The Annals of Probabi'ity
1981, Vol. 9, No. 3, 429-452

MINIMAL CONDITIONS FOR WEAK CONVERGENCE TO A DIFFUSION PROCESS ON THE LINE

BY INGE S. HELLAND

Agricultural University of Norway

By transforming a central limit theorem for dependent variables, we find conditions for a sequence of processes with paths in $D[0, \infty)$ to converge weakly to a diffusion process. Of the most important conditions, the first is related to (but weaker than) tightness, and in the next two we require that the first two conditional moments, given the past, of truncated increments in small time intervals, should stay close to the appropriate infinitesimal coefficients of the limiting diffusion times the length of the time interval. The limiting diffusions can have inaccessible or exit boundaries. We prove that our conditions are necessary and sufficient in order that: (1) the sequence of processes converges weakly in $D[0, \infty)$; (2) any finite number of conditional expectations given the past of bounded, continuous functionals of the processes converge jointly in distribution to the "correct" value.

1. Introduction. For each $n \geq 1$, let $X_n(t)$ be a real-valued random process, right-continuous on $[0, \infty)$, with left-hand limits on $(0, \infty)$, and adapted to an increasing family of σ-fields $\{\mathscr{F}_n(t); t \geq 0\}$. Our object is to give conditions under which $\{X_n(t)\}$ converges as $n \to \infty$ to a diffusion process $X(t)$ on the line. The mode of convergence will be weak convergence in the space $D[0, \infty)$ of right-continuous functions with left-hand limits, endowed with the Stone topology ([21], [32]).

This problem, and its generalization to vector-valued processes, has been attacked earlier by several authors, first under the assumption that each $X_n(t)$ is a Markov process ([10], [20], [25], [30], [34]); later it has emerged that essentially the same conditions suffice in the general case ([3], [17], [24], [26], [27]). The conditions given by different authors are of the same general nature: convergence of the first two moments of small increments given the past, and some additional conditions related to tightness and boundedness. However, there are great variations in details, and this leads one to suspect that some of the assumptions are unnecessarily strong.

Out of the many different techniques that have been used in deriving these limit theorems, two general classes of techniques should be mentioned. One is the semigroup-approach which has been developed in great generality by Kurtz [17]. (We will come back to the relationship between the results of [17] and those of the present paper later.) The other is the broad class of techniques that can be called martingale methods. Characterization of diffusion processes by means of certain martingales, is the basis for Stroock and Varadhan [33], [34]. A systematic development of martingale methods applied to weak limit theorems is given by Rebolledo [27].

The technique used in the present paper is different. Basically, we transform a functional central limit theorem for dependent variables via a random time change and a scale change. The central limit theorem that we use is of the "near martingale" type ([6], [7], [23]). The paper is limited in that we only consider one-dimensional, time-homogeneous diffusions as possible limit processes. But this limitation has enabled us to prove convergence under conditions that are both sufficient and necessary—in a sense to be made

Received May 8, 1979; revised October 1979.
AMS 1970 *subject classifications.* Primary 60F05; secondary 60J60, 60B10.
Key words and phrases. Weak convergence to diffusion process, random time change, continuous mapping theorem, minimal conditions for convergence.

precise in Section 3 below. Also, our results are valid without any Lipschitz condition on the diffusion coefficients. We assume in this paper that the limiting diffusion is without regular boundaries. Generalization of our results to include regular boundaries will be given elsewhere.

The paper is based on the well-known result that any one-dimensional, time-homogeneous diffusion process can be represented as a process derived from Brownian motion by a simple random time change combined with a scale change ([4], [22]). This result does not generalize to vector-valued processes. For processes that are inhomogeneous in time, the corresponding representation is somewhat more complicated. A recent representation theorem by Kurtz [19] gives some hope that the technique used here could be generalized to these cases, but then further technical difficulties have to be resolved. Special cases, such as convergence to one-dimensional Gaussian diffusions, are easily handled, however.

The plan of the paper is as follows: in Section 2 and Section 3 we formulate our conditions for convergence and give the main theorems. In Section 4 we formulate the two main results on which the proofs are based, and in Sections 5, 6 and 7, we carry out the proofs of the main theorems. In Section 8 we consider two examples of sequences where our conditions for convergence to diffusions hold.

One aspect of our method is that we avoid the concept of tightness. The approach is constructive in the sense that once one has a modulus of convergence for the central limit theorem for sums of dependent variables, one should in principle be able to find a modulus of convergence for our limit theorems, by retracing the steps in the proofs of the present paper and in those of Helland [12].

2. Conditions for convergence. Following Breiman [4], we define a diffusion process $X(t)$ to be a continuous, strong Markov process on an interval F (bounded or unbounded, closed at none, one or both boundaries) such that the following holds: there exist functions $\sigma^2(x)$, $\mu(x)$ defined and continuous on $G = \text{int } F$, $\sigma^2(x) > 0$ on G, such that for all $\epsilon > 0$

$$t^{-1}P^x[|X(t) - x| > \epsilon] \to_{\text{bp}} 0 \tag{2.1}$$

$$t^{-1}E^x(X(t) - x; \quad |X(t) - x| \le \epsilon) \to_{\text{bp}} \mu(x) \tag{2.2}$$

$$t^{-1}E^x((X(t) - x)^2; \quad |X(t) - x| \le \epsilon) \to_{\text{bp}} \sigma^2(x) \tag{2.3}$$

as $t \to 0$, where the convergence ($\to_{\text{bp}}$) is bounded pointwise on all compact intervals $K \subset G$. (In fact, one can show that the convergence in (2.1)–(2.3) is always uniform on compacts in G, cf., Problem 1, page 68 in Mandl [22] and the proof of Lemma 3.2, page 149 in Norman [25].) In addition to the drift coefficient $\mu(x)$ and the diffusion coefficient $\sigma^2(x)$, we have to specify boundary conditions at boundary points of F contained in F. This will be discussed below.

Let $G = (r_0, r_1)$ and let $c \in (r_0, r_1)$. We can define a scale function $u(x)$ by

$$u(c) = 0, \qquad u'(x) = \exp\left\{-\int_c^x 2\mu(y)\sigma^{-2}(y)\, dy\right\}. \tag{2.4}$$

It is well known and easy to see that $u(X(t))$ is a diffusion on natural scale, i.e., with vanishing drift coefficient. The speed measure $m(dx)$ is defined by

$$m(dx) = \sigma^{-2}(x) \exp\left\{\int_c^x 2\mu(y)\sigma^{-2}(y)\, dy\right\} dx, \tag{2.5}$$

and we put $m(x) = \int_c^x m(dx)$.

The boundary point r_i is called accessible if $R_i \equiv \inf\{t \ge 0 : X(t) = r_i\}$ is finite with positive P^x-probability for some (hence all) $x \in G$. A necessary and sufficient condition for r_1 to be accessible is given by

(2.6) $$\int_c^{r_1} m(x)u'(x)\,dx < \infty.$$

(See Breimann [4] or Mandl [22]). The condition for r_0 to be accessible is similar. Without loss of generality we can—and will—assume that F is the union of G and the accessible boundary points. Also we assume that all accessible boundary points, if any, are finite.

It is always true for an accessible boundary point r_i that $|u(r_i)| < \infty$. The boundary is called regular if in addition $|m(r_i)| < \infty$, otherwise it is called exit boundary. If r_i is an exit boundary, then the only boundary condition that gives a conservative, continuous Markov process $X(t)$, is the one where the process is stopped at r_i: on $[R_i < \infty]$ we have $X(t) = X(t \wedge R_i)$. This is well known, but can also be deduced from our results below. We will only work with conservative, continuous diffusion processes in this paper.

Thus we only need to specify boundary conditions for $X(t)$ at regular boundary points, if any. When giving conditions for $X_n(t)$ to converge to $X(t)$, we have to take these boundary conditions into consideration. This can be done, and the results will be discussed in a future publication. In the present paper, however, we will make the following crucial assumption: *the diffusion process $X(t)$ has no regular boundary points.*

Now let $\{X_n(t)\}$ and $\{\mathscr{F}_n(t)\}$ be as in the introduction. We can think of $\mathscr{F}_n(t)$ as $\sigma\{X_n(s); s \le t\}$, but other choices are possible. For each fixed n let $\{t_n^k;\ k = 0, 1, 2, \cdots\}$ be a sequence of stopping times relative to $\{\mathscr{F}_n(t);\ t \ge 0\}$ satisfying

(2.7) $$0 = t_n^0 \le t_n^1 \le \cdots;\ \lim_{k\to\infty} t_n^k = +\infty \text{ a.s.}$$

Also we assume that for each $t > 0$

(2.8) $$\max_{0\le k\le r_n(t)} \Delta t_n(k) \to 0 \quad \text{in probability as} \quad n \to \infty,$$

where we define

$$r_n(t) = \max\{k \ge 0 : t_n^k \le t\}; \qquad \Delta t_n(k) = t_n^{k+1} - t_n^k.$$

We will call $T = \{t_n^k;\ k = 0, 1, 2, \cdots\}_n$ a sequence of partitions. We are free to choose this sequence in the conditions below, but it will be shown later that if these conditions are satisfied for one choice of T, they will always be satisfied for a sequence of partitions of the form $\{k\delta_n;\ k = 0, 1, 2, \cdots\}_n$, where $\delta_n \downarrow 0$.

Define

(2.9) $$\Delta X_n(k) = X_n(t_n^{k+1}) - X_n(t_n^k)$$
$$\Delta_\epsilon X_n(k) = \Delta X_n(k) \cdot I(|\Delta X_n(k)| \le \epsilon),$$

where $I(\cdot)$ is the indicator function. Also we define

(2.10) $$VX_n(k) = \sup_{t_n^k \le s \le t_n^{k+1}} |X_n(s) - X_n(t_n^k)|.$$

We will let $\Rightarrow$ denote weak convergence in $D[0, \infty)$ with the Stone topology and $\to_p$ denote convergence in probability. From now on all convergence will be as $n \to \infty$, unless stated otherwise.

Let $X(t)$ be a diffusion process on F with given coefficients $\mu(x)$ and $\sigma^2(x)$, and given initial distribution. We assume that $X(0) \in F$ a.s. Let K be a closed set in F. Consider the following set of conditions.

A0 $\quad X_n(0) \to X(0) \quad$ in distribution.

A1 $\quad \sum_{k=0}^{r_n(t)} P[VX_n(k) > \epsilon \mid \mathscr{F}_n(t_n^k)] \to_p 0 \quad$ for all $\epsilon, t > 0$.

A1(K) $\quad \sum_{k=0}^{r_n(t)} P[VX_n(k) > \epsilon \mid \mathscr{F}_n(t_n^k)] I(X_n(t_n^k) \in K) \to_p 0 \quad$ for all $\epsilon, t > 0$.

A2(K) $\quad \sum_{k=0}^{r_n(t)} |E(\Delta_1 X_n(k) \mid \mathscr{F}_n(t_n^k)) - \mu(X_n(t_n^k))\Delta t_n(k)| I(X_n(t_n^k) \in K) \to_p 0$ for all $t > 0$.

A3(K) $\sum_{k=0}^{r_n(t)} | E(\{\Delta_1 X_n(k)\}^2 | \mathscr{F}_n(t_n^k)) - \sigma^2(X_n(t_n^k))\Delta t_n(k) | I(X_n(t_n^k) \in K) \to_p 0$ for all $t > 0$.

A4(0) $P[\inf_{0\le s\le t} X_n(s) < r_0 - \epsilon] \to 0$ for all $\epsilon, t > 0$.

A4(1) $P[\sup_{0\le s\le t} X_n(s) > r_1 + \epsilon] \to 0$ for all $\epsilon, t > 0$.

The first main result of this paper is the following

THEOREM 2.1. *Assume* A0 *and* A1. *Assume that* A2(K) *and* A3(K) *hold for all compact intervals K in G, and that* A4(i) *holds whenever r_i is accessible (hence exit) for $X(t)$. Then $X_n(t) \Rightarrow X(t)$.*

The single condition A1 *may be replaced by the condition that* A1(K) *should hold for all compact intervals K in some open set containing F.*

The proof of this Theorem will be completed in Section 6. Here we only insert the following remarks concerning the conditions.

1. As discussed in Helland [13], A1 is equivalent to the condition that

$$\max_{0\le k\le r_n(t)} VX_n(k) \to_p 0 \tag{2.11}$$

for all $t > 0$. In particular, A1 is independent of the choice of the family of σ-fields $\{\mathscr{F}_n(t)\}$, as long as $X_n(t)$ is adapted to this family. Also A1 is implied by the familiar tightness condition (see Billingsley [2], Theorem 15.5):

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty} P[\sup_{0\le s,u\le t; |s-u|\le\delta} |X_n(s) - X_n(u)| > \epsilon] \to 0 \tag{2.12}$$

for all $\epsilon, t > 0$. However, (2.12) is strictly stronger than A1; in fact, it is equivalent to the condition that A1 should hold for all sequences of partitions $\{t_n^k\}$ satisfying (2.7) and (2.8) (see [13]).

2. The condition that A2(K) and A3(K) should hold for all compact intervals K in G may often be replaced by the simpler, but slightly stronger conditions:

A2 $\sum_{k=0}^{r_n(t)} | E(\Delta_1 X_n(k) | \mathscr{F}_n(t_n^k)) - \mu(X_n(t_n^k))\Delta t_n(k) | \to_p 0$ for all $t > 0$.

A3 $\sum_{k=0}^{r_n(t)} | E(\{\Delta_1 X_n(k)\}^2 | \mathscr{F}_n(t_n^k)) - \sigma^2(X_n(t_n^k))\Delta t_n(k) | \to_p 0$ for all $t > 0$.

For instance, one can show (see Section 7) that A2 and A3—if meaningful—are always satisfied for some sequence of partitions when the other conditions of Theorem 2.1 hold and the diffusion process $X(t)$ is without accessible boundaries. For A2 and A3 to be meaningful, we may require: either $X_n(t_n^k) \in G$ a.s. for all k and n, or $X_n(t_n^k) \in F$ a.s. for all k and n and $\mu(\cdot)$ and $\sigma^2(\cdot)$ are continuous throughout F. (Note that to begin with $\mu(\cdot)$ and $\sigma^2(\cdot)$ were only defined on $G = \operatorname{int} F$.)

For a simple example where the set of conditions A2(K) can not be replaced by the condition A2, look at the following: Let $X(t)$ be a diffusion process on $F = [0, \infty)$ with $\mu(x) = -1$, $\sigma^2(x) = x$. For this process $x = 0$ is an exit boundary. Put $X_n(t) = X(t)$ ($n = 1, 2, \cdots$). Then of course $X_n(t) \Rightarrow X(t)$, and A2(K) holds for all compacts $K \subset G = (0, \infty)$ and for any sequence of partitions $\{t_n^k\}$ satisfying (2.7)–(2.8). This follows easily, since the convergence in (2.2) is uniform on compacts in G. However, A2 does not hold for any sequence of partitions, since $R_0 = \inf\{s \ge 0 : X(s) = 0\} < t$ with positive probability for any $t > 0$, and $E^0(X(\delta); |X(\delta)| \le 1) \doteq 0$. The same example shows that the assumption that A2(K) holds for all compact intervals $K \subset F$, will be too strong in general.

3. The conditions A4(i) ($i = 0, 1$) are trivially satisfied when $X_n(t) \in F$ a.s. for all $t \ge 0$. In general A4(i) is necessary for convergence also when r_i is inaccessible, but then it is automatic from the other conditions of Theorem 2.1.

4. A2(K) (alternatively A2) specifies the sense in which the conditional expectations of small increments given the past should be asymptotically proportional to the drift at every instant of time. A3(K) has a similar interpretation. We have truncated the increments arbitrarily at $\epsilon = 1$. It is easy to see that any other truncation value $\epsilon > 0$ gives an equivalent set of conditions, since by A1(K):

$$\sum_{k=0}^{r_n(t)} P[\,|\Delta X_n(k)| > \epsilon \mid \mathscr{F}_n(t_n^k)]I(X_n(t_n^k) \in K) \to_p 0. \tag{2.13}$$

If we insist upon using untruncated moments in A2(K) and A3(K), we have to add the Lindeberg condition

$$\sum_{k=0}^{r_n(t)} E(\{\Delta X_n(k)\}^2;\, |\Delta X_n(k)| > \epsilon \mid \mathscr{F}_n(t_n^k))I(X_n(t_n^k) \in K) \to_p 0 \tag{2.14}$$

for all ϵ, t > 0 and for all compact intervals K in G.

5. In many applications it is natural to use a sequence of partitions $\{t_n^k\}$ for which $X_n(t)$ is constant on each interval $[t_n^k, t_n^{k+1})$. For instance, when $X_n(t) = Y_n([nt])$, where $\{Y_n(k); k = 0, 1, \cdots\}_n$ is some sequence of pure jump Markov process, we will take $t_n^k = kn^{-1}$. In such cases A1(K) is implied by the simpler condition that (2.13) should hold for all ϵ, $t >$ 0, and (2.13) is implied by the Lindeberg condition (2.14).

6. Assuming that $X_n(t)$ is constant on each interval $[t_n^k, t_n^{k+1})$, A1–A3 are implied by

$$\sum_{k=0}^{r_n(t)} |e_n^j(k)| \to_p 0 \tag{2.15}$$

for $j = 1, 2, 3$ and $t > 0$, where

$$e_n^1(k) = E(\Delta X_n(k) \mid \mathscr{F}_n(t_n^k)) - \mu(X_n(t_n^k))\Delta t_n(k)$$
$$e_n^2(k) = E(\{\Delta X_n(k)\}^2 \mid \mathscr{F}_n(t_n^k)) - \sigma^2(X_n(t_n^k))\Delta t_n(k)$$
$$e_n^3(k) = E(|\Delta X_n(k)|^3 \mid \mathscr{F}_n(t_n^k)).$$

In this way Theorem 2.1 may be shown to generalize and unite several results in the literature. For instance, the "stochastic" conditions for convergence in Borovkov [3] are essentially of the form:

$$P[\cup_{k=0}^{r_n(t)} [e_n^j(k) > \epsilon \Delta t_n(k)]] \to 0$$

for all ϵ, $t > 0$ and $j = 1, 2, 3$, while his conditions for convergence "in the mean" are related to

$$E \sum_{k=0}^{r_n(t)} |e_n^j(k)| \to 0 \qquad t > 0;\; j = 1, 2, 3$$

both of which imply (2.15). (The exact forms of Borovkov's conditions are more complicated, but they can be shown to imply those of Theorem 2.1 by this type of reasoning.) Also, Theorem 2.1 generalizes the more recent results by Norman [26] (except that we do not consider convergence to processes with regular boundaries in this paper). This generalization goes in several directions: the state interval need not be bounded; we have no restriction on the coefficients beyond continuity of $\mu(\cdot)$ and $\sigma^2(\cdot)$ and positivity of $\sigma^2(\cdot)$ in G; in the conditions on the conditional moments of the increments (A2(K) and A3(K)) we have no requirement of uniformity near the boundaries of F; also, we require only convergence in probability in these conditions, not convergence in L^1-norm.

7. Using the semigroup approach, Kurtz [17] has developed general theorems on convergence to Markov processes (see also [18]). His results are applicable, e.g., when $C_0^2(F)$ (the space of twice continuously differentiable functions with compact support in F) is a core for the generator A of the limiting Markov process—i.e., the closure of A restricted to $C_0^2(F)$ is equal to A. Assuming this, it is possible to deduce results related to our Theorem 2.1 by combining Theorem (3.11) in [17] with a Taylor expansion and using

criteria for tightness from Section 4 in [17]. Indeed, it must be possible to deduce (essentially) Theorem 2.1 this way, since there exist unpublished versions of the conditions in [17] which are necessary and sufficient for convergence in a certain sense (Kurtz, private communication). Trying to carry out this program in practice, however, one runs into at least two difficulties. First, Kurtz's conditions for convergence concern each single time point t, while our conditions involve the behavior of $X_n(s)$ on the whole interval $[0, t]$. Secondly, Kurtz's conditions are of L^1-type (see his definition of p-lim, (1.7) in [17]), while we only require convergence in probability. Thus some truncation argument must be added. For an example where $X_n(t) \Rightarrow W(t)$ (Brownian motion), where the Lindeberg condition of the form (2.14) holds (with $K = R^1$), but where the corresponding L^1-condition does not hold, see page 623 in McLeish [23].

For deciding whether or not C_0^2 is a core for the generator of a given diffusion process, recent results by Ethier [8] are helpful. Most processes used in applications seem to satisfy this condition. However, one may also construct simple processes for which the condition does not hold. One example is given by $X(t) = | W(t) |^3$, another by $\tilde{X}(t) = W(t \wedge R_0)^3$, where $W(t)$ is Brownian motion and $R_0 = \inf\{t \geq 0 : W(t) = 0\}$. For both processes $\mu(x) = 3x^{1/3}$ and $\sigma^2(x) = 9x^{4/3}$, and $C_0^2[0, \infty)$ is not a core for any of the corresponding generators on $\hat{C}[0, \infty)$—the space of continuous functions vanishing at infinity. (I am grateful to T. Olsen for pointing out this to me.) Even if these examples are not quite relevant here (they involve a regular boundary point), they give some support to the idea of seeking alternative routes to limit theorems.

3. Necessity of the conditions. The conditions given in Section 2 are not necessary for weak convergence, even though they are very close to being so. An example for which $X_n(t) \Rightarrow X(t)$, but where the condition A2 does not hold for any sequence of partitions $\{t_n^k\}$, is given in Helland [13]. Using the same calculations as in [13], we can show that the condition A2(K) does not hold here either (for any compact interval K with nonempty interior). This example motivated the following

DEFINITION 3.1. Let $X_n(t)$ $(n = 1, 2, \cdots)$ and $X(t)$ be processes with paths in $D[0, \infty)$ and adapted to families of σ-fields $\mathscr{F}_n(t)$ and $\mathscr{F}(t)$ respectively. We say that $\{(X_n(t), \mathscr{F}_n(t))\}$ converges to $(X(t), \mathscr{F}(t))$ weakly and in conditional distributions and write $(X_n(t), \mathscr{F}_n(t)) \Rightarrow_c (X(t), \mathscr{F}(t))$ if the following holds: for every choice of $m \geq 1$, of m time points $t_i \geq 0 (i = 1, \cdots, m)$ and of m bounded, continuous functionals $g_i (i = 1, \cdots, m)$ on $D[0, \infty)$ the joint distribution of $E\{g_i(X_n(t_i + \cdot)) \mid \mathscr{F}_n(t_i)\}$ $(i = 1, \cdots, m)$ converges to that of $E\{g_i(X(t_i + \cdot)) \mid \mathscr{F}(t_i)\}$ $(i = 1, \cdots, m)$.

If this holds with $\mathscr{F}_n(t) = \sigma\{X_n(s); s \leq t\}$ and $\mathscr{F}(t) = \sigma\{X(s); s \leq t\}$, we simply write $X_n(t) \Rightarrow_c X(t)$.

It is clear that this type of convergence is at least as strong as weak convergence, and the example in [13] can be used to show that it is strictly stronger. It will be proved in Section 7 that the conditions of Theorem 2.1 are necessary and sufficient in order that $(X_n(t), \mathscr{F}_n(t)) \Rightarrow_c (X(t), \mathscr{F}(t))$ with $\mathscr{F}(t) = \sigma\{X(s); s \leq t\}$. A precise formulation will be given below, but first we will discuss some aspects of Definition 3.1.

Suppose that the joint distribution of $E\{g_i(X_n(t_i + \cdot)) \mid \mathscr{F}_n(t_i)\}$ $(i = 1, \cdots, m)$ converges to that of $E\{g_i(X(t_i + \cdot)) \mid \mathscr{F}(t_i)\}$ $(i = 1, \cdots, m)$ whenever the g_i's are uniformly continuous functions on D. Then $(X_n(t), \mathscr{F}_n(t)) \Rightarrow_c (X(t), \mathscr{F}(t))$. On the other hand, if we know that $(X_n(t), \mathscr{F}_n(t)) \Rightarrow_c (X(t), \mathscr{F}(t))$, then this same convergence in joint distribution holds whenever the g_i's are bounded, measurable functions that are continuous on some $C \subset D$ for which $P[X(t_i + \cdot) \in C] = 1$ for $i = 1, \cdots, m$. This is the content of Proposition 4.4 in [13]. The last result will be used in Section 7 to prove that $(X_n(t), \mathscr{F}_n(t)) \Rightarrow_c (X(t), \mathscr{F}(t))$ implies the conditions of Theorem 2.1.

Another useful result is the following (Proposition 4.9 on [13]): Let $X_n(t)$ and $\tilde{X}_n(t)$ be processes with paths in D, both adapted to the same family of σ-fields $\mathscr{F}_n(t)$. Suppose that

$(\tilde{X}_n(t), \mathscr{F}_n(t)) \Rightarrow_c (X(t), \mathscr{F}(t))$ for some $X(t)$ and $\mathscr{F}(t)$, and that $\sup_{s\le t} |X_n(s) - \tilde{X}_n(s)| \to_p 0$ for all $t > 0$. Then $(X_n(t), \mathscr{F}_n(t)) \Rightarrow_c (X(t), \mathscr{F}(t))$.

Next we show that the convergence $X_n(t) \Rightarrow_c X(t)$ is independent of the representation of the processes $X_n(t)$ and $X(t)$.

LEMMA 3.2. *Let $X(t)$ and $\tilde{X}(t)$ be two processes with paths in $D = D[0, \infty)$ and with the same distribution on D. Then for any m, any m time points $t_i \ge 0$ and any m bounded, measurable functionals g_i on D the joint distribution of $E(g_i(X(t_i + \cdot)) \mid \mathscr{F}(t_i))$ $(i = 1, \cdots, m)$ is equal to the joint distribution of $E(g_i(\tilde{X}t_i + \cdot)) \mid \tilde{\mathscr{F}}(t_i))$ $(i = 1, \cdots, m)$, where $\mathscr{F}(t) = \sigma\{X(s); s \le t\}$ and $\tilde{\mathscr{F}}(t) = \sigma\{\tilde{X}(s); s \le t\}$.*

PROOF. By Theorem 1.6 in the supplement of Doob [5] we may replace $\mathscr{F}(t_i)$ in $E(g_i(X(t_i + \cdot)) \mid \mathscr{F}(t_i))$ by $\mathscr{F}_0(t_i) = \sigma\{X(s); s \in I_i\}$ for some countable set $I_i \subset [0, t_i]$, similarly $\tilde{\mathscr{F}}_0(t_i)$ for $\tilde{\mathscr{F}}(t_i)$ (with the same I_i). Furthermore, we may suppose that $I_i \subset I_j$ whenever $t_i < t_j$. Now $\mathscr{F}_0(t_i)$ and $\tilde{\mathscr{F}}_0(t_i)$ may be generated by sequences of finite σ-fields $\mathscr{F}_1(t_i) \subset \mathscr{F}_2(t_i) \subset \cdots$, respectively $\tilde{\mathscr{F}}_1(t_i) \subset \tilde{\mathscr{F}}_2(t_i) \subset \cdots$, chosen in such a way that the joint distributions of $E(g_i(X(t_i + \cdot)) \mid \mathscr{F}_k(t_i))$ $(i = 1, \cdots, m)$ and $E(g_i(\tilde{X}(t_i + \cdot)) \mid \tilde{\mathscr{F}}_k(t_i))$ $(i = 1, \cdots, m)$ are equal. The lemma follows by letting $k \to \infty$ and applying martingale convergence as in [5], Chapter VII, Section 8.

The following theorem—the second main result of this paper—shows the precise sense in which the conditions of Theorem 2.1 are minimal. Let $X(t)$ be a diffusion process satisfying the conditions in Section 2, and put $\mathscr{F}(t) = \sigma\{X(s); s \le t\}$. For each n, let $X_n(t)$ be a process with paths in D and adapted to a family of σ-fields $\mathscr{F}_n(t)$. Recall A0–A4 of Section 2. Also recall that $G = \text{int } F$.

THEOREM 3.3. *Assume* A1 (*alternatively*: A1(K) *for all compacts K in some open set containing F*) *and* A0. *Assume that* A2(K) *and* A3(K) *hold for all compact intervals K in G, and that* A4(i) *holds whenever r_i is accessible for $X(t)$. Then $(X_n(t), \mathscr{F}_n(t)) \Rightarrow_c (X(t), \mathscr{F}(t))$.*

On the other hand, assume that $(X_n(t), \mathscr{F}_n(t)) \Rightarrow_c (X(t), \mathscr{F}(t))$. Then A0, A4(0) *and* A4(1) *hold, and* A1 *holds for every sequence of partitions satisfying* (2.7) *and* (2.8). *Furthermore, there exists a sequence of partitions of the form $\{k\delta_n; k \ge 0\}_n$ with $\delta_n \downarrow 0$ such that* A2(K) *and* A3(K) *with this sequence of partitions hold for all compact intervals K in G.*

In most cases in the literature where weak convergence to diffusion processes is proved and used, the conditions of Theorem 3.3 hold, at least with $\mathscr{F}_n(t) = \sigma\{X_n(s); s \le t\}$. Of course, in some cases, e.g., in connection with the results of Guess and Gillespie [11], the conditions may be difficult to verify, but constructed counterexamples of the type mentioned above, seem to be the only exceptions to the rule that A1, A2(K) and A2(K) ($K \subset G$) always hold for *some* sequence of partitions of the time axis. In particular, in all the usual applications in population genetics ([16], [26]), these conditions can be shown to hold. Two other examples are discussed in Section 8 below.

Thus $X_n(t) \Rightarrow_c X(t)$ in all these cases, and in many other cases where the limiting process has continuous paths. When $X(\cdot)$ has fixed discontinuities, our definition of $\Rightarrow_c$ seems to be too strong, as a referee pointed out: taking $g_i(x(\cdot)) = h_i(x(0))$ in Definition 3.1, we see that $\Rightarrow_c$ implies convergence of all finite-dimensional distributions. For appropriate definitions in the general case, the reader is referred to a recent paper by Aldous [1], where a related concept of convergence is discussed in great detail. Aldous argues that a process is described not merely by $\{X(t)\}$, but also by a filtration $\{\mathscr{F}(t); t \ge 0\}$, and that the concept of convergence should take this into account. His *extended weak convergence* coincides with our convergence $\Rightarrow_c$ when the limit is a diffusion process with its natural filtration.

4. Central limit theorem for dependent variables and random time change. A variant of Theorem 3.3 for the case where the limit $X(t)$ is Brownian motion, is proved in [13], using martingale central limit theorems. Here we only need the first half of this result. In the following $Z_n(t) \to_{pu} Z(t)$ means $\sup_{0\le s\le t} |Z_n(s) - Z(s)| \to_p 0$ for all $t > 0$.

PROPOSITION 4.1. *Let $W(t)$ be a standard Brownian motion process, and put $\mathscr{F}(t) = \sigma\{W(s); s \le t\}$. Suppose that $X_n(0) \to_p 0$, that* A1 *holds and that*

$$\sum_{k=0}^{r_n(t)} E(\Delta_1 X_n(k) \mid \mathscr{F}_n(t_n^k)) \to_{pu} 0 \tag{4.1}$$

$$\sum_{k=0}^{r_n(t)} \operatorname{Var}(\Delta_1 X_n(k) \mid \mathscr{F}_n(t_n^k)) \to_{pu} t. \tag{4.2}$$

Then $(X_n(t), \mathscr{F}_n(t)) \Rightarrow_c (W(t), \mathscr{F}(t))$.

In particular, under these conditions $X_n(t) \Rightarrow W(t)$. The object of the present paper is to transform this result by a random time change. The transformation that we need was discussed in Helland [12]. We recall the main result of [12] in the form that we will use it here.

Let $D = D[0, \infty)$ and let K be a compact interval on the line. Put $B = (\text{int } K)^c$. Let φ be a nonnegative continuous function on K, strictly positive on int K. We will define a mapping f from D into D as follows. For $x = \{x(t); t \ge 0\} \in D$, define

$$q_x = \inf\{t \ge 0 : x(t) \in B \quad \text{or} \quad x(t-) \in B\} \tag{4.3}$$

and

$$r_x = \int_0^{q_x} ds/\varphi(x(s)). \tag{4.4}$$

For $t < r_x$ let $\tau_x(t)$ be determined (uniquely) by

$$t = \int_0^{\tau_x(t)} ds/\varphi(x(s)). \tag{4.5}$$

Define $f = f_{\varphi,B}$ by

$$\begin{aligned} f(x)(t) &= x(\tau_x(t)) && \text{for} \quad 0 \le t < r_x \\ &= x(q_x-) && \text{for} \quad t \ge r_x \quad \text{if} \quad x(q_x-) \in B \\ &= x(q_x) && \text{for} \quad t \ge r_x \quad \text{if} \quad x(q_x-) \notin B. \end{aligned} \tag{4.6}$$

Note that in the present case $q_x = +\infty$ always implies $r_x = +\infty$. Thus f is always well defined; the problem with explosions is avoided because K is compact. The following proposition follows easily from Theorem 2.6 in Helland [12] and well-known properties of Brownian motion.

PROPOSITION 4.2. *Let P_1 be the probability measure on D induced by a Brownian motion without drift (with any initial distribution). Let P_2 be the probability measure on D induced by Brownian motion (without drift) stopped when hitting $B = (\text{int } K)^c$. Then $f_{\varphi,B}$ is continuous on a Borel subset $C \subset D$ with $P_1[C] = P_2[C] = 1$.*

To deduce weak convergence of processes from this result, we also need the fact that f is a Borel-measurable function from D into D. This was not discussed in [12], but it is relatively easy to see that f is always measurable. First we use the decomposition ((2.6) in [12]) $f = g \circ h$, where g is continuous everywhere (Proposition 2.5 in [12]) and h is defined as in (4.3)–(4.6) with $\varphi \equiv 1$. Thus we only have to prove measurability of h. From Billingsley [2], Theorem 14.5, the finite-dimensional sets generate the Borel sets of D, so it is enough to prove that for any Borel set A on the line the set $[h(x)(t) \in A]$ is a Borel set in D. But now

$$(4.7)\quad [h(x)(t) \in A] = ([q_x > t] \cap [x(t) \in A]) \cup ([q_x \le t] \cap [x(q_x-) \in A \cap B]) \cup ([q_x \le t] \cap [x(q_x-) \notin B] \cap [x(q_x) \in A]).$$

The mapping $x \to q_x$ is measurable since

$$(4.8)\quad [q_x \le t] = [x(t) \in B] \cup \left(\cap_{n=1}^{\infty} \cup_{r<t, r\,\mathrm{rational}} \left[d(x(r), B) < \frac{1}{n} \right] \right),$$

where $d(\xi, B) = (\xi - \alpha)^+ \wedge (\beta - \xi)^+$ if $K = [\alpha, \beta]$. The mappings $x \to x(q_x)$ and $x \to x(q_x-)$ are measurable by an argument similar to that of Proposition 12.39 in Breiman [4], and this completes the proof of measurability. Essentially the same proof works under the more general assumptions of [12].

5. Convergence to processes stopped at the boundaries of compacts. Let K be a compact subinterval of F such that $\mu(\cdot)$ and $\sigma^2(\cdot)$ are continuous on K and $\sigma^2(\cdot) > 0$ on K. In analogy with (4.3) and (4.6) define

$$(5.1)\quad R_n^K = \inf\{t \ge 0 : X_n(t) \notin \mathrm{int}\, K \quad \text{or} \quad X_n(t-) \notin \mathrm{int}\, K\}$$

$$(5.2)\quad X_n^K(t) = \begin{cases} X_n(t) & \text{for } t < R_n^K \\ X_n(R_n^K-) & \text{for } t \ge R_n^K \quad \text{if } X_n(R_n^K-) \notin \mathrm{int}\, K \\ X_n(R_n^K) & \text{for } t \ge R_n^K \quad \text{if } X_n(R_n^K-) \in \mathrm{int}\, K. \end{cases}$$

By an identity similar to (4.8) we prove that R_n^K is a stopping time relative to $\mathscr{F}_n(t)$ (i.e., $[R_n^K \le t] \in \mathscr{F}_n(t)$ for all t). It is technically convenient to stop the processes also when $X_n(t-)$ is at the boundary of K, even if the definitions seem complicated. (The corresponding aspect of Definition (4.6) is crucial for continuity of f.) If $\mathscr{F}_n(t)$ is a right-continuous family and $X_n^K(t)$ is quasileft-continuous, then $R_n^K = \inf\{t \ge 0 : X_n(t) \notin \mathrm{int}\, K\}$ a.s. and $X_n^K(t) = X_n(t \wedge R_n^K)$ a.s., where the exceptional set may be chosen independent of t.

Similarly, let $R^K = \inf\{t \ge 0 : X(t) \notin \mathrm{int}\, K\}$ and $X^K(t) = X(t \wedge R^K)$. In the next section, Theorem 2.1 will be proved from the basic

PROPOSITION 5.1. *Assume that* A0, A1(K), A2(K) *and* A3(K) *hold. Then* $X_n^K(t) \Rightarrow X^K(t)$.

One should note that this proposition, as well as the next one, also is valid for the case where F includes one or two regular boundary points for the diffusion $X(t)$.

We will first prove Proposition 5.1 for the case $\mu(\cdot) = 0$. In fact, for this case we will prove a formally stronger result. See Section 4 for the definition of $\to_{pu}$. Define $k_n^K = \max\{k : t_n^k < R_n^K\}$ when $R_n^K > 0$, otherwise let $k_n^K = -1$, which by definition means that any sum from 0 to k_n^K then is empty.

PROPOSITION 5.2. *Let K, $X_n^K(t)$ and $X^K(t)$ be as above with $\mu(\cdot) = 0$. Assume that* A0 *and* A1(K) *hold and that*

$$(5.3)\quad \textstyle\sum_{k=0}^{r_n(t) \wedge k_n^K} E(\Delta_1 X_n(k) \mid \mathscr{F}_n(t_n^k)) \to_{pu} 0$$

$$(5.4)\quad \textstyle\sum_{k=0}^{r_n(t) \wedge k_n^K} \mathrm{Var}(\Delta_1 X_n(k) \mid \mathscr{F}_n(t_n^k)) - \displaystyle\int_0^{t \wedge R_n^K} \sigma^2(X_n(s))\, ds \to_{pu} 0.$$

Then $X_n^K(t) \Rightarrow X^K(t)$.

PROOF. We will first prove Proposition 5.2 for the case $\sigma^2(\cdot) = 1$. To this end we introduce a new sequence of processes $\{Y_n(t)\}$ as follows. Define $S_n = t_n^j$ with $j = k_n^K + 1$

when $R_n^K < +\infty$, otherwise let $S_n = +\infty$. Let

$$Y_n(t) = \begin{cases} X_n(t) & \text{for} \quad t \le S_n \\ X_n(S_n) + W(t - S_n) & \text{for} \quad t > S_n, \end{cases} \tag{5.5}$$

where $W(t)$ is a standard Brownian motion ($\mu = 0$, $\sigma^2 = 1$, $W(0) = 0$) independent of $X_n(t)$ and of the σ-fields $\mathscr{F}_n(t)$. (Without loss of generality we may assume that the probability space is rich enough to support such a process.) Define $\tilde{\mathscr{F}}_n(t) = \mathscr{F}_n(t) \vee \sigma\{Y_n(s);\, s \le t\}$, $\Delta Y_n(k) = Y_n(t_n^{k+1}) - Y_n(t_n^k)$ and $\Delta_1 Y_n(k) = \Delta Y_n(k) \cdot I(|\Delta Y_n(k)| \le 1)$. Then it is easy to verify that

$$\textstyle\sum_{k=0}^{r_n(t)} P[VY_n(k) > \epsilon \mid \tilde{\mathscr{F}}_n(t_n^k)] \to_p 0 \qquad \text{for all} \quad \epsilon, t \tag{5.6}$$

$$\textstyle\sum_{k=0}^{r_n(t)} E(\Delta_1 Y_n(k) \mid \tilde{\mathscr{F}}_n(t_n^k)) \to_{pu} 0 \tag{5.7}$$

$$\textstyle\sum_{k=0}^{r_n(t)} \operatorname{Var}(\Delta_1 Y_n(k) \mid \tilde{\mathscr{F}}_n(t_n^k)) \to_{pu} t. \tag{5.8}$$

Therefore by Proposition 4.1 we have $(Y_n^0(t), \tilde{\mathscr{F}}_n(t)) \Rightarrow_c (W(t), \mathscr{F}_0(t))$, where $Y_n^0(t) = Y_n(t) - Y_n(0)$ and $\mathscr{F}_0(t) = \sigma\{W(s);\, s \le t\}$. In particular, $E\{g(Y_n(\cdot) - Y_n(0)) \mid \tilde{\mathscr{F}}_n(0)\}$ converges in distribution (hence in probability) to $E\{g(W(\cdot))\}$ for every bounded continuous functional g on D. Since $Y_n(0) = X_n(0)$ is measurable with respect to $\tilde{\mathscr{F}}_n(0)$ and $X_n(0) \to_D X(0)$, we find that $E\{g(Y_n(\cdot) - Y_n(0))h(Y_n(0))\} \to E\{g(W(\cdot))\}E\{h(X(0))\}$ for every bounded continuous function h on R^1. From this it follows that $Y_n(t) \Rightarrow X(t)$, where $X(t)$ is a Brownian motion with $\mu = 0$, $\sigma^2 = 1$ and initial distribution given by the distribution of $X(0)$. If we now use (4.3)–(4.6) with $\varphi \equiv 1$ and $B^c = \operatorname{int} K$, we find that $X_n^K(\cdot) = f_{\varphi, B}(Y_n(\cdot))$ and $X^K(\cdot) = f_{\varphi,B}(X(\cdot))$. Thus the proof for this case is completed by appealing to Proposition 4.2 and the continuous mapping theorem.

Next we prove Proposition 5.2 for the general case where $\sigma^2(\cdot)$ is any positive continuous function on K. (The proposition is true also when $\sigma^2(\cdot) = 0$ at one or both boundaries of K, but since we do not use this result, we omit the proof.) First we note that the sequence of processes $\{X_n^K(t)\}$ satisfies the assumptions of Proposition 5.2 along with $\{X_n(t)\}$. We will transform this sequence by a random time transformation, and define

$$Q_n^K = \int_0^{R_n^K} \sigma^2(X_n^K(s))\, ds \qquad \left(= \int_0^{R_n^K} \sigma^2(X_n(s))\, ds \right). \tag{5.9}$$

For $t < Q_n^K$ we define $T_n(t)$ by

$$t = \int_0^{T_n(t)} \sigma^2(X_n^K(s))\, ds, \tag{5.10}$$

and let $T_n(t) = R_n^K + t - Q_n^K$ for $t \ge Q_n^K$. From this we define a new sequence of processes $\{Z_n(t)\}$ by

$$Z_n(t) = X_n^K(T_n(t)). \tag{5.11}$$

A new sequence of partitions $\{s_n^k;\, k \ge 0\}_n$ is defined by

$$t_n^k = T_n(s_n^k) \qquad \left(\text{i.e., } s_n^k = \int_0^{t_n^k} \sigma^2(X_n(s))\, ds \quad \text{on} \quad [t_n^k \le R_n^K]\right).$$

Since $T_n(\cdot)$ is strictly increasing, this determines s_n^k uniquely. Also, since $\sigma^2(\cdot)$ is bounded away from 0, $s_n^k \to +\infty$ a.s. as $k \to +\infty$ for each n. By the boundedness of $\sigma^2(\cdot)$ on K it is easy to see that $\max_{0 \le k \le q_n(t)} \Delta s_n(k) \to 0$ in probability for all $t > 0$, where $\Delta s_n(k) = s_n^{k+1} - s_n^k$ and $q_n(t) = \max\{k \ge 0 : s_n^k \le t\}$. Let $\mathscr{G}_n(t) = \mathscr{F}_n(T_n(t))$. Then each s_n^k is a stopping time relative to $\{\mathscr{G}_n(t);\, t \ge 0\}$.

Since $\sigma^2(x) \ge \delta$ for all $x \in K$ for some $\delta > 0$, it is clear from (5.10) that $\{T_n(t);\, t \in [0, a],\, n = 1, 2, \cdots\}$ is contained in a compact interval $[0, \delta^{-1}a]$ for each $a > 0$. Therefore we may replace t by $T_n(t)$ in expressions like the left-hand side of (5.4). This gives, with the notation $\Delta_1 Z_n(k) = \{Z_n(s_n^{k+1}) - Z_n(s_n^k)\} \cdot I(|Z_n(s_n^{k+1}) - Z_n(s_n^k)| \le 1)$:

$$\sum_{k=0}^{q_n(t)\wedge k_n^K} P[\sup_{s_n^k \le s \le s_n^k} |Z_n(s) - Z_n(s_n^k)| > \epsilon \mid \mathscr{G}_n(s_n^k)] \to_p 0 \quad \text{for all} \quad \epsilon, t > 0, \tag{5.12}$$

$$\sum_{k=0}^{q_n(t)\wedge k_n^K} E(\Delta_1 Z_n(k) \mid \mathscr{G}_n(s_n^k)) \to_{pu} 0, \tag{5.13}$$

$$\sum_{k=0}^{q_n(t)\wedge k_n^K} \mathrm{Var}(\Delta_1 Z_n(k) \mid \mathscr{G}_n(s_n^k)) - t \wedge Q_n^K \to_{pu} 0, \tag{5.14}$$

where we have used the identity $q_n(t) = r_n(T_n(t))$. Note that $k_n^K = \max\{k : s_n^k < Q_n^K\}$ when $Q_n^K > 0$, otherwise $k_n^K = -1$. By the argument given in the first part of this proof, (5.12)–(5.14) imply that $Z_n(t) \Rightarrow Z(t)$ in $D[0, \infty)$, where $Z(t)$ is a Brownian motion with $\mu = 0$, $\sigma^2 = 1$, initial distribution given by the distribution of $X^K(0)$, and with the boundaries of K absorbing.

We may invert the transformation (5.9)–(5.11) as follows. We have

$$Q_n^K = \inf\{t \ge 0 : Z_n(t) \notin \text{int } K \quad [\text{or} \quad Z_n(t-) \notin \text{int } K]\} \tag{5.15}$$

$$R_n^K = \int_0^{Q_n^K} ds/\sigma^2(Z_n(s)). \tag{5.16}$$

For $t \le R_n^K$ we can determine $\tau_n := T_n^{-1}$ by

$$t = \int_0^{\tau_n(t)} ds/\sigma^2(Z_n(s)). \tag{5.17}$$

Finally

$$X_n^K(t) = \begin{cases} Z_n(\tau_n(t)) & \text{for} \quad t \le R_n^K \\ Z_n(Q_n^K) & \text{for} \quad t > R_n^K. \end{cases} \tag{5.18}$$

Comparing (5.15)–(5.18) with (4.3)–(4.6) we see that $X_n^K(t) = f(Z_n(\cdot))(t)$, where $f = f_{\varphi,B}$ with $\varphi(\cdot) = \sigma^2(\cdot)$, $B^c = \text{int } K$. By Proposition 4.2 we therefore have $X_n^K(t) \Rightarrow f_{\varphi,B}(Z(\cdot))(t)$ in $D[0, \infty)$. It is easy to see from standard diffusion theory (see Breiman [4], Chapter 16) that $f_{\varphi,B}(Z(\cdot))(t)$ with φ, B and $Z(t)$ as above is a diffusion on natural scale with infinitesimal variance $\sigma^2(\cdot)$, with the same initial distribution as $Z(t)$, and with boundaries of K absorbing.

PROOF OF PROPOSITION 5.1. For the case $\mu(\cdot) \equiv 0$ this is a consequence of Proposition 5.2, since A2(K) trivially implies (5.3), and A1(K), A2(K) and A3(K) together imply (5.4). Indeed, from A1(K) (cf., (2.11)) and the uniform continuity of $\sigma^2(\cdot)$ on K, we see that

$$\int_0^{t\wedge R_n^K} \sigma^2(X_n(s))\, ds - \sum_{k=0}^{r_n(t)\wedge k_n^K} \sigma^2(X_n(t_n^k))\Delta t_n(k) \to_{pu} 0, \tag{5.19}$$

and from A2(K) it follows that

$$\sum_{k=0}^{r_n(t)\wedge k_n^K} \{E(\Delta_1 X_n(k) \mid \mathscr{F}_n(t_n^k))\}^2 \to_{pu} 0. \tag{5.20}$$

By combining (5.19), (5.20) and A3(K) it is easy to prove (5.4), and thereby $X_n^K(t) \Rightarrow X^K(t)$. (In all these arguments we use the fact that $k \le k_n$ implies $X_n(t_n^k) \in K$.)

The general case (with $\mu(\cdot) \not\equiv 0$) is reduced to the above case by transforming to the natural scale of the limiting diffusion. Thus we define $U_n(t) = u(X_n(t))$ and $U(t) = u(X(t))$, where $u(\cdot)$ is determined by (2.4). Since $u(\cdot)$ is continuous on K, $u(K)$ is compact. By the continuity of $\mu(\cdot)\sigma^{-2}(\cdot)$ we even have that $u'(\cdot)$ and $u''(\cdot)$ are bounded and uniformly continuous on K. Therefore, if we put $\Delta U_n(k) = U_n(t_n^{k+1}) - U_n(t_n^k)$, we have by a Taylor-expansion

$$\Delta U_n(k) = u'(X_n(t_n^k))\Delta X_n(k) + \{\tfrac{1}{2} u''(X_n(t_n^k)) + H_n^k\}\{\Delta X_n(k)\}^2, \tag{5.21}$$

where H_n^k is small when $\Delta X_n(k)$ is small. By A1(K) (cf., (2.11)) we can find a sequence $\{\epsilon_n\}$, with $\epsilon_n \downarrow 0$ so slowly that

$$P[\max_{0\le k\le r_n(t)\wedge k_n^K} |\Delta X_n(k)| > \epsilon_n] \to 0 \tag{5.22}$$

$$P[\max_{0\le k\le r_n(t)\wedge k_n^K} |H_n^k| > \epsilon_n] \to 0.$$

Let $A_{n,k} = [|\Delta X_n(k)| > \epsilon_n] \cup [|H_n^k| > \epsilon_n]$. Then $P[\cup_{k=0}^{r_n(t)\wedge k_n^K} A_{n,k}] \to 0$, and by Lemma 3.1 in Helland [13] this implies

$$\textstyle\sum_{k=0}^{r_n(t)\wedge k_n^K} P[A_{n,k} \mid \mathscr{F}_n(t_n^k)] \to_{pu} 0. \tag{5.23}$$

By the boundedness of $u'(\cdot)$ and $u''(\cdot)$ on K we see from (5.21) that $\Delta_1 U_n(k) = \Delta U_n(k)$ on $A_{n,k}^c$ for n large enough, where as usual $\Delta_1 U_n(k) = \Delta U_n(k)\cdot I(|\Delta U_n(k)| \le 1)$. We also need the identity (cf., (2.4)) $u''(x) = -2\mu(x)\sigma^{-2}(x)u'(x)$. This gives from (5.21)

$$\begin{aligned}
&\textstyle\sum_{k=0}^{r_n(t)\wedge k_n^K} E(\Delta_1 U_n(k) \mid \mathscr{F}_n(t_n^k))\\
&= \textstyle\sum E(\Delta_1 U_n(k); A_{n,k}^c \mid \mathscr{F}_n(t_n^k)) + \sum E(\Delta_1 U_n(k); A_{n,k} \mid \mathscr{F}_n(t_n^k))\\
&= \textstyle\sum u'(X_n(t_n^k))E(\Delta_1 X_n(k) \mid \mathscr{F}_n(t_n^k))\\
&\quad - \textstyle\sum \mu(X_n(t_n^k))\sigma^{-2}(X_n(t_n^k))u'(X_n(t_n^k))E(\{\Delta_1 X_n(k)\}^2 \mid \mathscr{F}_n(t_n^k))\\
&\quad + \textstyle\sum E(H_n^k\{\Delta_1 X_n(k)\}^2; A_{n,k}^c \mid \mathscr{F}_n(t_n^k))\\
&\quad + O(\textstyle\sum P[A_{n,k} \mid \mathscr{F}_n(t_n^k)]),
\end{aligned} \tag{5.24}$$

where all summations are from $k = 0$ to $k = r_n(t) \wedge k_n^K$.

The first sum on the right-hand side of (5.24) is (pu)-asymptotically equal to

$$\textstyle\sum_{k=0}^{r_n(t)\wedge k_n^K} u'(X_n(t_n^k))\mu(X_n(t_n^k))\Delta t_n(k) \tag{5.25}$$

by A2(K). In the second sum we may replace $E(\{\Delta_1 X_n(k)\}^2 \mid \mathscr{F}_n(t_n^k))$ by $\sigma^2(X_n(t_n^k))\Delta t_n(k)$ since A3(K) holds. Therefore this sum is also (pu)-asymptotically equal to (5.25). Since the third sum is of order $O(\epsilon_n)$, this shows that

$$\textstyle\sum_{k=0}^{r_n(t)\wedge k_n^K} E(\Delta_1 U_n(k) \mid \mathscr{F}_n(t_n^k)) \to_{pu} 0.$$

By a similar calculation, using the simpler identity $\Delta U_n(k) = \{u'(X_n(t_n^k)) + G_n^k\}\Delta X_n(k)$, where G_n^k is small when $\Delta X_n(k)$ is small, we find that

$$\textstyle\sum_{k=0}^{r_n(t)\wedge k_n^K} \{E(\{\Delta_1 U_n(k)\}^2 | \mathscr{F}_n(t_n^k)) - u'(X_n(t_n^k))^2\sigma^2(X_n(t_n^k))\,\Delta t_n(k)\} \to_{pu} 0$$

$$\textstyle\sum_{k=0}^{r_n(t)\wedge k_n^K} \{E(\Delta_1 U_n(k) | \mathscr{F}_n(t_n^k))\}^2 \to_{pu} 0.$$

From this and the uniform continuity of $u'(\cdot)$ and $\sigma^2(\cdot)$ we see that (5.4) holds with $X_n(s)(\Delta_1 X_n(k))$ replaced by $U_n(s)(\Delta_1 U_n(k))$ and with $\sigma^2(x)$ replaced by $\tilde{\sigma}^2(y) = u'(x)^2\sigma^2(x)$ where $y = u(x)$. Furthermore it is immediate from A1(K) and uniform continuity that

$$\textstyle\sum_{k=0}^{r_n(t)} P[\sup_{t_n^k \le s \le t_n^{k+1}} |U_n(s) - U_n(t_n^k)| > \epsilon \mid \mathscr{F}_n(t_n^k)]I(U_n(t_n^k) \in u(K)) \to_p 0$$

for all $\epsilon, t > 0$. Hence $\{U_n(t)\}$ satisfies the conditions of Proposition 5.2 with $\tilde{\sigma}^2(\cdot)$ replacing $\sigma^2(\cdot)$ and with $u(K)$ replacing K. This shows that $U_n^K(t) \Rightarrow U^K(t)$, where $U(t) = u(X(t))$ and $U^K(t)$, $U_n^K(t)$ are the processes stopped at the boundaries of $u(K)$. Since $u(\cdot)$ has a continuous inverse, this implies that $X_n^K(t) \Rightarrow X^K(t)$.

6. Proof of Theorem 2.1. Assume that the conditions of Theorem 2.1 (first paragraph or second paragraph) hold. Look first at the case where both boundaries r_0 and r_1 are inaccessible for $X(t)$. By Proposition 5.1 we have $X_n^K(t) \Rightarrow X^K(t)$ whenever K is a compact subinterval of $G = (r_0, r_1)$. Fix $\epsilon > 0$ and $c > 0$. Since the boundaries of G are inaccessible, we can find $K = [\alpha, \beta]$ such that

$$P[\sup_{0\le s\le c} X(s) \ge \beta \quad\text{or}\quad \inf_{0\le s\le c} X(s) \le \alpha] < \epsilon. \tag{6.1}$$

For K' a compact interval in G such that $K \subset \text{int}(K')$, this means

$$P[\sup_{0\le s\le c} X^{K'}(s) \ge \beta \quad \text{or} \quad \inf_{0\le s\le c} X^{K'}(s) \le \alpha] < \epsilon.$$

Therefore, since $X_n^{K'}(t) \Rightarrow X^{K'}(t)$, we have for large n

$$P[\sup_{0\le s\le c} X_n^{K'}(s) \ge \beta \quad \text{or} \quad \inf_{0\le s\le c} X_n^{K'}(s) \le \alpha] < \epsilon,$$

which is equivalent to

$$P[\sup_{0\le s\le c} X_n(s) \ge \beta \quad \text{or} \quad \inf_{0\le s\le c} X_n(s) \le \alpha] < \epsilon. \tag{6.2}$$

With $K = [\alpha, \beta]$, (6.1) implies $P[X^k(s) \ne X(s)$ for some s in $[0, c]] < \epsilon$, while (6.2) implies that for large n we have $P[X_n^K(s) \ne X_n(s)$ for some s in $[0, c]] < \epsilon$. Since ϵ was arbitrary, it follows from $X_n^K(t) \Rightarrow X^K(t)$ that $X_n(t) \Rightarrow X(t)$ in $D[0, c]$. Since c was arbitrary, this implies $X_n(t) \Rightarrow X(t)$ in $D[0, \infty)$.

Suppose next that r_1 is inaccessible and r_0 is exit. Without loss of generality we can take $r_0 = 0$. Also, we may suppose that $\mu(\cdot) \equiv 0$, since the result may be transformed to any scale by the argument in the proof of Proposition 5.1. Let $\delta > 0$, and with $K = [\delta, \infty)$ define $R_n^\delta = R_n^K$ and $X_n^\delta(t) = X_n^K(t)$, where R_n^K and $X_n^K(t)$ are defined in (5.1) and (5.2). By combining Proposition 5.1 with an argument similar to that of the preceding paragraph to take care of the inaccessible boundary r_1, we see that $X_n^\delta(t) \Rightarrow X^\delta(t)$, where $X^\delta(t) = X(t \wedge R^\delta)$ with $R^\delta = \inf\{t \ge 0: X(t) \le \delta\}$.

Let $c > 0$, $\epsilon > 0$ be given. Most of the remaining part of this section will be used to prove that δ can be chosen so small that

$$\limsup_{n\to\infty} P[\sup_{R_n^\delta \le s\le c} X_n(s) > \epsilon] < \epsilon. \tag{6.3}$$

Once this is proved (assuming $\delta < \epsilon$), we can combine it with the condition A4(0) to see that

$$\limsup_{n\to\infty} P[\sup_{0\le s\le c} |X_n(s) - X_n^\delta(s)| > 2\epsilon] < \epsilon. \tag{6.4}$$

We also have $P[\sup_{0\le s\le c}|X(s) - X^\delta(s)| > 2\epsilon] < \epsilon$ when δ is small enough. This can be seen from the properties of exit boundaries, alternatively from the fact that A0–A4, which hold trivially if $X_n(t) \equiv X(t)$, imply (6.4). Therefore $X_n(t) \Rightarrow X(t)$ in $D[0, c]$ by Theorem 4.2 in [2], and in $D[0, \infty)$ by the arbitrariness of $c > 0$.

We prove (6.3) through a series of lemmas. Throughout we will let δ, λ and ϵ be constants satisfying $0 < \delta < \lambda < \epsilon$ and let $X_0(t)$ be a diffusion process on F which is on natural scale with diffusion coefficient $\sigma^2(\cdot)$ and $X_0(0) = \lambda$. Extend the probability space in such a way that it supports a sequence of pairs of random variables

$$\{(U^k, V^k); k = 1, 2, \cdots\} \tag{6.5}$$

which are independent of all the processes $X_n(t)$, and such that the pairs (U^k, V^k) are independent and each pair has the same distribution as (Z, Y), where

$$Z = \inf\{t \ge 0: X_0(t) \ge \epsilon \quad \text{or} \quad X_0(t) \le \delta\}; \qquad Y = X_0(Z). \tag{6.6}$$

Thus each U^k is nonnegative, and each V^k takes only two values (δ and ϵ).

Next we define by induction

$$S_n^1 = \inf\{t \ge R_n^\delta: X_n(t) \ge \lambda \quad \text{or} \quad X_n(t-) \ge \lambda\}$$

$$T_n^k = \inf\{t \ge S_n^k: X_n(t) \notin (\delta, \epsilon) \quad \text{or} \quad X_n(t-) \notin (\delta, \epsilon)\} \qquad k = 1, 2, \cdots$$

$$S_n^k \begin{cases} = \inf\{t \ge T_n^{k-1}: X_n(t) \ge \lambda \quad \text{or} \quad X_n(t-) \ge \lambda\} \\ \qquad \text{on} \quad [X_n(T_n^{k-1}-) \le \delta] \cup [X_n(T_n^{k-1}) \le \delta, X_n(T_n^{k-1}-) < \epsilon] \\ = \inf\{t \ge T_n^{k-1}: X_n(t) \le \lambda \quad \text{or} \quad X_n(t-) \le \lambda\} \\ \qquad \text{on} \quad [X_n(T_n^{k-1}-) \ge \epsilon] \cup [X_n(T_n^{k-1}) \ge \epsilon, X_n(T_n^{k-1}-) > \delta] \qquad k = 2, 3, \cdots. \end{cases} \tag{6.7}$$

Finally we take

$$Z_n^k = \begin{cases} T_n^k - S_n^k & \text{on } [S_n^k \le c] \\ U^k & \text{on } [S_n^k > c] \end{cases}$$

$$(6.8) \qquad Y_n^k = \begin{cases} X_n(T_n^k-) & \text{on } [S_n^k \le c] \cap [T_n^k < \infty] \cap [X_n(T_n^k-) \notin (\delta, \epsilon)] \\ X_n(T_n^k) & \text{on } [S_n^k \le c] \cap [T_n^k < \infty] \cap [X_n(T_n^k-) \in (\delta, \epsilon)] \\ 0 & \text{on } [S_n^k \le c] \cap [T_n^k = \infty] \\ V^k & \text{on } [S_n^k > c], \end{cases}$$

where of course $[S_n^k > c]$ is taken to include $[S_n^k = \infty]$.

LEMMA 6.1. *For all $k \ge 1$, the joint distribution of $(Z_n^1, Y_n^1, \cdots, Z_n^k, Y_n^k)$ converges to that of $(U^1, V^1, \cdots, U^k, V^k)$ as $n \to \infty$.*

PROOF. We use induction on k. Let $k \ge 2$ and let B^k be a Borel set in R^{2k-2} such that $P[(U^1, V^1, \cdots, U^{k-1}, V^{k-1}) \in B^k] > 0$ and $P[(U^1, V^1, \cdots, U^{k-1}, V^{k-1}) \in \partial B^k] = 0$. Put

$$(6.9) \qquad A_n^k = [(Z_n^1, Y_n^1, \cdots, Z_n^{k-1}, Y_n^{k-1}) \in B^k].$$

Let $\{n\}$ be an increasing sequence of integers such that

$$(6.10) \qquad \liminf_{n\to\infty} P[A_n^k, S_n^k \le c] > 0.$$

For any such sequence we will prove that

$$(6.11) \qquad P[(Z_n^k, Y_n^k) \in \cdot \mid A_n^k, S_n^k \le c] \Rightarrow P[(U^k, V^k) \in \cdot].$$

Taken together with the fact that the definition (6.8) gives $P[(Z_n^k, Y_n^k) \in \cdot \mid A_n^k, S_n^k > c] = P[(U^k, V^k) \in \cdot]$, this implies

$$(6.12) \qquad P[(Z_n^k, Y_n^k) \in \cdot \mid A_n^k] \Rightarrow P[(U^k, V^k) \in \cdot]$$

whenever $\liminf_{n\to\infty} P[A_n^k] > 0$. In exactly the same way it will follow that $P[(Z_n^1, Y_n^1) \in \cdot] \Rightarrow P[(U^1, V^1) \in \cdot]$, and Lemma 6.1 will follow by simple induction.

It remains to prove that (6.10) implies (6.11). Fix k and B^k, let A_n^k be as in (6.9) and define a new sequence of processes $\tilde{X}_n(t) = X_n(S_n^k + t)$ with distribution given by

$$(6.13) \qquad \tilde{P}[\tilde{X}_n(\cdot) \in C] = P[X_n(S_n^k + \cdot) \in C \mid A_n^k, S_n^k \le c]$$

for Borel sets C in $D[0, \infty)$. Let $\tilde{\mathscr{F}}_n(t) = \mathscr{F}_n(S_n^k + t)$, and define a new sequence of partitions $\{\tilde{t}_n^j; j = 0, 1, \cdots\}_n$ by $\tilde{t}_n^0 = 0$; $\tilde{t}_n^j = t_n^{r_n(S_n^k)+j} - S_n^k$ for $j \ge 1$.

First we remark that S_n^k is a stopping time relative to $\{\mathscr{F}_n(t)\}$ (cf., (4.3) and (4.8)). It is then easy to see that each $\tilde{t}_n^j$ is a stopping time relative to $\{\tilde{\mathscr{F}}_n(t)\}$, and that if $\tilde{r}_n(u) = \max\{j \ge 0: \tilde{t}_n^j \le u\}$, then $\tilde{r}_n(u) = r_n(S_n^k + u) - r_n(S_n^k)$.

Now for each compact interval K in G use A1(K), A2(K) and A3(K) with t replaced by $t + c$. Then it follows easily that these same conditions continue to hold if one replaces $\{X_n(t), \mathscr{F}_n(t), t_n^j, r_n(t), P\}$ by $\{\tilde{X}_n(t), \tilde{\mathscr{F}}_n(t), \tilde{t}_n^j, \tilde{r}_n(t), \tilde{P}\}$. Here we use (6.10) to show that convergence to zero in P-probability of a function of $X_n(S_n^k + \cdot)$ implies convergence to zero in $\tilde{P}$-probability of the corresponding function of $\tilde{X}_n(\cdot)$. Also we use the fact that $A_n^k \cap [S_n^k \le c]$ is measurable with respect to $\tilde{\mathscr{F}}_n(\tilde{t}_n^j)$ for all $j \ge 0$.

Furthermore, A1 (cf., (2.11)) implies $\max_{0\le j\le r_n(c)} VX_n(k) \to_p 0$. Since the maximal jump of $X_n(\cdot)$ in [0, c] never exceeds $2\max_{0\le j\le r_n(c)} VX_n(k)$, it follows from this and the definition (6.7) of S_n^k that $\tilde{X}_n(0) = X_n(S_n^k) \to \lambda$ in $\tilde{P}$-probability. Thus A0 holds for $\tilde{X}_n(t)$ with $X(0) = \lambda$. The same argument applies (by A4(0)) if A1(C) holds for all compacts $C \subset H$ for some open $H \supset F$. (But note that A1(C) for all compacts $C \subset G$ will not suffice.)

Therefore, from Proposition 5.1 we have $\tilde{X}_n^K(t) \Rightarrow X_0^K(t)$, where $\tilde{X}_n^K(t)$ is defined from $\tilde{X}_n(t)$ as in (5.1)–(5.2) and $X_0^K(t) = X_0(t \wedge R_0^K)$ with $R_0^K = \inf\{t \ge 0: X_0(t) \notin \text{int } K\}$. This holds for any compact $K = [\alpha, \beta] \subset G$, and by the argument of the first paragraph of this section, it also holds when $K = [\alpha, \infty)$, since r_1 is inaccessible for $X_0(t)$. With $K = [\alpha, \infty)$ where $0 < \alpha < \delta$, put $\tilde{X}_n^\alpha(t) = \tilde{X}_n^K(t)$ and $X_0^\alpha(t) = X_0^K(t)$. Thus $\tilde{X}_n^\alpha(t) \Rightarrow X_0^\alpha(t)$.

Consider the mapping g from $D[0, \infty)$ into $R^1 \times R^1$ defined by $g(x(\cdot)) = (z, y)$, where

$$z = \inf\{t \geq 0 : x(t) \notin (\delta, \epsilon) \quad \text{or} \quad x(t-) \notin (\delta, \epsilon)\}$$

$$\text{(6.14)} \qquad y = \begin{cases} x(z-) & \text{if} \quad z < \infty \quad \text{and} \quad x(z-) \notin (\delta, \epsilon) \\ x(z) & \text{if} \quad z < \infty \quad \text{and} \quad x(z-) \in (\delta, \epsilon) \\ 0 & \text{if} \quad z = \infty. \end{cases}$$

This mapping is measurable by the argument around (4.7)–(4.8), and it is continuous on a Borel set B in $D[0, \infty)$ with $P[X_0^\alpha(\cdot) \in B] = 1$. This latter fact is seen as follows: With Z given by (6.6) we have $Z = \inf\{t \geq 0 : X_0^\alpha(t) \notin (\delta, \epsilon)\}$. Since δ and ϵ are regular points for the diffusion process $X_0^\alpha(t)$, there are (random) times $T > Z$, arbitrarily close to Z, such that $X_0^\alpha(T) \notin [\delta, \epsilon]$. In the terminology of [12] this means that $X_0^\alpha(t)$ satisfies condition C3 a.s., and the a.s. continuity of g follows from Lemma 3.3 in [12].

From the definitions (6.6)–(6.8), and from (6.13)–(6.14), we see that $(Z, Y) = g(X_0^\alpha(\cdot))$ and $(Z_n^k, Y_n^k) = g(X_n(S_n^k + \cdot)) = g(\tilde{X}_n^\alpha(\cdot))$. Therefore the $\tilde{P}$-distribution of (Z_n^k, Y_n^k) converges to the distribution of (Z, Y) by the continuous mapping theorem, and (6.11) follows. This completes the proof of Lemma 6.1.

LEMMA 6.2. *With $W^0 = 0$ and $W^k = \sum_{j=1}^k U^j$ for $k \geq 1$ we have*

$$\text{(6.15)} \qquad \limsup_{n\to\infty} P[\sup_{R_n^\delta \leq s \leq c} X_n(s) > \epsilon] \leq P[\cup_{k=1}^\infty [W^{k-1} \leq c, V^k = \epsilon]].$$

PROOF. It is clear from the definitions that

$$\text{(6.16)} \qquad \begin{aligned} P[\sup_{R_n^\delta \leq s \leq c} X_n(s) > \epsilon] &\leq P[\cup_{k=1}^\infty [S_n^k \leq c, Y_n^k \geq \epsilon]] \\ &\leq P[\cup_{k=1}^\infty [\textstyle\sum_{j=1}^{k-1} Z_n^j \leq c, Y_n^k \geq \epsilon]], \end{aligned}$$

since $[S_n^k \leq c] \subset [\sum_{j=1}^{k-1} Z_n^j \leq c]$. From Lemma 6.1 we have for all $m \geq 1$ that

$$\text{(6.17)} \qquad \limsup_{n\to\infty} P[\cup_{k=1}^m [\textstyle\sum_{j=1}^{k-1} Z_n^j \leq c, Y_n^k \geq \epsilon]] \leq P[\cup_{k=1}^m [W^{k-1} \leq c, V^k \geq \epsilon]],$$

since the sets involved are closed in R^{2m}. Also, the right-hand side of (6.17) is less than or equal to the right-hand side of (6.15), since $[V^k \geq \epsilon] = [V^k = \epsilon]$ a.s. To see that the left-hand side of (6.17) can be made arbitrarily close to the lim sup of the right-hand side of (6.16), we remark that

$$\text{(6.18)} \qquad \begin{aligned} &\limsup_{n\to\infty} P[\cup_{k=m+1}^\infty [\textstyle\sum_{j=1}^{k-1} Z_n^j \leq c, Y_n^k \geq \epsilon]] \\ &\leq \limsup_{n\to\infty} P[\cup_{k=m+1}^\infty [\textstyle\sum_{j=1}^{k-1} Z_n^j \leq c]] \\ &= \limsup_{n\to\infty} P[\textstyle\sum_{j=1}^m Z_n^j \leq c] \leq P[W^m \leq c], \end{aligned}$$

since each Z_n^j $(j = 1, 2, \cdots, m)$ is nonnegative. Now $W^m = \sum_{j=1}^m U^k$, where $U^1, U^2, \cdots$ are i.i.d. with $P[U^k > 0] > 0$. Therefore the right-hand side of (6.18) can be made arbitrarily small by choosing m large enough, and (6.15) follows from (6.16)–(6.18).

LEMMA 6.3. *For $\delta < \epsilon/2$ we have*

$$\text{(6.19)} \qquad \limsup_{n\to\infty} P[\sup_{R_n^\delta \leq s \leq c} X_n(s) > \epsilon] \leq 2\left(c + 2\int_\delta^\epsilon \frac{y\,dy}{\sigma^2(y)}\right)\left(\epsilon \int_\delta^{\epsilon/2} \frac{dy}{\sigma^2(y)}\right)^{-1}.$$

PROOF. From Lemma 6.2 the left-hand side of (6.19) is bounded above by

(6.20) $$\sum_{k=1}^{\infty} P[W^{k-1} \le c] P[V^k = \epsilon] = P[Y = \epsilon] \sum_{k=0}^{\infty} P[W^k \le c].$$

With Y defined in (6.6) we get from the theory of diffusion processes ([4], Theorem 16.27) that $P[Y = \epsilon] = (\lambda - \delta)/(\epsilon - \delta)$. Now look at the renewal process $\{W^k\}$ with $W^k = \sum_{j=1}^{k} U^j$. From standard renewal theory we have $\sum_{k=0}^{\infty} P[W^k \le c] = E(N + 1)$, where N is the number of renewals in the time interval $(0, c]$. Furthermore, since $N + 1$ is a stopping time for the process, Wald's identity gives $E(N + 1) \cdot E(U^1) = E(W^{N+1}) \le c + E(U^{N+1})$. It is not true that $E(U^{N+1}) = E(U^1)$, so to get a bound on $E(U^{N+1})$, we use Wald's identity once more:

$$\{E(U^{N+1})\}^2 \le E(\{U^{N+1}\}^2) \le E(\textstyle\sum_{k=1}^{N+1} \{U^k\}^2) = E(N + 1) \cdot E(\{U^1\}^2).$$

This, together with the first inequality gives $E(N + 1) \cdot E(U^1) \le c + (E(N + 1) \cdot E(\{U^1\}^2))^{1/2}$. Now if z, a and b are nonnegative numbers satisfying the inequality $z \le a + bz^{1/2}$, it is easy to show that $z \le 2a + b^2$. Therefore

(6.21) $$\sum_{k=0}^{\infty} P[W^k \le c] = E(N + 1) \le \frac{2c}{E(U^1)} + \frac{E(\{U^1\}^2)}{\{E(U^1)\}^2}.$$

Let $G(x, y)$ be the Green's function for the operator $\frac{1}{2}\, d^2/dx^2$ on the interval (δ, ϵ), i.e., $G(x, y) = 2(x - \delta)(\epsilon - y)/(\epsilon - \delta)$ for $\delta \le x \le y \le \epsilon$ and $G(x, y) = G(y, x)$. Since $M_1(\lambda) = E(U^1)$ and $M_2(\lambda) = E(\{U^1\}^2)$ are solutions of the equations $\frac{1}{2}\sigma^2(\lambda) M_1''(\lambda) + 1 = 0$ and $\frac{1}{2}\sigma^2(\lambda) M_2''(\lambda) + 2M_1(\lambda) = 0$, we find

(6.22) $$E(U^1) = \int_{\delta}^{\epsilon} G(\lambda, y) \frac{dy}{\sigma^2(y)}$$

(6.23) $$E(\{U^1\}^2) = 2 \int_{y=\delta}^{\epsilon} \int_{z=\delta}^{\epsilon} G(\lambda, y) G(y, z) \frac{dy}{\sigma^2(y)} \cdot \frac{dz}{\sigma^2(z)}.$$

From (6.22) we see that $E(U^1) \ge \epsilon(\lambda - \delta)(\epsilon - \delta)^{-1} \int_{\lambda}^{\epsilon/2} \sigma^{-2}(y)\, dy$ when $\delta < \lambda < \epsilon/2$, and using the inequality $G(y, z) \le 2(z - \delta) < 2z$ we find from (6.23) that $E(\{U^1\}^2) \le 4E(U^1) \cdot \int_{\delta}^{\epsilon} z\sigma^{-2}(z)\, dz$. Inserting this into (6.20) – (6.21) and letting $\lambda \downarrow \delta$ gives (6.19).

Since $r_0 = 0$ is accessible, we see by the inequality corresponding to (2.6) that $\int_0^{\epsilon} y\sigma^{-2}(y)\, dy < \infty$. On the other hand, since r_0 is exit, $|m(0)| = \infty$ gives $\int_0^{\epsilon/2} \sigma^{-2}(y)\, dy = \infty$. Therefore, by (6.19) we can choose δ so small that (6.3) holds.

This completes the proof of Theorem 2.1 for the case where r_0 is exit. When r_1 is exit and r_0 inaccessible, the proof is similar. When both boundaries are exit, we can argue as follows: let $K = [\alpha, \beta] \subset (r_0, r_1)$ and $K' = (-\infty, \beta]$, and let $X_n^K(t)$, $X^K(t)$, $X_n^{K'}(t)$ and $X^{K'}(t)$ be the usual stopped processes. Now the argument of this section shows that we can make the sequence $\{X_n^K(t)\}$ close to $\{X_n^{K'}(t)\}$ in the sense of (6.4) by choosing α close enough to r_0. Similarly we can make $\{X_n^{K'}(t)\}$ close to $\{X_n(t)\}$ in the same sense by taking β close to r_1. Therefore $\{X_n^K(t)\}$ is close to $\{X_n(t)\}$, and $X_n(t) \Rightarrow X(t)$ follows again from Theorem 4.2 in [2].

7. Proof of Theorem 3.3. To prove the direct part of Theorem 3.3, we once again take Proposition 4.1 as our point of departure. Also we need the following facts about the convergence $\Rightarrow_c$ (Proposition 4.4b and 4.5 in [13]).

PROPOSITION 7.1. (a) *Suppose that* $\{X_n(t)\}$ is tight in $D = D[0, \infty)$. *Suppose further that* $E\{g_j(X_n(t_j + \cdot)) \mid \mathscr{F}_n(t_j)\}$ $(j = 1, \cdots, m)$ *converge jointly in distribution to* $E\{g_j(X(t_j + \cdot)) \mid \mathscr{F}(t_j)\}$ $(j = 1, \cdots, m)$ *for all choices of* m, *of* t_j $(j = 1, \cdots, m)$, *and for all choices of* g_j $(j = 1, \cdots, m)$ *that are of the form* $g_j(x(\cdot)) = \exp\{\sum_{k=1}^{p_j} i\lambda_k^j x(s_k^j)\}$ where $i = (-1)^{1/2}$. *Then* $(X_n(t), \mathscr{F}_n(t)) \Rightarrow_c (X(t), \mathscr{F}(t))$.

(b) *Suppose that* $(X_n(t), \mathscr{F}_n(t)) \Rightarrow_c (X(t), \mathscr{F}(t))$. *Then the joint convergence in distributions described under* (a) *holds for any choice functions* g_j $(j = 1, \cdots, m)$ *that are* *bounded and measurable, and are continuous on some Borel set C in D with* $P[X(t_j + \cdot) \in C] = 1$ $(j = 1, \cdots, m)$.

Our object will be to generalize the proofs of the preceding two sections, in each case replacing the convergence $\Rightarrow$ in the conclusion by the stronger convergence $\Rightarrow_c$. The main work will be to extend the proof of Proposition 5.2 in this way.

Suppose first that the assumptions of Proposition 5.2 hold with $\sigma^2(\cdot) = 1$, and let $Y_n(t)$ be defined as in (5.5). From (5.6)–(5.8) it follows again that $(Y_n^0(t), \tilde{\mathscr{F}}_n(t)) \Rightarrow_c (W(t), \mathscr{F}_0(t))$ with the notation as defined in Section 5. Now let $g_j(x(\cdot)) = \exp\{\sum_{k=1}^{p_j} i\lambda_k^j x(s_k^j)\}$ and let $t_j \geq 0$ $(j = 1, \cdots, m)$. Without loss of generality we can take $t_1 \leq t_2 \leq \cdots \leq t_m$. From the identity $Y_n(s) = Y_n(0) + Y_n^0(s)$ we see that

$$E\{g_j(Y_n(t_j + \cdot)) \mid \tilde{\mathscr{F}}_n(t_j)\} = E\{g_j(Y_n^0(t_j + \cdot) - Y_n^0(t_j)) \mid \tilde{\mathscr{F}}_n(t_j)\} \exp\{\textstyle\sum_{k=1}^{p_j} i\lambda_k^j Y_n(t_j)\}. \tag{7.1}$$

Since $(Y_n^0(t), \tilde{\mathscr{F}}_n(t)) \Rightarrow_c (W(t), \mathscr{F}_0(t))$, the first factor on the right-hand side of (7.1) converges in distribution—hence in probability—to $E\{g_j(W(t_j + \cdot) - W(t_j)) \mid \mathscr{F}_0(t_j)\} = E\{g_j(W(t_j + \cdot) - W(t_j))\}$, which also equals $E\{g_j(X(t_j + \cdot) - X(t_j)) \mid \mathscr{F}(t_j)\}$. In Section 5 we proved that $Y_n(t) \Rightarrow X(t)$. Hence the second factors on the right-hand side of (7.1) converge jointly (for $j = 1, 2, \cdots, m$) to the corresponding factors $\exp\{\sum_{k=1}^{p_j} i\lambda_k^j X(t_j)\}$. Hence $E\{g_j(Y_n(t_j + \cdot)) \mid \tilde{\mathscr{F}}_n(t_j)\}$ $(j = 1, \cdots, m)$ converge jointly in distribution to $E\{g_j(X(t_j + \cdot)) \mid \mathscr{F}(t_j)\}$, and by Proposition 7.1 (a) we have $(Y_n(t), \tilde{\mathscr{F}}_n(t)) \Rightarrow_c (X(t), \mathscr{F}(t))$.

The next step is to show that this implies $(X_n^K(t), \mathscr{F}_n(t)) \Rightarrow_c (X^K(t), \mathscr{F}(t))$ where $X_n^K(\cdot) = f(Y_n(\cdot))$ and $X^K(\cdot) = f(X(\cdot))$ with $f = f_{\varphi,B}(\varphi \equiv 1)$. To this end we use the identity

$$E\{g_i(X_n^K(t_i + \cdot)) \mid \tilde{\mathscr{F}}_n(t_i)\} = E\{g_i \circ f(Y_n(t_i + \cdot)) \mid \tilde{\mathscr{F}}_n(t_i)\}. \tag{7.2}$$

By Proposition 4.2, f is continuous a.s. with respect to the distribution of $X(\cdot)$, so by Proposition 7.1 (b), any m random variables of the form (7.2) (with g_i continuous and bounded) converge jointly in distribution to the corresponding random variables of the form $E\{g_i \circ f(X(t_i + \cdot)) \mid \mathscr{F}(t_i)\} = E\{g_i(X^K(t_i + \cdot)) \mid \mathscr{F}(t_i)\}$. Therefore $(X_n^K(t), \tilde{\mathscr{F}}_n(t)) \Rightarrow_c (X^K(t), \mathscr{F}(t))$, and since $\tilde{\mathscr{F}}_n(t) = \mathscr{F}_n(t) \vee \sigma\{Y_n(s); s \leq t\} = \mathscr{F}_n(t) \vee \sigma\{W(s); 0 \leq s \leq t - S_n\}$ where $W(t)$ is independent of $X_n(t)$ and of the σ-fields $\mathscr{F}_n(t)$, this implies $(X_n^K(t), \mathscr{F}_n(t)) \Rightarrow_c (X^K(t), \mathscr{F}(t))$.

In all that has been proved so far, the convergence $\Rightarrow_c$ may be replaced by a formally stronger type of convergence: suppose that $(Z_n(t), \mathscr{G}_n(t)) \Rightarrow_c (Z(t), \mathscr{G}(t))$ and in addition that the following holds: for every choice of m, of g_i (bounded and continuous), T_{ni}(stopping times with respect to $\mathscr{G}_n(t)$), T_i (stopping times with respect of $\mathscr{G}(t)$) such that $(Z_n(T_{ni}), \cdots, Z_n(T_{nm})) \rightarrow_D (Z(T_1), \cdots, Z(T_m))$ and such that $\max_{1\leq i\leq m} \limsup_{n\to\infty} P[T_{ni} > c]$ can be made arbitrarily small by choosing c large enough, we have that the joint distribution of $E\{g_i(Z_n(T_{ni} + \cdot)) \mid \mathscr{G}_n(T_{ni})\}$ $(i = 1, \cdots, m)$ converges to the joint distribution of $E\{g_i(Z(T_i + \cdot)) \mid \mathscr{G}(T_i)\} (i = 1, \cdots, m)$. This could suitably be called strong convergence in conditional distributions: $(Z_n(t), \mathscr{G}_n(t)) \Rightarrow_{sc} (Z(t), \mathscr{G}(t))$.

To show that $(X_n(t), \mathscr{F}_n(t)) \Rightarrow_{sc} (W(t), \mathscr{F}(t))$ holds in the conclusion of Proposition 4.1, one has to go through the proof of the direct half of Theorem 2.9 in [13] again, replacing everywhere t_i in $X_n(t_i + \cdot)$ and in $\mathscr{F}_n(t_i)$ by T_{ni}, and replacing t_i in $W(t_i + \cdot)$ and in $\mathscr{F}(t_i)$ by T_i. This works, since we again may use a martingale central limit theorem to deduce convergence in distribution of $X_n(T_{ni} + s + u) - X_n(T_{ni} + s)$ to $W(T_i + s + u) - W(T_i + s)$. Here the boundedness in probability of $\{T_{ni}\}$ is used, e.g., to prove that (4.2) implies

$$\textstyle\sum_{k=r_n(T_{ni}+s)}^{r_n(T_{ni}+s+u)} \operatorname{Var}(\Delta_1 X_n(k) \mid \mathscr{F}_n(t_n^k)) \rightarrow_p u.$$

The proof rests heavily on the generalization of Proposition 7.1 (a) to the convergence $\Rightarrow_{sc}$ (with the similar replacements as above). This generalization is proved exactly like Proposition 4.5 in [13]. The corresponding extension of Proposition 7.1 (b) also holds. Going through the beginning of the present section again, we finally deduce that $(X_n^K(t), \mathscr{F}_n(t)) \Rightarrow_{sc} (X^K(t), \mathscr{F}(t))$ under the assumptions of Proposition 5.2 with $\sigma^2(\cdot) = 1$.

To proceed, we also need the following

LEMMA 7.2. *Let K be a compact interval on the line, $B = (\text{int } K)^c$ and let φ be a positive continuous function on K. For $x \in D = D[0, \infty)$ let q_x, r_x, $\tau_x(t)$ and f be defined as in* (4.3)–(4.6). *Suppose that x has the properties $x(u) = x(q_x)$ for $u > q_x$ and $x(q_x -) = x(q_x)$ if $x(q_x -) \in B$. Then we have $f(x)(u) = x(\tau_x(u))$ for all $u \geq 0$ if we define $\tau_x(u) = q_x + u - r_x$ for $u \geq r_x$. Also, for all $s, t \geq 0$ we can write*

$$f(x)(t + s) = f(y_{x,t})(s), \qquad \text{where} \quad y_{x,t}(u) = x(\tau_x(t) + u). \tag{7.3}$$

PROOF. The first identity is trivial from (4.6) and the properties of x. These properties are shared by $y = y_{x,t}$. Then, using the definitions (4.3)–(4.5) combined with $\tau_x(u) = q_x + u - r_x$ for $u \geq r_x$, we can verify that $\tau_x(t + s) = \tau_x(t) + \tau_y(s)$ for $t, s \geq 0$ by considering separately the cases $t + s < r_x$, $t < r_x \leq t + s$ and $r_x \leq t$. From this, $f(x)(t + s) = x(\tau_x(t) + \tau_y(s)) = y(\tau_y(s))$, and (7.3) follows.

We are now in the position to prove that the conclusion $(X_n^K(t), \mathscr{F}_n(t)) \Rightarrow_c (X^K(t), \mathscr{F}(t))$ holds in Proposition 5.2 for the general case where $\sigma^2(\cdot)$ is any positive continuous function on K. We use the same notation as in (5.9)–(5.18), and we will define $\tau_n(t) = Q_n^K + t - R_n^K$ for $t > R_n^K$. Then $\tau_n(\cdot) = T_n^{-1}(\cdot)$, and each $\tau_n(t)$ is optional with respect to $\mathscr{G}_n(u) = \mathscr{F}_n(T_n(u))$. Also, from the definition (5.11) it is easy to see that each $Z_n(\cdot)$, considered as a path in $D[0, \infty)$, has the properties mentioned in Lemma 7.2 (with $\varphi(\cdot) = \sigma^2(\cdot)$). Therefore, for $s, t_i \geq 0$ we have $X_n^K(t_i + s) = f(Z_n(\cdot))(t_i + s) = f(Z_n(T_{ni} + \cdot))(s)$, where $f = f_{\varphi,B}$ and $T_{ni} = \tau_n(t_i)$. Letting $g_i(i = 1, \cdots, m)$ be bounded continuous functions on $D[0, \infty)$, we find

$$E\{g_i(X_n^K(t_i + \cdot)) \mid \mathscr{F}_n(t_i)\} = E\{g_i \circ f(Z_n(T_{ni} + \cdot)) \mid \mathscr{G}_n(T_{ni})\}. \tag{7.4}$$

By the extended Proposition 5.2 for the case $\sigma^2(\cdot) \equiv 1$, we have that $(Z_n(t), \mathscr{G}_n(t)) \Rightarrow_{sc} (Z(t), \mathscr{G}(t))$, where $\mathscr{G}(t) = \sigma\{Z(s); s \leq t\}$. Furthermore, in Section 5 we proved that $X_n^K(t) \Rightarrow X^K(t) = f(Z(\cdot))(t)$. Therefore $(Z_n(T_{n1}), \cdots, Z_n(T_{nm})) = (X_n^K(t_1), \cdots, X_n^K(t_m)) \to_D (X^K(t_1), \cdots, X^K(t_m)) = (Z(T_1), \cdots, Z(T_m))$, where the random variables T_i are defined by $t_i = \int_0^{T_i} du/\sigma^2(Z(u))$ if $t_i < R$, $T_i = Q + t_i - R$ if $t_i \geq R$, with $Q = \inf\{u \geq 0: Z(u) \in B\}$ and $R = \int_0^Q du/\sigma^2(Z(u))$. Thus by the definition of the convergence $\Rightarrow_{sc}$, and by the corresponding extension of Proposition 7.1 (b), the random variables in (7.4) converge jointly in distribution to $E\{g_i \circ f(Z(T_i + \cdot)) \mid \mathscr{G}(T_i)\} = E\{g_i(X^K(t_i + \cdot)) \mid \mathscr{F}^K(t_i)\}$, where $\mathscr{F}^K(t_i) = \mathscr{G}(T_i) = \sigma\{X^K(s); s \leq t_i\}$. But here we may replace $\mathscr{F}^K(t_i)$ by $\mathscr{F}(t_i) = \sigma\{X(s); s \leq t_i\}$, and this completes the proof of the conclusion $(X_n^K(t), \mathscr{F}_n(t) \Rightarrow_c (X^K(t), \mathscr{F}(t))$ in Proposition 5.2.

Then it follows easily that the same conclusion holds under the assumptions of Proposition 5.1. To prove this, we use Definition 3.1, the proof of Proposition 5.1 in Section 5 and the fact that the scale function $u(\cdot)$ has a continuous inverse.

From Section 6 we see that for all $c, \epsilon > 0$ one can find a compact interval $K \subset G$ so large that $\lim \sup_{n\to\infty} P[\sup_{0\leq s\leq c} |X_n(s) - X_n^K(s)| > \epsilon] < \epsilon$. The proof of the direct part of Theorem 3.3 is completed by combining this result with our extended Proposition 5.1 and the following general result on the convergence $\Rightarrow_c$. (Compare Billingsley [1], Theorem 4.2).

PROPOSITION 7.3. *For each $k \geq 0$ and $n \geq 1$ let $X_n^k(t)$ and $X^k(t)$ be processes with paths in $D = D[0, \infty)$ that are adapted to families of σ-fields $\mathscr{F}_n(t)$ and $\mathscr{F}(t)$ respectively. Let d be some metric for the Stone topology on D. Suppose that $(X_n^k(t), \mathscr{F}_n(t)) \Rightarrow_c (X^k(t), \mathscr{F}(t))$ for each $k \geq 1$ and that $d(X^k(\cdot), X^0(\cdot)) \Rightarrow_p 0$ as $k \to \infty$. Suppose further that*

$$\lim_{k\to\infty} \lim \sup_{n\to\infty} P[d(X_n^k(\cdot), X_n^0(\cdot)) \geq \epsilon] = 0$$

for all $\epsilon > 0$. Then $(X_n^0(t), \mathscr{F}_n(t)) \Rightarrow_c (X^0(t), \mathscr{F}(t))$.

PROOF. Let $g_i (i = 1, \cdots, m)$ be bounded uniformly continuous functions on D. We use first the estimate

$$(7.5)\quad E\,|\,E\{g_i(X_n^k(t_i + \cdot))\,|\,\mathscr{F}_n(t_i)\} - E\{g_i(X_n^0(t_i + \cdot))\,|\,\mathscr{F}_n(t_i)\}\,| \\ \le \rho_i(\epsilon) + P[d(X_n^k(\cdot), X_n^0(\cdot)) \ge \epsilon]\cdot 2\,\|g_i\|,$$

then the corresponding estimate of the difference between $E\{g_i(X^k(t_i + \cdot))\,|\,\mathscr{F}(t_i)\}$ and $E\{g_i(X^0(t_i + \cdot))\,|\,\mathscr{F}(t_i)\}$. Here $\rho_i(\cdot)$ is a modulus of continuity of the function g_i. The lim sup (as $n \to \infty$) of the right-hand side of (7.5) can be made arbitrarily small by choosing ϵ small and then k large. Therefore the convergence in joint distribution of $E\{g_i(X_n^0 \cdot (t_i + \cdot))\,|\,\mathscr{F}_n(t_i)\} (i = 1, \cdots, m)$ to $E\{g_i(X^0(_i + \cdot))\,|\,\mathscr{F}(t_i)\} (i = 1, \cdots, m)$ follows from the corresponding convergence of $E\{g_i(X_n^k(t_i + \cdot))\,|\,\mathscr{F}_n(t_i)\}$ to $E\{g_i(X^k(t_i + \cdot))\,|\,\mathscr{F}(t_i)\}$. This implies the conclusion by Proposition 4.4a in [13].

Thus we can conclude that $(X_n(t), \mathscr{F}_n(t)) \Rightarrow_c (X(t), \mathscr{F}(t))$ follows from the assumptions of the direct part of Theorem 3.3. By essentially the same proof we can in fact show that $(X_n(t), \mathscr{F}_n(t)) \Rightarrow_{sc} (X(t), \mathscr{F}(t))$ with $\Rightarrow_{sc}$ as defined above.

Conversely, suppose that $(X_n(t), \mathscr{F}_n(t)) \Rightarrow_c (X(t), \mathscr{F}(t))$. Then in particular $X_n(t) \Rightarrow X(t)$, and A0, A4(0) and A4(1) follow trivially. Furthermore, the function $g: D \to R^1$ defined by $g(x(\cdot)) = \sup_{0 \le s,u \le t, |s-u| \le \delta} |\,x(s) - x(u)\,|$ for fixed $\delta, t > 0$ is continuous a.s. with respect to the distribution of $X(\cdot)$, and $P[g(X(\cdot)) = \epsilon] = 0$ for ϵ small enough, so that

$$\lim_{n\to\infty} P[\sup_{0\le s,u \le t; |s-u|\le\delta} |\,X_n(s) - X_n(u)\,| > \epsilon] = P[\sup_{0\le s,u\le t; |s-u|\le\delta} |\,X(s) - X(u)\,| > \epsilon].$$

Letting $\delta \downarrow 0$, we see that (2.12) holds for all $\epsilon, t > 0$, and therefore the condition A1 holds for any sequence of partitions (see Section 3 in [13]).

Let K be a compact interval in G. To prove A2(K) and A3(K) for some sequence of partitions we first define

$$(7.6)\quad Z_n^1(t, \delta, K) = \textstyle\sum_{k=0}^{[t/\delta]} |\,E(\Delta_1^\delta X_n(k)\,|\,\mathscr{F}_n(k\delta)) - \delta\mu(X_n(k\delta))\,| \cdot I(X_n(k\delta) \in K)$$

$$(7.7)\quad Z_n^2(t, \delta, K) = \textstyle\sum_{k=0}^{[t/\delta]} |\,E(\{\Delta_1^\delta X_n(k)\}^2\,|\,\mathscr{F}_n(k\delta)) - \delta\sigma^2(X_n(k\delta))\,| \cdot I(X_n(k\delta) \in K),$$

where $\Delta_1^\delta X_n(k) = \{X_n((k+1)\delta) - X_n(k\delta)\}\cdot I(\,|\,X_n((k+1)\delta) - X_n(k\delta)\,| \le 1)$ and $[\cdot]$ denotes integral part. The proof will be completed if we can find a sequence $\{\delta_n\}$ with $\delta_n \downarrow 0$ such that $Z_n^j(t, \delta_n, K) \to_p 0 (j = 1, 2)$ for all $t > 0$ and for all compact intervals $K \subset G$.

We let $Z^j(t, \delta, K) (j = 1, 2)$ denote the random variables obtained by replacing $X_n(\cdot)$ by $X(\cdot)$ and $\mathscr{F}_n(\cdot)$ by $\mathscr{F}(\cdot)$ in (7.6) and (7.7). Then from Proposition 7.1 (b) it is immediate that $Z_n^j(t, \delta, K) (j = 1, 2)$ converge jointly in law to $Z^j(t, \delta, K) (j = 1, 2)$ for all $t, \delta > 0$ and all K. In particular, for all $\epsilon > 0$

$$(7.8)\quad \limsup_{n\to\infty} P[Z_n(t, \delta, K) \ge \epsilon] \le P[Z(t, \delta, K) \ge \epsilon],$$

where $Z_n(t, \delta, K) = \max(Z_n^1(t, \delta, K), Z_n^2(t, \delta, K))$ and $Z(t, \delta, K) = \max(Z^1(t, \delta, K), Z^2(t, \delta, K))$.

As already remarked in Section 2, the convergence in (2.2) and (2.3) is uniform on compacts in G. From this it is straightforward to verify that $Z^j(t, \delta, K) \to_p 0 (j = 1, 2; t > 0; K \subset G)$ as $\delta \downarrow 0$, so from (7.8)

$$(7.9)\quad \lim_{\delta\downarrow 0} \limsup_{n\to\infty} P[Z_n(t, \delta, K) \ge \epsilon] = 0$$

for all $\epsilon, t > 0$ and all $K \subset G$. Let $\{K_k\}_k$ be an increasing sequence of compact intervals such that $\cup_{k=1}^\infty K_k = G$. Then from (7.9) it follows that for each $k \ge 1$ we can find a $\delta_k^0 > 0$ such that $\limsup_{n\to\infty} P[Z_n(k, \delta_k^0, K_k) \ge k^{-1}] < k^{-1}$. Hence we can find $n_k \ge 1$ such that $P[Z_n(k, \delta_k^0, K_k) \ge k^{-1}] < k^{-1}$ for $n \ge n_k$. Furthermore, the sequences $\{\delta_k^0\}$ and $\{n_k\}$ can be chosen such that $\delta_k^0 \downarrow 0$ and $n_k \uparrow \infty$ as $k \to \infty$. If we now define $\delta_n = \delta_k^0$ for $n_k \le n < n_{k+1}$ $(k = 1, 2, \cdots)$, it follows that

$$P\{Z_n(t, \delta_n, K) \geq \epsilon] \to 0 \tag{7.10}$$

for all ϵ, $t > 0$ and all compacts $K \subset G$. (Recall from (7.6) and (7.7) that $Z_n(t, \delta, K)$ cannot decrease if t increases or K expands. Therefore the left-hand side of (7.10) is less than k^{-1} if k is so large that $k \geq t$, $\epsilon \leq k^{-1}$ and $K_k \supset K$, and if we then take $n \geq n_k$.) But (7.10) implies $Z_n^j(t, \delta_n, K) \to_p 0 (j = 1, 2;\ t > 0,\ K(\text{compact}) \subset G)$, and therefore A2($K$) and A3($K$) hold for the sequence of partitions $\{t_n^k = k\delta_n;\ k = 0, 1, \cdots\}_n$. This completes the proof.

If the limiting process $X(t)$ has no accessible boundaries, and if (say) $X_n(t) \in G$ a.s. for all n and t, then by a simple modification of the above proof, we can show that A2 and A3 also hold for a sequence of partitions of the same form. We let $Z_n^j(t, \delta) (j = 1, 2)$ be defined similarly to (7.6)–(7.7) without the factor $I(X_n(k\delta) \in K)$, and let $Z^j(t, \delta)$ be defined correspondingly. Let η, $t > 0$ be given. Then, since both boundaries of G are inaccessible for the process $X(t)$, we can find a compact interval $K \subset G$ such that $P[X(s) \in K$ for all s in $[0, t]] > 1 - \eta$. From this, combined with the uniform convergence on compacts in (2.2)–(2.3), it is easy to see that $\max(Z^1(t, \delta), Z^2(t, \delta)) \to_p 0$ as $\delta \downarrow 0$. The rest goes as above.

8. Two examples.

(a) *Branching processes in random environments.* Let $\{\zeta_k\}_k$ be a sequence of independent, identically distributed (i.i.d.) random variables taking values in some measurable space $(E, \mathscr{E})$. Let $\mathscr{M}$ be the σ-fields generated by $\{\zeta_k\}$, and for each $n \geq 1$ and $\zeta \in E$ let $\{p_n(i, \zeta);\ i = 0, 1, \cdots\}$ be a probability distribution on the nonnegative integers. Let z_n be a positive integer. For each $n \geq 1$ we can define a process $Z_n(k)$ by $Z_n(0) = z_n$ and for $k \geq 1$, $n \geq 1$

$$Z_n(k + 1) = \sum_{j=1}^{Z_n(k)} \xi_n(k, j), \tag{8.1}$$

where $\{\xi_n(k, j);\ k = 0, 1 \cdots, j = 1, 2, \cdots\}$ is a collection of random variables that are conditionally independent given the σ-field $\mathscr{M}$, and where $P[\xi_n(k, j) = i \mid \mathscr{M}] = p_n(i, \zeta_k)$ for $n, j \geq 1$ and $k, i \geq 0$. Then each $Z_n(k)$ is a branching process in random environments (BPRE) as defined by Smith and Wilkinson [31].

Suppose that $z_n/n \to x_0 > 0$ as $n \to \infty$, and define a sequence of processes $\{X_n(t)\}$ by $X_n(t) = Z_n([nt])/n$. We want to show that under certain conditions, $\{X_n(t)\}$ converges weakly in $D[0, \infty)$ to a diffusion process $X(t)$. This will generalize results on diffusion approximation of Galton-Watson processes that were proposed by Feller [9], and proved by Jiřina [14] and others. The present diffusion approximation was suggested without proof by Keiding [15]. The reader is referred to Keiding's paper for further discussion.

After the first version of the present paper was completed, the author was informed about an independent work by Kurtz [18] (based on [17]), where essentially the same results were proved and generalized to the case of weakly dependent environments. Nevertheless, our proof will be indicated here, both because it illustrates how to use our general results, and because some of our moment conditions are slightly weaker than those of Kurtz.

Denote the expectation and the variance of the off-spring distribution, conditioned upon the environment ζ by

$$\mu_n(\zeta) = \sum_{i=0}^{\infty} i\, p_n(i, \zeta) \tag{8.2}$$

$$\sigma_n^2(\zeta) = \sum_{i=0}^{\infty} (i - \mu_n(\zeta))^2 p_n(i, \zeta). \tag{8.3}$$

We make the following assumptions on the sequence of offspring distributions

$$E\{\mu_n(\zeta)\} = 1 + \alpha/n + o(n^{-1}) \tag{8.4}$$

$$E\{\mu_n(\zeta) - 1)^2\} = \omega^2/n + o(n^{-1}) \tag{8.5}$$

$$E\{\sigma_n^2(\zeta)\} = \tau^2 + o(1) \tag{8.6}$$

$$E\{\sum_{i=[n\epsilon]}^{\infty} i^2 p_n(i, \zeta)\} = o(1) \qquad \text{for all} \quad \epsilon > 0 \tag{8.7}$$

(8.8) $$E\{(\mu_n(\zeta)-1)^2;\,|\mu_n(\zeta)-1|>\epsilon\}=o(n^{-1}) \qquad \text{for all} \quad \epsilon>0$$

(8.9) $$E\{\sigma_n^2(\zeta);\,\sigma_n^2(\zeta)>n\,\epsilon\}=o(1) \qquad \text{for all} \quad \epsilon>0.$$

Here α is a real parameter, ω and τ are nonnegative and $\omega^2+\tau^2>0$.

THEOREM 8.1. *Under the conditions* (8.4)–(8.9), $\{X_n(t)\}$ *converges weakly in* $D[0,\infty)$ *as* $n\to\infty$ *to a diffusion process* $X(t)$, *started at* $X(0)=x_0$, *with drift coefficient* $\mu(x)=\alpha x$ *and with diffusion coefficient* $\sigma^2(x)=\tau^2x+\omega^2x^2$. *Also* $X_n(t)\Rightarrow_c X(t)$.

REMARKS. The state space of the diffusion process may be taken as $(0,\infty)$ if $\tau^2=0$, $[0,\infty)$ if $\tau^2>0$; in the latter case 0 is an exit boundary. The conditions for convergence (8.4)–(8.6) are the same as those proposed by Keiding. (8.7)–(8.9) are weak additional conditions. Both (8.7) and (8.9) follow if the expected third moment of the offspring distribution is $o(n^{1/2})$.

PROOF. We will need a general inequality concerning sums of random variables. Let $\{Y_j\}$ be a sequence of i.i.d. random variables with $\mu=E(Y_1)$ and $\sigma^2=\operatorname{Var} Y_1$. Put $S_r=\sum_{j=1}^{r}Y_j$ and let $\epsilon,\lambda,a>0$ be arbitrary. Then there is a universal constant $K>0$ such that

(8.10) $$n^{-1}\max_{1\le r\le na}E(S_r^2;\,|S_r|\ge n\epsilon)\le K\{\lambda\sigma^2+aE(Y_1^2;\,|Y_1|\ge n\epsilon/8)\}$$
$$+\,4na^2\mu^2I(|\mu|\ge\epsilon/2a)+4a\sigma^2I(\sigma^2\ge\lambda n\epsilon^2/4a^2).$$

This inequality can be derived as follows: first consider the case $\mu=0$. Then from (12.20) in Billingsley [2] (there is a minor error in that inequality: $\alpha/4$ should be $\alpha^{1/2}/2$) we deduce that the left-hand side of (8.10) is bounded above by $K'\{a^2\sigma^4/n\epsilon^2+aE(Y_1^2;\,|Y_1|\ge n\epsilon/2)$ for some universal constant K'. This bound is used for $a^2\sigma^4\le\lambda\sigma^2n\epsilon^2$, otherwise we use the trivial bound $a\sigma^2$. Finally, we generalize to $\mu\ne0$ by using the inequality preceding (12.20) in [2], the inequality $S_r^2\le2(S_r-r\mu)^2+2(r\mu)^2$ and the inequality in Dvoretzky [7], Lemma 3.3.

The sequence of processes $\{X_n(t)\}$ satisfies A0 and A4(0) trivially. By Theorem 2.1 it is enough to verify that A1(K), A2(K) and A3(K) hold for all intervals $K=[0,a]$ $(a>0)$. We take $t_n^k=k/n$. Using the Markov property of the processes $Z_n(k)$, we see that it is enough to prove (cf. Remark 5 in Section 2) that for all $a,t,\epsilon>0$ we have

(8.11) $$\max_{r\le na,k\le nt}|E\{Z_n(k+1)-Z_n(k)\,|\,Z_n(k)=r\}-\alpha r/n|\to0$$

(8.12) $$\max_{r\le na,k\le nt}|n^{-1}E\{(Z_n(k+1)-Z_n(k))^2\,|\,Z_n(k)=r\}-\tau^2r/n-\omega^2r^2/n^2|\to0$$

(8.13) $$\max_{r\le na,k\le nt}n^{-1}E\{(Z_n(k+1)-Z_n(k))^2;\,|Z_n(k+1)-Z_n(k)|>n\epsilon\,|\,Z_n(k)=r\}\to0.$$

By (8.1)–(8.3) it is easy to see that

$$E\{Z_n(k+1)-Z_n(k)\,|\,Z_n(k)=r\}=rE\{\mu_n(\zeta)-1]$$
$$E\{(Z_n(k+1)-Z_n(k))^2\,|\,Z_n(k)=r\}=r^2E\{(\mu_n(\zeta)-1)^2\}+rE\{\sigma_n^2(\zeta)\}.$$

Thus (8.11) and (8.12) follow from (8.4)–(8.6).

To prove (8.13), we condition upon $\mathscr{M}$ and use the estimate (8.10) with $Y_j=\xi_n(k,j)-1$. The right-hand side of (8.10) will then depend upon ζ_k. If we take expectation and use (8.6)–(8.9) in letting $n\to\infty$, we find that the upper limit of the left-hand side of (8.13) is at most $K\lambda\tau^2$. λ being arbitrarily small, this proves (8.13).

This same proof can also be used to prove diffusion approximaton theorems for more general models, where the off-spring distribution, in addition to depending upon the environment ζ, also depends upon the size of the present generation.

(b). *A convergent sequence of diffusion processes whose drift coefficient diverge.* (Rosenkrantz [28], [29].) Let $b(x)$ be a continuous function on R^1 such that $\int_{-\infty}^{\infty}|b(x)|\,dx$

$< \infty$ and $\int_{-\infty}^{\infty} b(x)\,dx = 0$. Let $X(t)$ be a diffusion process on $F = R^1$ with $X(0) = 0$, diffusion coefficient $\sigma^2(x) = 1$ and drift coefficient $b(x)$. Put $X_n(t) = n^{-1}X(n^2t)$. In [29] Rosenkrantz proved that $X_n(t) \Rightarrow W(t)$, standard Brownian motion. Since the sequence of drift coefficients $\mu_n(x) = nb(nx)$ need not converge here, one might question whether or not this can be strengthened to $X_n(t) \Rightarrow_c W(t)$; in particular one might ask whether the conditions A2(K) hold here. In fact these and the other conditions of Theorem 2.1 (Theorem 3.3) are satisfied for some sequence of partitions $\{t_n^k = k\delta_n\}(\delta_n \downarrow 0)$.

THEOREM 8.2. *Under the assumptions above we have* $X_n(t) \Rightarrow_c W(t)$.

PROOF. One can verify the conditions of Theorem 3.3 directly, but the following approach is simpler. Define a function u by $u(0) = 0$ and $u'(x) = \exp\{-\int_{-\infty}^{x} b(y)\,dy\}$ (cf., (2.4)), and let $U(t) = u(X(t))$, $U_n(t) = n^{-1}U(n^2t) = n^{-1}u(nX_n(t))$. We start by proving that $U_n(t) \Rightarrow_c W(t)$. To this end it is obviously enough to prove that for some sequence $\{\delta_n\}$ with $\delta_n \downarrow 0$ we have uniformly on compacts

(8.14) $$\delta_n^{-1}E^y(U_n(\delta_n) - y) = (n\delta_n)^{-1}E^{ny}(U(n^2\delta_n) - ny) \to 0$$

(8.15) $$\delta_n^{-1}E^y\{(U_n(\delta_n) - y)^2\} = n^{-2}\delta_n^{-1}E^{ny}\{(U(n^2\delta_n) - ny)^2\} \to 1$$

(8.16) $$\delta_n^{-1}E^y\{|U_n(\delta_n) - y|^3\} = n^{-3}\delta_n^{-1}E^{ny}\{|U(n^2\delta_n) - ny|^3\} \to 0$$

(8.17) $$\delta_n^{-1}P_n^y[\sup_{0\le s\le\delta_n}|U_n(s) - y| > \epsilon] \to 0 \qquad \text{for all} \quad \epsilon > 0.$$

Now $U(t)$ is a diffusion process on natural scale, so we have the representation (Breiman [4], Chapter 16)

(8.18) $$U(t) = W(T(t)); \text{ where } t = \int_0^{T(t)} ds/\tilde{\sigma}^2(W(s)),$$

and $\tilde{\sigma}^2(\cdot)$ is given by $\tilde{\sigma}^2(y) = u'(x)^2$ with $y = u(x)$. From the condition $\int_{-\infty}^{\infty} b(x)\,dx = 0$, we see that $u'(x) \to 1$ when $|x| \to \infty$. Hence $\tilde{\sigma}^2(y) \to 1$ when $|y| \to \infty$, and in particular $A^{-1} \le \tilde{\sigma}^2(y) \le A$ for some constant $A > 1$. Therefore $A^{-1}t \le T(t) \le At$. Since $W(t)$ and $W(t)^2 - t$ are martingales, we get $E^u(U(t) - u) = 0$ and $E^u\{(U(t) - u)^2\} = E^{W(0)=u}\{T(t)\}$ by (8.18) and the optional sampling theorem. Therefore (8.14) holds, and (8.15) follows by the following estimate: let $\epsilon > 0$, and let $c > 0$ be so large that $|\tilde{\sigma}^{-2}(y) - 1| \le \epsilon/A$ when $|y| > c$. Then (8.18) gives

(8.19) $$|T(t) - t| \le \epsilon t + A\int_0^{At} I(|W(s)| \le c)\,ds.$$

Since $P^u[|W(s)| \le c] \le P^0[|W(s)| \le c]$, we can choose $K > 0$ so large that $P^u[|W(s)| \le c] \le \epsilon$ for $s \ge K$ and for all u. Hence (8.19) gives $|E^{W(0)=u}\{T(t)\} - t| \le \epsilon(1 + A^2)t + A^2K$, and the left-hand side of (8.15) is bounded above by $\epsilon(1 + A^2) + n^{-2}\delta_n^{-1}A^2K$. Thus (8.15) holds if $\delta_n^{-1} = o(n^2)$.

Finally, the left-hand side of (8.16) is of the order $O(\sqrt{\delta_n})$ by the estimate $E^u\{|U(t) - u|^3\} \le \sup_{s\le At} E^u\{|W(s) - u|^3\}$, and (8.17) follows from (8.16) by Kolmogorov's inequality for martingales.

This shows that $U_n(t) \Rightarrow_c W(t)$, and by Proposition (4.9) in [13] (referred in Section 3 above) $X_n(t) \Rightarrow_c W(t)$ will follow if we can prove that $\sup_{s\le t}|X_n(s) - U_n(s)| \to_p 0$ for all $t > 0$. Let $v(\cdot) = u^{-1}(\cdot)$. Then $v'(y) \to 1$ as $|y| \to \infty$, and for all $\epsilon > 0$ one can find $K > 0$ such that $|v(y) - y| \le K + \epsilon|y|$ holds for all y. Thus $|X_n(s) - U_n(s)| \le Kn^{-1} + \epsilon|U_n(s)|$, and this completes the proof, since $U_n(t) \Rightarrow W(t)$.

NOTE. The results of this paper have now been generalized to cases where the limiting diffusion may have regular boundaries and to cases where the coefficients of the diffusion can have simple discontinuities. Also, a modified set of conditions should be mentioned:

Under weak assumptions it is possible to replace $\Delta t_n(k)$ in A2(K) and A3(K) by $E\{\Delta t_n(k) \mid \mathscr{F}_n(t_n^k)\}$. This is convenient for instance when the sequence under investigation consists of pure jump Markov processes in continuous time.

Acknowledgments. I am grateful to T. G. Kurtz for stimulating discussions during his visit to Bergen, and to a referee on an earlier version of this paper for several valuable comments.

REFERENCES

[1] ALDOUS, D. J. (1979). A concept of weak convergence for stochastic processes viewed in the Strasbourg manner. Preprint, Statist. Laboratory, Univ. Cambridge.
[2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.
[3] BOROVKOV, A. A. (1970). Theorems on the convergence to Markov diffusion processes. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* **16** 47–76.
[4] BREIMAN, L. (1968). *Probability.* Addison-Wesley, Reading.
[5] DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.
[6] DURRETT, R. and RESNICK, S. I. (1978). Functional limit theorems for dependent variables. *Ann. Probability* **6** 829–846.
[7] DVORETZKY, A. (1972). Asymptotic normality for sums of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **II** 513–535. Univ. California Press.
[8] ETHIER, S. N. (1978). Differentiability preserving properties of Markov semigroups associated with one-dimensional diffusions. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* **45** 225–238.
[9] FELLER, W. (1951). Diffusion processes in genetics. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 227–246. Univ. California.
[10] GIKHMAN, I. I. and SKOROKHOD, A. V. (1969). *Introduction to the Theory of Random Processes.* Saunders, Philadelphia.
[11] GUESS, H. A. and GILLESPIE, J. H. (1977). Diffusion approximations to linear stochastic difference equations with stationary coefficients. *J. Appl. Probability* **14** 58–74.
[12] HELLAND, I. S. (1978). Continuity of a class of random time transformations. *Stochastic Processes Appl.* **7** 79–99.
[13] HELLAND, I. S. (1980). On weak convergence to Brownian motion. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete.* **52** 251–265.
[14] JIŘINA, M. (1969). On Feller's branching diffusion processes. *Časopis Pěst. Mat.* **94** 84–90.
[15] KEIDING, N. (1975). Extinction and exponential growth in random environments. *Theor. Population Biology* **8** 49–63.
[16] KIMURA, M. (1964). Diffusion models in population genetics. *J. Appl. Probability* **1** 177–232.
[17] KURTZ, T. G. (1975). Semigroups of conditioned shifts and approximation of Markov processes. *Ann. Probability* **3** 618–642.
[18] KURTZ, T. G. (1977). Diffusion approximations for branching processes. Preprint, Univ. Wisconsin-Madison.
[19] KURTZ, T. G. (1978). Representations of Markov processes as multiparameter time change. Séries de Mathématiques Pures et Appliquées. Institut de Recherche Mathématique Avancée. Univ. Louis Pasteur, Strasbourg.
[20] KUSHNER, H. J. (1974). On the weak convergence of interpolated Markov chains to a diffusion. *Ann. Probability* **2** 40–50.
[21] LINDVALL, T. (1973). Weak convergence of probability measures and random functions in the function space $D[0, \infty)$. *J. Appl. Probability* **10** 109–121.
[22] MANDL, P. (1968). *Analytic Treatment of One-dimensional Markov Processes.* Springer, Berlin.
[23] MCLEISH, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probability* **2** 620–628.
[24] MORKVĖNAS, R. (1975). Weak convergence of random processes to the solution of a martingale problem. *Lithuanian Math. J.* **15** 247–253.
[25] NORMAN, M. F. (1972). *Markov Processes and Learning Models.* Academic, New York.
[26] NORMAN, M. F. (1975). Diffusion approximation of non-Markovian processes. *Ann. Probability* **3** 358–364.
[27] REBOLLEDO, R. (1979). La methode des martingales appliquee à l'etude de la convergence en loi de processus. *Mém. Soc. Math. France* **62**.
[28] ROSENKRANTZ, W. A. (1974). A convergent family of diffusion processes whose diffusion coefficients diverge. *Bull. Amer. Math. Soc.* **80** 973–976.
[29] ROSENKRANTZ, W. A. (1975). Limit theorems for solutions to a class of stochastic differential equations. *Indiana Univ. Math. J.* **24** 613–625.

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

[31] SMITH, W. L. and WILKINSON, W. E. (1969). On branching processes in random environments. *Ann. Math. Statist.* **40** 814–827.

[32] STONE, C. (1963). Weak convergence of stochastic processes defined on semi-infinite time intervals. *Proc. Amer. Math. Soc.* **14** 694–696.

[33] STROOCK, D. W. and VARADHAN, S. R. S. (1969). Diffusion processes with continuous coefficients I. *Comm. Pure Appl. Math.* **22** 345–400.

[34] STROOCK, D. W. and VARADHAN, S. R. S. (1969). Diffusion processes with continuous coefficients II. *Comm. Pure Appl. Math.* **22** 479–530.

AGRICULTURAL UNIVERSITY OF NORWAY
DEPARTMENT OF MATHEMATICS AND STATISTICS
1432 AAS-NLH
NORWAY