*The Annals of Probability*
1975, Vol. 3, No. 4, 618-642

# SEMIGROUPS OF CONDITIONED SHIFTS AND APPROXIMATION OF MARKOV PROCESSES

BY THOMAS G. KURTZ

*The Australian National University and*
*The University of Wisconsin, Madison*

Let $\mathscr{L}$ be the space of processes, progressively measurable with respect to an increasing family of $\sigma$-algebras $\{\mathscr{F}_t\}$ and having finite mean. Then $\mathscr{T}(s)f(t) = E(f(t+s)\,|\,\mathscr{F}_t)$, $f \in \mathscr{L}$, defines a semigroup of linear operators on $\mathscr{L}$. Using $\mathscr{T}(s)$ and known semigroup approximation theorems, techniques are developed for proving convergence in distribution of a sequence of (possibly non-Markov) processes to a Markov process. Results are also given which are useful in proving weak convergence. In particular for a sequence of Markov prcesses $\{X_n(t)\}$ it is shown that if the usual semigroups $(T_n(t)f(x) = E(f(X_n(t))\,|\,X(0) = x))$ converge uniformly in $x$ for $f$ continuous with compact support, then the processes converge weakly.

**1. Introduction.** A number of authors [13, 15, 27, 28, 32, 33, 34, 35] have applied operator semigroup approximation theorems to prove convergence of sequences of Markov processes. Our primary purpose in this paper is to develop techniques for applying these theorems to prove convergence of sequences of non-Markov processes to Markov processes. The work of Borovkov [6] and Gikhman [10] provide examples of the type of theorem we have in mind and served as the major motivation for the work in this paper.

These techniques, based on a semigroup of operators that may have other applications as well, should also be useful in dealing with sequences of Markov processes.

We have also included a number of results concerning weak convergence of sequences of Markov processes. In particular Theorem (4.29) states that under very general circumstances uniform convergence of the semigroups of a sequence of Markov processes implies weak convergence in the Skorokhod topology. This result is close to work of Skorokhod [28], and is also reminiscent of a theorem of Liggett [19] who shows that if the finite dimensional distributions of a sequence of diffusion processes (or birth and death processes) converge to those of a diffusion, then the processes converge weakly.

Let $(\Omega, \mathscr{F}, P)$ be a probability space and $\{\mathscr{F}_t : t \in [0, \infty)\}$ (or $\{\mathscr{F}_t : t \in (-\infty, \infty)\}$) be an increasing family of $\sigma$-algebras, $\mathscr{F}_t \subset \mathscr{F}$. Let $\mathscr{L}$ be the linear space of real valued, processes $f(t, w) \equiv f(t)$ progressively measurable with respect to $\{\mathscr{F}_t\}$ such that $f(t)$ has a finite expectation for all $t$. It can be shown (see Appendix) that for every $s$ $E(f(t+s)\,|\,\mathscr{F}_t)$ has a version that is progressively

Received October 29, 1973; revised January 9, 1975.

*AMS* 1970 *subject classifications*. Primary 60F99; Secondary 60J25, 60J35, 60J60, 60G45.

*Key words and phrases*. Markov process, approximation, operator semigroup, conditional expectation, weak convergence.

measurable. (Throughout, we will identify processes $f$ and $g$ if $P\{f(t) = g(t)\} = 1$ for all $t$.) We define

$$(1.1) \qquad \mathscr{T}(s)f(t) = E(f(t+s) \mid \mathscr{F}_t) .$$

It is easy to check that $\mathscr{T}(s)$ is a semigroup of linear operators on $\mathscr{L}$, and it is natural to call $\mathscr{T}(s)$ a semigroup of conditioned shifts.

There are a number of different possibilities for norms. We define three:

$$(1.2) \qquad \|f\|_1 = E(\textstyle\int_0^\infty e^{-t}|f(t)|\,dt) ;$$

$$(1.3) \qquad \|f\|_2 = \sup_t E(|f(t)|) ;$$

and

$$(1.4) \qquad \|f\|_3 = E(\textstyle\int_0^\infty e^{-t}|f(t)|^2\,dt)^{\frac{1}{2}} .$$

Each of the above is a norm on a space of equivalence classes of processes in $\mathscr{L}$ and each has certain advantages. The first is the weakest; under the third the space of processes with finite norm is a Hilbert space; and for the second $\mathscr{T}(s)$ is a semigroup of contractions. We observe that under $\|\cdot\|_1$, $\|\mathscr{T}(s)\| \leqq e^s$ and under $\|\cdot\|_3$, $\|\mathscr{T}(s)\| \leqq e^{1/2s}$.

In what follows we will concentrate on $\|\cdot\| \equiv \|\cdot\|_2$ and we will now use $\mathscr{L}$ to denote the space of progressively measurable processes having finite norm, (or more precisely the space of equivalence classes of processes having finite norm). We will note occasionally how a similar development is possible for $\|\cdot\|_1$ and $\|\cdot\|_3$. Following the development of semigroups in Dynkin [8], let $\mathscr{L}_0$ denote the subspace of $\mathscr{L}$ on which $\mathscr{T}(s)$ is strongly continuous, and let $\mathscr{A}$ denote the (strong) infinitesimal operator.

We define a notion of convergence weaker than strong convergence as follows: For $\{f_n\} \subset \mathscr{L}$

$$(1.5) \qquad p\text{-}\lim_{n\to\infty} f_n = f$$

if and only if

$$(1.6) \qquad \sup_n \|f_n\| < \infty$$

and

$$(1.7) \qquad \lim_{n\to\infty} E(|f_n(t) - f(t)|) = 0 \qquad \text{for every} \quad t .$$

Let $\hat{\mathscr{L}}_0$ denote the subspace of $\mathscr{L}$ on which $\mathscr{T}(s)$ is $p$-right continuous and define the $p$-infinitesimal operator by

$$\hat{\mathscr{A}}f = p\text{-}\lim_{s\to 0} \frac{1}{s}(\mathscr{T}(s)f - f)$$

if the limit exists and is in $\hat{\mathscr{L}}_0$. The properties of the $p$-infinitesimal operator are much the same as those of the weak-infinitesimal operator as discussed in Dynkin [8]. For example $\mathscr{D}(\hat{\mathscr{A}}) \subset \mathscr{L}_0$,

$$(1.8) \qquad (\lambda - \hat{\mathscr{A}})^{-1}f = \textstyle\int_0^\infty e^{-\lambda s}\mathscr{T}(s)f\,ds$$

for every $f \in \hat{\mathscr{L}}_0$, and $\lambda > 0$;

$$\mathscr{T}(s)f - f = \int_0^s \mathscr{T}(u)\hat{\mathscr{A}}f\,du \tag{1.9}$$

for every $f \in \mathscr{D}(\hat{\mathscr{A}})$; and

$$\mathscr{T}(s)f - f = \hat{\mathscr{A}} \int_0^s \mathscr{T}(u)f\,du \tag{1.10}$$

for every $f \in \hat{\mathscr{L}}_0$.

We observe that the fixed points of $\mathscr{T}(s)$ are precisely the processes that are martingales with respect to $\{\mathscr{F}_t\}$. (This observation essentially appears in Meyer [22], page 186.) More generally, any $f \in \mathscr{D}(\hat{\mathscr{A}})$ is a process of bounded variation in the sense of Föllmer [9] and

$$Z(t) = f(t) - f(0) - \int_0^t \hat{\mathscr{A}}f(u)\,du \tag{1.11}$$

is a zero-mean martingale with respect to $\{\mathscr{F}_t\}$. This follows from (1.9) by observing

$$\begin{aligned} E(Z(t+s) - Z(t) \mid \mathscr{F}_t) &= E(f(t+s) \mid \mathscr{F}_t) - f(t) - \int_t^{t+s} E(\hat{\mathscr{A}}f(u) \mid \mathscr{F}_t)\,du \\ &= \mathscr{T}(s)f(t) - f(t) - \int_0^s \mathscr{T}(u)\hat{\mathscr{A}}f(t)\,du \\ &= 0\,. \end{aligned}$$

We note that $f(t) = Z(t) + f(0) + \int_0^t \hat{\mathscr{A}}f(u)\,du$ is the (Fisk–Orey–Rao) decomposition of $f(t)$ into a martingale and a previsible process.

This leads to the following generalization of Dynkins Identity:

(1.12) PROPOSITION. Suppose $\tau$ is a stopping time with $\tau \geqq t$ almost surely, $f(t)$ is right continuous and $E(\sup_{t \leqq \tau} |Z(t)|) < \infty$.

Then

$$E(f(\tau) \mid \mathscr{F}_t) - f(t) = E\left(\int_t^\tau \hat{\mathscr{A}}f(s)\,ds \mid \mathscr{F}_t\right). \tag{1.13}$$

The proof is immediate since the Optional Sampling Theorem implies $E(Z(\tau) \mid \mathscr{F}_t) = Z(t)$. The conditions are satisfied if $E(\tau) < \infty$, $f(t)$ and $\hat{\mathscr{A}}f(t)$ are bounded and $f(t)$ is right continuous.

In Section 2, in the case of Markov processes, we examine the relationship of the above semigroup to the semigroup usually studied. In Section 3 we develop the general technique for proving convergence of the finite dimensional distributions of a sequence of processes to those of a Markov process and in Section 4 we consider the question of weak convergence. Section 5 is intended to be a user's guide to the application of semigroups in the proof of convergence theorems. If it is helpful thanks should be given to the referees and to Donald Iglehart whose suggestions led to its inclusion. In general the author is very appreciative of the careful reading given the original manuscript by the referees. Their efforts helped to clarify a number of points.

NOTE. For $\|\cdot\|_3$ the weaker notion of convergence corresponding to (1.6) and

(1.7) would require $\sup_n \|f_n\|_3 < \infty$ and

(1.14) $$\lim_{n\to\infty} m\{t \leqq T: E((f_n(t) - f(t))^2) > \varepsilon\} = 0$$

for all $T$, $\varepsilon > 0$, where $m$ denotes Lebesgue measure. In this case (and for the similar definition corresponding to $\|\cdot\|_1$) $\hat{\mathscr{L}}_0 = \mathscr{L}_0$ and hence $\hat{\mathscr{A}} = \mathscr{A}$.

**2. Markov processes.** We will only consider processes taking values on a complete, separable, locally compact space $E$ with metric $\rho$, although much of what we will say can be generalized if the need arises. Let $\mathscr{B}$ denote the Borel subsets of $E$ and $B \equiv B(E, \mathscr{B})$, the real valued, bounded, Borel measurable functions on $E$. Let $X(t)$ be a right continuous, $E$-valued stochastic process progressively measurable with respect to the increasing family of $\sigma$-algebras $\{\mathscr{F}_t\}$. Let $\mathscr{M}_1$ denote the linear space of processes $f(t, X(t))$, where $f(t, x)$: $[0, \infty) \times E \to \mathbb{R}$ is bounded and jointly measurable in $t$ and $x$, and let $\mathscr{M}_2$ denote the linear space of processes $f(X(t))$ where $f \in B(E, \mathscr{B})$. We observe that $X(t)$ is Markov if $\mathscr{T}(s)$ leaves $\mathscr{M}_1$ invariant and is Markov and temporally homogeneous if $\mathscr{T}(s)$ leaves $\mathscr{M}_2$ invariant. In the latter case $\mathscr{T}(s)$ induces a semigroup of operators $T(s)$ on equivalence classes of functions in $B(E, \mathscr{B})$: two functions $f_1, f_2 \in B(E, \mathscr{B})$ are equivalent if $P\{X(t) \in \Gamma\} = 0$ for all $t$, where $\Gamma = \{x : f_1(x) \neq f_2(x)\}$.

In this case $X(t)$ has a transition function $P(t, x, \Gamma)$, and $T(s)$ can be represented by

(2.1) $$T(s)f(x) = \int f(y)P(t, x, dy)\,.$$

The existence of $P(t, x, \Gamma)$ can be verified in much the same way as the existence of regular conditional distributions.

Since in general $X(t)$ does not uniquely determine a transition function, $T(s)$ may have more than one representation of this form. We will assume that we have selected one such representation. Let $\tilde{A}$ denote the weak-infinitesimal operator for $T(s)$ given by (2.1), considered as an operator on $B(E, \mathscr{B})$ under the supremum norm. (See Dynkin [8].) If $f \in \mathscr{D}(\tilde{A})$ then $g(t) \equiv f(X(t)) \in \mathscr{D}(\hat{\mathscr{A}})$ and $\hat{\mathscr{A}}g(t) = \tilde{A}f(X(t))$.

**3. Convergence to Markov processes.** In this section we consider a sequence of processes $\{X_n(t)\}$ which we assume to be defined on the same probability space and adapted to the same increasing family of $\sigma$-algebras $\{\mathscr{F}_t\}$. Since we are only concerned with convergence in distribution, this assumption serves only to simplify notation. Note that if $X_n(t)$ is defined on $(\Omega_n, \mathscr{F}_n, P_n)$ and adapted to $\{\mathscr{F}_{n,t}\}$, then we can always consider $(\Omega, \mathscr{F}, P) \equiv (\prod_n \Omega_n, \prod \mathscr{F}_n, \prod_n P_n)$ and $\mathscr{F}_t \equiv \prod_n \mathscr{F}_{n,t}$.

(3.1) PROPOSITION. Let $X_n(t)$ be a sequence of $E$-valued stochastic processes and let $X(t)$ be an $E$-valued Markov process with semigroup $T(s)$. Suppose $T(s)$ is strongly continuous (in the sup norm) on a subspace $K \subset B(E, \mathscr{B})$ which contains $\hat{C}$ (the space of continuous functions vanishing at infinity) and such that $f \in \hat{C}$, $g \in K$ implies $f \cdot g \in K$.

Suppose

$$\lim_{n\to\infty} E(f(X_n(0)) = E(f(X(0))) \tag{3.2}$$

for evey $f \in K$ and

$$\lim_{n\to\infty} E(|E(f(X_n(t+s))|\mathscr{F}_t) - T(s)f(X_n(t))|) = 0 \tag{3.3}$$

for all $s, t \geqq 0$ and every $f \in K$. Then the finite dimensional distributions of $X_n(t)$ converge to the finite dimensional distributions of $X(t)$.

REMARK. The assumptions on $K$ are not very restrictive. In many applications $K = \hat{C}$. In general if $L_0$ is the largest subspace of $B(E, \mathscr{B})$ on which $T(s)$ is strongly continuous and $\hat{C} \subset L_0$, then $f \in \hat{C}$ and $g \in L_0$ implies $f \cdot g \in L_0$. (Note we are assuming $P\{X(t) \in E\} = 1$ for all $t$).

PROOF. It is sufficient to show

$$\begin{aligned}\lim_{n\to\infty} E(f_1(X_n(t_1))f_2(X_n(t_2)) \cdots f_k(X_n(t_k))) \\ = E(f_1(X(t_1))f_2(X(t_2)) \cdots f_k(X(t_k)))\end{aligned} \tag{3.4}$$

for all $t_i \geqq 0$ and $f_i \in \hat{C}$.

We first observe that for $g \in K$

$$\begin{aligned}|E(g(X_n(t))) - E(g(X(t)))| \leqq |E(E(g(X_n(t))\,|\,\mathscr{F}_0) - T(t)g(X_n(0))| \\ + |E(T(t)g(X_n(0))) - E(T(t)g(X(0))|\,.\end{aligned} \tag{3.5}$$

The first term on the right goes to zero by (3.3) and the second by (3.2).

We now illustrate the proof of (3.4) in the case $k = 2$ and $0 < t_1 < t_2$:

$$\begin{aligned}&|E(f_1(X_n(t_1))f_2(X_n(t_2))) - E(f_1(X(t_1))f_2(X(t_2)))| \\ &\quad = |E(f_1(X_n(t_1))E(f_2(X_n(t_2))\,|\,\mathscr{F}_{t_1})) - E(f_1(X(t_1))T(t_2 - t_1)f_2(X(t_1)))| \\ &\quad \leqq |E(f_1(X_n(t_1))(E(f_2(X_n(t_2))\,|\,\mathscr{F}_{t_1}) - T(t_2 - t_1)f_2(X_n(t_1))))| \\ &\qquad + |E(f_1(X_n(t_1))T(t_2 - t_1)f_2(X_n(t_1)) - f_1(X(t_1))T(t_2 - t_1)f_2(X(t_1)))|\,.\end{aligned}$$

The first term on the right goes to zero by (3.3) and the second by (3.5), since $f_1(x)T(t_2 - t_1)f_2(x)$ is in $K$.

We observe that (3.3) describes a type of approximation of $T(s)$ by $\mathscr{T}(s)$. The results in [14] give semigroup approximation theorems for very general types of convergence. We will use these results to obtain conditions implying (3.3). Let $\{X_n(t)\}$ be a sequence of $E$-valued processes, and let $\mathscr{K}$ be the Banach space of bounded sequences $\{f_n\} \subset \mathscr{L}$ with $\|\{f_n\}\| = \sup_n \|f_n\|$. For $\{f_n\} \in \mathscr{K}$ and $f \in B(E, \mathscr{B})$ define

$$P\{f_n\} = f$$

if $p - \lim_{n\to\infty} f_n - f(X_n(\cdot)) = 0$.

For $f \in B(E, \mathscr{B})$ define

$$\|f\|_0 = \inf_{P\{f_n\}=f} \sup_n \|f_n\|\,. \tag{3.6}$$

One can check that $\|f\|_0 = \sup_t \limsup_{n\to\infty} E(|f(X_n(t))|)$.

We may consider $||f||_0$ as a norm on a space of equivalence classes of measurable functions, i.e., $f_1$ and $f_2$ are equivalent if $||f_1 - f_2||_0 = 0$. The equivalence class of a function $f$ will be denoted by $\hat{f}$. Let $L$ denote the completion of this space with respect to $||\cdot||_0$.

While $P$ is multivalued when considered as a mapping into $B(E, \mathscr{B})$, it is a contraction as a mapping into $L$ and hence it may be extended to a closed subspace of $\mathscr{K}$. Furthermore $P\{f_n\} = 0$ implies $P\{\mathscr{T}(s)f_n\} = 0$ and $P\{\int_0^\infty e^{-\lambda s}\mathscr{T}(s)f_n\, ds\} = 0$ for $\lambda > 0$. Consequently $P$ is precisely the type of operator considered in [14]. Theorem (2.13) of [14] implies

(3.7) THEOREM. *For $\hat{f}, \hat{g} \in L$, define $A\hat{f} = \hat{g}$ if there exists a sequence $\{f_n\} \subset \mathscr{D}(\hat{\mathscr{A}})$ such that $P\{f_n\} = \hat{f}$ and $P\{\hat{\mathscr{A}}f_n\} = \hat{g}$. Let $L_0$ be a subspace of $L$ and let $A_0$ be the restriction of $A$ to*

$$\mathscr{D}(A_0) = \{\hat{f} \in \mathscr{D}(A) \cap L_0 : A\hat{f} \in L_0\} .$$

*If $\mathscr{D}(A_0)$ and $\mathscr{R}(\lambda - A_0)$ are dense in $L_0$ for some $\lambda > 0$, then the closure of $A_0$ generates a strongly continuous semigroup $T_0(s)$ on $L_0$. If $f, g \in B(E, \mathscr{B})$, $\hat{f}, \hat{g} \in L_0$ and $\hat{g} = T_0(s)\hat{f}$, then*

$$\text{(3.8)} \qquad p\text{-lim}\,(E(f(X_n(t + s)) \,|\, \mathscr{F}_t) - g(X_n(t))) = 0 .$$

(3.9) COROLLARY. *Let $K$ be a Banach subspace of $B(E, \mathscr{B})$ with the supremum norm and suppose $T(s)$ is a semigroup of operators on $K$. Suppose each $f \in K$ is in some equivalence class in $L_0$ (denote it by $\hat{f}$) and $T(s)f$ is in $T_0(s)\hat{f}$ for all $s \geqq 0$. Then for $f \in K$*

$$\text{(3.10)} \qquad \lim_{n\to\infty} E(|E(f(X_n(t + s)) \,|\, \mathscr{F}_t) - T(s)f(X_n(t))|) = 0$$

*for all $s, t \geqq 0$.*

REMARK. Not a great deal can be said in general about $||\cdot||_0$ and $L$. However, what is important for applications is that $||\cdot||_0$ is weaker than the supremum norm. With this in mind the following is an immediate consequence of Theorem (3.7), Corollary (3.9) and Proposition (3.1).

(3.11) THEOREM. *Let $K$ be a Banach subspace of $B(E, \mathscr{B})$ with the supremum norm satisfying the conditions of Proposition (3.1) and suppose $T(s)$ is strongly continuous semigroup on $K$ with infinitesimal operator $A$ corresponding to a Markov process $X(t)$. Let $\{X_n(t)\}$ be a sequence of $E$-valued processes and let $D$ be the set of $f \in \mathscr{D}(A)$ such that there exists $\{f_n\} \subset \mathscr{D}(\hat{\mathscr{A}})$ with*

$$\text{(3.12)} \qquad p\text{-lim}\, f_n - f(X_n(\cdot)) = 0$$

*and*

$$\text{(3.13)} \qquad p\text{-lim}\, \hat{\mathscr{A}}f_n - Af(X_n(\cdot)) = 0 .$$

*Suppose $D$ and $\mathscr{R}(\lambda - A|_D)$ are dense in $K$ in the supremum norm for some $\lambda > 0$ ($A|_D$ denotes the restriction of $A$ to $D$.) Then for every $f \in K$ and all $s, t > 0$*

$$\text{(3.14)} \qquad \lim E(|E(f(X_n(t + s)) \,|\, \mathscr{F}_t) - T(s)f(X_n(t))|) = 0 .$$

*If in addition* $\lim_{n\to\infty} E(f(X_n(0))) = E(f(X(0)))$ *for every* $f \in K$, *then the finite dimensional distributions of* $X_n(t)$ *converge to those of* $X(t)$.

PROOF. Denoting the equivalence class of a function by $\hat{f}$, let $\hat{K} = \{\hat{f} : f \in K\}$ and let $L_0$ in Theorem (3.7) and Corollary (3.9) be the closure of $\hat{K}$ under $\|\cdot\|_0$. The fact that $\mathscr{D}(A_0)$ and $\mathscr{R}(\lambda - A_0)$ are dense in $L_0$ follows from the fact that $D$ and $\mathscr{R}(\lambda - A|_D)$ are dense in $K$. Since $\|\cdot\|_0$ is weaker than the supremum norm, these conditions also imply that $(\lambda - A)^{-1}f$ is in the equivalence class $(\lambda - A_0)^{-1}\hat{f}$ for all $f \in K$. This in turn implies the conditions of Corollary (3.9).

The following theorem specializes the above results to a sequence of Markov processes. In introducing the mappings $\eta_n : E_n \to E$ we have in mind such situations as $E_n$ being a subset of $E$ and $\eta_n$ being the natural injection; $E_n$ consisting of vectors and $\eta_n$ being a projection; or $E_n$ consisting of sets and $\eta_n$ giving the cardinality, perhaps normalized in some way.

(3.15) THEOREM. *Let* $K$ *be a Banach subspace of* $B(E, \mathscr{B})$ *satisfying the conditions of Proposition* (3.1) *and suppose* $T(s)$ *is a strongly continuous semigroup on* $K$, *with infinitesimal operator* $A$, *corresponding to a Markov process*, $X(t)$. *Let* $\{Y_n(t)\}$ *be a sequence of Markov processes with measurable state spaces* $(E_n, \mathscr{B}_n)$ *and weak infinitesimal operators* $\tilde{A}_n$, *and let* $\eta_n : E_n \to E$ *be measurable mappings. Define* $X_n(t) = \eta_n(Y_n(t))$ *and assume*

$$\lim_{n\to\infty} E(f(X_n(0))) = E(f(X(0))) \tag{3.16}$$

*for every* $f \in K$.

*Let* $D$ *be the subset of* $\mathscr{D}(A)$ *such that for* $f \in D$ *there are* $f_n \in \mathscr{D}(\tilde{A}_n)$ *such that*

$$\sup_n \sup_t E(|f_n(Y_n(t))|) < \infty\ , \tag{3.17}$$

$$\sup_n \sup_t E(|\tilde{A}_n f_n(Y_n(t))|) < \infty\ , \tag{3.18}$$

$$\lim_{n\to\infty} E(|f_n(Y_n(t)) - f(X_n(t))|) = 0 \qquad \textit{for every} \quad t\ , \tag{3.19}$$

*and*

$$\lim_{n\to\infty} E(|\tilde{A}_n f_n(Y_n(t)) - Af(X_n(t))|) = 0 \qquad \textit{for every} \quad t\ . \tag{3.20}$$

*If* $D$ *and* $\mathscr{R}(\lambda - A|_D)$ *are dense in* $K$ (i.e., $A$ *is the closure of* $A|_D$) *then the finite dimensional distributions of* $X_n(t)$ *converge to those of* $X(t)$.

REMARK. Most applications of semigroup approximation to the convergence of sequences of Markov processes up to this point have involved strong convergence which in our context would mean

$$\lim_{n\to\infty} \sup_y |f_n(y) - f(\eta_n(y))| = 0\ . \tag{3.21}$$

The type of convergence used above is obviously much weaker. However, we will see in the next section that the use of strong convergence can in many cases give us stronger results, namely weak convergence in the Skorokhod topology.

In order to be able to apply Theorem (3.11) to a sequence of non-Markov processes one must be able to find elements of $\mathscr{D}(\hat{\mathscr{A}})$. Potentially useful elements of $\mathscr{D}(\hat{\mathscr{A}})$ can be constructed as follows:

Suppose $g \in B(E, \mathscr{B})$ and $g(X_n(t))$ is right continuous. Then $g(X_n(t))$ is in $\hat{\mathscr{L}}_0$ and for $\varepsilon > 0$

$$f(t) \equiv \frac{1}{\varepsilon} \int_0^\varepsilon E(g(X_n(t+s)) \mid \mathscr{F}_t)\, ds \tag{3.22}$$

is in $\mathscr{D}(\hat{\mathscr{A}})$ with

$$\hat{\mathscr{A}} f(t) = \frac{1}{\varepsilon} (E(g(X_n(t+\varepsilon)) \mid \mathscr{F}_t) - g(X_n(t))) . \tag{3.23}$$

We could prove discrete parameter analogs of Theorem (3.11) and Theorem (3.15) but in the light of (3.22) and (3.23) it is simplest to show how to apply Theorem (3.11) to sequences of discrete parameter processes $\{X_n(k)\}$, $k = 0, 1, 2, \cdots$.

Ordinarily what one wants to do is to prove that $Y_n(t) = X_n([t/\varepsilon_n])$ converges to a Markov process $X(t)$ where $\varepsilon_n > 0$ and $\lim_{n\to\infty} \varepsilon = 0$. Observe that if $g(Y_n(t))$ is right continuous and bounded and

$$f(t) = \frac{1}{\varepsilon_n} \int_0^{\varepsilon_n} E(g(Y_n(t+s)) \mid \mathscr{F}_t)\, ds ,$$

then

$$\hat{\mathscr{A}} f(t) = \frac{1}{\varepsilon_n} (E(g(X_n([t/\varepsilon_n] + 1)) \mid \mathscr{F}_t) - g(X_n([t/\varepsilon_n]))) .$$

Ordinarily this is what will be needed to obtain the desired result. If $[t/\varepsilon] = k$ then typically

$$E(g(X_n(k+1)) \mid \mathscr{F}_t) = E(g(X_n(k+1)) \mid X_n(0) \cdots X_n(k)) .$$

**4. Weak convergence.** Many of the results in this section appear either explicitly or implicitly in other places, particularly in Skorokhod [28], Gikhman and Skorokhod [11] and Borovkov [6].

We assume that the reader is familar with the theory of weak convergence in the space $D(0, 1)$. (see Billingsley [3].) By weak convergence in $D(0, \infty)$ we mean weak convergence in $D(0, T_k)$ for each $T_k$ in some sequence with $\lim_{k\to\infty} T_k = \infty$. (See Lindvall [21].) Modifying the appropriate theorem in Billingsley we have the following criteria for tightness in $D(0, \infty)$. Recall that $E$ is a complete, separable, locally compact metric space.

(4.1) THEOREM. *Let $\{X_n(t)\}$ be a sequence of E-valued processes whose sample paths are right continuous and have left limits* (i.e., *whose sample paths are in $D_E(0, \infty)$*). *Let*

$$w'(X_n, \delta, T) = \inf_{\{t_i\}} \max_i \sup_{t_{i-1} \leqq s, t < t_i} \rho(X_n(s), X_n(t))$$

*where $\{t_i\}$ ranges over all partitions of the form $0 = t_0 < t_1 < \cdots < t_{r-1} < T \leqq t_r$ with $\min_{0 \leqq i \leqq r} (t_i - t_{i-1}) \geqq \delta$. The sequence of processes is tight if and only if*

(4.2) *for every $T > 0$ and $\eta > 0$ there is a compact set $K$ such that*

$$\liminf_{n\to\infty} P\{X_n(t) \in K \ \text{all}\ 0 \leqq t \leqq T\} > 1 - \eta ;$$

*and*

(4.3) *for every* $\varepsilon > 0$, $\eta > 0$, *and* $T > 0$ *there is a* $\delta > 0$ *such that*

$$\limsup_{n\to\infty} P\{w'(X_n, \delta, T) \geqq \varepsilon\} \leqq \eta .$$

We now state several lemmas. Some proofs are straightforward and are omitted.

(4.4) LEMMA. *Let* $\{X_n(t)\}$ *be a sequence of E-valued processes with sample paths in* $D_E(0, \infty)$. *Suppose Condition* (4.2) *is satisfied and that for every* $\varepsilon > 0$ *there is a sequence of processes* $\{X_n^\varepsilon(t)\}$ *that is tight such that* $\sup_t \rho(X_n(t), X_n^\varepsilon(t)) \leqq \varepsilon$. *Then* $\{X_n(t)\}$ *is tight.*

PROOF. Observe that $w'(X_n, \delta, T) \leqq 2\varepsilon + w'(X_n^\varepsilon, \delta, T)$. Consequently $P\{w'(X_n, \delta, T) \geqq 3\varepsilon\} \leqq P\{w'(X_n^\varepsilon, \delta, T) \geqq \varepsilon\}$.

(4.5) LEMMA. *Let* $X(t)$ *be a right continuous pure jump process. Let* $\tau_1, \tau_2, \cdots$ *be the jump times of* $X(t)$ *and let* $\Delta_k = \tau_k - \tau_{k-1}$ $(\tau_0 = 0)$. *Suppose there is a distribution function* $F(x)$ *such that* $P\{\Delta_k \leqq x\} \leqq F(x)$ *for all* $k$. *Let* $K(T) = \max\{k : \tau_{k-1} < T\}$. *Then for all integers* $L > 0$

$$P\{w'(X, \delta, T) > 0\} = P\{\min_{k \leqq K(T)} \Delta_k < \delta\} \leqq LF(\delta) + e^T \int_{0-}^{\infty} e^{-Lx}\, dF(x) . \tag{4.6}$$

REMARK. Observe that if $\lim_{x\to 0} F(x) = 0$ then the right hand side of (4.6) can be made arbitrarily small by taking $L$ large and $\delta$ small.

PROOF. The equality in (4.6) follows immediately from the definition of $w'(X, \delta, T)$ and $\Delta_k$.

$$\begin{aligned} P\{\min_{1\leqq k \leqq K(T)} \Delta_k < \delta\} &\leqq \textstyle\sum_{k=1}^{L} P\{\Delta_k < \delta\} + P\{K(T) > L\} \\ &\leqq LF(\delta) + e^T E(\exp(-\textstyle\sum_{k=1}^{L} \Delta_k)) \\ &\leqq LF(\delta) + e^T \textstyle\prod_{k=1}^{L} E(e^{-L\Delta_k})^{1/L} \\ &\leqq LF(\delta) + e^T \textstyle\int_{0-}^{\infty} e^{-Lx}\, dF(x) . \end{aligned} \tag{4.7}$$

To obtain weak convergence results from Lemma (4.4) and Lemma (4.5) we need an efficient method of approximating a process $X(t)$ by pure jump processes.

Let $\tau_0 = 0$ and for $k > 0$ define

$$\tau_k = \inf\{t > \tau_{k-1} : \rho(X(t), X(\tau_{k-1})) > \varepsilon\} , \tag{4.8}$$

and

$$s_k = \sup\{t < \tau_k : \rho(X(t), X(\tau_k)) \geqq \varepsilon\} . \tag{4.9}$$

Let

$$\begin{aligned} X^\varepsilon(t) &= X(0) && \text{for } t < \tfrac{1}{2}(s_1 + \tau_1) \\ &= X(\tau_k) && \text{for } \tfrac{1}{2}(s_k + \tau_k) \leqq t < \tfrac{1}{2}(s_{k+1} + \tau_{k+1}) . \end{aligned} \tag{4.10}$$

Since $\frac{1}{2}(s_1 + \tau_1) \geqq \frac{1}{2}\tau_1$ and

$$\frac{s_{k+1} + \tau_{k+1}}{2} - \frac{s_k + \tau_k}{2} \geqq \frac{\tau_k + \tau_{k+1}}{2} - \frac{s_k + \tau_k}{2} = \frac{\tau_{k+1} - s_k}{2}$$

in order to apply Lemma (4.5) it is sufficient to estimate the distributions of $\tau_1$ and $\tau_{k+1} - s_k$.

With this in mind we give the following lemma similar to a lemma of Skorokhod.

(4.11) LEMMA. *Let $X(t)$ be right continuous and let $r(x, y) = \rho(x, y) \wedge 1$. (Note that $r$ is still a metric.) Suppose $\gamma(\delta)$ is a random variable such that for some $\beta > 0$*

$$E(\gamma(\delta) \mid \mathscr{F}_t) \geqq E(r^\beta(X(t + u), X(t)) \mid \mathscr{F}_t) r^\beta(X(t), X(t - v)) \quad \text{a.s.} \tag{4.12}$$

*for all $t, u, v$ satisfying $0 \leqq t \leqq T$, $0 \leqq u \leqq \delta \wedge (T - t)$ and $0 \leqq v \leqq (2\delta) \wedge t$. Let $\tau$ be a stopping time with $\tau \leqq T - \delta$. Then*

$$\begin{aligned} P\{\sup\nolimits_{u \leqq \delta} r(X(\tau + u), X(\tau)) \geqq \varepsilon, \sup\nolimits_{v \leqq \delta \wedge \tau} r(X(\tau), X(\tau - v)) \geqq \varepsilon\} \\ \leqq \frac{(a_\beta + 2a_\beta^2) E(\gamma(\delta))}{\varepsilon^{2\beta}}, \end{aligned} \tag{4.13}$$

$$P\{\sup\nolimits_{u \leqq \delta} r(X(u), X(0)) \geqq \varepsilon\} \leqq \frac{3a_\beta E(\gamma(\delta)) + a_\beta^2 E(r^{2\beta}(X(\delta), X(0)))}{\varepsilon^{2\beta}} \tag{4.14}$$

*where $a_\beta$ is a constant such that*

$$(x + y)^\beta \leqq a_\beta(x^\beta + y^\beta) \qquad \textit{for} \quad x, y \geqq 0 .$$

PROOF. We will prove the lemma in the case $\beta = 1$. For general $\beta$ the proof is the same except that the triangle inequality for $r(x, y)$ must be replaced by

$$r^\beta(x, y) \leqq a_\beta(r^\beta(x, z) + r^\beta(z, y)) .$$

First observe that (4.12) holds with $t$ replaced by any stopping time bounded by $T$. This can be seen by approximating the stopping time by a decreasing sequence of discrete stopping times. Similarly if $\tau_1$ is a stopping time and $\tau_2$ is a stopping time with respect to $\{\mathscr{F}_{\tau_1 + u}\}$ such that $\tau_2 \leqq \delta$, then

$$\begin{aligned} E(\gamma(\delta) \mid \mathscr{F}_{\tau_1 + \tau_2}) \geqq E(r(X(\tau_1 + \delta), X(\tau_1 + \tau_2)) \mid \mathscr{F}_{\tau_1 + \tau_2}) \\ \times r(X(\tau_1 + \tau_2), X(\tau_1 + \tau_2 - v)) \quad \text{a.s.} \end{aligned} \tag{4.15}$$

Since the left hand side is independent of $v$, $v$ may be replaced by any random variable $V$ with $0 \leqq V \leqq (\tau_1 + \tau_2) \wedge 2\delta$. Let $\Delta = \inf\{t > 0 : r(X(\tau + t), X(\tau)) > \varepsilon\}$.

$$\begin{aligned} & E(r(X(\tau + \Delta \wedge \delta), X(\tau)) \mid \mathscr{F}_\tau) r(X(\tau), X(\tau - v)) \\ & \quad \leqq E(E(r(X(\tau + \delta), X(\tau + \Delta \wedge \delta)) \mid \mathscr{F}_{\tau + \Delta \wedge \delta}) \\ & \qquad \times r(X(\tau + \Delta \wedge \delta), X(\tau)) \mid \mathscr{F}_\tau) \\ & \qquad + E(E(r(X(\tau + \delta), X(\tau + \Delta \wedge \delta)) \mid \mathscr{F}_{\tau + \Delta \wedge \delta}) \\ & \qquad \times r(X(\tau + \Delta \wedge \delta), X(\tau - v)) \mid \mathscr{F}_\tau) \\ & \qquad + E(r(X(\tau + \delta), X(\tau)) \mid \mathscr{F}_\tau) r(X(\tau), X(\tau - v)) \\ & \quad \leqq 3E(\gamma(\delta) \mid \mathscr{F}_\tau) , \qquad \text{for} \quad v \leqq \delta \wedge \tau , \end{aligned} \tag{4.16}$$

and hence

$$(4.17)\qquad E(r(X(\tau + \Delta \wedge \delta), X(\tau))(\sup_{v \leqq \delta \wedge \tau} r(X(\tau), X(\tau - v)))) \leqq 3E(\gamma(\delta))\,.$$

The left hand side of (4.17) bounds

$$\varepsilon^2 P\{\sup_{u \leqq \delta} r(X(\tau + u), X(\tau)) \geqq \varepsilon,\ \sup_{v \leqq \delta \wedge \tau} r(X(\tau), X(\tau - v)) \geqq \varepsilon\}$$

and (4.13) follows.

Now let $\Delta = \inf\{t > 0 : r(X(t), X(0)) > \varepsilon\}$

$$(4.18)\qquad \begin{aligned} r(X(\Delta \wedge \delta), X(0))^2 \leqq{}& r(X(\delta), X(\Delta \wedge \delta))r(X(\Delta \wedge \delta), X(0)) \\ &+ r(X(\delta), X(0))r(X(\Delta \wedge \delta), X(0))\,. \end{aligned}$$

This gives

$$(4.19)\qquad \begin{aligned} &E(r^2(X(\Delta \wedge \delta), X(0))) \\ &\quad\leqq E(\gamma(\delta)) + [E(r^2(X(\delta), X(0))E(r^2(X(\Delta \wedge \delta), X(0)))]^{\frac{1}{2}}\,. \end{aligned}$$

A little algebraic manipulation gives (4.14).

REMARK. Relating Lemma (4.11) to $X^\varepsilon(t)$, note that the left hand side of (4.14) is $P\{\tau_1 \leqq \delta\}$ and the left hand side of (4.13) bounds

$$P\{\tau_{k+1} - s_k \leqq \delta, \tau_1 > \delta\}\,.$$

Lemmas (4.4), (4.5) and (4.11) imply

(4.20) THEOREM. *Let $\{X_n(t)\}$ be a sequence of processes with sample paths in $D_E(0, \infty)$ that satisfies Condition* (4.2). *Suppose for each $T > 0$ and $n$ there are random variables $\gamma_n(\delta)$ such that for some $\beta > 0$*

$$(4.21)\qquad E(\gamma_n(\delta) \mid \mathscr{F}_t) \geqq E(r^\beta(X_n(t + u), X_n(t)) \mid \mathscr{F}_t) r^\beta(X_n(t), X_n(t - v)) \quad \text{a.s.}$$

*for all $t$, $u$, $v$ satisfying $0 \leqq t \leqq T$, $0 \leqq u \leqq \delta \wedge (T - t)$ and $0 \leqq v \leqq (2\delta) \wedge t$.*

*If $\lim_{\delta\to 0} \limsup_{n\to\infty} E(\gamma_n(\delta)) = 0$ and $\lim_{\delta\to 0} \limsup_{n\to\infty} E(r(X_n(\delta), X_n(0))) = 0$ then the sequence $\{X_n(t)\}$ is tight.*

REMARK. Intuitively (4.21) says that if we have had a significant change over the last little time interval then we do not expect to have one over the next. In fact the conditions of this theorem are necessary for tightness. Of course $E(\gamma_n(\delta) \mid \mathscr{F}_t) \geqq E(r^\beta(X_n(t + u), X_n(t)) \mid \mathscr{F}_t)$ implies (4.21). Billingsley [4] contains conditions that imply (4.21).

Before turning our attention to Markov processes, we observe that if Condition (4.2) is satisfied then tightness of $\{X_n(t)\}$ is equivalent to tightness in $D_{\mathbf{R}}(0, \infty)$ of $\{g(X_n(t))\}$ for every real valued continuous function with compact support. Alternatively if the finite dimensional distributions of $\{X_n(t)\}$ converge to the finite dimensional distributions of a process $X(t)$ with sample paths in $D_E(0, \infty)$, then tightness of $\{g(X_n(t))\}$ for all continuous $g$ with compact support implies weak convergence of $\{X_n(t)\}$. Condition (4.2) follows by considering $\{g(X_n(t))\}$ where $g$ has compact support $K_1$ and $g \equiv 1$ on a compact set $K_2$.

Note

$$P\{X_n(t) \in K_1 \text{ all } t \leqq T\} \geqq P\{g(X_n(t)) = 1 \text{ all } t \leqq T\},$$

and

$$\liminf_{n\to\infty} P\{g(X_n(t)) = 1 \text{ all } t \leqq T\} \geqq P\{g(X(t)) > \tfrac{1}{2} \text{ all } t \leqq T\}.$$

To see this note that $\hat{C}(E)$, the space of continuous functions vanishing at infinity with the sup norm, is separable. For a countable dense subset $g_1, g_2, \cdots$ define

$$\hat{\rho}(x, y) = \sum_{i=1}^{\infty} \frac{1}{2^i} \frac{|g_i(x) - g_i(y)|}{1 + |g_i(x) - g_i(y)|}.$$

$\hat{\rho}$ is a metric and for any compact set $K$ there is an increasing, right continuous, function $\nu_K(z)$, $z \geqq 0$, with $\lim_{z\to 0} \nu_K(z) = 0$ such that

$$\rho(x, y) \leqq \nu_K(\hat{\rho}(x, y)) \qquad \text{for all} \quad x, y \in K.$$

If $X(s) \in K$ for $0 \leqq s \leqq T$ then

$$w'(X, \delta, T) \leqq \nu_K\left(\sum_{i=1}^{k} w'(g_i(X), \delta, T) + \frac{1}{2^{k-1}}\right)$$

for all $k$.

Of course in making use of the above observation we only need to consider a dense (in the sup norm) collection of $g$. In particular for $E = \mathbb{R}^r$ we can confine our attention to smooth $g$.

We now consider the special case of a sequence $\{X_n(t)\}$ of conservative Markov processes. Let $g$ be continuous with compact support $\Gamma$. We want to find $\gamma_n(\delta)$ for $g(X_n(t))$.

$$\begin{aligned} &E((g(X_n(t+u)) - g(X_n(t)))^2 | \mathscr{F}_t) \\ (4.22) \qquad &= T_n(u)g^2(X_n(t)) - g^2(X_n(t)) \\ &\quad - 2g(X_n(t))(T_n(u)g(X_n(t)) - g(X_n(t))). \end{aligned}$$

If we take

$$\begin{aligned} (4.23) \qquad \gamma_n(\delta) = \sup_{u\leqq\delta} [&\sup_x \chi_{K_T}(x)|T_n(u)g^2(x) - g^2(x)| \\ &+ \sup_{x\in\Gamma} 2|g(x)|\,|T_n(u)g(x) - g(x)|] \end{aligned}$$

where $K_T$ is the set visited by $X_n(t)$ up to time $T$ then (4.21) is satisfied with $\beta = 2$.

Note that $\gamma_n(\delta)$ is bounded by the constant $c_n(\delta)$ given by

$$\begin{aligned} (4.24) \qquad c_n(\delta) = &\sup_{u\leqq\delta} \sup_{x\in\Gamma} (|T_n(u)g^2(x) - g^2(x)| + 2|g(x)|\,|T_n(u)g(x) - g(x)|) \\ &+ \sup_{u\leqq\delta} \sup_{x\notin\Gamma} T_n(u)g^2(x). \end{aligned}$$

For a diffusion process or a birth and death process

$$(4.25) \qquad T_n(u)g^2(x) = E_x(T_n(u - \tau \wedge u)g^2(X(\tau \wedge u)))$$

where $\tau$ is the first hitting time for $\Gamma$ (in the case of a birth death process take

$\Gamma$ to be the set from which the support of $g$ is accessible in one jump) and since $g^2(X(\tau \wedge u)) = 0$

$$(4.26) \qquad \sup_{u \leq \delta} \sup_{x \notin \Gamma} T_n(u) g^2(x) \leqq \sup_{u \leq \delta} \sup_{x \in \Gamma} |T_n(u) g^2(x) - g^2(x)| .$$

If $\{T_n(t)\}$ converges strongly on a Banach space $K$ containing $\hat{C}$ to a semigroup $T(t)$ which is strongly continuous on $K$ then the convergence is uniform on bounded $t$-intervals and

$$(4.27) \qquad \lim_{\delta \to 0} \lim_{n \to \infty} c_n(\delta) = \lim_{\delta \to 0} c(\delta) = 0 .$$

Consequently $\{g(X_n(t)\}$ is tight.

If Condition (4.2) is satisfied and $T_n(t)g$ converges to $T(t)g$ uniformly on compact sets for all $g \in K$ the convergence is again uniform on bounded $t$-intervals and we still have

$$(4.28) \qquad \lim_{\delta \to 0} \lim_{n \to \infty} E(\gamma_n(\delta)) = 0$$

and hence tightness for $g\{X_n(t))\}$.

If we drop the assumption that Condition (4.2) holds then covergence of $T_n(t)g$ uniformly on compact sets for each $t$ no longer implies that the convergence is uniform on bounded $t$-intervals. However, if $\{X_n(t)\}$ are diffusions or birth and death processes and the limiting process is a diffusion, then the convergence is still uniform on bounded $t$-intervals. This coupled with (4.26) implies $\lim_{\delta \to 0} c_n(\delta) = 0$ for every continuous $g$ with compact support, and hence $\{g(X_n(t))\}$ is tight for every such $g$. Consequently, if the limiting diffusion has sample paths in $D(0, \infty)$ then $\{X_n(t)\}$ converges weakly. (Note: Weak convergence in $D(0, \infty)$ to a process with sample paths in $C(0, \infty)$ implies weak convergence in the uniform metric. See Billingsley [3] page 150.) This is Liggett's Theorem [19].

We summarize the above in the following

(4.29) THEOREM. *Let $K$ be a Banach subspace of $B(E, \mathscr{B})$ satisfying the conditions of Proposition (3.1) and suppose $T(s)$ is a strongly continuous semigroup on $K$, with infinitesimal operator $A$, corresponding to a Markov process $X(t)$ with sample paths in $D_E(0, \infty)$. Let $\{Y_n(t)\}$ be a sequence of Markov processes with measurable state spaces $(E_n, \mathscr{B}_n)$ and weak infinitesimal operators $\tilde{A}_n$, and let $\eta_n : E_n \to E$ be measurable mappings. Define $X_n(t) = \eta_n(Y_n(t))$ and suppose $X_n(t)$ has sample paths in $D_E(0, \infty)$ and*

$$(4.30) \qquad \lim_{n \to \infty} E(f(X_n(0)) = E(f(X(0))$$

*for every $f \in K$.*

*For $f \in K$ let $P_n f(y) \equiv f(\eta_n(y))$.*

*Suppose*

$$(4.31) \qquad \lim_{n \to \infty} \sup_y |T_n(t) P_n f(y) - T(t) f(\eta_n(y))| = 0$$

*for all $t \geqq 0$ and $f \in K$. Then $\{X_n(t)\}$ converges weakly to $X(t)$.*

*Let $D$ be the set of $f \in \mathscr{D}(A)$ such that there exist $f_n \in \mathscr{D}(\tilde{A}_n)$ with*

$$(4.32) \qquad \lim_{n \to \infty} \sup_y |f_n(y) - f(\eta_n(y))| = 0$$

*and*

(4.33) $$\lim_{n\to\infty} \sup_y |\tilde{A}_n f_n(y) - Af(\eta_n(y))| = 0 .$$

*It is necessary and sufficient for* (4.31) *that* $D$ *and* $\mathscr{R}(\lambda - A|_D)$ *be dense in* $K$ *for some* $\lambda > 0$.

If Condition (4.2) holds then uniform convergence in (4.31), (4.32) and (4.33) can be replaced by uniform convergence on compact subsets.

REMARK. This theorem is very close to results in [28]. The conditions for the convergence of the semigroups are a modification of those given by Trotter [32] and are given in [13].

The following is an example in which we have uniform convergence on compact sets of the semigroups but do not have weak convergence of the processes: Let $X_n(t)$ be the Markov chain on $\{0, 1, 2, \cdots\}$ with infinitesimal parameters $q^n_{0j} = 0$ all $j$; for $i \neq 0, n$, $q^n_{ij} = 0$ all $j \neq i, n$, $-q^n_{ii} = q^n_{in} = 1$; $q^n_{nj} = 0$ all $j \neq 0, n$, and $-q^n_{nn} = q^n_{n0} = n$. Let $X(t)$ be the Markov chain with infinitesimal parameters $q_{0j} = 0$ all $j$; for $i > 0$, $q_{ij} = 0$ all $j \neq i, 0$, and $-q_{ii} = q_{i0} = 1$. If $X_n(0) = X(0) = i > 0$ then the finite dimensional distributions converge but the sequence does not converge weakly.

We close this section with a simple lemma that should be quite useful in verifying that a sequence $\{g(X_n(t))\}$ is tight where $g$ is a bounded real valued function such that $g(X_n(t))$ has sample paths in $D_{\mathbb{R}}(0, \infty)$ and $\{X_n(t)\}$ is a sequence of Markov processes.

(4.34) LEMMA. *Let* $\{X_n(t)\}$ *be a sequence of Markov processes with weak infinitesimal operators* $\{\tilde{A}_n\}$ *and let* $g \in B(E, \mathscr{B})$ *be such that* $g(X_n(t))$ *has sample paths in* $D_{\mathbb{R}}(0, \infty)$ *for all* $n$.

*Suppose* $g_n \in \mathscr{D}(\tilde{A}_n)$ *satisfy* $g_n{}^2 \in \mathscr{D}(\tilde{A}_n)$,

$$\sup_n \sup_x |\tilde{A}_n g_n(x)| < \infty , \qquad \mathrm{snp}_n \sup_x |\tilde{A}_n g_n{}^2(x)| < \infty$$

*and*

(4.35) $$\lim_{n\to\infty} \sup_x |g_n(x) - g(x)| = 0 .$$

*Then* $\{g(X_n(t))\}$ *is tight.*

PROOF. By (4.26) it is enough to prove that $\{g_n(X_n(t))\}$ is tight. Furthermore by the uniform boundedness of $\tilde{A}_n g_n$ it is enough to prove the sequence

(4.36) $$Z_n(t) = g_n(X_n(t)) - g_n(X_n(0)) - \int_0^t \tilde{A}_n g_n(X_n(s))\, ds$$

is tight. But $Z_n(t)$ is a zero mean martingale hence it follows that for any partition $s = u_0 < u_1 < \cdots < u_n = t$ we have

(4.37) $$\begin{aligned} &E((Z(t) - Z(s))^2 | \mathscr{F}_s) \\ &\quad = \textstyle\sum_i E((Z(u_{i+1}) - Z(u_i))^2 | \mathscr{F}_s) \\ &\quad = \textstyle\sum_i E(E((g_n(X_n(u_{i+1})) - g_n(X_n(u_i)))^2 | \mathscr{F}_{u_i}) | \mathscr{F}_s) \\ &\qquad + O(\max_i (u_{i+1} - u_i)) \end{aligned}$$

But

$$
\begin{aligned}
&E((g_n(X_n(u_{i+1})) - g_n(X_n(u_i)))^2 \mid \mathscr{F}_{u_i}) \\
&\quad = E(g_n^2(X_n(u_{i+1})) \mid \mathscr{F}_{u_i}) - g_n^2(X_n(u_i)) \\
&\qquad - 2g_n(X_n(u_i))(E(g_n(X_n(u_{i+1})) \mid \mathscr{F}_{u_i}) - g_n(X_n(u_i))) \\
&\quad = E(\textstyle\int_{u_i}^{u_{i+1}} \tilde{A}_n g_n^2(X_n(v))\, dv \mid \mathscr{F}_{u_i}) \\
&\qquad - 2g_n(X_n(u_i))E(\textstyle\int_{u_i}^{u_{i+1}} \tilde{A}_n g_n(X_n(v))\, dv \mid \mathscr{F}_{u_i}) .
\end{aligned} \tag{4.38}
$$

Hence

$$
\begin{aligned}
&E((Z_n(t) - Z_n(s))^2 \mid \mathscr{F}_s) \\
&\quad = E(\textstyle\int_s^t \tilde{A}_n g_n^2(X_n(u))\, du \mid \mathscr{F}_s) \\
&\qquad - 2E(\textstyle\int_s^t g_n(X_n(u))\tilde{A}_n g_n(X_n(u))\, du \mid \mathscr{F}_s) \\
&\quad \leqq (t - s) \sup_x |\tilde{A}_n g_n^2(x) - 2g_n(x)\tilde{A}_n g_n(x)| \\
&\quad \equiv (t - s)M_n .
\end{aligned} \tag{4.39}
$$

The tightness of $\{Z_n(t)\}$ now follows by Theorem (4.20). Alternatively for $t_1 \leqq t \leqq t_2$

$$
E((Z_n(t_2) - Z_n(t))^2(Z_n(t) - Z_n(t_1))^2) \leqq M_n^2(t_2 - t)(t - t_1) . \tag{4.40}
$$

Since $\sup_n M_n < \infty$, this inequality implies $\{Z_n(t)\}$ is tight (see Billingsley [3], page 128).

**5. Applications and relationship to other work.** As was indicated in the introduction, the primary motivation behind the results in Section 3 was a desire to give a "semigroup" proof of the results of Borovkov and Gikhman. However, rather than reprove their general results using Theorem (3.11) we will consider two special results: Billingsley's Central Limit Theorem for martingales [2] and a result of Jagers [12] on diffusion approximations for age dependent branching processes. Each example illustrates how special properties of the processes involved can be used to verify the conditions of the general theorem. In addition we will apply Theorem (4.29) (actually its discrete parameter equivalent) to a sequence of infinite particle systems.

In order to be able to apply Theorem (3.11) one must know a good deal about the infinitesimal generator of the limiting semigroup. In particular one must know the form of the generator for a core, that is a subspace $D \subset \mathscr{D}(A)$ such that $A$ is the closure of $A|_D$.

For diffusion processes regularity theorems for partial differential equations frequently give information useful for finding cores. If $D$ is a subspace contained in $\mathscr{D}(A)$ and dense in $K$, and $T(t): D \to D$ then $D$ is a core. Consequently, if a regularity theorem implies $T(t): C^2 \to C^2$ (the bounded twice continuously differentiable functions) and $D = C^2 \cap \mathscr{D}(A)$ is dense (as it almost surely is) then $D$ is a core.

For example, let $T(t)$ be the semigroup on $\hat{C}(\mathbb{R}^n)$ corresponding to Brownian motion on $\mathbb{R}^n$. Then $\mathscr{D}(A) \supset C^2 \cap \hat{C}$ and $T(t): C^2 \cap \hat{C} \to C^2 \cap \hat{C}$. Hence $C^2 \cap \hat{C}$ is a core. (For $n = 1$ $\mathscr{D}(A) = C^2 \cap \hat{C}$ but this is not true for $n > 1$.)

Frequently even simpler subspaces are cores. For Brownian motion and many other diffusions $C_0^\infty$, the space of infinitely differentiable functions with compact support is a core.

(5.1) THEOREM (Billingsley). *Let* $\cdots Y_{-2}, Y_{-1}, Y_0, Y_1, Y_2 \cdots$ *be a stationary ergodic sequence with* $E(Y_k) = 0$ $E(Y_k^2) = \sigma^2$ *and suppose*

$$E(Y_{k+1} \mid Y_k, Y_{k-1}, \cdots) = 0 .$$

*Then* $X_n(t) = n^{-\frac{1}{2}} \sum_{k=1}^{[nt]} Y_k$ *converges weakly to Brownian motion* $B(t)$ *with* $E(B(t)) = 0$ *and* $\operatorname{Var}(B(t)) = \sigma^2 t$.

PROOF. There are two important properties to exploit: the fact that $X_n(t)$ is a martingale and the ergodicity of $\{Y_k\}$ which implies

$$\lim_{m\to\infty} E\left(\left|\frac{1}{m} \sum_{k=1}^{m} Y_k^2 - \sigma^2\right|\right) = 0 . \tag{5.2}$$

The elements in $\mathscr{D}(\hat{A})$ that we have at our disposal are of the form

$$f(t) = \frac{1}{\varepsilon} \int_0^\varepsilon E(g(X_n(t+s)) \mid \mathscr{F}_t)\, ds \tag{5.3}$$

where $\mathscr{F}_t = \sigma(Y_k : k \leqq [nt])$. If $g \in \hat{C}$ and $\varepsilon$ is small it is reasonable to expect that $f(t)$ is close to $g(X_n(t))$. Assuming $g \in C_0^2$ it is natural to attempt to study the limiting behavior of

$$\hat{\mathscr{A}} f(t) = \frac{1}{\varepsilon} E(g(X_n(t+\varepsilon)) - g(X_n(t)) \mid \mathscr{F}_t) \tag{5.4}$$

by expanding $g$ in a Taylor series about $X_n(t)$.

Doing this one obtains

$$\begin{aligned}
\hat{\mathscr{A}} f(t) = {} & \frac{1}{\varepsilon} E\left(\frac{1}{n^{\frac{1}{2}}} \sum_{k=[nt]+1}^{[n(t+\varepsilon)]} Y_k \mid \mathscr{F}_t\right) g'(X_n(t)) \\
& + \frac{1}{2n\varepsilon} E((\textstyle\sum_{k=[nt]+1}^{[n(t+\varepsilon)]} Y_k)^2 \mid \mathscr{F}_t) g''(X_n(t)) \\
& + \tfrac{1}{2} E(\textstyle\int_0^{X_n(t+\varepsilon)-X_n(t)} (X_n(t+\varepsilon) - X_n(t) - z)(g''(X_n(t)+z) \\
& - g''(X_n(t)))\, dz \mid \mathscr{F}_t) .
\end{aligned} \tag{5.5}$$

Using the Martingale property and defining $W_n$ to be the third term on the right this becomes

$$\hat{\mathscr{A}} f(t) = \tfrac{1}{2} E\left(\frac{1}{n\varepsilon} \sum_{k=[nt]+1}^{[n(t+\varepsilon)]} Y_k^2 \mid \mathscr{F}_t\right) g''(X_n(t)) + W_n . \tag{5.6}$$

If $n\varepsilon$ is large, (5.2) implies the conditional expectation multiplying $g''(X_n(t))$ is close to $\sigma^2$.

Using the stationarity the second term can be estimated by

$$E(|W_n|) \leqq \frac{1}{2\varepsilon} E((X_n(t+\varepsilon) - X_n(t))^2 \omega(|X_n(t+\varepsilon) - X_n(t)|)) \tag{5.7}$$
$$= \tfrac{1}{2} E\left(\frac{1}{m}(\textstyle\sum_{k=1}^m Y_k)^2 \omega\left(\frac{1}{n^{\frac{1}{2}}}|\textstyle\sum_{k=1}^m Y_k|\right)\right)$$

where $m = [n(t+\varepsilon)] - [nt]$ and $\omega(z) = \sup_x \sup_{0 \leqq y \leqq z} |g''(x+y) - g''(x)|$.

In order to prove convergence of the finite dimensional distributions using Theorem (3.11) we will produce a sequence $\{\varepsilon_n\}$ with $\lim_{n\to\infty} \varepsilon_n = 0$ such that $f_n(t)$ given by (5.3) satisfies

$$\sup_n \sup_t E(|f_n(t)|) < \infty\,, \tag{5.8}$$

$$\sup_n \sup_t E(|\hat{\mathscr{A}} f_n(t)|) < \infty\,, \tag{5.9}$$

$$\lim_{n\to\infty} E(|f_n(t) - g(X_n(t))|) = 0 \tag{5.10}$$

and

$$\lim_{n\to\infty} E(|\hat{\mathscr{A}} f_n(t) - \tfrac{1}{2}\sigma^2 g''(X_n(t))|) = 0\,. \tag{5.11}$$

Since for each fixed $m$ the right hand side of (5.7) goes to zero as $n$ goes to infinity, we may find a sequence $m_n$ with $\lim_{n\to\infty} m_n = \infty$ and

$$\lim_{n\to\infty} E\left(\frac{1}{m_n}(\textstyle\sum_{k=1}^{m_n} Y_k^2)\omega\left(\frac{1}{n^{\frac{1}{2}}}|\textstyle\sum_{k=1}^{m_n} Y_k|\right)\right) = 0\,. \tag{5.12}$$

We may assume $\lim_{n\to\infty} m_n/n = 0$, and we define $\varepsilon_n = m_n/n$. With this choice of $\varepsilon_n$ we have

$$\lim_{n\to\infty} E(|\hat{\mathscr{A}} f_n(t) - \tfrac{1}{2}\sigma^2 g''(X_n(t))|)$$
$$\leqq \lim_{n\to\infty} \tfrac{1}{2} \sup_x |g''(x)| E\left(\left|\frac{1}{m_n} \textstyle\sum_{k=[nt]+1}^{[nt+m_n]} Y_k^2 - \sigma^2\right|\right)$$
$$= 0\,.$$

The limit in (5.10) follows in much the same way and (5.8) and (5.9) are immediate since stationarity implies the expectations are independent of $t$.

Weak convergence follows from the results in Section 4 by noting

$$\begin{aligned} &E((X_n(t+u) - X_n(t))^2 | \mathscr{F}_t) \\ &\quad = E\left(\frac{1}{n} \textstyle\sum_{k=[nt]+1}^{[n(t+u)]} Y_k^2 | \mathscr{F}_t\right) \\ &\quad \leqq E\left(\sup_{l \leqq [nT]} \frac{1}{n} \textstyle\sum_{k=l+1}^{l+[n\delta]+1} Y_k^2 | \mathscr{F}_t\right) \\ &\quad \equiv E(\gamma_n(\delta) | \mathscr{F}_t)\,. \end{aligned} \tag{5.13}$$

We have

$$\lim_{n\to\infty} \frac{1}{n} \textstyle\sum_{k=1}^{[nt]} Y_k^2 = \sigma^2 t \tag{5.14}$$

almost surely and in $L_1$. Since the left hand side is increasing in $t$ convergence must be uniform on bounded $t$ intervals. This implies

$$\lim_{n\to\infty} E(\gamma_n(\delta)) = \sigma^2\delta$$

and hence the conditions of Theorem (4.20) are satisfied.

(5.15) THEOREM (Jagers). *Let $\{Z_n(t)\}$ be a sequence of age dependent branching processes with the same age distribution function $G(t)$, satisfying $G(0+) = 0$, and off spring distribution generating functions $f_n(z)$. Suppose*

$$\lambda = \textstyle\int_0^\infty t\,dG(t) < \infty\,, \tag{5.16}$$

$$\lim_{n\to\infty} n(f_n'(1) - 1) = \alpha\,, \tag{5.17}$$

$$\lim_{n\to\infty} f_n''(1) = \beta\,, \tag{5.18}$$

and

$$\sup_n f_n'''(1) < \infty\,. \tag{5.19}$$

*Let $a = \alpha/\lambda$ and $b = \beta/\lambda$.*

*Suppose $Z_n(0) = n$ and all initial particles are of age zero.*

*Define $X_n(t) = Z_n(nt)/n$. Then $X_n(t)$ converges weakly to the diffusion process $X(t)$ which is absorbing at zero, has $X(0) = 1$, and has generator $Af = \frac{1}{2}bxf'' + axf'$.*

PROOF. We may consider the semigroup for $X(t)$ on $\hat{C}(0, \infty)$, the space of continuous functions vanishing at zero and infinity. Then $\mathscr{D}(A) = \{f \in C^2 \cap \hat{C}: bxf'' + axf' \in \hat{C}\}$. A convenient core is the subspace of functions in $\mathscr{D}(A)$ that vanish for $x$ sufficiently large.

Let $m_n = f_n'(1)$ and $\beta_n = f_n''(1)$. The special properties of the branching processes we will need are properties of $E_s(Z_n(t))$ and $E_s(Z_n(t)(Z_n(t) - 1))$ where $E_s$ denotes the expectation under the assumption that the process begins with a single initial particle of age $s$. ($E$ without subscript, will denote the expectation under the original assumption of $n$ particles of age zero.) These properties may be obtained from renewal theory. Letting $M_n(t) = E_0(Z(t))$ and

$$B_n(t) = E_0(Z_n(t)(Z_n(t) - 1))$$

we have

$$M_n(t) = (1 - G(t)) + m_n \textstyle\int_0^t M_n(t-u)\,dG(u) \tag{5.20}$$

and

$$\begin{aligned} B_n(t) &= \beta_n \textstyle\int_0^t M_n(t-u)^2\,dG(u) + m_n \int_0^t B_n(t-u)\,dG(u) \\ &\equiv \xi_n(t) + m_n \textstyle\int_0^t B_n(t-u)\,dG(u)\,. \end{aligned} \tag{5.21}$$

(See [1], page 144). Letting $\tau_1, \tau_2, \cdots$ be independent random variables with distribution $G(t)$ and $Y(t) = \max\{k: \sum_{i=1}^k \tau_i \leqq t\}$ we may write

$$M_n(t) = E(m_n^{Y(t)}) \tag{5.22}$$

and

$$B_n(t) = E(\textstyle\sum_{k=1}^{Y(t)} m_n^k \xi_n(t - \sum_{i=1}^k \tau_i)). \tag{5.23}$$

From these identities it readily follows that

$$\lim_{n\to\infty} \sup_{t\leq T} |M_n(nt) - e^{at}| = 0 \tag{5.24}$$

and

$$\lim_{n\to\infty} \sup_{t\leq T} \left|\frac{B_n(nt)}{n} - \frac{\beta}{\alpha}(e^{2at} - e^{at})\right| = 0 \tag{5.25}$$

for every $T > 0$.

Furthermore

$$\sup_{t\leq T} |E_s(Z(nt)) - e^{at}| \equiv u_n(s) \tag{5.26}$$

and

$$\sup_{t\leq T} \left|\frac{1}{n} E_s(Z(nt)(Z(nt) - 1)) - \frac{\beta}{\alpha}(e^{2at} - e^{at})\right| \equiv v_n(s) \tag{5.27}$$

converge to zero as $n$ goes to infinity uniformly on bounded $s$ intervals.

For $f \in \mathscr{D}(A)$ and $f(x) = 0$ for $x$ sufficiently large

$$\begin{aligned}
&\frac{1}{\varepsilon} E(f(X_n(t+\varepsilon)) - f(X_n(t)) \,|\, \mathscr{F}_t) \\
&\quad = \frac{1}{\varepsilon} E(X_n(t+\varepsilon) - X_n(t) \,|\, \mathscr{F}_t) f'(X_n(t)) \\
&\qquad + \frac{1}{\varepsilon} E((X_n(t+\varepsilon) - X_n(t))^2 \,|\, \mathscr{F}_t) \tfrac{1}{2} f''(X_n(t)) \\
&\qquad + \frac{1}{\varepsilon} E(\textstyle\int_0^{X_n(t+\varepsilon)-X_n(t)} (X_n(t+\varepsilon) - X_n(t) - z) \\
&\qquad \times (f''(X_n(t)+z) - f''(X_n(t)))\, dz \,|\, \mathscr{F}_t)\,.
\end{aligned} \tag{5.28}$$

The branching property implies

$$\frac{1}{\varepsilon} E(X_n(t+\varepsilon) - X_n(t) \,|\, \mathscr{F}_t) = \frac{1}{n\varepsilon} \textstyle\sum_{i=1}^{Z_n(nt)} E_{s_i}(Z_n(n\varepsilon) - 1) \tag{5.29}$$

where $s_1, s_2, \cdots$ are the ages of the $Z_n(nt)$ particles. Consequently

$$\begin{aligned}
&E\left(\left|\frac{1}{\varepsilon} E(X_n(t+\varepsilon) - X_n(t) \,|\, \mathscr{F}_t) - X_n(t)\left(\frac{e^{a\varepsilon}-1}{\varepsilon}\right)\right|\right) \\
&\quad \leqq \frac{1}{\varepsilon} E\left(\frac{1}{n} \textstyle\sum_{i=1}^{Z_n(nt)} u_n(s_i)\right) = \frac{1}{\varepsilon} E_0(\textstyle\sum_{i=1}^{Z_n(nt)} u_n(s_i))\,.
\end{aligned} \tag{5.30}$$

Similarly

$$\begin{aligned}
&E\left(\left|\frac{1}{\varepsilon} E((X_n(t+\varepsilon) - X_n(t))^2 \,|\, \mathscr{F}_t) - X_n(t)\frac{\beta}{\alpha}\left(\frac{e^{2a\varepsilon} - e^{a\varepsilon}}{\varepsilon}\right)\right.\right. \\
&\qquad \left.\left. + \frac{1}{n} X_n(t)\left(\frac{e^{a\varepsilon}-1}{\varepsilon}\right) - \varepsilon X_n(t)\left(X_n(t) - \frac{1}{n}\right)\left(\frac{e^{a\varepsilon}-1}{\varepsilon}\right)^2\right|\right) \\
&\quad \leqq \frac{1}{\varepsilon} E\left(\frac{1}{n} \textstyle\sum_{i=1}^{Z_n(nt)} v_n(s_i)\right) + \frac{1}{\varepsilon n} E\left(\frac{1}{n} \textstyle\sum_{i=1}^{Z_n(nt)} u_n(s_i)\right) \\
&\qquad + \frac{1}{\varepsilon} E\left(\frac{1}{n^2} \textstyle\sum_{i\neq j} (u_n(s_i)u_n(s_j) + e^{a\varepsilon}(u_n(s_i) + u_n(s_j)))\right).
\end{aligned} \tag{5.31}$$

We note that $\gamma_n(t) \equiv E_0(\sum_{i=1}^{Z_n(nt)} u_n(s_i))$ satisfies

$$\gamma_n(t) = (1 - G(t))u_n(t) + m_n \int_0^t \gamma(t-u)\, dG(u) \tag{5.32}$$

and hence

$$\gamma_n(t) = E(m_n^{Y(t)} u_n(t - \textstyle\sum_{i=1}^{Y(t)} \tau_i)) \,. \tag{5.33}$$

Since the $u_n(s)$ are uniformly bounded and converge to zero uniformly on compact sets, $\lim_{n\to\infty} \gamma_n(t) = 0$. Similarly the right hand side of (5.31) converges to zero for every $\varepsilon > 0$. Therefore there is a sequence $\{\varepsilon_n\}$ with $\lim_{n\to\infty} \varepsilon_n = 0$ such that the right hand sides of (5.30) and (5.31) go to zero with $\varepsilon$ replaced by $\varepsilon_n$.

The third term on the right of (5.28) can be shown to be negligible by estimating conditional third moments in a manner similar to the above estimation of first and second moments.

Letting

$$g_n(t) = \frac{1}{\varepsilon_n} \int_0^{\varepsilon_n} E(f(X_n(t+s)))\, ds$$

we conclude from (5.30) and (5.31) that

$$\lim_{n\to\infty} E(|\hat{\mathscr{A}} g_n(t) - (\tfrac{1}{2} b X_n(t) f''(X_n(t)) + a X_n(t) f'(X_n(t)))|) = 0 \,,$$

and convergence of the finite dimensional distributions follows by Theorem (3.11).

To verify weak convergence we verify the conditions of Theorem (4.20) for $g(X_n(t))$ where $g$ is continuously differentiable with $g(x) = 0$ for $x > M > 0$. Suppose $|g(x)|$, $|g'(x)| \leqq C$. Then

$$\begin{aligned} &E((g(X_n(t+u)) - g(X_n(t)))^2 \mid \mathscr{F}_t) \\ &\quad \leqq \chi_{[0,2M]}(X_n(t)) C^2 E((X_n(t+u) - X_n(t))^2 \mid \mathscr{F}_t) \\ &\qquad + \chi_{(2M,\infty)}(X_n(t)) C^2 P\{X_n(t+u) < M \mid \mathscr{F}_t\} \,. \end{aligned} \tag{5.34}$$

Now

$$\begin{aligned} &E((X_n(t+u) - X_n(t))^2 \mid \mathscr{F}_t) \\ &\quad = \frac{1}{n^2} (\textstyle\sum_i E_{s_i}((Z_n(nu) - 1)^2) + \sum_{i \neq j} E_{s_i}(Z_n(nu) - 1) E_{s_j}(Z_n(nu) - 1)) \\ &\quad \leqq X_n(t)^2 \left[ \frac{1}{n} \sup_s E_s((Z_n(nu) - 1)^2) + \sup_s |E_s(Z_n(nu) - 1)|^2 \right] . \end{aligned} \tag{5.35}$$

Noting that

$$\begin{aligned} &\chi_{(2M,\infty)}(X_n(t)) P\{X_n(t+u) < M \mid \mathscr{F}_t\} \\ &\quad \leqq \chi_{(2M,\infty)}(X_n(t)) P\{(X_n(t+u) - X_n(t))^2 > (X_n(t) - M)^2 \mid \mathscr{F}_t\} \\ &\quad \leqq \chi_{(2M,\infty)}(X_n(t)) \frac{E((X_n(t+u) - X_n(t))^2 \mid \mathscr{F}_t)}{(X_n(t) - M)^2} \,, \end{aligned} \tag{5.36}$$

we have the left hand side of (5.34) bounded by

$$\gamma_n(\delta) = \sup_{u \leq \delta} 4C^2(M^2 + 1)\left[\frac{1}{n}\sup_s E_s((Z_n(nu) - 1)^2) + \sup_s |E_s(Z_n(nu) - 1)|^2\right].$$

Using (5.24) and (5.25) it can be shown that

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \sup_{u \leq \delta} \sup_s \frac{1}{n} E_s((Z_n(nu) - 1)^2) = 0$$

and

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \sup_{u \leq \delta} \sup_s |E_s(Z_n(nu) - 1)| = 0 .$$

Hence the conditions of Theorem (4.20) are satisfied.

We now consider a discrete parameter approximation to the simple exclusion model of an infinite particle system. (See Spitzer [29].) The basis for our result is the existence theorem of Liggett [20] that essentially characterizes the generator for the process. The model in which we are interested is one in which each of the particles is independently undergoing a Markov chain on a countable state space $S$ except that if a particle attempts a transition into an occupied state the transition is not made.

The state space for the infinite particle process can be thought of as the collection of functions $E = \{\eta(x) : \eta : S \to [0, 1]\}$ where $\eta(x) = 1$ means $x$ is occupied and $\eta(x) = 0$ means $x$ is unoccupied. Suppose $\alpha(x) > 0$ and $\sum_{x \in S} \alpha(x) < \infty$. Then $\rho(\eta, \tau) = \sum_{x \in S} \alpha(x)|\eta(x) - \tau(x)|$ defines a metric on $E$ under which $E$ is compact.

We will assume that the exponential waiting times in each state have parameters 1 and that at the end of the waiting time the probability of a particle in $x$ attempting a transition to $y$ is given by $p(x, y)$. We will also assume $p(x, x) = 0$.

We modify Liggett's approach somewhat in defining

$$\begin{aligned} \eta_{u,v}(x) &= \eta(x) && x \neq u, v \\ &= \eta(u) \wedge \eta(v) && x = u \\ &= \eta(u) \vee \eta(v) && x = v . \end{aligned}$$

Note that $\eta_{u,v}$ is the state obtained from $\eta$ if a transition is attempted from $u$ to $v$.

(5.37) THEOREM (Liggett). *Let $D$ be the collection of functions on $E$ that depend only on $\eta(x)$ for $x$ in a finite subset of $S$. (Note $D$ is dense in $C(E)$.) For $f \in D$ define*

(5.38) $$Af(\eta) = \sum_{x,y} p(x, y)(f(\eta_{x,y}) - f(\eta)) .$$

*Suppose*

(5.39) $$\sup_y \sum_x p(x, y) < \infty .$$

*Then the closure of $A$ generates a positive, strongly continuous, contraction semigroup on $C(E)$.*

Let $p_n(x, y) = (1 - n^{-1})\delta_{x,y} + n^{-1}p(x, y)$. We consider a sequence of discrete parameter processes $\{Y_n(k)\}$ with the following properties: at each time step all of the particles attempt transitions according to the transition probabilities $p_n(x, y)$. Transitions are suppressed to currently occupied states. It is possible for more than one particle to jump to an unoccupied state, but if this occurs we will assume the particles coalesce into a single particle.

For $\theta = \{(x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)\}$ with $x_i \neq x_j$ for $i \neq j$ define

$$\eta_\theta(z) = \eta(z) \min_i \eta_{x_i y_i}(z) + (1 - \eta(z)) \max_i \eta_{x_i y_i}(z) \tag{5.40}$$

and let $\theta_x = \{x_1, x_2, \cdots, x_m\}$.

For $f \in D$ let $\Gamma$ be the subset of $S$ upon which $f$ depends. Then

$$\begin{aligned} T_n f(\eta) &= E_\eta f(Y_n(1)) \\ &= \textstyle\sum_\theta f(\eta_\theta) \prod_i p_n(x_i, y_i) \prod_{z \in S-\Gamma-\theta_x} (1 - \sum_{\omega \in \Gamma} p_n(z, \omega)) \\ &\qquad \times \textstyle\prod_{z \in \Gamma - \theta_x} p_n(z, z) \\ &= \textstyle\sum_\theta f(\eta_\theta) \dfrac{1}{n^m} \prod_i p(x_i, y_i) \prod_{z \in S-\Gamma-\theta_x} \left(1 - \dfrac{1}{n} \sum_{\omega \in \Gamma} p(z, \omega)\right) \\ &\qquad \times \textstyle\prod_{z \in \Gamma - \theta_x} \left(1 - \dfrac{1}{n}\right) \end{aligned} \tag{5.41}$$

where the summation is over all $\theta = \{(x_1, y_1) \cdots (x_m, y_m)\}$ such that $x_i \neq x_j$ for $i \neq j$, $x_i \neq y_i$ and if $x_i \notin \Gamma$ then $y_i \in \Gamma$.

(5.42) THEOREM. *Let $Y_n(k)$ be defined as above, let $X_n(t) = Y_n([nt])$ and let $X(t)$ be the process corresponding to $A$ in* (5.38). *If $X_n(0) = X(0)$ then $X_n(t)$ converges weakly to $X(t)$ as $n$ goes to infinity.*

PROOF. A discrete parameter analog of Theorem (4.29) can be obtained in which the role of the infinitesimal operator is played by

$$A_n f(\eta) = n(T_n f(\eta) - f(\eta)) . \tag{5.43}$$

Alternatively, $A_n$ generates a continuous parameter Markov process and the results in [17] imply $X_n(t)$ converges weakly to $X(t)$ if and only if the corresponding continuous parameter processes converge weakly. Consequently, since $D$ is a core for $A$ we need only prove that

$$\lim_{n\to\infty} \sup_\eta |A_n f(\eta) - Af(\eta)| = 0 \tag{5.44}$$

for all $f \in D$.

But

$$\begin{aligned} |A_n f(\eta) - Af(\eta)| \leq & \left| \textstyle\sum_{x \notin \Gamma} \sum_{y \in \Gamma} (f(\eta_{xy}) - f(\eta)) p(x, y) \right. \\ & \left. \times \left( \textstyle\prod_{z \in S-\Gamma-x} \left(1 - \dfrac{1}{n} p(z, x)\right) - 1\right) \right| \\ & + \left| \textstyle\sum_{x \in \Gamma} \sum_{y \in S} (f(\eta_{xy}) - f(\eta)) p(x, y) \right. \\ & \left. \times \left( \textstyle\prod_{z \in S-\Gamma} \left(1 - \dfrac{1}{n} p(x, y)\right) \left(1 - \dfrac{1}{n}\right) - 1\right) \right| \\ & + 2 \sup_\eta |f(\eta)| \textstyle\sum_{\theta, m>1} \dfrac{1}{n^{m-1}} \prod_i p(x_i, y_i) . \end{aligned} \tag{5.45}$$

Let $M = \sup_x \sum_z p(z, x)$. The first two terms on the right of (5.45) go to zero since

$$\prod_{z\in S-\Gamma}\left(1 - \frac{1}{n} p(z, x)\right) \geqq e^{-M/n},$$

and the third term goes to zero since the coefficient of $1/n^{m-1}$ is

$$(5.46)\qquad \begin{aligned} &\sum_{y_1\in\Gamma} \cdots \sum_{y_m\in\Gamma} \prod_{i=1}^{m} \sum_{x\notin\Gamma} p(x, y_i) \\ &\quad + \sum_{x_1\in\Gamma} \cdots \sum_{x_m\in\Gamma} \prod_{i=1}^{m} \sum_{y\in S} p(x_i, y) \\ &\quad \leqq (\#\Gamma)^m M^m + (\#\Gamma)^m . \end{aligned}$$

There are of course a variety of other approaches for proving convergence to Markov processes. There is a vast literature on the central limit theorem for dependent random variables and for many of these results there is a corresponding invariance principle, that is convergence of the partial sum process to Brownian motion.

Rosen [26] uses characteristic functions to give convergence of partial sum processes to diffusions with time dependent generators of the form $a(t)f'' + b(t)xf'$. A similar approach is used in [16] and by Norman in [23, 24, 25] to give a central limit theorem for the deviation of certain Markov chain models from deterministic models given by ordinary differential equations. In [25], Norman obtains convergence of the one dimensional distributions that is uniform for all time not just for bounded time intervals. Stone [30] proves convergence theorems for sequences of one dimensional diffusions and birth and death processes using arguments involving local times. Rosenkrantz [27] obtains similar theorems using semigroups, but by considering the resolvents rather than the infinitesimal operators. Stroock and Varadhan [31] give very general conditions for the weak convergence of Markov chains to diffusions using their characterization of diffusion processes as solutions of a "martingale problem." Conditions for convergence of Markov chains to diffusions can be obtained using the stochastic integral representation of the diffusion. (See Gikhman and Skorokhod [11] page 459 and Kushner [18].)

The semigroup approach seems to give a unified method for a wide class of problems. However, it is frequently the case that special properties of the processes involved suggest special techniques (for example, Jagers' use of generating functions to obtain the result discussed above) that are either more direct or are in some sense more appropriate to the problem.

**6. Appendix.** The following lemma is essentially contained in Doob [7] page 358. It is not necessary to assume that the martingale is separable. A process $X(t, \omega)$ is progressively measurable with respect to $\{\mathscr{F}_t\}$ if $X(\cdot, \cdot) : [0, t] \times \Omega \to E$ is $\mathscr{B}([0, t]) \times \mathscr{F}_t$ measurable for every $t \geqq 0$. ($\mathscr{B}([0, t])$ is the Borel subsets of $[0, t]$.)

(6.1) LEMMA. *Every martingale has a version that is right continuous except at a countable set of points, and hence is progressively measurable.*

(6.2) THEOREM. *Let* $(\Omega, \mathscr{F}, P)$ *be a probability space and let* $\{\mathscr{F}_t\}$ *be an increasing family of* $\sigma$*-algebras in* $\mathscr{F}$. *If* $X(t)$ *is progressively measurable and* $E(|X(t)|) < \infty$ *for every* $t$, *then for every* $s \geqq 0$ $Y(t) \equiv E(X(t+s) \mid \mathscr{F}_t)$ *has a progressively measurable version.*

PROOF. Let $\mathscr{H}$ be the linear space of progressively measurable processes with finite mean such that $X(t) \in \mathscr{H}$ implies $Y(t) \equiv E(X(t+s) \mid \mathscr{F}_t)$ has a progressively measurable version for every $s \geqq 0$. Observe that $\mathscr{H}$ is closed under limits of monotone increasing sequences (i.e., $X_{n+1}(t) \geqq X_n(t)$ a.s. for all $t$) provided $E(\lim_{n\to\infty} X_n(t)) < \infty$ for all $t$. Consider a process of the form

$$X(t, \omega) = \sum_{i=1}^n c_i \mathscr{X}_{A_i}(t)\mathscr{X}_{B_i}(\omega)$$

where $A_i \in \mathscr{B}([0, \infty))$, $B_i \in \mathscr{F}$ and $t \in A_i$ implies $B_i \in \mathscr{F}_t$. Then

$$E(X(t+s) \mid \mathscr{F}_t) = \sum_{i=1}^n c_i \mathscr{X}_{A_i}(t+s)E(\mathscr{X}_{B_i} \mid \mathscr{F}_t)$$

has a progressively measurable version since $E(\mathscr{X}_{B_i} \mid \mathscr{F}_t)$ is a martingale and $X(t) \in \mathscr{H}$. The Monotone Class Theorem for functions (see Blumenthal and Getoor [5] page 5) implies $\mathscr{H}$ contains all progressively measurable processes with finite mean.

**Acknowledgments.** This work was completed while the author was a Visiting Fellow in the Statistics Department of the Institute for Advanced Studies at the Australian National University. The hospitality of that department was greatly appreciated.

REFERENCES

[1] ATHREYA, K. B. and Ney, P. E. (1972). *Branching Processes.* Springer-Verlag, Berlin.

[2] BILLINGSLEY, PATRICK (1961). The Lindeberg-Lévy theorem for martingales. *Proc. Amer. Math. Soc.* **12** 788-792.

[3] BILLINGSLEY, PATRICK (1968). *Convergence of Probability Measures.* Wiley, New York.

[4] BILLINGSLEY, PATRICK (1974). Conditional distributions and tightness. *Ann. Probability* **2** 480-485.

[5] BLUMENTHAL, R. M. and GETOOR, R. K. (1968). *Markov Processes and Potential Theory.* Academic Press, New York.

[6] BOROVKOV, A. A. (1970). Theorems on the convergence to Markov diffusion processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **16** 47-76.

[7] DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[8] DYNKIN, E. B. (1965). *Markov Processes.* Springer-Verlag, Berlin.

[9] FÖLLMER, HANS (1973). On the representation of semimartingales. *Ann. Probability* **1** 580-589.

[10] GIKHMAN, I. I. (1969). Convergence to Markov processes. *Ukrainian Math. J.* **21** 263-270.

[11] GIKHMAN, I. I. and SKOROKHOD, A. V. (1969). *Introduction to the Theory of Random Processes.* W. B. Saunders, Philadelphia.

[12] JAGERS, PETER (1971). Diffusion approximations of branching processes. *Ann. Math. Statist.* **42** 2074-2078.

[13] KURTZ, THOMAS G. (1969). Extensions of Trotter's operator semigroup approximation theorems. *J. Functional Analysis* **3** 111-132.

[14] KURTZ, THOMAS G. (1970). A general theorem on the convergence of operator semigroups. *Trans. Amer. Math. Soc.* **148** 23-32.

[15] KURTZ, THOMAS G. (1970). Solutions of ordinary differential equations as limits of pure jump Markov processes. *J. Appl. Probability* **7** 49–58.

[16] KURTZ, THOMAS G. (1971). Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J. Appl. Probability* **8** 344–356.

[17] KURTZ, THOMAS G. (1971). Comparison of semi-markov and Markov processes. *Ann. Math. Statist.* **42** 991-1002.

[18] KUSHNER, HAROLD J. (1974). On the weak convergence of interpolated Markov chains to a diffusion. *Ann. Probability* **2** 40–50.

[19] LIGGETT, THOMAS M. (1970). On convergent diffusions: The densities and the conditioned processes. *Indiana Univ. Math. J.* **20** 265–279.

[20] LIGGETT, THOMAS M. (1972). Existence theorems for infinite particle systems. *Trans. Amer. Math. Soc.* **165** 471-481.

[21] LINDVALL, TORGNY (1973). Weak convergence of probability measures and random functions in the function space $D[0, \infty)$. *J. Appl. Probability* **10** 109-121.

[22] MEYER, PAUL A. (1966). *Probability and Potentials.* Blaisdell, Waltham, Mass.

[23] NORMAN, M. FRANK (1968). Slow learning. *British J. Math. Statist. Psychology* **21** 141-159.

[24] NORMAN, M. FRANK (1972). *Markov Processes and Learning Models.* Academic Press, New York.

[25] NORMAN, M. FRANK (1974). A central limit theorem for Markov processes that move by small steps. *Ann. Probability* **2** 1065–1074.

[26] ROSEN, B. (1967). On the central limit theorem for sums of dependent random variables. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **7** 48–82.

[27] ROSENKRANTZ, WALTER A. (1975). Limit theorems for solutions of a class of stochastic differential equations. *Indiana Univ. Math. J.* **24** 613–625.

[28] SKOROKHOD, A. V. (1958). Limit theorems for Markov processes. *Theor. Probability Appl.* **3** 202–246.

[29] SPITZER, FRANK (1970). Interaction of Markov processes. *Advances in Math.* **5** 246–290.

[30] STONE, CHARLES (1963). Limit theorems for random walks, birth and death processes, and diffusion processes. *Illinois J. Math.* **7** 638–660.

[31] STROOCK, DANIEL W. and VARADHAN, S. R. S. (1967). Diffusion processes with continuous coefficients, II. *Comm. Pure Appl. Math.* **22** 479–530.

[32] TROTTER, H. F. (1958). Approximation of semigroups of operators. *Pacific J. Math.* **8** 887–919.

[33] WATANABE, SHINZO (1968). A limit theorem of branching processes and continuous state branching processes. *J. Math. Kyoto Univ.* **8** 141–167.

[34] WATANABE, SHINZO and WATANABE, TOITSU (1970). Convergence of isotropic scattering transport process to Brownian motion. *Nagoya Math. J.* **40** 161–171.

[35] WATANABE, TOITSU (1969). Convergence of transport process to diffusion. *Proc. Japan Acad.* **45** 470–472.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WISCONSIN, MADISON
MADISON, WISCONSIN 53706

**Added in proof.** Proposition (1.12) is essentially a special case of a result in Airault, Hélène, and Hans Föllmer (1974). Relative densities of semimartingales. *Invent. Math.* **27** 299–327.