A NEW CLASS OF PROBABILITY LIMIT THEOREMS, II

By

JOHN LAMPERTI

TECHNICAL REPORT NO. 7

March 2, 1961

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT G-9669
A NEW CLASS OF PROBABILITY LIMIT THEOREMS, II

by

John Lamperti

TECHNICAL REPORT NO. 7

March 2, 1961

PREPARED UNDER THE AUSPICES

OF

NATIONAL SCIENCE FOUNDATION GRANT G-9669

APPLIED MATHEMATICS AND STATISTICS LABORATORIES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
Preface

The first paper with the title "A new class of probability limit theorems" appeared as Technical Report No. 1 in this series. The present report is a successor rather than a sequel to that one, for all the previous results have been included here in §2, §7 and the Appendix. It is felt that a much more complete understanding of the phenomena in question has now been attained.

An abstract of this paper will appear shortly in the "research reports" section of the Bulletin of the American Mathematical Society.
A NEW CLASS OF PROBABILITY LIMIT THEOREMS

by

John Lamperti

1. Introduction.

The study of the asymptotic behavior of \( S_n = \xi_1 + \cdots + \xi_n \), where \( \xi_i \) are independent, identically distributed random variables with mean 0 and finite variance, is an important part of probability theory and one which has attained a considerable degree of perfection. The purpose of this paper is to begin to develop a theory of another rather extensive class of Markov processes in a somewhat similar manner.

This vague statement can be clarified with an example: suppose that \( \{X_n\} \) is a random walk on the non-negative integers with the transition probabilities

\[
(1.1) \quad P_{i,i+1} = \frac{1}{2} \left[ 1 + \alpha \frac{1}{i} + o\left(\frac{1}{i}\right) \right], \quad P_{i,i-1} = 1 - P_{i,i+1}, \quad i \neq 0, \\
P_{01} = 1 - P_{00} > 0, \quad P_{i,j} = 0 \text{ otherwise.}
\]

We shall prove in \$2\) that provided \( \alpha > -\frac{1}{2} \), so that \( \{X_n\} \) is null-recurrent or transient [6],

\[
(1.2) \quad \lim_{n \to \infty} \Pr \left( \frac{X_n}{\sqrt{n}} \leq y \right) = \int_0^y \frac{2^\alpha \frac{1}{2} \xi^\frac{1}{2}}{2^{\alpha+1/2} \Gamma(\alpha+1/2)} \, d\xi, \quad y \geq 0.
\]
This is our analogue of the central limit theorem giving the limit distribution of $S_n/\sqrt{n}$.

There are, of course, a large number of additional limit theorems for $\{S_n\}$. Many (but by no means all) of these are subsumed by the Erdos-Kac-Donsker invariance principle [3,2] which says that in a certain sense, the stochastic process $(S_i/\sqrt{n}, i = 0, 1, 2, \ldots)$ converges to a Brownian motion (Wiener) process as $n \to \infty$; a large class of path functionals, such as $\max(S_1/\sqrt{n}, \ldots, S_n/\sqrt{n})$, thus are shown to have limiting distributions. The greater part of the present paper is devoted to showing that a similar convergence occurs in the theory of processes like the random walks above. The limiting process is a Wiener process with reflecting barrier at the origin if $\alpha = 0$, but for other values of $\alpha$, the limit is a different type of diffusion. We thus obtain many limit-distribution theorems for functionals of $\{X_1/\sqrt{n}, \ldots, X_n/\sqrt{n}\}$.

There is much general theory which bears upon the problems we shall investigate. Khintchine in [9] proves a theorem which suggests that limits like that in (1.2) exist and satisfy a certain diffusion equation. The hypotheses of his theorem can not be satisfied in our cases, however, as we shall be led to singular (except when $\alpha = 0$) diffusion problems with reflecting barriers. More recently, Donsker's work has been much extended by Billingsley [1], Prohorov [12], Skorohod [14,15] and others. The first of these references proves convergence to Wiener process for sums of certain dependent variables. In [12] we
learn that often convergence of finite-dimensional distributions, plus an additional condition, implies weak convergence of a sequence of stochastic processes to a limit (invariance principle). Skorohod's work in [14] generalizes that of [12], and in [15] he further shows that under certain conditions convergence of the infinitesimal generators and initial distributions of a sequence of Markov processes implies an invariance principle. It might be possible to take this approach to obtain or improve the results to follow here, but again apparently not all of Skorohod's hypotheses are valid in our cases. In addition, his theorems are far from simple, and for the present purpose, methods will be used which depend mostly upon nothing deeper than the moment-convergence theorem.

The plan of our paper is as follows: §2 introduces the class of processes we shall investigate, and contains the proof of a theorem much more general than (1.2) together with two corollaries. §3 describes the diffusion processes which will become the limits of our Markov chains, and §4 proves the convergence of the finite dimensional distributions. The invariance principle, our main result, is proved in §5, and some applications follow. The proof does not use any of the general results in [12, 14, 15]; rather, after preparing the way with lemmas obtained by studying the moments of \( \{X_n\} \), we imitate the techniques of Erdős, Kac and Donsker with an assist from a lemma of Billingsley. Certain extensions and some problems for further research are discussed in §7, while an appendix contains a result concerning the verification
of hypotheses used in §2 and §4. The (partially converse) theorems (2.3) and (A.1) are of independent interest, as they offer a useful criterion for deciding whether or not certain Markov chains are positive recurrent. It is hoped that this criterion can be perfected and generalized in the future.

2. A first limit theorem.

All the stochastic processes we shall consider are Markov processes on the non-negative real axis; time is discrete and the transition probabilities are stationary. Let $X_n$ denote the state of the process at time $n$, and suppose that $X_0$ is a constant.

**Definition.** The process $\{X_n\}$ is null if for each finite interval $I$,

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \Pr(X_i \in I) = 0.
$$

(2.1)

Notice that this definition, which does not distinguish between null-recurrence and transience, implies that a null process has $\lim\sup X_n = +\infty$ with probability one, so the process can't be "trapped" in a finite interval. We shall assume that the conditional moments

$$
\mu_k(x) = \mathbb{E}[(X_{n+1} - X_n)^k | X_n = x]
$$

(2.2)

all exist and are bounded (not uniformly in $k$).
Theorem 2.1. Suppose that \( \{X_n\} \) is a null process as above such that

\[
\lim_{x \to \infty} \mu_2(x) = \beta > 0, \quad \lim_{x \to \infty} xu_{-1}(x) = \alpha > \frac{\beta}{2}.
\]

Then

\[
\lim_{n \to \infty} \Pr \left( \frac{X_n}{\sqrt{n}} \leq \xi \right) = \int_0^\xi \frac{2\alpha}{\beta} e^{-\frac{1}{2\beta} \xi^2} \frac{1}{(2\beta)^{\frac{\alpha-1}{2}}} \Gamma \left( \frac{\alpha}{\beta} + \frac{1}{2} \right) \, d\xi.
\]

Remarks. This theorem will take the place of the CLT in the convergence theory to follow. The result (1.2) for a random walk of the form (1.1) is, of course, a special case. It will be proved in the appendix to this paper that the null hypothesis is automatically satisfied when in addition to the other assumptions, \( \{X_n\} \) has a countable number of states with no finite point of accumulation. Regardless of whether the states are countable or not, however, if \( \alpha > \beta/2 \) in (2.3) the null assumption must hold. The reason is that in [10] it has been shown that under these conditions \( X_n \to \infty \) with probability one, which implies (2.1). The question of relaxing some of the other restrictions will be discussed in §7.

\[\star\]

We must now postulate that \( \lim \sup X_n = + \infty \text{ a.s.} \)
Lemma 2.1. Under the conditions of the theorem,

\[(2.5) \quad E(X_n^2) = (2\alpha + \beta)n + o(n) \ .\]

**Proof.** For any \( \epsilon > 0 \), let \( I \) be a compact interval such that
\[ |x_{\mu_1}(x) - \alpha| < \epsilon \quad \text{and} \quad |\mu_2(x) - \beta| < \epsilon \quad \text{if} \quad x \notin I \ . \]

Now

\[ E(X_n^2) = E\{E[(X_n + (X_{n+1} - X_n))^2 | X_n] \} = E(X_n^2) + 2E(X_n \mu_1(X_n)) + E(\mu_2(X_n)) \]

from definition (2.2) and properties of conditional expectations. Making use of the choice of \( I \) we have

\[ \Delta E(X_n^2) = 2(\alpha + \theta \epsilon) \ Pr(X_n \notin I) + (\beta + \theta \epsilon) \ Pr(X_n \notin I) + O[\Pr(X_n \notin I)], \]

where \( \theta \) means a quantity of absolute value \( \leq 1 \). Rearranging slightly and summing, this yields

\[ E(X_n^2) = (2\alpha + \beta + 3\theta \epsilon) n + O\left[ \sum_{i=0}^{n-1} \ Pr(X_i \notin I) \right] . \]

Since the process is null, the second term is \( o(n) \) by definition for any \( I \), and since \( \epsilon \) is arbitrary, (2.5) follows.
Lemma 2.2. Again under the conditions of Theorem 2.1, for each positive integer \( k \) we have

\[(2.6) \quad E(x_n^{2k}) = a_k n^k + o(n^k),\]

where

\[(2.7) \quad a_k = (2\beta)^k \frac{\Gamma\left(\frac{\alpha}{\beta} + \frac{k}{2}\right)}{\Gamma\left(\frac{\alpha}{\beta} + \frac{1}{2}\right)}.\]

Proof. Choose an interval \( I \) as in the proof of the preceding lemma. We proceed by induction and suppose (2.6) true for all integers up to and including \( k \). Then

\[(2.8) \quad E(x_{n+1}^{2k+2}) = E\left\{E\left[X_n + \Delta x^n\right]^{2k+2} | x_n\right\} = \]

\[E(x_n^{2k+2}) + (2k + 2) E[x_n^{2k+1} \mu_1(x_n)] + \left(\begin{array}{c} 2k+2 \nonumber \\
2 \nonumber \end{array}\right) E[x_n^{2k} \mu_2(x_n)] \]

\[+ \cdots + \left(\begin{array}{c} 2k+2 \nonumber \\
\ell \nonumber \end{array}\right) E[x_n^{2k+2-\ell} \mu_\ell(x_n)] + \cdots + E[\mu_{2k+2}(x_n)].\]

Since each function \( \mu_\ell(x) \) is bounded, the terms past the third in the above expression can be over-estimated by

\[(2.9) \quad c_1 E(x_n^{2k-1}) + c_2 E(x_n^{2k-2}) + \cdots + c_{2k} \cdot \]
Applying Schwartz's inequality and the induction hypothesis we have, for instance,

\[
E(X_n^{2k-1}) = E(X_n^k X_n^{k-1}) \leq \left[ E(X_n^{2k}) \ E(X_n^{2k-2}) \right]^{\frac{1}{2}}
\]

\[
= \sqrt{a_k a_{k-1}} n^{\frac{2k-1}{2}} (1 + o(1)) = o(n^k).
\]

Similarly the other terms in (2.9) can be shown to be \( o(n^k) \). Using this, assumption (2.3) and the definition of \( I \), expression (2.8) becomes

\[
\Delta E(x_n^{2k+2}) = (2k+2)\{\alpha + \theta \epsilon + \frac{2k+1}{2} (\beta + \theta \epsilon)\} \ E(x_n^{2k}) \{1 - Pr(X_n \in I)\} + o(n^k),
\]

where the integral over \( \{X_n \in I\} \) has been incorporated (since it is \( O(1) \)) into the \( o(n^k) \). Summing this and using again the induction hypothesis (2.6) and the assumption (2.1), we obtain finally

\[
E(x_n^{2k+2}) = a_k (2k+2)\{\alpha + \theta \epsilon + \frac{2k+1}{2} (\beta + \theta \epsilon)\} \sum_{i=0}^{n-1} i^k \{i + O(1)\}
\]

\[
= a_k \{2\alpha + 2\theta \epsilon + (2k+1)(\beta + \theta \epsilon)\} n^{k+1} + o(n^{k+1}).
\]

Because \( \epsilon > 0 \) is arbitrary, this yields the case \( k+1 \) of (2.6) and

\[
a_{k+1} = a_k \{2\alpha + (2k+1)\beta\}.
\]
Since the case $k = 1$ of (2.6) was proved and $a_1$ found in Lemma 2.1, (2.6) is true in general and

$$a_k = \prod_{i=1}^{k} \left\{ 2\alpha + (2i - 1)\beta \right\},$$

which is the same as (2.7)

**Proof of Theorem 2.1.** We have shown in the two lemmas above that all the moments of the random variables $X_n^2/n$ converge as $n \to \infty$ to the limits $a_k$ given in (2.7). It follows from the well-known criterion

$$\sum_{n=1}^{\infty} \frac{\mu_{2n}}{2n} = \infty$$

that the solution to the moment problem is unique in these cases, and so by the moment-convergence theorem we conclude that $X_n^2/n$ has a limiting distribution. It is easily seen that the $a_k$ are the moments of a gamma distribution, and this leads to (2.4) upon making a change of variables.

There are two simple corollaries (of the proof rather than the statement of the theorem) which perhaps have enough independent interest to be stated as theorems themselves:

**Theorem 2.2.** Under the hypothesis of Theorem 2.1, for every $\epsilon > 0$

$$\lim_{n \to \infty} \frac{X_n}{\frac{1}{n^2} + \epsilon} = 0 \text{ with probability one.}$$

(2.10)
Proof. Applying the Markov inequality to (2.6) gives for large $n$

$$\Pr(X_n \geq \frac{1}{2n} + \varepsilon) \leq a_k \frac{n^k}{n^{2k}} \cdot (\frac{1}{2n} + \varepsilon)^{-2k}$$

for any integer $k$. For any $\xi > 0$, $\varepsilon > 0$ it is only necessary to choose $k > 1/2\varepsilon$ and we obtain

$$\sum_{n=1}^{\infty} \Pr(X_n \geq \frac{\xi n^2}{2} + \varepsilon) < \infty;$$

the desired conclusion follows upon applying the first Borel-Cantelli lemma.

Theorem 2.3. If a non-negative Markov process $X_n$ satisfies (2.3) with $\alpha < -\beta/2$, the process can not be null in the sense of (2.1).

Proof. If the process were null the proof and conclusion of Lemma 2.1 would apply; the existence of the higher conditional moments $\mu_k$ was not used there. But with $\alpha < -\beta/2$ we would then have $E(X_n^2)$ negative for large $n$, a contradiction. If the Markov process has countable states without a point of accumulation, and is irreducible, we conclude that it is positive recurrent under these conditions. This generalizes to other chains a known fact concerning random walks [6]; however, the random walk theorem is much sharper. We mention again that the appendix contains a partial converse to this result.
3. A class of diffusions.

Before proceeding further, it is convenient to assemble some facts about certain diffusion processes which satisfy the backward differential equation

\[ u_t = \frac{\alpha}{x} u_x + \frac{\beta}{2} u_{xx}, \quad \alpha, \beta \text{ constant, } x \geq 0. \]

Here \( \beta > 0 \) and \( \alpha > -\beta/2 \). The relevance of equation (3.1) for our work is suggested by the theorem of Khintchine mentioned in the introduction, and by the fact that the adjoint of (3.1), namely

\[ v_t = -\frac{\alpha}{x} v_x + \frac{\beta}{2} v_{xx}, \]

admits as a solution the density obtained from a time-dependent version of (2.4):

\[ \lim_{n \to \infty} \Pr \left( \frac{X[n]}{\sqrt{n}} \leq y \right) = \int_0^y \frac{2\alpha}{\beta} e^{-\frac{1}{2\beta t} \frac{\alpha^2}{\beta + \frac{1}{2}}} \frac{\alpha + \frac{1}{2}}{\Gamma(\frac{\alpha + \frac{1}{2}}{2})} \, dx. \]

It is necessary to distinguish the cases \( 2\alpha < \beta \), when \( x = 0 \) is a regular boundary, and \( 2\alpha > \beta \) when \( x = 0 \) is an entrance boundary [4]. (In all cases \( x = \infty \) is a natural boundary.) In the entrance boundary

\[ \star \quad [x] \text{ means the greatest integer not exceeding } x. \]
case, there exists a unique diffusion process with $x_0 = x \geq 0$ and backward equation (3.1). The function

$$u(t,x) = \mathbb{E}(f(x_t) | x_0 = x)$$

is then the only bounded solution to (3.1) satisfying the initial condition $u(0,x) = f(x)$, where $f(x)$ is any bounded continuous function. In the regular case, the reflecting-barrier boundary condition can be imposed; this condition is

$$\lim_{x \to 0^+} x^{\beta} u_x(t,x) = 0 .$$

With the aid of this condition the initial-value problem for (3.1) again has unique solutions. Here also there is a unique diffusion process $x_t$ with $x_0 = x \geq 0$ such that (3.4) obeys (3.1) and (3.5); this process has everywhere continuous paths and conserves probability. In all cases, the transition function for the selected diffusion $x_t$,

$$p_t(x,y) = \Pr(x_{t+s} \leq y | x_s = x) ,$$

is a continuous function of its three arguments if $t > 0$ .

The statements above are not intended to be either sharp or complete; they simply summarize most of what is needed for the purposes of this paper. Almost all of the assertions "well-known," and the most
relevant references are Feller [4] and Ray [13]. The exception is the fact that \( x_t \) has continuous paths even in the reflecting barrier case. This has been shown by Ito and McKeen, but is not yet published.


For the present purposes a stronger hypothesis than that of §2 is required:

**Definition.** The process \( \{X_n\} \) is uniformly null if for each finite interval \( I \), and uniformly in \( x \),

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \Pr(X_i \in I \mid X_0 = x) = 0.
\]

(4.1)

(It is shown in the appendix that this is no additional restriction for irreducible Markov chains whose states have no point of accumulation.) The purpose of this section is to develop an appropriate analogue for the multidimensional C.L.T. We first prove

**Theorem 4.1.** If \( \{X_n\} \) is a uniformly null process obeying the hypotheses of Theorem 2.1, then

\[
\lim_{n \to \infty} \Pr[X_{nt} \leq y \sqrt{n} \mid X_0 = x \sqrt{n} + o(\sqrt{n})] = p_t(x,y)
\]

exists for all non-negative \( x, t, \) and \( y \); \( p_t(x,y) \) is the transition probability function of the diffusion process defined in §3 with the appropriate values of \( \alpha \) and \( \beta \).
The proof resembles that of Theorem 2.1 in that it depends on the calculation of asymptotic moments, but it is slightly more complicated. In writing (4.2) it is understood that the $o(\sqrt{n})$ is so chosen that $x\sqrt{n} + o(\sqrt{n})$ is a state of the Markov process, so that the left hand side is defined.

**Lemma 4.1.** Under the conditions of the theorem,

$$
(4.3) \quad \mathbb{E}[X_1^2 \mid X_o = x\sqrt{n} + o(\sqrt{n})] = (2\alpha + \beta) i + [x\sqrt{n} + o(\sqrt{n})]^2 + o(1),
$$

where $o(1)$ is uniform in $n$ and $x$.

**Proof.** Beginning as we did in the proof of Lemma 2.1, we have

$$
\Delta \mathbb{E}[X_1^2 \mid X_o] = \left\{2\alpha + \beta + 3\theta \epsilon\right\} \text{Pr}(X_1 \notin I \mid X_o) + O\left\{\text{Pr}(X_n \in I \mid X_o)\right\},
$$

where $\theta$, $\epsilon$ and $I$ have the same meaning as before. Summing up to $i$, we obtain

$$
\mathbb{E}[X_1^2 \mid X_o] = \left\{2\alpha + \beta + 3\theta \epsilon\right\} i + X_o^2 + O\left(\sum_{j=0}^{i-1} \text{Pr}(X_j \in I \mid X_o)\right).
$$

From the uniformly null hypothesis the last term is $o(1)$ uniformly in $X_o$, and since $\epsilon$ is arbitrary this yields (4.3). Of course if $i = O(n)$, the error term is also $o(n)$.
Lemma 4.2. Again under the conditions of Theorem 4.1, for each positive integer \( k \) and any \( M < \infty \) we have

\[
\mathbb{E}[x_1^{2k} | x_o] = x \sqrt{n} + o(\sqrt{n}) = \sum_{\ell=0}^{k-1} a_{\ell}^{(k)} x^{k-\ell} n^{\ell} + o(n^k)
\]

for all \( i \leq M \cdot n \), where

\[
a_{\ell}^{(k+1)} = a_{\ell}^{(k)} \frac{2(2\alpha + (2k+1)\beta)}{k+1-\ell} \quad \text{if} \quad \ell \leq k, \quad a_{k+1}^{(k+1)} = x^{2k+2}.
\]

If the \( o(\sqrt{n}) \) on the left side in (4.4) is uniform in \( x \) for \( x \) in a compact set, the \( o(n^k) \) term is uniform also.

Proof. The case \( k = 1 \) immediate from Lemma 4.1, and \( a_o^{(1)} = 2\alpha + \beta \), \( a_1^{(1)} = x^2 \) by (4.3). Assume (4.4) is true for all integers up to and including \( k \). Then

\[
\mathbb{E}[x_1^{2k+2} | x_o] = \sum_{\ell=0}^{2k+2} (2k+2-\ell) \mathbb{E}[x_1^{2k+2-\ell} | x_o],
\]

so that with \( I \) as before we have

\[
\Delta \mathbb{E}[x_1^{2k+2} | x_o] = (2k+2)\{2\alpha + 2(2k+1)\beta + 2\epsilon + \frac{2k+1}{2}(2\alpha + \beta)\epsilon\} \mathbb{E}[x_1^{2k} | x_o] \{1 - \text{Pr}(x_1 \notin I | x_o)\}
\]

\[
+ \sum_{\ell=3}^{2k+2} o_\ell \mathbb{E}[x_1^{2k+2-\ell} | x_o] + o\{\text{Pr}(x_1 \notin I | x_o)\}. \]
This is analogous to (2.8), and again it follows from the induction assumption, Schwartz' inequality and the assumption that \( i = O(n) \) that each summand in the next to last term is \( o(n^k) \). Using this, (4.1), and the induction assumption we obtain upon summing over \( i \) from 0 to \( i-1 \) that

\[
E[X_1^{2k+2} | X_0 = x\sqrt{n} + o(\sqrt{n})] = (2k+2) \left( \alpha + \beta \epsilon + \frac{2k+1}{2} \right) \sum_{\ell=0}^{k} a^{(k)}_\ell \frac{i^{k-\ell+1}n^{\ell}}{k-\ell+1}
+ x^{2k+2} n^{k+1} + o(n^{k+1})
\]

provided \( i = O(n) \). Since \( \epsilon \) is arbitrary this yields (4.4) for the case \( k+1 \) and also the recurrence relations (4.5). The final statement about uniformity in \( x \) follows from Lemma 4.1 if \( k = 1 \) and can be seen for all \( k \) by reexamining the above induction argument with this in mind.

It is now easy to show that the limit in (4.2) exists when \( t = 1 \): the lemmas state that all the moments of the distributions

\[
F_n(z) = \Pr\left[ \frac{X_n^2}{n} \leq z \mid \frac{X_0}{\sqrt{n}} = x + o(1) \right]
\]

have limits as \( n \to \infty \) and so it is only necessary to see that the limiting moments determine a unique distribution and apply the moment convergence theorem. It is easily verified by induction, using (4.5), that

\[
\text{Put } i = n \text{ in (4.4) and divide by } n^k.
\]
\[
(4.6) \quad a^{(k)}_\ell = (\frac{k}{\ell}) \prod_{j=\ell}^{k-1} \left\{ 2\alpha + (2j+1)\beta \right\} x^{2\ell},
\]

and the \( k \)th moment in question is, of course, \( \sum_{\ell=0}^{k} a^{(k)}_\ell \). If each term of the product in (4.6) were replaced by \( 2\alpha + (2k+1)\beta \) we should obtain the estimate

\[
(4.7) \quad \sum_{\ell=0}^{k} a^{(k)}_\ell < \sum_{\ell=0}^{k} (\frac{k}{\ell})^{k-\ell} x^{2\ell} = \left\{ 2\alpha + (2k+1)\beta + x^2 \right\}^{k}.
\]

This allows the same criterion used in proving Theorem 2.1 to be applied again; it shows that there is a unique solution to the moment problem which completes the proof of the existence of the limit when \( t = 1 \).

There is no difficulty with other values of \( t \), for

\[
(4.8) \quad \lim_{n \to \infty} E\left[ \left( \frac{X^2_{\text{int}}}{n} \right)^k \right] \bigg| \frac{X_0}{\sqrt{n}} = x + o(1) = t^k \lim_{m \to \infty} E\left[ \left( \frac{X^2_m}{m} \right)^k \right] \bigg| \frac{X_0}{\sqrt{m}} = \frac{x}{\sqrt{t}} + o(1)
\]

The factor \( t^k \) does not affect the unique solution of the moment problem so the same argument applies and proves (4.2) in general. Incidentally, we also see from (4.8) that

\[
(4.9) \quad p_t(x,y) = p_1\left( \frac{x}{\sqrt{t}}, \frac{y}{\sqrt{t}} \right).
\]

It remains to identify the limiting distribution \( p_t(x,y) \). The moment convergence theorem assures us that the limiting moments are, in fact, the moments of the limit and it is by using these that the
problem is solved. We have

\[
\mu_{2k}(x,t) = \int_0^\infty y^{2k} d\mu_t(x,y) = t^k \sum_{j=0}^k \binom{k}{j} \frac{k-1}{j!} \left\{ 2\alpha + (2j + 1)\beta \right\} \frac{x_j}{t^{j+1}}
\]

from (4.6) and (4.8). This function, a polynomial in \(x\) and \(t\), is a solution of the differential equation (3.1) as can be verified by direct substitution; it also satisfies condition (3.5) since \(2\alpha/\beta > -1\).

We can now see that

\[
g_t(x,u) = \int_0^\infty e^{-iuy^2} d\mu_t(x,y)
\]

satisfies the same differential equation and boundary condition. Indeed,

\[
g_t(x,u) = \sum_{n=0}^\infty \frac{(iu)^n}{n!} \mu_{2n}(x,t)
\]

where the expansion is valid when \(|ut| < (2\alpha)^{-1}\). (This can be shown by estimating \(\mu_{2n}\) from the obvious time-dependent extension of (4.7).) Since the coefficients satisfy (3.1) and (3.5), so also does \(g_t(x,u)\), at least for \(u\) and \(t\) in a neighborhood of the origin. A characteristic function is bounded for real \(u\) and the uniqueness of the solution of such initial value problems for (3.1) (with (3.5) if \(\alpha < \beta/2\)) shows that

\[
g_t(x,u) = E(e^{iux^2/t})_{x_0 = x},
\]
where \( x_t \) is the appropriate diffusion. The uniqueness of the Fourier-Stieltjes transform completes the identification of \( p_t(x, y) \) and the proof of the theorem.

Finally, we deduce the desired multi-dimensional limit theorem as a consequence:

**Theorem 4.2.** If the conditions of Theorem 4.1 are met, and if \( X_0 \) has any fixed value, then for \( 0 \leq t_1 \leq t_2 \leq \cdots \leq t_k \),

\[
(4.12) \quad \lim_{n \to \infty} \Pr \left( \frac{X_{[nt_1]}}{\sqrt{n}} \leq y_1, \ldots, \frac{X_{[nt_k]}}{\sqrt{n}} \leq y_k \right)
\]

exists and is the same as the joint distribution of \( \{x_{t_1}, \ldots, x_{t_k}\} \), where \( x_t \) is the appropriate diffusion from §3 with \( x_0 \neq 0 \).

**Proof.** Nothing is really lost by considering for simplicity the case \( k = 2 \). By the Markov property and the stationarity of the transition probabilities of \( \{X_n\} \), (4.12) may be written

\[
\lim_{n \to \infty} \int_0^{y_1} \Pr \left( \frac{X_{[n(t_2-t_1)]}}{\sqrt{n}} \leq y_2 \mid \frac{X_0}{\sqrt{n}} = x \right) \, d \Pr \left( \frac{X_{[nt_1]}}{\sqrt{n}} \leq x \mid X_0 = y_0 \right)
\]

By Theorem 4.1, the integrand converges to \( p_{t_2-t_1}(x, y_2) \) and the integrator to \( p_{t_1}(0, x) \). Thus the problem is one of justifying the convergence

\[
\int_0^{y_1} f_n(x) \, dg_n(x) \rightarrow \int_0^{y_1} f(x) \, dg(x)
\]
when \( f_n \to f \) and \( g_n \to g \), with \( g(x) \) a continuous distribution function, \( g_n(x) \) distributions, \( f_n(x) \) uniformly bounded, and \( f(x) \) continuous. It follows easily from standard theorems that this interchange of limiting operations is indeed correct if in addition we know that \( f_n(x) \to f(x) \) uniformly for \( x \in [0, y_1] \). Thus the proof will be finished when we have established the final

**Lemma 4.3.** Under the conditions of Theorem 4.1, the limit (4.2) holds uniformly in \( x \), for \( x \) in any compact interval \( I \).

**Proof.** It will be helpful to consider symmetrized versions of the distributions being studied. If \( F(y) \) is any distribution function with \( F(0) = 0 \) and \( F(\infty) = 1 \), let

\[
F^s(y) = \begin{cases} 
\frac{1-F(y)}{2} & \text{if } y \leq 0 \\
\frac{1+F(y)}{2} & \text{if } y > 0
\end{cases}
\]

Clearly the conclusion of the lemma is true if we can show that

\[
s_p^n(x, y) = \Pr[X_{nt} \leq \sqrt{n} x_0 = x \sqrt{n} + o(\sqrt{n})] \to s_p(x, y)
\]

uniformly for \( x \in I \). The reason for the symmetrization is that our grasp on the distributions comes only through the even moments; the odd ones are rendered harmless by the change. The alternative of studying \( X_n^2 \) as we have done earlier is less helpful because we lose the property that the characteristic functions are entire.
The main tool will be a result of Parzen [11], which gives the desired conclusion provided that \( \mathcal{P}_t(x, y) \) is equicontinuous in \( x \) (\( x \in I \), fixed \( t \)), and that the characteristic functions

\[
f_t^{(n)}(x, u) = \int_{-\infty}^{\infty} e^{iyu} d\mathcal{P}_t^{(n)}(x, y)
\]

converge to

\[
f_t(x, u) = \int_{-\infty}^{\infty} e^{iyu} d\mathcal{P}_t(x, y)
\]

uniformly in \( x \in I \) for each \( u \). The first condition is immediate since \( \mathcal{P}_t(x, y) \), and so \( \mathcal{P}_t(x, y) \), is jointly continuous in \( x \) and \( y \) (§3). To verify the second one, first observe that \( f_t(x, u) \) is an entire function of \( u \); this is the case since

\[
(4.13) \quad f_t(x, u) = \sum_{k=0}^{\infty} \mu_{2k}(x, t) \frac{(iu)^{2k}}{(2k)!}
\]

(the odd moments are 0) and the estimate (4.7), or the modified version of it if \( t \not= 1 \), shows that the series converges for all \( u \). The moments \( \mu_{2k}(x, t) \) are continuous in \( \alpha \), and hence bounded for \( x \in I \). We can therefore for any \( \epsilon \) and any \( u \), find \( K \) such that

\[
\mu_{2k}(x, t) \left| \frac{u^{2k}}{(2k)!} \right| < \epsilon
\]
for all $x \in I$, so that again for $x \in I$ we have

$$
(4.14) \quad f_t(x,u) = \sum_{k=0}^{K-2} \mu_{2k}(x,t) \left( \frac{i u}{2} \right)^{2k} \frac{(2k)!}{(2k)!} + \theta \epsilon, \quad |\theta| \leq 1.
$$

Now for the $f^{(n)}$ we have the limited expansion

$$
(4.15) \quad f_t^n(x,u) = \sum_{k=0}^{K-2} \mu_{2k}(x,t) \left( \frac{i u}{2} \right)^{2k} \frac{(2k)!}{(2k)!} + \theta \mu_{2k}(x,t) \frac{|u|^{2k}}{(2k)!}.
$$

The key fact is that by Lemmas 4.1 and 4.2, $\mu_{2k}(x,t) \rightarrow \mu_{2k}(x,t)$ uniformly for $x \in I$. Thus for large enough $n$, the last term in (4.15) is less than $2\epsilon$ for all $x \in I$, and each of the earlier moments is uniformly arbitrarily close to its limit. Knowing this, we see from (4.14) and (4.15) that the convergence of $f_t^n(x,u)$ to $f_t(x,u)$ is indeed uniform for $x \in I$. By Parzen's theorem, this completes the proof of the lemma and so that of Theorem 4.2.

5. The invariance principle.

Let $\{X_i\}$ be a Markov process of the kind we have been considering, with $X_0 = 0$. Define now

$$
(5.1) \quad x_t^{(n)} = \frac{X_i}{\sqrt{n}} \quad \text{if} \quad t = \frac{i}{n}, \quad i = 0, 1, \ldots, n,
$$

and determine $x_t^{(n)}$ by linear interpolation for other values of $t$ in $[0,1]$. The transition law of the Markov process determines, of
course, the joint probability distribution of $X_0, X_1, \ldots, X_n$; this in turn induces a probability measure on the Borel field of the space $C_0$ by virtue of the measurable mapping $\{x_0, x_1, \ldots, x_n\} \rightarrow x_t^{(n)}$. ($C_0$ is the metric space of all continuous functions on $[0,1]$ vanishing at the origin, with supremum norm.) We denote the induced measure on $C_0$ by $\mu_n$. The main result of this paper is the following:

**Theorem 5.1.** (Invariance principle.) Suppose that $\{X_n\}$ is a uniformly null (§4) process obeying (2.3) and with each $\mu_K(x)$ bounded. Then the induced measures $\mu_n$ converge weakly as $n \rightarrow \infty$ to a probability measure $\mu$ on $C_0$; $\mu$ is the measure of the diffusion process $x_t$ described in §3 for the appropriate values of $\alpha$ and $\beta$, with $x_0 = 0$.

The definition of weak convergence is that $\int f \, d\mu_n \rightarrow \int f \, d\mu$ for all bounded continuous functions $f$ on $C_o$, but it is easy to recast this condition slightly. One version [1] yields the

**Corollary.** Under the conditions of the theorem

$$\lim_{n \rightarrow \infty} \Pr[f(x_t^{(n)}) \leq \alpha] = \mu\{f(x_t) \leq \alpha\}$$

(5.2)

at each value of $\alpha$ where the right hand side is continuous, provided that $f(\ )$ is continuous a.e. ($\mu$) on $C_o$.

**Proof of the theorem.** Although it appears that our problem might fall within the scope of Prokhorov's or Skorohod's work, it is difficult to apply their theorems. It is rather easy, however, to use
the methods of Donsker and Billingsley. From the latter we need the following:

**Lemma 5.1.** (An adaptation of Theorem 2.3 of [1]). Let

\[ E = \{ x \mid 0 \neq \alpha_j \leq x_t \leq \beta_j > 0 \text{ for } \frac{j-1}{h} \leq t \leq \frac{j}{h}, \ j = 1, 2, \ldots, h \}^*, \text{ and} \]
\[ G_n = \{ (x_j) \mid \alpha_j \sqrt{n} \leq x_1 \leq \beta_j \sqrt{n} \text{ for } \frac{i-1}{n} \leq \frac{i}{n} \leq \frac{j}{n}, \ j = 1, \ldots, h; \ i = 1, \ldots, n \} . \]

Then if \( \Pr(G_n) = \Pr(\{X_0, X_1, \ldots, X_n\} \in G_n) \rightarrow \mu(E) \text{ as } n \rightarrow \infty \) for all \( h, \{\alpha_j\}, \{\beta_j\}, \) the measures \( \mu_n \) converge weakly to \( \mu \).

This result considerably simplifies our task. It is necessary next to provide an estimate which was immediate for sums of independent random variables with unit variance:

**Lemma 5.2.** For any \( \epsilon > 0 \) there exists \( A_\epsilon \) such that

\[ \Pr(\mid X_1 - x \mid > \epsilon \sqrt{n} \mid X_0 = x) \leq \frac{A_\epsilon}{k} \text{ for } 1 \leq \frac{n}{k} , \]
\[ \text{for all } n, k \text{ and all } x . \]

**Proof.** We use some moment estimates similar to those in §4.

Observe first that as \( X_1 \rightarrow \infty \),

\[ E[\mid X_{1+1}^2 - X_1^2 \mid X_1] = \beta + 2\alpha + o(1) \quad \text{and} \]

* We exclude \( \alpha_j = 0 \) in order that the boundary of \( E \) have \( (\mu) \) measure 0, but some of the \( \alpha \)'s may be negative.
\[ \mathbb{E}[(x_{i+1}^2 - x_i^2)^2 \mid x_i] = 4\beta x_i^2 + o(x_i^2) \] .

Proceeding in a familiar manner, we calculate

\[ \Delta \mathbb{E}[(x_1^2 - x^2)^2 \mid x_o = x] = \mathbb{E} \{ \mathbb{E}[(x_{i+1}^2 - x^2)^2 - (x_i^2 - x^2)^2 \mid x_i, x_o = x] \} \]

\[ = \mathbb{E} \{ \mathbb{E}[(x_{i+1}^2 - x_i^2)^2 + 2(x_i^2 - x^2)(x_{i+1}^2 - x_i^2) \mid x_i, x_o = x] \} \]

\[ = \mathbb{E} \{ 4\beta x_i^2 + 2(x_i^2 + 2(x_i^2 - x^2)(\beta + 2\alpha + o(1)) + o(x_i^2) \mid x_o = x) \} \] .

Applying Lemma 4.1 we have an estimate of the form

\[ |\Delta \mathbb{E}[(x_1^2 - x^2)^2 \mid x_o = x]| \leq c_1 + c_2 x^2 + c_3 \] ,

and summing this yields the inequality

\[ \mathbb{E}[(x_1^2 - x^2)^2 \mid x_o = x] \leq c_4 i^2 + c_5 i x^2 \] .

The Tchebychev inequality then gives

\[ (5.5) \quad \text{Pr}(\|x_1^2 - x^2\| > \lambda \mid x_o = x) \leq \frac{c_4 i^2 + c_5 i x^2}{\lambda^2} \] .

To convert (5.5) into (5.4), consider first the cases when

\[ x \geq \epsilon \sqrt{n}/2 \] . Put \( \lambda = x \epsilon \sqrt{n} \) and note that
\[ \Pr(\mid X_i - x \mid > \epsilon \sqrt{n}) \leq \Pr(\mid X_i - x \mid > \frac{\epsilon \sqrt{n} x}{X_i + x}) \leq \frac{C_4 i^2 + C_2 i x^2}{x \epsilon^2 n}. \]

Since \( i \leq n/k \), we obtain under these conditions the bound

\[ \Pr(\mid X_i - x \mid > \epsilon \sqrt{n} \mid X_o = x) \leq \frac{4C_4}{k \epsilon^2} + \frac{4C_2}{k \epsilon^2}, \]

which gives (5.4) with \( A_\epsilon \leq \frac{4C_4}{\epsilon} + \frac{4C_2}{\epsilon^2} \). For \( x < \epsilon \sqrt{n}/2 \), on the other hand, put \( \lambda = (x + \epsilon \sqrt{n})^2 - x^2 \). Since \( \lambda > x^2 \), the events \( X_i^2 - x^2 > \lambda \) and \( \mid X_i^2 - x^2 \mid > \lambda \) are identical and so

\[ \Pr(\mid X_i^2 - x^2 \mid > \lambda) = \Pr(X_i > \sqrt{x + x^2}) = \Pr(X_i > x + \epsilon \sqrt{n}). \]

Since \( x < \epsilon \sqrt{n} \), the last is \( \Pr(\mid X_i - x \mid > \epsilon \sqrt{n}) \). Thus

\[ \Pr(\mid X_i - x \mid > \epsilon \sqrt{n} \mid X_o = x) \leq \frac{C_4 i^2 + C_2 i x^2}{[(x + \epsilon \sqrt{n})^2 - x^2]^2} \leq \frac{1}{\epsilon} \left( \frac{C_4}{k^2} + \frac{\epsilon^2 C_2}{4k} \right). \]

and the previous choice of \( A_\epsilon \) still holds good; the lemma is proved.

The remainder of the proof is really just a paraphrase of a portion of Donsker's argument; it is included for the sake of completeness.

For any \( \epsilon > 0 \), let \( \mathcal{E}(\epsilon) \) and \( \Omega_n(\epsilon) \) be the sets defined in (5.3) with \( \alpha_j + \epsilon \) and \( \beta_j - \epsilon \) in place of \( \alpha_j \), \( \beta_j \) respectively for all \( j \). Choose any integer \( v \) and let

\[ n_{j,k} = [(j-1) \frac{n}{h} + \frac{k}{\nu} \frac{n}{h}], \quad j = 1, \ldots, h; k = 1, \ldots, v. \]
Let $G_{n,v}$ and $G_{n,v}(\epsilon)$ be sets defined by the same conditions as $G_n$ and $G_n(\epsilon)$, except that the inequalities need only hold for $X_i$ when $1$ is among the numbers $n_{j,k}$. We shall now show that

$$\Pr(G_{n,v}) \geq \Pr(G_n) \geq \Pr(G_{n,v}(\epsilon)) - \frac{A_\epsilon}{hv}.$$  

The first inequality is obvious. Let $H_{n,\ell}$ be the event that $X_{\ell}$ is the first random variable which fails to satisfy the defining conditions of $G_n$ (5.3); then clearly

$$1 - \Pr(G_n) = \sum_{\ell=1}^n \Pr(H_{n,\ell}).$$

Provided that $n_{j,k} < \ell \leq n_{j,k+1}$ we have

$$\Pr(H_{n,\ell}) = \Pr(H_{n,\ell} \text{ and } |X_{n_{j,k+1}} - X_{\ell}| > \epsilon \sqrt{n}) + \Pr(H_{n,\ell} \text{ and } |X_{n_{j,k+1}} - X_{\ell}| \leq \epsilon \sqrt{n}),$$

where $n_{j,k+1} - \ell \leq n/hv$. By the Markov property and Lemma 5.2, the first terms on the right are bounded by $\Pr(H_{n,\ell}) A_\epsilon/hv$, and so their sum by $A_\epsilon/hv$. Thus

$$1 - \Pr(G_n) \leq \frac{A_\epsilon}{hv} + \sum_{j,k} \sum_{\ell=n_{j,k+1}}^{n_{j,k+1}} \Pr(H_{n,\ell} \text{ and } |X_{n_{j,k+1}} - X_{\ell}| \leq \epsilon \sqrt{n}).$$

All of the (mutually exclusive) events making up the sum in (5.8) are contained in the complement of $G_{n,v}(\epsilon)$, and thus the second part of (5.7) follows.
To complete the proof that \( \Pr(G_n) \to \mu(E) \), first choose \( \varepsilon \) small enough so that \( \mu(E) - \mu(E(\varepsilon)) \leq \delta \), say. This is possible since \( E(\varepsilon) \not\subset \text{int}(E) \) as \( \varepsilon \searrow 0 \) and the boundary of \( E \) has measure 0. Then recall from §4 that the joint limiting distribution of
\[
\left\{ \frac{X_{n_j/k}}{\sqrt{n}} \right\}
\]
exists as \( n \to \infty \), and is equal to the joint distribution of \( \left\{ X_{t/v}, t=\frac{i}{hv} \right\} \). This implies that
\[
(5.9) \quad \lim_{n \to \infty} \Pr(G_{n,v}) = \mu\left\{ x_{v} | \alpha_j \leq x_{v}/hv \leq \beta_j \text{ for } \frac{j-1}{h} \leq \frac{j}{h} \right\},
\]
and the analogous thing for \( G_{n,v}(\varepsilon) \). The probability on the right, however, approaches \( \mu(E) \) as \( v \to \infty \). Choose a \( v \) making this difference and the corresponding one for \( E(\varepsilon) \), as well as \( A_{\varepsilon}/hv \), less than \( \delta \), and we have from (5.7)

\[
(5.10) \quad \mu(E) - 3\delta \leq \lim_{n \to \infty} \Pr(G_n) \leq \lim_{n \to \infty} \Pr(G_n) \leq \mu(E) + \delta
\]

By Lemma 5.1, the proof of the invariance principle is complete.

6. Some applications

There are two possible ways to apply the invariance principle to obtain specific limit theorems; both have been used in connection with Donsker's theorem and the Wiener process. The invariance principle insures that the limit distributions exist, and they may be identified by calculations with a particular process in the proper "domain of
attraction." Simple random walk has often served this purpose for sums of random variables [3]. Alternatively, the limiting stochastic process may be studied directly to determine the distribution of the functional $f(x_t)$. In this section we use the first approach to obtain the limiting distribution of the maximum functional, and we shall make a few additional observations. Other applications will be developed in the future.

**Theorem 6.1.** Let $\{X_n\}$ obey the invariance principle (§5) corresponding to the diffusion defined in §3 with certain values of $\alpha$ and $\beta$. Then

$$
(6.1) \quad \lim_{n \to \infty} \Pr \{ \max (X_1, X_2, \ldots, X_n) \leq x_0 \} = 1 - F \left( \frac{1}{x_0} \right),
$$

where $F(\cdot)$ is the distribution function with Laplace transform

$$
(6.2) \quad \varphi(s) = \left\{ \sum_{n=0}^{\infty} e_n^* s^n \right\}^{-1}, \text{ where } e_0^* = 1 \text{ and }
$$

$$
e_n^* = [(2\beta)^n n! \prod_{k=1}^{n} (k - \frac{1}{2} + \frac{\alpha}{\beta})]^{-1}.
$$

**Remark.** In case $\alpha = 0, \beta = 1$, the function $\sigma_2(x)$ of [3] is also equal to the limit in (6.1). These different appearing answers are proved equivalent by our work; see also the discussion at the end of this section.
Proof. The functional \( f(x_t) = \max \left\{ x_t \mid t \in [0,1] \right\} \) is continuous everywhere on \( C_0 \), so the limit in (6.1) exists and is independent of the choice of the process \( \{x_n\} \) for given values of \( \alpha \) and \( \beta \). A particular process in the domain of attraction will be constructed for which the limiting distribution (6.1) is known from Karlin and McGregor's theory of birth-and-death processes [8]. If the theory of [8] had been worked out for random walks as well as birth and death, as it undoubtedly could be, the process of (1.1) could furnish the necessary example and most of the labor of the present proof avoided. Our theorem, in particular, performs this task as far as the maximum functional is concerned.

Consider a Poisson process \( N_t \) with parameter \( \beta \). When an event occurs, a particle moving on the positive integers makes a transition one unit to left or right. If the particle was at \( j \), its probability of moving to \( j+1 \) will be \( \frac{1}{2}(1 + \frac{\alpha}{\beta j}) \) and to \( j-1 \), \( \frac{1}{2}(1 - \frac{\alpha}{\beta j}) \) provided \( j\beta < |\alpha| \). For the remaining states these probabilities can all be taken as \( \frac{1}{2} \) except that from 0 only transitions to \( j=1 \) are possible. This scheme describes a birth-and-death process with rates

\[
(6.3) \quad \lambda_j = \frac{\beta}{2j} + \frac{\alpha}{2j}, \quad \mu_j = \frac{\beta}{2j} - \frac{\alpha}{2j}
\]

for all large \( j \). Karlin and McGregor's work in [8] tells us that if \( y_t \) is the position of the particle at time \( t \) \((y_0 = 0)\), then
\[ \lim_{t \to \infty} \Pr(\max\{y_{\tau} \mid \tau \leq t\} \leq \sqrt{t}) = 1 - F\left( \frac{1}{x^2} \right), \]

where the Laplace transform of \( F(\cdot) \) is defined by (6.2). We shall define a discrete-time Markov process \( \{X_n\} \) by

\[ X_n = y_t \mid t = n, n = 0, 1, 2, \ldots. \]

There are two things to be done: we must show that \( \{X_n\} \) satisfies the hypotheses of Theorem 5.1, and that the maximum functional for \( X_n \) also obeys (6.4). The latter is simple, since clearly

\[ \max\{y_t \mid t \leq n\} - N \leq \max\{X_i \mid i \leq n\} \leq \max\{y_t \mid t \leq n\} \]

where \( N \) is the largest number of transitions to occur in a single unit of time before time \( n \). \( N \) is thus distributed as the largest among \( n \) independent random variables, each Poisson distributed with mean \( \beta \). The probability that \( N \) is small compared to \( \sqrt{n} \) can easily be seen to approach unity, so that (6.4) implies (6.1).

It remains to verify that (2.3) holds, for by the results of the appendix to this paper, that will imply the uniformly null condition in the present case. Since each transition of \( y_t \) is of length one, \( |X_{n+1} - X_n| \) is bounded by the number of transitions between \( t=n \) and \( t = n+1 \). Thus

\[ \mu_k(x) \leq \mathbb{E}[|X_{n+1} - X_n|^k \mid X_n = x] \leq \sum_{k=0}^{\infty} \frac{k^ke^{-\beta}}{k!} \beta^k. \]
The series is certainly finite, and is independent of \( x \). To study the precise asymptotic behavior of \( \mu_1(x) \) and \( \mu_2(x) \) is a little harder, and we first define

\[
\mu_k^{(\ell)}(x) = \mathbb{E}[(X_{n+1} - X_n)^k \mid X_n = x, N_{n+1} - N_n = \ell].
\]

\((N_{n+1} - N_n)\) is the number of transitions of \( y_t \) between \( t = n \) and \( t = n+1 \). Then

\[
\mu_2(x) = \sum_{\ell=0}^{\infty} \mu_2^{(\ell)}(x) \frac{e^{-\beta} \beta^\ell}{\ell!}.
\]

Since the transition probabilities approach \( \frac{1}{2} \) as \( x \to \infty \), \( \mu_2^{(\ell)}(x) \) approaches the corresponding value for simple random walk, namely \( \ell \).

Thus since \( \mu_2^{(\ell)}(x) \leq \ell^2 \), by dominated convergence,

\[
\lim_{x \to \infty} \mu_2(x) = \sum_{\ell=0}^{\infty} \frac{\ell e^{-\beta} \beta^\ell}{\ell!} = \beta.
\]

The first moment requires slightly more care. Note first that for values of \( \ell < x - |\alpha| \beta^{-1} \), if \( \alpha > 0 \),

\[
\frac{\ell \alpha}{\beta(x+\ell)} \leq \mu_1^{(\ell)}(x) \leq \frac{\ell \alpha}{\beta(x-\ell)}.
\]

This is true since \( \mu_1^{(\ell)}(x) \) is the mean of the sum of \( \ell \) random variables, each having a mean equal to \( \frac{\alpha}{\beta y} \) for some \( y \).
\[ x - \ell \] and \[ x + \ell \). Thus (still for \( \alpha \geq 0 \))

\[
\mu_1(x) = \sum_{k=0}^{\infty} \mu_1^{(k)}(x) \frac{e^{-\beta \frac{x}{k}}}{k^k} \geq \sum_{k=0}^{\infty} \frac{\ell k \alpha x}{x+k} \frac{e^{-\beta \frac{x}{k}}}{k^k} \\
= \sum_{k=0}^{\infty} \frac{\alpha}{x} \frac{\ell e^{-\beta \frac{x}{k}}}{k^k} + \sum_{k=0}^{\infty} \left( \frac{\alpha}{x+k} - \frac{\alpha}{x} \right) \frac{\ell e^{-\beta \frac{x}{k}}}{k^k} = \frac{\alpha}{x} + o(x^{-2}) .
\]

Estimating in the other direction, since \( \mu_1^{(k)}(x) \leq \ell \) we have

\[
\mu_1(x) = \left( \sum_{k=0}^{\left\lfloor \sqrt{x} \right\rfloor} + \sum_{k=\left\lfloor \sqrt{x} \right\rfloor + 1}^{\infty} \right) (\mu_1^{(k)}(x) \frac{e^{-\beta \frac{x}{k}}}{k^k}) \leq \\
\leq \sum_{k=0}^{\sqrt{x}} \frac{\ell k \alpha x}{\beta (x-k)} \frac{e^{-\beta \frac{x}{k}}}{k^k} + \sum_{k=\sqrt{x}}^{\infty} \frac{\ell e^{-\beta \frac{x}{k}}}{k^k} .
\]

and it is easy to see that this expression also is \( \frac{\alpha}{x} + o(x^{-2}) \). The case \( \alpha < 0 \) is entirely similar; this completes the verification that \( \{ X_n \} \) obeys the hypotheses of the invariance principle. Since (6.1) and (6.2) hold in this case, they must also hold whenever the invariance principle implies convergence of \( \{ X_n/\sqrt{n} \} \) to the same limiting process, and so the theorem is proved.

The result just proved has an immediate Corollary. For the diffusions defined in §3,

\[
\text{Pr}(\max \{ x_t | t \in [0,1] \} \leq x) = 1 - F\left( \frac{1}{x^2} \right) ,
\]

where the Laplace transform of \( F(\ ) \) is given by (6.2).
(This is not new in the case $\alpha = 0$). It might well be possible to prove the corollary directly; if this were done, Theorem 6.1 would, of course, follow. The present arrangement illustrates a remark of Skorohod to the effect that invariance principles may be useful for gaining information about the limiting process, as well as for proving limit theorems.

There are other functionals of interest for which the limit distributions are known if $\alpha = 0$. For instance, under the conditions of Theorem 5.1, the quantities

\[
\lim_{n \to \infty} \Pr \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \leq \frac{3}{2} x n \right) = \Phi_1(x) \quad \text{and} \quad \lim_{n \to \infty} \Pr \left( \sum_{i=1}^{n} X_i^2 \leq x n^2 \right) = \Phi_2(x)
\]

both exist; the appropriate functionals are

\[
f_1(x_t) = \int_0^t |x_t| \, dt \quad \text{and} \quad f_2(x_t) = \int_0^t x_t^2 \, dt.
\]

Actually, $f_1(x_t^{(n)})$ and $f_2(x_t^{(n)})$ are not quite the quantities we desire for (6.7), but the differences are asymptotically negligible. For example,

\[
f_1(x_t^{(n)}) = n^{3/2} \sum_{i=1}^{n} X_i - \frac{n^{3/2}}{2} X_n.
\]

which clearly has the same limiting distribution as the first quantity in (6.7). Limit theorems of this sort were proved and $F_1$ and $F_2$ described in [3] for the case of sums of independent random variables.
This gives the desired limiting distribution in the present context when \( \alpha = 0 \). The other cases have not yet been studied; perhaps it will be possible to use the birth-and-death process constructed for Theorem 6.1 to evaluate \( F_1 \) and \( F_2 \) when \( \alpha \neq 0 \). Another functional of interest is that value \( L_n \) of time when the maximum of \( \{x_1, \ldots, x_n\} \) is first attained. This is being studied in the theory of birth-and-death and the (forthcoming) results there should provide the necessary examples to obtain the analogue for \( L_n \) of Theorem 6.1.

7. Extensions and some further problems.

It is natural to ask what happens if the conditional moments \( \mu_1 \) and \( \mu_2 \) are not of the order of \( x^{-1} \) and a constant, respectively. Part of the answer is very easy: If for some number \( \delta < 2 \) we have

\[
(7.1) \quad \lim_{x \to \infty} x^{-\delta} \mu_2(x) = b > 0, \quad \lim_{x \to \infty} x^{-\delta+1} \mu_1(x) = a > \frac{b}{2(1-\delta)},
\]

while the higher moments \( \mu_k(x) \) are each \( O(x^\delta) \), and if the process \( \{X_n\} \) is uniformly null, then there is a convergence theory for \( \{x_n^{1/(2-\delta)}\} \) as \( n \to \infty \) entirely analogous to the one developed above for \( \delta = 0 \). This is shown by reducing the case (7.1) to the previous one with the following Lemma 7.1. If \( \{X_n\} \) satisfies (7.1) and \( \mu_k(x) = O(x^{k\delta/2}) \), then the process \( \{Y_n\} \) defined by \( Y_n = X_n^{1-\frac{\delta}{2}} \) satisfies
(7.2) \( \lim_{y \to \infty} \tilde{\mu}_2(y) = b(1 - \frac{5}{2})^2 \), \( \lim_{y \to \infty} y \tilde{\mu}_1(y) = (1 - \frac{5}{2})(a - \frac{b^2}{4}) \)

and \( \tilde{\mu}_k(y) = o(1) \), where \( \tilde{\mu}_k(y) \) are the conditional moments for \( \{ \tilde{y}_n \} \) defined in the usual way by (2.2).

Proof. By definition,

\[
(7.3) \quad \tilde{\mu}_k(y) = E[(y_{n+1} - y_n)^k | Y_n = y] = E[(X_n + \Delta X_n)^{1-\frac{5}{2}} - X_n^{1-\frac{5}{2}}]^k | X_n = \frac{2}{y^{2-5}}
\]

Expanding in binomial series we formally obtain

\[
\tilde{\mu}_1(y) = E[(1 - \frac{5}{2})X_n^{-\frac{5}{2}}(\Delta X_n)^{\frac{5}{2}} + \cdots | X_n = \frac{2}{y^{2-5}}] = (1 - \frac{5}{2})(a - \frac{b^2}{4}) y^{-1} + \cdots,
\]

the last by substituting from (7.1) the asymptotic values of \( \mu_1 \) and \( \mu_2 \). Similarly,

\[
\tilde{\mu}_2(y) = E[(1 - \frac{5}{2})^2 X_n^{-\frac{5}{2}}(\Delta X_n)^2 + \cdots | X_n = \frac{2}{y^{2-5}}] = (1 - \frac{5}{2})^2 b + \cdots,
\]

and the boundedness of \( \tilde{\mu}_k(y) \) for \( k > 2 \) is indicated in the same way by the growth conditions on \( \mu_k \). This, of course, is not a complete proof, but the tedious justification is quite similar to that in the appendix of [10] and will be omitted.
With the aid of this lemma, a process \( \{X_n\} \) satisfying (7.1) can be reduced to one obeying the hypotheses of our theorems above. The results can easily be translated back from \( \{Y_n\} \) to \( \{X_n\} \), justifying our earlier assertion. We shall not trouble to explicitly state all the results, but remark that in case \( \delta = 1 \), the limiting diffusion process has the backward equation

\[
(7.4) \quad u_t = - au_x + \frac{b}{2} uu_{xx},
\]

a type of diffusion which has some interest in its own right [5]. It might also be pointed out that Lemma 7.1 allows an extension of the criteria for null versus positive recurrent behavior in Theorems 2.3 and A.1 to process satisfying (7.1) instead of (2.3). This goes also for the recurrence-transience criteria of [10]. It is interesting that the recurrence criterion \( \alpha \leq \beta/2 \) for a process satisfying (2.3) is unchanged for one obeying (7.1), becoming simply \( a \leq b/2 \). One final thing worth noting in this regard is that the restriction \( \delta < 2 \) in (7.1) is essential. If \( \delta = 2 \) an entirely different type of phenomena occurs; this is indicated by the fact that certain branching processes [7] fall into this category. Their asymptotic behavior is of another sort altogether from that of the processes we study here.

Another variation consists of leaving \( \mu_2(x) \) of the order of a constant but altering \( \mu_1(x) \). Of course if \( \mu_1(x) \) goes to zero faster than \( x^{-1} \) we simply have the case \( \alpha = 0 \), but what if it
goes slower? A partial answer is given by

Theorem 7.1. If both \( \mu_2(x) \) and \( \mu_4(x) = O(1) \), if for some \( \delta \in (0,1) \)

\[
(7.5) \quad \lim_{x \to \infty} x^{1-\delta} \mu_1(x) = \alpha > 0 ,
\]

and if \( \lim \sup X_n = \infty \) a.s., then

\[
(7.6) \quad \lim_{n \to \infty} \frac{X_n}{\frac{1}{n^{2-\delta}}} = [\alpha(2-\delta)]^{\frac{1}{2-\delta}} \text{ in probability.}
\]

Proof. The assumptions of the theorem imply according to [10] that \( X_n \to \infty \) a.s. as \( n \to \infty \), so that \( \{X_n\} \) is null. The now familiar calculations with moments are therefore possible; it is expedient to begin with the moment of order \( 2-\delta \). We have

\[
\Delta \mathbb{E}(X_n^{2-\delta}) = \mathbb{E}(\mathbb{E}\left((X_n + \Delta X_n)^{2-\delta} \mid X_n\right)) = \mathbb{E}((2-\delta)X_n^{1-\delta} \mu_1(X_n) + o(X_n^{-\delta}) + \cdots)
\]

using the assumption that \( \mu_2(x) = O(1) \). Since the process is null we have using (7.5) that

\[
\lim_{n \to \infty} \Delta \mathbb{E}(X_n^{2-\delta}) = (2-\delta)\alpha ,
\]

and so

\[
(7.7) \quad \lim_{n \to \infty} \mathbb{E}\left(\frac{X_n^{2-\delta}}{n}\right) = (2-\delta)\alpha .
\]
Of course this formula has not been fully proved, but again the details are uninteresting and fairly standard and will be omitted. Similarly, we find that asymptotically for large \( n \),

\[
\Delta \mathbb{E}(X_n^{4-2\delta}) \sim (4-2\delta) \alpha \mathbb{E}(X_n^{2-\delta}) \sim 2\alpha^2 (2-\delta)^2 n.
\]

This implies that

\[
\lim_{n \to \infty} \frac{X_n^{4-2\delta}}{\mathbb{E}(\frac{X_n^{2-\delta}}{n})} = \alpha^2 (2-\delta)^2,
\]

which, together with (7.7), implies (7.6).

This result immediately suggests two additional questions: one is whether or not (7.6) can be strengthened to almost sure convergence; the other whether a non-degenerate limit distribution can be obtained under hypotheses related to those of Theorem 7.1 by subtracting an appropriate function of \( n \) from \( X_n \) before normalizing. A superficial look at the latter question has suggested that some estimate of the rate of convergence in (2.1) may be required to obtain a positive answer.

Another natural potential extension is to Markov processes on the whole real line. The difficulty here is that our work has all been with even moments, which will no longer determine distributions uniquely when \( \{X_n\} \) can take both positive and negative values.
Although the invariance principle has not been extended in this way, it is not difficult to generalize Theorem 2.1: if \( \{X_n\} \) is a null process, each \( \mu_k(x) \) is bounded, (2.3) holds as \( x \) tends to both \( +\infty \) and \( -\infty \), and if

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \Pr(X_n > 0) = p
\]

exists, then \( X_n/\sqrt{n} \) has a limiting distribution given by the density \( pg(y) \) for \( y \geq 0 \), \( (1-p) g(-y) \) for \( y < 0 \), where \( g(y) \) is the density in (2.4). The idea of the proof is to imitate the proof of Theorem 2.1 using the process \( \{X^+_n\} \). The limiting distribution of \( X^+_n/\sqrt{n} \) is even uniquely determined by its moments, and the previous techniques together with (7.9) suffice to calculate these. The same goes for \( \{X^-_n\} \), and the combined answer is as stated. It may perhaps be that (7.9) is always valid when the other hypotheses are satisfied. While this has not been proved, there are certain cases where the condition must hold.

Aside from the obvious case of symmetry \( (p=\frac{1}{2}) \), (7.9) holds if \( \alpha > \beta/2 \).

The reason is that, again by [10], \( |X_n| \to \infty \) in this case and there are only a finite number of changes of sign. The value of \( p \) is usually not easily ascertainable, however.

In conclusion, two other problems may be mentioned. One is, of course, the extension of these results to multivariate Markov processes. The other is the relaxing of the unpleasant assumption that all the \( \mu_k(x) \) exist. It would seem plausible that the boundedness of \( \mu_{2+\epsilon}(x) \)
for some \( \epsilon > 0 \) might be sufficient, as it was in [10] and as it is for the C.L.T. and Prokhorov's generalization in [12] of Donsker's theorem. Actually, our present approach does not really require all the \( \mu_k(x) \) to be bounded; a reexamination shows that

\[
(7.10) \quad \mu_k(x) = o(x^{k-2}) \quad k = 3, 4, \ldots
\]

is actually sufficient. However, since this is certainly not yet the "right" condition this bit of generality did not seem worth the trouble it would cost. The generalization to the Liapounov-type condition might be achieved by a truncation argument, but it appears more likely that a quite different method of proof will ultimately replace that used here.

Appendix - Null processes.

The purpose of this appendix is to show that under certain conditions the "null" and "uniformly null" hypotheses of Theorems 2.1 and 4.1 respectively are automatically fulfilled. This follows from the following:

**Theorem A.1.** Suppose \( \{X_n\} \) is an irreducible Markov chain with a countably-infinite number of states, which are positive numbers with no finite point of accumulation; suppose also that (2.3) holds and that \( \mu_k(x) = o(x) \). Then \( \{X_n\} \) is null.
Theorem A.2. If \( \{X_n\} \) is an irreducible, countable Markov chain either null-recurrent or transient, then for each state \( y \)

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{\ell=1}^{n} \Pr(X_\ell = y | X_0 = x) = 0
\]

(A.1)

holds uniformly in the initial state \( x \).

Remarks. Theorem A.1 is a partial converse to Theorem 2.3 and is also a generalization of facts known in the random walk case. These results establish (when they apply) \( \alpha = -\beta/2 \) as the borderline between positive and null recurrence; it is shown in [10] that under the conditions of Theorem 2.3, \( \alpha = \beta/2 \) is the borderline between null-recurrence and transience. The uniformity in Theorem A.2 does not hold in general without the Cesaro averaging, as can easily be shown by an example. Combining the two, we see that under the conditions of Theorem A.1 \( \{X_n\} \) is in fact uniformly null.

Proof of Theorem A.1. We need to show that \( \{X_n\} \) can not be positive-recurrent. Because of (2.3) we know that there is an \( \epsilon > 0 \) such that, for all large \( x \),

\[
\mu_2(x) + 2x\mu_1(x) \geq \epsilon ;
\]

(A.2)

assume temporarily that this inequality holds for all states \( x \).
Proceeding just as in the proof of Lemma 2.1, we find

$$\Delta E(X_n^2) = E\{\mu_2(X_n) + 2X_n\mu(X_n)\} \geq \epsilon$$

and summing shows that $E(X_n^2) \geq \epsilon n$. It is also clear that $E(X_n^2) = O(n^{1/2})$, and by continuing in the manner of Lemma 2.2, we find that $E(X_n^4) = O(n^2)$; in showing this we use the assumption that $\mu_4(x) = O(x)$.

We need an inequality due to Cantelli: if $X$ is a random variable with $E(X^2) = a$ and $E(X^4) = b$; then for $c < \sqrt{a}$,

$$\Pr(|X| \geq c) \geq \frac{(a-c^2)^2}{b-2ac+c^4}.$$

If this is applied to $X_n$ using the above estimates for the second and fourth moments and with $c = c_n = o(\sqrt{n})$, the result is

$$\Pr(X_n \geq c_n) \geq \frac{en^2 + o(n^2)}{o(n^2)} \geq \delta > 0.$$

But $c_n$ can be chosen to be unbounded, so that for any finite interval $I$

$$\lim \sup_{n \to \infty} \Pr(X_n \in I) \leq 1 - \delta < 1,$$

this shows that $\{X_n\}$ is not positive recurrent.

* The initial state is fixed independent of $n$. 
It remains to consider the chains for which (A.2) only holds for all \( x \geq M \) say. Given such a chain, we can construct another with the same states, say \( \{X'_n\} \), such that (A.2) holds for all states and such that the transition probabilities \( P'_{xy} \) are the same as they were for \( \{X_n\} \) unless \( x \), the initial state of the transition, is \( \leq M \).

To do this it is only necessary to select any state \( x_e \) such that \( x_e \geq M + \sqrt{\varepsilon} \), and replace the transitions from each \( x \leq M \) by certain passage into \( x_e \). It is clear that irreducibility is not affected, that (A.2) now holds without exception, and that the growth properties of \( \mu_1(x) \) are unchanged.

From the first part of the proof, we can conclude that \( \{X'_n\} \) is null. To see that the same is true of \( \{X_n\} \) the following is useful:

**Lemma.** If \( A \) and \( B \) are finite sets of states in a denumerable irreducible Markov chain, and if the expected first-passage times from \( x \) to \( B \) and from \( y \) to \( A \) are finite for every \( x \in A \) and \( y \in B \), then the chain is positive.*

**Proof.** The recurrence time \( T_A \) of set \( A \), starting from any \( x \in A \), is not greater than the sum of first-passage times from \( A \) to \( B \) and from \( B \) to \( A \), and hence \( E(T_A) \) is finite. So, then, is \( E(T_A^{(1)} + T_A^{(2)} + \cdots + T_A^{(n)}) \) for any \( n \); let \( n \) equal or exceed the number of states in \( A \). Then during the time of \( n \) recurrences of \( A \), there is a recurrence of some individual state of \( A \). In a null chain, however, this would require infinite expected time.

* This is part of a well-known theorem, if \( A \) and \( B \) consist of single states.
Apply the lemma to \( \{X_n'\} \) with \( A \) the set of states less than \( M \) and \( B = \{x_e\} \). Clearly the expected first-passage time from \( A \) to \( B \) is finite (equal to one!) but \( \{X_n'\} \) is null, so the first-passage time from \( x_e \) to \( A \) has infinite mean. The same must be true of the original chain because the transition probabilities entering into the first passage distribution from \( x_e \) to \( A \) were not altered in the construction of \( \{X_n'\} \). This would be impossible if \( \{X_n\} \) were positive recurrent.

\textbf{Proof of Theorem A.2.} Let us write the conditional probability in (A.1) as \( P_{xy}^{(\ell)} \) and the corresponding first-passage probability as \( f_{xy}^{(\ell)} \); of course, \( x \) and \( y \) belong to a countable set. Since the chain \( \{X_n\} \) is null recurrent or transient, \( P_{xy}^{(\ell)} \to 0 \) as \( \ell \to \infty \) for each \( x \) and \( y \). Choose \( M \) so that \( P_{yy}^{(\ell)} < \epsilon \) for all \( \ell \geq M \). Now for any \( \ell \),

\[
P_{xy}^{(\ell)} = \sum_{i=0}^{\ell-M} f_{xy}^{(i)} P_{yy}^{(\ell-i)} + \sum_{i=\ell-M+1}^{\ell} f_{xy}^{(i)} P_{yy}^{(\ell-i)}.
\]

In the first sum (which may be empty) each \( P_{yy}^{(\ell-i)} \leq \epsilon \), and since the \( f_{xy}^{(i)} \) sum to a value of \( \leq 1 \), the whole sum is \( \leq \epsilon \). The second sum is over-estimated by replacing \( P_{yy}^{(\ell-i)} \) by one, and so

\[
\sum_{\ell=1}^{n} P_{xy}^{(\ell)} \leq \epsilon \quad \sum_{\ell=1}^{n} \sum_{i=\ell-M+1}^{\ell} f_{xy}^{(i)} \leq \epsilon \quad n + M,
\]

the last inequality following upon reversing order of summation. Since this estimate can be made for any \( \epsilon > 0 \), uniformly in \( x \), the uniform limit (A.1) follows.
REFERENCES


* An English translation of this Journal has been prepared by the Society for Industrial and Applied Mathematics.