.14) <u>Remark</u>. This gives a useful technique for establishing extended weak convergence. Suppose $X^n \Rightarrow X$, and suppose X and Z are continuous. By the Skorohod representation, we may assume $X^n \to X$ a.s.. So $X^n_t \to X_t$ a.s. for each t. Suppose we can prove $Z^n_t \to Z_t$ a.s. for each t. Then $(Z^n_{t_1},\ldots,Z^n_{t_k}) \Rightarrow (Z_{t_1},\ldots,Z_{t_k})$, and so $X^n \Rrightarrow X$ by Proposition 16.13. Thus to improve weak convergence to extended weak convergence, when X and Z are continuous, we need only consider the behaviour of the prediction processes at a fixed time t.

In the spirit of Lemma 13.2 there is a reformulation of extended weak convergence which is more elementary, in that it does not explicitly use prediction processes.

(16.15) PROPOSITION $X^n \Rrightarrow X^\infty$ <u>if and only if, for all</u> $\phi_1,\ldots,\phi_k$ <u>in</u> C(D(S)),

$$(X^n_t,\ E(\phi_1(X^n)|\mathcal{F}^n_t),\ldots\ldots,\ E(\phi_k(X^n)|\mathcal{F}^n_t)\ )_{t\geq 0} \Rightarrow$$

$$(X^\infty_t,\ E(\phi_1(X^\infty)|\mathcal{F}^\infty_t),\ldots\ldots,\ E(\phi_k(X^\infty)|\mathcal{F}^\infty_t)\ )_{t\geq 0} \text{ on } D(S\times R^k).$$

Remark. In this form, extended weak convergence appears closely related to the type of convergence discussed in Helland (1980).

Though the prediction process form is usually more tractable, (16.15) is occasionally useful; for example, to prove the following analogue of a standard weak convergence result (B.4.1).

(16.16) LEMMA. *Suppose* $(X^n, F^n) \Rrightarrow (X,F)$. *Suppose* (Y^n) *is a sequence of processes such that*

(a) Y^n *is adapted to* F^n;

(b) $\sup_{t \leq L} d(X^n_t, Y^n_t) \to 0$ *in probability*, $L < \infty$.

Then $(Y^n, F^n) \Rrightarrow (X,F)$.

Proof. By (b) and (B.4.1), $(Y^n, X^n) \Rightarrow (X,X)$, and so $\phi(Y^n) - \phi(X^n) \to 0$ in probability, for $\phi \in C(D(S))$. So by (7.1),

$$\sup_t \left| E(\phi(Y^n) \mid \mathcal{F}^n_t) - E(\phi(X^n) \mid \mathcal{F}^n_t) \right| \to 0 \quad \text{in probability.}$$

Now use Proposition 16.15 and (B.4.1).

Proof of Proposition 16.15. Consider the continuous map $\Phi : S \times \Pi \to S \times R^k$ defined by

$$(s, \mu) \to (s, \phi^*_1(\mu), \ldots\ldots, \phi^*_k(\mu)).$$

Let $\Lambda : D(S \times \Pi) \to D(S \times R^k)$ be the map derived from Φ in the obvious way. Applying the continuous mapping theorem to Λ, we find:

(16.17) if $(X^n, Z^n) \Rightarrow (V,Y)$, say, on $D(S \times \Pi)$ then

$$(X^n_t, E(\phi_1(X^n) \mid \mathcal{F}^n_t), \ldots\ldots, E(\phi_k(X^n) \mid \mathcal{F}^n_t)) \Rightarrow$$

$$(V_t, \phi^*_1(Y_t), \ldots\ldots, \phi^*_k(Y_t)) \quad \text{on } D(S \times R^k).$$

This establishes the "only if" part of Proposition 16.15. And to prove the "if" part we need only show that $\{(X^n, Z^n) : n \geq 1\}$ is tight on $D(S \times \Pi)$, since (16.17) and Lemma 14.21 show that any subsequential limit must be (X^∞, Z^∞). Thus it suffices to verify the hypotheses of Lemma 16.12 for the $D(S\times\Pi)$-valued processes (X^n, Z^n). Since $X^n \Rightarrow X$, hypothesis (a) follows from Proposition and Lemma 16.6. To check (b), observe first that by Lemma 16.11 there exist countable sets $H_1 \subset C(S)$, $H_2 \subset C(D(S))$ such that

H_1 satisfies condition (16.10) for S;

$\{\phi^*: \phi \in H_2\}$ satisfies condition (16.10) for Π.

Define $H_3 \subset C(S\times\Pi)$ by $H_3 = \{\phi_1(f) + j\phi_2^*(\mu): \phi_i \in H_i,\ j \geq 0\}$. It is easily checked that H_3 satisfies (16.10) for $S\times\Pi$. Thus to get (b) it suffices to show

$$\phi_1(X_t^n) + j\phi_2^*(Z_t^n) \Rightarrow \phi_1(X_t) + j\phi_2^*(Z_t) \quad \text{on } D(R).$$

But this follows from the continuous mapping theorem, since $\phi^*(Z_t) = E(\phi(X)|\mathcal{F}_t)$.

For the rest of this section we reconsider condition (4.2). Call a sequence of processes (X^n, F^n) <u>taut</u> if that condition holds for stopping times in the set $\mathcal{T}^n$ of stopping times on F^n. That is,

$$(16.18) \qquad d(X^n_{T_n+\delta_n}, X^n_{T_n}) \xrightarrow[p]{} 0 \quad \text{for } T_n \in \mathcal{T}^n_L,\ \delta_n \downarrow 0.$$

Of course this implies (4.2), since any natural stopping time on X^n (i.e. one satisfying (4.1)) is a stopping time on F^n. We saw in Chapter 3 the use of tautness in proving weak convergence. Now tautness is not necessary for weak convergence, but Proposition 16.20 below says that tautness is necessary for extended weak convergence to a quasi left continuous limit.

(16.19) LEMMA. (X^n,F^n) <u>is taut if and only if</u> $d(X^n_{U_n}, X^n_{T_n}) \xrightarrow[p]{} 0$ <u>for all</u> $T_n, U_n \in \mathcal{T}^n_L$ <u>such that</u> $T_n \leq U_n \leq T_n + \delta_n$, $\delta_n \downarrow 0$.

(16.20) PROPOSITION. <u>Suppose</u> $(X^n,F^n) \Rightarrow (X,F)$, <u>and suppose</u> (X,F) <u>is quasi left continuous. Then</u> (X^n,F^n) <u>is taut</u>.

The proof of these results is deferred until Chapter 7, though we derive some consequences below. First, we record an obvious consequence of the definition of tautness.

(16.21) LEMMA. <u>Suppose</u> (X^n,F^n) <u>is taut on</u> $D(S_1)$, <u>and</u> (Y^n,F^n) <u>is taut on</u> $D(S_2)$. <u>Then</u> $((X^n,Y^n),F^n)$ <u>is taut on</u> $D(S_1 x S_2)$.

As mentioned earlier, $D(S_1 x S_2)$ is often more convenient than $D(S_1) \times D(S_2)$ for handling bivariate processes. Lemma 16.21 says that when we establish tightness via tautness, we automatically get tightness on $D(S_1 x S_2)$. Indeed, we have already used this technique in the proof of

(16.22) COROLLARY. <u>If</u> (X^n,F^n) <u>is taut then</u> $d(X^n_{U_n}, X^n_{U_n-}) \xrightarrow[p]{} 0$ <u>for predictable stopping times</u> $U_n \in \mathcal{T}^n_L$.

<u>Proof</u>. Given (U_n) we can choose $T_n \in \mathcal{T}^n_L$ such that

$$T_n < U_n \text{ on } \{U_n > 0\};$$
$$P(U_n - T_n > n^{-1}) < n^{-1};$$
$$P(d(X^n_{T_n}, X^n_{U_n-}) > n^{-1}) < n^{-1}.$$

Now apply Lemma 16.19 to T_n and $U_n \wedge (T_n + n^{-1})$.

(16.23) COROLLARY. <u>Suppose</u> (X^n,F^n) <u>is taut. Suppose</u> $(T_n,X^n) \Rightarrow (V,Y)$ <u>on</u> $R \times D(S)$, <u>where</u> $T_n \in \mathcal{T}^n$. <u>Then</u> $(T_n, X^n_{T_n}) \Rightarrow (V,Y_V)$ <u>on</u> $R \times S$.

<u>Proof</u>. The "process" form of Lemma   shows, without using any hypothesis about X^n or T_n:

(16.24) if $(T_n,X^n) \Rightarrow (V,Y)$ and if $V(\omega)$ is a.s. a continuity point of $Y(\omega)$ then $(T_n, X^n_{T_n}) \Rightarrow (V,Y_V)$.

Now for δ outside some countable set, $V(\omega)+\delta$ is a continuity point of $Y(\omega)$, and for such δ we have, by (16.24),

$$(T_n+\delta, X^n_{T_n+\delta}) \Rightarrow (V+\delta, Y_{V+\delta}).$$

Consider a sequence $\delta_n \downarrow 0$ and use (16.18).

17 OPTIMAL STOPPING

Example 11.4 showed that weak convergence $X^n \Rightarrow X$ is not sufficient for convergence of the optimally-stopped values $\sup\{Eg(X^n_T): T \in \mathcal{T}^n_1\}$ of the processes. To handle this type of problem within weak convergence theory it seems necessary to impose very restrictive conditions - e.g. each X^n is Markov and the generators converge. Theorem 17.2 says that extended weak convergence is sufficient, under mild conditions. This suggests that extended weak convergence provides a natural framework for discussing robustness of optimal stopping procedures.

Let $\gamma : [0,\infty) \times S \to R$ be bounded and continuous. Given a process (X,F) define

(17.1) $\Gamma(L) = \sup\{E\,\gamma(T,X_T): T \in \mathcal{T}_L\}$.

(17.2) THEOREM. Suppose $(X^n,F^n) \underset{=}{\Rightarrow} (X^\infty,F^\infty)$. Suppose (X^∞,F^∞) is quasi left continuous, and suppose

(17.3) F^∞ is the usual filtration for Z^∞.

Then $\Gamma_n(L) \to \Gamma_\infty(L)$.

Remark. Of course we could replace boundedness of γ by a suitable uniform integrability condition. Note also that (17.3) is a consequence of the simpler condition

(17.4) F^∞ is the usual filtration for X^∞.

Here is the idea behind the proof. For a process (X,F) let $\mathcal{S}(X,F) = \{\mathcal{L}(T,X_T): T \in \mathcal{T}\}$ be the set of stopped distributions. We want to say:

(17.5) $(X,F) \equiv (Y,G)$ implies $\mathcal{S}(X,F) = \mathcal{S}(Y,G)$,

and then to say that extended weak convergence implies convergence of $\mathcal{S}(X^n,F^n)$ to $\mathcal{S}(X^\infty,F^\infty)$ in some sense. The next example shows (17.5) may fail without condition (17.3).

(17.6) Example. Let U be a random variable with $P(U>u) = e^{1-u}$, $u \geq 1$. Let B, C be events with $P(B) = P(C) = \frac{1}{2}$, and suppose B, C and U are independent. Let $X_t = 1_{(t \geq U)}$, and let F be the usual filtration of X. Define filtrations G and H by

$$\mathcal{G}_t = \sigma(B,C) \qquad t < 1$$
$$\mathcal{H}_t = \sigma(B) \qquad t < 1$$
$$\mathcal{G}_t = \mathcal{H}_t = \sigma(B,C,\mathcal{F}_t,\ C \cap \{U \leq 2\}) \qquad 1 \leq t.$$

Plainly $(X,G) \equiv (X,H)$. Now we can define a stopping time T on G by

$$\begin{aligned} T &= 0 \quad \text{on } C^c \\ &= 1 \quad \text{on } C \cap \{U > 2\} \\ &= U \quad \text{on } C \cap \{U \leq 2\}. \end{aligned}$$

But there is no stopping time S on H such that $\mathcal{L}(S,X_S) = \mathcal{L}(T,X_T)$. Further, defining γ by

$$\gamma(t,1) = 1$$
$$\gamma(t,0) = -t \qquad t \leq 2,$$

then $E\,\gamma(T,X_T) > 0$ whereas $E\,\gamma(S,X_S) \leq 0$ for every stopping time $S \leq 2$ on H. Thus the optimally-stopped values (17.1) of synonymous processes may differ.

Remark. There is a simpler but less interesting counter-example to (17.5). Let $X = 0$, F the trivial filtration, and G a general filtration. Then $(X,F) \equiv (X,G)$, but F has no non-constant stopping times whereas G may have. This example is not significant, because in optimal stopping problems we may use randomised stopping times (see below), and then $\mathcal{S}(X,F)$ does equal $\mathcal{S}(X,G)$. Example 17.6 is more subtle, and the problem cannot be circumvented by considering randomised stopping times (the set B was included to show this).

We now start the proof of Theorem 17.2.

(17.7) LEMMA. Suppose $Y^n \Rightarrow Y^\infty$ on $D(S)$. Let F^n be the usual filtration for Y^n. Then given $T_\infty \in \mathcal{J}^\infty$ there exist $T_n \in \mathcal{J}^n$ such that

$$(T_n, Y^n_{T_n}) \Rightarrow (T^\infty, Y^\infty_{T_\infty}).$$

Proof. We may assume (Skorohod representation) that $Y^n \to Y^\infty$ a.s.. By approximating T_∞ from above, we may assume that T_∞ takes values in a discrete set (t_i) of continuity points of Y^∞, and that the set $A_i = \{T = t_i\}$ is in $\sigma(Y_r : r \le t_i)$. Fix $\varepsilon > 0$. By Lemma there exist bounded continuous $\mathcal{F}^D_{t_i}$-measurable functions ϕ_i such that

$$(17.8) \qquad E\left|\phi_i(Y^\infty) - 1_{A_i}\right| \le \varepsilon . 2^{-i}.$$

Define $T_n \in \mathcal{J}^n$ by

$$T_n = \min\{t_i : \phi_i(Y^n) > \tfrac{1}{2}\}.$$

Now $\phi_i(Y^n) \to \phi_i(Y^\infty)$ a.s.. So

$$(T_n = T \text{ for all sufficiently large } n) \quad \text{a.s. on } \bigcap_i \Omega_i$$

where

$$\Omega_i = A_i \cap \{\phi_i(Y^\infty) > \tfrac{1}{2}\} \quad \cup \quad A_i^c \cap \{\phi_i(Y^\infty) < \tfrac{1}{2}\} \quad .$$

But T_∞ is a.s. a continuity point of Y^∞, so

$$(T_n, Y^n_{T_n}) \to (T_\infty, Y^\infty_{T_\infty}) \quad \text{a.s. on } \bigcap \Omega_i.$$

Now (17.8) implies $P(\bigcap \Omega_i) \ge 1 - 3\varepsilon$. The result follows.

(17.9) LEMMA. Under the hypotheses of Theorem 17.2, $\Gamma_\infty(L) \le \liminf \Gamma_n(L)$.

Proof. Fix $T \in \mathcal{J}^\infty_L$. Applying Lemma 17.7 to (X^n), there exist $T_n \in \mathcal{J}^n$ such that

$$(T_n, X^n_{T_n}) \Rightarrow (T_\infty, X^\infty_{T_\infty}).$$

Since $T_\infty \leq L$ we see that $T_n \wedge L - T_n \underset{p}{\rightarrow} 0$. So by Proposition 16.20 and Lemma 16.19, $d(X^n_{T_n \wedge L}, X^n_{T_n}) \underset{p}{\rightarrow} 0$. This implies

$$(T_n \wedge L, X^n_{T_n \wedge L}) \Rightarrow (T_\infty, X^\infty_{T_\infty}),$$

and the result follows.

Lemma 17.9 is one half of Theorem 17.2. The other half, Lemma 17. below, requires the idea of randomised stopping times. Given a filtration F, a randomised stopping time on F is a stopping time on some enlarged filtration $\mathcal{F}^*_t = \sigma(\mathcal{F}_t, \mathcal{U})$, where $\mathcal{U}$ is independent of $\mathcal{F}_\infty$. The next lemma is well-known; (b) says that there is nothing to be gained by using randomised stopping times in optimal stopping problems.

(17.10) LEMMA (a) Let (X,F) be a process. Let $V \geq 0$ be a random variable such that $\{V \leq t\}$ and $\mathcal{F}_\infty$ are conditionally independent given $\mathcal{F}_t$. Then there exists a randomised stopping time T on F such that $\mathcal{L}(T, X_T) = \mathcal{L}(V, X_V)$.
(b) Define $\Gamma^*(L)$ as at (17.1), but using randomised stopping times. Then $\Gamma^*(L) = \Gamma(L)$.

Proof. Let A_t be a Skorohod version of the submartingale $P(V \leq t | \mathcal{F}_t)$. Then A is increasing, since for $s < t$ we have $A_t \geq P(V \leq s | \mathcal{F}_t) = P(V \leq s | \mathcal{F}_s) = A_s$, where the first equality is a consequence of the conditional independence. Let U be distributed uniformly on $[0,1]$, independent of $\mathcal{F}_\infty$. Define

(17.11) $T = \inf\{t : A_t \geq U\}$.

Then T is a randomised stopping time. For each t,

$$\begin{aligned} P(T \leq t | \mathcal{F}_\infty) &= P(A_t \geq U | \mathcal{F}_\infty) \\ &= A_t \\ &= P(V \leq t | \mathcal{F}_\infty). \end{aligned}$$

So $P(T \le t|X) = P(V \le t|X)$ for each t. Thus $\mathcal{L}(T,X) = \mathcal{L}(V,X)$, giving (a).

To prove (b), let $V \le L$ be a randomised stopping time, and define T as at (17.11). By (a), $E\,\gamma(V,X_V) = E\,\gamma(T,X_T)$. For $0 < u < 1$ let $T_u = \inf\{t: A_t \ge u\}$. Then $T_u \in \mathcal{T}_L$. And

$$E\,\gamma(T,X_T) = \int_0^1 E\gamma(T_u,X_{T_u})\,du \le \Gamma(L).$$

So $\Gamma^*(L) \le \Gamma(L)$, giving (b).

(17.12) LEMMA. Let (X^n,F^n), $1 \le n \le \infty$, satisfy the hypotheses of Theorem 17.2. Let $T_n \in \mathcal{T}^n$ be such that (T_n) is tight on R. Then there exists a randomised stopping time T_∞ on F^∞ and subsequences $Y^n = X^{j_n}$, $S_n = T_{j_n}$ such that $(S_n, Y^n_{S_n}) \Rightarrow (T_\infty, X^\infty_{T_\infty})$.

Remark. When we have only a single process (X,F), this "compactness of stopping times" result can be improved - see Baxter and Chacon (1977), Meyer (1978).

Proof. Using tightness, we can pass to a subsequence in which

(17.13) $(X^n,Z^n,T_n) = (X^\infty,Z^\infty,V)$, say, on $D(S)\times D(\Pi)\times R$.

By Proposition 16.20 and Corollary 16.23,

$$(T_n, X^n_{T_n}) \Rightarrow (V, X^\infty_V).$$

So by Lemma 17.10(a) it suffices to prove

(17.14) $\{V \le t\}$ and $\mathcal{F}^\infty_\infty$ are conditionally independent given $\mathcal{F}^\infty_t$.

Fix $t \ge 0$. Let $u > t$ be a continuity point of Z^∞. Consider $f \in C(R)$ and ψ, $\phi \in C(D(S))$ such that ψ is $\mathcal{F}^D_u$-measurable in the sense of Define ϕ^* as in Lemma 14.10. Since $\phi^*(Z^n_u) = E(\phi(X^n)|\mathcal{F}^n_u)$, we have

$$E\ \psi(Z^n).f(T_n \wedge u).(\phi^*(Z^n_u) - \phi(X^n)) = 0.$$

By (17.13)

$$E\ \psi(Z^\infty).f(V \wedge u).(\phi^*(Z^\infty_u) - \phi(X^\infty)) = 0.$$

This equality extends successively to:

(i) bounded measurable f, by Lemma

(ii) bounded measurable ϕ, by Lemma and Lemma 14.10;

(iii) bounded $\mathcal{F}^D_u$-measurable ψ, by Lemma

So by hypothesis (17.3)

$$E\, 1_A \cdot 1_{(V \leq t)} \cdot (\phi^*(Z^\infty_u) - \phi(X^\infty)) = 0, \qquad A \in \mathcal{F}^\infty_t.$$

Letting $u \downarrow t$,

$$E\, 1_A \cdot 1_{(V \leq t)} \cdot (E(\phi(X^\infty) \mid \mathcal{F}^\infty_t) - \phi(X^\infty)) = 0, \qquad A \in \mathcal{F}^\infty_t.$$

And this is equivalent to (17.14) (DM.II.45).

The proof of Theorem 17.2 is completed by the next lemma, an immediate consequence of Lemmas 17.12 and 17.10(b).

(17.15) LEMMA. Under the hypotheses of Theorem 17.2, $\Gamma_\infty(L) \geq \limsup \Gamma_n(L)$.

18 CONVERGENCE TO MARTINGALES

Proposition 5.1 showed that a weak limit of submartingales, or of processes which are approximately submartingales, is itself a submartingale. In this section we discuss the converse problem: if a sequence of processes converges to a submartingale, then are the processes themselves approximately submartingales (in some sense)? Example 11.4 showed this is not so under weak convergence, but Proposition 18.2 below gives a positive result under extended weak convergence.

Recall the definition (7.2) of a class (DL) process.

(18.1) Definition. A sequence of processes (X^n, F^n) is uniformly of class (DL) if $\{X^n_T : T \in \mathcal{T}^n_L, n \geq 1\}$ is uniformly integrable, for each $L < \infty$.

(18.2) PROPOSITION. Let (Y,F) be a quasi left continuous (sub)martingale. Let (X^n, F^n) be a sequence of processes, and let $1 \leq p < \infty$. Then the following are equivalent.

(a) $(X^n, F^n) \Rrightarrow (Y,F)$ and $(|X^n|^p, F^n)$ is uniformly of class (DL).

(b) There exist processes Y^n adapted to F^n such that

(i) (Y^n, F^n) is a (sub)martingale;

(ii) $(Y^n, F^n) \Rrightarrow (Y,F)$ and $(|Y^n|^p, F^n)$ is uniformly of class (DL);

(iii) $c^n_L = \sup_{T \in \mathcal{T}^n_L} E|Y^n_T - X^n_T|^p \to 0$, each $L < \infty$.

Lemma 16.16 and the elementary inequality

(18.3) $|x|^p \leq 2^p(|y|^p + |y-x|^p)$

show that (b) implies (a). The opposite implication is more interesting, since it shows that X^n may be approximated by a submartingale Y^n on the same filtration. For example, in the martingale case with $p = 1$, we obtain (from the optional sampling theorem for Y^n)

$$E\left|E(X^n_T \mid \mathcal{F}^n_S) - X^n_S\right| \leq 2\, C^n_L , \qquad S \leq T, \quad S,T \in \mathcal{T}^n_L.$$

This expresses the sense in which X^n is almost a martingale: it almost satisfies the optional sampling theorem.

Proposition 18.2 is the prototype for a class of "structure-preserving" results about extended weak convergence. These results take the form: if (X^n,F^n) converges to a limit with a certain structural property, then for large n the process X^n almost satisfies this property. Proposition 16.20 is of this form, for the property of quasi left continuity. Analogous results can be obtained for the Feller property and the predictable property.

Proof of Proposition 18.2. As observed ab ve, we need only prove (a) implies (b). The proof is a rather straightforward use of discrete approximation. We treat the submartingale case, and indicate the modifications needed in the martingale case.

Consider first a single process (X,F) of class (DL). For $m \geq 1$, $k \geq 0$ define

$$(18.4)\begin{cases} D_{j,m} = E(X_{j2^{-m}} - X_{(j-1)2^{-m}} \mid \mathcal{F}_{(j-1)2^{-m}}) \\ A^m_{k2^{-m}} = \sum_{j=1}^{k} (D_{j,m})^- , \quad \text{where } (x)^- = \max(0, -x) \\ S^m_{k2^{-m}} = A^m_{k2^{-m}} + X_{k2^{-m}} . \end{cases}$$

Then $(S^m_{k2^{-m}}, \mathcal{F}_{k2^{-m}})$, $k \geq 0$, is a discrete-time submartingale, so we can define a continuous-time submartingale (Y^m,F^m) by

$$Y^m_t = E(S^m_{k2^{-m}} \mid \mathcal{F}_t) , \qquad (k-1)2^{-m} \leq t < k2^{-m}.$$

Fix an integer L and a stopping time $T \in \mathcal{T}_L$. Then

$$Y^m_T = E(S^m_{r_m(T)} \mid \mathcal{F}_T), \quad \text{where } r_m(t) = \min\{k2^{-m} : k2^{-m} > t\}.$$

By Jensen's inequality,

$$(18.5) \quad E|Y^m_T - X_T|^p \leq E|S^m_{r_m(T)} - X_T|^p .$$

From (18.3), (18.4) and the fact that A^m is increasing, we get the estimate

$$(18.6) \quad E|S^m_{r_m(T)} - X_T|^p \leq 2^p \{E|A^m_L|^p + E|X_{r_m(T)} - X_T|^p\} .$$

Now consider a sequence of processes (X^n) satisfying hypothesis (a). We shall prove

$$(18.7) \quad \lim_{m\to\infty} \limsup_{n\to\infty} \sup_{T\in\mathcal{T}^n_L} E|X^n_{r_m(T)} - X^n_T|^p = 0$$

$$(18.8) \quad \lim_{n\to\infty} E|A^{n,m}_L|^p = 0, \quad \text{for each } m.$$

Then putting $V^n = Y^{m_n}$ for some sequence (m_n) converging to ∞ sufficiently slowly, we have from (18.5)-(18.8)

$$\sup_{T\in\mathcal{T}^n_L} E|V^n_T - X^n_T| \to 0 .$$

This is condition (iii) of (b), and (ii) follows by Lemma 16.16 and (18.3).

To prove (18.7), consider $T_n \in \mathcal{T}^n_L$ and $m_n \to \infty$. We must prove

$$E|X^n_{r_{m_n}(T_n)} - X^n_{T_n}|^p \to 0.$$

Now this sequence of random variables converges to zero in probability, by Proposition 16.20 and Lemma 16.19; convergence in L^p follows from the uniform integrability hypothesis in (a).

To prove (18.8), let $s < t$ be continuity points of the prediction process of (Y,F). By Lemma 16.2,

$$E(X_t^n - X_s^n | \mathcal{F}_s^n) \Rightarrow E(Y_t - Y_s | \mathcal{F}_s) \geq 0.$$

So $(E(X_t^n - X_s^n | \mathcal{F}_s^n))^- \to 0$ in probability, and also in L^p by the uniform integrability hypothesis. Then (18.8) follows from the definition of A^m, provided that $\{k2^{-m}: k \geq 0\}$ are continuity points. If not, we can carry out the entire construction using continuity points $(t_{k,m})$ in place of $(k2^{-m})$.

Remark. In the martingale case we alter (18.4) to:

$$S^m_{k2^{-m}} = X_{k2^{-m}} - \sum_{j=1}^{k} D_{j,m}$$

$$A^m_{k2^{-m}} = \sum_{j=1}^{k} |D_{j,m}|$$

and the argument is essentially unchanged.

19 THE DOOB-MEYER DECOMPOSITION

The Doob-Meyer decomposition, Theorem 7.4, can be regarded as a transformation taking submartingales to compensators. In this section we discuss the continuity "in distribution" of this transformation.

Observe first that the distribution of the compensator is not determined by the distribution of the submartingale. For example, the Poisson process N_t has compensator λt with respect to its usual filtration $\mathcal{F}_t$, but with respect to $\mathcal{G}_t = \mathcal{F}_\infty$ its compensator is N_t. The reader may object that this is artificial: why not restrict attention to processes with their usual filtrations? But to discuss convergence we are forced to allow unusual filtrations, as the next example shows.

(19.1) <u>Example</u>. Let $P(U > u) = e^{-u}$, and let F be the usual filtration for $1_{(t \geq U)}$. Let

$$X_t^n = n^{-1} 1_{(t \geq U)} + 1_{(t \geq U+1)}$$

Then (X^n, F) is a submartingale with compensator

$$A_t^n = n^{-1}.(t \wedge U) + 1_{(t \geq U+1)}.$$

Now $X^n \Rightarrow X$ and $A^n \Rightarrow A$, where

$$X_t = A_t = 1_{(t \geq U+1)}$$

And A is the compensator of (X,F). Now F is not the usual filtration of X: with respect to its usual filtration, X has compensator $(t-1)^+ \wedge U$. But F is plainly the "correct" filtration for X in this situation.

Perhaps the fundamental reason why weak convergence is unsatisfactory here is that the Doob-Meyer decomposition is not a sample-path transformation: there does not exist a map $\theta: D(R) \to D(R)$ such that $\theta(X)$ is the compensator of

X for each class (DL) submartingale. The next results show that extended weak convergence is more satisfactory.

(19.2) LEMMA. Let A be the compensator of the class (DL) submartingale (X,F). Suppose $(X,F) \equiv (X',F')$. Then (X',F') is a class (DL) submartingale, and its compensator A' is such that $\mathcal{L}(A') = \mathcal{L}(A)$.

(19.3) THEOREM. Let (X^n,F^n), $1 \leq n < \infty$, be a sequence of submartingales uniformly of class (DL), in the sense of (18.1). Suppose $(X^n,F^n) \Rrightarrow (X^\infty,F^\infty)$. Then (X^∞,F^∞) is a class (DL) submartingale: if it is quasi left continuous then $A^n \Rightarrow A^\infty$.

Before the proofs, we give two counter-examples to plausible strengthenings of these results.

(19.4) Example. Let $P(U > u) = e^{-u}$, $T_n = 1 + U/n$, $X_t^n = 1_{(t \geq T_n)}$ and $X_t = 1_{(t \geq 1)}$. With usual filtrations, $X^n \Rrightarrow X$. But X^n has compensator

$$A_t^n = n(t-1)^+ \wedge U ,$$

and X has compensator $A_t = 1_{(t \geq 1)}$, so A^n does not converge to A in any "distributional" sense. Thus to ensure weak convergence in Theorem 19.3, some restriction must be placed on the limit submartingale. In fact, the hypothesis of quasi left continuity can be weakened to regularity (see Definition 7.3), though we shall not prove this.

(19.5) Example. In Example 17.6, the compensator of X with respect to both G and H is

$$\begin{aligned} A_t &= (t-1)^+ \wedge U \qquad \text{on } C^c \\ &= (t-2)^+ \wedge U \qquad \text{on } C \cap \{U > 2\} \\ &= \int_1^{t \wedge U} \frac{e^{1-u}}{e^{1-u} - e^{-1}}\, du \qquad \text{on } C \cap \{U \leq 2\}. \end{aligned}$$

Now $(X,G) \equiv (X,H)$, but $(A,G) \not\equiv (A,H)$ because, for instance, $E(A_3 \mid \mathcal{G}_0)$ does not have the same distribution as $E(A_3 \mid \mathcal{H}_0)$. Thus the conclusion of Lemma 19.2 cannot be strengthened to $(A',F') \equiv (A,F)$.

We now start the proof of Theorem 19.3.

(19.6) LEMMA. <u>Under the hypotheses of Theorem 19.3</u>, (A^n) <u>is tight and</u>

$$P(\sup_{t \leq L} A^n_t - A^n_{t-} > \varepsilon) \to 0 \text{ \underline{as} } n \to \infty .$$

<u>Proof</u>. For any bounded stopping time T on F^n,

(19.7) $\quad EX^n_T = EA^n_T + EX^n_0$.

Consider $T_n \in \mathcal{T}^n_L$, $\delta_n \downarrow 0$. By Proposition 16.20,

$$X^n_{T_n + \delta_n} - X^n_{T_n} \to 0 \text{ in probability,}$$

and then we have convergence in L^1, since the processes are uniformly of class (DL). Applying (19.7) to (T_n) and $(T_n + \delta_n)$,

$$EA^n_{T_n + \delta_n} - EA^n_{T_n} \to 0.$$

But A^n is increasing, so $A^n_{T_n + \delta_n} - A^n_{T_n} \to 0$ in L^1. Now tightness follows from Theorem 4.4. The other assertion follows from Corollary 16.22 applied to $U_n = \min\{t : A^n_t - A^n_{t-} \geq \epsilon\} \wedge L$, which is predictable by Lemma 6.11.

<u>Remark</u>. Now the obvious method of attacking the proof of Theorem 19.3 is to consider a subsequential limit $(Z^\infty, X^\infty, A^\infty)$ of (Z^n, X^n, A^n). It is not hard to see that A^∞ is continuous increasing, and $X^\infty - A^\infty$ is a martingale with respect to G, the usual filtration of $(Z^\infty, X^\infty, X^\infty - A^\infty)$. Thus A^∞ is the compensator of X^∞ with respect to G. But we do not know how to show that A^∞ has the same distribution as the compensator of X^∞ with respect to F^∞.

Instead, we adopt the more pedestrian technique of discrete approximation, as in Section 18. Murali Rao (1969) has given a proof of the Doob-Meyer decomposition using this technique, and we use some of his estimates. For typographical convenience we write $X(t)$ for X_t.

Given an integrable process (X,F), define a process $A_m(t)$ by

$$A_m(0) = 0$$

$$(19.8) \quad A_m((i+1)2^{-m}) - A_m(i2^{-m}) = E(X((i+1)2^{-m}) - X(i2^{-m}) \mid \mathcal{F}_{i2^{-m}})$$

$$A_m(t) = A_m(i2^{-m}) \quad \text{on } i2^{-m} \le t < (i+1)2^{-m}.$$

(19.9) LEMMA. Suppose (X,F) is a submartingale.

(a) If $(A_m(L))_{m \geq 1}$ is uniformly integrable then X is class (DL).

(b) If X is class (DL) then $(A_m(L))_{m \geq 1}$ is uniformly integrable; explicitly,

$$E\, A_m(L).1_{(A_m(L) > 2\lambda)} \le 6 \sup\{E\, X(T).1_B : T \in \mathcal{T}_L,\ P(B) \le \lambda^{-1} E\, X(L)\} .$$

(c) Now suppose X is class (DL), with compensator A. Let w'' be as in , and let $0 < 3\varepsilon < \eta < 1 < \lambda < \infty$. Then

$$P(\sup_{t \le L} A(t) - A_m(t) > 2\eta) \le 3\eta^{-1}\Big[EA(L).1_{(A(L) > \lambda)} + \lambda \{3\varepsilon + P(w''(A, 2^{-m}, L) > \varepsilon) + P(\sup_{t \le L} A(t) - A(t-) > \varepsilon)\}^{\frac{1}{2}} \Big] .$$

Parts (a) and (b) are a reformulation of Lemma 2 of Murali Rao (1969); a proof can also be found in Chapter 3 of Liptser and Shiryayev (1977). Part (c) is a quantitative version of Rao's observation that A_m converges to A in L^1 when A is continuous. We shall prove (c) later.

Now let (X^n, F^n) be a sequence of submartingales, uniformly of class (DL). Let A^n be their compensators. Suppose $(X^n, F^n) \Rightarrow (X^\infty, F^\infty)$. Define A^n_m as at (19.8). We shall assume $\{i2^{-m} : i,m \ge 0\}$ are continuity points of

(X^∞, Z^∞), for otherwise we could use continuity points $(t_{m,i})$ in (19.8) in their place.

(19.10) LEMMA. $A_m^n \Rightarrow A_m^\infty$ <u>as</u> $n \to \infty$, <u>for each</u> m.

<u>Proof</u>. Fix m. Define $h_i: D(R) \to R$ by $h_i(f) = f((i+1)2^{-m}) - f(i2^{-m})$. Then

$$A_m^n((i+1)2^{-m}) - A_m^n(i2^{-m}) = E(h_i(X^n) \mid \mathcal{F}^n_{i2^{-m}}).$$

By Lemma 16.5,

$$(E(h_0(X^n)\mid\mathcal{F}^n_0), E(h_1(X^n)\mid\mathcal{F}^n_{2^{-m}}), \dots\dots\dots) \Rightarrow (E(h_0(X^\infty)\mid\mathcal{F}^\infty_0), E(h_1(X^\infty)\mid\mathcal{F}^\infty_{2^{-m}}), \dots\dots\dots)$$

and so

(19.11) $(A_m^n(0), A_m^n(2^{-m}), \dots\dots) \Rightarrow (A_m^\infty(0), A_m^\infty(2^{-m}), \dots\dots)$.

Since A_m is piecewise constant, the lemma is established.

(19.12) LEMMA. (X^∞, F^∞) <u>is a class</u> (DL) <u>submartingale</u>.

<u>Proof</u>. To prove X^∞ is a submartingale it suffices to show $A_m^\infty((i+1)2^{-m}) - A_m^\infty(i2^{-m}) \geq 0$. But this follows from (19.11). Since (X^n) is uniformly of class (DL), Lemma 19.9(b) shows that $\{A_m^n(L): m \geq 1, 1 \leq n < \infty\}$ is uniformly integrable. Then by (19.11),

(19.13) $\{A_m^n(L): m \geq 1, 1 \leq n \leq \infty\}$ is uniformly integrable.

Lemma 19.9(a) completes the proof.

Now assume (X^∞, F^∞) is quasi left continuous, so that its compensator A^∞ is continuous. Using (B.4.2), Theorem 19.3 follows from Lemma 19.10 and the next lemma.

(19.14) LEMMA. <u>Let</u> $L < \infty$, $\eta < 1$. <u>Then</u>

(i) $\lim_{m\to\infty} \limsup_{n\to\infty} P(\sup_{t \leq L} |A_m^n(t) - A^n(t)| > 2\eta) = 0$

(ii) $\lim_{m\to\infty} P(\sup_{t \leq L} |A_m^\infty(t) - A^\infty(t)| > 2\eta) = 0$

Proof. Fix $0 < \varepsilon < \eta/3$. By Lemma 19.6 and Proposition

$$P(\sup_{t \le L} |A^n_t - A^n_{t-}| > \varepsilon) \to 0$$

$$\lim_{m\to\infty} \limsup_{n\to\infty} P(w''(A^n, 2^{-m}, L) > \varepsilon) = 0.$$

So by Lemma 19.9(c) the quantity in (i) is at most

$$3\eta^{-1}\{\lambda(3\varepsilon)^{\frac{1}{2}} + \sup_{m,n} E\, A^n_m(L).1_{(A^n_m(L) > \lambda)}\}.$$

But by (19.3) we can make this arbitrarily small, by first choosing λ sufficiently large and then choosing ε sufficiently small. This establishes (i), and (ii) is similar.

Proof of Lemma 19.9(c). Without loss of generality take $m = 0$. Write $\hat{A}$ in place of A_m. Observe that definition (19.8) can be rephrased as

$$(19.15)\quad \begin{cases} \hat{A}_0 = 0 \\ \hat{A}_{i+1} - \hat{A}_i = E(A_{i+1} - A_i | \mathcal{F}_i) \\ \hat{A}_t = \hat{A}_i \quad \text{on} \quad i \le t < i+1, \end{cases}$$

where we revert to writing times as subscripts.

Because $\hat{A}$ is constant and A is increasing on each interval $[i, i+1)$,

$$(19.16)\quad P(\sup_{t \le L} |A_t - \hat{A}_t| > 2\eta) \le P(\max_{i \le L} |A_i - \hat{A}_i| > \eta) + P(\max_{i \le L} (A_i - A_{i-1}) > \eta).$$

By the maximal inequality for the martingale $(A_i - \hat{A}_i, \mathcal{F}_i)$,

$$(19.17)\quad P(\max_{i \le L} |A_i - \hat{A}_i| > \eta) \le \eta^{-1} E|A_L - \hat{A}_L|.$$

Now let $B_t = \lambda \wedge A_t$, and define $\hat{B}_t$ as at (19.15). Plainly $\hat{B}_t \le \hat{A}_t$. So

$$E|A_L - \hat{A}_L| \le (EA_L - EB_L) + (E\hat{A}_L - E\hat{B}_L) + E|B_L - \hat{B}_L|$$
$$= 2(EA_L - EB_L) + E|B_L - \hat{B}_L|$$

$$(19.18) \qquad \leq 2\, EA_L . 1_{(A_L > \lambda)} + \{E(B_L - \hat{B}_L)^2\}^{\frac{1}{2}}.$$

Since $(B_i - \hat{B}_i, \mathcal{F}_i)$ is a martingale,

$$E(B_L - \hat{B}_L)^2 = \sum_{i \leq L} E(\,(B_i - B_{i-1}) - (B_i - B_{i-1})\,)^2$$

$$\leq \sum E(B_i - B_{i-1})^2$$

$$(19.19) \qquad \leq E \sum (B_i - B_{i-1})^2 .$$

An elementary argument shows that, for constants $0 \leq c_0 \leq c_1 \leq \ldots\ldots \leq \lambda$,

$$\sum (c_i - c_{i-1})^2 \leq \lambda \max(c_i - c_{i-1})$$

$$\leq \lambda . 3\varepsilon + \lambda^2 . 1_{(\max(c_i - c_{i-1}) > 3\varepsilon)}$$

Applying this to (19.19),

$$(19.20) \quad E(B_L - \hat{B}_L)^2 \leq 3\lambda\varepsilon + \lambda^2 P(\max_{i \leq L} (B_i - B_{i-1}) > 3\varepsilon).$$

We now assert:

$$(19.21) \quad P(\max_i (B_i - B_{i-1}) > 3\varepsilon) \leq P(w''(A,1,L) > \varepsilon) + P(\max_t (A_t - A_{t-}) > \varepsilon).$$

To prove this, fix ω such that $B_i(\omega) - B_{i-1}(\omega) > 3\varepsilon$. Consider whether or not there exists $t \in [i-1, i)$ such that $B_{i-1}(\omega) + \varepsilon \leq B_t(\omega) < B_i(\omega) - \varepsilon$. If so, then $w''(B(\omega),1,L) > \varepsilon$; if not, then $B_t(\omega) - B_{t-}(\omega) > \varepsilon$ for some t in $[i-1, i]$.

Finally, consider the last term of (19.16).

$$P(\max (A_i - A_{i-1}) > \eta) \leq P(A_L > \lambda) + P(\max (B_i - B_{i-1}) > \eta)$$

$$(19.22) \qquad \leq EA_L . 1_{(A_L > \lambda)} + P(\max (B_i - B_{i-1}) > 3\varepsilon).$$

Inequalities (19.16)-(19.22) establish the lemma.

This completes our rather convoluted proof of Theorem 19.3: no doubt the proof could be simplified. To end the section we prove Lemma 19.2.

This requires familiarity with $\sigma(L^1,L^\infty)$-convergence - see (DM.II.24).

We need a straightforward technical lemma.

(19.23) LEMMA. Let Y be a random element. Let (h_n) be measurable real-valued functions such that $h_n(Y) \to V$ $\sigma(L^1,L^\infty)$. Suppose $\mathcal{L}(Y') = \mathcal{L}(Y)$. Then $h_n(Y') \to V'$ $\sigma(L^1,L^\infty)$, where $\mathcal{L}(V',Y') = \mathcal{L}(V,Y)$.

Proof of Lemma 19.2. Let (X,F) be a class (DL) submartingale, and define $A_m(t)$ by (19.8). Let Z be the prediction process of (X,F). Then

$$A_m(t) = H_{m,t}(Z)$$

for a certain function $H_{m,t}: D(\mathbb{T}) \to R$. Let (X',F') be some other process such that $(X',F') \equiv (X,F)$, that is to say $\mathcal{L}(Z') = \mathcal{L}(Z)$. Then $\mathcal{L}(A_m') = \mathcal{L}(A_m)$. So A_m' is increasing, and this implies (X',F') is a submartingale. Lemma 19.9 shows (X',F') is class (DL). Let A' be its compensator. Murali Rao (1969) shows that, for each t_1,

$$A_m(t_1) \to A(t_1) \quad \sigma(L^1,L^\infty)$$

$$A_m'(t_1) \to A'(t_1) \quad \sigma(L^1,L^\infty).$$

Lemma 19.23 shows that $\mathcal{L}(Z',A'(t_1)) = \mathcal{L}(Z,A(t_1))$. Repeating the argument with $t_2,\ldots\ldots,t_k$ gives

$$\mathcal{L}(Z',A'(t_1),\ldots\ldots,A'(t_k)) = \mathcal{L}(Z,A(t_1),\ldots\ldots,A(t_k)).$$

Hence $\mathcal{L}(A') = \mathcal{L}(A)$.

CHAPTER 6 - ESTABLISHING EXTENDED WEAK CONVERGENCE

In very simple cases where we have explicit expressions for the prediction process, it is possible to give ad hoc proofs of extended weak convergence dirextly from the definition: section 20 gives two instances of this. In section 21 we show how weak convergence results proved by the technique of Chapter 3 can be improved to extended weak convergence.

20 AD HOC PROOFS

We shall first consider what extended weak convergence means for "single point" processes. Let T be a random variable such that

(20.1) $\quad 0 < T < \infty$ a.s., $\quad P(T > q) > 0$ for each $q < \infty$.

Put $G(t) = P(T \leq t)$, $X_t = 1_{(t \geq T)}$ and let F be the usual filtration. Call such a process (X,F) a <u>single point process</u>.

(20.2) PROPOSITION. <u>For a sequence</u> (X^n) <u>of single point processes, the following are equivalent.</u>

(i) $\quad X^n \Rrightarrow X^\infty$

(ii) $\quad G_n \to G_\infty$ <u>in</u> D(R)

(iii)(a) $\quad T_n \Rightarrow T_\infty$; <u>and</u>

(b) <u>for each</u> t <u>such that</u> $P(T_\infty = t) \;\; 0$ <u>there exist</u> $t_n \to t$ <u>such that</u>

$P(T_n = t_n) \to P(T_\infty = t)$.

<u>Proof. (i) implies (ii)</u>. Let L be a continuity point of G_∞. Define $h \in C(D(R))$ by $h(f) = \int_L^\infty (|f(t) - 1| \wedge 1)\, e^{-t} dt$. Then

$$E(h(X^n)|\mathcal{F}^n_t) = \frac{a_n}{1-G_n(t)} 1_{(t < T)} \qquad \text{on } 0 \leq t \leq L;$$

where $a_n = Eh(X^n)$. By Lemma 16.2, $E(h(X^n)|\mathcal{F}^n_t) \Rightarrow E(H(X^\infty)|\mathcal{F}^\infty_t)$. Since $P(T_n > L) \to P(T_\infty > L) \;\; 0$, it is not hard to deduce

$$\frac{a_n}{1-G_n(t)} \to \frac{a_\infty}{1-G_\infty(t)} \qquad \text{in } D[0,L] .$$

But $a_n \to a_\infty$, so $G_n \to G_\infty$ in $D[0,L]$, and (ii) follows.

<u>(ii) implies (iii)</u>. Suppose $G_n \to G$ in $D(R)$. If t is a continuity point of G_∞, then by Lemma 2.4 $G_n(t) \to G_\infty(t)$: this proves (iii)(a). If t is a discontinuity point of G_∞, then Lemma 2.6(ii) proves the existence of (t_n) satisfying (iii)(b).

<u>(iii) implies (ii)</u> It suffices to prove that (G_n) is precompact in $D(R)$. Conditions (C1) and (C3) of section 2 are immediate, so it suffices to verify (C4). So consider (u_j^i) and a subsequence (G_j) such that $u_j^i \to t$ and $G_j(u_j^i) \to s_i$ for each $i = 1,2,3$: we must prove $s_1 = s_2$ or $s_2 = s_3$. This is immediate if t is a continuity point of G_∞, so suppose not. Fix $\varepsilon > 0$. Choose $v_1 < t < v_2$ such that v_1, v_2 are continuity points of G_∞ and

$$G_\infty(v_2) - G_\infty(v_1) \le G_\infty(t) - G_\infty(t-) + \varepsilon .$$

By hypothesis (iii)(b), there exist $t_j \to t$ such that

$$G_j(t_j) - G_j(t_j-) \to G_\infty(t) - G_\infty(t-), \text{ and so}$$

$$\limsup_{j\to\infty} \{G_j(v_2) - G_j(t_j)\} + \{G_j(t_j-) - G_j(v_1)\} \le \varepsilon .$$

For each sufficiently large j, either $(u_j^1, u_j^2) \subset (v_1, t_j)$ or $(u_j^2, u_j^3) \subset (t_j, v_2)$; so $\min(s_2-s_1,\ s_3-s_2) \le \varepsilon$, and the result follows.

(ii) implies (i). We need rather a lot of notation. Let μ be a probability measure with distribution function G. Let G^{-1} be the inverse distribution function $G^{-1}(x) = \inf\{t: G(t) \geq x\}$. Let U be uniform on (0,1), and set

(20.3) $\quad T = G^{-1}(U).$

Then T has distribution μ. Let $\mathcal{P} = \{\mu : T$ satisfies $(20.1)\} \subset \mathcal{P}(R)$. For $0 < t < \infty$ let $t^* \in D(R)$ be the function $t^*(u) = 1_{(u \geq t)}$. For $\mu \in \mathcal{P}$ let $\mu^* \in \Pi$ be the distribution of the process T^*, i.e. of the single point process $1_{(t \geq T)}$. Clearly the map $\mu \to \mu^*$ is continuous. For $0 \leq t < \infty$ define $c_t : \mathcal{P} \to \mathcal{P}$ by $c_t(\mathcal{L}(T)) = \mathcal{L}(T \mid T > t)$. The map $t \to c_t(\mu)$ is Skorohod. We now see that the prediction process Z_t of the single point process $X_t = 1_{(t \geq T)}$ is

$$Z_t = (c_t(\mu))^* , \quad t < T(\omega)$$
$$= \delta_{(T(\omega))^*} , \quad t \geq T(\omega).$$

Now consider a sequence (G_n) satisfying hypothesis (ii). Define T_n by (20.3), using the same U for each n. To prove extended weak convergence, we shall show

$$(X^n, Z^n) \to (X^\infty, Z^\infty) \quad \text{a.s.} \quad \text{in } D(R \times \Pi).$$

First, by easily-verified properties of inverse distribution functions,

(20.4) $\quad T_n \to T \quad$ a.s.

(20.5) $\quad G_n(T_n -) \to G(T -) \quad$ a.s.

Now (20.4) implies $X^n \to X^\infty$ a.s. in D(R), and $Z^n_{T_n} \to Z^\infty_{T_\infty}$ a.s. in Π. So by Lemma 3.5 it suffices to prove $Z^n \to Z^\infty$ in $D(\Pi)$ a.s.. Appealing to Lemma it suffices to prove

(20.6) $\quad c_t(\mu_n) \to c_t(\mu_\infty) \quad$ in $D(\mathcal{P})$

(20.7) $\quad c_{T_n -}(\mu_n) \to c_{T_\infty -}(\mu_\infty) \quad$ in $\mathcal{P}$ $\quad$ a.s.

From the definition of c_t,

(20.8) $c_t(\mu_n)$ has distribution function $u \to \frac{G_n(u \vee t) - G_n(t)}{1 - G_n(t)}$

Now (20.7) follows from (20.5) and the hypothesis $G_n \to G_\infty$. To prove (20.6), let (λ_n) be a scaling sequence for (G_n), so by Lemma 2.7

(20.9) $t_n \to t\pm$ implies $G_n(\lambda_n(t_n)) \to G_\infty(t\pm)$.

It suffices to prove (λ_n) is a scaling sequence for $(c_t(\mu_n))$; that is, by Lemma 2.7, to prove

$$t_n \to t\pm \text{ implies } c_{\lambda_n(t_n)}(\mu_n) \to c_{t\pm}(\mu_\infty) \text{ in } \mathcal{P} .$$

But this follows from (20.8) and (20.9).

FELLER PROCESSES

In section 13 we described the prediction process of a Feller process. We now prove this description is correct, and then show that for Feller processes extended weak convergence is essentially the same as weak convergence.

Recall some notation from section 13. (X,F) is Feller if there is a jointly continuous map $\phi:[0,\infty) \times S \to \Pi$ such that, for each t,

$\phi(t,s)$ is a regular conditional distribution for ${}^tX$ given $\mathfrak{F}_t$

where ${}^tX$ is the post-t process ${}^tX_u = X_{t+u}$.

Define $\Theta:D(S) \times [0,\infty) \times D(S) \to D(S)$ by

$\Theta(f,t,g)$ is the function $u \to f(u), \quad u < t$

(20.10) $\qquad\qquad \to g(u-t), \quad u \geq t.$

The results in this section are consequences of the continuity properties of Θ which were developed in section 2. Lemma 2. says:

if $f_n \to f_\infty$, $t_n \to t_\infty$, $g_n \to g_\infty$, then $\Theta(f_n,t_n,g_n) \to \Theta(f_\infty,t_\infty,g_\infty)$ provided

either (a) $f_\infty(t_\infty) = g_\infty(0)$;

or (b) $t_\infty > 0$, $f_n(t_n-) \to f_\infty(t_\infty-)$.

Now consider the induced map $\tilde{\Theta}:D(S) \times [0,\infty) \times \Pi \to \Pi$

(20.11) $\quad \tilde{\Theta}(f,t,\mathcal{L}(Y)) = \mathcal{L}(\Theta(f,t,Y)).$

The Skorohod representation theorem gives the following result.

(20.12) LEMMA. _If_ $f_n \to f_\infty$, $t_n \to t_\infty$, $\mu_n \to \mu_\infty$, _and if either_

(a) $\pi_0(\mu_\infty) = \delta_{f_\infty(t_\infty)}$ _; or_

(b) $t_\infty > 0$, $f_n(t_n-) \to f_\infty(t_\infty-)$,

then $\tilde{\Theta}(f_n,t_n,\mu_n) \to \tilde{\Theta}(f_\infty,t_\infty,\mu_\infty)$.

Now for a Feller process X, set

(20.13) $\quad Z_t = \tilde{\Theta}(X,t,\phi(t,X_t)).$

As mentioned in section 13, it is easy to check that Z_t is a regular conditional distribution for X given $\mathcal{F}_t$. To prove Z is the prediction process, we must check that Z is Skorohod. But this follows from Lemma 20.12, using (a) when $t_n \downarrow t$ and (b) when $t_n \uparrow\uparrow t$. Note that the left limit Z_{t-} equals $\theta(X,t,\phi(t,X_{t-}))$, so that each path $Z(\omega)$ is continuous wherever $X(\omega)$ is continuous. This establishes Proposition 13.8.

To state the convergence result, let $\rho = \mathcal{L}(X_0)$ be the initial distribution of a Feller process X. For a sequence of Feller processes (X^n), it is clear that $\rho_n \Rightarrow \rho_\infty$ and $\phi_n(0,s) \to \phi_\infty(0,s)$ uniformly on compacts imply $X^n \Rightarrow X^\infty$.

(20.14) PROPOSITION. <u>Suppose</u> (X^n) <u>is a sequence of Feller processes such that</u> $\rho_n \to \rho_\infty$ <u>and</u> $\phi_n(t,s) \to \phi_\infty(t,s)$ <u>uniformly on compact subsets of</u> $[0,\infty) \times S$. <u>Then</u> $X^n \Rrightarrow X^\infty$.

<u>Proof</u>. We know $X^n \Rightarrow X^\infty$, and so by the Skorohod representation theorem we may assume $X^n \to X^\infty$ in D(S) a.s.. We must prove

(20.15) $(X^n, Z^n) \to (X^\infty, Z^\infty)$ in $D(S \times \Pi)$ a.s..

Fix ω, write $f_n = X^n(\ )$, and suppose $f_n \to f_\infty$ in D(S). Let (λ_n) be a scaling sequence for (f_n). We shall prove (λ_n) is a scaling sequence for $(Z^n(\omega))$. By Lemma 2.7 and the definition of Z^n, we must prove:

if $t_n \to t_\infty \pm$ then $\theta(f_n, \lambda_n(t_n), \phi_n(\lambda_n(t_n), f_n(\lambda_n(t_n))))$

(20.16) $$\theta(f_\infty, t_\infty, \phi_\infty(t_\infty, f_\infty(t_\infty \pm))) .$$

But (λ_n) is a scaling sequence for (f_n), so $f_n(\lambda_n(t_n \pm)) \to f_\infty(t_\infty \pm))$. Now (20.16) follows from Lemma 20.12, using (a) when $t_n \downarrow t$ and (b) when $t_n \uparrow\uparrow t$.

INDEPENDENT INCREMENTS

Say a real-valued process (X,F) has independent increments if, for each t, the post-t process ${}^tX$ is independent of $\mathcal{F}_t$. For such a process, let

(20.17) $Z_t = \tilde{\theta}(X,t,\mathcal{L}({}^tX))$.

We state without proof analogues of Propositions 13.8 and 20.14.

(20.18) PROPOSITION. (a) If X has independent increments then (20.17) defines the prediction process.

(b) Suppose $X^n \Rightarrow X^\infty$, where X^n has independent increments for each n. Then $X^n \Rrightarrow X^\infty$.

The proof is rather tedious because here we do not have the continuity properties of ϕ.

Remarks. The results of this section are not useful, because when so much structure is assumed we do not benefit from knowing that extended weak convergence holds. Rather, they are intended to demonstrate that extended weak convergence is not an unduly restrictive notion in several familiar contexts.

21 THE BOOTSTRAP

Let us formulate a heuristic principle: any weak convergence result proved by the martingale technique of Chapter 3 can be improved to extended weak convergence. Instead of attempting to prove any general form of this principle, we shall illustrate it by considering two specific results, on convergence to diffusions (Theorem 8.22) and convergence to the Poisson process (Proposition 9.1).

CONVERGENCE TO DIFFUSIONS

As in section 8, let X be a diffusion with drift b(x) and variance a(x), and let $\Delta_x = \mathcal{L}(X|X_0=x)$. Proposition 13.8 (proved in the previous section) shows that the prediction process Z is continuous and may be written

(21.1) $\quad Z_t = \tilde{\theta}(X_t, t, \Delta_{X_t})$

where $\tilde{\theta}$ is defined as in (20.11).

(21.2) THEOREM. <u>Under the hypotheses of Theorem 8.22</u>, $X^n \Rrightarrow X$.

<u>Proof</u>. Recall thet ${}^tX$ denotes the post-t process $(X_{t+u})_{u\geq 0}$. For each t, n define $\hat{Z}^n_t$ by

(21.3) $\quad \hat{Z}^n_t$ is a regular conditional distribution for ${}^tX$ given $\mathcal{F}^n_t$.

Then the prediction process Z^n satisfies

(21.4) $\quad Z^n_t = \theta(X^n, t, \hat{Z}^n_t)$ a.s., for each t,n.

By Theorem 8.22 and the Skorohod representation theorem, we may assume

(21.5) $\quad X^n \to X$ in D(R) a.s.

and so in particular

(21.6) $\quad X^n_t \to X_t$ a.s., for each t.

Suppose we can prove

(21.7) $\quad Z^n_t \underset{p}{\to} \Delta_{X^n_t}$ in $\mathbb{P}$, for each t.

Then using (21.4) and (21.5) we can apply Lemma 20.12(a) to Z_t^n and deduce

$Z_t^n \underset{p}{\to} Z_t$ in Π, for each t.

So for any k-tuple (t_i), $(Z_{t_1}^n, \ldots, Z_{t_k}^n) \Rightarrow (Z_{t_1}, \ldots, Z_{t_k})$, and the theorem follows from Proposition 16.3.

Thus the main part of the proof is to establish (21.7). As a preliminary, we need to abstract the concept "hypotheses sufficient to impl weak convergence to the diffusion". Consider a map $H: \Pi \to [0,\infty]$ such that

(21.8) if $\mu_n \in \Pi$, $y \in R$, $H(\mu_n) \to 0$, $\pi_0(\mu_n) \to \delta_y$, then $\mu_n \to \Delta_y$ in Π.

Given such a map, (21.8) is a weak convergence theorem: conversely, given a weak convergence theorem we can define H so that (21.8) is a restatement of the theorem. Later (21.13) we shall define H so that (21.8) is Theorem 8.22.

The central idea is the following "bootstrap" technique. If (X^n, F^n) satisfies the hypotheses of Theorem 8.22, the the future processes ${}^tX^n$ conditioned on events G_n in $\mathcal{F}_t^n$ should also satisfy the hypotheses (because a martingale conditioned on the past remains a martingale). So the conditional distributions Z_t^n of ${}^tX^n$ given $\mathcal{F}_t^n$ should approach the distribution of the diffusion. In making these ideas precise we shall approximate regular conditional distributions by elementary conditional distributions: this requires the following technical lemma. Fix t for the rest of the proof.

(21.9) LEMMA. Suppose $H: \Pi \to [0,\infty]$ satisfies (21.8). Let (ξ_n) be random elements of Π such that

(i) $\pi_0(\xi_n) \underset{p}{\to} \delta_{X_t}$ in $\mathcal{P}(R)$.

Suppose that for each n there exist random elements $(\xi_{n,k})$ such that

(ii) $\xi_{n,k} \underset{p}{\to} \xi_n$ in Π as $k \to \infty$;

(iii) $\lim_{n\to\infty} \sup_k E\, H(\xi_{n,k}) = 0$

Then $\xi_n \underset{p}{\to} \Delta_{X_t}$ in Π.

<u>Proof</u>. Let ρ be a metrisation of Π. By (ii) we can choose $k_n \to \infty$ such that

(21.10) $\rho(\xi_{n,k_n}, \xi_n) \underset{p}{\to} 0$ as $n \to \infty$.

Since π_0 is continuous, (21.10) and (i) imply

(21.11) $\pi_0(\xi_{n,k_n}) \underset{p}{\to} \delta_{X_t}$.

And by (iii),

(21.12) $H(\xi_{n,k_n}) \underset{p}{\to} 0$.

By considering subsequences where convergence in (21.12) and (21.11) holds a.s., and applying (21.8), we deduce $\xi_{n,k_n} \underset{p}{\to} \Delta_{X_t}$. Now (21.10) gives the lemma.

We shall apply this lemma to $\xi_n = Z^n_t$ and elementary conditional distributions $(\xi_{n,k})$: once the conditions of the lemma are verified, (21.7) and the theorem follow. Condition (i) is immediate, since $\pi_0(\xi_n) = \delta_{X^n_t} \to \delta_{X_t}$ a.s. by (21.6).

To construct $\xi_{n,k}$, choose for each n an increasing family $(\mathcal{G}_{n,k})$ of finite σ-fields such that $\bigvee_k \mathcal{G}_{n,k} = \sigma(\xi_n) \subset \mathcal{F}^n_t$. For each $G_j \in \mathcal{G}_{n,k}$ define ξ^*_j to be the ordinary conditional distribution of ${}^tX^n$ given G_j. Define $\xi_{n,k}(\omega) = \xi^*_j$ on G_j. Then condition (ii) holds by Lemma 12.7.

Next, we must specify H so that (21.8) is Theorem 8.22. Let Θ denote a collection $(\Omega, \mathcal{B}, P, F, X, N, \bar{N})$, where X, N and $\bar{N}$ are processes adapted to a filtration F on a probability triple $(\Omega, \mathcal{B}, P)$. Call Θ <u>admissible</u> if conditions (c) and (d) of Theorem 8.22 hold. Say Θ <u>represents</u> μ if $\mathcal{L}(X) = \mu$. Define

$$\alpha(\Theta, L) = \sup\{E(N_T^2) : T \in \mathcal{T}_L\}$$
$$\beta(\Theta, L) = \sup\{E|\tilde{N}_T| : T \in \mathcal{T}_L\}$$
$$(21.13) \quad \gamma(\Theta, L) = E \sup_{t \le L} (X_t - X_{t-})^2$$
$$\delta(\Theta) = \int_0^\infty e^{-L} \wedge \{\alpha(\Theta, L) + \beta(\Theta, L) + \gamma(\Theta, L)\} \, dL$$

$$H(\mu) = \inf\{\delta(\Theta) : \Theta \text{ admissible and represents } \mu\}.$$

Now if $H(\mu_n) \to 0$ then there exist processes X^n satisfying hypotheses (b)-(f) of Theorem 8.22 and such that $\mathcal{L}(X^n) = \mu_n$. So (21.8) is indeed a restatement of Theorem 8.22.

It remains to check condition (iii) of Lemma 21.9, and this is the "bootstrap" argument. We are given processes (X^n, F^n) satisfying the hypotheses of Theorem 8.22. That is, we have admissible collections $(\Omega, \mathcal{B}, P, F^n, X^n, N^n, \bar{N}^n) = \Theta^n$ such that

(21.14) $\quad \delta(\Theta^n) \to 0.$

Fix n,k for the rest of the proof, and consider $G_j \in \mathcal{G}_{n,k}$. Put

$$\Theta_j = (\Omega, \mathcal{B}, P(\cdot|G_j), (\mathcal{F}^n_{t+u})_{u\geq 0}, (X^n_{t+u})_{u\geq 0}, (N^n_{t+u})_{u\geq 0}, (\bar{N}^n_{t+u})_{u\geq 0}).$$

Then Θ_j is admissible (because a martingale conditioned on a past event remains a martingale) and represents ξ^*_j (by definition of ξ^*_j). Given stopping times $T_j \leq L$ on $(\mathcal{F}^n_{t+u})_{u\geq 0}$, we can define a stopping time $T \leq L+t$ on F^n by $T = t + \sum T_j 1_{G_j}$. So using the definitions of α, β

$$\sum P(G_j)\, \alpha(\Theta_j, L) \leq \alpha(\Theta^n, L+t)$$

(21.15) $\quad \sum P(G_j)\, \beta(\Theta_j, L) \leq \beta(\Theta^n, L+t).$

Similarly, $\quad \sum P(G_j)\, \gamma(\Theta_j, L) \leq \gamma(\Theta^n, L+t).$

Finally,

$$\begin{aligned} E\,H(\xi_{n,k}) &\leq \sum P(G_j)\, H(\xi^*_j) \\ &\leq \sum P(G_j)\, \delta(\Theta_j) \\ &\leq e^t\, \delta(\Theta^n) \quad \text{by definition of } \delta \text{ and the estimates (21.15).} \end{aligned}$$

Now (21.14) gives condition (iii).

Theorem 21.2 gave sufficient conditions for extended weak convergence: we shall now show these conditions are essentially necessary. We need a routine truncation lemma. Define

$$h_k(x) = -k, \quad x \le -k$$
$$= x, \quad -k \le x \le k$$
$$= k, \quad k \le x.$$

(21.16) LEMMA. Let $V_n \Rightarrow V$ be real-valued random variables such that $EV^2 < \infty$. Then there exist $k_n \to \infty$ such that $\{h^2_{k_n}(V_n)\}$ is uniformly integrable.

(21.17) PROPOSITION. Let (Y^n, F^n) be a sequence of processes, and X the diffusion with drift $b(x)$ and variance $a(x)$. In order that $Y^n \Rrightarrow X$ it is necessary and sufficient that there exist X^n adapted to F^n such that

(i) $\sup_{t \le L} |X^n_t - Y^n_t| \xrightarrow[p]{} 0$

(ii) (X^n, F^n) satisfies the hypotheses of Theorem 8.22.

Proof. Sufficiency is immediate from Theorem 21.2 and Lemma 16.16. So suppose $Y^n \Rrightarrow X$. Then, for fixed L,

(21.18) $\sup_{t \le L} (Y^n_t)^2 \Rightarrow \sup_{t \le L} X^2_t$

Now $E \sup_{t \le L} X^2_t < \infty$, so by Lemma 21.16 we can choose $k_n \to \infty$ such that

(21.19) $\{k_n \wedge \sup_{t \le L} (Y^n_t)^2\}$ is uniformly integrable.

By a diagonal argument, we can choose (k_n) so that (21.19) holds for all L. Define $X^n_t = h_{k_n}(Y^n_t)$. By construction,

(21.20) $\{\sup_{t \le L} (X^n_t)^2\}$ is uniformly integrable.

From (21.18) we see that condition (i) is satisfied. Then by Lemma 16.16

(21.21) $(X^n, F^n) \Rrightarrow (X, F)$.

We must show that X^n satisfies the conditions of Theorem 8.22. Condition (a) is immediate, and (b) follows from Lemma 8.6. Applying the continuous mapping theorem for extended weak convergence (Lemma 16.1),

$$(21.22)\qquad (X^n_t - \int_0^t b(X^n_s)\,ds,\ F^n) \Rightarrow (M,F)$$

$$(21.23)\qquad ((X^n_t - \int_0^t b(X^n_s)\,ds)^2 - \int_0^t a(X^n_s)\,ds\ ,\ F^n) \Rightarrow (S,F),$$

for M, S as in (8.20). Because b(x) is bounded, (21.20) implies

$$((X^n_t - \int_0^t b(X^n_s)\,ds)^2,\ F^n) \quad \text{is uniformly of class (DL)}$$

in the language of Section 18. Now we can apply Proposition 18.2 (with p = 2) to (21.22) and deduce that there exist processes M^n, N^n adapted to F^n such that conditions (c) and (e) of Theorem 8.22 hold, and such that

$$(21.24)\qquad \{(M^n)^2\} \text{ is uniformly of class (DL).}$$

By (21.23) and Lemma 16.16,

$$(21.25)\qquad ((M^n_t)^2 - \int_0^t a(X^n_s)\,ds\ ,\ F^n) \Rightarrow (S,F).$$

Apply Proposition 18.2 (with p = 1) to (21.25) and (21.24), and deduce there exist processes S^n and $\widetilde{N}^n$ adapted to F^n such that conditions (d) and (f) of Theorem 8.22 hold.

CONVERGENCE TO THE POISSON PROCESS.

For the second illustration of the bootstrap technique we consider convergence of point processes to the Poisson process. The essence of the argument is almost identical to the argument for Theorem 21.2, but some preliminaries are different.

As in section 9, let (N^n) be sequence of point processes with compensators (A^n), and let N be the Poisson process of rate λ. Note that because N^n is an increasing process the assertions

(21.26)(i) (N^n) is uniformly of class (DL),

(ii) (N^n_t) is uniformly integrable, for each t,

are equivalent.

(21.27) THEOREM. Let (N^n) be a sequence of point processes satisfying (21.26). In order that $N^n \Rightarrow N$ it is necessary and sufficient that $A^n_t \underset{p}{\to} \lambda t$ for each t.

Proof. Necessity is immediate from Theorem 19.3, so we need only prove sufficiency. Proposition 9.1 establishes $N^n \Rightarrow N$. Before starting to prove extended weak convergence, we make two observations.

First, by Lemma 8.6 $\sup_{t \le L} | A^n_t - \lambda t | \underset{p}{\to} 0$. But $(A^n_L : n \ge 1)$ is uniformly integrable because $EA^n_L = EN^n_L \to EN_L = \lambda L$, and so

(21.28) $$\sup\{E|A^n_T - \lambda T| : T \in \mathcal{T}^n_L\} \to 0.$$

Second, for $y \in R$ define $\Delta_y \in \Pi$ to be the distribution of $(y + N_t)_{t \ge 0}$. Then the prediction process Z of N is given by

(21.29) $$Z_t = \tilde{\theta}(N, t, \Delta_{N_t})$$

for $\tilde{\theta}$ as in (20.11), using Lemma 20.12 to verify this process is Skorohod.

The prediction processes Z^n of N^n satisfy

(21.30) $\quad Z^n_t = \tilde{\theta}(N^n, t, \hat{Z}^n_t)$ a.s., for each t,

where $\hat{Z}^n_t$ is a regular conditional distribution for $(N^n_{t+u})_{u \geq 0}$ given $\mathcal{F}^n_t$.

To prove extended weak convergence, start by using the Skorohod representation theorem to suppose

(21.31) $\quad N^n \to N$ in $D(R)$ a.s.

We want to conclude

(21.32) $\quad (N^n, Z^n) \underset{p}{\to} (N, Z)$ in $D(R \times \Pi)$.

By analogy with the proof of Theorem 21.2, one might suppose it would suffice to prove $Z^n_t \underset{p}{\to} \Delta_{N_t}$; but unfortunately Proposition 16.3 is not available here.

Instead we use the following technical argument, which the reader may well omit.

LEMMA. <u>It suffices to prove</u>

(21.33) $\quad \sup\{\rho(Z^n_t, \Delta_{N^n_t}) : t \text{ rational}, t \leq L\} \underset{p}{\to} 0$

<u>where</u> ρ <u>is a metrisation of</u> Π.

<u>Proof</u>. By passing to a subsequence we may assume a.s. convergence in (21.33). Fix ω such that convergence holds in (21.33) and (21.31). Let (λ_n) be a scaling sequence for $(N^n(\omega))$: we shall prove (λ_n) is also a scaling sequence for $(Z^n(\omega))$, and then (21.32) follows. Note that the definition (2.2) of "$f_n \to f$ in (S,d) with scaling sequence (λ_n)" is equivalent to:

(21.34) $\quad d(f_n(t_n), f(\lambda_n^{-1}(t_n))) \to 0$ for any bounded sequence (t_n),

and it suffices to verify (21.34) for rational t_n. So to prove (λ_n) is a scaling sequence for $(Z^n(\omega))$ it suffices, using (21.29) and (21.30), to prove

(21.35) $\quad \rho\{\tilde{\theta}(N^n(\omega), t_n, Z^n_{t_n}(\omega)), \tilde{\theta}(N(\omega), \lambda_n^{-1}(t_n), N_{\lambda_n^{-1}(t_n)})\} \to 0$

for bounded rational t_n. Now omitting ω for typographical convenience,

$$\rho(Z^n_{t_n}, \Delta_{N_{\lambda_n^{-1}(t_n)}}) \leq \rho(Z^n_{t_n}, \Delta_{N^n_{t_n}}) + \rho(\Delta_{N^n_{t_n}}, \Delta_{N_{\lambda_n^{-1}(t_n)}})$$

and the first term tends to zero by (21.33), th second by (21.31) and (21.34). To prove (21.35), pass to a subsequence in which either $\lambda_n^{-1}(t_n) \downarrow t$ or $\lambda_n^{-1}(t_n) \uparrow\uparrow t$, and apply Lemma 20.12.

Note that to prove (21.33) it suffices to prove

(21.36) $\rho(Z^n_{T_n}, \Delta_{N^n_{T_n}}) \underset{p}{\to} 0$

for each sequence (T_n) of stopping times on F^n with $T_n \leq L$. Fix such a sequence (T_n) for the rest of the proof. Think of (21.36) as the analogue of (21.7). The rest of the proof closely follows the proof of Theorem 21.2, so we leave the reader to provide most of the details.

First, we need the analogue of (21.8). Let Π_0 denote the set of distributions of point processes. We shall define a map $H: \Pi_0 \to [0,\infty]$ such that

(21.37) if $(\mu_n) \in \Pi_0$, $y \in R$, $H(\mu_n) \to 0$, then $\mu_n \to \Delta_y$.

To do this, let Θ denote a collection $(\Omega, \mathcal{F}, P, N, A, F)$, where A_t is the compensator of a point process N_t with respect to a filtration F on a probability triple $(\Omega, \mathcal{F}, P)$. Define

$$\alpha(\Theta, L) = \sup\{E|A_T - \lambda T| : T \in \mathcal{T}_L\}$$

$$\delta(\Theta) = \int_0^\infty e^{-L} \wedge \alpha(\Theta, L)\, dL$$

$$H(\mu) = \inf\{\delta(\Theta) : \mathcal{L}(N) = \mu\}.$$

Now (21.37) follows from Proposition 9.1, because hypothesis (21.28) implies the hypothesis of that Proposition.

Here is the analogue of Lemma 21.9, with almost identical proof.

(21.38) LEMMA. Let (ξ_n) be random elements of Π_0 such that

(i) $\pi_0(\xi_n) = \Delta_{N^n_{T_n}}$ a.s.

Suppose that for each n there exist random elements $(\xi_{n,k})$ such that

(ii) $\xi_{n,k} \xrightarrow[p]{} \xi_n$ in Π_0 as $k \to \infty$;

(iii) $\lim_{n\to\infty} \sup_k EH(\xi_{n,k}) = 0$

Then $\rho(\xi_n, \Delta_{N^n_{T_n}}) \xrightarrow[p]{} 0.$

To complete the proof of the Theorem, we must show that $\xi_n = \hat{Z}^n_{T_n}$ satisfies the conditions of this Lemma.

CHAPTER 7

22 SAMPLE PATH PROPERTIES OF THE PREDICTION PROCESS

The purpose of this section is to show that, for Skorohod processes, some of the abstract concepts in the general theory of processes can be expressed more concretely in terms of the prediction process. In particular, quasi left continuity and predictability are sample path properties of the prediction process. These results are somewhat divorced from our main theme of weak convergence. But, as observed in Section 15, the "structure-preserving" feature of extended weak convergence is a consequence of structural properties of X appearing as sample path properties of Z.

In this section we presuppose greater familiarity with the general theory of processes than we suppose elsewhere.

Let F be a fixed filtration. Suppose X is a <u>measurable</u> process (DM.IV.3), not necessarily Skorohod or adapted. If X is real-valued and bounded or positive, then (DM.VI.43) it has a <u>predictable projection</u>, that is to say there exists a process Y such that

(a) Y is predictable;

(b) $Y_T = E(X_T | \mathcal{F}_{T-})$, T predictable.

This process Y is unique, up to indistinguishability. Unsurprisingly, there is an analogue of this result in which "conditional expectation" is replaced by "conditional distribution".

(22.1) <u>Definition</u>. Let X be a measurable S-valued process. Call a measurable $\mathcal{P}(S)$-valued process (ξ_t) an <u>extended predictable projection</u> of X onto F if

(a) ξ is predictable;

(b) ξ_T is a regular conditional distribution for X_T given $\mathcal{F}_{T-}$, for each predictable T.

(22.2) LEMMA. The extended predictable projection exists and is unique, up to indistinguishability. And X is predictable if and only if $P(\xi_t = \delta_{X_t}$ for all $t) = 1$.

Let us write $e: \mathcal{P}(R) \to R$ for the expectation map. When X is real-valued, and bounded or positive, then plainly the (usual) predictable projection Y is obtained from the extended predictable projection ξ by

$$Y_t = e(\xi_t).$$

Proof of Lemma 22.2. For each n partition S into sets $(A_i^n)_{i \geq 1}$ of diameter less than 2^{-n}, choosing (A_i^n) to be a refinement of (A_i^{n-1}). Fix $a_{n,i} \in A_i^n$ and define $\theta_n(s) = a_{n,i}$ for $s \in A_i^n$. We can define a $\mathcal{P}(S)$-valued process ξ^n by specifying

$\xi_t^n(\omega, \{a_{n,i}\})$ is the predictable projection of the process $1_{(X_t \in A_i^n)}$ onto F.

Clearly

(22.3) ξ^n is the extended predictable projection of $\theta_n(X)$.

Let ρ be the metrisation of $\mathcal{P}(S)$ defined by (1.5). Since $d(\theta_n(X), \theta_m(X)) \leq 2^{-n}$ for $m > n$, we can choose (ξ^n) such that

$$\rho(\xi_t^n(\omega), \xi_t^m(\omega)) \leq 2^{-n}, \quad m > n.$$

So $\xi^n \to \xi$, say, uniformly in (t, ω). Since also $\theta_n(X) \to X$ uniformly in (t, ω), it follows easily from (22.3) that ξ is an extended predictable projection of X.

Uniqueness follows from Lemmas 12.5 and 6.12. In the final assertion of the lemma, the "if" part is obvious and the "only if" part follows from the fact (Dellacherie (1972) IV.T31):

if X is predictable and T is predictable then X_T is $\mathcal{F}_{T-}$-measurable.

Let us now return to our usual assumptions that X is Skorohod and adapted to F. Let Z be the prediction process.

(22.4) PROPOSITION. $\pi_t(Z_{t-})$ <u>is the extended predictable projection of</u> X.

From this and Lemma 22.2 we derive two immediate corollaries.

(22.5) COROLLARY. X <u>is predictable if and only if</u> $P(\pi_t(Z_{t-}) = \delta_{X_t}$ <u>for all</u> $t) = 1$.

(22.6) COROLLARY. <u>If</u> X <u>is real-valued, and positive or bounded, then the process</u> $e\pi_t Z_{t-}$ <u>is the predictable projection of</u> X.

<u>Proof of Proposition 22.4.</u> The process Z_{t-} is left-continuous and hence is predictable. So the process $\pi_t(Z_{t-})$ is predictable. Let T be a predictable stopping time. We must verify condition (b) of (22.1), that is

(22.7) $\pi_T(Z_{T-})$ is a r.c.d. for X_T given $\mathcal{F}_{T-}$.

Since Z is Skorohod, Z_{T-} is $\mathcal{F}_{T-}$-measurable.

For $0 \leq a < b < \infty$ define $\pi_{a,b}: \Pi \rightarrow \mathcal{P}(S)$ by:

$\pi_{a,b}(\mathcal{L}(Y)) = \mathcal{L}(Y_\alpha)$, where α is uniform on $[a,b]$ and independent of Y.

The following continuity properties can be established by the usual Skorohod representation technique.

(22.8) The map $(\Theta, a, b) \rightarrow \pi_{a,b}(\Theta)$ is continuous at each point (Θ, a, b) such that $a < b$.

(22.9) $\pi_{a,b}(\Theta) \rightarrow \pi_a(\Theta)$ as $b \downarrow a$.

Let U be distributed uniformly on $[0,1]$, independent of F. Fix $\epsilon > 0$. For each t,

$\pi_{t,t+\varepsilon}(Z_t)$ is a r.c.d. for $X_{t+U\varepsilon}$ given $\mathcal{F}_t$.

So if R is a stopping time taking discrete values,

(22.10) $\pi_{R,R+\varepsilon}(Z_R)$ is a r.c.d. for $X_{R+U\varepsilon}$ given $\mathcal{F}_R$.

Since T is predictable, we can find a sequence (R_n) of discrete-valued stopping times announcing T (DM.IV.77). Then $\mathcal{F}_{R_n} \uparrow \mathcal{F}_{T-}$, $Z_{R_n} \to Z_{T-}$ and $X_{R_n+U\varepsilon} \to X_{T+U\varepsilon}$ a.s.. So by (22.8) $\pi_{R_n,R_n+\varepsilon}(Z_{R_n}) \to \pi_{T,T+\varepsilon}(Z_{T-})$ a.s..

Using Lemma 12.5, we obtain from (22.10)

$\pi_{T,T+\varepsilon}(Z_{T-})$ is a r.c.d. for $X_{T+U\varepsilon}$ given $\mathcal{F}_{T-}$.

Letting $\varepsilon \to 0$ and using (22.9) we obtain (22.7).

(22.11) COROLLARY. X _is quasi left continuous if and only if_ $P(\pi_t(Z_{t-}) = \delta_{X_t}$ _for all_ $t) = 1$.

Proof. Recall the definition of quasi left continuity:

$X_T = X_{T-}$ a.s., each predictable T.

Since X_{T-} is $\mathcal{F}_{T-}$-measurable, this is equivalent to the assertion:

$\delta_{X_{T-}}$ is a r.c.d. for X_T given $\mathcal{F}_{T-}$, each predictable T.

Since the process $\delta_{X_{t-}}$ is predictable, this is equivalent to the assertion:

$\delta_{X_{t-}}$ is the extended predictable projection of X.

The corollary now follows from Proposition 22.4.

Finally, consider the following property for processes:

(22.12) if T is a stopping time such that $X_T = X_{T-}$ a.s. then T is predictable.

This is a kind of converse of quasi left continuity. It is known that, under wide conditions, Markov processes have this property (Meyer (1967); Chung and Walsh (1974)). The prediction process throws more light on this result.

(22.13) PROPOSITION. Let (X,F) be a process with usual filtration, and let Z be its prediction process. Then Z has property (22.12).

From Proposition 13.8 we deduce the Markov process result.

(22.14) COROLLARY. A Feller process with usual filtration has property (22.12).

Proposition 22.13 could be deduced from the Markov process result, since it is not hard to show that Z is Markov. But it is simpler to give a direct proof.

Proof of Proposition 22.13. For $\phi \in C(D(S))$ define ϕ^* as in (14.10). Let T be a fixed stopping time such that $Z_T = Z_{T-}$ a.s.. Then

(22.15) $\phi^*(Z_T) = \phi^*(Z_{T-})$ a.s..

We must prove T is predictable. By (DM.VI.62) it suffices to prove that $M_T = M_{T-}$ a.s. for each bounded martingale (M_t). Since F is the usual filtration, any bounded martingale is of the form $M_t = E(\Theta(X)|\mathcal{F}_t)$ for some bounded measurable $\Theta: D(S) \to R$. By Lemma there exist $\phi_n \in C(D(S))$ such that $E|\Theta(X) - \phi_n(X)|^2 \leq 2^{-n}$. So

$$\begin{aligned}\sup_t |M_t - \phi_n^*(Z_t)| &= \sup_t |E(\Theta(X)|\mathcal{F}_t) - E(\phi_n(X)|\mathcal{F}_t)| \\ &\leq \sup_t E(|\Theta(X) - \phi_n(X)|\,|\mathcal{F}_t) \\ &\to 0 \text{ a.s. by (7.1).}\end{aligned}$$

Now (22.15) shows that $M_T = M_{T-}$ a.s., as required.

References

ALDOUS, D. J.: Two topics in probability theory. Ph.D. thesis, University of Cambridge (1977).

ALDOUS, D. J.: Stopping times and tightness. Ann. Probability 6, 335-340 (1978).

BAXTER, J. R., CHACON, R. V.: Compactness of stopping times. Z. Wahrscheinlichkeitstheorie verw. Geb. 40, 169-182 (1977).

BILLINGSLEY, P.: Convergence of probability measures. New York: Wiley (1968).

BILLINGSLEY, P.: Weak convergence of measures: applications in probability. S.I.A.M. Regional conference series, 5 (1971).

BROWN, B. M.: Martingale central limit theorems. Ann. Math. Statist. 42, 59-66 (1971).

BROWN, T. C.: Martingales and point process convergence. Ann. Probability 6, 615-628 (1978).

CANE, V. R.: A class of non-identifiable stochastic models. J. Appl. Prob. 14, 475-482 (1977).

CHUNG, K. L., WALSH, J. B.: Meyer's theorem on predictability. Z. Wahrscheinlichkeitstheorie verv. Geb. 29, 253-256 (1974).

DELLACHERIE, C.: Capacités et processus stochastiques. Berlin: Springer-Verlag (1972).

EAGLESON, G. K., WEBER, N. C.: Limit theorems for weakly exchangeable arrays. Proc. Cambridge Philos. Soc., to appear.

GAENSSLER, P., STUTE, W.: Wahrscheinlichkeitstheorie. Berlin: Springer-Verlag (1977).

HARLAMOV, B. P.: On convergence of semi-Markov walks to a continuous semi-Markov process. Theory Probability Appl. 21, 482-498 (1976).

IBRAGIMOV, I. A.: A note on the central limit theorem for dependent random variables. Theory Probability Appl. 20, 135-141 (1975).

JACOD, J., MEMIN, J.: Caractéristiques locales et conditions de continuité absolue pour les semimartingales. Z. Wahrscheinlichkeitstheorie verw. Geb. 35, 1-37 (1976).

JACOD, J.: A general theorem of representation for martingales. In: Proceedings of Symposia in Pure Mathematics 31, pp. 37-53. Providence: American Mathematical Society (1977).

KNIGHT, F. B.: A predictive view of continuous time processes. Ann. Probability 3, 573-596 (1975).

KNIGHT, F. B.: On prediction processes. In: A.M.S. Proceedings of Symposia in Pure Mathematics 31 (1977).

LINDVAAL, T.: Weak convergence on the function space $D[0,\infty)$. J. Appl. Prob. 10, 109-121 (1973).

LOYNES, R. M.: A criterion for tightness for a sequence of martingales. Ann. Probability 4, 859-862 (1976).

MCLEISH, D. L.: Dependent central limit theorems and invariance principles. Ann. Probability 2, 620-628 (1974).

MEYER, P. A.: Probability and potentials. Waltham, Mass: Blaisdell (1966).

MEYER, P. A.: Processus de Markov. Springer Lecture Notes in Mathematics 26 (1967).

MEYER, P. A.: La théorie de la prediction de F. Knight. In: Séminaire de Probabilités X. Springer Lecture Notes in Mathematics 511 (1976).

MEYER, P. A., YOR, M.: Sur la théorie de la prediction, et le probleme de decomposition des tribus $\mathcal{F}^{0}_{t+}$. In: Séminaire de Probabilités X. Springer Lecture Notes in Mathematics 511 (1976).

MEYER, P. A.: Convergence faible et compacité des temps d'arrêt, d'après Baxter et Chacòn. In: Séminaire de Probabilités XII, pp. 411-423. Springer Lecture Notes in Mathematics 649 (1978).

MURALI RAO, K.: On decomposition theorems of Meyer. Math. Scand. 24, 66-78 (1969).

PROHOROV, Y. V.: Probability distributions in functional spaces. Uspehi Matem. Nauk (N.S.) 55, 1967 (1953).

PROHOROV, Y. V.: Convergence of random processes and limit theorems in probability theory. Theory Probability Appl. 1, 157-214 (1956).

REBOLLEDO, R.: La méthode des martingales appliquée à l'étude de la convergence en loi de processus. To appear.

SKOROHOD, A. V.: Limit theorems for stochastic processes. Theory Probability Appl. 1, 261-290 (1956).

STONE, C.: Weak convergence of stochastic processes defined on semi-infinite time intervals. Proc. Amer. Math. Soc. 14, 694-696 (1963).

STRICKER, C., YOR, M.: Calcul stochastique dépendant d'un paramètre. Z. Wahrscheinlichkeitstheorie verw. Geb. 45, 109-133 (1978).

WHITT, W.: Weak convergence of probability measures on the function space $D[0,\infty)$. Technical report, Yale University (1971).