WEAK LIMIT THEOREMS FOR STOCHASTIC INTEGRALS AND STOCHASTIC DIFFERENTIAL EQUATIONS

by

Thomas G. Kurtz
Departments of Mathematics
and Statistics
University of Wisconsin–Madison
Madison, Wisconsin 53706

Philip Protter
Department of Mathematics
Purdue University
West Lafayette, Indiana 47907

Technical Report #89-11

Department of Statistics
Purdue University

May 1989

*

WEAK LIMIT THEOREMS FOR STOCHASTIC INTEGRALS AND STOCHASTIC DIFFERENTIAL EQUATIONS

by

Thomas G. Kurtz
Departments of Mathematics and Statistics
University of Wisconsin–Madison
Madison, Wisconsin 53706

Philip Protter
Department of Mathematics
Purdue University
West Lafayette, Indiana 47907

Abstract

Let X^n denote a sequence of adapted processes with paths with one right continuous and left limits. Let Y^n be a sequence of semimartingales. Simple sufficient conditions are given so that when (X^n, Y^n) converge weakly to (X, Y), the limit Y is a semimartingale and further $\int X^n_s - dY^n_s$ converges weakly to $\int X_s - dY_s$. Analogous results are given for stochastic differential equations. Examples are given showing how these results can be applied. Theorems of Jakubowski, Mémin, Pages and Slominski are generalized.

⋆

5/5/89

Weak limit theorems for stochastic integrals and stochastic differential equations

Thomas G. Kurtz
Departments of Mathematics and Statistics
University of Wisconsin-Madison
Madison, WI 53706

Philip Protter
Department of Mathematics
Purdue University
West Lafayette, IN 47907

1. Introduction. For $n = 1,2,\ldots$ let $\{Y_k^n : k \geq 0\}$ be a Markov chain. The classical assumptions leading to a diffusion approximation for such a sequence are that the increments of the chain satisfy

$$E[Y_{k+1}^n - Y_k^n | \mathcal{F}_t^n] = b(Y_k^n)\tfrac{1}{n} + o(\tfrac{1}{n}) \tag{1.1}$$

and

$$E[(Y_{k+1}^n - Y_k^n)^2 | \mathcal{F}_t^n] = a(Y_k^n)\tfrac{1}{n} + o(\tfrac{1}{n}) \tag{1.2}$$

Using these assumptions we can write

$$\begin{aligned} Y_k^n &= Y_0^n + \sum_{i=0}^{k-1}(Y_{i+1}^n - Y_i^n) \\ &= Y_0^n + \sum_{i=0}^{k-1} b(Y_i^n)\tfrac{1}{n} + \sum_0^{\infty}\sqrt{a(Y_i^n)}\, Z_i^n \tfrac{1}{\sqrt{n}} + \text{error} \end{aligned} \tag{1.3}$$

where

$$Z_k^n = \frac{Y_{k+1}^n - Y_k^n - E[Y_{k+1}^n - Y_k^n | \mathcal{F}_t^n]}{\sqrt{E[(Y_{k+1}^n - Y_k^n - E[Y_{k+1}^n - Y_k^n | \mathcal{F}_t^n])^2 | \mathcal{F}_t^n]}} \tag{1.4}$$

are martingale differences with conditional variance 1. If we define $X_n(t) = Y_{[nt]}^n$ and

$$W_n(t) = \frac{1}{\sqrt{n}} \sum_{i=0}^{[nt]} Z_i^n \tag{1.5}$$

then

$$X_n(t) = X_n(0) + \int_0^{\frac{[nt]}{n}} b(X_n(s))\, ds + \int_0^t \sqrt{a(X_n(s-))}\, dW_n(s) + \text{error} \tag{1.6}$$

Under mild additional assumptions, the martingale central limit theorem implies $W_n \Rightarrow W$, (throughout $\Rightarrow$ will denote convergence in distribution) where W is a standard Brownian motion. This convergence suggests that X_n should converge to a solution of the obvious limiting stochastic differential equation. This approach to deriving diffusion approximations has been taken by many authors (see, for example, Skorohod (1965), Chapter 6, Kushner (1974), and Strasser (1986)) although in recent years it has been largely replaced by methods which exploit the characterization of a Markov process as a solution of a martingale problem.

A key step in the application of the stochastic differential equation approach is to show that the sequence of stochastic integrals in the approximating equation converges to the corresponding stochastic integral in the limit. That there is a difficulty to be overcome is well-known from the work of Wong and Zakai (1965). See also Protter (1985).

Growing interest in stochastic differential equations driven by martingales (and more generally semimartingales) other than Brownian motion has led to renewed interest in this approach to the derivation of approximating processes. In addition, functionals of stochastic processes which can be represented by stochastic integrals arise in many areas of application including filtering and statistics. Limit theorems in these settings require conditions under which convergence of the integrand and integrator in a stochastic integral implies convergence of the integral.

Throughout, we will be considering cadlag processes (that is, processes X whose sample paths are right continuous and for which the left limit $X(t-)$ exists at each $t > 0$). This restriction to cadlag processes allows us to define stochastic integrals as limits of Riemann-Stieltjes-like sums, that is,

$$\int_0^t X(s-)\,dY(s) = \lim \sum X(t_i)(Y(t_{i+1}) - Y(t_i)) \tag{1.7}$$

where $\{t_i\}$ is a partition of $[0,t]$ and the limit is taken as the maximum of $t_{i+1} - t_i$ tends to zero. The integral exists if the limit exists in probability. Recall that the choice of the left end-point of $[t_i,t_{i+1})$ as the argument of X is critical even when Y is a Brownian motion. Indeed in the Brownian differential case, if we take the argument of X to be the midpoint, we obtain the Stratonovich integral. (We will, of course, assume that X is adapted (and hence the left continuous process $X(\cdot-)$ is predictable) and that Y is a semimartingale for the same filtration, but the uninitiated reader can follow much of what is going on without a thorough knowledge of these matters.) Throughout, we will use Protter (1989) as our basic reference for material on semimartingales and stochastic integration. See this volume for details and further references.

The following two examples will help motivate the assumptions of the main theorem.

<u>1.1 Example</u> Let $X = Y = X_n = \chi_{[1,\infty)}$ and $Y_n = \chi_{[1+\frac{1}{n},\infty)}$. Then

$$\int_0^t X_n(s-)\,dY_n(s) = 1 \tag{1.8}$$

but the limiting integral gives

$$\int_0^t X(s-)\,dY(s) = 0 \tag{1.9}$$

<u>1.2 Example</u> Let W be standard Brownian motion, and define W_n so that

$$\tfrac{d}{dt}W_n(t) = n(W(\tfrac{k+1}{n}) - W(\tfrac{k}{n})),\ \ t \in [\tfrac{k}{n},\tfrac{k+1}{n}) \tag{1.10}$$

Then

$$\begin{aligned} &\int_0^t W_n(s-)\,dW_n(s) \\ &= \int_0^t W_n(\tfrac{[ns]}{n})\,dW_n(s) + \int_0^t (W_n(s) - W_n(\tfrac{[ns]}{n}))\,dW_n(s) \\ &= \sum W(\tfrac{k}{n})(W(\tfrac{k+1}{n}) - W(\tfrac{k}{n})) + \sum \int_0^{\frac{1}{n}} (W(\tfrac{k}{n}+s) - W(\tfrac{k}{n}))(W(\tfrac{k+1}{n}) - W(\tfrac{k}{n}))\,ds \\ &\to \int_0^t W(s)\,dW(s) + \tfrac{1}{2}t \end{aligned} \tag{1.11}$$

Example 1.1 is indicative of problems that will arise whenever the integrand and the integrator have discontinuities which "coalesce" in the wrong way. We will avoid these difficulties by requiring that the pair of processes (X_n,Y_n) converge in the Skorohod topology on $D_{\mathbb{R}^2}[0,\infty)$ which is stronger than assuming convergence of each component in $D_{\mathbb{R}}[0,\infty)$. For future reference, let Λ denote the collection of continuous, strictly increasing functions mapping $[0,\infty)$ onto $[0,\infty)$. Recall that for any metric space E a sequence of cadlag, E-valued functions $\{x_n\}$ converges in the Skorohod topology to x, if there exists a sequence $\{\lambda_n\} \subset \Lambda$ such that $x_n \circ \lambda_n(t) \to x(t)$ and $\lambda_n(t) \to t$ uniformly for t in bounded intervals. Note that in Example 1.1, Y_n converges in the Skorohod topology with $E = \mathbb{R}$, but the pair (X_n,Y_n) does not converge in the Skorohod topology with $E = \mathbb{R}^2$, and in general, convergence in the Skorohod topology with $E = \mathbb{R}^2$ excludes the possibility of the type of coalescence of jumps that causes the problem in that example. In particular, for each n, let y_n be piecewise constant, and suppose the number of discontinuities of y_n in a bounded time interval is uniformly bounded in n. Then if $(x_n,y_n) \to (x,y)$ in the Skorohod topology on $D_{\mathbb{R}^2}[0,\infty)$, then

$$(1.12) \qquad \int_0^{\cdot} x_n(s-)\,dy_n(s) \to \int_0^{\cdot} x(s-)\,dy(s)$$

and

$$(1.13) \qquad \int_0^{\cdot} y_n(s-)\,dx_n(s) \to \int_0^{\cdot} y(s-)\,dx(s)$$

in the Skorohod topology on $D_{\mathbb{R}}[0,\infty)$. (Actually, the quadruple consisting of x_n, y_n, and the two integrals converges in $D_{\mathbb{R}^4}[0,\infty)$).

Example 1.2 points to more subtle problems, and we will come back to it when we discuss the hypotheses of the main theorem.

We will formulate the main theorem, Theorem 2.2, in Section 2. This theorem is essentially the same as that given by Jakubowski, Memin, and Pages (1988), but we believe that our formulation and proof are more readily accessible to researchers without extensive expertise in the theory of semimartingales and stochastic integration. Section 3 will be devoted to further examples and applications. Section 4 contains some relative compactness results for stochastic integrals and some variations on the main theorem. Applications to stochastic differential equations will be discussed in Section 5. In particular, we generalize results of Słomiński (1989). Some technical results will be given in Section 6.

2. Weak convergence of stochastic integrals. Throughout we will be making various transformations of the processes involved. We will need to have information about the continuity properties of these transformations, and the following lemma will be useful in obtaining this information.

2.1 Lemma Let E_1 and E_2 be metric spaces, and let $F:D_{E_1}[0,\infty) \to D_{E_2}[0,\infty)$. Suppose $F(x\circ\lambda) = F(x)\circ\lambda$ for all $x \in D_{E_1}[0,\infty)$ and all $\lambda \in \Lambda$. Suppose $x_n(t) \to x(t)$ uniformly for t in bounded intervals implies $F(x_n) \to F(x)$ in the Skorohod topology. Then $x_n \to x$ in the Skorohod topology implies that $F(x_n) \to F(x)$ in the Skorohod topology. If $x_n(t) \to x(t)$ uniformly on bounded intervals implies $F(x_n)(t) \to F(x)(t)$ uniformly on bounded intervals, then $x_n \to x$ in the Skorohod topology implies $(x_n,F(x_n)) \to (x,F(x))$ in the Skorohod topology on $D_{E_1\times E_2}[0,\infty)$.

Proof Suppose $x_n \to x$ in the Skorohod topology. Then there exist $\lambda_n \in \Lambda$ such that $x_n\circ\lambda_n(t) \to x(t)$ and $\lambda_n(t) \to t$ uniformly on bounded intervals. It follows that $F(x_n\circ\lambda_n) \to F(x)$ in the Skorohod topology, so there exist $\eta_n \in \Lambda$ such that $\eta_n(t) \to t$ and $F(x_n\circ\lambda_n)\circ\eta_n(t) \to F(x)(t)$ uniformly on bounded intervals. Since $\lambda_n\circ\eta_n(t) \to t$ and $F(x_n)\circ\lambda_n\circ\eta_n(t) = F(x_n\circ\lambda_n)\circ\eta_n(t) \to F(x)(t)$ uniformly on bounded intervals, it follows that $F(x_n) \to F(x)$ in the Skorohod topology. The last statement is immediate from the definition of the Skorohod topology. □

The following functional gives a good example of an application of the lemma. Fix m, and define $h_\delta:[0,\infty) \to [0,\infty)$ by $h_\delta(r) = (1 - \delta/r)^+$. Define $J_\delta:D_{\mathbb{R}^m}[0,\infty) \to D_{\mathbb{R}^m}[0,\infty)$ by

(2.1)
$$J_\delta(x)(t) = \sum_{s\le t} h_\delta(|x(s) - x(s-)|)(x(s) - x(s-))$$

Lemma 2.1 shows that $x \to J_\delta(x)$ and $x \to x - J_\delta(x)$ are continuous. Consequently, by (1.12), if $(x_n,y_n) \to (x,y)$, then

(2.2)
$$\int_0^\cdot x_n(s-)\,dJ_\delta(y_n)(s) \to \int_0^\cdot x(s-)\,dJ_\delta(y)(s)$$

Let $\{\mathcal{F}_t\}$ be a filtration. A cadlag, $\{\mathcal{F}_t\}$-adapted process Y is a semimartingale if it can be decomposed as $Y = M + A$ where M is an $\{\mathcal{F}_t\}$-local martingale and the sample paths of A have finite variation on bounded time intervals, that is, there exists a sequence of $\{\mathcal{F}_t\}$-stopping times, τ_k, such that $\tau_k \to \infty$ a.s and for each k, $M^{\tau_k} \equiv M(\cdot\wedge\tau_k)$ is a uniformly integrable martingale, and for every $t > 0$, $T_t(A) = \sup \sum|A(t_{i+1}) - A(t_i)| < \infty$ a.s (where the supremum is over partitions of $[0,t]$).

An $\mathbb{R}^m$-valued process is an $\{\mathcal{F}_t\}$-semimartingale, if each component is a semimartingale. Let $\mathbb{M}^{km}$ denote the real-valued, k×m matrices. Throughout, $\int X\,dY$ will denote $\int X(s-)\,dY(s)$.

2.2 Theorem For each n, let (X_n,Y_n) be an $\{\mathcal{F}_t^n\}$-adapted process with sample paths in $D_{\mathbb{M}^{km}\times\mathbb{R}^m}[0,\infty)$, and let Y_n be an $\{\mathcal{F}_t^n\}$-semimartingale. Fix $\delta > 0$ (allowing $\delta = \infty$), and define $Y_n^\delta = Y_n - J_\delta(Y_n)$. (Note that Y_n^δ will also be a semimartingale.) Let $Y_n^\delta = M_n^\delta + A_n^\delta$ be a decomposition of Y_n^δ into an $\{\mathcal{F}_t^n\}$-local martingale and a process with finite variation. Suppose

C2.2(i) For each $\alpha > 0$, there exist stopping times $\{\tau_n^\alpha\}$ such that $P\{\tau_n^\alpha \le \alpha\} \le \frac{1}{\alpha}$ and $\sup_n E[[M_n^\delta]_{t\wedge\tau_n^\alpha} + T_{t\wedge\tau_n^\alpha}(A_n^\delta)] < \infty$.

If $(X_n,Y_n) \Rightarrow (X,Y)$ in the Skorohod topology on $D_{\mathbb{M}^{km}\times\mathbb{R}^m}[0,\infty)$, then Y is a semimartingale with respect to a filtration to which X and Y are adapted, and $(X_n,Y_n,\int X_n\,dY_n) \Rightarrow (X,Y,\int X\,dY)$ in the Skorohod topology on $D_{\mathbb{M}^{km}\times\mathbb{R}^m\times\mathbb{R}^k}[0,\infty)$. If $(X_n,Y_n) \to (X,Y)$ in probability, then the triple converges in probability.

2.3 Remark For $c > 0$, define $\tau_n^c = \inf\{t:|M_n^\delta(t)|\vee|M_n^\delta(t-)| \ge c \text{ or } T_t(A_n^\delta) \ge c\}$. Suppose the following conditions hold.

C2.2(ii) $\{T_t(A_n^\delta)\}$ is stochastically bounded for each $t > 0$.

C2.2(iii) For each $c > 0$, $\sup_n E[M_n^\delta(t\wedge\tau_n^c)^2 + T_{t\wedge\tau_n^c}(A_n^\delta)] < \infty$

Since $\sup_{t\le\alpha}|M_n^\delta(t)| = \sup_{t\le\alpha}|Y_n^\delta(t) - A_n^\delta(t)| \le \sup_{t\le\alpha}|Y_n(t)| + T_\alpha(A_n^\delta)$ is stochastically bounded in n for each α, there exists c_α so that $P\{\tau_n^{c_\alpha} \le \alpha\} \le \frac{1}{\alpha}$. In addition $E[[M_n^\delta]_{t\wedge\tau_n^{c_\alpha}}] = E[(M_n^\delta(t\wedge\tau_n^{c_\alpha})^2]$, and C2.2(i) is satisfied with $\tau_n^\alpha = \tau_n^{c_\alpha}$.

For $\delta < \infty$, C2.2(iii) will usually be immediate since the discontinuities of Y_n^δ are bounded in magnitude by δ (making Y_n^δ a special semimartingale) and there will exist a decomposition with the discontinuities of each term bounded by 2δ (see Jacod and Shiryaev (1987), Lemma I.4.24).

2.4 Remark To see that Y is a semimartingale it is enough to show that Y^δ is a semimartingale. Without loss of generality, we can assume that for $\alpha = 1,2,\ldots,$ $\tau_n^\alpha \le \tau_n^{\alpha+1}$. Let $Y_n^{\delta\alpha} =$

$Y_n^\delta(\cdot\wedge\tau_n^\alpha)$. Then $\{(X_n,Y_n,Y_n^\delta,Y_n^{\delta 1},Y_n^{\delta 2},\ldots,\tau_n^1,\tau_n^2,\ldots)\}$ is relatively compact in $D_{\mathbb{M}^{km}\times\mathbb{R}^m\times\mathbb{R}^m}[0,\infty)\times D_{\mathbb{R}^m}[0,\infty)^\infty\times[0,\infty]^\infty$. Let $(X,Y,Y^\delta,Y^{\delta 1},Y^{\delta 2},\ldots,\tau^1,\tau^2\ldots)$ be some limit point, and let $\{\mathcal{F}_t\}$ be the filtration generated by the limiting processes and random times. For each $T > 0$, let

(2.3) $$V_T(Y_n^{\delta\alpha}) \equiv \sup E[\sum|E[Y_n^{\delta\alpha}(t_{i+1}) - Y_n^{\delta\alpha}(t_i)|\mathcal{F}_t^n]|]$$

where the supremum is over all partitions of $[0,T]$. Then

(2.4) $$\sup_n V_T(Y_n^{\delta\alpha}) \leq \sup_n E[T_{T\wedge\tau_n^\alpha}(A_n^\delta)] < \infty$$

and hence $V_T(Y^{\delta\alpha}) < \infty$ (V_T defined using $\{\mathcal{F}_t\}$). (See for example Meyer and Zheng (1984) Theorem 4 or Kurtz (1989) ???) It follows that $Y^{\delta\alpha}$ is a local $\{\mathcal{F}_t\}$-quasi-martingale and hence an $\{\mathcal{F}_t\}$-semimartingale. But

(2.5) $$Y^\delta(t\wedge\tau^\alpha) = Y^{\delta\alpha}(t) + (Y^\delta(\tau^\alpha) - Y^{\delta\alpha}(\tau^\alpha))\chi_{\{\tau^\alpha\leq t\}}$$

so Y^δ is a local $\{\mathcal{F}_t\}$-semimartingale and hence an $\{\mathcal{F}_t\}$-semimartingale.

2.5 Remark Note that if $\{(X_n,Y_n)\}$ satisfies the conditions of the theorem, then $\{\int X_n\,dY_n\}$ satisfies C2.2(i).

2.6 Remark With reference to Example 1.2, note that $T_t(W_n) = O(\sqrt{n})$.

Proof Let $Z_n = (X_n,Y_n,J_\delta(Y_n),Y_n^\delta)$. Z_n has sample paths in $D_E[0,\infty)$ for $E = \mathbb{M}^{km}\times\mathbb{R}^m\times\mathbb{R}^m\times\mathbb{R}^m$. The limit in (1.13) suggests attempting to approximate X_n by a piecewise constant process. The problem is to find such an approximation that converges in distribution along with X_n (in fact, along with Z_n). Furthermore, the approximation must be adapted to a filtration with respect to which Y_n is a semimartingale. By Lemma 6.1, there exists a (random) mapping $I_\epsilon:D_E[0,\infty) \to D_E[0,\infty)$ such that $|z(t) - I_\epsilon(z)(t)| \leq \epsilon$ for all $z \in D_E[0,\infty)$ and $t \geq 0$, $I_\epsilon(z)$ is a step function, and the mapping $z \to (z,I_\epsilon(z))$ is continuous at z a.s for each $z \in D_E[0,\infty)$. Furthermore, $I_\epsilon(Z_n)$ is adapted to a filtration $\mathcal{G}_t^n = \mathcal{F}_t^n\vee\mathcal{H}$, where $\mathcal{H}$ is independent of $\{\mathcal{F}_t^n\}$ (and hence Y_n will be a $\{\mathcal{G}_t^n\}$-semimartingale. Let X_n^ϵ denote the first, $\mathbb{M}^{km}$-valued component of $I_\epsilon(Z_n)$. Then $|X_n - X_n^\epsilon| \leq \epsilon$, and $(X_n,Y_n,J_\delta(Y_n),Y_n^\delta,X_n^\epsilon) \Rightarrow (X,Y,J_\delta(Y),Y^\delta,X^\epsilon)$.

Define $U_n = \int X_n \, dY_n$ and $U_n^\epsilon = \int X_n^\epsilon \, dY_n^\delta + \int X_n \, dJ_\delta(Y_n)$ with similar definitions for U and U^ϵ. Then it follows as in (1.12) and (1.13) that $(X_n, Y_n, U_n^\epsilon) \Rightarrow (X, Y, U^\epsilon)$ in $D_{\mathbb{M}^{km} \times \mathbb{R}^m \times \mathbb{R}^k}[0,\infty)$. Observing that

$$R_n^\epsilon \equiv U_n - U_n^\epsilon = \int (X_n - X_n^\epsilon) \, dY_n^\delta = \int (X_n - X_n^\epsilon) \, dM_n^\delta + \int (X_n - X_n^\epsilon) \, dA_n^\delta \tag{2.6}$$

we see that for any stopping time τ

$$E[\sup_{s \le t \wedge \tau} |R_n^\epsilon(s)|] \le \epsilon \Big(2E[[M_n^\delta]_{t \wedge \tau}]^{\frac{1}{2}} + E[T_{t \wedge \tau}(A_n^\delta)] \Big) \tag{2.7}$$

with similar estimates holding for $U - U^\epsilon$. Applying C2.2(i), it follows that $(X_n, Y_n, U_n) \Rightarrow (X, Y, U)$.

A review of the proof shows that if convergence in distribution is replaced by convergence in probability in the hypotheses, then convergence in probability will hold in the conclusion. □

The transformation J_δ provides a convenient, continuous way to eliminate the large jumps from Y_n in Theorem 2.2. Occasionally, however, it may be useful to apply some other truncation of the large jumps. For example, if Y_n is a martingale it may be possible to truncate the large jumps in such a way that the truncated process is still a martingale, simplifying the verification of the hypotheses of the theorem. With these possibilities in mind, we state a slightly more general version of the theorem.

<u>2.7</u> <u>Theorem</u> For each n, let (X_n, Y_n) be an $\{\mathcal{F}_t^n\}$-adapted process with sample paths in $D_{\mathbb{M}^{km} \times \mathbb{R}^m}[0,\infty)$, and let Y_n be an $\{\mathcal{F}_t^n\}$-semimartingale. Suppose that $Y_n = M_n + A_n + Z_n$, where M_n is a local $\{\mathcal{F}_t^n\}$-martingale, A_n is an $\{\mathcal{F}_t^n\}$-adapted, finite variation process, and Z_n is constant except for finitely many discontinuities in any finite time interval. Let $N_n(t)$ denote the number of discontinuities of Z_n in the interval $[0,t]$. Suppose $\{N_n(t)\}$ is stochastically bounded for each $t > 0$, and

C2.7 For each $\alpha > 0$, there exist stopping times $\{\tau_n^\alpha\}$ such that $P\{\tau_n^\alpha \le \alpha\} \le \frac{1}{\alpha}$ and $\sup_n E[[M_n]_{t \wedge \tau_n^\alpha} + T_{t \wedge \tau_n^\alpha}(A_n)] < \infty$.

If $(X_n,Y_n,Z_n) \Rightarrow (X,Y,Z)$ in the Skorohod topology on $D_{\mathbb{M}^{km}\times\mathbb{R}^m}[0,\infty)$, then Y is a semimartingale with respect to a filtration to which X and Y are adapted, and $(X_n,Y_n,\int X_n\,dY_n) \Rightarrow (X,Y,\int X\,dY)$ in the Skorohod topology on $D_{\mathbb{M}^{km}\times\mathbb{R}^m\times\mathbb{R}^k}[0,\infty)$. If $(X_n,Y_n,Z_n) \to (X,Y,Z))$ in probability, then convergence in probability holds in the conclusion.

3. Examples and applications

3.1 Example As a simple first example, we consider limit theorems for sums of products of independent random variables which arise in the study of U-statistics. Let $\{\xi_i\}$ be i.i.d. real-valued random variables with mean zero and variance σ^2. Define

(3.1) $$W_n^{(k)}(t) = \frac{1}{n^{k/2}} \sum_{1\le i_1<\cdots i_k\le[nt]} \xi_{i_1}\cdots\xi_{i_k}$$

and $Z_n = (W_n^{(1)},\ldots,W_n^{(m)})$. Note that $W_n^{(1)} \Rightarrow \sigma W$, where W is standard Brownian motion, and observe that we can write

(3.2) $$W_n^{(k)}(t) = \int_0^t W_n^{(k-1)}(s-)\,dW_n^{(1)}(s)$$

It follows (by induction) that $Z_n \Rightarrow Z = (W^{(1)},\ldots,W^{(m)})$, where $W^{(1)} = \sigma W$ and $W^{(k)}$ is the corresponding interated integral. (Note that $X_n \Rightarrow X$ in $D_E[0,\infty)$ implies that $(X_n,X_n) \Rightarrow (X,X)$ in $D_{E\times E}[0,\infty)$).

3.2 Example (Bobkoski (1983)) Let $\{\xi_i\}$ be as above. For a constant ϕ, let $\{Y_k\}$ satisfy

(3.3) $$Y_{k+1} = \phi Y_k + \xi_{k+1}$$

Given $Y_1,\ldots,Y_m$, the least squares estimate $\hat\phi$ for an unknown ϕ is the value of ϕ minimizing $\sum(Y_{k+1} - \phi Y_k)^2$, that is, the solution of

(3.4) $$\sum Y_k(Y_{k+1} - \phi Y_k) = 0$$

given by

(3.5) $$\hat\phi = \frac{\sum Y_k Y_{k+1}}{\sum Y_k^2}$$

Now consider a sequence of such processes $\{Y_k^n\}$ in which the true $\phi_n = (1 - \frac{\beta}{n})$. If we define $X_n(t) = \frac{1}{\sqrt{n}} Y^n_{[nt]}$

(3.6) $$X_n(t) = \phi_n^{[nt]} X_n(0) + \int_0^t \phi_n^{[nt]-1-[ns-]} dW_n(s)$$

where $W_n = W_n^{(1)}$, and if $X_n(0) \to X(0)$, it follows that $X_n \Rightarrow X$ given by

(3.7) $$X(t) = e^{-\beta t} X(0) + \int_0^t e^{-\beta(t-s)} \sigma \, dW(s)$$

Note that X is an Ornstein-Uhlenbeck process satisfying $dX = -\beta X \, dt + \sigma \, dW$. For the least squares estimate of ϕ_n at time t, we have

(3.8) $$\sum_{k=0}^{[nt]-1} Y_k^n \Big((\phi_n - \hat{\phi}_n) Y_k^n + \xi_{k+1} \Big) = 0$$

which implies

(3.9) $$n(\phi_n - \hat{\phi}_n) \int_0^{\frac{[nt]}{n}} X_n(s)^2 \, ds = \int_0^t X_n(s-) \, dW_n(s)$$

and it follows that

(3.10) $$n(\phi_n - \hat{\phi}_n) \Rightarrow \frac{\int_0^t \sigma X(s) \, dW(s)}{\int_0^t X(s)^2 \, ds}$$

More general results along these lines have been given by Llatas (1987) and Cox and Llatas (1989).

3.3 Example Work on approximation of nonlinear filters, DiMasi and Rungaldier (1981, 1982), Johnson (1983), Goggin (1988), involves studying the limiting behavior of a sequence of Girsanov-type densities, each of which typically includes the exponential of a stochastic integral. For example, let $\{X_n\}$ be a sequence of processes with sample paths in $D_E[0,\infty)$, such that $X_n \Rightarrow X$. Let N be a unit Poisson process independent of the X_n, let the observation process Y_n be given by

(3.11) $$Y_n(t) = N\Big(n \int_0^t \Big(\lambda + n^{-\frac{1}{2}} h(X_n(s)) \Big) ds \Big)$$

and define

(3.12) $$U_n(t) = n^{-\frac{1}{2}} (Y_n(t) - \lambda n t)$$

Note that $\mathcal{F}_t^{Y_n} = \mathcal{F}_t^{U_n}$ and observe that $(X_n,U_n) \Rightarrow (X,U)$ where for a standard Brownian motion W independent of X

$$U(t) = \sqrt{\lambda}W(t) + \int_0^t h(X(s))\,ds \tag{3.13}$$

Suppose that (X_n,U_n) is defined on a probability space $(\Omega,\mathcal{F},P_n)$. Then there exists a probability measure Q_n on the same measurable space, $(\Omega,\mathcal{F})$, under which X_n has the same distribution as under P_n, Y_n is independent of X_n and is a Poisson process with parameter $n\lambda$, and $P_n \ll Q_n$ on $\mathcal{G}_t^n = \sigma(X_n(s),U_n(s):s\leq t)$ with

$$\begin{aligned} L_n(t) &= \frac{dP_n}{dQ_n}\bigg|_{\mathcal{G}_t^n} \\ &= \exp\Big\{\int_0^t \ln\Big(1+n^{-\frac{1}{2}}\lambda^{-1}h(X_n(s-))\Big)dY_n(s) - \int_0^t n^{\frac{1}{2}}h(X_n(s))\,ds\Big\} \\ &= \exp\Big\{\int_0^t n^{\frac{1}{2}}\ln\Big(1+n^{-\frac{1}{2}}\lambda^{-1}h(X_n(s-))\Big)dU_n(s) \\ &\qquad + \int_0^t\Big(n\lambda\ln\Big(1+n^{-\frac{1}{2}}\lambda^{-1}h(X_n(s-))\Big) - n^{\frac{1}{2}}h(X_n(s))\Big)ds\Big\} \end{aligned} \tag{3.14}$$

Similarly, if (X,U) is defined on a probability space $(\Omega,\mathcal{F},P)$, there exists a measure Q on $(\Omega,\mathcal{F})$ such that, under Q, X has the same distribution as under P, U is independent of X with the same distribution as $\sqrt{\lambda}W$, and $P \ll Q$ on $\mathcal{G}_t = \sigma\{X(s),U(s):s\leq t\}$ with

$$L(t) = \frac{dP}{dQ}\bigg|_{\mathcal{G}_t} = \exp\Big\{\int_0^t \lambda^{-1}h(X(s))\,dU(s) - \int_0^t \tfrac{1}{2}\lambda^{-1}h^2(X(s))\,ds\Big\} \tag{3.15}$$

Expanding the logarithm in (3.14) in a Taylor series and applying Theorem 2.2, we see that $L_n \Rightarrow L$ under $\{P_n\}$, P and under $\{Q_n\}$, Q.

Results of Goggin (1988) can then be applied to show that the conditional distribution $\mu_n(t)$ of $X_n(t)$ given $\mathcal{F}_t^{Y_n}$ converges in distribution to the conditional distribution $\mu(t)$ of $X(t)$ given $\mathcal{F}_t^U$ as a process in $D_{\mathcal{P}(E)}[0,\infty)$.

3.4 Example (Meyer (1989), Emery (1989)) Next we consider the problem of showing existence of solutions of the structure equation arising in the study of chaotic representations formulated by Meyer. Given $F \in C(\mathbb{R})$, the problem is to show existence of a martingale X satisfying

$$d[X]_t = dt + F(X(t-))dX(t) \tag{3.16}$$

or, equivalently,

$$X(t)^2 - X(0)^2 - 2\int_0^t X(s-)dX(s) = t + \int_0^t F(X(s-))dX(s) \tag{3.17}$$

Of course, if X is standard Brownian motion, then (3.16) is satisfied for $F(x) = 0$. If X is a martingale with $|X(t)| = \sqrt{t}$, then, obviously from (3.17), (3.16) holds with $F(x) = -2x$. See Protter and Sharpe (1979) and Emery (1989) for a construction of such a martingale. For Azema's martingale (Protter (1989) §IV.6), $F(x) = -x$.

Following Meyer (1989), we define a sequence of discrete time martingales and show that the sequence is relatively compact and that the limit satisfies (3.16). Setting $\Delta Y_n(k) = Y_n(k+1) - Y_n(k)$ and assuming for simplicity that $Y_n(0) = 0$, the discrete time analogue of (3.16) becomes

$$\Delta Y_n(k)^2 = \tfrac{1}{n} + F(Y_n(k))\Delta Y_n(k) \tag{3.18}$$

Consequently,

$$\Delta Y_n(k) = \frac{F(Y_n(k)) \pm \sqrt{F(Y_n(k))^2 + \frac{4}{n}}}{2} \equiv \Delta_n^{\pm}(k) \tag{3.19}$$

and since we want Y_n to be a martingale, we must have

$$P\{\Delta Y_n(k) = \Delta_n^{+}(k)\} = 1 - P\{\Delta Y_n(k) = \Delta_n^{-}(k)\} = \frac{\Delta_n^{-}(k)}{\Delta_n^{-}(k) - \Delta_n^{+}(k)} \tag{3.20}$$

Define $X_n(t) = Y_n([nt])$. Note that $E[X_n(t)^2] = \frac{[nt]}{n}$ and more generally

$$E[(X_n(t+h) - X_n(t))^2|\mathcal{F}_t^{X_n}] = \frac{[n(t+h)]}{n} - \frac{[nt]}{n} \tag{3.21}$$

The relative compactness of $\{X_n\}$ (and hence for $\{(X_n, F \circ X_n)\}$) follows easily. (See, for example, Ethier and Kurtz (1986), Remark 3.8.7.) Since X_n satisfies

$$X_n(t)^2 - X_n(0)^2 - 2\int_0^t X_n(s-)\,dX_n(s) = \frac{[nt]}{n} + \int_0^t F(X_n(s-))\,dX_n(s) \tag{3.22}$$

we see that any limit point of the sequence $\{X_n\}$ satisfies (3.17).

More generally, the above construction will give solutions of

$$d[X]_t = dt + F(X,t-)\,dX(t) \tag{3.23}$$

for any $F:D_{\mathbb{R}}[0,\infty) \to D_{\mathbb{R}}[0,\infty)$ satisfying C5.2(ii) and C5.2(iii) below and $F(x,t) = F(x^t,t)$ for all $x \in D_{\mathbb{R}}[0,\infty)$ and $t \geq 0$ where $x^t = x(\cdot \wedge t)$.

<u>3.5</u> <u>Example</u> (Neuhaus (1977)) Let $\xi_1, \xi_2, \ldots$ be i.i.d. uniform-[0,1] random variables, and let h be a measurable, symmetric function defined on $[0,1]\times[0,1]$ satisfying

$$\int_0^1\int_0^1 h^2(x,y)\,dx\,dy < \infty \tag{3.24}$$

and

$$\int_0^1 h(x,y)\,dx = \int_0^1 h(x,y)\,dy = 0 \tag{3.25}$$

Define

$$Z_n^h = \frac{1}{n} \sum_{1\leq i<j\leq n} h(\xi_i, \xi_j) \tag{3.26}$$

Then $\{Z_n^h\}$ is asymptotically Gaussian. To see that this is the case and to identify the limit, we follow a suggestion of Lajos Horvath and represent (3.26) in terms of the empirical distribution function F_n

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} \chi_{[\xi_i,\infty)}(t) \tag{3.27}$$

In terms of F_n, Z_n^h can be written

(3.28) $$Z_n^h = n \iint_{s<t} h(s,t)\, dF_n(s)\, dF_n(t)$$

and defining $B_n(t) = \sqrt{n}\,(F_n(t) - t)$, the symmetry of h and (3.25) give

(3.29) $$Z_n^h = \iint_{s<t} h(s,t)\, dB_n(s)\, dB_n(t)$$

If g satisfies the same conditions as h, then

(3.30) $$E[(Z_n^h - Z_n^g)^2] = \frac{n(n-1)}{2n^2} \int_0^1 \int_0^1 (h(x,y) - g(x,y))^2\, dx\, dy$$

Since any $h \in L^2([0,1]\times[0,1])$ can be approximated by smooth g, we may as well assume that h is continuously differentiable. Under this assumption we can write

(3.31) $$X_n(t) = \int_0^t h(s,t)\, dB_n(s) = h(t,t)\, B_n(t) - \int_0^t h_s(s,t) B_n(s)\, ds$$

and, since $B_n \Rightarrow B$, the Brownian bridge, (see, for example, Billingsley (1968), §13 and §19, or Protter (1989) §V.6), the continuous mapping theorem implies that $X_n \Rightarrow X$ given by

(3.32) $$X(t) = \int_0^t h(s,t)\, dB(s)$$

More precisely, $(X_n, B_n) \Rightarrow (X,B)$ in $D_{\mathbb{R}\times\mathbb{R}}[0,\infty)$.

The process B_n is a semimartingale with decomposition

(3.33) $$\begin{aligned} B_n(t) &= \sqrt{n}\,(F_n(t) - t) = \sqrt{n}\Big(F_n(t) - \int_0^t \frac{1 - F_n(s)}{1-s}\, ds\Big) - \sqrt{n}\int_0^t \frac{F_n(s) - s}{1-s}\, ds \\ &= M_n(t) - \int_0^t \frac{1}{1-s} B_n(s)\, ds \end{aligned}$$

Note that $E[M_n(t)^2] = E[[M_n]_t] = t$. In fact, $[M_n]_t \to t$, implying, by the martingale central theorem, that $M_n \Rightarrow W$ and yielding, in the limit, the classical stochastic differential equation for B. For this decomposition we have

$$(3.34) \qquad E\left[T_t\left(\int_0^{\cdot} \frac{1}{1-s} B_n(s)\,ds\right)\right] = E\left[\int_0^t \frac{1}{1-s}|B_n(s)|\,ds\right]$$
$$\leq \int_0^t \frac{1}{1-s}\sqrt{E[B_n(s)^2]}\,ds = \int_0^t \sqrt{\frac{s}{1-s}}\,ds < \infty$$

for $t \leq 1$. Consequently, the conditions of Theorem 2.2 are satisfied, and Z_n^h converges in distribution to

$$(3.35) \qquad Z^h = \int_0^1\int_0^t h(s,t)\,dB(s)\,dB(t)$$

For related results see Hall (1979). Rubin and Vitale (1980) and Dynkin and Mandelbaum (1983) consider more general symmetric statistics. Rubin and Vitale represent the limiting random variables as series of products of Hermite polynomials of Gaussian random variables. Dynkin and Mandelbaum represent the limits as multiple Wiener integrals. These higher order limit theorems can also be obtained by the techniques used above with the limiting random variables represented as multiple integrals of B. Filippova (1961) obtained limits represented as multiple integrals of Brownian bridge in special cases. □

3.6 Example (Duffie and Protter (1989)) Theorem 2.2 is useful in the derivation and justification of models in continuous time finance theory as limiting cases of discrete time models. For example, let the sequence of random variables $\xi_1^n, \xi_2^n, \ldots$ denote the periodic rate of return on a security with initial price S_0. After k periods the price of the security will be

$$(3.36) \qquad S_k^n = S_0^n \prod_{i=1}^{k} (1 + \xi_i^n)$$

Let $Y_n(t) = \sum_{i \leq [nt]} \xi_i^n$ and $S_n(t) = S_{[nt]}^n$. Noting that $S_{k+1}^n - S_k^n = S_k^n \xi_k^n$, we can write

$$(3.37) \qquad S_n(t) = S_n(0) + \int_0^t S_n(s-)\,dY_n(s)$$

If θ_k^n units of the security are held during the (k+1)th period, the financial gain for the period is $\theta_k^n(S_{k+1}^n - S_k^n)$, and the cumulative gain up to time t can be written

$$(3.38) \qquad G_n(t) = \int_0^t \theta_n(s-)\,dS_n(s)$$

where $\theta_n(t) = \theta_{[nt]}^n$. Suppose that $\{Y_n\}$ satisfies C2.2(i) for some δ and that $(Y_n, \theta_n, S_n(0)) \Rightarrow$

$(Y,\theta,S(0))$ (in $D_{\mathbb{R}^2}[0,\infty)\times\mathbb{R}$). Then the limiting equation

$$S(t) = S(0) + \int_0^t S(s-)\,dY(s) \tag{3.39}$$

has a (locally) unique global solution, so by Theorem 5.4 below (see also Avram (1988)), $S_n \Rightarrow S$. (More precisely $(Y_n,\theta_n,S_n) \Rightarrow (Y,\theta,S)$.) It follows that $\{S_n\}$ also satisfies C2.2(i), so that $G_n \Rightarrow G$ given by

$$G(t) = \int_0^t \theta(s-)\,dS(s) \tag{3.40}$$

The solution of (3.39) with $S(0) = 1$ is called the stochastic or Doléans-Dade exponential and is denoted $\mathcal{E}(X)$. The general solution is then given by $S = S(0)\mathcal{E}(X)$. (Protter (1989) §II.8.)

4. Relative compactness and additional convergence results

4.1 Proposition Let $\{(U_n,Y_n)\}$ be relatively compact (in the sense of convergence in distribution) in $D_{\mathbb{R}^k\times\mathbb{R}^m}[0,\infty)$ with (U_n,Y_n) adapted to $\{\mathcal{F}_t^n\}$, and $\{Y_n\}$ satisfying C2.2(i) for some $\delta > 0$. Suppose that X_n has sample paths in $D_{\mathbb{M}^{km}}[0,\infty)$ and is adapted to $\{\mathcal{F}_t^n\}$. Let $H_n(t) = \sup_{s\le t}|X_n(s)|$, and suppose that $\{H_n(t)\}$ is stochastically bounded for each t. Define

$$Z_n(t) = U_n(t) + \int_0^t X_n(s-)dY_n(s) \tag{4.1}$$

Then $\{(U_n,Y_n,Z_n)\}$ is relatively compact in $D_{\mathbb{R}^k\times\mathbb{R}^m\times\mathbb{R}^k}[0,\infty)$.

4.2 Remark The result will also hold under the assumption that X_n is predictable and H_n is a right continuous, adapted, increasing process satisfying $|X_n(s)| \le H_n(t)$ for $s \le t$ with the usual extension of the stochastic integral to predictable integrands. This result is very close to part (ii) of Theorem 2.3 in Jacod, Memin, and Metivier (1983).

Proof The relative compactness of $\{(U_n,Y_n,\int X_n\,dJ_\delta(Y_n))\}$ is immediate. Since the stochastic integral on the right of (4.1) has a discontinuity only when Y_n has a discontinuity, and $\{(U_n,Y_n)\}$ is relatively compact, the proposition will follow if we show that $\{\int X_n\,dY_n^\delta\}$ is relatively compact (see, for example, Kurtz (1989), Lemma 2.2). By the same observation, we can, in fact, treat the summands in the matrix multiplication separately. Consequently, to simplify notation we will assume that $k = m = 1$.

Define $\eta_n^b = \inf\{t:H_n(t) \ge b\}$. Let φ be C^2, convex and symmetric with $\varphi(0) = \varphi'(0) = 0$, $\varphi''(0) = 1$, φ'' decreasing, and $\varphi''(1) = 0$. Let $G_n = \chi_{[0,\eta_n^b)}X_n$, $V_n = \int G_n dY_n^\delta$, and $W_n = \int G_n\,dJ_\delta(Y_n)$. Then setting $\Delta V_n(s) = V_n(s) - V_n(s\text{-})$ and $[Y_n^\delta]_t^c = [Y_n^\delta]_t - \sum_{s\le t}\Delta Y_n^\delta(s)^2$ (note that $[Y_n^\delta]^c = [M_n^\delta]^c$), Ito's formula gives

$$(4.2)\quad \varphi(V_n(t)-V_n(t_0))$$

$$= \int_{t_0}^t \varphi'(V_n(s\text{-})-V_n(t_0))\,G_n(s\text{-})\,dY_n^\delta(s) + \int_{t_0}^t \tfrac{1}{2}\varphi''(V_n(s\text{-})-V_n(t_0))G_n(s\text{-})^2\,d[Y_n^\delta]_s^c$$

$$+ \sum_{t_0<s\le t}\Big(\varphi(V_n(s)-V_n(t_0)) - \varphi(V_n(s\text{-})-V_n(t_0)) - \Delta V_n(s)\,\varphi'(V_n(s\text{-})-V_n(t_0))\Big)$$

$$\leq \int_{t_0}^{t} \varphi'(V_n(s\text{-}) - V_n(t_0))\, G_n(s\text{-})\, dM_n^\delta(s)$$

$$+ b(T_t(A_n^\delta) - T_{t_0}(A_n^\delta)) + \tfrac{1}{2}b^2([M_n^\delta]_t^c - [M_n^\delta]_{t_0}^c) + \sum_{t_0<s\leq t} 4\varphi\Big(\frac{\Delta V_n(s)}{2}\Big)$$

$$\leq \int_{t_0}^{t} \varphi'(V_n(s\text{-}) - V_n(t_0))\, G_n(s\text{-})\, dM_n^\delta(s)$$

$$+ C\Big[[M_n^\delta]_t - [M_n^\delta]_{t_0} + T_t(A_n^\delta) - T_{t_0}(A_n^\delta)\Big]$$

for some constant C. (The last inequality uses the fact that $\varphi(u) \leq u^2 \wedge |u|$.) Then

(4.3) $$\varphi(V_n(t) - V_n(t_0)) - C\Big[[M_n^\delta]_t - [M_n^\delta]_{t_0} + T_t(A_n^\delta) - T_{t_0}(A_n^\delta)\Big]$$

is a local supermartingale, and, with reference to C2.2(i) and the observation that $\{[M_n^\delta]_t\}$ is stochastically bounded for each t, we see that the conditions of Theorem 1.2 of Kurtz (1989) are satisfied. Let C_n be the dual predictable projection of $t + [M_n^\delta]_t + T_t(A_n^\delta)$, and define $\hat{V}_n(t) = \lim_{s\to t+} V_n(C_n^{-1}(s)\text{-})$, with similar definitions for $\hat{Y}_n$, $\hat{U}_n$, and $\hat{W}_n$. Then $\{(\hat{U}_n,\hat{Y}_n,\hat{V}_n,\hat{W}_n)\}$ is relatively compact in the Skorohod topology, and $\{C_n\}$ is relatively compact in the topology corresponding to convergence at every point of continuity. Let $(\hat{U},\hat{Y},\hat{V},\hat{W},\hat{C})$ be some limit point. Since (U_n,Y_n) converges in distribution in the Skorohod topology, Lemma 2.3 of Kurtz (1989) implies that in any interval on which $\hat{C}^{-1}$ is constant, $(\hat{U},\hat{Y})$ is constant except for at most one jump. The same observation must hold for $(\hat{U},\hat{Y},\hat{V},\hat{W})$, since the jumps of $\hat{V}$ and $\hat{W}$ are bounded by b times the magnitude of the jumps of $\hat{Y}$. Consequently, by Kurtz (1989), Lemma 2.3(b), $\{(U_n,Y_n,V_n,W_n)\}$ is relatively compact in the Skorohod topology, and since b is arbitrary and $\lim_{b\to\infty} \sup_n P\{\eta_b^n \leq T\} = 0$ for each $T > 0$, the proposition follows. □

This general relative compactness result leads to the problem of identifying the limit under more general assumptions on the limiting behavior of $\{X_n\}$ than in Theorem 2.2. First assume that $(X_n,Y_n) \Rightarrow (X,Y)$ in $D_{\mathbb{M}^{km}}[0,\infty)\times D_{\mathbb{R}^m}[0,\infty)$ (rather than in $D_{\mathbb{M}^{km}\times\mathbb{R}^m}[0,\infty)$) and that $\{Y_n\}$ satisfies C2.2(i). For all but countably many $\epsilon > 0$, $(X_n(\cdot-\epsilon),Y_n) \Rightarrow (X(\cdot-\epsilon),Y)$ in $D_{\mathbb{M}^{km}\times\mathbb{R}^m}[0,\infty)$. Consequently, for each such ϵ,

(4.4) $$\int_0^t X_n(s-\epsilon\text{-})\, dY_n(s) \Rightarrow \int_0^t X(s-\epsilon\text{-})\, dY(s)$$

and hence there exists a sequence $\epsilon_n \to 0$ slowly enough such that

(4.5) $$\int_0^t X_n(s-\epsilon_n-)\,dY_n(s) \Rightarrow \int_0^t X(s-)\,dY(s)$$

Noting that $\{\int X_n\,dY_n\}$ is relatively compact by Proposition 4.1, assume that $\int X_n\,dY_n \Rightarrow Z$. Consequently,

(4.6) $$\int_0^{\cdot} (X_n(s-) - X_n(s-\epsilon_n-))\,dY_n(s) \Rightarrow Z(\cdot) - \int_0^{\cdot} X(s-)\,dY(s)$$

Note that the sequence on the left in (4.6) is relatively compact by Proposition 4.1.

Let $J_\delta(X_n)$ denote the $\mathbb{M}^{km}$-valued process whose ijth component is $J_\delta(X_n^{ij})$ where X_n^{ij} is the ijth component of X_n, and let $X_n^\delta = X_n - J_\delta(X_n)$. Let $V_n^\delta(t) = \sup_{s\le t}|X_n^\delta(s) - X_n^\delta(s-\epsilon_n)|$. Then $V_n^\delta \Rightarrow V^\delta$ given by $V^\delta(t) = \sup_{s\le t}|X^\delta(s) - X^\delta(s-)| \le \sqrt{km}\,\delta$. By the same type of estimate as in (2.7), to identify the right side of (4.6) it is enough to identify the limit of

(4.7) $$U_n^\delta(t) = \int_0^t \Big(J_\delta(X_n)(s-) - J_\delta(X_n)(s-\epsilon_n-)\Big)dY_n(s)$$

(along a subsequence if necessary) and then to let $\delta \to 0$. Let $\{\tau_{in}^\delta\}$ denote the times of discontinuity of $J_\delta(X_n)$ with $\tau_{0n}^\delta = 0$. Note that $\{\tau_{in}^\delta\}$ are just the times when at least one component of X_n has a discontinuity larger than δ. Then U_n^δ can be written

(4.8) $$\sum_{\tau_{in}^\delta \le t} (Y_n(\tau_{in}^\delta+\epsilon_n) - Y_n(\tau_{in}^\delta))(J_\delta(X_n)(\tau_{in}^\delta) - J_\delta(X_n)(\tau_{in}^\delta-))$$

and any limit point U^δ of $\{U_n^\delta\}$ satisfies

(4.9) $$U^\delta(t) = \sum_{\beta_i^\delta \le t} (J_\delta(X)(\beta_i^\delta) - J_\delta(X)(\beta_i^\delta-))(Y(\beta_i^\delta) - Y(\beta_i^\delta-))$$

where $\{\beta_i^\delta\}$ is some <u>subset</u> of the times at which some component of X has a discontinuity larger that δ. Letting $\delta \to 0$, we see that

$$U(t) \equiv Z(t) - \int_0^t X(s-)\,dY(s) = \sum_{\beta_i^\delta \le t} (Y(\beta_i) - Y(\beta_i-))(X(\beta_i) - X(\beta_i-)) \tag{4.10}$$

where $\{\beta_i\}$ is some subset of the times at which both Y and X have discontinuities. From (4.8) it is clear that $\{\beta_i\}$ is empty unless some discontinuities of Y_n "coalesce" with discontinuities of X_n from above. The following theorem gives conditions under which no such coalescence occurs.

4.3 Theorem For each n, let (X_n,Y_n) be an $\{\mathcal{F}_t^n\}$-adapted process with sample paths in $D_{\mathbb{M}^{km}\times\mathbb{R}^m}[0,\infty)$, and let Y_n be an $\{\mathcal{F}_t^n\}$-semimartingale. Suppose that for some $0 < \delta \le \infty$, C2.2(i) holds and that for all $T > 0$ and $\eta > 0$ there exist random variables $\{\gamma_n^T(\eta)\}$ such that

$$E[1\wedge|Y_n(t+u) - Y_n(t)||\mathcal{F}_t^n] \le E[\gamma_n^T(\eta)|\mathcal{F}_t^n], \qquad 0\le u\le\eta,\ 0\le t\le T \tag{4.11}$$

and $\lim_{\eta\to 0}\overline{\lim}_{n\to\infty} E[\gamma_n^T(\eta)] = 0$.

If $(X_n,Y_n) \Rightarrow (X,Y)$ in $D_{\mathbb{M}^{km}}[0,\infty)\times D_{\mathbb{R}^m}[0,\infty)$, then Y is a semimartingale with respect to a filtration to which X and Y are adapted, and $(X_n,Y_n,\int X_n\,dY_n) \Rightarrow (X,Y,\int X\,dY)$ in $D_{\mathbb{M}^{km}\times\mathbb{R}^m\times\mathbb{R}^k}[0,\infty)$. If $(X_n,Y_n) \to (X,Y)$ in probability, then the triple converges in probability.

4.4 Remark See Ethier and Kurtz (1986) Theorem 3.8.6 and Remark 3.8.7 for the connection of (4.11) to conditions for the relative compactness of $\{Y_n\}$. These conditions imply a type of uniform quasi-left continuity on the sequence $\{Y_n\}$. Consequently, this theorem is related to Theorem 5.1 of Jakubowski, Memin, and Pages (1989).

Proof We need only show that $U \equiv 0$ in (4.10). The inequality in (4.11) holds with t replaced by a stopping time. Consequently we have (with reference to (4.8)) for $\epsilon_n \le \eta$

$$E\left[\sum_{i=1}^m 1\wedge\left|(Y_n(\tau_{in}^\delta\wedge T+\epsilon_n) - Y_n(\tau_{in}^\delta\wedge T))\right|1\wedge\left|(J_\delta(X_n)(\tau_{in}^\delta\wedge T) - J_\delta(X_n)(\tau_{in}^\delta\wedge T-))\right|\right] \tag{4.12}$$

$$\le E\left[\sum_{i=1}^m \gamma_n^T(\eta)1\wedge\left|(J_\delta(X_n)(\tau_{in}^\delta\wedge T) - J_\delta(X_n)(\tau_{in}^\delta\wedge T-))\right|\right]$$

$$\le mE[\gamma_n^T(\eta)]$$

Since the number of discontinuities of $J_\delta(X_n)$ in any finite time interval is stochastically bounded in n, it follows that $U^\delta(t) = 0$ for each $t > 0$. Consequently, $U \equiv 0$ and the theorem follows.

Noting that if a sequence $\{U_n\}$ is defined on a single sample space and $U_n \Rightarrow 0$, then $U_n \to 0$ in probability, we see that convergence in distribution can be replaced by convergence in probability in the statement of the theorem. □

In the next theorem we weaken the assumption that the integrands converge in the Skorohod topology at the cost of adding the requirement that the limiting integrator be continuous. The conditional variation on [0,t] of a process X with respect to a filtration $\{\mathcal{F}_t\}$ is defined by

$$V_t(X) = \sup E[\sum_i |E[X(t_{i+1}) - X(t_i)|\mathcal{F}_{t_i}]|] \tag{4.13}$$

where the supremum is over all partitions of [0,t]. $M_E[0,\infty)$ denotes the space of (equivalence classes of) measurable E-valued functions topologized by convergence in measure.

<u>4.5</u> <u>Theorem</u> For each n, let (X_n,Y_n) be an $\{\mathcal{F}_t^n\}$-adapted process with sample paths in $D_{\mathbb{M}^{km}\times\mathbb{R}^m}[0,\infty)$, and let X_n and Y_n be $\{\mathcal{F}_t^n\}$-semimartingales. Suppose that for some $0 < \delta \le \infty$, C2.2(i) holds for $\{Y_n\}$ and that for each $t > 0$

$$\sup_n (V_t(X_n) + E[|X_n(t)|]) < \infty \tag{4.14}$$

where $V_t(X_n)$ is the conditional variation with respect to the filtration $\{\mathcal{F}_t^n\}$. If $(X_n,Y_n) \Rightarrow (X,Y)$ in $M_{\mathbb{M}^{km}}[0,\infty)\times D_{\mathbb{R}^m}[0,\infty)$ and Y is continuous, then X has a version with sample paths in $D_{\mathbb{M}^{km}}[0,\infty)$, Y is a semimartingale with respect to a filtration to which X and Y are adapted, and $(X_n,Y_n,\int X_n\,dY_n) \Rightarrow (X,Y,\int X\,dY)$ in $M_{\mathbb{M}^{km}}[0,\infty)\times D_{\mathbb{R}^m\times\mathbb{R}^k}[0,\infty)$. If $(X_n,Y_n) \to (X,Y)$ in $M_{\mathbb{M}^{km}}[0,\infty)\times D_{\mathbb{R}^m}[0,\infty)$ in probability, then the triple converges in probability.

<u>Proof</u> Let φ be convex, symmetric and C^2 on $\mathbb{R}$, and suppose that $\varphi(0) = 0$, $\varphi''(0) = 1$, φ'' is decreasing on $[0,\infty)$, and $\varphi''(1) = 0$, and define ψ on $\mathbb{M}^{km}$ by $\psi(x) = \sum\varphi(x_{ij})$. As in (1.4) of Kurtz (1989), there exists an increasing process B_n such that

$$\psi(X_n(t_0+t) - X_n(t_0)) - (B_n(t_0+t) - B_n(t_0)) \tag{4.15}$$

is a local $\{\mathcal{F}_{t_0+t}\}$-supermartingale for each $t_0 \geq 0$.

Then setting $C_n(t) = B_n(t) + [M_n^\delta]_t + T_t(A_n^\delta)$ It follows from C2.2(i), that

$$\psi(X_n(t_0+t) - X_n(t_0)) + |M_n^\delta(t_0+t) - M_n^\delta(t_0)|^2 + |A_n^\delta(t_0+t) - A_n^\delta(t_0)| - (C_n(t_0+t) - C_n(t_0)) \tag{4.16}$$

is a local $\{\mathcal{F}_{t_0+t}\}$-supermartingale for each $t_0 \geq 0$ and that the conditions of Theorem 1.1 of Kurtz (1989) are satisfied for $Z_n = (X_n, M_n^\delta, A_n^\delta)$. Let D_n denote the dual predictable projection of $C_n(t) + t$, and let γ_n denote the inverse of D_n. Define $\tilde{Z}_n(t) = (\tilde{X}_n, \tilde{M}_n^\delta, \tilde{A}_n^\delta)) = \lim_{s\to t+} Z_n(\gamma_n(s)-)$. Then $\{(\tilde{Z}_n, \gamma_n)\}$ is relatively compact in $D_{\mathbb{M}^{km}\times\mathbb{R}^m\times\mathbb{R}^m\times\mathbb{R}}[0,\infty)\times[0,\infty]$. Furthermore, $\tilde{Y}_n^\delta = \tilde{M}_n^\delta + \tilde{A}_n^\delta$ is a sequence of semimartingales satisfying C2.2(i) (with $\tilde{M}_n^\delta$ in place of M_n^δ and $\tilde{A}_n^\delta$ in place of A_n^δ). It follows that if a subsequence of $\{(\tilde{Z}_n, \gamma_n)\}$ converges to a process $(\tilde{Z}, \gamma)$, then, setting $\tilde{Z} = (\tilde{X}, \tilde{M}, \tilde{A})$ and $\tilde{Y} = \tilde{M} + \tilde{A}$, along that subsequence

$$V_n \equiv \int \tilde{X}_n \, d\tilde{Y}_n^\delta \Rightarrow \int \tilde{X} \, d\tilde{Y} \equiv V \tag{4.17}$$

But $V_n \circ D_n = \int X_n \, dY_n^\delta$, and defining $\gamma^{-1}(t) = \inf\{u : \gamma(u) > t\}$, then (X,Y) has the same distribution as $(\tilde{X}\circ\gamma^{-1}, \tilde{Y}\circ\gamma^{-1})$ (since the continuity of Y ensures that $Y_n^\delta \Rightarrow Y$). The fact that $Y_n^\delta \Rightarrow Y$ and that Y is continuous implies that $\tilde{Y}$ is constant on any interval on which γ is constant (by Kurtz (1989), Lemma 2.3(c)) and hence V is also. Furthermore, the continuity of $\tilde{Y}$ ensures the continuity of V, and it follows from Kurtz (1989), Lemma 2.3, that $V_n \circ D_n \Rightarrow V\circ\gamma^{-1} = \int \tilde{X}\circ\gamma^{-1} \, d\tilde{Y}\circ\gamma^{-1}$, which gives the theorem. □

The above theorem still is not optimal even in the case of continuous integrands. For example, if each Y_n is a standard Brownian motion and $(X_n, Y_n) \Rightarrow (X,Y)$ in $L^2_{\mathbb{R}}[0,\infty)\times D_{\mathbb{R}}[0,\infty)$, then $\int X_n \, dY_n \Rightarrow \int X \, dY$. The following theorem comes close to covering this situation at the cost of placing strong conditions on the relationship between X_n and Y_n. Of course, other approximations of X_n could be used in place of X_n^h defined below.

4.6 Theorem Let $\{(X_n,Y_n)\}$ satisfy the conditions of Proposition 4.1, and $Y_n = M_n + A_n + Z_n$, where (M_n,A_n,Z_n) satisfy the conditions of Theorem 2.7. Define X_n^h by

$$X_n^h(t) = h^{-1}\int_{t-h}^{t} X_n(s)\,ds \tag{4.18}$$

Suppose that for each $t > 0$ and $\epsilon > 0$

$$\lim_{h\to 0}\overline{\lim_{n\to\infty}} P\left\{\int_0^t |X_n^h(s-)-X_n(s-)|^2 d[M_n]_s + \int_0^t |X_n^h(s-)-X_n(s-)|\, d(T_s(A_n)+T_s(Z_n)) \geq \epsilon\right\} = 0 \tag{4.19}$$

If $(X_n,Y_n,Z_n) \Rightarrow (X,Y,Z)$ in $M_{\mathbb{M}^{km}}[0,\infty)\times D_{\mathbb{R}^m\times\mathbb{R}^m}[0,\infty)$, then

$$U(t) \equiv \lim_{h\to 0}\int_0^t X^h\, dY \tag{4.20}$$

exists, and $(X_n,Y_n,\int X_n\, dY_n) \Rightarrow (X,Y,U)$ in $M_{\mathbb{M}^{km}}[0,\infty)\times D_{\mathbb{R}^m\times\mathbb{R}^k}[0,\infty)$. If $(X_n,Y_n,Z_n) \to (X,Y,Z)$ in $M_{\mathbb{M}^{km}}[0,\infty)\times D_{\mathbb{R}^m\times\mathbb{R}^m}[0,\infty)$ in probability, then $(X_n,Y_n,\int X_n\, dY_n) \to (X,Y,U)$ converges in probability.

Proof Since X_n^h is locally Lipschitz, the conditions on H_n in Proposition 4.1 ensure that $(X_n^h,Y_n,Z_n) \Rightarrow (X^h,Y,Z)$ in $D_{\mathbb{M}^{km}\times\mathbb{R}^m\times\mathbb{R}^m}[0,\infty)$ and hence that $\int X_n^h\, dY_n \Rightarrow \int X^h\, dY$. Consequently, (4.19) implies the result. □

5. Stochastic differential equations In this section we generalize results of Slomiński (1989). (See also Hoffman (1989) for results assuming the limiting semimartingale is continuous.) Note that Slomiński also considers Stratonovich equations. Avram (1988) considered the special case of stochastic exponentials, that is solutions of equations of the form ($k = m = 1$)

$$X(t) = 1 + \int_0^t X(s-)\,dY(s) \tag{5.1}$$

For $n = 1,2,\ldots$ let $F_n: D_{\mathbb{R}^k}[0,\infty) \to D_{\mathbb{M}^{km}}[0,\infty)$, let U_n and Y_n be processes with sample paths in $D_{\mathbb{R}^k}[0,\infty)$ and $D_{\mathbb{R}^m}[0,\infty)$ respectively, adapted to a filtration $\{\mathcal{F}_t^n\}$. Suppose Y_n is a semimartingale and that F_n is nonanticipating in the sense that $F_n(x,t) = F_n(x^t,t)$ for all $t \geq 0$ and $x \in D_{\mathbb{R}^k}[0,\infty)$, where $x^t(\cdot) = x(\cdot \wedge t)$. Let X_n be adapted to $\{\mathcal{F}_t^n\}$ and satisfy

$$X_n(t) = U_n(t) + \int_0^t F_n(X_n,s-)\,dY_n(s) \tag{5.2}$$

In order to apply Theorem 2.2 to the study of the weak convergence of solutions of this sequence of equations to the solution of a limiting equation

$$X(t) = U(t) + \int_0^t F(X,s-)\,dY(s) \tag{5.3}$$

we need conditions under which weak convergence of the pair $(X_n,Y_n) \Rightarrow (X,Y)$ implies $(Y_n,F_n(X_n)) \Rightarrow (Y,F(X))$. We could, of course, simply assume that $(x_n,y_n) \to (x,y)$ in $D_{\mathbb{R}^k\times\mathbb{R}^m}[0,\infty)$ implies $(x_n,y_n,F_n(x_n)) \to (x,y,F(x))$ in $D_{\mathbb{R}^k\times\mathbb{R}^m\times\mathbb{M}^{km}}[0,\infty)$, and under that assumption we have the following proposition.

5.1 Proposition Suppose that (U_n,X_n,Y_n) satisfies (5.2), that $\{(U_n,X_n,Y_n)\}$ is relatively compact in $D_{\mathbb{R}^k\times\mathbb{R}^k\times\mathbb{R}^m}[0,\infty)$, that $(U_n,Y_n) \Rightarrow (U,Y)$, and that $\{Y_n\}$ satisfies C2.2(i) for some $0 < \delta \leq \infty$. Assume that $\{F_n\}$ and F satisfy

C5.1 If $(x_n,y_n) \to (x,y)$ in the Skorohod topology, then $(x_n,y_n,F_n(x_n)) \to (x,y,F(x))$ in the Skorohod topology.

Then any limit point of the sequence $\{X_n\}$ satisfies (5.3).

<u>Proof</u> First note that if a subsequence of $\{X_n\}$ converges in distribution, then along a further subsequence the triple will converge in distribution to a process (U,X,Y). Theorem 2.2 then implies that (5.3) is satisfied. □

The following lemma, a generalization of Lemma 2.1, shows that the assumption on the sequence $\{F_n\}$ is valid for many interesting examples. Let Λ^1 be the subset of absolutely continuous functions in Λ for which $\gamma(\lambda) \equiv \|\ln\dot{\lambda}\|_\infty$ is finite.

<u>5.2 Lemma</u> Suppose that $\{F_n\}$ and F satisfy the following conditions:

C5.2(i) For each compact subset $\mathcal{K} \subset D_{\mathbb{R}^k}[0,\infty)$ and $t > 0$, $\sup_{x\in\mathcal{K}}\sup_{s\le t}|F_n(x,s) - F(x,s)| \to 0$.

C5.2(ii) For $\{x_n\}$ and x in $D_{\mathbb{R}^k}[0,\infty)$ and each $t > 0$, $\sup_{s\le t}|x_n(s) - x(s)| \to 0$ implies $\sup_{s\le t}|F(x_n,s) - F(x,s)| \to 0$.

C5.2(iii) For each compact subset $\mathcal{K} \subset D_{\mathbb{R}^k}[0,\infty)$ and $t > 0$, there exists a continuous function $\omega:[0,\infty) \to [0,\infty)$ with $\omega(0) = 0$ such that for all $\lambda \in \Lambda^1$, $\sup_{x\in\mathcal{K}}\sup_{s\le t}|F(x\circ\lambda,s) - F(x,\lambda(s))| \le \omega(\gamma(\lambda))$.

Then $(x_n,y_n) \to (x,y)$ in the Skorohod topology implies $(x_n,y_n,F_n(x_n)) \to (x,y,F(x))$ in the Skorohod topology.

<u>Proof</u> If $(x_n,y_n) \to (x,y)$ in the Skorohod topology, then there exist $\lambda_n \in \Lambda^1$ such that $\gamma(\lambda_n) \to 0$ and $(x_n\circ\lambda_n,y_n\circ\lambda_n) \to (x,y)$ uniformly on bounded time intervals. Consequently,

(5.4) $F_n(x_n,\lambda_n(s)) - F(x,s) =$

$$F_n(x_n,\lambda_n(s)) - F(x_n,\lambda_n(s)) + F(x_n,\lambda_n(s)) - F(x_n\circ\lambda_n,s) + F(x_n\circ\lambda_n,s) - F(x,s)$$

goes to zero uniformly in s on bounded intervals. □

<u>5.3 Examples</u> Let $g:\mathbb{R}^k\times[0,\infty) \to \mathbb{M}^{km}$ and $h:[0,\infty) \to [0,\infty)$ be continuous. The following functions satisfy C5.2(ii) and C5.2(iii).

a) $F(x,t) = g(x(t),t)$

b) $F(x,t) = \int_0^t h(t-s)g(x(s),s)\,ds$

For $k = m = 1$

c) $F(x,t) = \sup_{s\leq t} h(t-s)g(x(s),s)$

d) $F(x,t) = \sup_{s\leq t} h(t-s)g(x(s)-x(s-),s)$

One shortcoming of Proposition 5.1 is the apriori assumption that the sequence of solutions is relatively compact. We can avoid this assumption by localizing the result and applying Proposition 4.1. We say that (X,τ) is a local solution of (5.3) if there exists a filtration $\{\mathcal{F}_t\}$ to which X, U, and Y are adapted, Y is an $\{\mathcal{F}_t\}$-semimartingale, τ is an $\{\mathcal{F}_t\}$-stopping time, and

(5.5) $$X(t\wedge\tau) = U(t\wedge\tau) + \int_0^{t\wedge\tau} F(X,s-)\,dY(s)$$

We say that local uniqueness holds for (5.3) if any two local solutions (X_1,τ_1), (X_2,τ_2) satisfy $X_1(t) = X_2(t)$, $t \leq \tau_1\wedge\tau_2$, a.s. See Protter (1989), Chapter V, for sufficient conditions for uniqueness.

5.4 Theorem Suppose that (U_n,X_n,Y_n) satisfies (5.1), $(U_n,Y_n) \Rightarrow (U,Y)$ in the Skorohod topology, that $\{Y_n\}$ satisfies C2.2(i) for some $0 < \delta \leq \infty$, and that $\{F_n\}$ and F satisfy C5.1 (see Lemma 5.2). For $b > 0$, define $\eta_n^b = \inf\{t:|F_n(X_n,t)|\vee|F_n(X_n,t-)| \geq b\}$ and let X_n^b denote the solution of

(5.6) $$X_n^b(t) = U_n(t) + \int_0^t \chi_{[0,\eta_n^b)}(s-)F_n(X_n^b,s-)\,dY_n$$

that agrees with X_n on $[0,\eta_n^b)$. Then $\{(U_n,X_n^b,Y_n)\}$ is relatively compact and any limit point, (U,X^b,Y), gives a local solution (X^b,τ) of (5.3) with $\tau = \eta^c \equiv \inf\{t:|F(X^b,t)|\vee|F(X^b,t-)| \geq c\}$ for any $c < b$. If there exists a global solution X of (5.3) and local uniqueness holds, then $(U_n,X_n,Y_n) \Rightarrow (U,X,Y)$.

5.5 Remark If U and Y are continuous then, then C5.1 can be replaced by

C5.4 If $(x_n,y_n) \to (x,y)$ in the compact uniform topology (that is $(x_n(t),y_n(t)) \to (x(t),y(t))$ uniformly on bounded time intervals), then $(x_n,y_n,F_n(x_n)) \to (x,y,F(x))$ in the Skorohod topology.

Recall that if $z_n \to z$ in the Skorohod topology and z is continuous, then $z_n \to z$ in the compact uniform topology.

<u>Proof</u> The relative compactness of $\{(U_n, X_n^b, Y_n)\}$ is an immediate consequence of Proposition 4.1. The sequence $\{(U_n, X_n^b, Y_n, \eta_n^b)\}$ will be relatively compact in $D_{\mathbb{R}^k \times \mathbb{R}^k \times \mathbb{R}^m}[0,\infty) \times [0,\infty]$. Let (U, X^b, Y, η_0^b) denote a weak limit point. To simplify notation, assume that the original sequence converges and (with reference to the Skorohod representation theorem) assume that the convergence is almost sure rather than in distribution. Note that $\eta^b \leq \eta_0^b$.

It follows that $U_n + \int F_n(X_n^b)\,dY_n \to U + \int F(X^b)\,dY$ and since

$$X_n^b(t) = U_n(t) + \int_0^t F_n(X_n^b, s-)\,dY_n(s) \tag{5.7}$$

for $t \leq \eta_n^b$,

$$X^b(t) = U(t) + \int_0^t F(X^b, s-)\,dY(s) \tag{5.8}$$

for $t < \eta_0^b$. Let $c < b$. If $\eta^c < \eta^b$, then (5.8) holds for $t \leq \eta^c$. If $\eta^c = \eta^b$, then $F(X^b)$ has a discontinuity at η^c with $|F(X^b, \eta^c-)| \leq c$ and $|F(X^b, \eta^c)| \geq b$. It follows that for $c < d < b$, $(U_n(\eta_n^d), X_n^b(\eta_n^d), Y_n(\eta_n^d), Y_n(\eta_n^d-), F_n(X_n^b, \eta_n^d), F_n(X_n^b, \eta_n^d-), \eta_n^d)$ converges to $(U(\eta^d), X^b(\eta^d), Y(\eta^d), Y(\eta^d-), F(X^b, \eta^d), F(X^b, \eta^d-), \eta^d)$ and

$$X^b(\eta^d) = U(\eta^d) + \int_0^{\eta^d} F(X^b, s-)\,dY(s) \tag{5.9}$$

so that (5.8) holds for $t \leq \eta^c$ $(= \eta^d)$. Consequently, (X^b, η^c) is a local solution of (5.3).

If local uniqueness holds for (5.3) and there exists a global solution X, then X^b must agree with X on the interval $[0, \eta^c]$ for all c and b with $c < b$, and since X is a global solution, it follow that $\eta^b \to \infty$ as $b \to \infty$. The convergence in distribution of X_n to X follows. □

6. Technical results

Uniform approximation by step functions Let E be a metric space with metric r. Let $\{\theta_k\}$ be a sequence of independent random variables, uniformly distributed on the interval $[\frac{1}{2},1]$. Fix $\epsilon > 0$, and for $z \in D_E[0,\infty)$ define $\tau_0(z) = 0$ and $\tau_{k+1}(z) = \inf\{t > \tau_k(z): r(z(t),z(\tau_k(z))) \vee r(z(t-),z(\tau_k(z))) \geq \epsilon\theta_k\}$ and set $\gamma_k(z) = z(\tau_k(z))$. Finally, define $I_\epsilon(z)$ by $I_\epsilon(z)(t) = \gamma_k(z)$ for $\tau_k(z) \leq t < \tau_{k+1}(z)$. Note that $r(z(t),I_\epsilon(z)(t)) \leq \epsilon$ for all t. Let $U_1 = \{u:u = r(z(t),z(0)) \text{ or } r(z(t-),z(0)) \text{ for some } t \text{ such that } z(t) \neq z(t-)\}$, and defining $m(t) = \sup_{s\leq t} r(z(s),z(0))$, let $U_2 = \{m(t):m \text{ is not strictly increasing at } t\}$. U_1 and U_2 are countable, so with probability one, $\epsilon\theta_0 \notin U_1 \cup U_2$. Let $z_n \to z$, and assume that $\epsilon\theta_0 \notin U_1 \cup U_2$. Either m is strictly increasing at $\tau_1(z)$ or $r(z(\tau_1(z)-),z(0)) < \epsilon\theta_0 < r(z(\tau_1(z)),z(0))$, and it follows that $\tau_1(z_n) \to \tau_1(z)$. Either z is continuous at $\tau_1(z)$ or $r(z(\tau_1(z)-),z(0)) < \epsilon\theta_0 < r(z(\tau_1(z)),z(0))$, and it follows that $\gamma_1(z_n) \to \gamma_1(z)$. In general, if $z_n \to z$ in the Skorohod topology, $t_n \to t$ and $z_n(t_n) \to z(t)$, then $z_n(t_n+\cdot) \to z(t+\cdot)$ in the Skorohod topology. Consequently, $z_n \to z$ implies $z_n(\tau_1(z_n)+\cdot) \to z(\tau_1(z)+\cdot)$ a.s. An induction argument then shows that $z_n \to z$ implies $\tau_k(z_n) \to \tau_k(z)$ and $\gamma_k(z_n) \to \gamma_k(z)$ a.s for all k. With these observations, we can prove the following lemma.

6.1 Lemma Let I_ϵ be defined as above. If $z_n \to z$ in the Skorohod topology on $D_E[0,\infty)$, then $(z_n,I_\epsilon(z_n)) \to (z,I_\epsilon(z))$ a.s. in the Skorohod topology on $D_{E\times E}[0,\infty)$.

To carry out the proof, we need the following (see Proposition 3.6.5 of Ethier and Kurtz (1986)).

6.2 Lemma For an arbitrary metric space (E',r'), $v_n \to v$ in the Skorohod topology on $D_{E'}[0,\infty)$ if and only if the following conditions hold:

C6.2(i) If $t_n \to t$, then $\lim_{n\to\infty} r'(v_n(t_n),v(t)) \wedge r'(v_n(t_n),v(t-)) = 0$

C6.2(ii) If $s_n \geq t_n$, $s_n,t_n \to t$, and $v_n(t_n) \to v(t)$, then $v_n(s_n) \to v(t)$.

Proof of Lemma 6.1 Suppose $z_n \to z$ in $D_E[0,\infty)$ and $t_n \to t$. If $\tau_k(z) < t < \tau_{k+1}(z)$, then $I_\epsilon(z)$ is continuous at t, $I_\epsilon(z_n)(t_n) \to \gamma_k(z) = I_\epsilon(z)(t)$, and C6.2(i) and (ii) follow for $\{(z_n,I_\epsilon(z_n))\}$ by the analogous conditions for $\{z_n\}$. If $t = \tau_k(z)$, we can assume that either z is continuous at $\tau_k(z)$ or $r(z(\tau_k(z)-),z(\tau_{k-1}(z))) < \epsilon\theta_{k-1} < r(z(\tau_k(z)),z(\tau_{k-1}(z)))$. The convergence of $\tau_{k-1}(z_n)$,

$\tau_k(z_n)$, $\gamma_{k-1}(z_n)$, and $\gamma_k(z_n)$ implies C6.2(i) and (ii) for $\{I_\epsilon(z_n)\}$, and if z is continuous at $\tau_k(z)$, C6.2(i) and (ii) follow for $\{(z_n,I_\epsilon(z_n))\}$. If $r(z(\tau_k(z)-),z(\tau_{k-1}(z))) < \epsilon\theta_{k-1} < r(z(\tau_k(z)),z(\tau_{k-1}(z)))$, then, with probability one, for n sufficiently large the same inequality holds with z replaced by z_n. Consequently, if $t_n \geq \tau_k(z_n)$ and $t_n \to \tau_k(z)$, then $z_n(t_n)$ and $I_\epsilon(z_n)(t_n)$ both converge to $\gamma_k(z)$, and if $t_n < \tau_k(z_n)$ and $t_n \to t$, then $z_n(t_n)$ converges to $z(\tau_k(z)-)$ and $I_\epsilon(z_n)(t_n)$ converges to $\gamma_{k-1}(z) = I_\epsilon(z)(\tau_k(z)-)$. C6.2(i) and (ii) follow for $\{(z_n,I_\epsilon(z_n))\}$. □

Uniform tightness Jakubowski, Memin, and Pages (1989) and Słomiński (1989) develop their results under a "uniform tightness" condition. We discuss this condition for a sequence of one-dimensional semimartingales $\{Y_n\}$ satisfying $Y_n(0) = 0$. The results below are essentially contained in Lemma 3.1 of Jakubowski, Memin, and Pages (1989). They are presented here for completeness.

Let $\mathcal{H}_n$ denote the collection of cadlag $\{\mathcal{F}_t^n\}$-adapted, $\mathbf{R}$-valued processes satisfying $|H_n(t)| \leq 1$ for all $t \geq 0$. Then $\{Y_n\}$ is uniformly tight if for each $t > 0$

$$\left\{\int_0^t H_n(s-)\,dY_n(s): H_n \in \mathcal{H}_n,\ n = 1,2,\ldots\right\} \tag{6.1}$$

is stochastically bounded.

Assume that $\{Y_n\}$ is uniformly tight. Let $\mathcal{T}_n$ denote the collection of $\{\mathcal{F}_t^n\}$-stopping times. For $\tau \in \mathcal{T}_n$ and $\epsilon > 0$, let $H_n = \chi_{[0,\tau)}$. Then the integral in (6.1) gives $Y_n(t\wedge\tau)$, and we see that for each $t > 0$, $\{Y_n(t\wedge\tau): \tau \in \mathcal{T}_n,\ n = 1,2,\ldots\}$ is stochastically bounded. Considering the collection of stopping times of the form $\tau = \inf\{s: |Y_n(s)| \geq c\}$, it follows that $\{\sup_{s\leq t}|Y_n(s)|: n = 1,2,\ldots\}$ is stochastically bounded. Recalling that

$$[Y_n]_t = Y_n(t)^2 - \int_0^t 2Y_n(s-)\,dY_n(s) \tag{6.2}$$

and using the stochastic boundedness of the suprema, we see that $\{[Y_n]_t: n = 1,2,\ldots\}$ is stochastically bounded.

The stochastic boundedness of the quadratic variations ensures that the uniform tightness of $\{Y_n\}$ implies uniform tightness of $\{Y_n^\delta\}$ for each $0 < \delta < \infty$. Fix $0 < \delta < \infty$ and let $Y_n^\delta = M_n^\delta + A_n^\delta$ be the canonical decomposition of Y_n^δ (Protter (1989) §III.5). Then the discontinuities of M_n^δ and A_n^δ are bounded by 2δ, and $E[[Y_n^\delta]_\tau] = E[[M_n^\delta]_\tau] + E[[A_n^\delta]_\tau]$ for any stopping time τ (with the possibility of $\infty = \infty$) (Protter (1989) §IV.2.) Let $\gamma_n^c = \inf\{s:[Y_n^\delta]_s \geq c\}$. Fix t and for k = 1,2,..., let $\{t_i^k\}$ be a partition of [0,t] with $\lim_{k\to\infty}\max_i(t_{i+1}-t_i) = 0$. Define

$$H_n^k = \sum_i \text{sign}\Big(E[A_n^\delta(t_{i+1}^k\wedge\gamma_n^c) - A_n^\delta(t_i^k\wedge\gamma_n^c)|\mathcal{F}_{t_i}^n]\Big)\chi_{[t_i^k\wedge\gamma_n^c,t_{i+1}^k\wedge\gamma_n^c)} \tag{6.3}$$

The first term on the right of

$$\int_0^u H_n^k(s-)\,dY_n^\delta(s) = \int_0^u H_n^k(s-)\,dM_n^\delta(s) + \int_0^u H_n^k(s-)\,dA_n^\delta(s) \equiv U_n^k(u) + V_n^k(u) \tag{6.4}$$

satisfies

$$E[\sup_{s\leq t} U_n^k(s)^2] \leq 4\,E[M_n^\delta(t\wedge\gamma_n^c)^2] \leq 4(c+(2\delta)^2) \tag{6.5}$$

so $\{U_n^k(t):k,n = 1,2,\ldots\}$ is stochastically bounded which, by the stochastic boundedness of (6.1) (with Y_n replaced by Y_n^δ), implies the stochastic boundedness of $\{V_n^k(t):k,m = 1,2,\ldots\}$. But the predictability of A_n^δ implies

$$\begin{aligned}(6.6)\quad & T_{t\wedge\gamma_n^c}(A_n^\delta) \\ &= \lim_{k\to\infty}\sum \text{sign}\Big(E[A_n^\delta(t_{i+1}^k\wedge\gamma_n^c) - A_n^\delta(t_i^k\wedge\gamma_n^c)|\mathcal{F}_{t_i}^n]\Big)\Big(A_n^\delta(t_{i+1}^k\wedge\gamma_n^c) - A_n^\delta(t_i^k\wedge\gamma_n^c)\Big) \\ &= \lim_{k\to\infty} V_n^k(t)\end{aligned}$$

(see Dellacherie and Meyer (1982), page 423) so $\{T_{t\wedge\gamma_n^c}(A_n^\delta)\}$ is stochastically bounded for each c. But the stochastic boundedness of $\{[Y_n^\delta]_t\}$ for each t implies that for each $\epsilon > 0$, there exists a c such that $P\{\gamma_n^c \leq t\} \leq \epsilon$ and hence there exists an $a > 0$ such that $P\{T_t(A_n^\delta) \geq a\} \leq P\{T_{t\wedge\gamma_n^c}(A_n^\delta) \geq a\} + P\{\gamma_n^c \leq t\} \leq 2\epsilon$, verifying the stochastic boundedness of $\{T_t(A_n^\delta)\}$ and C2.2(ii). C2.2(iii) is immediate, so C2.2(i) holds.

If $\{Y_n\}$ is relatively compact and satisfies C2.2(i) for some $\delta > 0$, then Proposition 4.1 implies uniform tightness for $\{Y_n\}$. Actually the relative compactness is not needed. If there exists a δ for which $\{J_\delta(Y_n)\}$ is stochastically bounded and C2.2(i) holds, then $\{Y_n\}$ is uniformly tight.

7. References

Avram, F. (1988). Weak convergence of the variations, iterated integrals, and Doléans-Dade exponentials of sequences of semimartingales. *Ann. Probab.* 16, 246-250.

Bobkoski, Mark J. (1983). Hypothesis testing in nonstationary time series. PhD Dissertation, University of Wisconsin-Madison.

Dellacherie, Claude and Meyer, Paul-André (1982). *Probabilities and Potential B.* North Holland, Amsterdam.

DiMasi, G. B. and Runggaldier, W. J. (1981). Continuous-time approximations for the nonlinear filtering problem. *Appl. Math. Optim.* 7, 233-245.

DiMasi, G. B. and Runggaldier, W. J. (1982). On approximation methods for nonlinear filtering. *Nonlinear Filtering and Stochastic Control, Lect. Notes Math.* 972. S. K. Mitter and A. Moro eds. Springer Verlag, Berlin, ??-??

Duffie, Darrell and Protter, Philip (1989). From discrete to continuous time finance: Weak convergence of the financial gain process. (to appear)

Dynkin, E. B. and Mandelbaum, A. (1983). Symmetric statistics, Poisson point processes, and multiple Wiener integrals. *Ann. Statist.* 11, 739-745.

Billingsley, Patrick (1968). *Convergence of Probability Measures.* Wiley, New York.

Cox, Dennis D. and Llatas, Isabel (1989). M-estimation for nearly nonstationary autoregressive time series. (to appear)

Emery, M. (1989). On the Azéma martingales. Séminaire de Probabilités XXIII (to appear).

Ethier, Stewart N. and Kurtz, Thomas G. (1986). *Markov Processes: Characterization and Convergence.* Wiley, New York.

Filippova, A. A. (1961). Mises' theorem on the asymptotic behavior of functionals of empirical distribution functions and its statistical applications. *Theor. Probab. Appl.* 7, 24-56.

Goggin, Eimear Mary (1988). Weak convergence of conditional probabilities. PhD dissertation, University of Wisconsin-Madison.

Hall, Peter (1979). On the invariance principle for U-statistics. *Stochastic Process. Appl.* 9, 163-174.

Hoffman, Karin (1989). Approximation of stochastic integral equations by martingale difference arrays. (preprint).

Jacod, J., Memin, J. and Metivier, M. (1983). On tightness and stopping times. *Stochastic Process. Appl.* 14, 109-146.

Jacod, Jean and Shiryaev, Albert (1987). *Limit Theorems for Stochastic Processes.* Springer-Verlag, Berlin.

Jakubowski, A, Memin, J., and Pages, G. (1989). Convergence en loi des suites d'integrales stochastique sur l'espace D^1 de Skorokhod. *Probab. Th. Rel. Fields* 81, 111-137.

Johnson, Daniel P. (1983). Diffusion approximations for optimal filtering of jump processes and for queuing networks. PhD dissertation, University of Wisconsin-Madison.

Kurtz, Thomas G. (1989). Random time changes and convergence in distribution under the Meyer-Zheng conditions.

Kushner, Harold J. (1974). On the weak convergence of interpolated Markov chains to a diffusion. *Ann. Probab.* 2, 40-50.

Llatas, Isabel (1987). Asymptotic inference for nearly non-stationary time series. PhD Dissertation, University of Wisconsin-Madison.

Meyer, Paul-André (1989). Note sur les martingales d'Azema. Séminaire de Probabilités XXIII (to appear).

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

Neuhaus, G. (1977). Functional limit theorems for U-statistics in the degenerate case. *J. Multivariate Anal.* 7, 424-439.

Protter, Philip (1985). Approximations of solutions of stochastic differential equations driven by semimartingales. *Ann. Probab.* 13, 716-743.

Protter, Philip (1989). *Stochastic Integration and Differential Equations: A New Approach* (to appear).

Protter, Philip and Sharpe, Michael (1979). Martingales with given absolute value. *Ann. Probab.* 7, 1056-1058.

Rubin, H. and Vitale, R. A. (1980). Asymptotic distribution of symmetric statistics. *Ann. Statist.* 8, 165-170.

Skorohod, A. V. (1965). *Studies in the Theory of Random Processes.* Addison-Wesley, Reading, MA.

Slomiński, Leszek (1989). Stability of strong solutions of stochastic differential equations. (preprint)

Strasser, H. (1986). Martingale difference arrays and stochastic integrals. *Probab. Th. Rel. Fields* 72, 83-98.

Wong, Eugene, and Zakai, Moshe (1965). On the convergence of ordinary integrals to stochastic integrals. *Ann. Math. Statist.* 36, 1560-1564.

Acknowledgements The research of both authors is supported in part by the National Science Foundation. This work was completed while the first author was visiting the Department of Mathematics at the University of Utah. The hospitality of that department is appreciated and gratefully acknowledged.