Adv. Appl. Prob. **18**, 20–65 (1986)
Printed in N. Ireland

WEAK CONVERGENCE OF SEQUENCES OF SEMIMARTINGALES WITH APPLICATIONS TO MULTITYPE BRANCHING PROCESSES

A. JOFFE*, *Université de Montréal*
M. METIVIER**, *École Polytechnique, Palaiseau*

Abstract

The paper is devoted to a systematic discussion of recently developed techniques for the study of weak convergence of sequences of stochastic processes. The methods described make essential use of the semimartingale structure of the processes. Sufficient conditions for tightness including the results of Rebolledo are derived. The techniques are applied to a special class of processes, namely the D-semimartingales. Applications to multitype branching processes are given.

WEAK CONVERGENCE; TIGHTNESS; SEMIMARTINGALES; D-SEMIMARTINGALES; MULTITYPE BRANCHING PROCESS

Introduction

The purpose of this paper is to study weak convergence of sequences of stochastic processes; we rely on recently developed techniques which depend on a special structure of the processes considered, namely the semimartingale property. In order to reach a wider audience, a significant part of the paper is expository but some of the results and many of the applications are new.

We are concerned with the following problems: let $(X^n)_{n\in N}$ be a sequence of processes whose trajectories $X^n(\cdot, w): t \to X(t, w)$ are right-continuous mappings from R^+ into a Polish space H, with left limits at every point $t \in R^+$. In other words, the paths $X^n(\cdot, w)$ are elements of a function space $D(R^+, H)$; thus the $X^{(n)}$'s are random elements with values in $D(R^+; H)$, which is assumed to be endowed with the classical Skorokhod topology. The law $\tilde{P}^n$ of the process X^n is the law induced by X^n on the σ-algebra of Borel subsets in $D(R^+; H)$.

Received 12 March 1984; revision received 31 January 1985.
* Postal address: Département de Mathématiques et Statistique, Université de Montréal, P.O. Box 6128, Succ. A, Montréal, Québec H3C 3J7, Canada.
** Postal address: Centre de Mathématiques Appliquées, École Polytechnique, Cedex 91128, Palaiseau, France.
Research supported in part by the Natural Science and Engineering Research Council, Canada, the FCAC Programme of the Ministère de l'Education du Québec, and the Air Force Office of Scientific Research Grant No. F49620-82-C-0009, USA.

The fundamental problem of weak convergence of processes is the study of conditions under which the laws $\tilde{P}^n$ of a sequence of processes converge weakly and also to identify the limit law $\tilde{P}$. (Weak convergence will be understood in the following sense: for every continuous bounded real function ϕ on $D(R^+; H)$, $\lim_{n\to\infty} \int \phi \, d\tilde{P}^n = \int \phi \, d\tilde{P}$.) The study of the convergence of a given sequence (X^n) is usually accomplished in two stages:

(i) One first shows that the sequence $\tilde{P}^n$ has weak limits, i.e. it is weakly compact. Since $D(R^+, H)$ is complete, metrisable and separable, this conclusion is known to be equivalent to the following one called 'tightness': for every $\varepsilon > 0$ there exists a compact set K_ε in $D(R^+, H)$ such that $\inf_n \tilde{P}^n(K_\varepsilon) \geqq 1 - \varepsilon$.

(ii) One proves that the limit of any convergent subsequence $(\tilde{P}^{n_k})_{k>0}$ of $(\tilde{P}^n)_{n>0}$ satisfies properties which imply its uniqueness, thus obtaining the existence of the limit $\tilde{P}$ and, at the same time, its characterization.

Necessary and sufficient conditions for tightness are well known; the reader is referred for example to the book of Billingsley (1969). These conditions do not refer to any special structure of the process (except for the regularity of their paths). This general approach is often difficult to apply in particular situations. For these reasons sufficient conditions, particularly those involving moments, have been derived and extensively used.

The point of view we take is that, in most particular situations, the processes X^n have a special structure, which allows us to write them as $X^n = A^n + M^n$ where M^n is a local martingale and A^n is a process with paths of finite variation. Such a process X^n is called a semimartingale and often we have specific information about A^n and M^n which can be used to verify tightness and in many cases, to obtain the limit of the sequence; $\langle M^n \rangle$, the increasing process associated to M^n by the Doob–Meyer decomposition, will play an important role.

The main purpose of the paper is to present and apply results which use the semimartingale structure; our work is divided into four parts.

In Section 1, classical situations illustrate the main ideas. Since some elementary calculations of stochastic calculus are needed, a brief account of this subject is included.

In Section 2, sufficient conditions for tightness are obtained and applied to some examples. The most significant result is the theorem of Rebolledo which gives a tightness criterion for X^n in terms of the processes A^n and $\langle M^n \rangle$, which in the usual examples, have an integral expression allowing an easy verification of Rebolledo's hypothesis. This part uses materials included in the unpublished manuscript by Métivier (1979).

The class of D-martingales discussed in Section 3 is a class of processes for which a formula analogous to the Dynkin formula for a Markov process holds. Roughly speaking, a process X with values in R^d is a D-semimartingale if to every bounded function φ of class C^2 on R^d is associated a process $L(\varphi, z, t, \cdot)$ such that for some non-decreasing function A the process

$$M_t^\varphi := \varphi(X_t) - \varphi(X_0) - \int_0^t L(\varphi, X_{s-}, s, \cdot)\, dA_s$$

is a square-integrable martingale.

For sequences of such processes tightness is implied by simple assumptions on the local coefficients $b^i(x, t, \cdot)$ and $a^{ij}(x, t, \cdot)$ which are defined in terms of $L(\varphi^i, x, t, \cdot)$ and $L(\varphi^i\varphi^j, x, t, \cdot)$ where $\varphi^i(x) = x^i$ is the ith coordinate of x. The limits are identified by looking at the corresponding martingale problem in the sense of Stroock and Varadhan.

Section 4 deals with multitype branching processes. In his pioneering work Feller (1951) showed that a sequence of critical or approximately critical branching processes properly normalized has finite joint distributions which converge weakly to the corresponding distribution of a certain diffusion. Tightness and weak convergence were proved much later. Most of the literature was devoted to the study of one-type processes.

The above methods are here applied to the study of multitype branching processes, and results are readily obtained which are valid under minimal assumptions on the moments of the processes considered.

1. Why semimartingales? A brief review of some stochastic calculus

1.1. *Why semimartingales*?

1.1.1. *Typical elementary classical case.* Let $\boldsymbol{X}$ be a pure homogeneous jump Markov process in R^d with law of jumps μ. This means that the infinitesimal generator L of $\boldsymbol{X}$ has the form

$$L\phi(x) = \lambda \int_{\mathbb{R}^d} [\phi(x+y) - \phi(x)]\mu(dy).$$

We change the units of length and time so that at time t the position of the particle previously described by $\boldsymbol{X}_t$ is now described by $\boldsymbol{X}_t^n = \varepsilon_n \boldsymbol{X}_{t/\alpha_n}$, (where the unit of length has been multiplied by $1/\varepsilon_n$ and the unit of time by $1/\alpha_n$) whose generator is given by

$$L^n\phi(\boldsymbol{x}) = \lambda/\alpha_n \int_{\mathbb{R}^d} [\phi(\boldsymbol{x} + \varepsilon_n \boldsymbol{y}) - \phi(\boldsymbol{x})]\mu(d\boldsymbol{y}).$$

What are the limit laws of $\boldsymbol{X}^n$ as $n \uparrow \infty$?

1.1.2. *The semimartingale property.* Let us consider a sequence of Markov processes $\boldsymbol{X}^n$ with generator L^n. Set $\phi^i(\boldsymbol{x}) = x^i$ and assume ϕ^i and $\phi^i\phi^j \in \mathscr{D}(L^n)$, the domain of L^n. Set

$$b^{n,i}(\boldsymbol{x}) = L^n\phi^i, \tag{1.1.1}$$

$$a^{n,ij}(\boldsymbol{x}) = L^n\phi^i\phi^j - \phi^i L^n\phi^j - \phi^j L^n \phi^i. \tag{1.1.2}$$

The Dynkin formula says that for all $\phi \in \mathscr{D}(L^n)$ with $E(\phi(\boldsymbol{X}_t^n)) < \infty$ and for all t,

$$\phi(\boldsymbol{X}_t^n) = \phi(\boldsymbol{X}_0^n) + \int_0^t L^n\phi(\boldsymbol{X}_{s-}^n)\, ds + M_t^n(\phi) \tag{1.1.3}$$

where $M^n(\phi)$ is a martingale. (See for instance Métivier (1982).) If one does not know whether $E(\phi(\boldsymbol{X}_t^n)) < \infty$, but assumes for example that the jumps of $\boldsymbol{X}_t^n$ are bounded by $a > 0$, consider a smooth function $\phi^{k,i}$ such that

$$\phi^{k,i}(\boldsymbol{x}) := \begin{cases} x^i & \text{if } |\boldsymbol{x}| \leqq k + a \\ 0 & \text{if } |\boldsymbol{x}| > k + a + 1, \end{cases}$$

$\phi^{k,i}(\boldsymbol{x})$ being twice continuously differentiable and bounded. Set

$$\tau_k^n := \inf\{t : |\boldsymbol{X}_t^n| \geqq k\}.$$

Because of the assumption on jumps

$$\sup_{t \leqq \tau_k^n} |\boldsymbol{X}_t^n| \leqq k + a,$$

and therefore, using the Dynkin formula again and also the stopping theorem for martingales,

$$\begin{aligned} \boldsymbol{X}_{t\wedge\tau_k^n}^{n,i} = \phi^{k,i}(\boldsymbol{X}_{t\wedge\tau_k^n}^n) &= \boldsymbol{X}_0^n + \int_0^{t\wedge\tau_k^n} L^n\phi^{k,i}(\boldsymbol{X}_s^n)\, dx + \text{martingale} \\ &= \boldsymbol{X}_0^n + \int_0^{t\wedge\tau_k^n} b^{n,i}(\boldsymbol{X}_s^n)\, ds + \text{martingale}. \end{aligned}$$

In other words, for an increasing sequence $(\tau_k^n)_{k\geqq 0}$ of stopping times, with $\lim_n \tau_n = +\infty$ a.s.

$$\begin{cases} \boldsymbol{X}_t^n = \boldsymbol{X}_0^n + \int_0^t b^n(\boldsymbol{X}_s^n)\, ds + \boldsymbol{M}_t^n, \text{ where } \boldsymbol{b}^n \text{ denotes the vector } b^{n,i}, \\ \qquad\qquad\qquad i = 1, \cdots, d \\ \boldsymbol{M}_{t\wedge\tau_k^n}^n : \text{martingale for all } k \text{ (i.e. } \boldsymbol{M}^n \text{ is a local martingale).} \end{cases} \tag{1.1.4}$$

Definition. A process $\boldsymbol{X}$ which can be written as $\boldsymbol{X} = \boldsymbol{V} + \boldsymbol{M}$, where V has path with finite variation and $\boldsymbol{M}$ is a local martingale, is called a *semimartingale*.

1.1.3 *More general remarks.* When the processes (X^n) are not Markovian, representations of the type (1.1.3) may be replaced in many cases by a formula of the form

$$\phi(X_t^n) = \phi(X_0^n) + \int_0^t L^n(\phi, X_s^n, s, \omega)\, dA_s^n + M_t^n(\phi), \tag{1.1.5}$$

where $L^n(\phi, x, s, \omega)$ still depends linearly on ϕ but may depend also on (s, ω) in a suitable measurable way. For example, for many pure-jump non-Markovian processes we have an L^n of the form

$$L^n(\phi, x, s, \omega) = \int (\phi(x+y) - \phi(x))\nu^n(s, \omega, x; dy).$$

Therefore defining $\quad b^{n,i}(x, s, \omega) = L^n(\phi^i, x, s, \omega),$

we get as above the semimartingale property of X^n.

Example. $(X_k^n)_{k>0}$ is a sequence of Markov chains with transition Π^n. $(X_t^n)_{t\geq 0}$ is the right-continuous step interpolation of X^n, and we set

$$\tilde{X}_t^n = \varepsilon_n X_{[nt]}^n$$

and denote by $\tilde{\Pi}^n$ the transition of the Markov chain $(\varepsilon_n X_k^n)_{k\geq 0}$. We may write for every bounded measurable ϕ

$$\phi(\tilde{X}_t^n) = \phi(\tilde{X}_0^n) + \int_0^t n(\tilde{\Pi}^n - I)\phi(\tilde{X}_{s^-}^n)\, dA_s^n,$$

with $A_t^n = [\eta t]/\eta$. Setting $L^n\phi(x) = n(\tilde{\Pi}^n - I)\phi(x)$, we see that we are in the situation just described.

1.2. *A brief review of stochastic calculus.* In practice, the processes that one meets are most often semimartingales. There is a stochastic calculus for these processes: representation formulas of considerable usefulness will play a fundamental role in the practical rules derived below. For more details the reader is referred to Jacod (1979) or Métivier (1982).

1.2.1. *Stochastic integrals.* We recall that if $\boldsymbol{Z}$ is a semimartingale we have a stochastic integration theory available. This means that, for a wide class of processes Φ, in particular those which are predictable† and locally bounded,

† Predictability is actually defined as follows. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(\mathcal{F}_t)_{t\geq 0}$ be a filtration of sub-σ-algebras of $\mathcal{F}$. Consider on $\mathbb{R}^+ \times \Omega$ the σ-algebra $\mathcal{P}$ generated by the sets of the form $(s, t] \times F$ where $0 \leq s \leq t$, $F \in \mathcal{F}_s$ ('predictable rectangles'). The elements of $\mathcal{P}$ are called predictable sets, and a process $X:(t, \omega) \to X(t, \omega)$ on Ω is called predictable if it is a $\mathcal{P}$-measurable function.

For most of what follows, it is enough to remember that *processes which are adapted to* $(\mathcal{F}_t)$ and left-continuous, are predictable.

the stochastic integral $\int \Phi\, d\boldsymbol{Z} := (\int_0^t \Phi_s\, d\boldsymbol{Z}_s)_{t\geq 0}$ is defined as a right-continuous semimartingale, the mapping $\Phi \to \int \Phi\, d\boldsymbol{Z}$ being linear with continuity properties analogous to those of an ordinary integral (e.g. the dominated convergence theorem, see Métivier (1982)).

Moreover, if $\boldsymbol{Z}$ is a local martingale, then $\int \Phi\, d\boldsymbol{Z}$ is also a local martingale (even an L^p-martingale if Φ is bounded and $\boldsymbol{Z}$ is an L^p-martingale).

1.2.2. *Quadratic variation.* If Z^i and Z^j are two real martingales (or two coordinate processes of an $\mathbb{R}^d$-valued semimartingale $\boldsymbol{Z}$) then we have the following formula of *integration by parts*:

$$Z_t^i Z_t^j = Z_0^i Z_0^j + \int_0^t Z_{s^-}^i\, dZ_s^j + \int_0^t Z_{s^-}^j\, dZ_s^i + [Z^i, Z^j]_t \tag{1.2.1}$$

where $[Z^i, Z^j]_t$ is a process with paths of finite variation known as the *mutual variation* of Z^i and Z^j. If $i = j$, $[Z^i, Z^j]_t$ is an increasing process called the *quadratic variation* of Z^i.

It may be helpful to remember that for every t

$$[Z^i, Z^j]_t = \lim_{\delta(\Pi)\downarrow 0} \Pr \sum_{t_n \in \Pi} (Z_{t_{n+1}\wedge t}^i - Z_{t_n\wedge t}^i)(Z_{t_{n+1}\wedge t}^j - Z_{t_n\wedge t}^j) \tag{1.2.2}$$

where Π is a subdivision $0 = t_0 < t_1 < \cdots < t_n < \cdots$ of $\mathbb{R}$ whose mesh $\delta(\Pi) = \sup_n (t_{n+1} - t_n)$ tends to 0. As a consequence of this formula it is easily seen that if Z^i has paths with finite variation,

$$[Z_i, Z_j] = \sum_{s\leq t} \Delta Z_s^i \Delta Z_s^j, \quad \text{(where} \quad \Delta Z_s = Z_s - Z_{s^-}). \tag{1.2.3}$$

When $\boldsymbol{Z}$ is an $\mathbb{R}^d$-valued process the formula (1.2.1) can be written

$$\boldsymbol{Z}_t \otimes \boldsymbol{Z}_t = \boldsymbol{Z}_0 \otimes \boldsymbol{Z}_0 + \int_0^t \boldsymbol{Z}_{s^-} \otimes d\boldsymbol{Z}_s + \int_0^t d\boldsymbol{Z}_s \otimes \boldsymbol{Z}_{s^-} + [\![\boldsymbol{Z}]\!]_t \tag{1.2.4}$$

where $[\![\boldsymbol{Z}]\!]_t$ denotes the matrix-valued process $([Z^i, Z^j]_t)$.

For such a vector process one also defines

$$[\boldsymbol{Z}]_t := \operatorname{trace} [\![\boldsymbol{Z}]\!]_t = \sum_{i=1}^d [Z^i, Z^i]_t \tag{1.2.5}$$

and observes that

$$|\boldsymbol{Z}_t|^2 = |\boldsymbol{Z}_0|^2 + 2\int_0^t \boldsymbol{Z}_{s^-} \cdot d\boldsymbol{Z}_s + [\boldsymbol{Z}]_t \tag{1.2.6}$$

(denoting by $\cdot$ the scalar product in $\mathbb{R}^d$).

1.2.3. *The Meyer process of a semimartingale.* According to a decomposition theorem due to Meyer (see Grigelionis (1973), Métivier (1982)), and assuming that the process $([Z_i, Z_j]_t)_{t\geq 0}$ is locally integrable, there exists a unique predictable process, with paths of finite variation, denoted by $\langle Z_i, Z_j\rangle_t$, such that

$$([Z^i, Z^j]_t - \langle Z^i, Z^j\rangle_t)_{t\geq 0} \tag{1.2.7}$$

is a local martingale.

For an $\mathbb{R}^d$-valued semimartingale Z we write $\langle\!\langle Z\rangle\!\rangle$ for the matrix-valued process $(\langle Z^i, Z^j\rangle)_{i,j=1,\cdots,d}$ and

$$\langle \boldsymbol{Z}\rangle := \text{trace } \langle\!\langle Z\rangle\!\rangle. \tag{1.2.8}$$

Consequently

$$[\boldsymbol{Z}] - \langle \boldsymbol{Z}\rangle \text{ is a local martingale.} \tag{1.2.9}$$

The interest of the Meyer process is that it is rather regular and usually carries a lot of information about the law of $\boldsymbol{Z}$ (usually more than $[\boldsymbol{Z}]$). For example, if N is a Poisson process, it follows from (1.2.3) that $[N] = N$, but $N^t - t$ is immediately seen to be a martingale, which gives $\langle N_t\rangle = t$ and it can be proven (see, for example, Grigelionis (1973) or Métivier (1982)) that the only pure jump process N with jumps of amplitude 1 such that $\langle N\rangle = t$ is precisely the Poisson process.

1.2.4. *A small calculation.* Let us return to the situation described in 1.1.3 with $A_t^n = t$ for simplicity. Thus (dropping the n, for the rest of the calculation) we have

$$\phi(\boldsymbol{X}_t) = \phi(\boldsymbol{X}_0) + \int_0^t L(\phi, s, \omega, \boldsymbol{X}_s)\, ds + \boldsymbol{M}_t(\phi). \tag{1.2.10}$$

Set

$$b^i(\boldsymbol{x}, s, \omega) := L(\phi^i, s, \omega, \boldsymbol{x})$$

$$a^{ij}(\boldsymbol{x}, s, \omega) := L(\phi^i\phi^j, s, \omega, \boldsymbol{x}) - x^i b^j(s, \omega, \boldsymbol{x}) - x^j b^i(s, \omega, \boldsymbol{x}).$$

As already seen

$$\boldsymbol{X}_t = \boldsymbol{X}_0 + \int_0^t \boldsymbol{b}(\boldsymbol{X}_s, s, \omega)\, ds + \boldsymbol{M}_t. \tag{1.2.11}$$

Applying (1.2.10) again to $\phi^i\phi^j$:

$$\begin{aligned} X_t^i X_t^j = X_0^i X_0^j &+ \int_0^t (b^i(\boldsymbol{X}_s, s, \omega) X_s^j + b^j(\boldsymbol{X}_s, s, \omega) X_s^i)\, ds \\ &+ \int_0^t a^{ij}(\boldsymbol{X}_s, s, \omega)\, ds + \text{loc. mart.} \end{aligned} \tag{1.2.12}$$

But, the integration by parts formula (1.2.1) and (1.2.10) together tell us

$$\begin{aligned} X_t^i X_t^j = X_0^i X_0^j &+ \int_0^t X_s^i b^j(\boldsymbol{X}_s, s, \omega)\, ds + \int_0^t X_{s-}^i\, dM_s^j \\ &+ \int_0^t X_{s-}^j b^i(\boldsymbol{X}_s, s, \omega)\, ds + \int_0^t X_{s-}^j\, dM_s^i + [X^i X^j]_s. \end{aligned} \tag{1.2.13}$$

From (1.2.3) and the continuity of the paths of $(\int_0^t b(\boldsymbol{X}_s, s, \omega)\, ds)_{t\geq 0}$ we obtain

$$[M^i, M^j]_t = [X^i, X^j]_t.$$

Dropping the local martingale part in (1.2.13) we can write

$$X_t^i X_t^j = X_0^i X_0^j + \int_0^t (b^i(\boldsymbol{X}_s, s, \omega) X_s^j + b^j(\boldsymbol{X}_s, s, \omega) X_s^i)\, ds + [M^i, M^j]_t. \tag{1.2.14}$$

This formula compared with (1.2.12) gives

$$[M^i, M^j]_t - \int_0^t a^{ij}(\boldsymbol{X}_s, s, \omega)\, ds = \text{loc. mart.}$$

From this formula and the definition in Section 1.2.3 we obtain (since $(\int_0^t a^{ij}(\boldsymbol{X}_s, s, \omega)\, ds)_{t\geq 0}$ is continuous and therefore predictable):

$$\langle M^i, M^j\rangle_t = \int_0^t a^{ij}(\boldsymbol{X}_s, s, \omega)\, ds. \tag{1.2.15}$$

Thus we obtain the process $\langle\!\langle \boldsymbol{M}\rangle\!\rangle_t$. Its trace

$$\langle \boldsymbol{M}\rangle_t = \int_0^t \text{trace } \boldsymbol{a}(\boldsymbol{X}_s, s, \omega)\, ds, \tag{1.2.16}$$

which will turn out to play a decisive role in what follows, has a simple explicit form.

We shall illustrate this immediately on the example of jump processes given at the beginning.

1.3. *Two simple examples of weak convergence*

1.3.1. *An elementary 'invariance principle'.* We return to the example of a sequence of pure-jump Markov processes as in Section 1.1.1. Take

$$\alpha_n = \varepsilon_n, \qquad \varepsilon_n \downarrow 0.$$

Set

$$\boldsymbol{\beta} := \lambda \int_{\mathbb{R}^d} y\mu(dy)$$

$$\alpha^{ij} := \lambda \int_{\mathbb{R}^d} (y^i - \beta^i)(y^j - \beta^j)\mu(dy) \qquad \bar{\alpha} := \text{trace } (\alpha^{ij}).$$

The calculation of L^n, b^n and a^n as in Section 1.1.2 immediately gives

$$b^n(x) = \beta$$

$$a^{n,i,j}(x) = \varepsilon_n \alpha^{ij}.$$

Therefore

(1.3.1) $$X_t^n := \varepsilon_n X_{t/\varepsilon_n} = X_0^n + \beta t + M_t^n$$

and

(1.3.2) $$\langle M^n \rangle_t = \varepsilon_n \bar{\alpha} t.$$

The Doob inequality gives for every $N > 0$

$$E\Big(\sup_{0 \leqq t \leqq N} |M_t^n|^2\Big) \leqq 4\varepsilon_n \bar{\alpha} N$$

which yields

(1.3.3) $$\lim_{n\to\infty} E\Big(\sup_{t \leqq N} |X_t^n - X_0^n - \beta t|^2\Big) = 0.$$

We want now to study the 'fluctuation process':

(1.3.4) $$\frac{X_t^n - X_0^n - \beta t}{\sqrt{\varepsilon_n}} = (1/\sqrt{\varepsilon_n}) M_t^n.$$

We know that $Z^n := M^n/\sqrt{\varepsilon_n}$ is a martingale and from (1.3.2)

(1.3.5) $$\langle Z^n \rangle_t = \bar{\alpha} t.$$

From the Doob inequality again

(i) $$E \sup_{t \leqq N} |Z_t^n|^2 \leqq 4\bar{\alpha} N \quad \text{for every } n,$$

(ii) $\sup_{\theta \leqq \delta} E\,|Z_{\tau_n+\theta}^n - Z_{\tau_n}^n| \leqq 4\bar{\alpha}\delta$ for every n and for every stopping time τ_n.

The Rebolledo theorem will imply that just those two inequalities (in fact weaker inequalities) yield the weak compactness (or tightness) of the sequence of laws of the processes M^n.

Let us show now that there is only one possible limit. The Dynkin formula applied to the process (X_t^n) and the function $x \to \phi((x - x_0 - \beta t)/\sqrt{\varepsilon_n})$ gives

$$\begin{aligned}\phi(Z_t^n) &= \phi((X_t^n - x_0^n - \beta t)/\sqrt{\varepsilon_n}) \\ &= \int_0^t -(1/\sqrt{\varepsilon_n})\beta \cdot \nabla\phi(M_s)\, ds \\ &\quad + \int_0^t ds \int_{\mathbb{R}^d} (\lambda/\varepsilon_n)[\phi(M_s^n/\sqrt{\varepsilon_n} + \sqrt{\varepsilon_n}\, y) - \phi(M_s^n/\sqrt{\varepsilon_n})]\mu(dy).\end{aligned}$$

Therefore

(1.3.6) $$\phi(Z_t^n) = \int_0^t \sum_{i,j=1}^d \alpha^{ij}(\partial^2\phi)/(\partial x^i \partial x^j)\phi''(Z_s^n)\, ds + \psi(\varepsilon_n, t) + \text{martingale}$$

where $\lim_{n\to\infty} \|\psi(\varepsilon_n, t)\|_\infty = 0$.

Let us call $\bar{P}_n$ the law of Z^n in the space $\mathbb{D}(\mathbb{R}^+; \mathbb{R}^d)$ and set as usual $\xi_t(\bar{\omega}) = \bar{\omega}(t)$ for each $\bar{\omega} \in \mathbb{D}(\mathbb{R}^+; \mathbb{R}^d)$. On the space $\mathbb{D}(\mathbb{R}^+; \mathbb{R}^d)$ we consider the right-continuous filtration $(\mathscr{D}_t^0)_{t\geqq 0}$ generated by the 'canonical process ξ'. In other words, if $\mathscr{D}_t$ is the σ-algebra generated by $\{\xi_s : s \leqq t\}$ one sets $\mathscr{D}_t = \bigcap_{s>t} \mathscr{D}_s^0$.

Equation (1.3.6) allows us to write for every $F \in \mathscr{D}_s$ and $s \leqq t$

$$\bar{E}_n\Big\{\Big[\phi(\xi_t) - \phi(\xi_s) - \int_0^t \sum_{i,j=1}^d \alpha^{ij}((\partial^2\phi)/(\partial x^i \partial x^j))(\xi_s)\, ds\Big] 1_F\Big\} \leqq \|\psi(\varepsilon_n, t)\|_\infty.$$

Set

$$\Phi(\bar{\omega}) = 1_F(\bar{\omega})[\phi(\xi_t(\bar{\omega}) - \phi(\xi_s(\bar{\omega})))] - \int_s^t \sum_{i,j=1}^d \alpha^{ij}((\partial^2\phi)/(\partial x^i \partial x^j))(\xi_s(\bar{\omega}))\, ds.$$

Consider any convergent subsequence $(\bar{P}_{n_k})_{k\in\mathcal{N}}$ and call $\bar{P}$ its limit. From (ii) it follows easily that for every t, $\bar{E}(|\Delta\xi_t|^2) = 0$ and that for $\bar{P}$-almost all $\bar{\omega}$ the mapping $\bar{\omega} \to \Phi(\bar{\omega})$ is continuous at point $\bar{\omega}$. Therefore the convergence of $(\bar{P}_{n_k})_{k\geqq 0}$ to $\bar{P}$ allows us to write

$$\bar{E}(\Phi) = \lim_n \bar{E}^n(\Phi) = 0.$$

This proves that for the law $\bar{\mathscr{P}}$ and every ϕ twice differentiable with compact support,

$$\Big(\phi(\xi_t) - \phi(\xi_0) - \int_0^t \sum_{i,j=1}^d \alpha^{ij}((\partial^2\phi)/(\partial x^i \partial x^j))(\xi_s)\, ds\Big)_{t\geqq 0}$$

is a martingale.

It is known that there is only one probability law with this property: it is the law of the diffusion with generator

$$\sum_{i,j=1}^d \alpha^{ij}((\partial^2)/(\partial x^i \partial x^j)),$$

(see, for instance, Stroock and Varadhan (1979)).

1.3.2. *An example in branching processes.* As an introduction to Section 4 we consider an example of a Markov jump process which is not homogeneous in space (the intensity of jumps depends on x). We take for X the

Galton–Watson process with one type of particle. Such a process is an $\mathcal{N}$-valued Markov process whose generator L is given by

$$L\phi(x) = \lambda x \sum_{k \geq 0} [\phi(x+k-1) - \phi(x)]\nu(k) \tag{1.3.7}$$

where λ is the 'rate of death' of one particle and $\nu(k)$ is the probability that a dying particle gives birth to k descendants.

We can clearly consider L as operating on functions on $\mathbb{R}^+$ (not only on $\mathcal{N}$). In this way all the processes $X_t^n := \varepsilon_n X_{t/\varepsilon_n}$ have $\mathbb{R}^+$ as state space.

Writing L^n, b^n, $a^{n,j}$ as before and assuming

$$\gamma := \lambda \sum_{k \geq 0} (k-1)\nu(k) < \infty, \qquad \beta := \lambda \sum_{k \geq 0} (k-1)^2 \nu(k) < \infty$$

we easily obtain

$$X_t^n = X_0^n + \int_0^t \gamma X_s^n \, ds + M_t^n,$$

M_t^n being a martingale with

$$\langle M^n \rangle_t = \int_0^t \beta X_s^n \, ds.$$

Let us assume that we are in the *critical case*: ($\gamma = 0$) and $X_0^n = x$ (we consider a system of particles with elementary mass ε_n, but with a fixed total mass at the beginning). Then

$$X_t^n = x + M_t^n$$

is a martingale,

$$\langle M^n \rangle_t = \int_0^t \beta X_s^n \, ds,$$

since $E(X_t^n) = x$ and therefore

$$E\langle M^n \rangle_t \leq \beta x t.$$

One obtains again from the Doob inequality

$$\text{(i)} \qquad E\Big(\sup_{t \leq N} |X_t^n|^2\Big) \leq N\beta x$$

and for every stopping time τ_n

$$\langle M^n \rangle_{\tau_n + \delta} - \langle M^n \rangle_{\tau_n} = \int_{\tau_n}^{\tau_n + \delta} \beta \cdot X_s^n \, ds$$

and for each $\tau_n \leq N$

$$\text{(ii)} \qquad P\{\langle M_n \rangle_{\tau_n + \delta} - \langle M^n \rangle_{t_n} \geq \eta\} \leq P\Big\{\sup_{s \leq N} X_s^n \geq \eta/\beta\delta\Big\} \leq (4\beta\delta^2/\eta^2)Nx.$$

These properties (i) and (ii) will immediately imply tightness.

As regards the convergence, it is easy to derive (as we shall do in Section 4), using the martingale property of $\phi(X_t^n) - \phi(X_0^n) - \int_0^t L^n\phi(X_s)\, ds$, that the sequence $\tilde{P}^n$ of laws of the processes X^n converges weakly to the law $\tilde{P}$ characterized by the following property. For every ϕ twice differentiable bounded on $[0, \infty[$, the process

$$\Big(\phi(\xi_t) - \phi(x) - \int_0^t \beta \xi_s (\partial^2\phi/\partial x^2)(\xi_s)\, ds\Big)_{t \geq 0}$$

is a martingale. This is the law $\tilde{P}$ of diffusion with generator $\beta x(\partial^2/\partial x^2)$ (with absorption at the boundary 0).

2. Sufficient conditions for tightness

In this section we give a detailed exposition of the results which provide sufficient conditions for tightness in terms of the semimartingale structure of the processes considered. The result of widest and most immediate applicability in this respect is the theorem of Rebolledo (1979). The initial Rebolledo proof is quite long and imbedded in a paper containing many variations on the theme; at the same time, Aldous (1978) gave a criterion for tightness which we exploit to present a simpler proof here.

For the sake of completeness, we start by recalling a few facts on the Skorokhod topology; then present the theorems of Aldous and Rebolledo, and briefly mention new developments, without proof, since they will not be used in our applications.

2.1. *Weak compactness of a sequence of probability measures on* $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$. $\mathcal{H}$ will denote a complete separable metric space with a distance d. We write $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$ ($\mathbb{D}([0, N], \mathcal{H})$) for the set of $\mathcal{H}$-valued functions defined on $\mathbb{R}^+$ ($[0, N]$), which are right continuous and have left limits at every $t \in \mathbb{R}^+$ ($t \in [0, N]$). These functions are called *cadlag* in the French terminology, and we shall keep this useful abbreviation.

The space $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$ will be endowed with a topology analogous to the Skorokhod J-topology introduced in Skorokhod (1956) for $\mathbb{D}([0, N]; \mathcal{H})$. The reader is referred to this paper, or to Billingsley (1969) or Parthasarathy (1967) for the properties of the Skorokhod topology on $\mathbb{D}([0, N]; \mathcal{H})$ and to Lindvall

(1973) and Whitt (1980) for details concerning the Skorokhod topology on $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$.

This topology, which turns out to be completely metrizable, is quickly understood by its convergence structure: a sequence $(x_n)_{n\in\mathcal{N}}$ in $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$ converges to x in $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$ if and only if there exists a sequence (λ_n) of homeomorphisms of $\mathbb{R}^+$ such that $\lim_n \lambda_n =$ identity mapping and $x = \lim_n x_n \circ \lambda_n$, both convergences being uniform on every compact subset of $\mathbb{R}^+$.

Let us briefly recall one possible definition of the topology on $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$, as given in Lindvall (1973).† For each $N > 0$ and $x \in \mathbb{D}(\mathbb{R}^+; \mathcal{H})$, call $\Pi_N x$ the function:

$$\Pi_N x(t) := \begin{cases} x(t) & \text{if } t \leqq N \\ [(N+1)-t]x(t) & \text{if } N \leqq t \leqq N+1 \\ 0 & \text{if } t \geqq N+1 \end{cases}$$

and let for $x, y \in \mathbb{D}(\mathbb{R}^+; \mathcal{H})$

$$\delta_N(x, y) := \inf_{\lambda\in\Lambda} \left[\sup_{t\in\mathbb{R}^+} d(\Pi_N x(t), \Pi_n g \circ \lambda(t)) + \sup_{t\in\mathbb{R}^+} |t - \lambda(t)| + \sup_{s\neq t} \left| \log \frac{\lambda(t)-\lambda(s)}{t-s} \right| \right] \tag{2.1.1}$$

where Λ is the set of homeomorphisms of $[0, \infty[$.

It is easily seen that the functions δ_N are pseudo-metrics (they are symmetric and satisfy the triangle inequality) on $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$. The metric structure, which we shall consider on $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$, is the one defined by this family of pseudo-metrics. It may be defined by the metric

$$\delta(x, y) := \sum_{N\in\mathcal{N}} \frac{1}{2^N} (\delta_N(x, y) \wedge 1). \tag{2.1.2}$$

The Skorokhod topology on $\mathbb{D}([0, N]; \mathcal{H})$ can be defined by the metric

$$\tilde{\delta}_N(x, y) := \inf_{\lambda\in\Lambda_N} \left[\sup_{t\leqq N} d(x(t), y \circ \lambda(t)) + \sup_{t\leqq N} |t - \lambda(t)| + \sup_{\substack{s\neq t \\ s,t\leqq N}} \left| \log \frac{\lambda(t)-\lambda(s)}{t-s} \right| \right], \tag{2.1.3}$$

† An alternative rapid way of defining the topology of $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$, suggested to us by N. Dinculeanu, is the following. Let us call $\mathbb{D}([0, \infty]; \mathcal{H})$ the set of cadlag functions from $[0, \infty]$ to $\mathcal{H}$ with the metric complete structure homeomorphic to that of $\mathbb{D}([0, 1]; \mathcal{H})$ through the homeomorphism $\phi \to \phi \circ \psi$ with $\psi(x) = |\log(1+x)|$. For each $N > 0$ consider the mapping Π_N from $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$ into $\mathbb{D}([0, \infty]; \mathcal{H})$ defined by $\Pi_N f(t) = f(t)$ for $t < N$ and $\lim_{s\uparrow N} f(s)$ for $t \geqq N$. The Skorokhod topology on $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$ is the coarsest one for which the mappings Π_N are continuous.

where Λ_N is the set of all increasing homeomorphisms of $[0, N]$. For this metric $\mathbb{D}([0, N], \mathcal{H})$ is complete separable. It is easily seen that, for each N, Π_N is a continuous mapping of $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$ onto a complete subspace of $\mathbb{D}([0, N+1]; \mathcal{H})$, namely the space of cadlag functions from $[0, N+1]$ into $\mathcal{H}$, which are continuous and vanishing at $N+1$.

This implies immediately the completeness of $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$ for the metric (2.1.2) and shows also that a set A in $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$ is compact in $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$ if and only if its projections $\Pi_N(A)$ are compact. This makes it possible to easily deduce compactness criteria in $\mathbb{D}(\mathbb{R}^+: \mathcal{H})$ from known compactness criteria in $\mathbb{D}([0, N]; \mathcal{H})$ (as for example in Billingsley (1974)).

2.1.1. *Remark.* It should be remarked that, if a sequence (x_n) in $\mathbb{D}([0, N], \mathcal{H})$ converges to x, for the metric $\tilde{\delta}_N$, then $\lim_n x_n(N) = x(N)$. For this reason the topology on $D(R^+; \mathcal{H})$ is *not* defined by the family $(\tilde{\delta}_N)$ of pseudometrics.

2.1.2. *Compactness criteria in* $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$. We state a necessary and sufficient condition for the compactness of the closure of a set A in $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$, which, as remarked above, can easily be derived from compactness criteria in the spaces $\mathbb{D}([0, N], \mathcal{H})$.

If $x \in \mathbb{D}(\mathbb{R}^+; \mathcal{H})$ we define for each $N > 0$ and $\delta > 0$

$$w^N(x, \delta) := \inf_{\Pi_\delta} \max_{t_i\in\Pi} \sup_{t_i\leqq s<t<t_{i+1}} d(x(t), x(s))$$

where Π_δ is the set of all increasing sequences $t = t_0 < t_1 < \cdots < t_n < N$ in $\mathbb{R}^+$ with the property $\inf_{t_i\in N} |t_{i+1} - t_i| \geqq \delta$.

A set $A \subset \mathbb{D}(\mathbb{R}^+; \mathcal{H})$ has a compact closure if and only if it satisfies the following two conditions:

(i) There exists a dense subset T of $\mathbb{R}^+$ such that for every $t \in T$ the set $\{x(t) : x \in A\}$ has compact closure in $\mathcal{H}$.

(ii) For every $N > 0$, $\lim_{\delta\downarrow 0} \sup_{x\in A} w^N(x, \delta) = 0$.

2.1.3. *Weak compactness of sequences of probability measures.* On $\tilde{\Omega} := \mathbb{D}(\mathbb{R}^+; \mathcal{H})$ we consider the σ-algebra $\tilde{\mathcal{B}}$ of Borel sets associated with the previously defined topology. We denote also by $(\xi_t)_{t\geqq 0}$ the 'projection process' or 'canonical process': i.e., for $\tilde{\omega} \in \tilde{\Omega}$, $\xi_t(\tilde{\omega}) := \tilde{\omega}(t)$.

A family $(\tilde{P}_n)_{n\geqq 0}$ of probability measures on $(\tilde{\Omega}, \tilde{\mathcal{B}})$ is said to *converge weakly* to $\tilde{P}$ if for every bounded continuous function Φ on $\tilde{\Omega}$

$$\lim_{n\to\infty} \int \Phi \, d\tilde{P}_n = \int \Phi \, d\tilde{P}.$$

A family $(\tilde{P}_n)_{n\geqq 0}$ is said to be relatively weakly compact if every subsequence admits a weakly convergent subsequence. A well-known necessary and sufficient condition for the relative weak compactness of the sequence $(\tilde{P}_n)_{n\geqq 0}$, due to Prokhorov (see Parthasarathy (1967)) is the 'tightness' of the sequence $(\tilde{P}_n)_{n\geqq 0}$: for every $\varepsilon>0$ there exists a compact subset K_ε of $\tilde{\Omega}$ such that $\inf_n \tilde{P}_n(K_\varepsilon)\geqq 1-\varepsilon$.

The above characterization of sets with compact closure in $\tilde{\Omega}$ leads to the following criteria for tightness (for the analogous statement in $\mathbb{D}([0,1];\mathscr{H})$ see Billingsley (1968), p. 117).

Theorem (basic criterion for tightness). A family $(\tilde{P}_n)_{n\geqq 0}$ of probabilities on $\tilde{\Omega}:=\mathbb{D}(\mathbb{R}^+;\mathscr{H})$ is tight if and only if the two following conditions hold:

(i) There exists a dense set T in $\mathbb{R}^+$ such that for every $t\in T$ and $\varepsilon>0$ there exists a compact set $C(t,\varepsilon)$ in $\mathscr{H}$ such that $\sup_n \tilde{P}_n\{\tilde{\omega}: \xi_t(\tilde{\omega})\notin C(t,\varepsilon)\}\leqq\varepsilon$.

(ii) $\lim_{\delta\to 0}\lim_n \sup \tilde{P}_n\{\tilde{\omega}: w^N(\tilde{\omega},\delta)>\eta\}=0$ for all $\eta>0$ and $N>0$.

Remark. Condition (i) is easily reformulated as follows: for every t in a dense set $T\in\mathbb{R}^+$, the family of probability measures $(P_n\circ\xi_t^{-1})_{n\in\mathcal{N}}$ on $\mathscr{H}$ is a tight family. The conditions (i) and (ii) are easily seen to be implied by the conditions for compactness in $\tilde{\Omega}$. Conversely, if (i) and (ii) hold, take a countable dense subset $(\Gamma_k)_{k\in\mathcal{N}}$ in T and find for each k a compact $C_k\subset\mathscr{H}$ and $\delta_k>0$ with $\sup_n \tilde{P}_n\{\xi_{t_k}\notin C_k\}\leqq\varepsilon/2^{k+2}$ and for some n_0

$$\sup_{n\geqq n_0}\tilde{P}_n\{w^k(\cdot,\delta_k)>1/k\}\leqq\varepsilon/2^{k+2}.$$

If we let

$$K_\varepsilon:=\tilde{\Omega}-\bigcup_{k>n_0}(\{\xi_{t_k}\notin C_k\}\cup\{w^k(\cdot,\delta_k)>1/k\})$$

the set K_ε has compact closure and $\inf_{n\geqq n_0}\tilde{P}_n(K_\varepsilon)\geqq 1-\varepsilon$.

Corollary. Let $(X^n)_{n\in\mathcal{N}}$ be a sequence of processes defined on their respective probability spaces $(\Omega^n,\mathscr{A}^n,P^n)$, with values in the complete separable metric space $\mathscr{H}$. The sequence $(\tilde{P}^n)$ of laws of the processes (X^n) form a tight sequence if and only if:

[T_1] For every t in some dense subset T of $\mathbb{R}^+$ the laws of the random variables $(X_t^n)_{n\in\mathcal{N}}$ form a tight sequence of laws in $\mathscr{H}$.

[T_2] For every $N>0$, $\eta>0$, $\varepsilon>0$, there exists $\delta>0$ such that $\lim_n\sup P_n\{\omega:\omega\in\Omega_n, w^N(X^n(\cdot,\omega),\delta)>\eta\}\leqq\varepsilon$.

2.1.4. *Particular case*: $\mathscr{H}=\mathbb{R}^d$. In this case sets of compact closure in $\mathbb{R}^d$ are bounded sets. Moreover, condition [T_1], when [T_2] holds, can be replaced by a condition on the jumps of the processes (X^n):

[T_1'] For each t the laws of the random variables $(\Delta X_t^n)_{n\in\mathcal{N}}$ where $\Delta X_t^n:=X_t^n-\lim_{s\uparrow t}X_s^n$ form a tight sequence.

2.1.5. *Particular case*: *$\mathscr{H}$ is a separable Hilbert space.* We recall that in this case a subset K in $\mathscr{H}$ has compact closure if and only if for every $\rho>0$ there exists a finite-dimensional subspace F of $\mathscr{H}$ such that $\mathrm{proj}_F(K)$ has a compact closure in F and $\sup_{h\in K} d(h,F)\leqq\rho$. If the processes $(X^n)_{n\geqq 0}$ take their values in the separable Hilbert space $\mathscr{H}$, their laws $(\tilde{P}^n)$ form a tight sequence if and only if the conditions [T_1''], [T_1'''] and [T_2] hold, where:

[T_1''] For each t in a dense subset T of $\mathbb{R}^+$ and every $h\in\mathscr{H}$ the laws of the real random variables $(h, X_t^n)_{n\in\mathcal{N}}$ form a tight sequence.

[T_1'''] For each $\varepsilon>0$, $\rho>0$, and $t\in T$ there exists a finite-dimensional subspace F of $\mathscr{H}$, such that $\lim_n\sup P^n\{d(X_t^n,F)>\rho\}\leqq\varepsilon$.

2.2. *The Aldous condition for tightness.* The notations are those of Section 2.1. We shall give in this section a sufficient condition for tightness, due to Aldous. An important part of this condition is the property which we now introduce and denote by [A].

In this section $(X^n)_{n\geqq 0}$ will denote a sequence of cadlag processes defined respectively on their own probability spaces $(\Omega^n,\mathscr{A}^n,P^n)$ and taking values in the complete separable metric space $\mathscr{H}$ (with distance d). It is assumed that on each space Ω^n a filtration $(\mathscr{F}_t^n)_{t\geqq 0}$ of sub-σ-algebras of $\mathscr{A}^n$ is given, with respect to which X^n is adapted.

2.2.1. *Definition.* We say that the sequence $(X^n)_{n\in\mathcal{N}}$ of processes satisfies the Aldous condition [A] if:

[A] For each $N>0$, $\varepsilon>0$, $\eta>0$ there exists a $\delta>0$ and n_0 with the property that whatever be the family of stopping times $(\tau_n)_{n\in\mathcal{N}}$ (τ_n being an $\mathscr{F}^n$-stopping time on Ω^n) with $\tau_n\leqq N$,

$$\sup_{n\geqq n_0}\sup_{\theta\leqq\delta}P^n\{d(X_{\tau_n}^n,X_{\tau_n+\theta}^n)\geqq\eta\}\leqq\varepsilon.$$

We say that the sequence $(X^n)_{n\in\mathcal{N}}$ satisfies the condition [A′] if:

[A′] For each $N>0$, $\varepsilon>0$, $\eta>0$ there exists $\delta>0$ and n_0 such that for any sequence $(\sigma_n,\tau_n)_{n\in\mathcal{N}}$ of pairs of stopping times (τ_n and σ_n being an $\mathscr{F}^n$-stopping time on Ω^n) with $\sigma_n\leqq\tau_n\leqq N$,

$$\sup_{n\geqq n_0}P^n\{d(X_{\sigma_n}^n,X_{\tau_n})\geqq\eta,\ \tau_n<\sigma_n+\delta\}\leqq\varepsilon.$$

2.2.2. *Theorem* (Aldous). Conditions [A] and [A′] are equivalent and imply the tightness condition [T_2].

Proof. [A′] trivially implies [A]. The fact that [A] implies [A′] is a consequence of the following lemma.

Lemma. Let X be a process with paths in $D(\mathscr{H})$. Let τ_1 and τ_2 be bounded random variables with $\tau_1 \leqq \tau_2$ such that for all $\theta \in [0, 2\delta]$,

$$P(d(X_{\tau_i}, X_{\tau_i+\theta}) \geqq \eta) \leqq \varepsilon, \quad \text{for} \quad i = 1, 2.$$

Then

$$P(d(X_{\tau_1}, X_{\tau_2}) \geqq 2\eta,\ \tau_2 < \tau_1 + \delta) \leqq 8\varepsilon.$$

Proof. Set $I = [0, 2\delta]$. Let f be any function on R_+ and $0 \leqq t_1 \leqq t_2 < t_1 + \delta$. We claim that if $d(f(t_1), f(t_2)) \geqq 2\eta$ then one of the two sets $\{\theta \in I : d(f(t_i), f(t_i + \theta)) \geqq \eta\}$ has Lebesgue measure $\geqq \delta/2$.

Indeed, if both have measure $<\delta/2$, their complements in $I: \Delta_i = \{\theta \in I : d(f(t_i), f(t_i + \theta)) < \eta\}$ each have measure $>3\delta/2$. But since the two translates $t_1 + \Delta_1$ and $t_2 + \Delta_2$ are contained in the set $[t_1, t_2 + 2\delta]$, which has measure $<3\delta$, they must intersect. Thus there exists t' with $d(f(t_1), f(t')) < \eta$ and $d(f(t_2), f(t')) < \eta$, establishing the claim. Applying this to the functions $X(\omega)$ and the times $\tau_1(\omega)$, $\tau_2(\omega)$ we have (l is the Lebesgue measure)

$$\begin{aligned} P(d(X_{\tau_1}, X_{\tau_2}) \geqq 2\eta,\ \tau_1 \leqq \tau_2 < \tau_1 + \delta) &\\ \leqq P(l\{\theta \in I : d(X(\tau_1), X(\tau_1 + \theta)) \geqq \eta\} \geqq \delta/2) &\\ + P(l\{\theta \in I : d(f(\tau_2), f(\tau_2 + \theta)) \geqq \eta\} \geqq \delta/2). \end{aligned}$$

By the Chebychev inequality and Fubini's theorem, the first term is at most

$$2/\delta \int_0^{2\delta} P\{d(f(\tau_1), f(\tau_1 + \theta)) \geqq \eta\}\, d\theta \leqq 4\varepsilon.$$

Similarly for the second term. We finish the proof of Theorem 2.2 by showing that Condition [A′] implies [T_2].

For each n, define the increasing sequence of stopping times

$$0 = \tau_0^n, \cdots, \tau_{i+1}^n = \inf\{t : t > \tau_i^n d(X_t^n, X_{\tau_i^n}^n) > \eta\}.$$

From this definition

$$[\tau_i^n < N] \subset [d(X_{\tau_i^n}, X_{\tau_{i-1}^n}) \geqq \eta],$$

and by hypothesis there exists δ such that $P_n[\tau_i^n < N,\ \tau_i^n < \tau_{i-1}^n + \delta] \leqq \varepsilon/2$, for all n.

Choose an integer q such that $q\delta \geqq 2N$. Then $P_n(\tau_q^n < N) \leqq \varepsilon$, for all n.

Indeed,

$$\begin{aligned} NP(\tau_q < N) &\geqq E[\tau_q; \tau_q < N] = \sum_{i=1}^{q} E[\tau_i - \tau_{i-1}; \tau_q < N] \\ &\geqq \sum_{i=1}^{q} \delta P(\tau_q < N, (\tau_i - \tau_{i-1}) \geqq \delta) \\ &\geqq \sum_{i=1}^{q} \delta (P(\tau_q < N) - P(\tau_q < N, \tau_i - \tau_{i-1} < \delta) \\ &\geqq q\delta[P(\tau_q < N) - \varepsilon/2]. \end{aligned}$$

From $q\delta \geqq 2N$ this yields

$$P_n(\tau_q^n < N) \leqq \varepsilon \text{ for all } n.$$

Now using this hypothesis again we can find a $\sigma > 0$ such that

$$P_n(\tau_q^n < N,\ \tau_i^n < \tau_{i-1}^n + \sigma) \leqq \varepsilon/q \text{ for all } n$$

so that

$$P_n\Big(\bigcup_{i=1}^{q} [\tau_i^n < N,\ \tau_i^n < \tau_{i-1}^n + \sigma]\Big) \leqq \varepsilon \text{ for all } n.$$

But on the set $[\tau_q^n = N] \backslash \bigcup_{i=1}^q [\tau_i^n < N,\ \tau_i^n < \tau_{i-1}^n + \sigma]$, $w^N(\delta)(X^n) \leqq \eta$ so that $P^n(w^N(\sigma)(X^n) \geqq \eta) \leqq 2\varepsilon$ for all n.

2.2.3. *Remark.* In Aldous (1978) the following trivial example is given which proves that [A] is far from being a necessary condition for tightness. Take X_n as being the deterministic process $X_n(t) = 0$ for $0 \leqq t \leqq \frac{1}{2}$, $X_n(t) = 1$ for $\frac{1}{2} \leqq t \leqq 1$.

The reader will notice that the same reasoning can be applied if $X_n := 1_{[\tau, \infty[}$ where τ is a predictable stopping time, and more generally if X_n converges in law toward a process X which has a discontinuity on an accessible stopping time. In other words the *Aldous condition implies the quasi-continuity of the limit X.* (No jump at accessible stopping times.)

2.2.4. *Remark.* One may wonder whether the 'strong uniform right equicontinuity in probability' expressed by condition [A] for the processes is strictly stronger than the mere 'uniform right equicontinuity in probability' which would read in the case $H = R$:

[ECP] For every $N > 0$, $\varepsilon > 0$, $\eta > 0$ there exist $\delta > 0$ and n_0 such that

$$\sup_{n \geqq n_0} \sup_{\theta \in [0, \delta]} \sup_{t \in [0, N]} P_n\{|X^n(t + \theta) - X^n(t)| \geqq \eta\} \leqq \varepsilon.$$

The following example with $X^n := X$ shows that it is not the case. Take

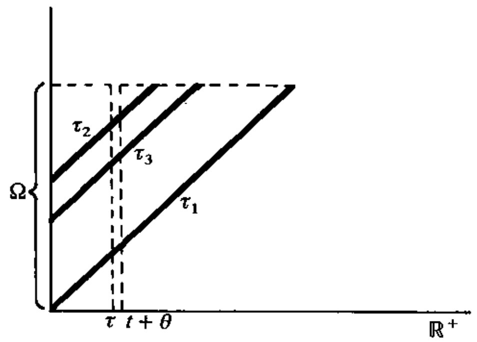

Figure 1.

$\Omega = [0, 1]$ with P: Lebesgue measure. Take $\tau_1(\omega) = \omega$, $\tau_k(\omega) = (\omega - 1/k)^+$ if $k \geqq 2$. If X is defined by

$$X(t, \omega) = \sum_{k \geqq 1} \frac{1}{2^k} 1_{[\tau_k, \infty[}^{(t,\omega)},$$

clearly the τ_k's are stopping times of the process X.

For every t and $\theta \leqq 1/2^r$,

$$P\left\{ X(t+\theta) - X(t) \Big| > \frac{1}{2^{r-1}} \right\} < r\theta.$$

This shows that [ECP] is fulfilled.

Conversely, whatever the value of θ, as soon as $1/k < \delta$ we have $P\{|X(\tau_k + \delta) - X(\tau_k)| \geqq \frac{1}{2}\} = 1$, and property [A] does not hold for (X^n).

2.2.5. *Remark.* A natural question, too, is to ask whether the weaker assumption [ECP] is necessary for weak convergence, while [A] is not, as shown in Remark 2.2.3. The following example given to us by S. Orey shows that this is not the case.

For each n, Ω_n is $[0, 1]$ with the Lebesgue measure, while $X_n(\omega) = f(n(t - \omega)^+)$, where f is the triangular function

$$f(x) = \begin{cases} |1 - x|, & x < 1 \\ 0, & x \geqq 1. \end{cases}$$

Since for every $\eta > 0$, n_0, $n \geqq n_0$, t and $\theta \leqq 1/n_0$,

$$P\{|X^n(t+\theta) - X^n(t)| \geqq \eta\} \leqq 2/n_0,$$
$$P\{w^1(X^n(\omega), \delta) \geqq 1\} = 1 \qquad \text{as soon as } n > 1/\delta.$$

The necessary condition $[T_2]$ for tightness is therefore not fulfilled.

2.3. *Tightness of sequences of laws of Hilbert-valued semimartingales.* In this section, H is a separable Hilbert space. We now give a sufficient condition for tightness for the laws (P_n) of a sequence (X^n) of semimartingales due to Rebolledo (1979).

2.3.1. *Lemma* (Lenglart (1977)). Let X be a cadlag adapted positive process on a stochastic basis $(\Omega, \mathscr{A}, (\mathscr{F}_t)_{t \in \mathbb{R}^+}, P)$ and Y a real adapted increasing process on the same basis such that for every finite stopping time $\tau: E(X_\tau) \leqq E(Y_\tau)$. Then

(i) For every finite stopping time τ and $\varepsilon > 0$,

(2.3.1) $$P\left\{ \sup_{s \leqq \tau} X_s > \varepsilon \right\} \leqq 1/\varepsilon E(Y_\tau).$$

(ii) If moreover Y is predictable, for every stopping time τ, every ε and $n > 0$:

(2.3.2) $$P\left\{ \sup_{s \leqq \tau} X_s > \varepsilon \right\} \leqq 1/\varepsilon E(Y_\tau \wedge \eta) + P\{Y_\tau \geqq \eta\}.$$

Proof. (i) Set $S := \inf\{t: X_t > \varepsilon\} \wedge \tau$. Since clearly $\{\sup_{s \leqq t} X_s > \varepsilon\} \subset \{X_S \geqq \varepsilon\}$, we have

$$\varepsilon P\left\{ \sup_{s \leqq \tau} X_s > \varepsilon \right\} \leqq \varepsilon P\{X_S \geqq \varepsilon\} \leqq E(X_S) \leqq E(Y_S) \leqq E(Y_\tau).$$

(ii) We may write

(2.3.3) $$P\left\{ \sup_{s \leqq t} X_s > \varepsilon \right\} \leqq P\left\{ \sup_{s \leqq \tau} X_s > \varepsilon,\ Y_\tau < \eta \right\} + P\{Y_\tau \geqq \eta\}.$$

Define

$$S := \inf\{t: t > 0,\ Y_t \geqq \eta\}, \text{ and let } [S] \text{ denote the graph of } S.$$

If Y is predictable, S is predictable, being the beginning of the predictable set $\{(t, \infty): Y_t(\omega) \geqq \eta\}$ and since $[S] \subset \{(t, \omega): Y_t(\omega) \geqq \eta\}$. Let S_n be an announcing sequence for S: i.e. an increasing sequence (S_n) of finite stopping times such that $S_n < S$ a.s. and $\lim_n S_n = S$. Since $\{Y_\tau < \eta\} \subset \{\tau < S\}$ we may write

$$\left\{ \sup_{s \leqq \tau} X_s > \varepsilon,\ Y_\tau < \eta \right\} \subset \bigcup_n \left\{ \sup_{t \leqq \tau \wedge S_n} X_t \geqq \varepsilon \right\}$$

and from the properties of X and Y

$$P\left(\left\{ \sup_{s \leqq \tau} X_s > \varepsilon,\ Y < \eta \right\} \right) \leqq \lim_n 1/\varepsilon E(Y_{\tau \wedge S_n}) \leqq 1/\varepsilon E(Y_\tau \wedge \eta).$$

This inequality, with (2.3.3) proves (2.3.2).

2.3.2. *Theorem* (Rebolledo). Let (M^n) be a sequence of H-valued processes, which are right-continuous locally square-integrable martingales, defined on their own probability space $(\Omega^n, (\mathcal{F}_t^n), P^n)$. Let $\langle M^n \rangle$ $([M^n])$ be the associated Meyer increasing process (quadratic variation). (See Sections 1.2.2 and 1.2.3.) Then if the processes $(\langle M^n \rangle)_{n\in\mathcal{N}}$ satisfy the condition [A], the same condition holds for the sequence $(M^n)_{n\in\mathcal{N}}$ and $([M^n])_{n\in\mathcal{N}}$. If H is finite-dimensional and if the processes $(\langle M^n \rangle)_{n\in\mathcal{N}}$ satisfy condition $[T_1]$, then the same condition holds for $(M^n)_{n\in\mathcal{N}}$ and $([M^n])_{n\in\mathcal{N}}$.

Proof. We first remark that if M is a locally square-integrable martingale, then so is

$$L_t^T = M_t - M_{T\wedge t}$$

for every stopping time T. Moreover

$$\langle L^T \rangle_t = \langle M \rangle_t - \langle M \rangle_{T\wedge t}, \; [L^T]_t = [M]_t - [M]_{T\wedge t}.$$

If (U_k) is an increasing sequence of stopping times such that $U_k \uparrow \infty$ and $(L^T_{t\wedge U_k})$ is for each k a square-integrable martingale, we have the following well-known equalities for every finite stopping time S:

$$E(\langle L^T \rangle_{S\wedge U_k}) = E[L^T]_{S\wedge U_k}$$

$$= E(\|L^T_{S\wedge U_k}\|^2) \leqq E\Big(\sup_{0\leqq s\leqq S\wedge U_k} \|L^T_s\|^2\Big) \leqq 4E(\langle L^T \rangle_{S\wedge U_k}).$$

Therefore, for every finite stopping time,

$$E(\|L^T_S\|^2) \leqq E(\langle L^T \rangle_S) = E([L^T]_S) \leqq E\Big(\sup_{0\leqq s\leqq S} \|L^T_s\|^2\Big) \leqq 4E(\langle L^T \rangle_S).$$

We can therefore in particular apply Lemma 3.1 to $X = \|L^T\|^2$ and $Y = \langle L^T \rangle$.
We get for every a and b and every $\delta > 0$,

$$P\Big\{\sup_{T\leqq s\leqq T+\delta} \|M_s - M_T\| \geqq b\Big\}$$

$$\leqq 1/b^2 E[(\langle M \rangle_{T+\delta} - \langle M \rangle_T) \wedge a] + P\{(\langle M \rangle_{T+\delta} - \langle M \rangle_T) \geqq a\}.$$

Choose $a \leqq b^2\varepsilon/2$. Then for every δ such that

$$P\{|\langle M \rangle_{T+\delta} - \langle M_T \rangle| \geqq a\} \leqq \varepsilon/2$$

we obtain

$$P\Big\{\sup_{T\leqq s\leqq T+\delta} \|M_s - M_T\| \geqq b\Big\} \leqq \varepsilon.$$

It is therefore clear that, if condition [A] holds for $\langle M^n \rangle$, it holds for M^n. The same reasoning shows that it holds for $[M]$ too. The inequalities

$$P\Big\{\sup_{0\leqq s<N} \|M_s\| \geqq b\Big\} \leqq 1/b^2 E(\langle M_N \rangle \wedge a) + P\{\langle M \rangle_N \geqq a\}$$

$$P\{[M]_N \geqq b\} \leqq 1/b E(\langle M_N \rangle \wedge a) + P\{\langle M \rangle_N \geqq a\}$$

show in the same way that condition $[T_1]$ of Theorem 1 holds for M^n and $[M^n]$ as soon as it holds for $\langle M^n \rangle$. This proves Theorem 3.2.

2.3.3. *Corollary.* If $(X^n)_{n\geqq 0}$ is a sequence of finite-dimensional semimartingales of the form $X^n = A^n + M^n$ with A^n a process with finite variation and M^n a locally square-integrable martingale and if the sequences (A^n) and $(\langle M^n \rangle)$ satisfy the conditions $[T_1]$ and [A] of the Aldous theorem, the probability laws $(\tilde{P}^n)$ of the sequence (X^n) form a tight family. (We assume $X_0^n = A_0^n$.)

2.3.4. *Remark.* Here is an example to show that the tightness of $(\langle M^n \rangle)_{n>0}$ is not enough to imply the tightness of $(M_n)_{n>0}$.

Take $\Omega^n = R^+ \times \{-1, +1\}$, and let τ^n be the projection of Ω on R^+ and u^n the projection on $\{-1, +1\}$. The probability law P^n on Ω^n is such that the random variables τ^n and u^n are independent, with $P^n\{\tau^n > t\} = \exp(-t)$ and $P^n\{u^n = +1\} = P^n\{u^n = -1\} = \frac{1}{2}$. We set $X_t^n := 1_{\{\tau^n\leqq t\}} + u^n 1_{\{\tau^n+(1/n)\leqq t\}}$. The filtration $(\mathcal{F}_t^n)_{t\geqq 0}$ is the right-continuous one generated by X^n. We observe that $\mathcal{F}^n_{\tau^n}$ is the σ-algebra generated by τ^n and that $E^n(u^n \mid \mathcal{F}_{(\tau^n+1/n)-}) = E^n(u^n \mid \mathcal{F}^n_{\tau^n}) = 0$. Let us note also that τ^n is a totally inaccessible jump, while $\tau^n + (1/n)$ is predictable. Let us set $M_t^n := 1_{\{\tau^n\leqq t\}} - \tau^n \wedge t + u^n 1_{\{\tau^n+(1/n)\leqq t\}}$. It is easily seen that $\langle M_t^n \rangle = \tau^n \wedge t + 1_{\{\tau^n+(1/n)\wedge t\}}$. The sequence $\langle M_t^n \rangle$ converges weakly towards the law of the increasing process $A_t := \tau \wedge t + 1_{\{\tau\leqq t\}}$ where τ is an exponentially distributed random variable, while the laws of (M^n) do not form a tight sequence (because of the two jumps of magnitude 1 which come together with probability 1 at time τ^n).

2.4. *Extensions of the Aldous and Rebolledo results.* In order to obtain conditions closer to necessity than the previous Aldous and Rebolledo hypotheses (see Remark 2.2.3) and in particular to obtain tightness conditions where the limits are not necessarily quasi-continuous, extensions of the theorems described in Sections 2.2 and 2.3 have been worked out, in particular by Jacod *et al.* (1983) and also by Platen and Rebolledo (1981) and Rebolledo (1979).

For the sake of completeness we mention a few results from Jacod *et al.* (1983), but we shall not use these results later in our applications. We say that an increasing process G^n dominates the semimartingale $X^n = A^n + M^n$ if for P^n almost all ω, the measure $dG^n(\cdot, \omega)$ is greater than $d(|A^n| + \langle M \rangle)(\omega)$ where $|A^n|$ is the variation of A^n.

2.4.1. *Theorem* (Jacod *et al.* (1983)). Each one of the following conditions is sufficient for the tightness of the sequence $\tilde{P}^n_{Z^n}$, where G^n dominates the semimartingale Z^n.

(C_1) The sequence $(\tilde{P}^n_{G^n})$ converges weakly towards a $\tilde{P}$ such that $\tilde{P}\{\tilde{\omega}:\tilde{\omega}$ continuous$\}=1$.
(C_2) $(\tilde{P}^n_{G^n})$ converges weakly towards a Dirac probability measure.
(C_3) $(\Omega^n, (\mathcal{F}^n_t), P^n)=(\Omega, (\mathcal{F}^n_t), P)$
  (i) G^n converges in probability to G as $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$-valued random variables.
  (ii) G is predictable with respect to the filtration $\bigcap_n \mathcal{F}^n_t$.
(C_4) $\tilde{P}^n_{G^n}$ converges weakly to a limit $\tilde{P}$ such that the canonical process ξ is $\mathcal{D}(\tilde{P})$-predictable where $\mathcal{D}(\tilde{P})$ denotes the natural filtration of $\mathbb{D}(\mathbb{R}^+; \mathcal{H})$ completed for $\tilde{P}$.

Another type of extension concerns the infinite-dimensional case. The last statement in Theorem 2.3.2 applies only to the finite-dimensional case. The following result may be found in Métivier (1984).

Recall that $\langle\langle M\rangle\rangle$ is the $\mathcal{H}\hat{\otimes}_1\mathcal{H}$-valued process (matrix-valued process) such that $M\otimes M-\langle\langle M\rangle\rangle$ is a martingale, with $\langle\langle M\rangle\rangle$ predictable and with finite variation. (See Métivier (1982), Chapter 4, for details.) Let us recall that if (h_i) is an orthonormal basis of $\mathcal{H}$, the elements of $\mathcal{H}\hat{\otimes}_1\mathcal{H}$ are of the form $\hat{y}:=\sum_{i,j\in\mathcal{N}}\lambda_{ij}h_i\otimes h_j$ with $\|\hat{y}\|_1=\sum|\lambda_{ij}|<\infty$, and $\mathcal{H}\hat{\otimes}_1\mathcal{H}$ is a Banach space for the norm $\|\hat{y}\|_1$. This space is included in the Hilbert–Schmidt tensor product

$$\mathcal{H}\hat{\otimes}_2\mathcal{H}:=\Big\{\hat{y}:\hat{y}=\sum_{i,j}\lambda_{i,j}h_i\otimes h_j, \sum_{ij}|\lambda_{ij}|^2<\infty\Big\}.$$

The space $\mathcal{H}\hat{\otimes}_2\mathcal{H}$ is a Hilbert space with $h_i\otimes h_j$ as an orthonormal basis and norm $\|\hat{y}\|_2=(\sum_{ij}|\lambda_{ij}|^2)^{\frac{1}{2}}$. The injection from $\mathcal{H}\hat{\otimes}_1\mathcal{H}$ into $\mathcal{H}\hat{\otimes}_2\mathcal{H}$ is continuous.

2.4.2. *Theorem* (Infinite-dimensional semimartingales). Let (M^n) be a sequence of locally square-integrable martingales, with values in the separable Hilbert space $\mathcal{H}$. If the sequence of $\mathcal{H}\hat{\otimes}_2\mathcal{H}$-valued processes $(\langle\langle M^n\rangle\rangle^{\frac{1}{2}})_{n\in\mathcal{N}}$ satisfies the condition $[T_1]$, the same condition holds for the processes $(M^n)_{n\in\mathcal{N}}$.

3. Weak convergence of $\mathcal{D}$-semimartingales

3.1. *$\mathcal{D}$-semimartingales in* $\mathbb{R}^d$. In this section we introduce the class of processes for which one can write a formula similar to the Dynkin formula for Markov processes. As a consequence of this formula, they turn out to be particular semimartingales. For this reason we call them Dynkin semimartingales or, for short, $\mathcal{D}$-semimartingales.

3.1.1. *Definition.* A cadlag adapted process $\boldsymbol{X}$, defined on the stochastic basis $(\Omega, \mathcal{A}, (\mathcal{F}_t)_{t\geq 0}, P)$, with values in $\mathbb{R}^d$, will be called a $\mathcal{D}$-semimartingale if there exists an increasing cadlag function $A(t)$, a vector space $\mathcal{C}$ of continuous functions on $\mathbb{R}^d$ and a mapping $L:(\mathcal{C}\times\mathbb{R}^d\times\mathbb{R}^+\times\Omega)\to\mathbb{R}$ with the following properties:

(D_1) The functions $\boldsymbol{x}\to\phi^i(\boldsymbol{x}):=x^i$ and $\phi^i\phi^j$, $i, j=1,\cdots,d$ belong to $\mathcal{C}$.
(D_2) (i) For every $(\boldsymbol{x}, t, \omega)\in\mathbb{R}^d\times\mathbb{R}^+\times\Omega$ the mapping $\phi\to L(\phi, \boldsymbol{x}, t, \omega)$ is a linear form on $\mathcal{C}$ and $L(\phi, \cdot, t, \omega)\in\mathcal{C}$.
  (ii) For every $\phi\in\mathcal{C}$, $(\boldsymbol{x}, t, \omega)\to L(\phi, \boldsymbol{x}, t, \omega)$ is $\mathcal{B}_{\mathbb{R}^d}\otimes\mathcal{P}$-measurable for the Borel σ-algebra $\mathcal{B}_{\mathbb{R}^d}$ of $\mathbb{R}^d$ and the σ-algebra $\mathcal{P}$ of predictable sets.
(D_3) For every $\phi\in\mathcal{C}$ the process $\boldsymbol{M}^\phi$ defined by $\boldsymbol{M}^\phi(t, \omega):=\phi(\boldsymbol{x}_t(\omega))-\phi(\boldsymbol{x}_0(\omega))-\int_0^t L(\phi, \boldsymbol{x}_{s^-}(\omega), s, \omega)\,dA_s$ is a locally square-integrable martingale on $(\Omega, (\mathcal{F}_t)_{t\geq 0}, P)$.

Now, we introduce the first- and second-order 'local coefficients' b_i and $a_{i,j}$.

3.1.2. *Definition.* For every $i, j=1,\cdots,d$ we set

$$b_i(\boldsymbol{x}, t, \omega):=L(\phi^i, \boldsymbol{x}, \omega, t)$$

$$a_{i,j}(\boldsymbol{x}, t, \omega):=L(\phi^i\phi^j, \boldsymbol{x}, \omega, t)-x^ib_j(\boldsymbol{x}, \omega, t)-x^jb_i(\boldsymbol{x}, \omega, t).$$

The b_i's and $a_{i,j}$'s will be called the *local coefficients of first and second order.* We write $\boldsymbol{b}(\boldsymbol{x}, t, \omega)$ for the vector $(b_i(\boldsymbol{x}, t, \omega))_{i=1,\cdots,d}$ and $\boldsymbol{a}(\boldsymbol{x}, t, \omega)$ for the matrix $(a_{i,j}(\boldsymbol{x}, t, \omega))_{i,j=1,\cdots,d}$.

3.1.3. *Lemma.* Let $\boldsymbol{X}$ be a $\mathcal{D}$-semimartingale with local coefficients $\boldsymbol{b}$ and $\boldsymbol{a}$. Let us define

$$\boldsymbol{M}_t:=\boldsymbol{X}_t-\boldsymbol{X}_0-\int_0^t \boldsymbol{b}(\boldsymbol{X}_{s^-}, s, \cdot)\,dA_s.$$

Then $(\boldsymbol{M}_t)_{t\geq 0}$ is a locally square-integrable martingale and

$$\langle\boldsymbol{M}\rangle_t=\int_{]0,t]}\operatorname{trace}\boldsymbol{a}(\boldsymbol{X}_{s^-}, s, \cdot)\,dA_s-\sum_{s\leq t}\|\boldsymbol{b}(\boldsymbol{X}_{s^-}, s, \cdot)\|^2\,|\Delta A_s|^2.$$

Proof. The fact that $(\boldsymbol{M}_t)$ is a locally square-integrable martingale follows immediately from the definition of b and (D_1) and (D_3).

Now the formula on quadratic variation gives

$$\|\boldsymbol{X}_t\|^2=\|\boldsymbol{X}_0\|^2+2\int_{]0,t]}\boldsymbol{X}_{s^-}\,d\boldsymbol{X}_s+[\boldsymbol{X}]_t. \quad (3.1.1)$$

But, writing $\boldsymbol{V}_t$ for the process $\int_{]0,t[}\boldsymbol{b}(\boldsymbol{X}_{s^-}, s, \cdot)\,dA_s$ (with paths of finite variation) we obtain

$$[X]_t = [M]_t + 2[M, V]_t + [V]_t = [M]_t + 2\sum_{s\leqq t} \Delta M_s \Delta V_s + \sum_{s\leqq t} \|\Delta V_s\|^2$$

and hence

$$(3.1.2) \qquad [X]_t = [M]_t + 2\sum_{s\leqq t} (b(X_{s^-}, s, \cdot)\Delta M_s)\Delta A_s + \sum_{s\leqq t} \|b(X_{s^-}, s, \cdot)\|^2(\Delta A_s)^2.$$

Therefore

$$(3.1.3) \qquad \begin{aligned} \|X_t\|^2 = \|X_0\|^2 + 2\int_{]0,t[} (X_s \cdot b(X_{s^-}, s, \cdot))\, dA_s + [N]_t \\ + \sum_{s\leqq t} \|b(X_{s^-}, s, \cdot)\|^2 |\Delta A_s|^2 + N_t \end{aligned}$$

where (N_t) is a locally square-integrable martingale. Now using property (D_3) with $\phi(x) = \|x\|^2 = \text{trace}\,(x^t x)$, we obtain

$$(3.1.4) \qquad \begin{aligned} \|X_t\|^2 = \|X_0\|^2 + 2\int_{]0,t]} X_s \cdot b(X_{s^-}, s, \cdot)\, dA_s \\ + \int_{]0,t]} \text{trace}\, a(X_{s^-}, s, \cdot)\, dA_s + Y_t \end{aligned}$$

where (Y_t) is a locally square-integrable martingale. Comparing (3.1.2) and (3.1.3) we see that

$$\left([M_t] - \int_{]0,t]} \text{trace}\, a(X_{s^-}, s, \cdot)\, dA_s + \sum_{s\leqq t} \|b(X_{s^-})\|^2 |\Delta A_s|^2\right)_{t\geqq 0}$$

is a locally square-integrable martingale. Formula (1.2.9) then gives the lemma.

We shall make use of the following version of the classical Gronwall inequality.

3.1.4. *Lemma.* Let A be a cadlag increasing function on $[0, T]$ with $A(0) = 0$, $A(T) \leqq \ell$. If Φ is a positive left-continuous function such that for all $t \leqq T$

$$\Phi(t) \leqq K + \rho \int_{]0,t]} \Phi(s)\, dA_s$$

then

$$\Phi(T) \leqq 2K \sum_{j=0}^{[2\rho\ell]} (2\rho\ell)^j$$

where $[x]$ denotes the integer part of x.

Proof. Define $\sigma_0 = 0 \cdots \sigma_{k+1} := \inf\{t : A_t - A_{\sigma_k} \geqq 1/2\rho\} \wedge T$ and set $x_k := \Phi_{\sigma_k}$. Note that, because of the left continuity of Φ

$$\Phi(t) \leqq K + \rho \int_{]0,t[} \Phi(s)\, dA_s \text{ for all } t \leqq T$$

and therefore

$$x_{k+1} \leqq K + \rho \int_{]0,\sigma_k]} \Phi(s)\, dA(s) + \rho \int_{]\sigma_k, \sigma_{k+1}[} \Phi(s)\, dA(s) \leqq K + \rho\ell x_k + \tfrac{1}{2} x_{k+1}$$

or

$$x_{k+1} \leqq 2K + 2\rho\ell x_k.$$

Noticing that $T \leqq \sigma_{[2\rho\ell]}$ leads immediately to the inequality of the lemma.

3.2. *Tightness of sequences of $\mathscr{D}$-semimartingales*

3.2.1. *Hypotheses.* We consider a sequence (X^n) of $\mathscr{D}$-semimartingales, each X^n being defined on its own probability space $(\Omega^n, \mathscr{A}^n, (\mathscr{F}_t^n)_{t\geqq 0}, P^n)$. We assume that these $\mathscr{D}$-semimartingales are associated with the same space $\mathscr{C}$ of continuous functions on $\mathbb{R}^d$ and we call L^n (or A^n) the mapping $(\phi, x, t, \omega) \to L^n(\phi, x, t, \omega)$ (or the increasing function) attached to X^n (see Definition 3.1.1).

For easy further reference we now list some hypotheses which will be used later.

(H_1) There exists a constant K and a sequence of positive adapted processes $(C_t^n)_{t>0}$ such that for every n, x and ω
 (i) $(\|b^n(x, \omega, t)\|^2 + \text{trace}\, a^n(x, \omega, t)) \leqq K(C_t^n + \|x\|^2)$.
 (ii) For every $T > 0$

$$\sup_n \sup_{t\leqq T} E(C_t^n) < \infty \quad \text{and} \quad \lim_{\rho\to\infty} \sup_n P^n\left\{\sup_{t\leqq T} C_t^n > \rho\right\} = 0.$$

(H_2) The sequence $(X_0^n)_{n\geqq 0}$ of random variables is such that

$$\sup_n E(\|X_0^n\|^2) < \infty.$$

(H_3) There exists a positive function α on $\mathbb{R}^+$ and a decreasing sequence of numbers (ρ_n) such that $\lim_{t\downarrow 0} \alpha(t) = 0$, $\lim_{n\uparrow\infty} \rho_n = 0$ and for all $0 < s < t$ and all n:

$$(A^n(t) - A^n(s)) \leqq \alpha(t - s) + \rho_n.$$

In the practical applications which will be mentioned in subsequent sections the sequence (A^n) will be either $A^n(t) = t$ for all n and t or $A^n(t) = (1/n)[nt]$

where $[u]$ denotes the integral part of u. Let us remark also that condition (H_3) implies that A^n has all its jumps smaller than ρ_n.

3.2.2. *Lemma*. Let $(X^n)_{n\geqq 0}$ be a sequence of $\mathcal{D}$-semimartingales satisfying the assumptions (H_1), (H_2) and (H_3). We set

$$M_t^n := X_t^n - X_0^n - \int_{]0,t]} b^n(X_{s-}^n, s, \cdot)\, dA_s^n. \tag{3.2.1}$$

Then for each $T > 0$ there exists a constant K_T and n_0 such that for all $n \geqq n_0$

$$E\Big(\sup_{t\leqq T} \|X_t^n\|^2\Big) \leqq K_T(1 + E\,\|X_0^n\|^2). \tag{3.2.2}$$

and

$$E\Big(\sup_{t\leqq T} \|M_t^n\|^2\Big) \leqq K_T(1 + E\,\|X_0^n\|^2). \tag{3.2.3}$$

Proof. Let $(\tau_k^n)_{k>0}$ be a sequence of stopping times converging monotonically to $+\infty$, such that $(M_{\tau_k^n\wedge t}^n)_{t\geqq 0}$ is a square-integrable martingale for each k. Writing X_t^n as $M_t^n + X_0^n + \int_{]0,t[} b^n(X_{s-}^n, s, \cdot)\, dA_s^n$ we derive

$$E\Big(\sup_{t\leqq T} \|X_{\tau_k^n\wedge t}^n\|^2\Big) \leqq 3E(\|X_0^n\|^2) + 3E(\alpha(T)+\rho_n)\Big(\int_{]0,T\wedge\tau_k^n]} \|b^n(X_{s-}, s)\|^2\, dA_s^n\Big)$$
$$+ 3E\Big(\sup_{t\leqq T} \|M_{\tau_k^n\wedge t}^n\|^2\Big).$$

Using the Doob inequality,

$$E\Big(\sup_{t\leqq T} \|M_{\tau_k^n\wedge t}\|^2\Big) \leqq 4E(\langle M\rangle_{\tau_k^n}). \tag{3.2.4}$$

Using Lemma 3.1.3 we derive, for some constant $\bar{K}$,

$$E\Big(\sup_{t\leqq T} \|X_{\tau_k^n\wedge t}^n\|^2\Big) \leqq 3E(\|X_0^n\|^2) + \bar{K}\Big[\int_{]0,T]} (C_s^n + \|X_{s-\wedge\tau_k^n}^n\|^2)\, dA_s^n\Big].$$

Setting $\gamma_T := \sup_{t\leqq T} E(C_t^n)$ and using (H_3) one obtains for every $t < T$

$$E\Big(\sup_{s\leqq t} \|X_{\tau_k^n\wedge s}^n\|^2\Big)$$
$$\leqq 3E(\|X_0^n\|^2) + 12K(\alpha(T)+\rho_0)\gamma_T + \bar{K}\int_{]0,t]} \Big(\sup_{u<s} \|X_{u\wedge\tau_k^n}^n\|^2\, dA_s^n\Big).$$

Applying Lemma 3.1.4 to $\Phi(t) := E(\sup_{s<t} \|X_{\tau_k^n\wedge s}^n\|^2)$ we obtain immediately

the existence of a constant K_T' such that

$$E\Big(\sup_{s<T} \|X_{\tau_k^n\wedge s}^n\|^2\Big) \leqq K_T'(1 + E\,\|X_0^n\|^2).$$

The Fatou lemma for $k \uparrow \infty$ leads immediately to

$$E\Big(\sup_{t<T} \|X_t^n\|^2\Big) \leqq K_T'(1 + E\,\|X_0^n\|^2)$$

and therefore to formula (3.2.2). Formula (3.2.4) and Lemma 3.1.3 imply for $n \geqq n_0$

$$E\Big(\sup_{t\leqq T} \|M_t^n\|^2\Big) \leqq 4\int_{[0,T]} E(\operatorname{trace} a^n(X_{s-}^n, s, \cdot))\, dA_s^n$$
$$\leqq 4K\int_{[0,T]} E(C_s^n + \|X_s^n\|^2)\, dA_s^n$$
$$\leqq 4K(T + \rho_{n_0})\Big(\gamma_T + \sup_{t\leqq T} E\,\|X_t^n\|^2\Big).$$

This gives formula (3.3.3).

3.2.3. *Proposition*. Every sequence (X^n) of $\mathcal{D}$-semimartingales satisfying the hypotheses (H_1), (H_2) and (H_3) is tight.

Proof. The inequality (3.2.2) implies the tightness of the laws of the $\mathbb{R}^d$-valued random variables $\{X_t^n : n \in \mathcal{N}\}$ for every t.

We have only to prove, using the theorem of Rebolledo, that the processes $(B^n)_{n\in\mathcal{N}}$ and $(\langle M^n\rangle)_{n\in\mathcal{N}}$ satisfy the Aldous condition [A], where

$$B_t^n := \int_{]0,T]} b^n(X_{s-}^n, s)\, dA_s^n.$$

From (3.2.2) and (H_1) we derive

$$P^n\Big\{\sup_{s\leqq T} \|b^n(X_{s-}^n, s, \cdot)\| > \frac{\eta}{\alpha(\delta)}\Big\} \leqq \frac{\alpha(\delta)^2}{\eta^2}\bar{K}_T(1 + E\,\|X_0^n\|^2) \tag{3.2.5}$$

and

$$P^n\Big\{\sup_{s\leqq T} \operatorname{trace} a^n(X_{s-}^n, s, \cdot) > \frac{\eta}{\alpha(\delta)}\Big\} \leqq \frac{\alpha(\delta)}{\eta}\bar{K}_T(1 + E\,\|X_0^n\|^2) \tag{3.2.6}$$

for some $\bar{K}_T$ and all $n \in \mathcal{N}$. Since $A_{\tau^n+\delta}^n - A_{\tau^n}^n \leqq \alpha(\delta) + \rho_n$ the following inequalities follow from (3.2.5) for any family (τ_n) of stopping times such that $\tau_n \leqq T$:

$$\sup_n P^n\Big\{\|B_{\tau_n+\delta}^n - B_{\tau_n}^n\| > \eta\Big(1 + \frac{\rho_n}{\alpha(\delta)}\Big)\Big\} \leqq \frac{\alpha(\delta)^2}{\eta^2}\bar{K}_T\Big(1 + \sup_n E\,\|X_0^n\|^2\Big) \tag{3.2.7}$$

and

$$\sup_n P^n\left\{\int_{]\tau_n,\tau_n+\delta]} \|b^n(X^n_{s-}, s, \cdot)\|^2\, dA^n_s > \eta\left(1+\frac{\rho_n}{\alpha(\delta)}\right)\right\}$$
(3.2.8)
$$\leqq \frac{\alpha(\delta)}{\eta}\tilde{K}_T\left(1+\sup_n E\,\|X^n_0\|^2\right).$$

From the inequality (3.2.6) we derive

$$\sup_n P^n\left\{\int_{]\tau_n,\tau_n+\delta]} \operatorname{trace} a^n(X^n_{s-}, s, \cdot)\, dA^n_s > \eta\left(1+\frac{\rho_n}{\alpha(\delta)}\right)\right\}$$
(3.2.9)
$$\leqq \frac{\alpha(\delta)}{\eta}\tilde{K}_T\left(1+\sup_n E\,\|X^n_0\|^2\right).$$

The Aldous condition for (B^n) follows readily from (3.2.7) while the Aldous condition for $\langle M^n\rangle$ follows from (3.2.8), (3.2.9) and from the fact that as soon as $\rho_n \leqq 1$

$$\sum_{\tau_n<s\leqq\tau_n+\delta} \|b^n(X^n_{s-}, s, \cdot)\|^2\, \|\Delta A^n_s\|^2 \leqq \int_{]\tau_n,\tau_n+\delta]} \|b^n(X^n_{s-}, s, \cdot)\|^2\, dA^n_s.$$

3.3. *Convergence of sequences of $\mathscr{D}$-semimartingales.* Once the tightness of a sequence (X^n) of $\mathscr{D}$-semimartingales has been derived, the study of the convergence in many cases goes as follows. Assume that there exists a subset of bounded functions $\mathscr{C}_0$ of $\mathscr{C}$ and on $\mathscr{C}_0$ a linear operator L, mapping $\mathscr{C}_0$ into bounded continuous functions on $\mathbb{R}^d$. Let us also assume that we are able to prove (this is sometimes one of the most difficult steps!) that for every t

(H₄) $$\lim_{n\to\infty}\int_0^t E^n\,|L^n(\phi, X^n_{s-}, s, \cdot) - L\phi(X^n_{s-})|\,dA^n_s = 0.$$

Let us finally make the hypotheses

(H_5) The measure dA^n converges weakly to the Lebesgue measure and strengthen (H_2) to

(H_2') The laws of $(X^n_0)_{n\geqq 0}$ converge weakly to a probability law μ_0 and $\sup_n E\,|X^n_0|^2 < \infty$.

Then the limits of the laws $(\tilde{P}^n)$ of the sequence (X^n) can be characterized as solutions of a 'martingale problem' associated with (L, C_0, μ_0). Let us recall the definition. On the space $\tilde{\Omega} := \mathbb{D}(\mathbb{R}^+; \mathbb{R}^d)$ (see Section 1.1.4) we consider the right-continuous filtration $(\tilde{\mathscr{F}}_t)_{t\geqq 0}$ generated by the canonical process ξ. We say that a law $\tilde{P}$ on $\tilde{\mathscr{B}}$ is solution of the martingale problem (L, C_0, μ_0) if

(M_1) The law of ξ_0 is μ_0 under $\tilde{P}$ and

(M_2) For every $\phi \in \mathscr{C}_0$ the process M^ϕ defined by $M^\phi_t := \phi(\xi_t) - \phi(\xi_0) - \int_0^t L\phi(\xi_s)\,ds$ is a martingale on $(\Omega, (\tilde{\mathscr{F}}_t), \tilde{P})$.

3.3.1. *Theorem.*

(i) Let (X^n) be a sequence of $\mathscr{D}$-semimartingales satisfying (H_1), (H_2'), (H_3). Then for any weak limit $\tilde{P}$ of the sequence $(\tilde{P}^n)$ the canonical process is continuous in probability.

(ii) If moreover (H_4) and (H_5) hold, the weak limits $\tilde{P}$ are solutions of the martingale problem $(L, \mathscr{C}_0, \mu_0)$. In particular if the martingale problem $(L, \mathscr{C}_0, \mu_0)$ has a unique solution $\tilde{P}$ then $(\tilde{P}^n)_{n\geqq 0}$ converges weakly to $\tilde{P}$.

Proof. (i) The first part of the theorem follows from the following lemma. (An analogous proof allows us to show that under the same conditions one has $\tilde{P}\{|\Delta\xi_\tau| > 0 = 0\}$ for every predictable stopping time τ.)

Lemma. If $\tilde{P}$ is the weak limit of the sequence of the probability laws $(\tilde{P}^n)$ of processes (X^n) satisfying the Aldous condition, the canonical process ξ is continuous in probability for $\tilde{P}$.

Proof of the lemma. If the sequence $\tilde{P}^n$ converges to $\tilde{P}$, there exists a dense set $T \in \mathbb{R}^+$ such that, for every $u \in T$, the laws of X^n_u converge to the law of ξ_u for $\tilde{P}$. Let us assume that for some η, $\varepsilon > 0$ and $t \in \mathbb{R}^+$, for every $h > 0$, one can find t_1 and t_2 with $t-h<t_1<t<t_2<t+h$ and $\tilde{P}\{|\xi_{t_2}-\xi_{t_1}|>n\}>\varepsilon$. Because of the right continuity of ξ, t_1 and t_2 may be assumed to be in T. Then

$$\lim_{n\to\infty} P^n\{|\xi_{t_1}-\xi_{t_2}|>\eta\}>\varepsilon.$$

But this contradicts the Aldous condition, which implies the existence of h such that for every t_1

$$\lim_{n\to\infty}\sup_{t_1\leqq t\leqq t_1+2h} P\{|\xi_t-\xi_{t_1}|>\eta/2\}=0.$$

(ii) As a consequence of (i) the functions $\tilde{\omega}\to\xi_t(\tilde{\omega})$ are continuous at $\tilde{P}$-almost all points $\tilde{\omega}$. For every $s<t\in\mathbb{R}^+$, $\phi\in\mathscr{C}_0$ and bounded $\tilde{\mathscr{B}}$-measurable function Φ on $\tilde{\Omega}$, the mapping

$$\tilde{\omega}\to\left[\phi(\xi_t(\tilde{\omega}))-\phi(\xi_s)-\int_s^t L\phi(\xi_s(\omega))\,ds\right]\Phi(\omega)$$

is $\tilde{P}$-a.s. continuous and bounded.

If $(\tilde{P}^{n'})$ is a convergent subsequence of $\tilde{P}^n$ we can therefore write

$$\lim_{n'}\tilde{E}^{n'}\left[\Phi(\cdot)\left[\Phi(\xi_t)-\phi(\xi_s)-\int_s^t L\phi(\xi_s)\,ds\right]\right]$$
$$=\tilde{E}\left[\Phi(\cdot)\left[\phi(\xi_t)-\phi(\xi_s)-\int_s^t L\phi(\xi_s)\,ds\right]\right].$$

To prove the martingale property we only have to show that, when Φ is $\tilde{\mathscr{F}}_s$-measurable, this limit is 0.

But, by hypothesis,

$$\bar{E}^{n'}\left\{\Phi(\cdot)\left[\phi(\xi_t)-\phi(\xi_s)-\int_s^t L^n(\phi,\xi_{s^-},s,\cdot)\,dA_s^n\right]\right\}=0$$

for all n'.

The required nullity is then a consequence of (H_4) and of the following relations which are implied by (H_5):

$$\lim_n \Phi\left|\int_s^t L\phi(\xi_s)\,ds-\int_s^t L\phi(\xi_s)\,dA_s^n\right|=0 \qquad \tilde{P}\text{-a.s.}$$

(Note that $s\to L\phi(\xi_s)$ is cadlag.)

3.3.2. *Example.* Let us return to the example in Section 1.3.2 with $\mathscr{C}=\mathscr{C}^2(\mathbb{R})$. In this (critical) case $b^n=0$ and $a^n(x,s,\cdot)=\beta x$. Tightness is a very trivial consequence of Proposition 3.2.3. Now take for $\mathscr{C}_0$ the subset of functions in $\mathscr{C}$ with compact support and for $\phi\in\mathscr{C}_0$

$$L\phi(x)=\beta x\frac{\partial^2\phi}{\partial x^2}(x) \quad \text{for} \quad x\geqq 0 \quad \text{and} \quad L\phi(x)=0 \quad \text{for} \quad x<0.$$

One can write

$$L^n\phi(x)=L\phi(x)+L_3^n\phi(x)$$

with

$$L_3^n\phi(x)=\frac{1}{\varepsilon_n^2}\beta x\sum_k[\phi(x+\varepsilon_n(k-1))-\phi(x)-\varepsilon_n(k-1)\phi'(x)-\tfrac{1}{2}\varepsilon_n^2(k-1)^2\phi''(\xi_x)]\nu(k)$$

for $x\geqq 0$ and $L^n\phi(x)=0$ for $x\leqq 0$. Since ϕ, ϕ' and ϕ'' are bounded, one may find for each $\varepsilon_n>0$ an integer k_ε such that

$$\sup_x\left|\sum_{k>k_\varepsilon}[\phi(x+\varepsilon_n(k-1)-\phi(x)-\varepsilon_n(k-1)\phi'(x)-\tfrac{1}{2}\varepsilon_n(k-1)^2\phi''(\xi_x)]\nu(k)\right|$$
$$\leqq \varepsilon_n^2\sup_x|\phi''(x)|\left(\sum_{k>k_\varepsilon}(k-1)^2\nu(k)\right)\leqq\varepsilon_n^2.$$

For such a k_ε one can moreover write

$$\frac{1}{\varepsilon_n^2}\limsup_n\sup_x\left|\sum_{k\leqq k_\varepsilon}[\phi(x+\varepsilon_n(k-1)-\phi(x)-\varepsilon_n(k-1)\phi'(x)-\tfrac{1}{2}\varepsilon_n^2(k-1)^2\phi''(\xi_x)]\nu(k)\right]\leqq\lim_n\varepsilon_n(k_\varepsilon-1)^3\|\phi'''\|_\infty.$$

One has therefore

$$\lim_{n\to\infty}\sup_x|L_3^n\phi(x)|=0$$

for every $\phi\in\mathscr{C}_0$ and convergence in law to the diffusion with generator $1_{[x\geqq 0]}\beta x(\partial^2/\partial x^2)$.

4. Multitype branching processes

4.1. *Notation.* We introduce the following notation:

R^d: the d-dimensional Euclidean space; its elements will be denoted by $\boldsymbol{x}=(x_1,\cdots,x_d)$,

R^{+d}: the subset of R^d whose elements have non-negative coordinates:

$$\{\boldsymbol{x}\in R^d: x_i\geqq 0,\ i=1,\cdots,d\},$$

N^d: the lattice points of R^d; its elements are $\boldsymbol{i}=(i_1,\cdots,i_d)$, i_j being integers,

$N^{+d}=R^{+d}\cap N^d$,

$\delta_{i,j}$: the Kronecker symbol,

$\boldsymbol{e}_j$: the j-unit vector, $j=1,\cdots,k$; $\boldsymbol{e}_j=(\delta_{1,j},\cdots,\delta_{d,j})$,

φ^l: $R^d\to R$, the projection map $\varphi^l(x_1,\cdots,x_d)=x_l$,

$\boldsymbol{I}$: the unit $d\times d$ matrix.

For a matrix $\boldsymbol{M}$ (or a vector $\boldsymbol{x}$), $\boldsymbol{M}^{\mathrm{T}}$ (or $\boldsymbol{x}^{\mathrm{T}}$) will denote the transpose.

For each i, $i=1,\cdots,d$ let $\boldsymbol{\xi}_i=(\xi_{i,1},\cdots,\xi_{i,d})$ be a random element of N^{+d} whose probability distribution on N^{+d} is given by $p_i(\boldsymbol{j})=p_i(j_1,\cdots,j_d)$.

We shall assume that $E\boldsymbol{\xi}_i$ and $E\xi_{i,j}^2$ are finite.

The elements of the mean matrix $\boldsymbol{M}$ are given by

$$M_{i,j}=\sum_{\boldsymbol{l}} l_j p_i(\boldsymbol{l}), \qquad i,j=1,\cdots,d;$$

and the elements of the covariance matrices by

$$\sigma_{i,j}(l)=\sum_{\boldsymbol{r}}(r_i-M_{li})(r_j-M_{lj})p_l(\boldsymbol{r}).$$

$\boldsymbol{\sigma}_{ij}$ will denote the column vector given by

$$\boldsymbol{\sigma}_{ij}^{\mathrm{T}}=(\sigma_{ij}(1),\cdots,\sigma_{ij}(d)), \qquad \sigma_k:=\sigma_{k,k}(k).$$

$\boldsymbol{x}\cdot\boldsymbol{y}$ will denote the scalar product of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$.

4.2. *Discrete time*

4.2.1. *The model.* A multitype Galton–Watson process is a Markov chain $\mathbf{Z}_l = (Z_{l,1}, \cdots, Z_{l,d})$ whose state space is N^{+d} and whose transition probability $P(\boldsymbol{i}, \boldsymbol{j})$ is given by

$$P(\boldsymbol{i}, \boldsymbol{j}) = p_1^{*i_1} * p_2^{*i_2} \cdots * p_d^{*i_d}(\boldsymbol{j}), \tag{4.2.1}$$

where $*$ denotes the product of convolution.

The quantities defined in Section 3.1 are easily computed, and we obtain

$$L\varphi(\boldsymbol{i}) = \sum_{\boldsymbol{j}} (\varphi(\boldsymbol{j}) - \varphi(\boldsymbol{i}))p_1^{*i_1} \cdots p_d^{*i_d}(\boldsymbol{j}) \tag{4.2.2}$$

$$\boldsymbol{b}(\boldsymbol{i}) = (b_1(\boldsymbol{i}), \cdots, b_d(\boldsymbol{i})) = \boldsymbol{i}(\boldsymbol{M} - \boldsymbol{I}) \tag{4.2.3}$$

$$a_{i,j}(\boldsymbol{l}) = ((\boldsymbol{M}^{\mathrm{T}} - \boldsymbol{I})\boldsymbol{l}^{\mathrm{T}}\boldsymbol{l}(\boldsymbol{M} - \boldsymbol{I}))_{ij} + \boldsymbol{l}\boldsymbol{\sigma}_{i,j}. \tag{4.2.4}$$

We consider a sequence of such processes $\mathbf{Z}_l^{(n)}$, $n = 1, 2, \cdots$, using the above notation with superscript n, and we introduce the sequence $\boldsymbol{X}^{(n)}(t)$ of normalized processes:

$$\boldsymbol{X}^{(n)}(t) = \varepsilon_n \mathbf{Z}[nt], \qquad t \geqq 0 \tag{4.2.5}$$

where ε_n is a sequence of constants decreasing to 0. For the processes $\boldsymbol{X}^{(n)}(t)$ we have (proceeding as in Example of Section 1.1.3: $A_n(t) = 1/n[nt]$):

$$L^{(n)}(\boldsymbol{\varphi}, \boldsymbol{x}) = nL\left(\psi, \frac{1}{\varepsilon_n}\boldsymbol{x}\right) \quad \text{where} \quad \psi(\boldsymbol{x}) = \boldsymbol{\varphi}(\varepsilon_n \boldsymbol{x}) \tag{4.2.6}$$

$$\boldsymbol{b}^{(n)}(\boldsymbol{x}) = n\boldsymbol{x}(\boldsymbol{M}^{(n)} - \boldsymbol{I}) \tag{4.2.7}$$

$$a_{i,j}^{(n)}(\boldsymbol{x}) = n((\boldsymbol{M}^{(n)\mathrm{T}} - \boldsymbol{I})\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}(\boldsymbol{M}^{(n)} - \boldsymbol{I}))_{i,j} + n\varepsilon_n \boldsymbol{x}\boldsymbol{\sigma}_{i,j}^{(n)}. \tag{4.2.8}$$

From this it follows that

$$\operatorname{trace} a^{(n)}(\boldsymbol{x}) = \frac{\|\boldsymbol{b}^{(n)}(\boldsymbol{x})\|^2}{n} + n\varepsilon_n \boldsymbol{x} \sum_{i=1}^{d} \boldsymbol{\sigma}_{i,i}^{(n)} \tag{4.2.9}$$

and the sequence of martingales

$$\mathcal{M}^{(n)}(t) = \boldsymbol{X}^{(n)}(t) - \boldsymbol{X}^{(n)}(0) - n\int_0^t \boldsymbol{X}_n(\tau -)(\boldsymbol{M}^{(n)} - \boldsymbol{I})\, dA_n(\tau) \tag{4.2.10}$$

will have their Doob–Meyer increasing process given by

$$\langle \mathcal{M}^{(n)} \rangle_t = n\varepsilon_n \int_0^t \boldsymbol{X}^{(n)}(s -) \sum_{i=1}^{d} \boldsymbol{\sigma}_{ii}^{(n)}\, dA_n(s) \tag{4.2.11}$$

(use formula (3.1.3), and observe that the quadratic terms cancel each other).

From Proposition 3.2.3 it follows that the processes $\boldsymbol{X}^{(n)}$ and $\mathcal{M}^{(n)}$ will be tight as soon as $n(\boldsymbol{M}^{(n)} - \boldsymbol{I})$, $\boldsymbol{\sigma}_{ii}^{(n)}$ and $n\varepsilon_n$ are bounded.

It is now easy to derive sufficient conditions for the weak convergence of $\boldsymbol{X}^{(n)}$ and to identify the limit.

4.2.2. *Theorem.* Let $\mathbf{Z}^{(n)}$ be a sequence of multitype Galton–Watson processes such that

(i) $$\boldsymbol{M}^{(n)} = \boldsymbol{I} + \frac{1}{n}\boldsymbol{C}_n \quad \text{with} \quad \lim_{n\to\infty} \boldsymbol{C}_n = C$$

(ii) $$\lim_{n\to\infty} \sigma_{ll}^{(n)}(l) = \sigma_l, \qquad l = 1, \cdots, d, \qquad \boldsymbol{\sigma}_{ij}^{(n)} \leqq K \quad \text{for some } K,$$

(iii) $$\lim_n \sum_{\boldsymbol{r}:\|\boldsymbol{r}\| > \varepsilon\sqrt{n}} (r_i - m_{l,i}^{(n)})^2 p_l^{(n)}(\boldsymbol{r}) = 0, \quad \begin{matrix} l = 1, \cdots, d, & \text{for any } \varepsilon > 0 \\ i = 1, \cdots, d. & \end{matrix}$$

(That is, we consider a family $\{\xi_l^{(n)} : n \in N, l = 1, \cdots, d\}$ of independent random variables with respective laws $p_l^{(n)}$, satisfying Lindeberg's condition.) Then, given a sequence ε_n such that $\lim_{n\to\infty} n\varepsilon_n = \alpha$, $\alpha > 0$, the processes

$$\boldsymbol{X}^{(n)}(t) = \varepsilon_n \mathbf{Z}^{(n)}[nt], \qquad \boldsymbol{X}^n(0) = \boldsymbol{x}_0^{(n)}, \qquad \lim \boldsymbol{x}_0^{(n)} = \boldsymbol{x}_0, \ \boldsymbol{x}_0 \neq 0$$

will converge weakly to the unique diffusion process in R^{+d} starting at $\boldsymbol{x}_0$ with generator given on C^2 functions by:

$$L := \sum_{i=1}^{d} (\boldsymbol{x}\boldsymbol{C})_i \frac{\partial}{\partial x_i} + \frac{1}{2}\alpha \sum_{l=1}^{d} x^l \sigma_l \frac{\partial^2}{\partial x_l^2}. \tag{4.2.12}$$

Proof. We have only to check the hypothesis (H_4) of Theorem 3.3.1 and to establish the uniqueness of the martingale problem associated with (4.2.12). First, observe that our assumptions imply

$$\lim_n \sigma_{ij}^{(n)}(l) = 0 \quad \text{if} \quad (i, j) \neq (l, l).$$

Next we shall make use of the following lemma.

Lemma. Let $\boldsymbol{\eta}_i^{(n)}$, $i = 1, \cdots, n$ be a sequence of independent identically distributed d-dimensional random vectors with mean 0 and covariance matrix $\boldsymbol{\sigma}^{(n)}$ such that $\lim_{n\to\infty} \boldsymbol{\sigma}^{(n)} = \boldsymbol{\sigma}$. Let $\boldsymbol{S}_n = \sum_{i=0}^{n} \boldsymbol{\eta}_i^n$; if the $\boldsymbol{\eta}_i^{(n)}$ satisfy the central limit theorem, namely the Lindeberg condition: for any $\varepsilon > 0$

$$\lim_{n\to\infty} E\|\boldsymbol{\eta}_i^{(n)}\|^2 I_{\{\|\boldsymbol{\eta}^{(n)}\| > \varepsilon\sqrt{n}\}} = 0$$

then $\|S_n\|^2/n$ is uniformly integrable and for any bounded continuous function f

vanishing at 0

$$\lim_{n\to\infty} E\frac{\|\boldsymbol{S}_n\|^2}{n} f\left(\frac{\boldsymbol{S}_n}{n}\right) = 0.$$

Proof. $\boldsymbol{S}_n/\sqrt{n}$ converges weakly to a d-dimensional Gaussian vector G with mean 0 and covariance $\boldsymbol{\sigma}$. By the Skorokhod representation theorem (Skorokhod (1956)) there are random variables $\boldsymbol{S}_n'$, $\boldsymbol{G}'$ with the same law as $\boldsymbol{S}_n$, $\boldsymbol{G}$ such that $\boldsymbol{S}_n'/\sqrt{n}$ converges almost surely to $\boldsymbol{G}'$. Since $E(\|\boldsymbol{S}_n\|^2/n)$ converges to $E\|\boldsymbol{G}\|^2$, it follows that $\|\boldsymbol{S}_n\|^2/n$ is uniformly integrable and so is $(\|\boldsymbol{S}_n'\|^2/n)f(\boldsymbol{S}_n'/n)$, since f is bounded. This last sequence converges almost surely to 0 and, taking expectations, the lemma is established.

We return to the proof of the theorem. We take for $\mathscr{C}_0$ the class of bounded C^2-functions φ on R^{+d}. The generators are given by

$$L^{(n)}(\varphi, \boldsymbol{x}) = nE(\varphi(\varepsilon_n \boldsymbol{Z}_n) - \varphi(\boldsymbol{x}))$$

where

$$\boldsymbol{Z}^{(n)} = \sum_{k=1}^{d} \sum_{j=1}^{x_k/\varepsilon_n} \xi_{k,j}^{(n)}$$

and the $\xi_{k,j}^{(n)}$ are independent random vectors distributed according to $p_k^{(n)}$.

Let us observe that

$$E\boldsymbol{Z}^{(n)} = \frac{1}{\varepsilon_n}\boldsymbol{x}\left(I + \frac{1}{n}\boldsymbol{C}_n\right) = \frac{1}{\varepsilon_n}\boldsymbol{x}_n$$

and

$$(\mathrm{Cov}\, \boldsymbol{Z}^{(n)})_{ij} = \frac{1}{\varepsilon_n}\boldsymbol{x}\boldsymbol{\sigma}_{ij}^{(n)}.$$

Writing

$$L^{(n)}(\varphi, \boldsymbol{x}) = nE(\varphi(\varepsilon_n \boldsymbol{Z}_n) - \varphi(\boldsymbol{x}_n)) + n(\varphi(\boldsymbol{x}_n) - \varphi(\boldsymbol{x}))$$

and using Taylor's series with remainder, we obtain

$$L^{(n)}(\varphi, \boldsymbol{x}) = \sum_{i=1}^{d} (\boldsymbol{x}C_n)_i \frac{\partial\varphi}{\partial x_i}(\boldsymbol{x}) + \frac{n\varepsilon_n}{2}\sum_{l=1}^{d}\sum_{i,j} x_l \sigma_{ij}^{(n)}(l)\frac{\partial^2\varphi}{\partial x_i \partial x_j}(x_n) + R_n(\boldsymbol{x})$$

where $R_n(\boldsymbol{x})$ is given by

$$R_n(\boldsymbol{x}) = \frac{n}{2}\sum_{ij} E(\varepsilon_n \boldsymbol{Z}_n - \boldsymbol{x}_n)_i(\varepsilon_n \boldsymbol{Z}_n - \boldsymbol{x}_n)_j\left(\frac{\partial\varphi}{\partial x_i \partial x_j}(\lambda_n^{ij}) - \frac{\partial\varphi}{\partial x_i \partial x_j}(\boldsymbol{x}_n)\right)$$

where $\|\lambda_n^{ij}\| \leqq \|\boldsymbol{x}_n - \varepsilon_n \boldsymbol{Z}_n\|$. It follows from the lemma that $R_n(\boldsymbol{x})$ converges to 0,

and moreover

$$|R_n(\boldsymbol{x})| \leqq C\,\|\boldsymbol{x}\|.$$

To finish the proof we have to check condition (H_4) and prove the uniqueness of the solution of the martingale problem $(L, \mathscr{C}_0, x)$.

We write (H_4) in the form

$$\lim_{n\to\infty}\int_0^t E^n\,|L^n(\varphi, \xi_{s^-}) - L(\varphi, \xi_{s^-})|\,dA_n(s) = 0;$$

knowing that $\lim_{n\to\infty} R_n(\xi_{s^-}) = 0$ and

$$R_n(\xi_{s^-}) \leqq C\,\|\xi_{s^-}\| \quad \text{for some constant } C$$

and using the fact that

$$E(R^n(\xi_{s^-})) \leqq K\exp(\alpha t) \quad \text{for some constants } K \text{ and } \alpha,$$

the property (H_4) follows immediately from the Lebesgue dominated convergence theorem and

$$\lim_{n\to\infty}\int_0^t K\exp(\alpha s)(dA_{s^-}^n\, ds) = 0.$$

The uniqueness of the solution of the martingale problem in R^{+d} associated with the diffusion operator L and with initial condition $\boldsymbol{x}_0$ is easily seen in the following way. It is a standard result in the study of martingale problems (see Strook and Varadhan (1979), Theorem 6.2.3, p. 142) that, if the solutions Φ_x of the martingale problem $(L, \mathscr{C}_0, \boldsymbol{x}_0)$ for all $\boldsymbol{x}_0 \in R^{+d}$ are such that for any $t > 0$ the law $\xi_t^{-1} \circ \tilde{P}_{\boldsymbol{x}_0}$ is uniquely determined, then the marginals $(\xi_{t_1}^{-1} \circ \tilde{P}_{\boldsymbol{x}_0}, \cdots, \xi_{t_n}^{-1} \circ \tilde{P}_{\boldsymbol{x}_0})$ are uniquely determined for all $\boldsymbol{x}_0$ and any finite family $t_0 < t_1 < \cdots < t_n$. Therefore $\tilde{P}_{\boldsymbol{x}_0}$ is unique for any initial condition $\boldsymbol{x}_0$.

If $\tilde{P}$ is any limit law of the laws $\tilde{P}_n$ of the processes $\boldsymbol{X}^n$, we need only to check that the moments of ξ_t under $\tilde{P}$ are uniquely determined. But, if ψ is a monomial of degree k, the particular form of L gives

$$E(\psi(\xi_n)) - \boldsymbol{x}_0 - \int_0^t \sum_i a_i^k E(\psi^{k,i}(\xi_n))\,du - \int_0^t \sum_i b_i^k E(\psi^{k-1,i}(\xi_n))\,du = 0$$

where $\{\psi^{k,i} : i \in I_k\}$ is the family of all monomials of degree k, and a_i^k and b_i^k are constants bounded by $K\,k(k-1)$. The moments of order k are therefore recursively determined.

Remarks. (a) The statement of the theorem also appears in Buckholtz and Wasan (1982), under stronger assumptions and with a formal proof which seems to contain some gaps.

(b) The assumptions of Theorem 2.2 are essentially necessary for weak convergence to a diffusion; the following example shows that if we keep all the above assumptions except (iii) we shall still have weak convergence of $\boldsymbol{X}^{(n)}$ to a Markov jump process.

4.2.3. *Example.* We consider the sequence $Z^{(n)}$ of multitype Galton–Watson processes associated with the probabilities $p_l(j)$ whose Fourier transforms are given by

$$E\exp(i\boldsymbol{u}\cdot\xi_l^{(n)})=\left(1-\frac{\lambda}{n^2}\right)\exp(iu_l)+\frac{\lambda}{n^2}\psi_l(n\boldsymbol{u}),\qquad l=1,\cdots,d$$

where $\psi_l(\boldsymbol{u})$ is the Fourier transform of a non-negative random vector $\boldsymbol{Y}_l$ with finite covariance. The conditions (i) and (ii) of the theorem being satisfied, the associated sequence $\boldsymbol{X}^{(n)}(t)=Z^{(n)}_{[nt]}/n$ is tight; we note, however, that condition (iii) is violated.

The action of the generator L_n on the functions $\exp(i\boldsymbol{u}\cdot\boldsymbol{x})$ using formula (4.2.2)–(4.2.6) yields:

$$L(\exp(i\boldsymbol{u}\cdot\boldsymbol{x}),\boldsymbol{x})=\lim_{n\uparrow\infty}L_n(\exp(i\boldsymbol{u}\cdot\boldsymbol{x}),\boldsymbol{x})=\exp(i\boldsymbol{u}\cdot\boldsymbol{x})\lambda\sum_{l=1}^{k}x_l(\psi_l(\boldsymbol{u})-1).$$

In general, for any function f continuous and bounded,

$$L(f,\boldsymbol{x})=-\lambda\sum_{l=1}^{d}x_lf(\boldsymbol{x})+\lambda\sum_{l=1}^{d}x_l\int_{R^{+d}}f(\boldsymbol{x}+\boldsymbol{y})\,dG_l(\boldsymbol{y})\tag{4.2.13}$$

where G_l denotes the distribution of $\boldsymbol{Y}_l$.

(4.2.13) is the generator of the process $\boldsymbol{X}(t)$ in R^{+d}, which starting at the point $\boldsymbol{x}$, will stay there for a random time T_x whose distribution is exponential with parameter $\lambda\sum_{l=1}^{k}x_l$, then it will jump to $\boldsymbol{x}+\boldsymbol{y}$ where the law of $\boldsymbol{y}$ is given by $\sum_l x_lG_l(\boldsymbol{y})/\sum x_l$.

From Theorem 3.3.1 it will follow that $\boldsymbol{X}^{(n)}$ converges weakly to $\boldsymbol{X}$ provided one can show that the martingale problem associated with L has a unique solution. This follows easily from the fact that the martingale problem associated with L characterizes the distributions of the waiting time in a state and the jumps.

4.3. *Limit of the accompanying martingale of a critical process (discrete time).* In the critical case we are inspired by the work of Kurtz (1975) to prove another limit theorem. This is a case where we study the limit of a family of martingales which are not Markovian. Let $\boldsymbol{Z}_l$ be a critical process with mean matrix $\boldsymbol{M}$ which is assumed to be primitive, i.e. there is an integer m such that all the elements of $\boldsymbol{M}^m$ are strictly positive and the largest eigenvalue of $\boldsymbol{M}$ is 1. It then follows from the Perron–Frobenius theory that to the eigenvalue 1

there will correspond a right eigenvector $\boldsymbol{\mu}^{\mathrm{T}}$ and a left eigenvector $\boldsymbol{v}$ whose components can be chosen to be strictly positive and normalized in such a way that

$$\sum_{i=1}^{d}\mu_i=1,\qquad \boldsymbol{v}\boldsymbol{\mu}^{\mathrm{T}}=1\tag{4.3.1}$$

$$M^n=\boldsymbol{R}+\boldsymbol{Q}^n\quad\text{with}\quad R_{ij}=\mu_iv_j,\qquad \boldsymbol{RQ}=\boldsymbol{QR}=0$$

and there are constants c and ρ, $0<\rho<1$ such that for all n

$$|Q_{ij}^n|\leqq c\rho^n.$$

A multitype Galton–Watson process whose mean matrix is primitive is called positively regular. We assume also the existence of $\boldsymbol{\sigma}$, (see Section 4.4.1). As above let us consider the sequence of processes

$$\boldsymbol{X}^{(n)}(t)=\varepsilon_n\boldsymbol{Z}_{[nt]}$$

with $\lim_{n\uparrow\infty}n\varepsilon_n=\alpha$, $\alpha>0$. Then since $\boldsymbol{X}^{(n)}(t)$ has non-negative coordinates and $E_x\boldsymbol{X}^{(n)}(t)=\boldsymbol{x}M^{[nt]}$ is bounded by (4.3.1) it follows easily from (4.2.11) that $\langle\mathcal{M}^{(n)}\rangle_t$ will satisfy the assumptions of Aldous's theorem and from Rebolledo's theorem (4.3.2) $\mathcal{M}^{(n)}$ will be a tight sequence as soon as $\boldsymbol{X}^{(n)}(0)=\boldsymbol{x}_0^{(n)}$ are bounded. Let us assume that $\lim_{n\to\infty}\boldsymbol{x}_0^{(n)}=\boldsymbol{x}_0$.

Let

$$\boldsymbol{Y}^{(n)}(t)=\boldsymbol{y}+n\int_0^t\boldsymbol{X}^{(n)}(\tau-)(\boldsymbol{M}-\boldsymbol{I})\,dA_n(\tau).$$

Then the couple $(\boldsymbol{X}^{(n)}(t),\boldsymbol{Y}^{(n)}(t))$ defines for each n a discrete Markov chain with time parameter j/n, $j=0,1,2,\cdots$ and state space R^{2d}; its transition probability is given by

$$P_n(\boldsymbol{x},\boldsymbol{y},\boldsymbol{u},\boldsymbol{v})=P_n(\boldsymbol{x},\boldsymbol{u})\delta_{\boldsymbol{v},\boldsymbol{y}+\boldsymbol{x}(\boldsymbol{M}-\boldsymbol{I})}$$

where $P_n(\boldsymbol{x},\boldsymbol{u})$ is the transition probability of the process $\boldsymbol{X}^{(n)}$.

It follows that if

$$L_nf(\boldsymbol{x},\boldsymbol{y})=\sum_{\boldsymbol{u},\boldsymbol{v}}P_n(\boldsymbol{x},\boldsymbol{y};\boldsymbol{u},\boldsymbol{v})[f(\boldsymbol{u},\boldsymbol{v})-f(\boldsymbol{x},\boldsymbol{y})]$$

then

$$f(\boldsymbol{X}^{(n)}(t),\boldsymbol{Y}^{(n)}(t))-f(\boldsymbol{X}^{(n)}(0),\boldsymbol{Y}^{(n)}(0))-n\int_0^tL_n(f(\boldsymbol{X}^{(n)}(\tau-),\boldsymbol{Y}^{(n)}(\tau-))\,dA_n(\tau)$$

will be a martingale. Choosing $f(\boldsymbol{x},\boldsymbol{y})$ of the form $f(\boldsymbol{x},\boldsymbol{y})=\varphi(\boldsymbol{x}-\boldsymbol{y})$ with φ in

$C_2(\mathbb{R}^d)$, and using the Taylor formula with remainder, one obtains:

$$\varphi(\boldsymbol{X}^{(n)}(t)-\boldsymbol{Y}^{(n)}(t))-\varphi(\boldsymbol{X}^{(n)}(0)-\boldsymbol{Y}^{(n)}(0))$$

$$(4.3.2)\qquad -\tfrac{1}{2}n\varepsilon_n\int_0^t\sum_{i,j}\boldsymbol{X}^{(n)}(-\tau)\boldsymbol{\sigma}_{i,j}(\cdot)\frac{\partial^2\varphi}{\partial X_i\partial X_j}(\boldsymbol{X}^{(n)}(\tau-)-\boldsymbol{Y}^{(n)}(\tau-))\,dA_n(\tau)+E_n$$

is a martingale. The remainder E_n is of the form

$$E_n=n\int_0^t R_n(\boldsymbol{X}^{(n)}(\tau-),\,\boldsymbol{Y}^{(n)}(\tau-))\,dA_n(\tau)$$

with

$$R_n=\sum_j R_j,$$

$$R_j(\boldsymbol{x},\boldsymbol{y})=\frac{1}{2}\sum_{l,m}(\varepsilon_n\boldsymbol{j}-\boldsymbol{x}\boldsymbol{M})_l(\varepsilon_n\boldsymbol{j}-\boldsymbol{x}\boldsymbol{M})_m P\Big(\frac{x}{\varepsilon_n},\boldsymbol{j}\Big)\Big[\frac{\partial^2\varphi}{\partial x_l\partial x_m}(\boldsymbol{\xi}_j)-\frac{\partial^2\varphi}{\partial x_l\partial x_m}(\boldsymbol{x}-\boldsymbol{y})\Big]$$

where $\|\boldsymbol{\xi}_j-(\boldsymbol{x}-\boldsymbol{y})\|\leqq\|\varepsilon_n\boldsymbol{j}-\boldsymbol{x}\boldsymbol{M}\|$.

Using the lemma as in the proof of Theorem 4.2.2 one obtains

$$\lim_{n\to\infty}R_n(\boldsymbol{x},\boldsymbol{y})=0\quad\text{and}\quad\lim_{n\to\infty}E_n=0.$$

To check (H_4) with

$$L\varphi(\boldsymbol{z})=\alpha\frac{1}{2}\sum_{i,j}\boldsymbol{z}\boldsymbol{R}\boldsymbol{\sigma}_{ij}\frac{\partial^2\varphi}{\partial z_i\partial z_j}$$

we have now to show that

$$\lim_{n\to\infty}\int_0^t E^n\alpha\sum_{i,j}\boldsymbol{Y}^{(n)}(\tau-)\boldsymbol{R}\boldsymbol{\sigma}_{ij}\frac{\partial^2\varphi}{\partial z_i\partial z_j}(\boldsymbol{x}_{\tau-}^{(n)}-\boldsymbol{y}_{\tau-}^{(n)})\,dA_z^n=0$$

and this follows from the fact that $\lim_{n\to\infty}E\,\|Y_z^n\|^2=0$ which is obtained from the explicit form of the covariance of Z_n (see for instance Harris (1963), p. 37).

Therefore any weak limit $\bar{P}$ of the laws $P^{(n)}$ of the processes $\mathcal{M}^{(n)}$ is a solution of the martingale problem $(L,\mathscr{C}_0,x_0)$ on the half space $\boldsymbol{x}\boldsymbol{R}\geqq 0$ (where all the processes $\mathcal{M}^{(n)}$ take their values), with $\mathscr{C}_0$ the set of C^2 functions. Exactly as in the proof of Theorem 4.2.2 the uniqueness of $\bar{P}$ follows from the argument on the uniqueness of the moments.

We have thus obtained the following theorem.

4.3.1. *Theorem.* Let $\boldsymbol{Z}_n$ be a multitype critical positive regular Galton–Watson process with mean matrix $\boldsymbol{M}$ and covariance matrices $\boldsymbol{\sigma}_{i,j}$. Let $\boldsymbol{R}$ be the projection matrix in the Perron–Frobenius decomposition of M. Then if $\boldsymbol{X}^{(n)}(t)=\varepsilon_n\boldsymbol{Z}^n_{[nt]}$ with $\boldsymbol{X}^{(n)}(0)=\boldsymbol{x}_0^{(n)}$ converging to $x_0\neq 0$, $n\varepsilon_n\to\alpha$, $\alpha>0$, the sequence

$$\boldsymbol{\xi}^{(n)}(t)=\boldsymbol{X}^{(n)}(t)-n\int_0^t\boldsymbol{X}^{(n)}(\tau-)(\boldsymbol{M}-\boldsymbol{I})\,dA_n(\tau)$$

will converge weakly to the unique diffusion in the half space $\boldsymbol{x}\boldsymbol{R}\geqq 0$ with diffusion operator L given by

$$L\varphi=\boldsymbol{\alpha}\boldsymbol{X}\boldsymbol{R}\frac{1}{2}\sum_{ij}\boldsymbol{\sigma}_{ij}\frac{\partial^2\varphi}{\partial x_i\partial x_j}\quad(\text{all }\varphi\ C^2\text{-functions on }\boldsymbol{x}\boldsymbol{R}\geqq 0)$$

and initial condition $\boldsymbol{x}_0$.

4.3.2. *Remarks.* (a) Since $\boldsymbol{\xi}_n^{(n)}\boldsymbol{R}=\boldsymbol{X}^{(n)}\boldsymbol{R}$ and $\boldsymbol{X}^{(n)}(\boldsymbol{I}-\boldsymbol{R})$ goes to 0 it follows that $n\int_0^t\boldsymbol{X}^{(n)}(\tau-)(\boldsymbol{M}-\boldsymbol{I})\,dA_n(\tau)$ behaves like the diffusion $\boldsymbol{\xi}^{(n)}(\boldsymbol{I}-\boldsymbol{R})$.

$\boldsymbol{X}\boldsymbol{R}$ is the projection of the vector $\boldsymbol{X}$ on the vector $\boldsymbol{v}$ in the direction perpendicular to $\boldsymbol{\mu}$.

(b) A particular case of an age-dependent process. Let $\boldsymbol{Z}(t)$ be an age-dependent process whose lifetime distribution T is discrete with a finite number of jumps at the integers, i.e.

$$P(T\leqq x)=\sum_1^k g_j\delta_j(x).$$

If the process is critical we can describe $\mathbf{Z}$ by a $(k+1)$-type Galton–Watson process where the type denotes the age of the particle. The mean matrix M will be given by

$$M=\begin{pmatrix} 0 & 1 & 2 & & & k\\ 0 & 1 & 0 & & & 0\\ 1-r_1 & 0 & r_1 & & & \\ 1-r_2 & 0 & 0 & r_2 & & \\ \vdots & & & & & \\ 1-r_k & & & & & r_{k-1}\\ 1 & 0 & & & & 0 \end{pmatrix}$$

where

$$r_j=\frac{g_{j+1}+\cdots+g_k}{g_j+\cdots+g_k}=P(T\geqq j+1\mid T\geqq j).$$

Letting $r_0=1$, $r_k=0$, we have

$$M_{ij}=\begin{cases}1-r_i & \text{if } j=0\\ r_i & \text{if } j=i+1\\ 0 & \text{otherwise.}\end{cases}$$

The right eigenvector is $\mu=(1,\cdots,1)$ and the left eigenvector ν is given by $\nu_0=1/(1+ET)$, $\nu_j=g_{j+1}+\cdots+g_k=P(T\geqq j)/(1+ET)$ and our theorem applies to this situation.

4.4 *Continuous time*

4.4.1. *The model.* We consider a system of d types of particles. λ_i denotes the rate of death of a particle of type i and $p_i(\boldsymbol{j})=p_i(j_1,\cdots,j_d)$ the probability that when it dies $j_1,\cdots,j_d$ particles of type $1,\cdots,d$ are created. Let $\boldsymbol{Z}(t)=(Z^{(1)}(t),\cdots,Z^{(d)}(t))$ denote the random vector of the number of particles of type $1,\cdots,d$ in the system at time t.

We adapt the notation of 4.4.1 by setting in an analogous way

$$M_{ij}:=\sum_l l_j p_i(\boldsymbol{l})$$

$$\boldsymbol{\Lambda}=\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_d\end{pmatrix}$$

$$\sigma_{m,l}(i)=\sum_r (r_m-\delta_{im})(r_l-\delta_{il})p_i(\boldsymbol{r}),\qquad \sigma_l=\sigma_{l,l}(l)$$

$\boldsymbol{\sigma}^{\mathrm{T}}_{m,l}=(\sigma_{m,l}(1)\cdots\sigma_{m,l}(d))$, where T denotes the transpose

$\boldsymbol{e}_j=(\delta_{1j},\cdots,\delta_{dj})$ the jth unit vector

$\boldsymbol{\alpha}=\boldsymbol{\Lambda}(\boldsymbol{M}-\boldsymbol{I})$

N^{+d} the set of d-dimensional vectors whose components are non-negative integers.

Under those assumptions it can be shown that this system defines a unique Markov process $\boldsymbol{Z}(t)$ whose state space is N^{+d}. The quantities defined in Section 3.1 are given by

$$L(\varphi,\boldsymbol{x})=-(\lambda_1x_1+\cdots+\lambda_dx_d)\varphi(\boldsymbol{x})+\sum_{i=1}^d\lambda_ix_i\sum_{\boldsymbol{j}\in N^{+d}}\varphi(\boldsymbol{x}+\boldsymbol{j}-\boldsymbol{e}_i)p_i(\boldsymbol{j}),\tag{4.4.1}$$

$$\boldsymbol{b}(\boldsymbol{x})=\boldsymbol{x}\boldsymbol{\alpha}\tag{4.4.2}$$

and

$$a_{ij}(\boldsymbol{x})=\boldsymbol{x}\cdot\boldsymbol{\Lambda}\cdot\boldsymbol{\sigma}_{ij}(\cdot).\tag{4.4.3}$$

Moreover

$$E\boldsymbol{Z}(t)=\boldsymbol{Z}(0)\exp(\boldsymbol{\alpha}t).\tag{4.4.4}$$

(This implies in particular the positivity of the matrix $\exp(\boldsymbol{\alpha}t)$.)

We consider now a sequence $(\boldsymbol{Z}^{(n)}(\cdot))_{n>0}$ of such processes using the above notation with superscript n. We introduce the sequence $\boldsymbol{X}^{(n)}(t)=\varepsilon_n\boldsymbol{Z}^{(n)}(nt)$ (noting that the $\boldsymbol{Z}^n(nt)$ are as $\boldsymbol{Z}^n(t)$ but with an increasing intensity of jumps which come to the same as accelerating the time) and we obtain for their generator

$$L^n\varphi(\boldsymbol{x})=n\sum\lambda_i\frac{x_i}{\varepsilon_n}\sum_{\boldsymbol{j}\in N^d}[\varphi(\boldsymbol{x}+\varepsilon_n\boldsymbol{j}-\varepsilon_n\boldsymbol{e}_i)-\varphi(\boldsymbol{x})]p_i(\boldsymbol{j}).\tag{4.4.5}$$

We have all the ingredients to establish an analogue of Theorem (4.2.2), but, instead, we shall modify the model to deal with population-size dependent multitype branching processes, generalizing the results of Lipow (1977).

Keeping the above notation, we let the quantities $p_i^{(n)}$ and $\lambda_i^{(n)}$ be functions of $\boldsymbol{x}\in N^{+d}$. The new processes are still Markovian, but (4.4.5) will take the form:

$$\begin{aligned}L^{(n)}(\varphi,\boldsymbol{x})&=n\sum_{i=1}^d\lambda_i^{(n)}(\boldsymbol{x}\varepsilon_n^{-1})\,x_i\varepsilon_n^{-1}\\&\quad\times\sum_{\boldsymbol{j}\in N^{+d}}(\varphi(\boldsymbol{x}+\varepsilon_n\boldsymbol{j}-\varepsilon_n\boldsymbol{e}_i)-\varphi(\boldsymbol{x}))p_i^{(n)}(\boldsymbol{j},\boldsymbol{x}\varepsilon_n^{-1})\end{aligned}\tag{4.4.6}$$

from which it follows that

$$\boldsymbol{b}^{(n)}(\boldsymbol{x})=n\boldsymbol{x}\boldsymbol{\alpha}_n(\varepsilon_n^{-1}\boldsymbol{x})\tag{4.4.7}$$

$$a_{ij}^{(n)}(\boldsymbol{x})=n\varepsilon_n\boldsymbol{x}\boldsymbol{\Lambda}^{(n)}(\varepsilon_n^{-1}\boldsymbol{x})\boldsymbol{\sigma}_{ij}^{(n)}(\varepsilon_n^{-1}\boldsymbol{x}).\tag{4.4.8}$$

We make the following assumptions:

(i) $\sup_n\sup_x\Lambda^{(n)}(\varepsilon_n^{-1}\boldsymbol{x})<\infty$, $\lim_{n\uparrow\infty}\Lambda^{(n)}(\varepsilon_n^{-1}\boldsymbol{x})=\Lambda(\boldsymbol{x})$;

(ii) $\sup_n\sup_x n\boldsymbol{\alpha}_n(\varepsilon_n^{-1}\boldsymbol{x})<\infty$ $\lim_{n\uparrow\infty}n\boldsymbol{\alpha}_n(\varepsilon_n^{-1}\boldsymbol{x})=\boldsymbol{C}(\boldsymbol{x})$;

(iii) $\lim_{N\uparrow\infty}\sup_n\sup_x\sum_{\boldsymbol{j}:\|\boldsymbol{j}\|>N}\|\boldsymbol{j}\|^2p_l^{(n)}(\boldsymbol{j},\boldsymbol{x})=0$, $l=1,\cdots,d$;

(iv) $\sup_n\sup_x\sigma_l^{(n)}(\boldsymbol{x})<\infty$, $\lim_{n\uparrow\infty}\lambda_l^{(n)}(\varepsilon_n^{-1}\boldsymbol{x})\sigma_l^{(n)}(\varepsilon_n^{-1}\boldsymbol{x})=\sigma_l(\boldsymbol{x})$;

(v) Let $\mathscr{C}_0$ be the space of C^2 bounded continuous functions on R^{+d} and L be defined on $\mathscr{C}_0$ by

$$L=\sum_{i=1}^{d}(\boldsymbol{x}C(\boldsymbol{x}))_i\frac{\partial}{\partial x_i}+\frac{\alpha}{\varepsilon}\sum_{i=1}^{d}x_i\sigma_i(\boldsymbol{x})\frac{\partial^2}{\partial x_i^2},\qquad \alpha>0. \tag{4.4.9}$$

Then the martingale problem associated with $(L, \mathscr{C}_0, \boldsymbol{x}_0)$, $\boldsymbol{x}_0\in R^{+d}$ has a unique solution (this is true in particular if $d=1$ or if $C(\boldsymbol{x})$ and $\sigma(\boldsymbol{x})$ are constants).

4.4.2. *Theorem.* Let $\boldsymbol{Z}^{(n)}(t)$ be a sequence of population-size-dependent multitype branching processes, satisfying the assumptions (i) to (v), above, and let $\boldsymbol{X}^{(n)}(t)=\varepsilon_n\boldsymbol{Z}^{(n)}(nt)$ with $\lim \varepsilon_n n=\alpha$, $\alpha>0$ and $\boldsymbol{X}^{(n)}(0)=\boldsymbol{x}_0^{(n)}$ with $\lim \boldsymbol{x}_0^{(n)}=\boldsymbol{x}_0$, $\boldsymbol{x}_0\neq 0$. Then the sequence $\boldsymbol{X}^{(n)}(t)$ converges weakly to the unique diffusion in R^{+d} starting at $\boldsymbol{x}_0$ with generator L given by (4.4.9).

Proof. The proof follows the steps of the proof of Theorem 4.2.2 but the computations are much easier. However, we need an estimate on $E\boldsymbol{X}^{(n)}(t)$ which is easily obtained from the Gronwall inequality: since

$$\boldsymbol{X}^{(n)}(t)-\boldsymbol{X}^{(n)}(0)-n\int_0^t \boldsymbol{X}^{(n)}(t)\boldsymbol{\alpha}^{(n)}(\varepsilon_n^{-1}\boldsymbol{X}^{(n)}(\tau))\,d\tau$$

is a martingale, taking expectations and using (ii) one obtains that there is a constant K such that

$$E\boldsymbol{X}^{(n)}(t)\leqq \boldsymbol{X}^{(n)}(0)+K\int_0^t E\boldsymbol{X}^{(n)}(\tau)\,d\tau$$

from which it follows that

$$E\boldsymbol{X}^{(n)}(t)\leqq \boldsymbol{X}^{(n)}(0)+K\exp(Kt).$$

Remark. Being in the continuous-time case, the branching structure of the above processes has completely disappeared. One deals here with diffusion approximations of a relative large class of Markov processes.

4.4.3. *The critical case.* Theorem 4.3.1 has also a version for multitype branching processes in the continuous-time case. Going back to the notation of the beginning of this section, the process $\boldsymbol{Z}(t)$ whose generator is given by (4.4.1) will be assumed to be positive regular, i.e. there exists a $t_0>0$ such that for all $i, j=1,\cdots,d$, $(\exp \boldsymbol{\alpha}t_0)_{ij}>0$. Under that assumption the eigenvalues of $\exp(\boldsymbol{\alpha}t)$ are given by $\exp(\lambda_i t)$, where λ_i are the eigenvalues of $\boldsymbol{\alpha}$ and they can be arranged in such a way that

$$\lambda_1>\operatorname{Re}\lambda_2\geqq\cdots\geqq\operatorname{Re}\lambda_k.$$

The left and right eigenvectors $\boldsymbol{v}$ and $\boldsymbol{u}$ of λ_1 can be chosen with all coordinates

strictly positive and normalized so that

$$\boldsymbol{u}\cdot\boldsymbol{v}=1,\qquad \boldsymbol{u}\cdot 1=1.$$

We shall assume that the process is critical, i.e. $\lambda_1=0$. Then the following facts are well known:

$$\lim_{t\to\infty}\exp(\alpha t)=\boldsymbol{R}=((u_i, v_j))$$

and the $\sigma_{ij}(l)$ being finite, imply that the covariance matrices of $\boldsymbol{Z}(t)$ grow linearly with t. (Details can be found, for instance, in Chapter V of Athreya and Ney (1972).)

Theorem. Let $\boldsymbol{Z}(t)$ be a critical positive regular multitype branching process, with parameters $\boldsymbol{\alpha}=\Lambda(M-I)$ and $\boldsymbol{\sigma}$. Let $\boldsymbol{R}$ be the projection matrix in the Perron–Frobenius decomposition of $\exp(\boldsymbol{\alpha}t)$. Then, if $\boldsymbol{X}^{(n)}(t)=\varepsilon_n\boldsymbol{Z}(nt)$ with $\boldsymbol{X}^{(n)}(0)=\boldsymbol{x}_0^{(n)}$ converging to $\boldsymbol{x}_0$, $\boldsymbol{x}_0\neq 0$ and $\lim n\varepsilon_n=\alpha$, $\alpha>0$, the sequence of martingales

$$\xi^{(n)}(t)=\boldsymbol{X}^{(n)}(t)-n\int_0^t \boldsymbol{X}^{(n)}(\tau^-)\boldsymbol{\alpha}\,d\tau \tag{4.4.10}$$

will converge weakly to the unique diffusion in the half space $\boldsymbol{xR}\geqq 0$, with generator L given by

$$L\varphi=\alpha\boldsymbol{XR}\boldsymbol{\Lambda}\frac{1}{2}\sum_{ij}\boldsymbol{\sigma}_{ij}\frac{\partial^2\varphi}{\partial x_i\partial x_j}$$

and initial condition $\boldsymbol{x}_0$ (φ is any C^2-function on $\boldsymbol{xR}\geqq 0$).

Proof. One can easily imitate the proof of Theorem 4.3.1; we sketch the main steps of a different argument.

For any function φ on $\boldsymbol{R}^d$ of class C^1 we recall the elementary formula

$$\varphi(\xi_t)-\varphi(\xi_0)=\sum_{s\leqq t}\varphi(\xi_s)-\varphi(\xi_{s^-})-\operatorname{grad}\varphi(\xi_{s^-})\cdot\Delta\xi_s+\int_0^t \operatorname{grad}\varphi(\xi_{s^-})\,d\xi_s$$

where ξ_s is a function of bounded variation. For φ in C^2 we apply this formula to the martingale $\xi^{(n)}$ in (4.4.10) to obtain, using Taylor's formula with remainder,

$$\varphi(\xi^{(n)}(t))-\varphi(\xi^{(n)}(0))$$
$$=\frac{1}{2}\sum_{i,j=1}^{d}\sum_{s\leqq t}(\xi^{(n)}(s)-\xi^{(n)}(s^-))_i(\xi^{(n)}(s)-\xi^{(n)}(s^-))_j\frac{\partial^2\varphi}{\partial x_i\partial x_j}(\xi^{(n)}(s^-))+R_n$$

(first-order terms cancel in the expansion). Observing that the sum over s can be written as

$$\int_0^t \frac{\partial^2\varphi}{\partial x_i \partial x_j}(\xi_s^{(n)})\,d[(X^{(n)})_i(X^{(n)})_j]$$

(where [ ] denotes the mutual quadratic variation) and one obtains, using (1.2.8),

$$\varphi(\xi^{(n)}(t)) - \varphi(\xi^{(n)}(0)) = \frac{1}{2}\sum_{i,j=1}^{d} \int_0^t \frac{\partial^2\varphi}{\partial x_i \partial x_j}(\xi_s^{(n)})\,d\langle X_i^{(n)}, X_j^{(n)}\rangle + R_n + \text{martingale}.$$

But from the proof of Lemma 3.1.3

$$\langle X_i^{(n)} X_j^{(n)}\rangle_t = n\varepsilon_n \int_0^t a_{ij}(X_s^{(n)})\,ds$$

where $\boldsymbol{a}$ is given by (4.4.3).

Finally we obtain

$$\varphi(\xi_t^{(n)}) - \varphi(\xi_0^{(n)}) - \frac{1}{2}\sum_{i,j} n\varepsilon_n \int_0^t \frac{\partial^2\varphi}{\partial x_i \partial x_j}(\xi_{\tau-}^{(n)})X^{(n)}(\tau^-)\Lambda\sigma_{ij}\,d\tau$$
$$= \text{martingale} + R_n$$

from which the proof follows as in Theorem 4.2.2.

Acknowledgements

We are grateful to many colleagues for valuable conversations on this work; among them, let us mention T. Kurtz, S. Orey, D. Stroock and T. Traynor.

References

ALDOUS, D. (1978) Stopping times and tightness. *Ann. Prob.* **6,** 335–340.

ALDOUS, D. (1981) Weak convergence and the general theory of process. Preprint, University of California, Berkeley.

ATHREYA, K. B. AND NEY, P. (1972) *Branching Processes.* Springer-Verlag, Berlin.

BILLINGSLEY, P. (1969) *Convergence of Probability Measures.* Wiley, New York.

BILLINGSLEY, P. (1974) Conditional distributions and tightness. *Ann. Prob.* **2,** 480–485.

BUCKHOLTZ, P. G. AND WASAN, M. T. (1982) Diffusion approximation of the two types Galton-Watson process with mean matrix close to the identity matrix. *J. Multivariate Anal.* **12,** 493–507.

DELLACHERIE, C. AND MEYER, P. A. (1979) *Probabilités et potentiel* II. Hermann, Paris.

FELLER, W. (1951) Diffusion in genetics. *Proc. 2nd Berkeley Symp. Math. Statist. Prob.,* 227–246.

GRIGELIONIS, B. (1973) On relative compactness of sets of probability measures in $D_{[0,\infty[}(R)$. *Lit. Math. Sb.* **13,** 83–96.

HARRIS, T. E. (1963) *The Theory of Branching Processes.* Springer-Verlag, Berlin.

JACOD, J. (1979) *Calcul stochastique et problèmes de martingales.* Lecture Notes in Mathematics **714,** Springer-Verlag, Berlin.

JACOD, J., MEMIN, J. AND MÉTIVIER, M. (1983) On tightness and stopping times. *Stoch. Proc. Appl.* **14,** 109–141.

JAGERS, P. (1971) Diffusion approximation of branching processes. *Ann. Math. Statist.* **42,** 2074–2078.

JIRINA, M. (1969) On Feller's branching diffusion processes. *Casopis Pest. Mat.* **94,** 84–90.

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

KURTZ, T. (1978) Diffusion approximations of branching processes. In *Advances in Probability* **5,** ed. A. Joffe and P. Ney, Dekker, New York, 262–292.

KURTZ, T. (1981) *Approximation of Population Processes.* CBMS-NSF Regional Conference Series in Applied Mathematics 36, SIAM, Philadelphia.

LAMPERTI, J. (1962) Limiting distributions for branching processes. *Proc. 5th Berkeley Symp. Math. Statist. Prob.* **2**(2), 225–242.

LAMPERTI, J. (1967) The limit of a sequence of branching processes. *Z. Wahrscheinlichkeitsth.* **7,** 271–288.

LENGLART, E. (1977) Relations de domination entre deux processus. *Ann. Inst. H. Poincaré* B **13,** 171–179.

LINDVALL, T. (1973) Weak convergence of probability measures and random functions in the space $D[0, \infty[$. *J. Appl. Prob.* **10,** 109–121.

LIPOW, C. (1977) Limiting diffusions for population-size dependent branching processes. *J. Appl. Prob.* **14,** 14–24.

MÉTIVIER, M. (1979) On weak convergence of processes. Internal report, Dept of Mathematics, University of Minnesota, Minneapolis.

MÉTIVIER, M. (1982) *Semimartingales.* De Gruyter, Berlin.

MÉTIVIER, M. (1984) Convergence faible et principe d'invariance pour des martingales à valeurs dans des espaces de Sobolev. Rapport interne de l'Ecole Polytechnique.

PARTHASARATHY, K. R. (1967) *Probability Measures on Metric Spaces.* Academic Press, New York.

PLATEN, E. AND REBOLLEDO, R. (1981) Weak convergence of semi-martingales and discretisation methods. Akad. Wiss. DDR. Preprint P. Maths 15181.

REBOLLEDO, R. (1979) La méthode des martingales appliquée à la convergence en loi des processus. *Mem. Soc. Math. France Suppl.,* 1–125.

SKOROKHOD, S. V. (1956) Limit theorems for stochastic processes. *Theory Prob. Appl.* **1,** 261–290.

STROOCK, D. W. AND VARADHAN, S. R. S. (1979) *Multidimensional Diffusion Processes.* Springer-Verlag, Berlin.

WHITT, W. (1980) Some useful functions for functional limit theorems. *Maths. Operat. Res.* **5,** 67–85.