*The Annals of Probability*
1991, Vol. 19, No. 3, 1010–1034

# RANDOM TIME CHANGES AND CONVERGENCE IN DISTRIBUTION UNDER THE MEYER–ZHENG CONDITIONS[1]

BY THOMAS G. KURTZ

*University of Wisconsin–Madison*

An analog of conditions of Meyer and Zheng for the relative compactness (in the sense of convergence in distribution) of a sequence of stochastic processes is formulated for general separable metric spaces and the corresponding notion of convergence is characterized in terms of the convergence in the Skorohod topology of time changes of the original processes. In addition, convergence in distribution under the topology of convergence in measure is discussed and results of Jacod, Mémin and Métivier on convergence under the Skorohod topology are extended.

**1. Introduction.** Meyer and Zheng (1984) give conditions under which a sequence of stochastic processes $\{X_n\}$ is relatively compact (in the sense of convergence in distribution) when the space of sample paths is topologized by convergence in measure. It is clear from their paper that these conditions, involving boundedness of the conditional variations of the processes, imply much greater uniform regularity of the sample paths of the processes than is implied by convergence in measure. In this paper we extend and refine the results of Meyer–Zheng in a number of ways. We formulate an analog of the Meyer–Zheng conditions for a general separable metric space; we capture the greater regularity of the convergence under these conditions by showing that they imply the existence of a sequence of random time transformations $\{\gamma_n\}$ such that $\{X_n \circ \gamma_n\}$ is relatively compact under the Skorohod topology; and we formulate simpler conditions, immediately implied by the Meyer–Zheng conditions, that imply relative compactness under the topology of convergence in measure.

For a cadlag, real-valued process $X$ adapted to a filtration $\{\mathscr{F}_t\}$ (we will assume that all filtrations are right-continuous and complete) define the conditional variation $V_t(X)$ on the interval $[0, t]$ by

$$V_t(X) = \sup E\left[\sum_i |E[X(t_{i+1}) - X(t_i)|\mathscr{F}_{t_i}]|\right], \tag{1.1}$$

where the supremum is over all partitions of the interval $[0, t]$. Considering a sequence of cadlag processes $\{X_n\}$, with $X_n$ adapted to a filtration $\{\mathscr{F}_t^n\}$ and

Received May 1989; revised May 1990.
[1]Research supported by the National Science Foundation.
*AMS* 1980 *subject classifications*. Primary 60F17; secondary 60G99.
*Key words and phrases*. Weak convergence, tightness, time change, Skorohod topology, conditional variation.

$V_t(X_n)$ defined relative to this filtration, Meyer and Zheng require that

$$\sup_n \Big( V_t(X_n) + \sup_{s \le t} E\big[|X_n(s)|\big] \Big) < \infty, \qquad t > 0. \tag{1.2}$$

Under this condition, each $X_n$ is a local quasimartingale. Consequently, for each $n$ there exists a predictable, finite variation process $B_n$ such that $X_n - B_n$ is a local martingale and $E[T_t(B_n)] \le V_t(X_n)$. [$T_t(b)$ will denote the total variation of a function $b$ on $[0, t]$.] The quasimartingale property also implies

$$P\Big\{ \sup_{s \le t} |X_n(s)| \ge c \Big\} \le c^{-1}\big( V_t(X_n) + E\big[|X_n(t)|\big] \big). \tag{1.3}$$

Let $\varphi$ be convex, symmetric, $C^2$ and satisfy $\varphi(0) = 0$, $\varphi''(0) = 1$, $\varphi''$ nonincreasing on $[0, \infty)$ and $\varphi''(1) = 0$. Then there exist increasing processes $A_n$ such that for each $b > 0$ and $\tau_b^n = \inf\{t\colon |X_n(t)| \vee |X_n(t-)| \ge b\}$,

$$E\big[A_n(\tau_b^n \wedge T)\big] \le E\big[|X_n(t)|\big] + 2(1 + b)V_T(X_n) + b^2 + 2b,$$

and for each $a \ge 0$,

$$\varphi(X_n(a + t) - X_n(a)) - (A_n(a + t) - A_n(a)) \tag{1.4}$$

is a local supermartingale (Theorem 5.7). It is the analog of this observation that we use to extend the Meyer–Zheng conditions to a general separable metric space $(E, r)$. Note also that $A_n$ generalizes the notion of a dominating process employed by Jacod, Mémin and Métivier (1983), and the next results extend their work as well.

$D_E[0, \infty)$ will denote the set of cadlag, $E$-valued functions on $[0, \infty)$ with the Skorohod topology. $T[0, \infty)$ will denote the set of right-continuous, nondecreasing functions $a$ on $[0, \infty)$ with $a(0) = 0$ and $\lim_{t \to \infty} a(t) = \infty$, $T_s[0, \infty)$ will denote the subset of strictly increasing functions and $T_c[0, \infty)$ will denote the subset of nondecreasing, continuous functions. For $a \in T[0, \infty)$, we define $a^{-1}(t) = \inf\{u\colon a(u) > t\}$.

We will prove the following theorem.

1.1 THEOREM. *Let $\{X_n\}$ be a sequence of processes with sample paths in $D_E[0, \infty)$. Let $\varphi_n\colon E \times E \to [0, \infty)$, $n = 1, 2, \dots$, be continuous and have the property that for each compact $K \subset E$, there exists a nondecreasing $\alpha_K\colon [0, \infty) \to [0, \infty)$ with $\alpha_K(0) = 0$, $\alpha_K(r) > 0$ for $r > 0$ and $\inf_n \varphi_n(x, y) \ge \alpha_K(r(x, y) \wedge 1)$ for $x, y \in K$. Suppose that the following conditions hold:*

C1.1(i) (*Compact containment.*) *For each $\varepsilon > 0$ and $T > 0$, there exists a compact $K \subset E$ such that*

$$\liminf_{n \to \infty} P\{X_n(t) \in K,\ t \le T\} \ge 1 - \varepsilon. \tag{1.5}$$

C1.1(ii) *For each $n$, there exists an increasing process $A_n$ such that for each $t_0 \geq 0$,*

$$\varphi_n(X_n(t_0 + t), X_n(t_0)) - (A_n(t + t_0) - A_n(t_0)) \tag{1.6}$$

*is a local $\{\mathscr{F}^n_{t_0+t}\}$-supermartingale.*

C1.1(iii) *For each $\alpha > 0$, there exist stopping times $\{\tau_n^\alpha\}$ such that $P\{\tau_n^\alpha \leq \alpha\} \leq 1/\alpha$ and for each $t \geq 0$, $\sup_n E[A_n(t \wedge \tau_n^\alpha)] < \infty$.*

*Then the following hold:*

(a) *There exist processes $\{\gamma_n\}$ in $T_c[0, \infty)$, with $\{\gamma_n^{-1}(t)\}$ stochastically bounded for each $t$, and $E$-valued, cadlag processes $\{Y_n\}$ such that $X_n(t) = Y_n(\gamma_n^{-1}(t))$ and $\{(Y_n, \gamma_n)\}$ is relatively compact in the Skorohod topology on $D_{E\times\mathbb{R}}[0, \infty)$.*

(b) *Suppose the sequence $\{(Y_n, \gamma_n)\}$ in part* (a) *converges in distribution to $(Y, \gamma)$ in the Skorohod topology. Then (by the Skorohod representation theorem) there exists a probability space on which are defined processes $\{(\hat{Y}_n, \hat{\gamma}_n)\}$ converging almost surely to a process $(\hat{Y}, \hat{\gamma})$ in the Skorohod topology on $D_{E\times\mathbb{R}}[0, \infty)$ such that $(\hat{Y}_n, \hat{\gamma}_n)$ has the same distribution as $(Y_n, \gamma_n)$ (and hence $\hat{X}_n \equiv \hat{Y}_n \circ \hat{\gamma}_n^{-1}$ has the same distribution as $X_n$) and with probability* 1, *$\hat{X}_n(t) \to \hat{X}(t) \equiv \hat{Y} \circ \hat{\gamma}^{-1}(t)$ for all but countably many $t$.*

(c) *If $\gamma$ in part* (b) *is strictly increasing, then $X_n \Rightarrow Y \circ \gamma^{-1}$ in the Skorohod topology.*

1.2 Remarks. (a) The conditions of the theorem can be weakened to the assertion that for each $T > 0$, there exist $\varphi_n$, $\alpha_K$ and $A_n$ with C1.1(ii) and C1.1(iii) holding for $t_0 + t \leq T$ and $t \leq T$, respectively. C1.1(iii) is satisfied if for each $t \geq 0$ and $c > 0$, $\{A_n(t)\}$ is stochastically bounded and $\sup_n E[A_n(t \wedge \tau_n^c)] < \infty$, where $\tau_n^c = \inf\{s: A_n(s) \geq c\}$.

(b) It is tempting to define a notion of convergence which states that $x_n \to x$ if there exist $y_n, y \in D_E[0, \infty)$ and $\gamma_n, \gamma \in T_c[0, \infty)$ such that $x_n = y_n \circ \gamma_n^{-1}$, $x = y \circ \gamma^{-1}$ and $(y_n, \gamma_n) \to (y, \gamma)$ in the Skorohod topology. Unfortunately, this notion of convergence does not correspond to a metric.

(c) Without loss of generality, we can assume $A_n(t + h) - A_n(t) \geq h$ (otherwise replace $A_n(t)$ by $A_n(t) + t$). Let $A_n^p$ be the dual predictable projection of $A_n$. Then we will show that we can take $\gamma_n$ to be the inverse of $A_n^p$ and $Y_n(u) = \lim_{v \to u+} X_n(\gamma_n(v) - )$. The conclusions of the theorem also hold with this $Y_n$ replaced by $Y_n^+$ defined by $Y_n^+(u) = X_n(\gamma_n(u))$.

1.3 Corollary. *Let $\{X_n\}$ be a sequence of processes with sample paths in $D_{\mathbb{R}^k}[0, \infty)$. Assume that the following condition holds.*

C1.3 *For each $\alpha > 0$, there exist stopping times $\{\tau_n^\alpha\}$ with $P\{\tau_n^\alpha \leq \alpha\} \leq 1/\alpha$ such that for each $t \geq 0$, $\sup_n E[|X_n^{\tau_n^\alpha}(t)|] < \infty$ and $\sup_n V_t(X_n^{\tau_n^\alpha}) < \infty$.*

*Then the conclusions of Theorem* 1.1 *hold.*

1.4 COROLLARY. *Let* $\{f_i\} \subset \overline{C}(E)$ *separate points in* $E$. *Suppose that the compact containment condition*, C1.1(i), *holds and that for each* $i$ *and* $t > 0$,

$$\sup_n V_t(f_i \circ X_n) < \infty. \tag{1.7}$$

*Then the conclusions of Theorem* 1.1 *hold.*

Let $M_E[0,\infty)$ be the space of (equivalence classes of) $E$-valued, Borel measurable functions topologized by convergence in measure. Clearly, by part (b) of Theorem 1.1, the Meyer–Zheng conditions imply that the sequence $\{X_n\}$ is relatively compact considered as a sequence of $M_E[0,\infty)$-valued random variables. This conclusion, however, follows under a much weaker set of conditions.

1.5 THEOREM. *Let* $\{X_n\}$ *be a sequence of measurable, E-valued processes. Let* $\varphi_n$: $E \times E \to [0,\infty)$, $n = 1, 2, \dots$, *have the property that for each compact* $K \subset E$ *there exists a nondecreasing* $\alpha_K$: $[0,\infty) \to [0,\infty)$ *with* $\alpha_K(0) = 0$, $\alpha_K(r) > 0$ *for* $r > 0$ *and* $\inf_n \varphi_n(x,y) \geq \alpha_K(r(x,y) \wedge 1)$ *for* $x, y \in K$. *Suppose that the following conditions hold*:

C1.5(i) *For each* $\varepsilon > 0$ *and* $T > 0$, *there exists a compact* $K \subset E$ *such that*

$$\limsup_{n\to\infty} \int_0^T P\{X_n(t) \in K^c\}\, dt \leq \varepsilon. \tag{1.8}$$

C1.5(ii) *For each* $n$, *there is an increasing function* $a_n$ *such that for each* $t_0 \geq 0$,

$$E[\varphi_n(X_n(t_0 + t), X_n(t_0))] \leq a_n(t + t_0) - a_n(t_0). \tag{1.9}$$

C1.5(iii) *For each* $t \geq 0$, $\sup_n a_n(t) < \infty$.

*Then* $\{X_n\}$ *is relatively compact in* $M_E[0,\infty)$.

In Section 2, we examine time transformations and Skorohod convergence and prove Theorem 1.1. In Section 3, we consider conditions more general than those of part (c) of Theorem 1.1 under which the convergence is actually in the Skorohod topology. In particular, we extend results of Jacod, Mémin and Métivier (1983) to the present setting. Section 4 is devoted to the study of convergence in $M_E[0,\infty)$ and the proof of Theorem 1.5. Much of this material appears in some form elsewhere [e.g., Dellacherie and Meyer (1978)]. We collect it here as a convenient reference. Section 5 is an appendix collecting proofs of the properties of quasimartingales mentioned earlier.

**2. Convergence under random time changes.** We recall certain facts about convergence in the Skorohod topology.

2.1 LEMMA. *Let $\{x_n\} \subset D_E[0,\infty)$. Then $x_n \to x$ in the Skorohod topology if and only if the following conditions hold*:

C2.1(i) *If $t_n \to t$, then* $\lim_{n\to\infty} r(x_n(t_n), x(t)) \wedge r(x_n(t_n), x(t-)) = 0$.
C2.1(ii) *If $s_n \ge t_n$, $s_n, t_n \to t$ and $x_n(t_n) \to x(t)$, then $x_n(s_n) \to x(t)$.*

PROOF. See Ethier and Kurtz (1986), Proposition 3.6.5. □

2.2 LEMMA. *Let $\{x_n\} \subset D_{E_1}[0,\infty)$ and $\{y_n\} \subset D_{E_2}[0,\infty)$. Suppose $x_n \to x$ in $D_{E_1}[0,\infty)$ and $y_n \to y$ in $D_{E_2}[0,\infty)$. Then $(x_n, y_n) \to (x,y)$ in $D_{E_1\times E_2}[0,\infty)$ if and only if for all sequences $\{s_n\}$ and $\{t_n\}$ converging to $t \in [0,\infty)$,*

$$(2.1)\qquad \lim_{n\to\infty} r_1(x_n(s_n), x_n(s_n-)) \wedge r_2(y_n(t_n), y_n(t_n-)) \wedge \chi_{\{t_n \ne s_n\}} = 0.$$

PROOF. The necessity of (2.1) follows from the necessity of C2.1(i).
Suppose $t_n \to t$. If either $x$ or $y$ are continuous at $t$, then C2.1(i) and C2.1(ii) are easily verified for the sequence $\{(x_n, y_n)\}$. Suppose both $x$ and $y$ have discontinuities at $t$. Then there exist $\{s_n\}$ and $\{t_n\}$ converging to $t$ such that $r_1(x_n(s_n), x_n(s_n-)) \to r_1(x(t), x(t-))$ and $r_2(y_n(t_n), y_n(t_n-)) \to r_2(y(t), y(t-))$. Consequently, by (2.1), $t_n \ne s_n$ for only finitely many $n$ and the conditions of Lemma 2.1 follow. □

Recall that $T_s[0,\infty)$ denotes the space of right-continuous, strictly increasing functions $a$ on $[0,\infty)$ with $a(0) = 0$ and $a(t) \to \infty$ and note that $a \in T_s[0,\infty)$ implies that $a^{-1} \in T_c[0,\infty)$ and vice versa.

2.3 LEMMA. *Let $\{x_n\}, \{y_n\} \subset D_E[0,\infty)$, $\{a_n\} \subset T_s[0,\infty)$ and $x_n(t) = y_n(a_n(t))$ for $t \ge 0$. Suppose there exist $y \in D_E[0,\infty)$ and $a \in T_s[0,\infty)$ such that $y_n \to y$ in the Skorohod topology and $a_n(t) \to a(t)$ at each continuity point of $a$. Define $x(t) = y(a(t))$. Then*

(a) *$x_n(t) \to x(t)$ for all but countably many $t$.*
(b) *If on any interval $[u,v]$ on which $a^{-1}$ is constant, $y$ is constant except for at most one jump, then $x_n \to x$ in the Skorohod topology. In particular, the conclusion holds if $a^{-1}$ is strictly increasing, that is, if $a$ is continuous.* (*Note that since $a$ is right-continuous and $a(0) = 0$, $a^{-1}$ is not constant on any interval $[0,v]$. If we drop the assumption that $a(0) = 0$, then we must add the requirement that $y$ be constant in the interval $[0, a(0)]$.*)
(c) *If $y_n(u) = x_n(a_n^{-1}(u))$ or $y_n(u) = \lim_{v\to u+} x_n(a_n^{-1}(v)-)$, then the conditions in* (b) *are necessary as well as sufficient for $x_n \to x$ in the Skorohod topology.* [*Note that since $a_n$ is strictly increasing, $a_n^{-1}(a_n(t)) = t$ and hence in either case $y_n(a_n(t)) = x_n(t)$.*]

(d) *If* $a_n \to a$ *in the Skorohod topology and* $y_n$ *is given by one of the formulas in* (c), *then* $x_n \to x$ *in the Skorohod topology if and only if for all sequences* $\{s_n\}$ *and* $\{t_n\}$ *converging to* $t \in [0, \infty)$,

$$\text{(2.2)} \qquad \lim_{n \to \infty} r(x_n(s_n), x_n(s_n -)) \wedge r(x_n(t_n), x_n(t_n -)) \wedge \chi_{\{t_n \neq s_n\}} = 0.$$

2.4 REMARK. (a) For $a \in T_s[0, \infty)$ and $x \in D_E[0, \infty)$, define $y(u) = \lim_{v \to u+} x(a^{-1}(v) - )$. Then $y(u) = x(t)$ if $u = a(t)$, and $y(u) = x(t - )$ if $a(t - ) \le u < a(t)$.

(b) The limit in (2.2) essentially says that discontinuities cannot coalesce.

PROOF. If $t_n \to t$, any limit point of $\{a_n(t_n)\}$ must be in the interval $[a(t - ), a(t)]$. Consequently, any limit point of $\{x_n(t_n)\} = \{y_n(a_n(t_n))\}$ must be in the set $\Gamma_t = \{y(u), y(u - ): u \in [a(t - ), a(t)]\}$. Since there are at most countably many values of $t$ for which $a(t) \neq a(t - )$ and at most countably many values of $t$ for which $y(a(t)) \neq y(a(t) - )$ (recall $a$ is strictly increasing), part (a) follows.

Under the hypotheses of part (b), $\Gamma_t = \{y(a(t - ) - ), y(a(t))\} = \{x(t - ), x(t)\}$ and if $x_n(t_n) = y_n(a_n(t_n)) \to y(a(t)) = x(t)$, $s_n \ge t_n$ and $s_n \to t$, then by Lemma 2.1, $x_n(s_n) = y_n(a_n(s_n)) \to y(a(t)) = x(t)$ and (b) follows by the same lemma. Part (c) also follows by Lemma 2.1.

To prove part (d) we verify the hypotheses of part (b). If $a$ has a discontinuity at $t$, the conditions on $\{a_n\}$ and $\{y_n\}$ ensure that $y$ is constant on $(a(t - ), a(t))$, so to verify the hypotheses of part (b) we need only show that $y$ has a discontinuity at no more than one of the endpoints of the interval $[a(t - ), a(t)]$. Suppose $y$ has a discontinuity at both endpoints. Then there exist $u_n \to a(t - )$ and $v_n \to a(t)$ such that $r(y_n(u_n), y_n(u_n - )) \to r(y(a(t - )), y(a(t - ) - ))$ and $r(y_n(v_n), y_n(v_n - )) \to r(y(a(t)), y(a(t) - ))$. If $y_n(u) = x_n(a_n^{-1}(u))$, then taking $s_n = a_n^{-1}(v_n)$ and $t_n = a_n^{-1}(u_n)$, $y_n(v_n) = x_n(s_n)$, $y_n(u_n) = x_n(t_n)$ and by the continuity of $a_n^{-1}$ we must have $y_n(v_n - ) = x_n(s_n - )$ and $y_n(u_n - ) = x_n(t_n - )$ for $n$ sufficiently large. Since $s_n \to t$ and $t_n \to t$, we derive a contradiction of (2.2). If $y_n(u) = \lim_{v \to u+} x_n(a_n^{-1}(v) - )$, first observe that if $y_n(u) = x_n(s - )$ for some $s$, then $y_n$ is continuous at $u$. Consequently, we can obtain a contradiction of (2.2) by the same argument as before. □

2.5 LEMMA. *For* $a \in T[0, \infty)$, *let* $b(u) = \lim_{v \to u+} a(a^{-1}(v) - )$. *Then* $b(u) \le u$ *and if* $b(u) > b(u - h)$, *then* $b(u) > u - h$.

PROOF. By Remark 2.4(a), if $u = a(t)$ for some $t$, then $b(u) = a(t) = u$. Otherwise, $a(t - ) \le u < a(t)$ and $b(u) = a(t - ) \le u$. If $b(u) > b(u - h)$, then there must be a $v \in (u - h, u]$ such that $b(v) = v \le b(u)$. □

We will need the following result from Kurtz (1975) [see Ethier and Kurtz (1986), Theorem 3.8.6]. Let $q(x, y) = 1 \wedge r(x, y)$.

2.6 LEMMA. *Let $(E, r)$ be complete and separable and let $\{X_n\}$ be a sequence of cadlag, E-valued processes, $X_n$ adapted to $\{\mathscr{F}_t^n\}$. Then $\{X_n\}$ is relatively compact in the Skorohod topology if and only if the compact containment condition holds* [C1.1(i)],

$$\lim_{\delta \to 0} \limsup_{n \to \infty} E\left[q(X_n(\delta), X_n(0))\right] = 0 \tag{2.3}$$

*and for each $\delta > 0$, $T > 0$ and n there exist random variables $\gamma_n^T(\delta)$ such that*

$$\begin{aligned} &E\left[q(X_n(t+u), X_n(t)) | \mathscr{F}_t^n\right] q(X_n(t), X_n(t-v)) \\ &\qquad \le E\left[\gamma_n^T(\delta) | \mathscr{F}_t^n\right], \end{aligned} \tag{2.4}$$

*$0 \le t \le T$, $0 \le u \le \delta$, $0 \le v \le \delta \wedge t$, and*

$$\lim_{\delta \to 0} \limsup_{n \to \infty} E\left[\gamma_n^T(\delta)\right] = 0. \tag{2.5}$$

PROOF OF THEOREM 1.1. Let $A_n^p$ denote the $\{\mathscr{F}_t^n\}$-dual predictable projection of $A_n$. Then the assumptions of the theorem hold with $A_n$ replaced by $A_n^p$. Without loss of generality, we can assume that $A_n^p(t+h) - A_n^p(t) \ge h$ [otherwise replace $A_n^p(t)$ by $A_n^p(t) + t$]. Let $\gamma_n$ denote the inverse of $A_n^p$ [note that $\gamma_n(t+h) - \gamma_n(t) \le h$] and define $Y_n(t) = \lim_{s \to t+} X_n(\gamma_n(s)-)$ and $B_n(t) = \lim_{s \to t+} A_n^p(\gamma_n(s)-)$. Since $\gamma_n$ is continuous, for each $t$, $\gamma_n(t)$ is an $\{\mathscr{F}_t^n\}$-predictable stopping time.

The fact that (1.6) is a local $\{\mathscr{F}_{t_0+t}^n\}$-supermartingale for each $t_0$ implies that for bounded stopping times $\tau_1 \le \tau_2$,

$$E\left[\varphi_n(X_n(\tau_2), X_n(\tau_1)) | \mathscr{F}_{\tau_1}^n\right] \le E\left[A_n^p(\tau_2) - A_n^p(\tau_1) | \mathscr{F}_{\tau_1}^n\right] \tag{2.6}$$

(with the possibility of one or both sides being infinite) and the predictability of $\gamma_n(t)$ implies

$$\begin{aligned} &E\left[\varphi_n(X_n(\gamma_n(t+u)-), X_n(\gamma_n(t)-)) | \mathscr{F}_{\gamma_n(t)-}^n\right] \\ &\qquad \le E\left[A_n^p(\gamma_n(t+u)-) - A_n^p(\gamma_n(t)-) | \mathscr{F}_{\gamma_n(t)-}^n\right] \end{aligned} \tag{2.7}$$

and hence, letting $\mathscr{G}_t^n = \cap_{s>t} \mathscr{F}_{\gamma_n(s)-}^n$,

$$E\left[\varphi_n(Y_n(t+u), Y_n(t)) | \mathscr{G}_t^n\right] \le E\left[B_n(t+u) - B_n(t) | \mathscr{G}_t^n\right]. \tag{2.8}$$

Fix $K$ compact and define $\eta_n^K = \inf\{t \colon Y_n(t) \notin K\}$. Note also that $Y_n$ is constant on any interval on which $B_n$ is constant, that is, on intervals of the form $[A_n^p(t-), A_n^p(t))$. Consequently, for $\delta > 0$, $0 \le u \le \delta$ and $0 \le v \le t \wedge \delta$,

$$\begin{aligned} &E\left[\varphi_n(Y_n(t+u), Y_n(t)) | \mathscr{G}_t^n\right] q(Y_n(t), Y_n(t-v)) \\ &\qquad \le E\left[B_n(t+u) - B_n(t) | \mathscr{G}_t^n\right] \chi_{\{B_n(t) > B_n(t-v)\}} \\ &\qquad \le 2\delta, \end{aligned} \tag{2.9}$$

where the last inequality follows by Lemma 2.5.

Fix $T > 0$ and $\delta > 0$. For any $\varepsilon > 0$, $t \le T$, $0 \le u \le \delta$ and $0 \le v \le t \wedge \delta$,

$$\begin{aligned}
&E\left[q(Y_n(t+u), Y_n(t))|\mathscr{G}_t^n\right]q(Y_n(t), Y_n(t-v)) \\
&\quad \le E\left[\varepsilon + \chi_{\{\eta_n^K \le T\}} + \chi_{\{\eta_n^K > T\}}\alpha_K(\varepsilon)^{-1}\alpha_K(q(Y_n(t+u), Y_n(t)))|\mathscr{G}_t^n\right] \\
&\qquad \times q(Y_n(t), Y_n(t-v)) \\
&\quad \le E\left[\varepsilon + \chi_{\{\eta_n^K \le T\}} + 2\delta\alpha_K(\varepsilon)^{-1}|\mathscr{G}_t^n\right].
\end{aligned} \tag{2.10}$$

Since the compact containment condition for $\{X_n\}$ implies the condition for $\{Y_n\}$, defining $\gamma_n^T(\delta) = \varepsilon + \chi_{\{\eta_n^K \le T\}} + 2\delta\alpha_K(\varepsilon)^{-1}$, by an appropriate choice of $\varepsilon$ and $K$ depending on $\delta$, we see that (2.4) and (2.5) are satisfied for $\{Y_n\}$. A similar estimate gives (2.3) and we can conclude that $\{Y_n\}$ is relatively compact which completes the proof of part (a).

Part (b) follows by Lemma 2.3(a) and part (c) follows by Lemma 2.3(b). □

PROOF OF COROLLARY 1.3. The compact containment condition C1.1(i) follows by Lemma 5.3. (Note that the vector-valued case follows by treating individual components.) Letting $\varphi$ be as in Theorem 5.7, for $x \in \mathbb{R}^k$ define $\hat{\varphi}(x) = \sum_{i=1}^k \varphi(x_i)$. $A_n$ satisfying C1.1(ii) and C1.1(iii) is then constructed as in the proof of Theorem 5.7. □

PROOF OF COROLLARY 1.4. As in (1.4) (see Section 5), for each $n$ and $i$, there exists an increasing process $A_n^i$ such that

$$\left(f_i(X_n(t_0+t)) - f_i(X_n(t_0))\right)^2 - \left(A_n^i(t_0+t) - A_n^i(t_0)\right) \tag{2.11}$$

is an $\{\mathscr{F}_{t_0+t}\}$-supermartingale and $E[A_n^i(t)] \le 6\|f_i\|_\infty V_t(f_i \circ X_n) + \|f_i\|_\infty^2$. For $T > 0$, there exist $\{a_i\}$ such that

$$\sum a_i \|f_i\|_\infty^2 < \infty \qquad \text{and} \qquad \sum a_i \sup_n \|f_i\|_\infty V_T(f_i \circ X_n) < \infty.$$

Define $\varphi_n(x,y) = \varphi(x,y) = \sum a_i(f_i(x) - f_i(y))^2$, $A_n = \sum a_i A_n^i$ and $\alpha_K(u) = \inf\{\varphi(x,y)\colon r(x,y) \ge u,\ x,y \in K\}$. Then C1.1(ii) and C1.1(iii) hold for $t$ and $t_0 + t$ restricted to the interval $[0,T]$. An examination of the proof of Theorem 1.1 shows that this is sufficient to verify the conclusions of the theorem. □

## 3. Additional conditions for Skorohod convergence.

Let $\{X_n\}$ satisfy the conditions of Theorem 1.1 and let $A_n^p$, $Y_n$ and $\gamma_n$ be as in the proof of the theorem. We would like to know what additional conditions are necessary to ensure relative compactness of $\{X_n\}$ in the Skorohod topology. With reference to Lemma 2.3(c), we have the following.

3.1 PROPOSITION. *Under the conditions of Theorem* 1.1, *suppose*

$$\lim_{h\to 0} \limsup_{n\to\infty} E\left[1 \wedge A_n^p(h)\right] = 0. \tag{3.1}$$

*Then $\{X_n\}$ is relatively compact in the Skorohod topology if and only if for each*

*$T > 0$ and $c > 0$,*

$$\lim_{h\to 0}\limsup_{n\to\infty}\sup E\Big[\sum_i q(X_n(t_{i+1}), X_n(\sigma_i))q(X_n(\tau_i), X_n(t_i)) \times (c\wedge A_n^p(t_{i+1}) - c\wedge A_n^p(t_i))\Big] = 0, \tag{3.2}$$

*where the supremum is over all choices of $\{t_i\}$ satisfying $0 = t_0 < \cdots < t_{m-1} < T \le t_m$ and $t_{i+1} - t_i \le h$ and all choices of stopping times satisfying $t_i \le \tau_i \le \sigma_i \le t_{i+1}$.*

*Let $\gamma_n$ and $Y_n$ be defined as in Remark 1.2(c). Then along any subsequence on which $(Y_n, \gamma_n) \Rightarrow (Y, \gamma)$, the left side of (3.2) is bounded by*

$$E\Big[\sum_{t\le T} Q(t)(A^p(t) - A^p(t-))\Big], \tag{3.3}$$

*where $A^p = \gamma^{-1}$ and $Q(t) = \sup\{q(Y(u_1-), Y(u_2))q(Y(v_1), Y(v_2))\colon A^p(t-) \le u_1 \le u_2 \le v_1 < v_2 \le A^p(t)\}$.*

3.2 REMARKS. (a) It is sufficient to show that the limit in (3.2) holds for each $\varepsilon > 0$, with the supremum replaced by the maximum over the two partitions given by taking $t_i = ih$ or $t_i = h/2 + (i-1)h$ and over $\tau_i = t_{i+1} \wedge \inf\{s > t_i\colon q(X_n(s), X_n(t_i)) \ge \varepsilon\}$ and $\sigma_i = \tau_i$ or $\sigma_i = \eta_i = t_{i+1} \wedge \inf\{s > \tau_i\colon q(X_n(s), X_n(\tau_i)) \ge \varepsilon\}$.

(b) By Lemma 2.3(c), $\{X_n\}$ relatively compact implies (3.3) is zero.

PROOF. The necessity of (3.2) is immediate, since $\{X_n\}$ relatively compact implies that

$$\lim_{h\to 0}\limsup_{n\to\infty}\sup E\Big[\sup_i q(X_n(t_{i+1}), X_n(\sigma_i))q(X_n(\tau_i), X_n(t_i))\Big] = 0. \tag{3.4}$$

To prove sufficiency, we only need to consider the two families of $\{t_i\}$ defined in Remark 3.2. Note that any interval $[r, s]$ with $s - r \le h/2$ is a subinterval of some interval in one of these partitions.

To show that $\{X_n\}$ is relatively compact, it is enough to consider subsequences along which $\{(Y_n, \gamma_n)\}$ converges in distribution, and hence, as in Theorem 1.1, part (b), we may as well assume that $\{(Y_n, \gamma_n)\}$ converges almost surely to $(Y, \gamma)$. Let $A^p = \gamma^{-1}$. Then, with probability 1, $A_n^p(t) \to A^p(t)$ at every point of continuity of $A^p$. Suppose $\gamma$ is constant on an interval $[u, v] = [A^p(t-), A^p(t)]$ and the condition of Lemma 2.3(b) fails for $Y$ on this interval. Then there exist $\varepsilon > 0$ and $u \le u_1 < u_2 \le v$ such that $(v-u) \wedge q(Y(u), Y(u_1)) \wedge q(Y(u_1), Y(u_2)) \ge \varepsilon$. If this is the case, then there exist $t_n < r_n^1 < r_n^2 \le s_n$ such that $t_n \to t$, $s_n \to t$, $A_n^p(t_n) \to A^p(t-)$, $A_n^p(s_n) \to A^p(t)$, $X_n(r_n^1) \to Y(u_1)$ and $X_n(r_n^2) \to Y(u_2)$. Let $\Gamma_\varepsilon$ be the event that such an interval exists. Then the probability of the event on which $X_n$ does not converge in the Skorohod topology is $\lim_{\varepsilon\to 0} P(\Gamma_\varepsilon)$. Consequently, we are done if we show that $P(\Gamma_\varepsilon) = 0$ for each $\varepsilon > 0$.

Fix $\varepsilon > 0$. For each $t_i$, define $\tau_i = t_{i+1} \wedge \inf\{s > t_i\colon q(X_n(s), X_n(t_i)) \geq \varepsilon\}$ and $\eta_i = t_{i+1} \wedge \inf\{s > \tau_i\colon q(X_n(s), X_n(\tau_i)) \geq \varepsilon\}$. Note that

$$q(X_n(\eta_i), X_n(\tau_i)) \leq q(X_n(t_{i+1}), X_n(\eta_i)) + q(X_n(t_{i+1}), X_n(\tau_i)).$$

On the event $\Gamma_{2\varepsilon}$, for $n$ sufficiently large, one of the two partitions previously designated contains an interval $[t_i, t_{i+1}]$ on which $\eta_i < t_{i+1}$. In this case, for $c$ sufficiently large, the random variable in the expectation in (3.2) is greater than $\varepsilon^3$. Since we only need consider the two partitions for each $h$, (3.2) implies $P(\Gamma_{2\varepsilon}) = 0$, and we conclude that $X_n$ converges in the Skorohod topology.

The bound (3.3) follows from the fact that $X_n(t) = Y_n(A_n^p(t))$. □

The following lemmas are useful in verifying (3.2).

3.3 Lemma. *Suppose $A_n^p \Rightarrow A^p$ in the Skorohod topology. Then* (3.2) *holds if for each $T > 0$ and $c > 0$,*

$$\lim_{h\to 0} \limsup_{n\to\infty} \sup E\Big[\sum_i (c \wedge A_n^p(t_{i+1}) - c \wedge A_n^p(\tau_i)) \times q(X_n(\tau_i), X_n(t_i))\Big] = 0, \tag{3.5}$$

*where the supremum is over all partitions of $[0, T]$ satisfying $t_{i+1} - t_i \leq h$ and all choices of stopping times satisfying $t_i \leq \tau_i \leq t_{i+1}$.*

*With $Y_n$ and $\gamma_n$ as in Remark* 1.2(c), *if $(Y_n, \gamma_n) \Rightarrow (Y, \gamma)$ and $A_n^p \Rightarrow A^p$ in the Skorohod topology, then the limit on the left in* (3.2) *equals*

$$E\Big[\sum_{t\leq T} q(Y(A^p(t)), Y(A^p(t)-))q(Y(A^p(t-)), Y^p(A(t-)-)) \times (c \wedge A^p(t) - c \wedge A^p(t-))\Big]. \tag{3.6}$$

Proof. Writing

$$(c \wedge A_n^p(t_{i+1}) - c \wedge A_n^p(t_i)) = (c \wedge A_n^p(t_{i+1}) - c \wedge A_n^p(\tau_i)) + (c \wedge A_n^p(\tau_i) - c \wedge A_n^p(t_i)), \tag{3.7}$$

we see that the expectation in (3.2) equals the sum of a term bounded by the expectation in (3.5) and

$$E\Big[\sum_i q(X_n(t_{i+1}), X_n(\sigma_i))(c \wedge A_n^p(\tau_i) - c \wedge A_n^p(t_i))\Big]. \tag{3.8}$$

But as in the proof of Theorem 1.1, letting $\beta_n^K = \inf\{t\colon X_n(t) \notin K\}$ and $\tau_n^c = \inf\{t\colon A_n^p(t) \geq c\}$ and using the predictability of $\tau_n^c$,

$$E\big[q(X_n(t_{i+1}), X_n(\sigma_i))|\mathscr{F}_{\sigma_i}^n\big] \leq E\big[\varepsilon + \chi_{\{\tau_n^c \wedge \beta_n^K \leq T\}} + \alpha_K(\varepsilon)^{-1}(c \wedge A_n^p(t_{i+1}) - c \wedge A_n^p(\sigma_i))|\mathscr{F}_{\sigma_i}^n\big]. \tag{3.9}$$

Consequently, (3.8) is bounded by

$$
\begin{aligned}
&c\left(\varepsilon + P\{\tau_n^c \wedge \beta_n^K \le T\}\right) + \alpha_K(\varepsilon)^{-1} \\
&\qquad \times E\left[\sum_i (c \wedge A_n^p(t_{i+1}) - c \wedge A_n^p(\tau_i))(c \wedge A_n^p(\tau_i) - c \wedge A_n^p(t_i))\right]
\end{aligned}
\tag{3.10}
$$

and the convergence of $\{A_n^p\}$ in the Skorohod topology ensures that the expectation in (3.10) goes to zero as $h \to 0$, uniformly in $n$. The probability in (3.10) can be made small uniformly in $n$ by taking $c$ large and applying the compact containment condition. Finally, $\varepsilon$ is arbitrary, so the first part of the lemma follows.

It follows from (3.3) that (3.6) is an upper bound for the limit in (3.2). To see that (3.6) is achieved, drop the supremum on the left of (3.2), select the $t_i$ to satisfy $P\{A(t_i) = A(t_i -)\} = 1$ and let $\sigma_i = \tau_i$ be defined as in Remark 3.2. Then the limit on the left of (3.2) is within $\varepsilon c$ of (3.6). □

3.4 Lemma. *Suppose $A_n^p \Rightarrow A^p$ in the Skorohod topology. Let $\sigma_0^n(\varepsilon) = 0$ and $\sigma_{i+1}^n(\varepsilon) = \inf\{t > \sigma_i^n(\varepsilon)\colon A_n^p(t) - A_n^p(t-) \ge \varepsilon\}$. Then* (3.2) *holds if*

$$
\begin{aligned}
\lim_{h \to 0} \limsup_{n \to \infty} E\Bigg[ &\sum_{\sigma_i^n(\varepsilon) \le T} q(X_n(\sigma_i^n(\varepsilon) -), X_n(\sigma_i^n(\varepsilon) - h)) \\
&\times (c \wedge A_n^p(\sigma_i^n(\varepsilon)) - c \wedge A_n^p(\sigma_i^n(\varepsilon) -))\Bigg] = 0
\end{aligned}
\tag{3.11}
$$

*for each $\varepsilon > 0$, and hence if*

$$
\lim_{h \to 0} \limsup_{n \to \infty} E\left[q(X_n(\sigma_i^n(\varepsilon) -), X_n(\sigma_i^n(\varepsilon) - h))\chi_{\{\sigma_i^n(\varepsilon) \le T\}}\right] = 0 \tag{3.12}
$$

*for each $\varepsilon > 0$ and $i > 0$.*

*Suppose for each $\varepsilon > 0$, $i > 0$ and $\delta > 0$, there exist $0 < h_\delta < \delta$ and stopping times $\tau_i^{n,\delta}$ such that $\liminf_{n \to \infty} P\{h^\delta \le \sigma_i^n(\varepsilon) - \tau_i^{n,\delta} \le \delta\} \ge 1 - \delta$; then* (3.12) *holds.*

Proof. Let $Y_n$ and $\gamma_n$ be as in Remark 1.2(c), and, as before, assume that the convergence is almost sure. Suppose $A^p(\sigma) - A^p(\sigma -) > \varepsilon$. Then there exists $i$ depending on $n$ such that $\sigma_i^n(\varepsilon) \to \sigma$. Recall that $Y_n(u) = Y_n(A_n^p(\sigma_i^n(\varepsilon) -)) = X_n(\sigma_i^n(\varepsilon) -)$ for $A_n^p(\sigma_i^n(\varepsilon) -) \le u < A_n^p(\sigma_i^n(\varepsilon))$. It follows that

$$
\begin{aligned}
\lim_{h \to 0} \lim_{n \to \infty} &q(X_n(\sigma_i^n(\varepsilon) -), X_n(\sigma_i^n(\varepsilon) - h)) \\
&\times (c \wedge A_n^p(\sigma_i^n(\varepsilon)) - c \wedge A_n^p(\sigma_i^n(\varepsilon) -)) \\
&= q(Y(A^p(\sigma -)), Y(A^p(\sigma -) -))(c \wedge A^p(\sigma) - c \wedge A^p(\sigma -))
\end{aligned}
\tag{3.13}
$$

and hence (3.11) implies that (3.6) is zero.

For $\tau_i^{n,\delta}$ as in the statement of the lemma, let $\beta_i^{n,\delta} = \inf\{t > \tau_i^{n,\delta} \wedge T: r(X_n(t), X_n(\tau_i^{n,\delta} \wedge T)) > \eta\}$. Then

$$\begin{aligned}(3.14)\qquad &\limsup_{n\to\infty} E\left[q\left(X_n(\sigma_i^n(\varepsilon)-), X_n(\sigma_i^n(\varepsilon)-h^\delta)\right)\chi_{\{\sigma_i^n(\varepsilon)\le T\}}\right]\\ &\quad\le \delta + 2\eta + \limsup_{n\to\infty} P\{\beta_i^{n,\delta} < \sigma_i^n(\varepsilon) \le T\}.\end{aligned}$$

The inequality in (2.6) and the predictability of $\sigma_i^n(\varepsilon)$ imply that the probability on the right of (3.14) is bounded by

$$\begin{aligned}(3.15)\qquad &\alpha_K(\eta)^{-1}E\left[A_n^p(\sigma_i^n(\varepsilon)-) - A_n^p(\tau_i^{n,\delta})\chi_{\{\tau_i^{n,\delta}<T\}}\right]\\ &\quad + P\{X_n(t)\notin K, \text{ some } t \le T\}.\end{aligned}$$

Since $\{A_n^p\}$ converges in the Skorohod topology, we must have

$$\begin{aligned}(3.16)\qquad &\lim_{\delta\to 0}\limsup_{n\to\infty} E\left[(c\wedge A_n^p(\sigma_i^n(\varepsilon)) - c\wedge A_n^p(\sigma_i^n(\varepsilon)-))\right.\\ &\qquad\left.\times(c\wedge A_n^p(\sigma_i^n(\varepsilon)-) - c\wedge A_n^p(\sigma_i^n(\varepsilon)-\delta))\right] = 0.\end{aligned}$$

Consequently, (3.15) can be made arbitrarily small by taking $K$ large and $\delta$ small. □

The condition in the last statement of Lemma 3.4 can be thought of as a uniform predictability condition. This type of condition plays a key role in the results of Jacod, Mémin and Métivier (1983). We now extend Theorem 2.3 of that paper to the present setting.

3.5 THEOREM. *Under the conditions of Theorem* 1.1, *assume that there exists a continuous function* $\varphi$: $E\times E \to [0,\infty)$ *such that* $\varphi_n \to \varphi$ *uniformly on compact subsets of* $E\times E$. *Let* $Y_n$, $A_n^p$, $\gamma_n$ *and* $B_n$ *be as in the proof of Theorem* 1.1 *and assume that* $(Y_n, \gamma_n) \Rightarrow (Y, \gamma)$ *and* $A_n^p \Rightarrow A^p$ *in the Skorohod topology. Define* $B(t) = \lim_{s\to t+} A^p(\gamma(s)-)$. *Then the following hold.*

(a) *There exists a complete, right-continuous filtration* $\{\mathscr{G}_t\}$ *to which* $(Y,\gamma,B)$ *is adapted such that for each* $a \ge 0$,

$$(3.17)\qquad Z_a(t) \equiv \varphi(Y(t+a), Y(a)) - (B(t+a) - B(a))$$

*is a* $\{\mathscr{G}_{a+t}\}$-*supermartingale*.

(b) *Let* $\mathscr{H}_t = \mathscr{G}_{A^p(t)}$. *If* $A^p$ *is* $\{\mathscr{H}_t\}$-*predictable* (*or more generally if the jump times of* $A^p$ *are predictable in the sense that* $\{(t,\omega): A(t,\omega)\ne A(t-,\omega)\} = \bigcup_k\{(\tau_k(\omega),\omega): \omega\in\Omega\}$ *for some sequence of predictable stopping times* $\{\tau_k\}$), *then* $X_n \Rightarrow X$ *in the Skorohod topology*.

3.6 REMARKS. (a) In part (a), we can always take $\mathscr{G}_t = \mathscr{G}_t^0 \equiv \sigma\{(Y(s),\gamma(s),B(s)): s \le t+\}$, but we may need $\mathscr{G}_t$ to be larger, if possible, in order to obtain the predictability of $A^p$ with respect to $\{\mathscr{G}_{A^p(t)}\}$.

(b) Conditions C1–C4 of Jacod, Mémin and Métivier (1983) imply the $\{\mathscr{H}_t\}$-predictability of $A^p$ (with $\mathscr{G}_t = \mathscr{G}_t^0$). In particular, the conditions on $A^p$

hold if $A^p$ is continuous, if the jump times of $A^p$ are deterministic, or if $A^p$ is predictable with respect to the filtration $\mathscr{H}_t^{A^p} = \sigma\{A^p(s): s \le t\}$.

PROOF. Observe that $(Y_n, \gamma_n, B_n)$ is adapted to $\{\mathscr{G}_t^n\}$ and that

$$Z_a^n(t) \equiv \varphi_n(Y_n(t+a), Y_n(a)) - (B_n(t+a) - B_n(a)) \tag{3.18}$$

is a $\{\mathscr{G}_t^n\}$-supermartingale. Since $Z_a^n \Rightarrow Z_a$ and $Z_a^n(t) \ge -(t+a)$, it follows that $Z_a$ is a $\{\mathscr{G}_{t+a}^0\}$-supermartingale which proves part (a).

Note that if $\tau$ is an $\{\mathscr{H}_t\}$-stopping time, then $A^p(\tau)$ is a $\{\mathscr{G}_t\}$-stopping time. Fix $t_0 < T$, and let $\tau = T \wedge \inf\{t > t_0: A^p(t) \ne A^p(t-)\}$. If $\tau$ is predictable, there exists a sequence of stopping times $\tau_k < \tau$ such that $\tau_k \to \tau$ and $A^p(\tau -)$ and $A^p(\tau_k)$ will be $\{\mathscr{G}_t\}$-stopping times. As in the proof of Theorem 1.1, it follows that

$$\begin{aligned} &E\big[\varphi(Y(A^p(\tau-)), Y(A^p(\tau_k)))\big] \\ &\quad \le E\big[B(A^p(\tau-)) - B(A^p(\tau_k))\big] \\ &\quad \le E\big[A^p(\tau-) - A^p(\tau_k)\big], \end{aligned} \tag{3.19}$$

where the last inequality follows by Lemma 2.5. Since the right side goes to zero as $\tau_k \to \tau$ and the left side converges to $E[\varphi(Y(A^p(\tau-)), Y(A^p(\tau-)-))]$ (recall that $A^p$ is strictly increasing), it follows that (3.6) holds. Consequently, Lemma 3.3 and Proposition 3.1 give part (b). □

**4. Convergence in measure.** Let $(E, r)$ be a complete, separable metric space and let $M_E[0,\infty)$ be the space of equivalence classes of Borel-measurable, $E$-valued functions on $[0,\infty)$ (two functions being equivalent if they are equal Lebesgue a.e.). For $x, y \in M_E[0,\infty)$, define

$$d_m(x,y) = \int_0^\infty e^{-t}[1 \wedge r(x(t), y(t))]\, dt. \tag{4.1}$$

(Note that we will use the same notation for an equivalence class $x$ in $M_E[0,\infty)$ and for an element of the equivalence class provided it makes no difference which element we use.) Then $d_m$ is a metric corresponding to convergence in measure and $(M_E[0,\infty), d_m)$ is a complete, separable metric space. We leave the verification of separability to the reader. To see that the space is complete, let $\{x_n\}$ be a Cauchy sequence. Select a subsequence such that

$$\sum_{k=1}^\infty d_m(x_{n_k}, x_{n_{k+1}}) = \int_0^\infty e^{-t} \sum_{k=1}^\infty 1 \wedge r\big(x_{n_k}(t), x_{n_{k+1}}(t)\big)\, dt < \infty \tag{4.2}$$

and note that (4.2) implies that there is a set $T \subset [0,\infty)$ of full Lebesgue measure on which

$$\sum_{k=1}^\infty 1 \wedge r\big(x_{n_k}(t), x_{n_{k+1}}(t)\big)\, dt < \infty. \tag{4.3}$$

For $t \in T$, $x(t) \equiv \lim_{k\to\infty} x_{n_k}(t)$ exists, and, fixing $x_0 \in E$, define $x(t) \equiv x_0$ for

$t \notin T$. Then

$$d_m(x_n(t), x(t)) \le d_m(x_n(t), x_{n_k}(t)) + \sum_{i=k}^{\infty} d_m(x_{n_i}, x_{n_{i+1}}) \tag{4.4}$$

and it follows that $x_n \to x$ in $M_E[0, \infty)$.

By Prohorov's theorem, a critical factor in the study of weak convergence is the characterization of the compact subsets of the metric space under consideration.

4.1 THEOREM. *A subset $A \subset M_E[0, \infty)$ is relatively compact if and only if the following conditions hold.*

C4.1(i) *For every $\varepsilon, T > 0$, there exists a compact set $K \subset E$ such that*

$$\sup_{x \in A} m\{t \le T: x(t) \notin K\} \le \varepsilon. \tag{4.5}$$

C4.1(ii) *For every $T > 0$,*

$$\lim_{h \to 0} \sup_{x \in A} \int_0^T 1 \wedge r(x(t+h), x(t))\, dt = 0. \tag{4.6}$$

4.2 REMARK. (a) Note that if $\alpha$ is a strictly increasing function with $\alpha(0) = 0$, then

$$\lim_{h \to 0} \sup_{x \in A} \int_0^T \alpha(1 \wedge r(x(t+h), x(t)))\, dt = 0 \tag{4.7}$$

implies (4.6).

(b) An examination of (4.12) shows that C4.1(ii) of Theorem 4.1 can be weakened to the following.

C4.1(iii) *For every $T > 0$, there exists a sequence $h_k \to 0$ such that*

$$\lim_{k \to \infty} \sup_{x \in A} h_k^{-1} \int_0^{h_k} \int_0^T 1 \wedge r(x(t+u), x(t))\, dt\, du = 0. \tag{4.8}$$

PROOF. The proof of necessity is left to the reader. To prove sufficiency, we will make use of the following lemma which will also have application elsewhere.

4.3 LEMMA. *Suppose $A$ satisfies* C4.1(i). *Then $A$ is relatively compact in $M_E[0, \infty)$ if and only if $\{f \circ x: x \in A\}$ is relatively compact in $M_{\mathbb{R}}[0, \infty)$ for every $f \in \overline{C}(E)$.*

PROOF. Again the proof of necessity is left to the reader. Let $K_1 \subset K_2 \subset \cdots$ be a sequence of compact subsets of $E$ such that for every $T > 0$, $\lim_{k \to \infty} \sup_{x \in A} m\{t \le T: x(t) \notin K_k\} = 0$. Let $\{f_i\} \subset \overline{C}(E)$ satisfy $\|f_i\| \le 1$ and separate points in $E$. Then for each $k$, there exists a continuous function $w_k$

with $w_k(0) = 0$, such that

$$r(x,y) \le w_k\left(\sum_{i=1}^{\infty} 2^{-i}|f_i(x) - f_i(y)|\right), \qquad x,y \in K_k. \tag{4.9}$$

Given a sequence in $A$, by a diagonalization argument we can select a subsequence $\{x_n\}$ such that for each $f_i$, $\{f_i \circ x_n\}$ is a Cauchy sequence in $M_{\mathbb{R}}[0,\infty)$. Consequently, by (4.9),

$$\begin{aligned}\limsup_{j,n\to\infty} d_m(x_j, x_n) &\le \limsup_{j,n\to\infty}\left(\int_0^\infty e^{-t} w_k\left(\sum_{i=1}^{\infty} 2^{-i}|f_i \circ x_j(t) - f_i \circ x_n(t)|\right) dt\right.\\ &\qquad\qquad \left. + \int_0^\infty e^{-t}\chi_{\{x_j(t)\notin K_k \text{ or } x_n(t)\notin K_k\}}\, dt\right)\\ &= \limsup_{j,n\to\infty} \int_0^\infty e^{-t}\chi_{\{x_j(t)\notin K_k \text{ or } x_n(t)\notin K_k\}}\, dt\end{aligned} \tag{4.10}$$

and by the definition of $\{K_k\}$, the right side of (4.10) goes to zero as $k \to \infty$. The relative compactness of $A$ follows. □

Proof of Theorem 4.1. It follows from Lemma 4.3 that it is sufficient to verify the compactness of $B_f = \{z \in M_{\mathbb{R}}[0,\infty)\colon z = f \circ x,\ x \in A\}$ for each $f \in \overline{C}(E)$. Let $h > 0$. For $z \in B_f$, define $z^{(h)}$ by

$$z^{(h)}(t) = h^{-1}\int_0^h z(t+u)\, du. \tag{4.11}$$

Then for compact $K \subset E$,

$$\begin{aligned}&\sup_{z\in B_f}\int_0^\infty e^{-t}(1 \wedge |z^{(h)}(t) - z(t)|)\, dt\\ &\quad\le \sup_{z\in B_f} h^{-1}\int_0^h\int_0^\infty e^{-t}|z(t+u) - z(t)|\, dt\, du\\ &\quad\le \sup_{x\in A} h^{-1}\int_0^h\int_0^\infty e^{-t}|f\circ x(t+u) - f\circ x(t)|\, dt\, du\\ &\quad\le \sup_{x\in A} h^{-1}\int_0^h\int_0^\infty e^{-t}\big[\omega_K(r(x(t+u), x(t)))\\ &\qquad\qquad + \chi_{\{x(t+u)\notin K \text{ or } x(t)\notin K\}}\big]\, dt\, du,\end{aligned} \tag{4.12}$$

where $\omega_K(u) = \sup\{|f(x) - f(y)|\colon x, y \in K,\ r(x,y) \le u\}$. Given $\varepsilon > 0$, by making $K$ large and $h$ small, the right side of (4.12) can be made less than $\varepsilon$. Furthermore, for each $h > 0$, $\Gamma^{(h)} = \{z^{(h)}\colon z \in B_f\}$ is relatively compact by Ascoli's theorem (compactness in $C_{\mathbb{R}}[0,\infty)$ implies compactness in $M_{\mathbb{R}}[0,\infty)$). Consequently, we have shown that for each $\varepsilon > 0$, there exists a compact subset $\Gamma \subset M_{\mathbb{R}}[0,\infty)$ such that $B_f \subset \{z\colon \inf_{y\in\Gamma} d_m(z,y) \le \varepsilon\}$ which implies the relative compactness of $B_f$. □

4.4 COROLLARY. *Let $K_1, K_2, \dots$ be compact sets in $E$, and let $h_k > 0$ satisfy $h_k \to 0$. Let $\mathscr{H}(K_1, K_2, \dots, h_1, h_2, \dots)$ be the set of $x \in M_E[0,\infty)$ satisfying*

$$m\{t \le k : x(t) \notin K_k\} \le \frac{1}{k}, \tag{4.13}$$

$$h_k^{-1} \int_0^{h_k} \int_0^T 1 \wedge r(x(t+u), x(t))\, dt\, du \le \frac{1}{k}. \tag{4.14}$$

*Then $\mathscr{H}(K_1, K_2, \dots, h_1, h_2, \dots)$ is compact.*

PROOF. Note that $\mathscr{H}(K_1, K_2, \dots, h_1, h_2, \dots)$ is closed and satisfies C4.1(i) and C4.1(iii). □

Let $X$ be an $E$-valued stochastic process defined on a probability space $(\Omega, \mathscr{F}, P)$. Then $X$ is *measurable* if the mapping $(t, \omega) \to X(t, \omega)$ is $\mathscr{B}[0,\infty) \times \mathscr{F}$-measurable. A measurable stochastic process determines an $M_E[0,\infty)$-valued random variable $\tilde{X}$ ($\tilde{X}(\omega)$ is the equivalence class containing $X(\cdot, \omega)$) and therefore induces a probability distribution on $M_E[0,\infty)$. To see that $\tilde{X}$ is a random variable, note that

$$\{\omega : d_m(\tilde{X}, x) < a\} = \left\{\omega : \int_0^\infty e^{-t}[1 \wedge r(X(t,\omega), x(t))]\, dt < a\right\}.$$

The converse also holds in the sense that for any $M_E[0,\infty)$-valued random variable $\tilde{X}$, there is a measurable process $X$ such that $X(\cdot, \omega) \in \tilde{X}(\omega)$. To see that such an $X$ exists, we construct a canonical $\hat{x} \in x$ for each $x \in M_E[0,\infty)$.

First note that for $f \in B(E)$ and $0 \le a < b$, the mapping $x \to \int_a^b f(x(t))\, dt$ is well-defined on $M_E[0,\infty)$ since the integral is the same for all functions in the equivalence class. Furthermore, the mapping $(x,t) \to h^{-1}\int_t^{t+h} f(x(s))\, ds$ is continuous for $f \in \bar{C}(E)$ and hence $\mathscr{B}(M_E[0,\infty)) \times \mathscr{B}[0,\infty)$-measurable for all $f \in B(E)$. It follows then that the mapping

$$(x,t) \to y_f(x,t) \equiv \limsup_{n\to\infty} n \int_t^{t+n^{-1}} f(x(s))\, ds \tag{4.15}$$

is $\mathscr{B}(M_E[0,\infty)) \times \mathscr{B}[0,\infty)$-measurable for all $f \in B(E)$. Note that by the Lebesgue differentiation theorem, $y_f(x, \cdot\,) \in f \circ x$.

Let $D \subset E$ be a countable dense set and let $\{F_i\}$ be some ordering of the collection of all closed balls in $E$ with center in $D$ and radius $k^{-1}$ for some positive integer $k$. Define $h: E \to \mathbb{R}^\infty$ by $h(x) = (\chi_{F_1}(x), \chi_{F_2}(x), \dots)$. Then $h$ is 1–1 and the range $\Gamma$ of $h$ is a Borel subset of $\mathbb{R}^\infty$. Fix $x_0 \in E$ and define

$$g(y) = \begin{cases} h^{-1}(y), & y \in \Gamma, \\ x_0, & y \in \mathbb{R}^\infty - \Gamma. \end{cases} \tag{4.16}$$

Finally, define $y_i: M_E[0,\infty) \times [0,\infty) \to \mathbb{R}$ by

$$y_i(x,t) = \limsup_{n\to\infty} n \int_t^{t+n^{-1}} \chi_{F_i}(x(s))\, ds \tag{4.17}$$

and $G$: $M_E[0,\infty) \times [0,\infty) \to E$ by

$$G(x,t) = g(y_1(x,t), y_2(x,t),\dots). \tag{4.18}$$

Then $G$ is Borel measurable and $G(x,\cdot) \in x$ for all $x \in M_E[0,\infty)$. It follows that for any $M_E[0,\infty)$-valued random variable $X$, $\hat{X}(t,\omega) \equiv G(X(\omega),t)$ defines a measurable process and $\hat{X}(\cdot,\omega) \in X(\omega)$.

4.5 PROPOSITION. *Let $X$ and $Y$ be measurable stochastic processes. Then $X$ and $Y$ induce the same distribution on $M_E[0,\infty)$ if and only if for $m = 1,2,\dots$ and (Lebesgue) almost every $(t_1,\dots,t_m) \in [0,\infty)^m$, $(X(t_1),\dots,X(t_m))$ and $(Y(t_1),\dots,Y(t_m))$ have the same distribution on $E^m$.*

PROOF. Consider the collection of functions on $M_E[0,\infty)$ of the form

$$F(x) = \prod_{i=1}^{m} \int_0^{T_i} f_i(t, x(t))\,dt, \tag{4.19}$$

where for $i = 1,\dots,m, T_i > 0$ and $T_i \in \overline{C}([0,\infty) \times E)$. This collection of functions is closed under multiplication and separates points in $M_E[0,\infty)$ and therefore it separates measures on $M_E[0,\infty)$ [see Theorem 3.4.5 of Ethier and Kurtz (1986)]. Consequently, $X$ and $Y$ have the same distribution on $M_E[0,\infty)$ if and only if $E[F(X)] = E[F(Y)]$ for each such function, and the proposition follows easily. □

Theorem 4.1 and Prohorov's theorem give the following.

4.6 THEOREM. *Let $\{X_\beta\}$ be a family of measurable $E$-valued processes. Then $\{X_\beta\}$ is relatively compact in $M_E[0,\infty)$ (that is, the set of corresponding distributions is relatively compact) if and only if the following conditions hold.*

C4.6(i) *For every $\varepsilon, T > 0$, there exists a compact $K \subset E$ such that*

$$\sup_\beta \int_0^T P\{X_\beta(t) \notin K\}\,dt \le \varepsilon. \tag{4.20}$$

C4.6(ii) *For every $T > 0$,*

$$\lim_{h\to 0} \sup_\beta E\left[\int_0^T 1 \wedge r(X_\beta(t+h), X_\beta(t))\,dt\right] = 0. \tag{4.21}$$

PROOF. Fix $\varepsilon > 0$ and select $K_k$ such that

$$\sup_\beta \int_0^k P\{X_\beta(t) \notin K_k\}\,dt \le \frac{\varepsilon}{k2^k} \tag{4.22}$$

and $h_k$ such that

$$\sup_\beta E\left[h_k^{-1}\int_0^{h_k}\int_0^k 1 \wedge r(X_\beta(t+u), X_\beta(t))\,dt\,du\right] \le \frac{\varepsilon}{k2^k}. \tag{4.23}$$

Let $\mathscr{H} = \mathscr{H}(K_1, K_2, \dots, h_1, h_2, \dots)$ be as in Corollary 4.4. Then

$$
\begin{aligned}
&P\{X_\beta \notin \mathscr{H}\} \\
(4.24)\qquad &\le \sum_{k=1}^{\infty} P\left\{m\{t \le k: X_\beta(t) \notin K_k\} > \frac{1}{k}\right\} \\
&\quad + \sum_{k=1}^{\infty} P\left\{h_k^{-1}\int_0^{h_k}\int_0^k 1 \wedge r\big(X_\beta(t+u), X_\beta(t)\big)\,dt\,du > \frac{1}{k}\right\} \\
&\le 2\varepsilon
\end{aligned}
$$

and the theorem follows by Prohorov's theorem. □

4.7 Corollary. *Let $\{X_\beta\}$ be a family of measurable E-valued processes. Suppose there exists a dense set $Q \subset [0,\infty)$ such that for each $t \in Q$, $\{X_\beta(t)\}$ is relatively compact, and for every $T > 0$,*

$$
(4.25)\qquad \lim_{h\to 0}\sup_\beta \int_0^T \sup_{u\le h} E\big[1 \wedge r\big(X_\beta(t+u), X_\beta(t)\big)\big]\,dt = 0.
$$

*Then $\{X_\beta\}$ is relatively compact.*

Proof. We must verify C4.6(i). Fix $T > 0$, $0 < \varepsilon < 1$, and select $h$ so that the quantity in the limit in (4.25) is less than $\varepsilon^2$. Select $\{t_i\} \subset Q$ so that $t_1 < h \le t_2 < \cdots < t_{m-1} < T \le t_m$ and $t_{i+1} - t_i < h$, $i = 1, \dots, m-1$. For compact $K \subset E$, let $K^\varepsilon = \{y: \inf_{x\in K} r(x,y) < \varepsilon\}$. Then setting $\gamma(t) = \min\{t_i: t \le t_i\}$,

$$
\begin{aligned}
&\int_0^T P\{X_\beta(t) \notin K^\varepsilon\}\,dt \\
(4.26)\qquad &\le \int_0^T \big[P\{X_\beta(\gamma(t)) \notin K\} + P\{r\big(X_\beta(t), X_\beta(\gamma(t))\big) > \varepsilon\}\big]\,dt \\
&\le T\max_i P\{X_\beta(t_i) \notin K\} + \varepsilon^{-1}\int_0^T \sup_{u\le h} E[1 \wedge r(X(t), X(t+u))]\,dt \\
&\le T\max_i P\{X_\beta(t_i) \notin K\} + \varepsilon.
\end{aligned}
$$

By the definition of $Q$, $K$ can be selected so that the right side of (4.26) is less than $2\varepsilon$ for all $\beta$. For $\delta > 0$ and $n = 1, 2, \dots$, let $K_n$ be a compact set selected in this manner for $\varepsilon = \delta/2^n$, and let $K_0$ be the closure of $\bigcap_{n=1}^\infty K_n^{\varepsilon_n}$. Then $K_0$ is compact (it is complete and totally bounded),

$$
(4.27)\qquad \int_0^T P\{X_\beta(t) \notin K_0\}\,dt \le \sum_{n=1}^\infty \int_0^T P\{X_\beta(t) \notin K_n^{\varepsilon_n}\}\,dt \le \sum_{n=1}^\infty 2\varepsilon_n = 2\delta,
$$

and C4.6(i) follows. □

Theorem 1.5 now follows from Theorem 4.6.

PROOF OF THEOREM 1.5. C4.6(i) follows from C1.5(i); consequently, it remains to show how to estimate the expectation on the left of (4.21) using C1.5(ii). For $\varepsilon > 0$ and $K \subset E$ compact,

$$
\begin{aligned}
E\left[\int_0^T 1 \wedge r(X_n(t+h), X_n(t))\,dt\right]
&\le \varepsilon T + 2E\left[m\{t \le T+h : X_n(t) \notin K\}\right] \\
&\quad + \alpha_K(\varepsilon)^{-1} E\left[\int_0^T \alpha_K(1 \wedge r(X_n(t+h), X_n(t)))\,dt\right] \\
&\le \varepsilon T + 2E\left[m\{t \le T+h : X_n(t) \notin K\}\right] \\
&\quad + \alpha_K(\varepsilon)^{-1} E\left[\int_0^T \varphi_n(X_n(t+h), X_n(t))\,dt\right] \qquad (4.28)\\
&\le \varepsilon T + 2E\left[m\{t \le T+h : X_n(t) \notin K\}\right] \\
&\quad + \alpha_K(\varepsilon)^{-1} \int_0^T (a_n(t+h) - a_n(t))\,dt \\
&\le \varepsilon T + 2E\left[m\{t \le T+h : X_n(t) \notin K\}\right] \\
&\quad + \alpha_K(\varepsilon)^{-1}\left(\int_T^{T+h} a_n(t)\,dt - \int_0^h a_n(t)\,dt\right).
\end{aligned}
$$

The first term on the right can be made arbitrarily small by taking $\varepsilon$ small, the second term can be made small by the choice of $K$ and the third term goes to zero uniformly in $n$ as $h \to 0$ by C1.5(iii). Consequently, C4.6(ii) holds and the theorem follows. □

If a sequence of measurable processes $\{X_n\}$ converges in distribution in $M_E[0,\infty)$, then $\{X_n(t)\}$ need not converge in distribution for any $t$; however, relative compactness of $\{X_n\}$ and convergence in distribution of $\{(X_n(t_1),\dots,X_n(t_m))\}$ for each $m$ and almost every $(t_1,\dots,t_m) \in [0\,\infty)^m$ imply convergence in distribution of $\{X_n\}$. More generally, we have the following.

4.8 THEOREM. *Let $\{X_n\}$ be a sequence of measurable processes and suppose that $\{X_n\}$ is relatively compact in $M_E[0,\infty)$. Then $\{X_n\}$ converges in distribution in $M_E[0,\infty)$ if and only if for each $m = 1,2,\dots$ and each $f \in \overline{C}(E^m)$, the sequence of functions $F_n$: $(t_1,\dots,t_m) \to E[f(X_n(t_1),\dots,X_n(t_m))]$ converges in measure.*

4.9 REMARK. If $\{X_n\}$ converges in distribution in $M_E[0,\infty)$, then there exists a set $B \subset [0,\infty)$ of measure zero and a subsequence along which $(X_n(t_1),\dots,X_n(t_m))$ converges in distribution for all $t_i \notin B$.

4.10 COROLLARY. *Let $\{X_n\}$ be a sequence of measurable processes and suppose that $\{X_n\}$ is relatively compact in $M_E[0,\infty)$ and for each $T>0$,*

$$\lim_{h\to 0}\sup_n\sup_{0\le t\le T} E\big[1\wedge r\big(X_n(t+h),X_n(t)\big)\big]=0. \tag{4.29}$$

*Then $\{X_n\}$ converges in distribution in $M_E[0,\infty)$ if and only if $\{(X_n(t_1),\dots,X_n(t_m))\}$ converges in distribution in $E^m$ for each $m=1,2,\dots$ and each $(t_1,\dots,t_m)\in[0,\infty)^m$.*

Note that the conditions of Theorem 4.6 only involve the finite dimensional distributions of the processes (in fact, only the one- and two-dimensional distributions). It is natural to conjecture that if a consistent family of finite dimensional distributions satisfies these conditions, then the corresponding process has a measurable version. Unfortunately, this conjecture is not correct. Take, for example, the finite dimensional distributions corresponding to a process $X$ such that the random variables $X(t)$ are i.i.d. nondegenerate for $t$ in the Cantor ternary set $\Gamma$ and $X(t)=0$ for $t\notin\Gamma$. Of course, these finite dimensional distributions uniquely determine a probability distribution on $M_{\mathbb{R}}[0,\infty)$ (the zero process), and this observation is true in general.

4.11 THEOREM. *Let $\{\nu_{t_1,\dots,t_m}\in\mathscr{P}(E^m)\colon (t_1,\dots,t_m)\in[0,\infty)^m,\ m=1,2,\dots\}$ be a consistent family of finite dimensional distributions such that $(t_1,\dots,t_m)\mapsto\nu_{t_1,\dots,t_m}(B)$ is Borel-measurable for each $B\in\mathscr{B}(E^m)$ and the following conditions are satisfied.*

C4.11(i) *For each $\varepsilon>0$ and $T>0$, there exists a compact set $K\subset E$ such that*

$$\int_0^T \nu_t(K^c)\,dt\le\varepsilon. \tag{4.30}$$

C4.11(ii) *For each $T>0$,*

$$\lim_{h\to 0}\int_0^T\int 1\wedge r(x,y)\nu_{t,t+h}(dx,dy)=0. \tag{4.31}$$

*Then there exists an $M_E[0,\infty)$-valued random variable $X$ such that*

$$\begin{aligned}E\left[\prod_{i=1}^m\int_0^{T_i}f_i(t,X(t))\,dt\right]\\=\int_0^{T_1}\cdots\int_0^{T_m}\int\prod_{i=1}^m f_i(t_i,x_i)\nu_{t_1\cdots t_m}(dx_1,\dots,dx_m)\,dt_1\cdots dt_m\end{aligned} \tag{4.32}$$

*for all $T_i\ge 0$ and $f_i\in B([0,\infty)\times E)$.*

PROOF. For $n=1,2,\dots,$ let $\tau_0^n=0$ and let $\tau_1^n<\tau_2^n<\cdots$ be the jump times of a Poisson process $N_n$ with parameter $n$. Let $\{Y_i^n\}$ be $E$-valued random variables such that the conditional distribution of $(Y_0^n,Y_1^n,\dots,Y_m^n)$

given $N_n$ is $\nu_{0,\tau_1^n,\ldots\tau_2^n}$. Define $X_n(t) = Y_k^n$, $\tau_k^n \le t < \tau_{k+1}^n$. C4.11(i) and C4.11(ii) imply that $\{X_n\}$ satisfies the conditions of Theorem 4.6 so $\{X_n\}$ is relatively compact and a tedious calculation shows that

$$
\begin{aligned}
&\lim_{n\to\infty} E\left[\prod_{i=1}^m \int_0^{T_i} f_i(t, X_n(t))\,dt\right] \\
&\qquad = \int_0^{T_1}\cdots\int_0^{T_m}\int \prod_{i=1}^m f_i(t_i, x_i)\nu_{t_1\ldots t_m}(dx_1,\ldots,dx_m)\,dt_1\ldots dt_m
\end{aligned}
\tag{4.33}
$$

for all $T_i > 0$ and $f_i \in \overline{C}([0,\infty)\times E)$. □

**5. Processes with finite conditional variation.** In this section we collect some results on processes for which $V_t(X)$ is finite. Throughout the section, $X$ will be a cadlag, real-valued process adapted to a filtration $\{\mathscr{F}_t\}$.

5.1. *Proposition*. Suppose that $E[|X(t)|] < \infty$ and $V_t(X) < \infty$ for each t $\ge 0$. Then $X$ has a unique decomposition $X = M + B$, where $M$ is a local martingale and $B$ is a predictable finite variation process and $B$ satisfies $E[T_t(B)] \le V_t(X)$.

PROOF. Since $X$ is a local quasimartingale [see, e.g., Protter (1990), Section 3], the desired decomposition $X = M + B$ exists and is unique. Let $X^T(t) = X(t) - E[X(T)|\mathscr{F}_t]$ for $0 \le t < T$ and $X^T(t) = 0$ for $t \ge T$. Then $X^T$ is a quasimartingale with $V_\infty(X^T) = V_T(X)$ and hence $X^T = U_1^T - U_2^T$, where $U_1^T$ and $U_2^T$ are positive supermartingales with $E[U_1^T(0) + U_2^T(0)] = V_T(X)$ [see the construction in Protter (1990), Section III.3]. There exist predictable increasing processes $A_i^T$ with $A_i^T(0) = 0$ such that $U_i^T + A_i^T$ is a local martingale. Since for a localizing sequence of stopping times $\{\tau_k\}$, $E[A_i^T(\tau_k \wedge T)] \le E[U_i^T(\tau_k \wedge T) + A_i^T(\tau_k \wedge T)] \le E[U_i^T(0)]$, it follows by the monotone convergence theorem that $E[A_1^T(T) + A_2^T(T)] \le V_T(X)$. Since $X(t) + A_1^T(t) - A_2^T(t) = X^T(t) + A_1^T(t) - A_2^T(t) + E[X(T)|\mathscr{F}_t]$ is a local martingale for $0 \le t \le T$, the uniqueness of the decomposition implies $B(t) = A_1^T(t) - A_2^T(t)$ for $0 \le t \le T$, and hence $E[T_T(B)] \le E[A_1^T(T) + A_2^T(T)] \le V_T(X)$. □

5.2 LEMMA. *The conditional variation of X satisfies*

$$
V_T(X) = \sup E\left[\sum_i \left|E\left[X(\tau_{i+1}) - X(\tau_i)|\mathscr{F}_{\tau_i}\right]\right|\right],
\tag{5.1}
$$

*where the supremum is over all collections of stopping times satisfying* $0 \le \tau_0 \le \cdots \le \tau_m \le T$. *For any stopping time* $\tau$, $E[|X(\tau \wedge T)|] \le V_T(X) + E[|X(T)|]$.

PROOF. Let $X^T = U_1^T - U_2^T$ be as in the proof of Proposition 5.1. Then

$$
\begin{aligned}
& E\Big[\sum_i \big|E\big[X(\tau_{i+1}) - X(\tau_i)|\mathscr{F}_{\tau_i}\big]\big|\Big] \\
&\quad = E\Big[\sum_i \big|E\big[X^T(\tau_{i+1}) - X^T(\tau_i)|\mathscr{F}_{\tau_i}\big]\big|\Big] \\
&\quad \le E\big[U_1^T(0) + U_2^T(0)\big] \\
&\quad = V_T(X)
\end{aligned} \tag{5.2}
$$

and (5.1) follows. For any stopping time $\tau$, $E[|X(\tau \wedge T)|] \le E[|E[X(T) - X(\tau \wedge T)|\mathscr{F}_{\tau\wedge T}]| + |X(T)|] \le V_T(X) + E[|X(T)|]$. □

5.3 LEMMA. *Suppose that $E[|X(t)|] < \infty$ and $V_t(X) < \infty$ for each $t \ge 0$. Then for each $c > 0$,*

$$
cP\Big\{\sup_{0\le t\le T} |X(t)| \ge c\Big\} \le V_T(X) + E\big[|X(T)|\big]. \tag{5.3}
$$

PROOF. Let $\tau = \inf\{t\colon |X(t)| \ge c\}$. Then by Lemma 5.2,

$$
\begin{aligned}
cP\Big\{\sup_{0\le t\le T} |X(t)| \ge c\Big\} &\le E[|X(\tau \wedge T)|] \\
&\le V_T(X) + E[|X(T)|]
\end{aligned} \tag{5.4}
$$

and the lemma follows. □

5.4 LEMMA. *Suppose that $E[|X(t)|] < \infty$ and $V_t(X) < \infty$ for each $t \ge 0$. Let $\tau$ be a stopping time and define $\hat{X}(t) = \chi_{[0,\tau)}X(t)$. Then*

$$
V_t(\hat{X}) + E\big[|\hat{X}(t)|\big] \le V_t(X) + E\big[|X(t)|\big]. \tag{5.5}
$$

PROOF. For any partition of $[0, t]$,

$$
\begin{aligned}
& \sum_i \big|E\big[\hat{X}(t_{i+1}) - \hat{X}(t_i)|\mathscr{F}_{t_i}\big]\big| \\
&\quad \le \sum_i \big|E\big[X(t_{i+1}) - X(t_i)|\mathscr{F}_{t_i}\big]\big|\chi_{\{\tau > t_i\}} \\
&\qquad + \sum_i \big|E\big[\hat{X}(t_{i+1}) - X(t_{i+1})|\mathscr{F}_{t_i}\big]\big|\chi_{\{\tau > t_i\}} \\
&\quad \le \sum_i \big|E\big[X(t_{i+1}) - X(t_i)|\mathscr{F}_{t_i}\big]\big|\chi_{\{\tau > t_i\}} \\
&\qquad + \sum_i \big|E\big[(X(t) - X(t_{i+1}))\chi_{\{t_i < \tau \le t_{i+1}\}}|\mathscr{F}_{t_i}\big]\big| \\
&\qquad + \sum_i \big|E\big[X(t)\chi_{\{t_i < \tau \le t_{i+1}\}}|\mathscr{F}_{t_i}\big]\big|,
\end{aligned} \tag{5.6}
$$

and taking expectations it follows that

$$(5.7)\quad E\left[\sum_i \left|E\left[\hat{X}(t_{i+1}) - \hat{X}(t_i)|\mathscr{F}_{t_i}\right]\right|\right] \le V_t(X) + E\left[|X(t)|\chi_{\{\tau \le t\}}\right].$$

Noting that $E[|\hat{X}(t)|] = E[|X(t)|\chi_{\{\tau > t\}}]$, the lemma follows.

5.5 LEMMA. *Let $\varphi$ be convex and satisfy $|\varphi'| \le K$. Then for $L^1$ random variables $X$ and $Y$ and any $\sigma$-algebra $\mathscr{D}$,*

$$(5.8)\quad |E[\varphi(X) - \varphi(Y)|\mathscr{D}]| \le 2K|E[X - Y|\mathscr{D}]| + E[\varphi(X) - \varphi(Y)|\mathscr{D}].$$

PROOF. Let $\theta = \chi_{\{E[\varphi(X)-\varphi(Y)|\mathscr{D}]<0\}}$. Then

$$(5.9)\quad \begin{aligned} |E[\varphi(X) - \varphi(Y)|\mathscr{D}]| &= E[\varphi(X) - \varphi(Y)|\mathscr{D}] - 2\theta E[\varphi(X) - \varphi(Y)|\mathscr{D}] \\ &= E[\varphi(X) - \varphi(Y)|\mathscr{D}] \\ &\quad - 2\theta E[\varphi(X) - \varphi(Y) - \varphi'(Y)(X - Y)|\mathscr{D}] \\ &\quad - 2\theta\varphi'(Y)E[X - Y|\mathscr{D}] \\ &\le E[\varphi(X) - \varphi(Y)|\mathscr{D}] + 2K|E[X - Y|\mathscr{D}]|. \end{aligned}$$

□

5.6 LEMMA. *Suppose that $E[|X(t)|] < \infty$ and $V_t(X) < \infty$ for each $t \ge 0$. Let $\varphi$ be convex and satisfy $|\varphi'| \le K$. Then*

$$(5.10)\quad V_t(\varphi \circ X) \le 2KV_t(X) + E[\varphi(X(t)) - \varphi(X(0))].$$

PROOF. The inequality follows immediately from (5.8). □

5.7 THEOREM. *Suppose that $E[|X(t)|] < \infty$ and $V_t(X) < \infty$ for each $t \ge 0$. Let $\varphi$ be convex, symmetric, $C^2$ and satisfy $\varphi(0) = 0$, $\varphi''(0) = 1$, $\varphi''$ nonincreasing on $[0, \infty)$ and $\varphi''(1) = 0$. Then there exists an increasing process $A$ such that for each $a \ge 0$,*

$$(5.11)\quad \varphi(X(a + t) - X(a)) - (A(a + t) - A(a))$$

*if a local $\{\mathscr{F}_{a+t}\}$-supermartingale and for $\tau_b = \inf\{t\colon |X(t)| \vee |X(t-)| \ge b\}$,*

$$(5.12)\quad E[A(\tau_b \wedge T)] \le E[|X(T)|] + 2(1 + b)V_T(X) + b^2 + 2b.$$

PROOF. Let $X = M + B$ be the decomposition of Proposition 5.1, let $\Delta X(s) = X(s) - X(s-)$ and let $[X]^c$ denote the continuous part of the quadratic variation of $X$. Define

$$(5.13)\quad A(t) = T_t(B) + \frac{1}{2}[X]_t^c + 4\sum_{s \le t}\varphi\left(\frac{\Delta X(s)}{2}\right).$$

Note that $\varphi(u) \le |u| \wedge u^2$, so $A(t) \le T_t(B) + [X]_t < \infty$ for each $t > 0$. More precisely,

$$
\begin{aligned}
A(t) &\le T_t(B) + 2|X(t) - X(t-)| + [X]_{t-} \\
&\le T_t(B) + 2|X(t) - X(t-)| + X^2(t-) + 2\int_0^t X(s-)\,dX(s) \\
&\le T_t(B) + 2|X(t) - X(t-)| + X^2(t-) \\
&\quad + 2\int_0^t X(s-)\,dB(s) + 2\int_0^t X(s-)\,dM(s) \\
&\le T_t(B) + 2|X(t) - X(t-)| + X^2(t-) \\
&\quad + 2\int_0^t |X(s-)|\,dT_s(B) + 2\int_0^t X(s-)\,dM(s).
\end{aligned}
\tag{5.14}
$$

Consequently,

$$
\begin{aligned}
A(\tau_b \wedge T) &\le (1 + 2b)T_T(B) + 2|X(\tau_b \wedge T)| + 2b + b^2 \\
&\quad + 2\int_0^{\tau_b \wedge T} X(s-)\,dM(s)
\end{aligned}
\tag{5.15}
$$

and (5.12) follows by Proposition 5.1 and Lemma 5.2.

We verify that (5.11) is a local supermartingale assuming that $a = 0$ and $X(0) = 0$. By Itô's formula, for $t > u$,

$$
\begin{aligned}
\int_u^t \varphi'(X(s-))\,dM(s) &= \varphi(X(t)) - \varphi(X(u)) \\
&\quad - \int_u^t \varphi'(X(s-))\,dB(s) - \tfrac{1}{2}\int_u^t \varphi''(X(s-))\,d[X]_s^c \\
&\quad - \sum_{u<s\le t} \big(\varphi(X(s)) - \varphi(X(s-)) \\
&\qquad - \varphi'(X(s-))\,\Delta X(s)\big) \\
&\ge \varphi(X(t)) - \varphi(X(u)) - (A(t) - A(u))
\end{aligned}
\tag{5.16}
$$

and since the left side is an increment of a local martingale and the right side is locally integrable, the right side is an increment of a local supermartingale. □

The following theorem is essentially Theorem 4 of Meyer and Zheng (1984). Let $V_t^0(X)$ denote the conditional variation of $X$ defined relative to the natural filtration $\mathscr{F}_t^X = \sigma(X(s): s \le t)$.

5.8 THEOREM. *Let $\{X_n\}$ be a sequence of cadlag, real-valued processes such that for each $t > 0$,*

$$
C(t) = \sup_n \big(V_t(X_n) + E[|X_n(t)|]\big) < \infty. \tag{5.17}
$$

*Then $\{X_n\}$ is relatively compact in $M_E[0,\infty)$ and any limit point $X$ has a cadlag version satisfying $V_t^0(X) + E[|X(t)|] \le C(t)$ for all but countably many $t$.*

PROOF. The relative compactness of $\{X_n\}$ and the existence of a cadlag version for any limit point is a consequence of Theorem 1.1. Let $Y_n$ and $\gamma_n$ be defined as in Remark 1.2(c). By Theorem 1.1, part (b), we can assume (at least along a subsequence) that $X_n = Y_n \circ \gamma_n^{-1}$, $(Y_n, \gamma_n) \to (Y, \gamma)$ a.s. in $D_{\mathbb{R}^2}[0,\infty)$ and that $X_n(t) \to X(t) = Y \circ \gamma^{-1}(t)$ a.s. for all but countably many $t$. Let $\eta_n^c = \inf\{t\colon |Y_n(t)| \vee |Y_n(t-)| \ge c\}$. For all but countably many $c$, $\eta_n^c \to \eta^c = \inf\{t\colon |Y(t)| \vee |Y(t-)| \ge c\}$ a.s. Select such a $c$ and define $\tau_n = \gamma_n \circ \eta_n^c$ ($= \inf\{t\colon |X_n(t)| \vee X_n(t-)| > c$) and $X_n^c(t) = X_n(t)\chi_{[0,\tau_n)}(t)$, with $X^c$ defined similarly using $\tau = \gamma \circ \eta^c = \lim_{n\to\infty} \tau_n$. Then $X_n^c(t) \to X^c(t)$ a.s. for all but countably many $t$. Let $D$ be the countable exceptional set.

For $i = 1, 2, \dots$, let $g_i \in \overline{C}(\mathbb{R}^{m_i})$ satisfy $\|g_i\|_\infty \le 1$. Then for any partition $\{t_i\}$ of the interval $[0,t]$ with $t_i \notin D$ and any choice of $s_i^j \notin D$,

$$\begin{aligned}
&\sum_i E\big[\big(X^c(t_{i+1}) - X^c(t_i)\big)g_i\big(X(s_i^1), \dots, X(s_i^{m_i})\big)\big] + E\big[|X^c(t)|\big] \\
&\quad = \lim_{n\to\infty}\Big(\sum_i E\big[\big(X_n^c(t_{i+1}) - X_n^c(t_i)\big)g_i\big(X_n(s_i^1), \dots, X_n(s_i^{m_i})\big)\big] \\
&\qquad\qquad + E\big[|X_n^c(t)|\big]\Big) \qquad (5.18)\\
&\quad \le C(t),
\end{aligned}$$

where the last inequality follows from Lemma 5.4. Letting $c \to \infty$ on the left and then taking the supremum over the $g_i$ gives the desired inequality. □

**Acknowledgments.** This work was completed while the author was visiting the Department of Mathematics at the University of Utah. The hospitality of that department is gratefully acknowledged. The author would also like to thank Philip Protter for many helpful discussions.

## REFERENCES

DELLACHERIE, C. and MEYER, P.-A. (1978). *Probabilities and Potential*. North-Holland, Amsterdam.

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley, New York.

JACOD, J., MÉMIN, J. and MÉTIVIER, M. (1983). On tightness and stopping times. *Stochastic Process. Appl.* **14** 109–146.

KURTZ, T. G. (1975). Semigroups of conditioned shifts and approximation of Markov processes. *Ann. Probab.* **3** 618–642.

MEYER, P.-A. and ZHENG, W. A. (1984). Tightness criteria for laws of semimartingales. *Ann. Inst. H. Poincaré Probab. Statist.* **20** 353–372.

PROTTER, P. (1990). *Stochastic Integration and Differential Equations: A New Approach*. Springer, New York.

DEPARTMENTS OF MATHEMATICS AND STATISTICS
UNIVERSITY OF WISCONSIN–MADISON
MADISON, WISCONSIN 53706