WEAK LIMIT THEOREMS FOR STOCHASTIC INTEGRALS
AND STOCHASTIC DIFFERENTIAL EQUATIONS

BY THOMAS G. KURTZ AND PHILIP PROTTER

University of Wisconsin–Madison and Purdue University

Assuming that \((X_n, Y_n)\) is a sequence of cadlag processes converging in distribution to \((X, Y)\) in the Skorohod topology, conditions are given under which the sequence \((X_n dY_n)\) converges in distribution to \(X dY\). Examples of applications are given drawn from statistics and filtering theory. In particular, assuming that \((U_n, Y_n) \Rightarrow (U, Y)\) and that \(F_n \rightarrow F\) in an appropriate sense, conditions are given under which solutions of a sequence of stochastic differential equations \(dX_n = dU_n + F_n(X_n) dY_n\) converge to a solution of \(dX = dU + F(X) dY\), where \(F_n\) and \(F\) may depend on the past of the solution. As is well known from work of Wong and Zakai, this last conclusion fails if \(Y\) is Brownian motion and the \(Y_n\) are obtained by linear interpolation; however, the present theorem may be used to derive a generalization of the results of Wong and Zakai and their successors.

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

\[
E[Y^n_{k+1} - Y^n_k | \mathcal{F}^n_k] = b(Y^n_k) \frac{1}{n} + o\left(\frac{1}{n}\right)
\]

and

\[
E\left[(Y^n_{k+1} - Y^n_k)^2 | \mathcal{F}^n_k\right] = a(Y^n_k) \frac{1}{n} + o\left(\frac{1}{n}\right).
\]

Using these assumptions we can write

\[
Y^n_k = Y^n_0 + \sum_{i=0}^{k-1} (Y^n_{i+1} - Y^n_i)
\]

\[
= Y^n_0 + \sum_{i=0}^{k-1} b(Y^n_i) \frac{1}{n} + \sum_{i=0}^{k-1} a(Y^n_i)^{1/2} Z^n_{i+1} \frac{1}{\sqrt{n}} + \text{error},
\]

where

\[
Z^n_{k+1} = \frac{Y^n_{k+1} - Y^n_k - E[Y^n_{k+1} - Y^n_k | \mathcal{F}^n_k]}{\sqrt{E[(Y^n_{k+1} - Y^n_k - E[Y^n_{k+1} - Y^n_k | \mathcal{F}^n_k])^2 | \mathcal{F}^n_k]}}
\]

are martingale differences with conditional variance 1. If we define \(X^n(t) = Y^n_{[nt]}\)
and
\begin{equation}
W_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} Z_i^n,
\end{equation}
then
\begin{equation}
X_n(t) = X_n(0) + \int_0^{[nt]/n} b(X_n(s)) \, ds
+ \int_0^t \sqrt{a(X_n(s-))} \, dW_n(s) + \text{error}.
\end{equation}

[Note that $X_n$ is constant on intervals of length $1/n$, so the first sum in (1.3) equals the first integral in (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, e.g., 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 c\-d\-a\-g \-p\-r\-o\-c\-e\-s [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,
\begin{equation}
\int_0^t X(s-) \, dY(s) = \lim \sum X(t_i)(Y(t_{i+1}) - Y(t_i)),
\end{equation}
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 endpoint 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 (1990) 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.

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

$$
\int_0^t X_n(s-) \, dY_n(s) = 1,
$$

(1.8)

but the limiting integral gives

$$
\int_0^t X(s-) \, dY(s) = 0.
$$

(1.9)

1.2 Example. Let $W$ be standard Brownian motion and define $W_n$ so that

$$
\frac{d}{dt} W_n(t) = n \left( W\left( \frac{k+1}{n} \right) - W\left( \frac{k}{n} \right) \right), \quad t \in \left[ \frac{k}{n}, \frac{k+1}{n} \right).
$$

(1.10)

Then

$$
\int_0^t W_n(s-) \, dW_n(s)
= \int_0^t W_n\left( \frac{[ns]}{n} \right) \, dW_n(s) + \int_0^t \left( W_n(s) - W_n\left( \frac{[ns]}{n} \right) \right) \, dW_n(s)
$$

(1.11)

$$
= \sum W\left( \frac{k}{n} \right) \left( W\left( \frac{k+1}{n} \right) - W\left( \frac{k}{n} \right) \right)
$$

$$
+ \sum \int_0^{1/n} \left( W\left( \frac{k}{n} + s \right) - W\left( \frac{k}{n} \right) \right) \left( W\left( \frac{k+1}{n} \right) - W\left( \frac{k}{n} \right) \right) \, ds
$$

$$
\to \int_0^t W(s) \, dW(s) + \frac{1}{2} t.
$$

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 $D_{\mathbb{R}}[0, \infty)$ which is stronger than assuming convergence of each component in $D_{\mathbb{R}}[0, \infty)$. For future reference, let $\Lambda$ 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 \( D_{\mathbb{R}}[0, \infty) \),

\[
\int_0^s x_n(s) \, dy_n(s) \to \int_0^s x(s) \, dy(s),
\]

\[
\int_0^s y_n(s) \, dx_n(s) \to \int_0^s y(s) \, dx(s),
\]

in the Skorohod topology on \( D_{\mathbb{R}_1}[0, \infty) \). (Actually, the quadruple consisting of \( x_n, y_n \) and the two integrals converges in \( D_{\mathbb{R}}[0, \infty) \)). 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, Mémin and Pages (1989), 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 Slomiński (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 \( x_n \to x \) is 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 $J_\delta: D_{R^m}[0, \infty) \to D_{R^m}[0, \infty)$ by

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

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

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

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 $\tau_k \to \infty$ 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, $k \times m$ 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 $D_{M_{\infty}^m \times \mathbb{R}^m}[0, \infty)$ 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 $Y_n^\delta$ will also be a semimartingale.) Let $Y_n^\delta = M_n^\delta + A_n^\delta$ be a decomposition of $Y_n^\delta$ 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

$$P(\tau_n^\alpha \leq \alpha) \leq 1/\alpha \text{ and } \sup_n E[|M_{\tau_n^\alpha}^{\delta} + T_{\tau_n^\alpha}^{\tau_n^\alpha}(A_n^{\delta})|] < \infty.$$

If $(X_n, Y_n) \Rightarrow (X, Y)$ in the Skorohod topology on $D_{M_{\infty}^m \times \mathbb{R}^m}[0, \infty)$, 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 $D_{M_{\infty}^m \times \mathbb{R}^m \times \mathbb{R}^m}[0, \infty)$. If $(X_n, Y_n) \Rightarrow (X, Y)$ in probability, then the triple converges in probability.

2.3 Remark. If there exist decompositions of $\{Y_n^\delta\}$ such that C2.2(i) holds, we will simply say that $(Y_n)$ satisfies C2.2(i) for $\delta$. For $c > 0$, define
\[ \tau^c_\delta = \inf\{ t : |M_n^\delta(t)| \vee |M_n^\delta(t - )| \geq c \text{ or } T_i(A_n^\delta) \geq c \}. \]

Suppose the following conditions hold.

C.2.1 (ii) \( (T_i(A_n^\delta)) \) is stochastically bounded for each \( t > 0 \).

C.2.2 (iii) For each \( c > 0 \),

\[ \sup_n E[M_n^\delta(t \wedge \tau^c_n)^2 + T_n \wedge \tau_n(A_n^\delta)] < \infty. \]

Since convergence in distribution of \( \{Y_n\} \) in the Skorohod topology implies stochastic boundedness for \( \{\sup_{t \leq \alpha} Y_n(t)\} \), \( \sup_{t \leq \alpha} |Y_n^\delta(t) - A_n^\delta(t)| \leq \sup_{t \leq \alpha} |Y_n(t)| + T_n(A_n^\delta) \) is stochastically bounded in \( n \) for each \( \alpha \) and hence there exists \( c_\alpha \) so that \( P(\tau^c_n \leq \alpha) \leq 1/\alpha \). In addition, \( E[|M_n^\delta(t \wedge \tau^c_n)^2|] \) and C.2.2(i) is satisfied with \( \tau^c_n = \tau_n^c \). For \( \delta < \infty \), C.2.2(iii) will usually be immediate since the discontinuities of \( Y_n^\delta \) are bounded in magnitude by \( \delta \) (making \( Y_n^\delta \) a special semimartingale) and there will exist a decomposition with the discontinuities of each term bounded by \( 2\delta \) [see Jacod and Shiryaev (1987), Lemma 1.4.24].

2.4 Remark. To see that \( Y^\delta \) is a semimartingale, it is enough to show that \( Y^\delta \) is a semimartingale. Without loss of generality, we can assume that for \( \alpha = 1, 2, \ldots, \tau^c_n \leq \tau_n^{\alpha+1} \). Let \( Y_n^{\delta \alpha} = Y_n^\delta(\cdot \wedge \tau_n^\alpha) \). Then \( \{(X_n, Y_n, Y_n^\delta, Y_n^{\delta 1}, Y_n^{\delta 2}, \ldots, \tau_n^{\alpha-1}, \tau_n^\alpha, \tau_n^\beta, \ldots)\} \) is relatively compact in \( D_{M^m \times R^m \times R^m}[0, \infty] \times D_{R^m}[0, \infty] \times [0, \infty] \). Let \( (X, Y, Y^\delta, Y^{\delta 1}, Y^{\delta 2}, \ldots, \tau^1, \tau^2, \ldots) \) be some limit point and let \( \{\mathcal{T}_i\} \) be the filtration generated by the limiting processes and random times. For each \( T > 0 \), let

\[ V_T(Y_n^{\delta \alpha}) = \sup E\left[ \sum |E[Y_n^\delta(t_{i+1}) - Y_n^\delta(t_i)|\mathcal{T}_i]\right], \]

where the supremum is over all partitions of \([0, T]\). Then

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

and hence \( V_T(Y^{\delta \alpha}) < \infty \) (\( V_T \) defined using \( \{\mathcal{T}_i\} \)). [See, e.g., Meyer and Zheng (1984), Theorem 4 or Kurtz (1991), Theorem 5.8.] It follows that \( Y^{\delta \alpha} \) is a local \( \{\mathcal{T}_i\} \)-quasimartingale and hence an \( \{\mathcal{T}_i\} \)-semimartingale.

2.5 Remark. If \( Y_n = Y \) for each \( n \), then \( \{Y_n\} \) satisfies C.2.2(i) for all finite \( \delta \). If \( \{Y_n\} \) is relatively compact in the Skorohod topology and satisfies C.2.2(i) for some \( \delta \in (0, \infty) \), then \( \{Y_n\} \) satisfies C.2.2(i) for all finite \( \delta \). If \( \{(X_n, Y_n)\} \) is relatively compact in the Skorohod topology and \( \{Y_n\} \) satisfies C.2.2(i) for some \( \delta \in (0, \infty) \), then \( \{X_n, dY_n\} \) satisfies C.2.2(i) for all finite \( \delta \).

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

Proof. Let \( Z_n = (X_n, Y_n, J_i(Y_n), Y_n^\delta) \). \( Z_n \) has sample paths in \( D_{M^m \times R^m \times R^m \times R^m} \). 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 $I_z: D_E[0, \infty) \to D_E[0, \infty)$ such that $|x(t) - I_z(x)(t)| \leq \varepsilon$ for all $z \in D_E[0, \infty]$ and $t \geq 0$, $I_z(x)$ is a step function and the mapping $z \to (z, I_z(x))$ is continuous at $z$ a.s. for each $z \in D_E[0, \infty)$. Furthermore, $I_z(Z_n)$ is adapted to a filtration $\mathcal{F}_t^n = \mathcal{F}_t^n \vee \mathcal{H}$, where $\mathcal{H}$ is independent of $\{\mathcal{F}_t^n\}$ (and hence $Y_n$ will be a $\{\mathcal{F}_t^n\}$-semimartingale). Let $X_n^\varepsilon$ denote the first, $M^{k,m}$-valued component of $I_z(Z_n)$. Then $|X_n - X_n^\varepsilon| \leq \varepsilon$ and $(X_n, Y_n, J_\delta(Y_n), Y_\delta, X_n^\varepsilon) \Rightarrow (X, Y, J_\delta(Y), Y_\delta, X^\varepsilon)$.

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

$$R_n^\varepsilon = U_n - U_n^\varepsilon = \int (X_n - X_n^\varepsilon) \, dY_\delta$$

(2.6)

$$= \int (X_n - X_n^\varepsilon) \, dM_\delta + \int (X_n - X_n^\varepsilon) \, dA_\delta,$$

we see that for any stopping time $\tau$,

$$E\left[\sup_{s \leq \tau} R_n^\varepsilon(s)\right] \leq \varepsilon \left(2E\left[\left[ M_\delta^2\right]_{t \wedge \tau}\right]^{1/2} + E\left[T_{t \wedge \tau} A_\delta^2\right]\right)$$

(2.7)

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

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. $\square$

The transformation $J_\delta$ 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.

2.7 THEOREM. For each $n$, let $\{X_n, Y_n\}$ be an $\{\mathcal{F}_t^n\}$-adapted process with sample paths in $D_{M^{k,m} \times R^m}[0, \infty)$ 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
\[ P(\tau_n^\alpha \leq \alpha) \leq 1/\alpha \text{ and } \sup_n E[|M_n|_{t \wedge \tau_n^\alpha} + T_{t \wedge \tau_n^\alpha}(A_n)] < \infty. \]

If \( (X_n, Y_n, Z_n) \Rightarrow (X, Y, Z) \) in the Skorohod topology on \( D_{\mathbb{R}^d} \), then \( Y \) is a semimartingale with respect to a filtration to which \( X \) and \( Y \) are adapted and \( (X_n, Y_n, /X_n dY_n) \Rightarrow (X, Y, /X dY) \) in the Skorohod topology on \( D_{\mathbb{R}^d} \). If \( (X_n, Y_n, Z_n) \rightarrow (X, Y, Z) \) 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 \( \sigma^2 \). Define

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

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

\[
W_n^{(k)}(t) = \int_0^t W_n^{(k-1)}(s -) \, dW_n^{(1)}(s).
\]

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 iterated integral. [Note that \( X_n \Rightarrow X \) in \( D_\mathbb{E} \) implies that \( (X_n, X_n) \Rightarrow (X, X) \) in \( D_{\mathbb{R}^{2d}} \).]

3.2 Example [Bobkoski (1983)]. Let \( \{ \xi_i \} \) be as before. For a constant \( \phi \), let \( \{ Y_k \} \) satisfy

\[
Y_{k+1} = \phi Y_k + \xi_{k+1}.
\]

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

\[
\sum Y_k (Y_{k+1} - \phi Y_k) = 0,
\]
given by

\[
\hat{\phi} = \frac{\sum Y_k Y_{k+1}}{\sum Y_k^2}.
\]

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

\[
X_n(t) = \phi_n^{[nt]} X_n(0) + \int_0^t \phi_n^{[nt]-1} dW_n(s),
\]
where \( W_n = W_n^{(1)} \) and if \( X_n(0) \to X(0) \), it follows that \( X_n \Rightarrow X \) given by

\[
X(t) = e^{-\beta t}X(0) + \int_0^t e^{-\beta(t-s)} \sigma \, dW(s).
\]

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

\[
\sum_{k=0}^{[nt]-1} Y^n_k ( (\phi_n - \hat{\phi}_n) Y^n_k + \xi_{k+1} ) = 0,
\]

which implies

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

and it follows that

\[
n(\phi_n - \hat{\phi}_n) = \frac{\int_0^t \sigma X(s) \, dW(s)}{\int_0^t X(s)^2 \, ds}.
\]

More general results along these lines have been given by Llatas (1987), Chan and Wei (1988) and Cox and Llatas (1989).

3.3 Example. Work on approximation of nonlinear filters [DiMasi and Runggaldier (1981, 1982), Johnson (1983) and 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 \( D_{\mathbb{R}}[0, \infty) \), 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

\[
Y_n(t) = N \left( n \int_0^t \left( \lambda + n^{-1/2} h\left( X_n(s) \right) \right) \, ds \right)
\]

and define

\[
U_n(t) = n^{-1/2}(Y_n(t) - \lambda nt).
\]

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

\[
U(t) = \sqrt{\lambda} W(t) + \int_0^t h(X(s)) \, ds.
\]

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 $\mathcal{F}_t^n = \sigma(X_n(s), U_n(s); s \leq t)$ with
\[
L_n(t) = \left. \frac{dP_n}{dQ_n} \right|_{\mathcal{F}_t^n} = \exp\left( \int_0^t \ln(1 + n^{-1/2}\lambda^{-1}h(X_n(s))) \, dY_n(s) - \int_0^t n^{1/2}h(X_n(s)) \, ds \right)
\]
(3.14)
\[
= \exp\left( \int_0^t n^{1/2} \ln(1 + n^{-1/2}\lambda^{-1}h(X_n(s))) \, dU_n(s) \right.
\]
\[
+ \left. \int_0^t (n\lambda \ln(1 + n^{-1/2}\lambda^{-1}h(X_n(s))) - n^{1/2}h(X_n(s))) \, ds \right).
\]
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 $\mathcal{F}_t = \sigma(X(s), U(s); s \leq t)$ with
\[
L(t) = \left. \frac{dP}{dQ} \right|_{\mathcal{F}_t} = \exp\left( \int_0^t \lambda^{-1}h(X(s)) \, dU(s) - \int_0^t \frac{1}{2}\lambda^{-1}h^2(X(s)) \, ds \right).
\]
(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 $\mathcal{F}_Y^n$ converges in distribution to the conditional distribution $\mu(t)$ of $X(t)$ given $\mathcal{F}_t^U$ as a process in $D(\mathbb{R})[0, \infty)$.

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
\[
d[X]_t = dt + F(X(t-)) \, dX(t),
\]
or, equivalently,
\[
X(t)^2 - X(0)^2 - 2 \int_0^t X(s-) \, dX(s) = t + \int_0^t F(X(s-)) \, dX(s).
\]
(3.16)
(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 (1990), Section 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
\[
\Delta Y_n(k)^2 = \frac{1}{n} + F(Y_n(k)) \Delta Y_n(k).
\]
(3.18)
Consequently,

\begin{equation}
\Delta Y_n(k) = \frac{F(Y_n(k)) \pm \sqrt{F(Y_n(k))^2 + 4/n}}{2} = \Delta_n^\pm(k)
\end{equation}

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

\[
P[\Delta Y_n(k) = \Delta_n^+(k)] = 1 - P[\Delta Y_n(k) = \Delta_n^-(k)]
\]

\begin{equation}
= \frac{\Delta_n^-(k)}{\Delta_n^-(k) - \Delta_n^+(k)}.
\end{equation}

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

\begin{equation}
E[(X_n(t + h) - X_n(t))^2|\mathcal{F}_t^{X_n}] = \frac{[n(t + h)]}{n} - \frac{[nt]}{n}.
\end{equation}

The relative compactness of \( \{X_n\} \) (and hence for \( \{X_n, F^\alpha X_n\} \)) follows easily [see, e.g., Ethier and Kurtz (1986), Remark 3.8.7]. Since \( X_n \) satisfies

\begin{equation}
X_n(t)^2 - X_n(0)^2 - 2\int_0^t X_n(s -) dX_n(s)
\end{equation}

\begin{equation}
= \frac{[nt]}{n} + \int_0^t F(X_n(s -)) dX_n(s),
\end{equation}

we see that any limit point of the sequence \( \{X_n\} \) satisfies (3.17). More generally, the above construction will give solutions of

\begin{equation}
d[X]_t = dt + F(X, t -) dX(t)
\end{equation}

for any \( F: D_p[0, \infty) \to D_p[0, \infty) \) satisfying C5.4(i) and C5.4(ii) and \( F(x, t) = F(x^t, t) \) for all \( x \in D_p[0, \infty) \) and \( t \geq 0 \), where \( x^t = x(\cdot \wedge t) \).

3.5 Example [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

\begin{equation}
\int_0^1 \int_0^1 h^2(x, y) \, dx \, dy < \infty
\end{equation}

and

\begin{equation}
\int_0^1 h(x, y) \, dx = \int_0^1 h(x, y) \, dy = 0.
\end{equation}

Define

\begin{equation}
Z_n^h = \frac{1}{n} \sum_{1 \leq i < j \leq n} h(\xi_i, \xi_j).
\end{equation}

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 \( F_n \):

\begin{equation}
F_n(t) = \frac{1}{n} \sum_{i=1}^n \chi_{(\xi_i, \infty)}(t).
\end{equation}
In terms of $F_n$, $Z^h_n$ can be written

\[(3.28) \quad Z^h_n = n \int_0^t h(s,t) \, dF_n(s) \, dF_n(t)\]

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

\[(3.29) \quad Z^h_n = \int_0^t h(s,t) \, dB_n(s) \, dB_n(t).\]

If $g$ satisfies the same conditions as $h$, then

\[(3.30) \quad E\left[\left( Z^h_n - Z^g_n \right)^2 \right] = \frac{n(n-1)}{2n^2} \int_0^1 \int_0^1 (h(x,y) - g(x,y))^2 \, dx \, dy.\]

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

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

and, since $B_n \Rightarrow B$, the Brownian bridge [see, e.g., Billingsley (1968), Sections 13 and 19, or Protter (1990), Section V.6], the continuous mapping theorem implies that $X_n \Rightarrow X$ given by

\[(3.32) \quad X(t) = \int_0^t h(s,t) \, dB(s).\]

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

The process $B_n$ is a semimartingale with decomposition

\[(3.33) \quad B_n(t) = \sqrt{n} \, (F_n(t) - t) = \sqrt{n} \left( F_n(t) - \int_0^t \frac{1 - F_n(s)}{1 - s} \, ds \right) \]

\[- \sqrt{n} \int_0^t \frac{F_n(s) - s}{1 - s} \, ds = M_n(t) - \int_0^t \frac{1}{1 - s} \, B_n(s) \, ds.\]

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

\[(3.34) \quad E\left[ T \left( \int_0^t \frac{1}{1 - s} \, B_n(s) \, ds \right) \right] = E\left[ \int_0^t \frac{1}{1 - s} |B_n(s)| \, ds \right] \leq \int_0^t \frac{1}{1 - s} \sqrt{E[B_n(s)^2]} \, ds \]

\[= \int_0^t \frac{s}{1 - s} \, ds < \infty.\]
for $t \leq 1$. Consequently, the conditions of Theorem 2.2 are satisfied and $Z_n^h$ converges in distribution to

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

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 previously used techniques 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

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

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

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

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

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

where $\theta_n(t) = \theta_{\lfloor nt \rfloor}^n$. Suppose that $\{Y_n\}$ satisfies C2.2(i) for some $\delta$ and that $(Y_n, \theta_n, S_n(0)) \Rightarrow (Y, \theta, S(0))$ (in $D_{R_+}(0, \infty) \times \mathbb{R}$). Then the limiting equation

$$S(t) = S(0) + \int_0^t S(s-) \, dY(s)$$

has a (locally) unique global solution, so by Theorem 5.4 [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

$$G(t) = \int_0^t \theta(s-) \, dS(s).$$

The solution of (3.39) with $S(0) = 1$ is called the stochastic or Doléans–Dade
exponential and is denoted \( \varphi(X) \). The general solution is then given by \( S = S(0)\varphi(X) \) [Protter (1990), Section II.8].

3.7 Example. For each \( n \), let \( Y_n \) be an \( \{\mathcal{F}_t^n\} \)-semimartingale and let \( \{\tau_k^n\} \) be a sequence of \( \{\mathcal{F}_t^n\} \)-stopping times with \( \tau_0^n = 0 \) and \( \lim_{t \to \infty} \tau_k^n = \infty \). Define

\[
\tilde{Y}_n(t) = Y_n(\tau_k^n), \quad \tau_k^n \leq t < \tau_{k+1}^n.
\]

Suppose that \( Y_n \Rightarrow Y \) and that \( \{Y_n\} \) satisfies C2.2(i) for some \( \delta \in (0, \infty) \). (In particular, this last statement holds if \( Y_n = Y \) for all \( n \).) Then \( \{(\tilde{Y}_n)\} \) is relatively compact in the Skorohod topology and satisfies C2.2(i). If \( \sup_n \tau_k^n - \tau_k^n \to 0 \), then \( \tilde{Y}_n \Rightarrow Y \). Since the increments of \( \tilde{Y}_n \) can be estimated in terms of the increments of \( Y_n \), the relative compactness of \( \{(\tilde{Y}_n)\} \) follows easily [see Theorem 3.7.2 of Ethier and Kurtz (1986)]. The convergence assertion follows from Proposition 3.6.5 of Ethier and Kurtz (1986). To see that C2.2(i) is satisfied, define

\[
\tilde{A}_n^\delta(t) = A_n^\delta(\tau_k^n), \quad \tilde{M}_n^\delta(t) = M_n^\delta(\tau_k^n), \quad \tau_k^n \leq t < \tau_{k+1}^n.
\]

If \( \delta = \infty \), then \( \tilde{Y}_n = \tilde{M}_n^\infty + \tilde{A}_n^\infty \), \( \tilde{E}[|\tilde{M}_n^\infty|_{\tau_k^n \wedge \tau_{k+1}^n}] \leq \tilde{E}[|M_n^\infty|_{\tau_k^n \wedge \tau_{k+1}^n}] \) and \( \tilde{E}[T_{t \wedge \tau_k^n}(\tilde{A}_n^\infty)] \leq \tilde{E}[T_{t \wedge \tau_k^n}(A_n^\infty)] \) so C2.2(i) holds. If \( \delta < \infty \), then

\[
\tilde{Y}_n = J_\delta(\tilde{Y}_n) + \tilde{M}_n^\delta + \tilde{A}_n^\delta + (J_\delta(Y_n) - J_\delta(\tilde{Y}_n)),
\]

where \( J_\delta(Y_n) = J_\delta(Y_n)(\tau_k^n) \) for \( \tau_k^n \leq t < \tau_{k+1}^n \). As before, \( \tilde{E}[|\tilde{M}_n^\delta|_{\tau_k^n \wedge \tau_{k+1}^n}] \leq \tilde{E}[|M_n^\delta|_{\tau_k^n \wedge \tau_{k+1}^n}] \) and we claim that there exist stopping times \( \tilde{\tau}_n^\alpha \) satisfying \( P(\tilde{\tau}_n^\alpha \leq \alpha) \leq 1/\alpha \) and \( \sup_n \tilde{E}[T_{t \wedge \tilde{\tau}_n^\alpha}(\tilde{A}_n^\delta + J_\delta(Y_n) - J_\delta(\tilde{Y}_n))] < \infty \). The relative compactness of \( \{Y_n\} \) and \( \{(\tilde{Y}_n)\} \) implies that \( T_{t}(J_\delta(Y_n) - J_\delta(\tilde{Y}_n)) \) is stochastically bounded for each \( t \). For each \( \alpha > 0 \), select \( \alpha \) such that \( P(T_n(J_\delta(Y_n) - J_\delta(\tilde{Y}_n)) \geq \alpha) \leq 1/2\alpha \), and define \( \eta_n^\alpha = \inf(t: T_n(J_\delta(Y_n) - J_\delta(\tilde{Y}_n)) \geq \alpha) \) and \( \tilde{\tau}_n^\alpha = \tau_{2\alpha} \wedge \eta_n^\alpha \). Then, noting that the magnitude of the discontinuities of \( \tilde{M}_n^\delta + \tilde{A}_n^\delta + (J_\delta(Y_n) - J_\delta(\tilde{Y}_n)) \) is at most \( \delta \),

\[
\tilde{E}[T_{t \wedge \tilde{\tau}_n^\alpha}(\tilde{A}_n^\delta + J_\delta(Y_n) - J_\delta(\tilde{Y}_n))]
\]

\[
\leq \tilde{E}[T_{t \wedge \tilde{\tau}_n^\alpha}(A_n^\delta)] + c_\alpha + \delta + \tilde{E}[|J_\delta(Y_n) - J_\delta(\tilde{Y}_n)|_{t \wedge \tilde{\tau}_n^\alpha}].
\]

(3.44)

\[
\leq 2\tilde{E}[T_{t \wedge \tilde{\tau}_n^\alpha}(A_n^\delta)] + c_\alpha + \delta + \sqrt{E[|M_n^\delta|_{t \wedge \tilde{\tau}_n^\alpha}]},
\]

and C2.2(i) follows.
4. Relative compactness and additional convergence results. The conditional variation on \([0,t]\) of a process \(X\) with respect to a filtration \(\mathcal{F}_t\) is defined by \(V_t(X) = \sup E[\sum |E[X(t_{i+1}) - X(t_i)| |\mathcal{F}_t]]\), where the supremum is over all partitions of \([0,t]\). (For vector-valued \(X\) we take \( |x| = \sum |x_i| \).) For a stopping time \(\tau\), \(X^\tau\) will denote the stopped process given by \(X^\tau(t) = X(t \wedge \tau)\).

4.1 Lemma. For \(n = 1, 2, \ldots\), let \(X_n\) and \(Y_n\) be \((\mathcal{F}_t^n)\)-adapted, \(X_n\) in \(D_{\mathbb{R}^n}[0, \infty)\) and \(Y_n\) in \(D_{\mathbb{R}^n}[0, \infty)\). Assume the following condition:

\[
\text{C4.1} \quad \text{For each } \alpha > 0, \text{ there exist stopping times } (\tau_n^\alpha) \text{ with } \Pr(\tau_n^\alpha \leq \alpha) \leq 1/\alpha \text{ such that for each } t \geq 0, \\
\sup_n E[|Y_n^{\tau_n^\alpha}(t)|] < \infty \text{ and } \sup_n V_t(Y_n^{\tau_n^\alpha}) < \infty \text{ (where the conditional variation for } Y_n \text{ is with respect to } \mathcal{F}_t^n). 
\]

Let \(H_n(t) = \sup_{s \leq t} |X_n(s)|\) and suppose that \(\{H_n(t)\}\) is stochastically bounded for each \(t\). Define

\[
Z_n(t) = \int_0^t X_n(s -) \, dY_n(s).
\]

Then \(\{Z_n\}\) satisfies C4.1 and there exist strictly increasing, \((\mathcal{F}_t^n)\)-adapted processes \(C_n\) with \(C_n(0) = 0\), \(C_n(t + h) - C_n(t) \geq h\) and \(\{C_n(t)\}\) stochastically bounded for all \(t, h \geq 0\), such that, defining \(\gamma_n = C_n^{-1}\), \(Y_n(t) = Y_n(\gamma_n(t))\) and \(Z_n(t) = Z_n(\gamma_n(t))\), \((Y_n, Z_n, \gamma_n)\) is relatively compact in \(D_{\mathbb{R}^m} \times \mathbb{R}^k \times \mathbb{R}[0, \infty)\).

4.2 Remark. (a) Note that \(Z_n(t) = Z_n(\gamma_n^{-1}(t))\).

(b) Theorem 3.5 of Kurtz (1991) gives conditions on the sequence \(\{C_n\}\) which imply relative compactness for \((Y_n, Z_n)\). This theorem is an extension of Theorem 2.3 of Jacod, Mémin and Métrivier (1983).

(c) The result will also hold under the assumption that \(X_n\) is predictable and \(H_n\) is a right-continuous, adapted, increasing process satisfying \(|X_n(s)| \leq H_n(t)\) for \(s \leq t\), with the usual extension of the stochastic integral to predictable integrands.

(d) Let

\[
Y_n = \sum_{k=0}^{n-1} \frac{(-1)^k}{2n} X(1 + k/n^2, 1 + (k+1)/n^2)
\]

and \(X_n = -\text{sign}(Y_n)\). Then the conditions of the lemma are satisfied and

\[
Z_n = \sum_{k=1}^{n-1} \frac{1}{n} X(1 + k/n^2, \omega) + \frac{1}{2n} X(1 + 1/n, \omega) .
\]

The \(\gamma_n\) can be selected so that \(\dot{\gamma}_n = 1/n\) on the interval \([1, 2)\) and \(\dot{\gamma}_n = 1\) otherwise. The sequence \(\{\dot{Z}_n\}\) then converges in \(D_{\mathbb{R}^n}[0, \infty)\) to a continuous, piecewise linear function. Note that \(\{Z_n\}\) does not converge in the Skorohod topology. (We thank Mémin and Slominski for bringing this example to our attention and pointing out a serious error in an earlier version of this paper.)
PROOF. Each $Y_n$ has a unique decomposition $Y_n = M_n + B_n$, where $M_n$ is a local martingale and $B_n$ is a predictable finite variation process satisfying $E[T_{\tau_n}(B_n)] \leq V_{\alpha}(Y_n)$ [see Kurtz (1991), Proposition 5.1]. If we write

$$Z_n(t) = \int_0^t X_n(s-) \, dM_n(s) + \int_0^t X_n(s-) \, dB_n(s),$$

the first term is a local martingale and the total variation of the second term up to time $t$ is bounded by $H_n(t - T_Y(B_n))$. Consequently, if $\sigma_n$ is a stopping time so that the first term stopped is a martingale, $H_n(\sigma_n -) \leq c$ and $\sigma_n \leq \tau_n$, then $V_t(Z_n^\alpha) \leq cV_t(Y_n^\alpha)$. But for any $\beta > 0$, $c$ and $\alpha$ can be selected so that $P(\sigma_n \leq \beta) \leq 1/\beta$ and it follows that $(Z_n)$ satisfies C4.1. This in turn implies that $(Y_n, Z_n)$ satisfies C4.1. Corollary 1.3 of Kurtz (1991) then gives the other conclusions. □

4.3 Proposition. Let $((U_n, Y_n))$ be relatively compact (in the sense of convergence in distribution) in $D_{R^k \times R^m}(0, \infty)$ with $(U_n, Y_n)$ adapted to $(\mathcal{F}_n^\alpha)$ and $(Y_n)$ satisfying C2.2(i) for some $\delta > 0$. Suppose that $X_n$ has sample paths in $D_{M^k \times M^m}(0, \infty)$ and is adapted to $(\mathcal{F}_n^\alpha)$. Define

$$Z_n(t) = U_n(t) + \int_0^t X_n(s-) \, dY_n(s).$$

Suppose there exist strictly increasing, $(\mathcal{F}_n^\alpha)$-adapted processes $C_n$, with $C_n(t + h) - C_n(t) \geq h$ and $(C_n(t))$ stochastically bounded for all $t, h \geq 0$ such that, defining $\gamma_n = C_n^{-1}$, $U_n(t) = U_n(\gamma_n(t))$, etc., $((U_n, Z_n, X_n, Y_n, \gamma_n))$ is relatively compact in $D_{R^k \times M^k \times R^m \times R^\alpha}(0, \infty)$. Then $((Z_n, U_n, Y_n))$ is relatively compact in $D_{R^k \times R^k \times R^m}(0, \infty)$.

PROOF. For technical reasons, we extend the definition of the processes to the time interval $[-1, \infty)$ by setting $U_n(t) = U_n(0), X_n(t) = X_n(0), Y_n(t) = Y_n(0)$ and $C_n(t) = t$ for $-1 \leq t < 0$. These definitions ensure that $((U_n, Z_n, X_n, Y_n, \gamma_n))$ is relatively compact in $D_{R^k \times M^k \times R^m \times R^\alpha}[-1, \infty)$.

The fact that $(Y_n)$ satisfies C2.2(i) implies that $(\hat{Y}_n)$ satisfies C2.2(i). Consequently, selecting a convergent subsequence from $((U_n, Z_n, X_n, Y_n, \gamma_n))$ with limit $(\hat{U}, \hat{X}, \hat{Y}, \gamma)$, by Theorem 2.2, $((\hat{Z}, \hat{U}, \hat{X}, \hat{Y}, \gamma_n))$ converges to $(\hat{Z}, \hat{U}, \hat{X}, \hat{Y}, \gamma)$, where

$$\hat{Z}(t) = \hat{U}(t) + \int_{-1}^t \hat{X}(s-) \, d\hat{Y}(s) = \hat{U}(t) + \int_0^t \hat{X}(s-) \, d\hat{Y}(s).$$

We may assume that $(U_n, Y_n)$ converges along the same subsequence and the limit must be $(\hat{U}, Y) = (\hat{U} \circ \gamma^{-1}, \hat{Y} \circ \gamma^{-1})$, where $\gamma^{-1}(t) = \inf(u: \gamma(u) > t)$. [Note that $\gamma^{-1}$ is defined so that it is right-continuous and that the conditions on $C_n$ imply $\gamma^{-1}(t) = t$ for $t \leq 0$.] Lemma 2.3 of Kurtz (1991) (with the obvious modification for the time interval $[-1, \infty)$) implies that $(U_n, Y_n) \Rightarrow (\hat{U}, Y)$ in the Skorokhod topology if and only if on any interval on which $\gamma$ is constant, $(\hat{U}, \hat{Y})$ is constant except for at most one jump. But on any interval on which $(\hat{U}, \hat{Y})$ is constant, $\hat{Z}$ is constant and $\hat{Z}$ jumps only when $\hat{U}$ or $\hat{Y}$
jumps. Consequently, on any interval on which $\gamma$ is constant, $(\dot{Z}, \dot{U}, \dot{Y})$ is constant except for at most one jump. Applying the cited lemma again, we have that, along the subsequence, $(Z_n, U_n, Y_n) \Rightarrow (Z, U, Y)$ in the Skorohod topology on $D_{\mathbb{R}^k \times [0, \infty]} \times \mathbb{R}^m \times \mathbb{R}^k \times \mathbb{R}$ $[-1, \infty)$. But $(Z, U, Y)$ must be continuous at 0, so the convergence holds in $D_{\mathbb{R}^k \times [0, \infty]} \times \mathbb{R}^m \times \mathbb{R}$ as well, and the proposition follows.

4.4 Corollary. Let $((U_n, X_n, Y_n))$ have sample paths in $D_{\mathbb{R}^k \times [0, \infty]} \times \mathbb{R}^m \times \mathbb{R}$ and be adapted to $(\mathcal{F}_t^n)$ and let $Z_n$ be given by (4.5). Suppose that $((U_n, X_n))$ is relatively compact in $D_{\mathbb{R}^k \times [0, \infty]} \times \mathbb{R}^m \times \mathbb{R}$, that $(Y_n)$ satisfies C2.2(i) for some $\delta > 0$ and that $(X_n)$ satisfies C4.1. Then $((Z_n, U_n, Y_n))$ is relatively compact in $D_{\mathbb{R}^k \times [0, \infty]} \times \mathbb{R}^m \times \mathbb{R}$.

Proof. The relative compactness of $((U_n, Y_n, (X_n, dX_n^\delta(Y_n))))$ is immediate. Since the stochastic integral on the right of (4.5) 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 $(dX_n^\delta)$ is relatively compact [see, e.g., Kurtz (1991), Lemma 2.2].

C2.2(i) implies C4.1 for $(Y_n^\delta)$. Consequently, Corollary 1.3 of Kurtz (1991) implies the existence of strictly increasing, $(\mathcal{F}_t^n)$-adapted processes $C_n$, with $C_n(0) = 0$, $C_n(t + h) - C_n(t) \geq h$ and $(C_n(t))$ stochastically bounded for all $t, h \geq 0$, such that, defining $\gamma_n = C_n^{-1}$, $\hat{Y}_n^\delta(t) = Y_n^\delta(\gamma_n(t))$ and $\hat{X}_n(t) = X_n(\gamma_n(t))$, $(\hat{Y}_n^\delta, \hat{X}_n, \gamma_n)$ is relatively compact in $D_{\mathbb{R}^m \times [0, \infty]}$. Defining

$$V_n(t) = \int_0^t X_n(s -) \, dY_n^\delta(s),$$

(4.7)

Proposition 4.3 implies $((Y_n^\delta, V_n))$ is relatively compact in $D_{\mathbb{R}^m \times [0, \infty]}$. □

4.5 Corollary. Suppose $((U_n, Y_n, X_n))$ is relatively compact in $D_{\mathbb{R}^k \times [0, \infty]} \times D_{\mathbb{R}^m \times [0, \infty]} \times \mathbb{R}^k$, $(Y_n^\delta)$ satisfies C2.2(i) for some $\delta > 0$ and $Z_n$ is given by (4.5). Then $((U_n, Y_n, Z_n))$ is relatively compact in $D_{\mathbb{R}^k \times [0, \infty] \times [0, \infty]}$.

Proof. Let $W_n = (U_n, Y_n, X_n)$. The idea of the proof is to define a positive function $h(r, s)$ which is nondecreasing in $r$ and nonincreasing in $s$ such that

$$C_n(t) = t + \sum_{s \leq t} h(\lvert W_n(s -) - W_n(s -) \lvert, s)$$

(4.8)

satisfies the hypotheses of Proposition 4.3. Note that $C_n$ is designed so that the successive discontinuities of $W_n \circ C_n^{-1}$ are separated by a deterministic function of the size of the first discontinuity. Lemma 2.2 of Kurtz (1991) then gives the relative compactness. The difficulty arises in ensuring that $C_n(t)$ is stochastically bounded for each $t$. For $k = 1, 2, \ldots$, let $N_n^k(t)$ be the number of discontinuities of $W_n$ before time $t$ satisfying $1/k \leq \lvert W_n(s) - W_n(s -) \lvert < 1/(k - 1)$. The relative compactness of $(W_n)$ in $D_{\mathbb{R}^k \times [0, \infty]} \times D_{\mathbb{R}^m \times [0, \infty]}$ ensures that $(N_n^k(t))$ is stochastically bounded for each $t$ and that

$$\lim_{s \to 0} \sup_n P(N_n^k(s) > 0) = 0.$$
independent of $n$ such that

\[(4.9) \hspace{1cm} \sup_n P \left( a_k(t) N^k_n(t) > \frac{1}{2^k} \right) \leq \frac{1}{2^k}. \]

Without loss of generality, we can take $a_k(t)$ to be nonincreasing in $t$ and $k$. Define

\[(4.10) \hspace{1cm} C_n(t) = t + \sum_{m=0}^{\infty} \sum_{k=1}^{\infty} a_k(m + 1) (N^k_n((m + 1) \wedge t) - N^k_n(m \wedge t)). \]

Note that the first sum in (4.10) is in fact finite and that (4.9) implies by Borel–Cantelli that only finitely many terms in the second sum exceed $1/2^k$. To check the stochastic boundedness of $\{C_n(t)\}$, it is enough to check the stochastic boundedness of

\[(4.11) \hspace{1cm} K^m_n = \sum_{k=1}^{\infty} a_k(m) N^k_n(m) \]

for each $m$. We have

\[(4.12) \hspace{1cm} P \left( K^m_n > a + 1 \right) \leq \sum_{k=1}^{l} P \left( a_k(m) N^k_n(m) > \frac{a}{l} \right) + P \left( \sum_{k=l+1}^{\infty} a_k(m) N^k_n(m) > \frac{1}{2^l} \right) \leq \sum_{k=1}^{l} P \left( a_k(m) N^k_n(m) > \frac{a}{l} \right) + \frac{1}{2^l} \]

and the stochastic boundedness of $\{K^m_n\}$ follows easily from the stochastic boundedness of the $N^k_n(m)$. \( \Box \)

These relative compactness results lead 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 $D_{\mathbb{R}^m \times [0, \infty)} \times D_{\mathbb{R}^m \times [0, \infty)}$ (rather than in $D_{\mathbb{R}^m \times \mathbb{R}^m \times [0, \infty)}$ 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 $D_{\mathbb{R}^m \times \mathbb{R}^m \times [0, \infty)}$. Consequently, for each such $\epsilon$,

\[(4.13) \hspace{1cm} \int_0^t X_n(s - \epsilon - ) \, dY_n(s) \Rightarrow \int_0^t X(s - \epsilon - ) \, dY(s) \]

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

\[(4.14) \hspace{1cm} \int_0^t X_n(s - \epsilon_n - ) \, dY_n(s) \Rightarrow \int_0^t X(s - \epsilon - ) \, dY(s). \]

Noting that $\{/X_n, dY_n\}$ is relatively compact by Corollary 4.5, assume that
\[ \int_0^t (X_n(s) - X_n(s - \epsilon_n -)) \, dY_n(s) = Z(\cdot) - \int_0^t X(s) \, dY(s). \]

Note that the sequence on the left in (4.15) is relatively compact by an argument similar to that used in the proof of Corollary 4.5.

Let \( J_{\delta}^i(X_n) \) denote the \( M^k \times m \)-valued process whose \( ij \)-th component is \( J_{\delta}^{ij}(X_n) \), where \( X_n^{ij} \) is the \( ij \)-th component of \( X_n \) and let \( X_n^\delta = X_n - J_{\delta}(X_n) \). Let \( V_n^\delta(t) = \sup_{s \leq t} |X_n^\delta(s) - X_n^\delta(s - \epsilon_n)| \). Then \( V_n^\delta \to V^\delta \) given by \( V^\delta(t) = \sup_{s \leq t} |X^\delta(s) - X^\delta(s - )| \leq \sqrt{kn} \delta \). By the same type of estimate as in (2.7), to identify the right side of (4.15) it is enough to identify the limit of

\[ U_n^\delta(t) = \int_0^t (J_{\delta}(X_n)(s) - J_{\delta}(X_n)(s - \epsilon_n -)) \, dY_n(s) \]

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

\[ \sum_{\tau_n^\delta \leq t} (Y_n(\tau_n^\delta + \epsilon_n) - Y_n(\tau_n^\delta))(J_{\delta}(X_n)(\tau_n^\delta) - J_{\delta}(X_n)(\tau_n^\delta - )) \]

and any limit point \( U^\delta \) of \( \{U_n^\delta\} \) satisfies

\[ U^\delta(t) = \sum_{\beta_i \leq t} (J_{\delta}(X)(\beta_i^\delta) - J_{\delta}(X)(\beta_i^\delta - ))(Y(\beta_i^\delta) - Y(\beta_i^\delta - )), \]

where \( \{\beta_i^\delta\} \) is some subset of the times at which some component of \( X \) has a discontinuity larger than \( \delta \). Letting \( \delta \to 0 \), we see that

\[ U(t) = Z(t) - \int_0^t X(s) \, dY(s) \]

\[ = \sum_{\beta_i \leq t} (Y(\beta_i) - Y(\beta_i - ))(X(\beta_i) - X(\beta_i - )), \]

where \( \{\beta_i\} \) is some subset of the times at which both \( Y \) and \( X \) have discontinuities. From (4.17) 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.

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

\[ E[1 \wedge |Y_n(t + u) - Y_n(t)| F^n_t] \leq E[\gamma_n^T(\eta)|F^n_t], \]

\[ 0 \leq u \leq \eta, 0 \leq t \leq T, \]

and \( \lim_{n \to 0} \limsup_{n \to \infty} E[\gamma_n^T(\eta)] = 0. \)
If $(X_n, Y_n) \Rightarrow (X, Y)$ in $D_{bkh}[0, \infty) \times D_R[0, \infty)$, 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 $D_{bkm} \times R^m \times R^d[0, \infty)$. If $(X_n, Y_n) \rightarrow (X, Y)$ in probability, then the triple converges in probability.

4.7 Remark. See Ethier and Kurtz (1986), Theorem 3.8.6 and Remark 3.8.7 for the connection of (4.20) 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, Mémin and Pages (1989).

Proof. We need only show that $U \equiv 0$ in (4.19). The inequality in (4.20) holds with $t$ replaced by a stopping time. Consequently, we have [with reference to (4.17)] for $\varepsilon_n \leq \eta$,

$$E \left[ \sum_{i=1}^m 1 \wedge \left| (Y_n(\tau^\delta_{in} \wedge T + \varepsilon_n) - Y_n(\tau^\delta_{in} \wedge T)) \right| \right. \left. \times 1 \wedge \left| (J_\delta(X_n)(\tau^\delta_{in} \wedge T) - J_\delta(X_n)(\tau^\delta_{in} \wedge T -)) \right| \right]$$

(4.21)

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

$$\leq mE[\gamma_n^T(\eta)].$$

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 $U_n \rightarrow 0$ 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. $M_R[0, \infty)$ denotes the space of (equivalence classes of) measurable $E$-valued functions topologized by convergence in measure.

4.7 Theorem. For each $n$, let $(X_n, Y_n)$ be an $(\mathcal{F}_t^n)$-adapted process with sample paths in $D_{bkm} \times R^m[0, \infty)$. Suppose that $(Y_n)$ satisfies C2.2(i) for some $0 < \delta \leq \infty$ and that $(X_n)$ satisfies C4.1. If $(X_n, Y_n) \Rightarrow (X, Y)$ in $M_{bkm}[0, \infty) \times D_R[0, \infty)$ and $Y$ is continuous, then $X$ has a version with sample paths in
$D_{M^{h,m}}[0,\infty), Y$ is a semimartingale with respect to a filtration to which $X$ and $Y$ are adapted and $(X_n, Y_n, jX_n \, dY_n) \Rightarrow (X, Y, jX \, dY)$ in $M_{M^{h,m}}[0,\infty) \times D_{\mathbb{R}^m \times \mathbb{R}^k}(0,\infty)$. If $(X_n, Y_n) \rightarrow (X, Y)$ in $M_{M^{h,m}}[0,\infty) \times D_{\mathbb{R}^m}(0,\infty)$ in probability, then the triple converges in probability.

PROOF. Let $C_n, \gamma_n, \hat{X}_n, \text{ and } \hat{Y}_n$ be as in Corollary 4.4 and Proposition 4.3 and set $Z_n = jX_n \, dY_n$ and $\hat{Z}_n = Z_n \circ \gamma_n = j\hat{X}_n \, d\hat{Y}_n$. Then $\{(X_n, Y_n, Z_n, \hat{X}_n, \hat{Y}_n, \hat{Z}_n, \gamma_n)\}$ is relatively compact in $M_{M^{h,m}}[0,\infty) \times D_{\mathbb{R}^m \times \mathbb{R}^k}(0,\infty) \times D_{M^{h,m} \times \mathbb{R}^m \times \mathbb{R}^k \times \mathbb{R}^m}(0,\infty)$. If $(X, Y, Z, \hat{X}, \hat{Y}, \hat{Z}, \gamma)$ is a limit point, then $X = \hat{X} \circ \gamma^{-1}$, $Y = \hat{Y} \circ \gamma^{-1}$ and $Z = \hat{Z} \circ \gamma^{-1}$. Since $\hat{Y}$ is constant on any interval on which $\gamma$ is constant, which implies

$$Z(t) = \hat{Z} \circ \gamma^{-1}(t) = \int_0^t \hat{X} \circ \gamma^{-1}(s-) \, d\hat{Y} \circ \gamma^{-1}(s) = \int_0^t X(s-) \, dY(s)$$

and the theorem follows. □

The previous 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 $jX_n \, dY_n \Rightarrow jX \, dY$. 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 next.

4.8 THEOREM. Let $Y_n = M_n + A_n + Z_n$, where $\{(M_n, A_n, Z_n)\}$ satisfies the conditions of Theorem 2.7. Let $H_n(t) = \sup_{s \leq t} |X_n(s)|$ and suppose that $\{H_n(t)\}$ is stochastically bounded for each $t$. Define $X_n^h$ by

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

Suppose that for each $t > 0$ and $\varepsilon > 0$,

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

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

$$U(t) = \lim_{h \to 0} \int_0^t X_n^h \, dY$$

exists and $(X_n, Y_n, jX_n \, dY_n) \Rightarrow (X, Y, U)$ in $M_{M^{h,m}}[0,\infty) \times D_{\mathbb{R}^m \times \mathbb{R}^k}(0,\infty)$. If $(X_n, Y_n, Z_n) \rightarrow (X, Y, Z)$ in $M_{M^{h,m}}[0,\infty) \times D_{\mathbb{R}^m \times \mathbb{R}^k}(0,\infty)$ in probability, then $(X_n, Y_n, jX_n \, dY_n) \rightarrow (X, Y, U)$ converges in probability.
Proof. Since $X^h_n$ is locally Lipschitz, the conditions on $H_n$ ensure that $(X^h_n, Y_n, Z_n) \Rightarrow (X^h, Y, Z)$ in $D_{R^k \times R^k \times R^k}[0, \infty)$ and hence that $\int X^h_n \, dY_n \Rightarrow \int X^h \, dY$. Consequently, estimating as in (2.7), (4.24) implies the result. \[ \square \]

5. Stochastic differential equations. In this section we generalize results of Slomiński (1989) concerning convergence of sequences of solutions of stochastic differential equations [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 ($k = m = 1$):

\begin{equation}
X(t) = 1 + \int_0^t X(s -) \, dY(s).
\end{equation}

For $n = 1, 2, \ldots$ let $F_n: D_{\bar{R}^k}[0, \infty) \to D_{\bar{R}^k}[0, \infty)$, let $U_n$ and $Y_n$ be processes with sample paths in $D_{\bar{R}^k}[0, \infty)$ and $D_{\bar{R}^k}[0, \infty)$, 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 $F_n(x, t) = F_n(x', t)$ for all $t \geq 0$ and $x \in D_{\bar{R}^k}[0, \infty)$, where $x'(\cdot) = x(\cdot \wedge t)$. Let $X_n$ be adapted to $(\mathcal{F}_t^n)$ and satisfy

\begin{equation}
X_n(t) = U_n(t) + \int_0^t F_n(X_n, s -) \, dY_n(s).
\end{equation}

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

\begin{equation}
X(t) = U(t) + \int_0^t F(X, s -) \, dY(s),
\end{equation}

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 $D_{\bar{R}^k \times R^k}[0, \infty)$ implies $(x_n, y_n, F_n(x_n)) \to (x, y, F(x))$ in $D_{\bar{R}^k \times R^k \times R^k}[0, \infty)$ 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 $D_{\bar{R}^k \times \bar{R}^k \times \bar{R}^k}[0, \infty)$, 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).

Proof. 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. \[ \square \]
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 \( \Lambda^1 \) be the subset of absolutely continuous functions in \( \Lambda \) for which \( \gamma(\lambda) \equiv ||\ln \lambda||_\infty \) is finite.

5.2 Lemma. 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 \( t > 0 \),
\[
\sup_{x \in \mathcal{K}} \sup_{s \leq t} |F_n(x, s) - F(x, s)| \to 0.
\]

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

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 \( \lambda \in \Lambda^1 \),
\[
\sup_{x \in \mathcal{K}} \sup_{s \leq t} |F(x \circ \lambda, s) - F(x, \lambda(s))| \leq \omega(\gamma(\lambda)).
\]

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.

Proof. 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 bound time intervals. Consequently,
\[
F_n(x_n, \lambda_n(s)) - F(x, s)
\]
\[
= F_n(x_n, \lambda_n(s)) - F(x_n, \lambda_n(s)) + F(x_n, \lambda_n(s))
\]
\[
- F(x_n \circ \lambda_n, s) + F(x_n \circ \lambda_n, s) - F(x, s)
\]
goes to zero uniformly in \( s \) on bounded intervals. \( \square \)

5.3 Examples. Let \( g: \mathbb{R}^k \times [0, \infty) \to \mathbb{R}^{km} \) 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) \( F(x, t) = \int_0^t h(t - s)g(x(s), s) \text{ds} \).

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 a priori assumption that the sequence of solutions is relatively compact [see Theorem 2.3 of Jacod, Mémin and Métévier (1983) for general conditions on \( \{Y_n\} \) under which the desired relative compactness will hold]. We can avoid this assumption by localizing the result and applying Proposition 4.3. We say that \( (X, \tau) \) 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, \( \tau \) is an \( \{ \mathcal{F}_t \} \)-stopping time and

\[
X(t \wedge \tau) = U(t \wedge \tau) + \int_0^{t \wedge \tau} F(X, s -) \, dY(s).
\]

We say that strong local uniqueness holds for (5.3) if any two local solutions \((X_1(t), \tau_1), (X_2(t), \tau_2)\) satisfy \(X_1(t) = X_2(t), \ t \leq \tau_1 \wedge \tau_2\), a.s. To define a notion of weak local uniqueness (that is, uniqueness of distributions), we need to require that the stopping time associated with the solution be a measurable function of the solution. We say that \((\tilde{U}, \tilde{Y}, \tilde{X}, \tilde{\tau})\) is a weak local solution of (5.3) if \((\tilde{U}, \tilde{Y})\) is a version of \((U, Y)\) and (5.5) holds with \((U, Y, X, \tau)\) replaced by \((\tilde{U}, \tilde{Y}, \tilde{X}, \tilde{\tau})\). We say that weak local uniqueness holds for (5.3) if for any two weak local solutions \((U_1, X_1, \tau_1)\) and \((U_2, X_2, \tau_2)\) with \(\tau_1 = h_1(X_1)\) and \(\tau_2 = h_2(X_2)\) for measurable functions \(h_1, h_2\) on \(D_{\mathbb{R}}[0, \infty)\), \((X_1, h_1 \wedge h_2(X_1))\) and \((X_2, h_1 \wedge h_2(X_2))\) have the same distribution [see Protter (1990), Chapter V, for sufficient conditions for uniqueness].

In order to apply Proposition 4.3, we need assumptions on the properties of \(F_n(x)\) and \(F(x)\) under transformations of the time scale. Let \(T_1[0, \infty)\) denote the collection of nondecreasing mappings \(\lambda\) of \([0, \infty)\) onto \([0, \infty)\) in particular, \(\lambda(0) = 0\) such that \(\lambda(t + h) - \lambda(t) \leq h\) for all \(t, h \geq 0\). Let \(\iota\) denote the identity map \(\iota(s) = s\). We will assume that there exist mappings \(G_\lambda, G: D_{\mathbb{R}}[0, \infty) \times T_1[0, \infty) \to D_{\mathbb{R}}^+[0, \infty)\) such that \(F_n(x) \circ \lambda = G_n(x \circ \lambda, \lambda)\) and \(F(x) \circ \lambda = G(x \circ \lambda, \lambda)\) for \((x, \lambda) \in D_{\mathbb{R}}[0, \infty) \times T_1[0, \infty)\). We need the following strengthening of C5.2:

C5.4(i) For each compact subset \(\mathcal{K} \subset D_{\mathbb{R}}[0, \infty) \times T_1[0, \infty)\) and \(t > 0\),
\[
\sup_{(x, \lambda) \in \mathcal{K}} \sup_{s \leq t} |G_n(x, \lambda, s) - G(x, \lambda, s)| \to 0.
\]

C5.4(ii) For \((x_n, \lambda_n) \in D_{\mathbb{R}}[0, \infty) \times T_1[0, \infty), \sup_{s \leq t} |x_n(s) - x(s)| \to 0\) and \(\sup_{s \leq t} |\lambda_n(s) - \lambda(s)| \to 0\) for each \(t > 0\) implies
\[
\sup_{s \leq t} |G(x_n, \lambda_n, s) - G(x, \lambda, s)| \to 0.
\]

We note that each of the examples in 5.3 has a representation in terms of a \(G\) satisfying C5.4(ii) and that C5.4 implies C5.2.

5.4 Theorem. Suppose that \((U_n, X_n, Y_n)\) satisfies (5.2), \((U_n, Y_n) \Rightarrow (U, Y)\) in the Skorohod topology and that \((Y_n)\) satisfies C2.2(i) for some \(0 < \delta \leq \infty\). Assume that \(\{F_n\}\) and \(F\) have representations in terms of \(\{G_n\}\) and \(G\) satisfying C5.4. 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

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

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 \((X^b, \tau)\) of (5.3) with \(\tau = \eta^c = \inf(t: |F(X^b, t)| \vee |F(X^b, t^-)| \geq c)\) for any \(c < b\). If there exists a global solution \(X\) of (5.3) and weak local uniqueness holds, then \((U_n, X_n, Y_n) \Rightarrow (U, X, Y)\).
PROOF. By Lemma 4.1, there exist \( \gamma_n \) such that \((U_n \circ \gamma_n, X_n^b \circ \gamma_n, Y_n \circ \gamma_n)\) is relatively compact in \(D_{R^+ \times R^+ \times R^+}[0, \infty)\). C5.4 then implies that \(\{\chi_{(0, \eta_n^b)} \circ \gamma_n F_n(X_n^b) \circ \gamma_n\} = \{\chi_{(0, \eta_n^0)} \circ \gamma_n G_n(X_n^b \circ \gamma_n, \gamma_n)\}\) is relatively compact in \(D_{R^+}[0, \infty)\). The relative compactness of \((U_n, X_n^b, Y_n)\) then follows by Corollary 4.5 and Proposition 4.3. The sequence \((U_n, X_n^b, Y_n, \eta_n^b)\) will be relatively compact in \(D_{R^+ \times R^+ \times R^+}[0, \infty) \times [0, \infty)\). Let \((U_n, X_n^b, Y, \eta_0^b)\) 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
\[
X_n^b(t) = U_n(t) + \int_0^t F_n(X_n^b, s-) \, dY_n(s)
\]
for \(t \leq \eta_n^b\),
\[
X^b(t) = U(t) + \int_0^t F(X^b, s-) \, dY(s)
\]
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 \(\eta^c\) with \(|F(X^b, \eta^c^-) - F(X^b, \eta^c)| \leq c\) and \(|F(X^b, \eta^c)| \geq b\). It follows that for \(c \leq d < b\),
\[(U_n(\eta_n^d), X_n^b(\eta_n^d), Y_n(\eta_n^d), Y_n(\eta_n^d -), F_n(X_n^b, \eta_n^d), F_n(X_n^b, \eta_n^d -), \eta_n^d)\]
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
\[
X^b(\eta^d) = U(\eta^d) + \int_0^{\eta^d} F(X^b, s-) \, dY(s),
\]
so that (5.8) holds for \(t \leq \eta^c\) (\(= \eta^d\)). Consequently, \((X^b, \eta^c)\) is a local solution of (5.3).

Note that \(\eta^c\) is a measurable function of \(X^b\) [say \(h_c(X^b)\)]. Consequently, if weak local uniqueness holds for (5.3) and there exists a global weak solution \(\bar{X}\), then \((X^b, \eta^c)\) must have the same distribution as \((\bar{X}, h_c(\bar{X}))\) for all \(c\) and \(b\) with \(c < b\). Since \(\bar{X}\) is a global solution, \(h_c(\bar{X}) \to \infty\) as \(c \to \infty\). Convergence in distribution of \((U_n, X_n, Y_n)\) follows. \(\square\)

Unlike Theorem 2.2, Theorem 5.4 does not immediately hold with convergence in distribution replaced by convergence in probability. In particular, we must assume a strong uniqueness for the limiting equation (5.3) or convergence in probability could fail to hold even with \((U_n, Y_n) \equiv (U, Y)\). If \(X\) and \(\bar{X}\) are solutions of (5.3) which are not almost surely equal and \((\xi_n)\) are i.i.d. with \(P(\xi_n = 1) = P(\xi_n = 0) = \frac{1}{2}\), then take \(X_n = \xi_n X + (1 - \xi_n) \bar{X}\). We need the following lemma.

5.5 LEMMA. Assume that \(F\) has a representation in terms of a \(G\) satisfying C5.4(ii). Suppose that there exists a global (weak) solution of (5.3) and that strong uniqueness holds for (5.3) for any version of \((U, Y)\). Then any solution of (5.3) is a measurable function of \((U, Y)\) that is, if \(X\) satisfies (5.3),
then there exists a measurable mapping \( g: D_{\mathbb{R}^+ \times \mathbb{R}^m}[0, \infty) \to D_{\mathbb{R}}[0, \infty) \) such that \( X = g(U, Y) \) a.s.].

**Proof.** Define \( (U_n, Y_n) \) by \( (U_n(t), Y_n(t)) = (U([nt]/n), Y([nt]/n)) \) and let \( X_n \) satisfy (5.2). Then \( X_n \) is a measurable function of \( (U_n, Y_n) \) and hence of \( (U, Y) \). Let \( n_k \to \infty \) and \( m_k \to \infty \). Then by Theorem 5.4, \( (U_{n_k}, U_{m_k}, X_{n_k}, X_{m_k}) \) converges in distribution to \( (U, U, X, \tilde{X}, Y, Y) \), where \( X \) and \( \tilde{X} \) satisfy (5.3). But, strong local uniqueness implies \( X = \tilde{X} \) a.s. Consequently, if \( d \) is a metric for \( D_{\mathbb{R}}[0, \infty) \), then

\[
\lim_{k \to \infty} E[1 \wedge d(X_{n_k}, X_{m_k})] = E[1 \wedge d(X, X)] = 0
\]

and hence \( (X_n) \) is a Cauchy sequence for convergence in probability. Since \( X_n \) is a measurable function of \( (U, Y) \), the lemma follows. \( \Box \)

**5.6 Corollary.** If in Theorem 5.4, we assume that \( (U_n, Y_n) \) converges in probability to \( (U, Y) \), that there exists a global solution of (5.3), and that strong local uniqueness holds for (5.3) for any version of \( (U, Y) \), then \( X_n \) converges in probability.

**Proof.** Let \( f \) be a bounded, continuous function on \( D_{\mathbb{R}}[0, \infty) \) and \( g \) be a bounded continuous function on \( D_{\mathbb{R}^+ \times \mathbb{R}^m}[0, \infty) \). Then since \( (U_n, X_n, Y_n) \Rightarrow (U, X, Y) \),

\[
\lim_{n \to \infty} E[f(X_n)g(U_n, Y_n)] = E[f(X)g(U, Y)].
\]

The convergence in probability of \( (U_n, Y_n) \) then implies

\[
\lim_{n \to \infty} E[f(X_n)g(U, Y)] = E[f(X)g(U, Y)]
\]

and \( L^1 \)-approximation of measurable functions by continuous functions implies that (5.12) holds for all bounded, measurable \( g \). Lemma 5.5 ensures the existence of a bounded measurable \( g \) such that \( f(X) = g(U, Y) \) a.s. Consequently,

\[
\lim_{n \to \infty} E\left[ (f(X_n) - f(X))^2 \right] = \lim_{n \to \infty} \left( E[f(X_n)^2] - 2E[f(X_n)f(X)] + E[f(X)^2] \right) = 0
\]

and convergence in probability for \( (X_n) \) follows. \( \Box \)

Theorem 5.4 perhaps makes the theory look more simple and benign than it really is. Example 1.2 of the Introduction reveals a pathology originally discovered by Wong and Zakai (1965): that certain naive approximations of semimartingale differentials lead to a lack of continuity of the corresponding solutions of stochastic differential equations. Indeed, it was this pathology that led McShane to develop his integral and to his proposal of a canonical form [McShane (1975)], though these can now be recognized as special cases of the
semimartingale integral. The Wong–Zakai pathology has also led people to pay increased attention to Stratonovich and more generally symmetrized integrals and their differential equations [e.g., Mackevicius (1987)].

Examples 5.7, 5.8 and 5.9, that follow, motivate Theorem 5.10, which extends the results of Wong and Zakai. This extension is by no means the first; however, almost all the previous ones [e.g., Nakao and Yamato (1976), Doss (1977), Sussman (1978), Krener (1979), Ikeda and Watanabe (1981), Marcus (1981), Konecny (1983), Protter (1985), Mackevicius (1987), Picard (1989) and Bally (1989)] are concerned with $L^p$, almost sure, or in probability convergence, always on only one probability space. The one exception is Śłomiński (1989), who deals with weak convergence. Moreover, the level of generality in previous work is only that of Proposition 5.12 (albeit for more general approximation schemes; we have not bothered with the obvious modifications needed to include all of the previous results), and hence Theorem 5.10 is new even on the level of convergence in probability.

5.7 Example. Let $W_n$ be as in Example 1.2. Then clearly $\{W_n\}$ does not satisfy C2.2(i) (otherwise $\langle W_n, dW_n \rangle$ would converge to $\langle W, dW \rangle$); however, if we define $Y_n(t) = W([nt] + 1)/n$ and $Z_n = W_n - Y_n$, then $\{Y_n\}$ satisfies C2.2(i) ($Y_n$ is a martingale with respect to the filtration defined by $\mathcal{F}_t^n = \sigma(W(s); s \leq ([nt] + 1)/n)$) and $Z_n \to 0$. Furthermore, we observe that

$$\int_0^t Z_n \, dZ_n \to -\frac{1}{2} t,$$

(5.14)

$$[Z_n]_t = -[Y_n, Z_n] = Z_n^2(t) - 2\int_0^t Z_n \, dZ_n \to t$$

(5.15)

and [noting that $Z_n(t-) = 0$ at each discontinuity of $Z_n$]

$$T_t \left( \int Z_n \, dZ_n \right) = \int_0^t (Z_n(s) - W(s)) \, ds \to Ct,$$

(5.16)

where $C = E[|W(1)|\int_0^1 |W(1) - W(s)| \, ds]$. Setting $H_n = \|Z_n \, dZ_n$ and $I_n = \|[Z_n]$, it follows that $\{H_n\}$ and $\{I_n\}$ satisfy C2.2(i).

5.8 Example. Let $V$ be the Ornstein–Uhlenbeck process satisfying

$$dV = dW - V \, dt,$$

(5.17)

where $W$ is a standard Brownian motion. Let

$$W_n(t) = \frac{1}{n} \int_0^{tn^2} V(s) \, ds = \int_0^t n V(n^2 s) \, ds.$$

(5.18)

It follows that

$$W_n(t) = \frac{1}{n} W(n^2 t) - \frac{1}{n} V(n^2 t)$$

(5.19)

and defining $Y_n(t) = (1/n)W(n^2 t)$ and $Z_n(t) = -(1/n)V(n^2 t)$ we see that, as in Example 5.7, $\{Y_n\}$ satisfies C2.2(i) (each $Y_n$ is a standard Brownian motion)
and \( Z_n \to 0 \). Again, setting \( H_n = \int Z_n \, dZ_n \), \( K_n = [Y_n, Z_n] \) and \( I_n = [Z_n] \), we see that

\[
H_n(t) = \frac{1}{n^2} \int_0^{n^2t} V(s) \, dW(s) - \frac{1}{n^2} \int_0^{n^2t} V(s)^2 \, ds.
\]

The first term on the right of (5.20) is a martingale with quadratic variation

\[
\frac{1}{n^4} \int_0^{n^2t} V(s)^2 \, ds,
\]

while the second term obviously has finite variation. It follows that \((H_n)\) satisfies C2.2(i), and \( H_n(t) \to -\frac{1}{2} t \). Note, in addition, that \( I_n(t) = -K_n(t) = t \).

5.9 Example. Let \((U_k, k \geq 0)\) be a finite, irreducible Markov chain with transition matrix \( P = ((p_{ij})) \). Let \( \pi = (\pi_1, \ldots, \pi_M) \) give the stationary distribution and let \( f \) be a function satisfying

\[
\sum_m f(m) \pi_m = 0.
\]

Define

\[
W_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} f(U_k).
\]

Letting \( P g(i) = \sum_j g(j) p_{ij} \), by (5.22) there exists a function \( h \) such that \( Ph = f \). Substituting in (5.23), we obtain

\[
W_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} (Ph(U_k) - h(U_k))
\]

\[
= \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} (Ph(U_{k-1}) - h(U_k)) + \frac{1}{\sqrt{n}} (Ph(U_{[nt]}) - Ph(U_0))
\]

\[
= Y_n(t) + Z_n(t).
\]

As in Examples 5.7 and 5.8, \((Y_n)\) is a sequence of martingales satisfying C2.2(i) which, by the martingale central limit theorem [see, e.g., Ethier and Kurtz (1986), Theorem 7.1.4], converges in distribution to \( \sigma W \), where

\[
\sigma^2 = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{[nt]} (Ph(U_{k-1}) - h(U_k))^2
\]

\[
= \sum_{i,j} \pi_i p_{ij} (Ph(i) - h(j))^2.
\]
Again \( Z_n \to 0 \), \([Z_n] \to Ct\), \( \int Z_n \, dZ_n \to -\frac{1}{2} Ct \) and \([Y_n, Z_n] \to Dt\), where

\[
C = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{[nt]} (Ph(U_k) - Ph(U_{k-1}))^2
= \sum_{i,j} \pi_{ij} (Ph(j) - Ph(i))^2,
\]

\[
D = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{[nt]} (Ph(U_{k-1}) - h(U_k))(Ph(U_k) - Ph(U_{k-1}))
= \sum_{i,j} \pi_{ij} (Ph(i) - h(j))(Ph(j) - Ph(i))^2
\]

and \( \{\int Z_n \, dZ_n\} \) and \( \{[Z_n]\} \) satisfy C2.2(i).

Clearly Theorem 5.4 does not apply directly to

\[
X_n(t) = U_n(t) + \int_0^t F(X_n(s-)) \, dW_n(s)
\]

for \( \{W_n\} \) as in any of the above examples; however, if we specialize to

\[
X_n(t) = X_n(0) + \int_0^t F(X_n(s-)) \, dW_n(s)
= X_n(0) + \int_0^t F(X_n(s-)) \, dY_n(s) + \int_0^t F(X_n(s-)) \, dZ_n(s),
\]

we can apply Theorem 5.4 to obtain the following extension of the classical results of Wong and Zakai (1965).

5.10 Theorem. Let \( Y_n \) and \( Z_n \) be \( (\mathcal{F}_n) \)-semimartingales and let \( X_n(0) \) be \( \mathcal{F}_0 \)-measurable. Let \( F: \mathbb{R}^k \to \mathbb{R}^k \) in (5.29) be bounded and have bounded first and second order derivatives. Define \( H_n = (H_n^{\beta\gamma}) \) and \( K_n = (K_n^{\beta\gamma}) \) by

\[
H_n^{\beta\gamma}(t) = \int_0^t Z_n^{\beta}(s-) \, dZ_n^{\gamma}(s)
\]

and

\[
K_n^{\beta\gamma}(t) = [Y_n^{\beta}, Z_n^{\gamma}].
\]

Suppose that \( \{Y_n\} \) and \( \{H_n\} \) satisfy C2.2(i) and that \( (X_n(0), Y_n, Z_n, H_n, K_n) \Rightarrow (X(0), Y, 0, H, K) \). Then \( (X_n(0), Y_n, Z_n, H_n, K_n, X_n) \) is relatively compact and any limit point \( (X(0), Y, 0, H, K, X) \) satisfies

\[
X(t) = X(0) + \int_0^t F(X(s-)) \, dY(s)
+ \sum_{\alpha, \beta, \gamma} \int_0^t \partial_{\alpha} F_\beta(X(s-)) F_{\alpha\gamma}(X(s-)) \, d(H^{\beta\gamma}(s) - K^{\beta\gamma}(s)),
\]
where \( \partial_\alpha \) denotes the partial derivative with respect to the \( \alpha \)th variable and \( F_\beta \) denotes the \( \beta \)th column of \( F \).

5.11 Remark. (a) The boundedness assumptions on \( F \) and its derivatives may be dropped to obtain a localized result with a statement analogous to that of Theorem 5.4.

(b) The theorem can be extended to equations of the form

\[
X_n(t) = U_n(t) + \int_0^t F(X_n(s -)) \, dY_n(s) + \int_0^t F(X_n(s -)) \, dZ_n(s),
\]

by writing \( U_n = \hat{Y}_n + \hat{Z}_n \) and forming the system

\[
\begin{pmatrix}
U_n(t) \\
X_n(t)
\end{pmatrix}
= \begin{pmatrix}
U_n(0) \\
X_n(0)
\end{pmatrix}
+ \int_0^t \begin{pmatrix}
I & 0 \\
0 & F(X_n(s -))
\end{pmatrix}
\begin{pmatrix}
\hat{Y}_n(s) \\
\hat{Y}_n(s)
\end{pmatrix}
+ \int_0^t \begin{pmatrix}
I & 0 \\
0 & F(X_n(s -))
\end{pmatrix}
\begin{pmatrix}
\hat{Z}_n(s) \\
\hat{Z}_n(s)
\end{pmatrix},
\]

(5.34)

(c) Note that since \( Z_n \Rightarrow 0 \), \( \sup_{s \leq \tau} |Z_n(s) - Z_n(s -)| \Rightarrow 0 \) and \( H \) and \( K \) must be continuous. We are not, however, assuming that \( Y \) is continuous.

(d) Let \( I^{\beta\gamma}_n(t) = [Z^{\beta}_n, Z^{\gamma}_n]_t \). Since

\[
[Z^{\beta}_n, Z^{\gamma}_n]_t = Z^{\beta}_n(t)Z^{\gamma}_n(t) - Z^{\beta}_n(0)Z^{\gamma}_n(0)
\]

\[
- \int_0^t Z^{\beta}_n(s -) \, dZ^{\gamma}_n(s) - \int_0^t Z^{\gamma}_n(s -) \, dZ^{\beta}_n(s),
\]

(5.35)

it follows that \( I^{\beta\gamma}_n \Rightarrow -(H^{\beta\gamma} + H^{\gamma\beta}) \). Since \( I^{\beta\beta}_n \) is nondecreasing and converges in distribution to a continuous process, it follows that \( \{I^{\beta\beta}_n\} \) satisfies C2.2(i) (at least for any finite \( \delta \)) and, by estimating the increments of \( I^{\beta\gamma}_n \) by the increments of \( I^{\beta\beta}_n \) and \( I^{\gamma\gamma}_n \), that \( \{I^{\beta\gamma}_n\} \) satisfies C2.2(i).

(e) Since

\[
|K^{\beta\gamma}(t + h) - K^{\beta\gamma}(t)| \leq \frac{1}{2} ([Y^{\beta}_n]_{t+h} + [Z^{\gamma}_n]_{t+h} - [Y^{\beta}_n]_t - [Z^{\gamma}_n]_t),
\]

(5.36)

it follows that \( \{K_n\} \) satisfies C2.2(i).

Proof. The result is obtained by integrating the second term on the right of (5.28) by parts. Note that by Itô's formula

\[
F_{\beta}(X_n(t)) = F_{\beta}(X_n(0)) + \sum_\alpha \int_0^t \partial_\alpha F_{\beta}(X_n(s -)) \, dX^n_\alpha(s) + R^{i\beta}_n(t),
\]

(5.37)

where the increments of \( R^{i\beta}_n \) are dominated by a linear combination of the increments of \([Y^{\alpha}_n] \) and \([Z^{\gamma}_n] \) (which implies that \( \{R_n\} \) satisfies C2.2(i). Inte-
grating by parts, we obtain
\[
\int_0^t F^\beta_{i\beta}(X_n(s-)) \, dZ^\beta_n(s) = F^\beta_{i\beta}(X_n(t))Z^\beta_n(t) - F^\beta_{i\beta}(X_n(t\theta))Z^\beta_n(0)
\]
\[= \sum_{\alpha} \int_0^t \partial_{\alpha} F^\beta_{i\beta}(X_n(s-))Z^\beta_n(s-) \, dX^\alpha_n(s)
\]
\[= \int_0^t Z^\beta_n(s-) \, dR^i_{n\beta}(t) - \sum_{\alpha} \int_0^t \partial_{\alpha} F^\beta_{i\beta}(X_n(s-)) \, dX^\alpha_n(s)
\]
\[\times d\left[ X^\alpha_n, Z^\beta_n \right]_s + \left[ R^i_{n\beta}, Z^\beta_n \right]_t
\]
\[= \eta_n(t) - \sum_{\alpha, \gamma} \int_0^t \partial_{\alpha} F^\beta_{i\beta}(X_n(s-))F^\alpha_{\gamma\alpha}(X_n(s-))Z^\beta_n(s-) \, dZ^\gamma_n(s)
\]
\[= \sum_{\alpha, \gamma} \int_0^t \partial_{\alpha} F^\beta_{i\beta}(X_n(s-))F^\alpha_{\gamma\alpha}(X_n(s-)) \, d\left( \left[ Y^\gamma_n, Z^\beta_n \right]_s + \left[ Z^\gamma_n, Z^\beta_n \right]_s \right)
\]
\[= \eta_n(t) - \sum_{\alpha, \gamma} \int_0^t \partial_{\alpha} F^\beta_{i\beta}(X_n(s-))F^\alpha_{\gamma\alpha}(X_n(s-)) \, d\left( H^\beta_{\gamma\beta}(s) + K^\gamma_{\gamma\beta}(s) + I^\gamma_{\beta\beta}(s) \right),
\]
where \( \eta_n \to 0 \). Substituting (5.38) into (5.29), the theorem follows from Theorem 5.4. \( \Box \)

Much of the work on approximation of solutions of stochastic differential equations has been concerned with linear interpolations of the integrator. The next result shows that Theorem 5.10 applies to these approximations.

5.12 PROPOSITION. For each \( n \), let \( V_n \) be an \( (\mathcal{F}^n_t) \)-semimartingale and \( \{\tau^*_k\} \) be a sequence of \( (\mathcal{F}^n_t) \)-stopping times with \( \tau^*_0 = 0 \) and \( \lim_{k \to \infty} \tau^*_k = \infty \). Suppose that \( \lim_{n \to \infty} \sup_k \tau^*_{k+1} - \tau^*_k = 0 \). Define the linear interpolation
\[
(5.39) \quad \hat{V}_n(t) = \frac{\tau^*_k - t}{\tau^*_{k+1} - \tau^*_k} V_n(\tau^*_k) + \frac{t - \tau^*_k}{\tau^*_{k+1} - \tau^*_k} V_n(\tau^*_{k+1}), \quad \tau^*_k \leq t < \tau^*_{k+1},
\]
and define
\[
(5.40) \quad Y_n(t) = V_n(\tau^*_k) = \hat{V}_n(\tau^*_k), \quad \tau^*_k \leq t < \tau^*_{k+1},
\]
and
\[
(5.41) \quad Z_n(t) = \hat{V}_n(t) - Y_n(t)
\]
\[= \frac{\tau^*_k - t}{\tau^*_{k+1} - \tau^*_k} (V_n(\tau^*_k) - V_n(\tau^*_{k+1})), \quad \tau^*_k \leq t < \tau^*_{k+1},
\]
Define \( H_n \) and \( K_n \) as in Theorem 5.10. Suppose that \( V_n \to Y \), where \( Y \) is
continuous and that \( \{V_n\} \) satisfies C2.2(i) for some \( \delta \in (0, \infty) \). (In particular, this last statement holds if \( V_n = Y \) for all \( n \).) Then \( Y_n \Rightarrow Y \) and \( Z_n \Rightarrow 0 \), \( \{Y_n\} \) and \( \{H_n\} \) satisfy C2.2(i) for some \( \delta \in (0, \infty) \), and

\[
(5.42) \quad H_n \Rightarrow -\frac{1}{2} \langle [Y^\gamma, Y^\gamma] \rangle,
\]

\[
(5.43) \quad K_n \Rightarrow -\langle [Y^\beta, Y^\gamma] \rangle.
\]

**Proof.** Let \( \tau_n(t) = \min \{ \tau_n^a : \tau_n^a > t \} \). Then \( \tau_n(t) \) is an \( \{\mathcal{F}_t^n\} \)-stopping time for each \( t \) and \( V_n^\beta, Y_n \) and \( Z_n \) are adapted to the filtration \( \{\mathcal{F}_t^n\} \) given by \( \mathcal{F}_t^n = \mathcal{F}_{\tau_n(t)}^n \). The proof that \( \{Y_n\} \) satisfies C2.2(i) is essentially the same as the proof that \( \{Y_n^\beta\} \) satisfies C2.2(i) in Example 3.7. The fact that \( Y_n \Rightarrow Y \) follows from the convergence of \( V_n \) and Proposition 3.6.5 of Ethier and Kurtz (1986). The convergence of \( Z_n \) follows from the continuity of \( Y \) and the continuous mapping theorem.

Note that

\[
H_n^\beta(t) = -\frac{1}{2} \sum_{\tau_n+1 \leq t} \left( \tau_n^a + t - \tau_n^a \right) \frac{(\tau_n^a - t)^2}{(\tau_n^a + t)^2} \times (V_n^\beta(\tau_n^a) - V_n^\beta(\tau_n^a))(V_n^\gamma(\tau_n^a) - V_n^\gamma(\tau_n^a))
\]

\[
= -\frac{1}{2} \sum_{\tau_n+1 \leq t} (V_n^\beta(\tau_n^a) - V_n^\beta(\tau_n^a))(V_n^\gamma(\tau_n^a) - V_n^\gamma(\tau_n^a))
\]

\[
= -\frac{1}{2} \langle Y_n^\beta, Y_n^\gamma \rangle_t
\]

\[
= -\frac{1}{2} \langle Y_n^\beta(t) Y_n^\gamma(t) - \int_0^t Y_n^\beta(s) \, dY_n^\gamma(s) - \int_0^t Y_n^\gamma(s) \, dY_n^\beta(s) \rangle
\]

and (5.42) follows by Theorem 2.2. A similar calculation gives (5.43). The fact that \( \{H_n\} \) satisfies C2.2(i) follows from the monotonicity and convergence of \( H_n^\beta \) and the fact that the total variation of \( H_n^\beta \) can be estimated in terms of the total variation of \( H_n^\beta \) and \( H_n^\gamma \). \( \square \)

6. Technical results.

**Uniform approximation by step functions.** Let \( E \) be a metric space with metric \( r \). Let \( \{\theta_n\} \) be a sequence of independent random variables, uniformly distributed on the interval \( [\frac{1}{2}, 1] \). Fix \( \varepsilon > 0 \) and for \( z \in D \), \( z(0) = 0 \) and \( \tau_+(z) = \inf \{ t > 0 : r(z(t), z(\tau_+(z))) \vee r(z(t - ), z(\tau_+(z))) > \varepsilon \theta_0 \} \) and set \( \gamma(x) = z(\tau_+(z)) \). Finally, define \( I(x) \) by \( I(x)(t) = \gamma(x) \) for \( \tau_+(z) \leq t < \tau_+(z) \). Note that \( r(z(t), I(x)(t)) \leq \varepsilon \) for all \( t \). Let \( U_1 = \{ u : u = r(z(t), z(0)) \) or \( r(z(t - ), z(0)) \) for some \( t \) such that \( z(t) \neq z(t - ) \) and defining \( m(t) = \sup_{s < t} r(z(s), z(0)) \), let \( U_2 = \{ m(t) : m \) is not strictly increasing at \( t \} \). \( U_1 \) and \( U_2 \) are countable, so with probability 1, \( \varepsilon \theta_0 \notin U_1 \cup U_2 \). Let \( z_n \to z \) and assume that \( \varepsilon \theta_0 \notin U_1 \cup U_2 \). Either \( m \) is strictly increasing at \( \tau_+(z) \) or
r(z(τ_l(z) − ), z(0)) < εθ_0 < r(z(τ_l(z)), z(0)) and it follows that τ_l(z_n) → τ_l(z).
Either z is continuous at τ_l(z) or r(z(τ_l(z) − ), z(0)) < εθ_0 < r(z(τ_l(z)), z(0)), and it follows that γ_l(z_n) → γ_l(z). In general, if z_n → z in the Skorohod topology, t_n → t and z_n(t_n) → z(t), then z_n(t_n + ·) → z(t + ·) in the Skorohod topology. Consequently, z_n → z implies z_n(τ_l(z_n) + ·) → z(τ_l(z) + ·) a.s.
An induction argument then shows that z_n → z implies τ_k(z_n) → τ_k(z) and γ_k(z_n) → γ_k(z) a.s. for all k. With these observations, we can prove the following lemma.

6.1 Lemma. Let I_n be defined as before. If z_n → z in the Skorohod topology on D_E[0,∞), then (z_n, I_n(z_n)) → (z, I(z)) a.s. in the Skorohod topology on D_E×E[0,∞).

To carry out the proof, we need the following [see Proposition 3.6.5 of Ethier and Kurtz (1986); note that the third condition in that proposition is implied by the other two].

6.2 Lemma. For an arbitrary metric space (E', r'), u_n → u in the Skorohod topology on D_E[0,∞) if and only if the following conditions hold.

C6.2(i) If t_n → t, then lim_n→∞ r'(u_n(t_n), u(t)) ∧ r'(u_n(t_n), u(t − )) = 0.

C6.2(ii) If s_n ≥ t_n, s_n, t_n → t and u_n(t_n) → u(t), then u_n(s_n) → u(t).

Proof of Lemma 6.1. Suppose z_n → z in D_E[0,∞) and t_n → t. If τ_k(z) < t < τ_{k+1}(z), then I_n(z) is continuous at t, I_n(z_n)(t_n) → γ_k(z) = I_n(z)(t) and C6.2(i) and (ii) follow for (z_n, I_n(z_n)) by the analogous conditions for (z_n). If t = τ_k(z), we can assume that either z is continuous at τ_k(z) or r(z(τ_k(z) − ), z(τ_{k-1}(z))) < εθ_{k-1} < r(z(τ_k(z)), z(τ_{k-1}(z))). The convergence of τ_{k-1}(z_n), τ_k(z_n), γ_k(z_n) and γ_k(z) implies C6.2(i) and (ii) for (z_n, I_n(z_n)), and if z is continuous at τ_k(z), C6.2(i) and (ii) follow for (z_n, I_n(z_n)). If r(z(τ_k(z) − ), z(τ_{k-1}(z))) < εθ_{k-1} < r(z(τ_k(z)), z(τ_{k-1}(z))), then, with probability 1, for n sufficiently large the same inequality holds with z replaced by z_n. Consequently, if t_n ≥ τ_k(z_n) and t_n → τ_k(z), then z_n(t_n) and I_n(z_n)(t_n) both converge to γ_k(z), and if t_n < τ_k(z_n) and t_n → t, then z_n(t_n) converges to z(τ_k(z) − ) and I_n(z_n)(t_n) converges to γ_{k-1}(z) = I_n(z(τ_k(z) − )). C6.2(i) and (ii) follow for (z_n, I_n(z_n)). □

Uniform tightness. Jakubowski, Mémin and Pages (1989) and Slomiński (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 following are essentially contained in Lemma 3.1 of Jakubowski, Mémin and Pages (1989).

Let H_n denote the collection of cadlag (F^1_n)-adapted, R-valued processes satisfying |H_n(t)| ≤ 1 for all t ≥ 0. Then (Y_n) is uniformly tight if for each
$t > 0$,

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

is stochastically bounded.

Assume that $(Y_n)$ is uniformly tight. Let $\mathcal{T}_n$ denote the collection of $(\mathcal{T}_n)$-stopping times. For $\tau \in \mathcal{T}_n$ and $\varepsilon > 0$, 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 $\tau = \inf\{s : |Y_n(s)| \geq \varepsilon\}$, it follows that $(\sup_{s \leq t}|Y_n(s)| : n = 1, 2, \ldots)$ is stochastically bounded. Recalling that

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

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 $(Y^\delta_n)$ for each $0 < \delta < \infty$. Fix $0 < \delta < \infty$ and let $Y^\delta_n = M^\delta_n + A^\delta_n$ be the canonical decomposition of $Y^\delta_n$ [Protter (1990), Section III.5]. Then the discontinuities of $M^\delta_n$ and $A^\delta_n$ are bounded by $2\delta$ and $E[[Y^\delta_n]]_\tau = E[[M^\delta_n]]_\tau + E[[A^\delta_n]]_\tau$ for any stopping time $\tau$ (with the possibility of $\infty = \infty$) [Protter (1990), Section IV.2]. Let $\gamma^\delta_n = \inf\{s : |Y^\delta_n(s)| \geq \varepsilon\}$. Fix $t$ and for $k = 1, 2, \ldots$, let $(t^k_n)$ be a partition of $[0, t]$ with

$$\lim_{k \to \infty} \max_i (t^k_{i+1} - t^k_i) = 0.$$  

Define

$$H^k_n = \sum_i \text{sign} \left( E \left[ A^\delta_n(t^k_{i+1} \wedge \gamma^\delta_n) - A^\delta_n(t^k_i \wedge \gamma^\delta_n) | \mathcal{F}^n_{t^k_i} \right] \right) \chi_{(t^k_i \wedge \gamma^\delta_n, t^k_{i+1} \wedge \gamma^\delta_n)}.$$  

The first term on the right of

$$\int_0^u H^k_n(s - ) \, dY^\delta_n(s) = \int_0^u H^k_n(s - ) \, dM^\delta_n(s) + \int_0^u H^k_n(s - ) \, dA^\delta_n(s)$$  

$$= U^k_n(u) + V^k_n(u)$$

satisfies

$$E \left[ \sup_{s \leq t} U^k_n(s)^2 \right] \leq 4 E \left[ M^\delta_n(t \wedge \gamma^\delta_n)^2 \right] \leq 4(c + (2\delta)^2),$$

so $(U^k_n(t) : k, n = 1, 2, \ldots)$ is stochastically bounded which, by the stochastic boundedness of (6.1) (with $Y_n$ replaced by $Y^\delta_n$), implies the stochastic boundedness of $(V^k_n(t) : k = 1, 2, \ldots)$. But the predictability of $A^\delta_n$ implies

$$T_{t \wedge \gamma^\delta_n} (A^\delta_n)$$

$$= \lim_{k \to \infty} \sum \text{sign} \left( E \left[ A^\delta_n(t^k_{i+1} \wedge \gamma^\delta_n) - A^\delta_n(t^k_i \wedge \gamma^\delta_n) | \mathcal{F}^n_{t^k_i} \right] \right) \times (A^\delta_n(t^k_{i+1} \wedge \gamma^\delta_n) - A^\delta_n(t^k_i \wedge \gamma^\delta_n))$$

$$= \lim_{k \to \infty} V^k_n(t)$$

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

If there exists a $\delta$ for which $(J_{\delta}(Y_n))$ is stochastically bounded and C2.2(i) holds, then $(Y_n^s)$ satisfies C4.1 and Lemma 4.1 implies $(Y_n)$ is uniformly tight.

Acknowledgments. 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.

REFERENCES


