Stochastic Processes and their Applications 1 (1973) 335–368. © North-Holland Publ. Co.

ON THE TRANSITION FROM A MARKOV CHAIN TO A CONTINUOUS TIME PROCESS

Anders GRIMVALL

Department of Mathematics, Linköping University, Linköping, Sweden

Received 19 January 1973
Revised 13 June 1973

Abstract. Starting from a real-valued Markov chain $X_0, X_1, ..., X_n$ with stationary transition probabilities, a random element $\{Y(t); t \in [0, 1]\}$ of the function space D[0, 1] is constructed by letting $Y(k/n) = X_k$, $k = 0, 1, ..., n$, and assuming $Y(t)$ constant in between. Simple tightness criteria for sequences $\{Y(t);\ t \in [0, 1]\}_n$ of such random elements in D[0, 1] are then given in terms of the one-step transition probabilities of the underlying Markov chains. Applications are made to Galton–Watson branching processes.

AMS Subj. Classif.: Primary 60B10, 60J05; Secondary 60J80

measures on function spaces	tightness
Markov chains	branching processes

1. Introduction and summary

Let $\{X_{n,0}, X_{n,1}, X_{n,2}, ..., X_{n,n}\}_n$ be a sequence of real-valued Markov chains with stationary transition probabilities $p^{(n)}(a, \cdot)$; that is, for every Borel set E, the relation

$$\mathrm{P}[X_{n,k+1} \in E \mid X_{n,1}, X_{n,2}, ..., X_{n,k}] = p^{(n)}(X_{n,k}, E) \tag{1.1}$$

is satisfied with probability 1. With each of these Markov chains we associate a continuous-time process $Y_n(t)$ defined by

$$Y_n(t) = \begin{cases} X_{n,k} & \text{for } k/n \le t < (k+1)/n\ , \\ X_{n,n} & \text{for } t = 1\ . \end{cases} \tag{1.2}$$

Then $\{Y_n(t);\ t \in [0, 1]\}$ can be considered as a random element of the space D[0, 1] consisting of all functions on [0, 1] with no discontinuities of the second kind. With the Skorokhod topology (see [1, p. 111]), this space becomes a complete separable metric space. In this paper we will study $\mathcal{D}$-convergence of the sequence $\{Y_n(t);\ t \in [0,1]\}_n$; that is, weak convergence of the corresponding sequence of probability measures on D[0, 1].

Assume that $\{Y_n(t);\ t \in [0,1]\}_n$ is $\mathcal{D}$-convergent with limit $\{Y(t);\ t \in [0,1]\}$ and let h be a functional on D[0, 1] which is continuous with respect to the Skorokhod topology. Then we have (see [1, p. 30])

$$h(Y_n) \overset{w}{\to} h(Y) \quad \text{as } n \to \infty . \tag{1.3}$$

This shows that $\mathcal{D}$-convergence can be a useful tool when we want to study properties of the processes $\{Y_n(t);\ t \in [0,1]\}$ and $\{Y(t);\ t \in [0,1]\}$ that can not be expressed in terms of their finite-dimensional distributions. If the distribution of $h(Y)$ is known, (1.3) gives an approximate distribution of $h(Y_n)$ for large $n \in N$. On the other hand, if the distribution of $h(Y)$ is unknown, we can sometimes choose the approximating processes $\{Y_n(t);\ t \in [0,1]\}$ so simple that (1.3) yields some information about the distribution of $h(Y)$. The last method is particularly important when simulation techniques are employed.

By a famous theorem due to Prokhorov (see [1, p. 37]), a sequence $\{Y_n\}_n$ of random elements in D[0, 1] is conditionally compact if and only if it is tight. This suggests a useful method to establish $\mathcal{D}$-convergence. First we show that the finite-dimensional distributions converge and then we prove that $\{Y_n\}_n$ is tight (see [1, p. 124]). From classical probability theory we have a rich supply of tools for determining convergence of finite-dimensional distributions. Therefore, we will in this paper confine our interest to tightness criteria.

$\mathcal{D}$-convergence in connection with Markov processes, in particular diffusion processes, has been treated by Skorokhod, Gikhman, Borovkov and others. Since the infinitesimal approach to a diffusion process is the most convenient one, their conditions for $\mathcal{D}$-convergence usually have been based on the asymptotic behaviour of the two first moments of the increments within a short time-interval. Here we will mainly emphasize "continuity properties" of the transition probabilities $p^{(n)}(a,\cdot)$ considered as functions of a. It has also been our aim to give our tightness criteria a simple form. Therefore, they have, to the greatest possible extent, been based on properties of the one-dimensional projections of our processes and the one-step transition probabilities of our Markov chains.

The plan for this paper is as follows. In Section 2 we start by showing how the general tightness conditions in [1] can be simplified, when the processes $\{Y_n(t);\ t \in [0,1]\}$ are constructed from Markov chains as in (1.2). At the end of the same section we make our first attempt to relate the tightness of the sequence $\{Y_n\}_n$ to the properties of the projections $Y_n(t)$. The main results here are generalizations of corresponding results in [14].

Even if the $\{Y_n(t);\ t \in [0,1]\}$ are constructed from Markov chains, all limit processes need not be Markov processes. In Section 3 we will give sufficient conditions for this to occur. These conditions will take a particularly simple form if the $\{Y_n(t);\ t \in [0,1]\}$ are constructed from stochastically monotone Markov chains; that is, Markov chains such that the transition probabilities $p(a, \{x;\ x \leq y\})$ are non-increasing in a for each fixed y.

In Section 4 we will continue to study the relations between the properties of the projections $Y_n(t)$ and the tightness of the sequence $\{Y_n\}_n$. All Markov chains considered in that section are stochastically monotone.

Section 5 is devoted to an application of the theory in earlier sections. We will study $\mathcal{D}$-convergence of a sequence of normalized critical Galton–Watson processes. In fact, we will be able to show that $\mathcal{D}$-convergence in this case is equivalent to convergence of the finite-dimensional distributions, provided we make an exception for degenerate limits.

2. Conditional compactness of a sequence of Markov chains

From now on, $\{X_{n,0}, X_{n,1}, \ldots, X_{n,n}\}_n$ will always denote a sequence of Markov chains with stationary transition probabilities. If nothing else is stated, we will assume that $\mathbf{P}[X_{n,0} = 0] = 1$. The one-step transition probabilities of the n^{th} Markov chain are denoted by $p^{(n)}(a, \cdot)$. Transition probabilities corresponding to several steps are denoted by $q_\Delta^{(n)}(a, \cdot)$, where $n\,\Delta$ is the number of steps and Δ is assumed to be chosen from the set $\{j/n:\ j = 0, 1, 2, \ldots, n\}$. Thus, for every Borel set E, the relation

$$\mathbf{P}[X_{n,k+n\Delta} \in E \mid X_{n,k}] = q_\Delta^{(n)}(X_{n,k}, E)$$

is fulfilled with probability 1.

The continuous-time process $\{Y_n(t);\ t \in [0,1]\}$ defined by (1.2) will be called the process or the "random Markov line" associated with $\{X_{n,0}, X_{n,1}, ..., X_{n,n}\}$. When $\{Y_n(t); t \in [0,1]\}$, considered as a random element of D[0, 1], converges weakly to the random element $\{Y(t); t \in [0,1]\}$ of D[0, 1], we will write

$$Y_n \xrightarrow{\mathcal{D}} Y \qquad \text{as} \quad n \to \infty . \tag{2.2}$$

All our theorems will be stated for Markov chains with the real line as common state space. But there should be no difficulty to give corresponding results when the state space is the half-line $[0,\infty)$ or a compact interval.

We are now ready to give the fundamental theorem on conditional compactness of a sequence of random Markov lines.

Theorem 2.1. *Let $\{X_{n,0}, X_{n,1}, ..., X_{n,n}\}_n$ be a sequence of Markov chains with transition probabilities $q_\Delta^{(n)}(a, E)$ satisfying* (2.1), *and let $\{Y_n(t); t \in [0,1]\}_n$ be the associated sequence of continuous-time processes. Assume that*

(i) $\mathrm{P}[\sup_{0 \le t < 1} |Y_n(t)| > \lambda] \to 0$ *as* $\lambda \to \infty$, *uniformly in* $n \in \mathrm{N}$;

(ii) *for every compact set C and every $\epsilon > 0$ there exists $\delta = \delta(C, \epsilon) > 0$ such that*

$$q_\Delta^{(n)}(a, (a-\epsilon, a+\epsilon)^c) < \epsilon$$

for all $n \in \mathrm{N}$, $a \in C$ and $\Delta \le \delta$.

Then the sequence $\{Y_n(t); t \in [0,1]\}_n$ of random elements in D[0, 1] *is tight.*

Proof. We shall show that the conditions for tightness given in [1, Theorem 15.2] are satisfied. But this can be done by an almost verbatim repetition of the arguments in [14, p. 182]. Further details are therefore omitted.

Let $p^{(n)}(a, a + \mathrm{d}x)$ denote the probability measure which to each Borel set E assigns the number $p^{(n)}(a, E_a)$, where $E_a = \{a+x; x \in E\}$. Intuitively, $p^{(n)}(a, a+\mathrm{d}x)$ corresponds to the conditional distribution of $X_{n,k} - X_{n,k-1}$, given $X_{n,k-1} = a$. Although the transition probabilities $q_\Delta^{(n)}(a, E)$ always can be expressed directly in terms of the one-step transition probabilities, it is in many cases easier to calculate the convolutions of the measures $p^{(n)}(a, a+\mathrm{d}x)$. Therefore, we shall state and prove two theorems where the tightness conditions are given in terms of these convolutions. But first we consider the case when $X_{n,k}$ is the k^{th} partial sum of a sequence of independent equally distributed random variables. Then the measures $\mu_a^{(n)}(\mathrm{d}x) = \mu^{(n)}(\mathrm{d}x)$, defined as in Theorem 2.2 below, can be taken independent of $a \in \mathbf{R}$, and by a theorem due to Prokhorov (see [14, p. 197]), the sequence $\{Y_n\}_{n\in\mathbf{N}}$ in $\mathrm{D}[0,1]$ is tight if and only if $\{(\mu^{(n)})^{n*}\}_{n\in\mathbf{N}}$ is tight. The following theorem generalizes this fact.

Theorem 2.2. *Let* $\{X_{n,0}, X_{n,1}, X_{n,2}, ..., X_{n,n}\}_n$ *be a sequence of Markov chains with transition probabilities* $p^{(n)}(a, E)$, *and denote the measure* $p^{(n)}(a, a+\mathrm{d}x)$ *by* $\mu_a^{(n)}(\mathrm{d}x)$. *Assume that*

(i) $\{(\mu_a^{(n)})^{n*}\}_{n\in\mathbf{N},a\in\mathbf{R}}$ *is tight.*

Then the random elements $\{Y_n(t)\colon t \in [0,1]\}$ *associated with the Markov chains form a tight sequence in* $\mathrm{D}[0,1]$.

Proof. Let $\varphi_a^{(n)}$ be the characteristic function of $\mu_a^{(n)}$. The family $\{(\mu_a^{(n)})^{n*}\}_{n\in\mathbf{N},a\in\mathbf{R}}$ is tight if and only if $\{(\varphi_a^{(n)})^{n}\}_{n\in\mathbf{N},a\in\mathbf{R}}$ is equicontinuous at zero. Thus, for every $\epsilon > 0$, there exists a $\delta = \delta(\epsilon) > 0$ such that

$$|(\varphi_a^{(n)}(t))^n - 1| < \epsilon \qquad \text{for all } t \in [-\delta,\delta],\ a \in \mathbf{R},\ n \in \mathbf{N}\,. \tag{2.3}$$

Using the inequalities $\log(1+x) < x$ and $\mathrm{e}^x < 1 + 2x$, valid for all x in some neighbourhood of zero, we conclude that

$$|\varphi_a^{(n)}(t)| > 1 - 2\epsilon/n \qquad \text{for all } t \in [-\delta,\delta],\ a \in \mathbf{R} \text{ and } n \in \mathbf{N}\,. \tag{2.4}$$

Similarly, for all $t \in [-\delta,\delta]$, $a \in \mathbf{R}$ and $n \in \mathbf{N}$, there exists an integer $j \in \{0, 1, 2, ..., n\}$ such that

$$\arg \varphi_a^{(n)}(t) \in [2j\pi/n - 2\epsilon/n,\ 2j\pi/n + 2\epsilon/n]\,. \tag{2.5}$$

But each $\varphi_a^{(n)}(t)$ is continuous, and $\arg \varphi_a^{(n)}(0) = 0$. Thus we must choose $j = 0$ in (2.5), and it follows that

$$|\varphi_a^{(n)}(t) - 1| < 4\epsilon/n \qquad \text{for all } t \in [-\delta,\delta],\ a \in \mathbf{R},\ n \in \mathbf{N}\,.$$

(Some of the arguments above might fail if ϵ is large but we need only consider sufficiently small ϵ.)

If $\psi_k^{(n)}$ denotes the characteristic function of $X_{k,n}$, we can easily show by induction that

$$|\psi_k^{(n)}(t) - 1| \leq 4\epsilon k/n \qquad \text{for all } t \in [-\delta, \delta] .$$

Hence, by a well-known inequality for characteristic functions (see [10, p. 65]), we get

$$\mathrm{P}[|X_{k,n}| > 2\delta^{-1}] \leq \delta^{-1} \int_{-\delta}^{\delta} |1 - \psi_k^{(n)}(t)|\, \mathrm{d}t \leq 8\epsilon k/n . \tag{2.6}$$

In order to show that condition (i) in Theorem 2.1 is satisfied we need the following Kolmogorov type inequality.

Proposition 2.3. *Let* $\{Z_0, Z_1, Z_2, ..., Z_n\}$ *be a homogeneous Markov chain such that, for some* $\epsilon > 0$, $m \in \mathrm{N}$ *and* $\lambda_0 > 0$,

$$\mathrm{P}[\sup\{|Z_j - Z_i|: 0 \leq (j-i)/n < 1/m\} > \lambda_0 \mid Z_i] < \epsilon \quad a.s.$$

Assume that we can choose $\lambda_1 \geq \lambda_0$ *so large that*

$$\sup\{\mathrm{P}[|Z_j| > \lambda_1]: j = 1, 2, ..., n\} < \epsilon/2m .$$

Then we have

$$\mathrm{P}[\sup\{|Z_j|.\ j = 1, 2, ..., n\} > 2\lambda] < 2\epsilon$$

for all $\lambda \geq \lambda_1$.

Proof. We need only consider the case $n > 2m$. Let τ be the hitting-time for the set $(2\lambda_1, \infty) \cup (-\infty, -2\lambda_1)$ and put $\tau = n+1$ if $\sup\{|Z_j|: j=1,2,...,n\} \leq 2\lambda_1$. Since $n > 2m$, we can choose integers n_i such that $(i-1)/2m < n_i/n \leq i/2m$ for $i = 1, 2, 3, ..., 2m-1$ and $n_{2m}/n = 1$. Then we get

$$\begin{aligned} &\mathrm{P}[\tau \leq n, \sup\{|Z_j - Z_\tau|: 0 \leq (j-\tau)/n < 1/m\} \leq \lambda_1] \leq \\ &\leq \sum_{i=1}^{2m} \mathrm{P}[|Z_{n_i}| > \lambda_1] < \epsilon . \end{aligned} \tag{2.7}$$

On the other hand,

$$\mathbf{P}[\tau \leq n, \sup\{|Z_j - Z_\tau|: 0 \leq (j-\tau)/n < 1/m\} > \lambda_0] =$$

$$= \sum_{i=1}^{n} \mathbf{P}[\tau = i, \sup\{|Z_j - Z_i|: 0 \leq (j-i)/n < 1/m\} > \lambda_0] . \quad (2.8)$$

By conditioning with $\mathfrak{B}(Z_0, Z_1, \ldots, Z_i)$ and using the Markov property we get

$$\mathbf{P}[\tau \leq n, \sup\{|Z_j - Z_\tau|: 0 \leq (j-\tau)/n < 1/m\} > \lambda_0] =$$

$$= \sum_{i=1}^{n} \mathbf{E}\{\mathbf{P}[\tau = i, \sup\{|Z_j - Z_i|: 0 \leq (j-i)/n < 1/m\} > \lambda_0 \mid \mathfrak{B}(Z_0, Z_1, \ldots, Z_i)]\}$$

$$= \sum_{i=1}^{n} \mathbf{E}\{I_{\{i\}}(\tau)\, \mathbf{P}[\sup\{|Z_j - Z_i|: 0 \leq (j-i)/n < 1/m\} > \lambda_0 \mid Z_i]\}$$

$$< \epsilon \sum_{i=1}^{n} \mathbf{P}[\tau = i] \leq \epsilon . \quad (2.9)$$

Hence, by (2.7) and (2.9),

$$\mathbf{P}[\sup\{|Z_j|: j = 1, 2, \ldots, n\} > 2\lambda] \leq \mathbf{P}[\tau \leq n]$$

$$= \mathbf{P}[\tau \leq n, \sup\{|Z_j - Z_\tau|: 0 \leq (j-\tau)/n < 1/m\} \leq \lambda_1]$$

$$+ \mathbf{P}[\tau \leq n, \sup\{|Z_j - Z_\tau|: 0 \leq (j-\tau)/n < 1/m\} > \lambda_1] < 2\epsilon , \quad (2.10)$$

for all $\lambda \geq \lambda_1$, and this completes the proof of the proposition.

We now return to the proof of Theorem 2.2. By the same arguments as those preceding Proposition 2.3, we can prove that

$$|\psi_k^{(n)}(t) - 1| \leq 4\epsilon k/n \qquad \text{for all } |t| \leq \delta ,$$

where $\psi_k^{(n)}$ now denotes the characteristic function of $X_{n,k} - X_{n,0}$ and the distribution of $X_{n,0}$ is arbitrary. Starting from the inequality $|\varphi(t+h) - \varphi(t)| \leq \sqrt{2|\varphi(h) - 1|}$, valid for all characteristic functions $\varphi(t)$, we can easily prove that

$$|\varphi(t) - 1| \leq k\sqrt{2|\varphi(t/k) - 1|} , \qquad k = 1, 2, 3, \ldots . \quad (2.11)$$

Applying (2.11) to the characteristic function $\psi_k^{(n)}$, we get

$$|\psi_k^{(n)}(t) - 1| \le ([\delta^{-1}] + 1)\, j\sqrt{2 \cdot 4\epsilon\, k/n} \qquad \text{for all } t \in [-j, j]\,, \tag{2.12}$$

and from (2.12) we obtain

$$\mathrm{P}[|X_{n,k} - X_{n,0}| > 2/j] \le j^{-1} \int_{-j}^{j} |\psi_k^{(n)}(t) - 1|\,\mathrm{d}t \le 2j([\delta^{-1}] + 1)\sqrt{8\epsilon\, k/n}$$

$$\text{for all } j \in \mathbf{N}\,. \tag{2.13}$$

Since the distribution of $X_{n,0}$ is arbitrary, (2.13) is equivalent to

$$\sup_{a\in\mathbf{R}} \int \cdots \int_{|x_k - a| > 2/j} p^{(n)}(a, \mathrm{d}x_1)\, p^{(n)}(x_1, \mathrm{d}x_2) \ldots p^{(n)}(x_{k-1}, \mathrm{d}x_k) \le$$

$$\le 2j([\delta^{-1}] + 1)\sqrt{8\epsilon\, k/n}\,. \tag{2.14}$$

By Kolmogorov's inequality for Markov chains (see [15, p. 157]), we can easily show that, for every $j \in \mathbf{N}$,

$$\sup_{a\in\mathbf{R}} \int \cdots \int_{\sup\{|x_i - a|:\ 0 \le i \le k\} > 4/j} p^{(n)}(a, \mathrm{d}x_1)\, p^{(n)}(x_1, \mathrm{d}x_2) \ldots p^{(n)}(x_{k-1}, \mathrm{d}x_k) \le$$

$$\le 4j([\delta^{-1}] + 1)\sqrt{8\epsilon\, k/n}\,, \tag{2.15}$$

for all k/n sufficiently small. For our original Markov chains (2.15) means that

$$\mathrm{P}[\sup\{|X_{n,m+i} - X_{n,m}|:\ 0 \le i \le k\} > 4/j \mid X_{n,m}] \le$$

$$\le 4j([\delta^{-1}] + 1)\sqrt{8\epsilon\, k/n} \qquad \text{a.s.} \tag{2.16}$$

That condition (i) of Theorem 2.1 is fulfilled now follows from (2.6), (2.16) and Proposition 2.1, while (ii) follows directly from (2.14). Thus $\{Y_n\}_n$ is tight.

Condition (i) of Theorem 2.2 is in general too strong to be useful in applications. Therefore, we shall prove two simple generalizations of that theorem. The first one, Theorem 2.2', is natural to use when we are dealing with sequences of normalized branching Markov processes. The second one, Theorem 2.4, can be applied when we are studying convergence to a Brownian motion with reflecting barrier and similar processes.

Theorem 2.2'. *Assume that*

(i) $\mathbf{P}[\sup\{|Y_n(t)|: 0 \le t \le 1\} > \lambda] \to 0$ *as* $\lambda \to \infty$, *uniformly in* $n \in \mathrm{N}$;

(ii) $\{(\mu_a^{(n)})^{n*}\}_{a\in C, n\in \mathrm{N}}$ *is tight for every compact* C.

Then the sequence $\{Y_n\}_n$ *is tight in* D[0, 1].

Proof. Consider a Markov chain with one-step transition probabilities

$$q^{(n)}(a,\cdot) = \begin{cases} p^{(n)}(\lambda,\cdot) & \text{for } a > \lambda\,, \\ p^{(n)}(a,\cdot) & \text{for } |a| \le \lambda\,, \\ p^{(n)}(-\lambda,\cdot) & \text{for } a < -\lambda\,, \end{cases}$$

and let $Y'_n(t)$ be the corresponding random Markov line. By Theorem 2.2, $\{Y'_n\}_n$ is tight in D[0, 1]. Observing that

$$\mathbf{P}[w'_\delta(Y_n) \ge \epsilon] \le \mathbf{P}[w'_\delta(Y'_n) \ge \epsilon] + \mathbf{P}[\sup\{|Y_n(t)|: 0 \le t \le 1\} > \lambda]\,,$$

where w' is the continuity modulus defined in [1, p. 110], and using [1, Theorem 15.2], we can easily complete the proof.

Application. For sequences of normalized Galton–Watson processes the conditions of Theorem 2.2' become very simple. Let, for each $n \in \mathrm{N}$, $\{Z_j^{(n)}\}_j$ denote the variables of a Galton–Watson process, where the number of off-spring of one individual is determined by the probabilities $\{p_k^{(n)}\}_k$. Define a sequence of continuous-time processes $\{Y_n\}_n$ by

$$Y_n(t) = Z^{(n)}_{[nt]}/b_n \qquad \text{for } t \in [0,1]; \quad Z_0^{(n)} = b_n\,,$$

where $b_n > 0$ are normalizing constants. If ν_n is a probability measure that gives mass $p_k^{(n)}$ to the point $(k-1)/b_n$, condition (ii) of Theorem 2.2' is satisfied if

(ii') $\{(\nu_n)^{n b_n *}\}_n$ is tight.

Condition (i) can easily be checked if we observe that $\{Y_n(t);\, t \in [0,1]\}$ is a supermartingale (submartingale).

Sometimes it is convenient to consider the subspace C[0, 1] of D[0, 1]. This subspace consists of all continuous functions on [0, 1] and the

Skorokhod topology relativized to C[0, 1] is equivalent to the topology of uniform convergence. Here we will only give an example which indicates how sufficient conditions for $\mathcal{D}$-convergence in C[0, 1] can be obtained. As before, $\{X_{n,0}, X_{n,1}, X_{n,2}, ..., X_{n,n}\}_n$ denotes a sequence of Markov chains with transition probabilities $p^{(n)}(a,\cdot)$ and $\mu_a^{(n)}(\mathrm{d}x) = p^{(n)}(a, a+\mathrm{d}x)$.

Theorem 2.4. *Assume that*

(i) *there exists a point* a_0 *such that, for every* δ*-neighbourhood* S_δ *of* a_0, *the family* $\{(\mu_a^{(n)})^{n^*}\}_{a\in S_\delta^c, n\in\mathrm{N}}$ *is tight.*

Then the sequence of measures on C[0, 1] *corresponding to the random polygonal lines* $\{Y_n(t)\colon t \in [0, 1]\}$ *defined by*

$$Y_n(t) = X_{n,k} + n(t-k/n)(X_{n,k+1} - X_{n,k}) \quad for \quad t \in [k/n, (k+1)/n]$$

is conditionally compact, provided

(ii) *for each* $\delta > 0$, $n \cdot p^{(n)}(a, (a-\delta, a+\delta)^c) \to 0$ *as* $n \to \infty$, *uniformly in* $a \in \mathbf{R}$.

Proof. Let Y_n' denote the random element in D[0, 1] defined by (1.2). We shall use Theorem 2.1 to prove that $\{Y_n'\}_n$ is tight. Let $\epsilon > 0$ be given. From the proof of Theorem 2.2 it follows that we can choose Δ_0 so small that

$$q_\Delta^{(n)}(a, (a-\epsilon, a+\epsilon)^c) < \epsilon \quad \text{for all } \Delta \le \Delta_0,\ n \in \mathrm{N} \text{ and } a \in S_{2\epsilon}^c(a_0)\,.$$

Let us now consider a Markov chain $\{Z_0, Z_1, ..., Z_{n\Delta}\}$ with transition probabilities $p^{(n)}(a,\cdot)$ and such that $\mathrm{P}[Z_0 = a] = 1$, where $a \in S_{2\epsilon}(a_0)$. Denote by τ_n the hitting time for the set $(-\infty, a_0 - 2\epsilon) \cup (a_0 + 2\epsilon, \infty)$. By the assumption (ii) we get

$$\mathrm{P}[|\tau_n - a_0| \ge 3\epsilon] < \epsilon$$

for all sufficiently large $n \in \mathrm{N}$. The strong Markov property then shows that there exists an integer n_0 such that

$$q^{(n)}(a, (a-6\epsilon, a+6\epsilon)^c) < 2\epsilon \quad \text{for all } \Delta \le \Delta_0,\ n \le n_0,\ a \in \mathbf{R}\,.$$

Similarly, we show that condition (i) in Theorem 2.1 is satisfied so that $\{Y_n'\}_n$ is tight in D[0, 1]. Observing that for all n sufficiently large the probability that $\{Y'_n(t);\, t \in [0,1]\}$ has a jump exceeding ϵ is less than ϵ, we can use [1, Theorems 8.2 and 15.2] to complete the proof.

Assume that the measures $\mu_a^{(n)}(\mathrm{d}x) = p^{(n)}(a, a+\mathrm{d}x)$ are independent of a; that is, $X_{n,k}$ is the k^{th} partial sum of a sequence of identically distributed independent random variables. Then $\{Y_n\}_n$ is tight if and only if $\{Y_n(1)\}_n$ is tight (see [14, p. 197]). We shall give a rather natural generalization of this theorem.

Theorem 2.5. *Assume that*

(i) $n \cdot p^{(n)}(a, [a-1, a+1]^c) \to 0$ *as* $n \to \infty$, *uniformly in* $a \in \mathbf{R}$;

(ii) *for every bounded interval* $[t_1, t_2]$ *there exists a constant* $K = K(t_1, t_2)$ *such that* $|\varphi_a^{(n)}(t) - \varphi_b^{(n)}(t)| \le K/n$ *for all* $t \in [t_1, t_2]$ *and all* $a, b \in \mathbf{R}$ (*here* $\varphi_a^{(n)}$ *is the characteristic function of the measure* $\mu_a^{(n)}$);

(iii) $\{Y_n(1)\}_n$ *is tight.*

Then the sequence $\{Y_n\}_n$ *of random elements in* D[0, 1] *is also tight.*

Proof. Since $p^{(n)}(a, [a-1, a+1]^c) = o(1/n)$, it is no restriction to assume that $p^{(n)}(a, [a-1, a+1]^c) = 0$. Let $\psi_a^{(n,m)}$ be the characteristic function of the measure

$$q_{m/n}^{(n)}(a, \mathrm{d}x_m) = \int_{x_1 \in \mathbf{R}} \cdots \int_{x_{m-1} \in \mathbf{R}} p^{(n)}(a, \mathrm{d}x_1)\, p^{(n)}(x_1, \mathrm{d}x_2) \ldots p^{(n)}(x_{m-1}, \mathrm{d}x_m)$$

Some simple calculations show that

$$|\psi_a^{(n,m)}(t) - [\varphi_b^{(n)}(t)]^m| \le m\, K/n \tag{2.17}$$

for all $a, b \in \mathbf{R}$ and all $t \in [t_1, t_2]$.

Put $\alpha_b^{(n)} = \int x\, \mu_b^{(n)}(\mathrm{d}x)$ and $(\sigma_b^n)^2 = \int (x - \alpha_b^n)^2\, \mu_b^{(n)}(\mathrm{d}x)$ and let $S_b^{(n,m)} = \Sigma_{i=1}^m Z_i$, where $\{Z_i\}_i$ are independent random variables, each one with distribution $\mu_b^{(n)}$. Then by the Berry–Esseen theorem on normal approximation (see [6, p. 542]) we have (notice that $|Z_i - \alpha_b^n| \le 2$)

$$\begin{aligned}|\psi_a^{(n,m)}(t/(\sigma_b^n\sqrt{m}))\, &\exp[-\mathrm{i}\, t\, m\, \alpha_b^n/(\sigma_b^n\sqrt{m})] \\ &- [\varphi_b^{(n)}(t/(\sigma_b^n\sqrt{m}))]^m \exp[-\mathrm{i}\, t\, m\, \alpha_b^n/(\sigma_b^n\sqrt{m})]| \le m\, K/n\end{aligned} \tag{2.19}$$

for all $a, b \in \mathbf{R}$ and all $t \in [t_1\, \sigma_b^n \sqrt{m},\, t_2\, \sigma_b^n \sqrt{m}]$.

Let us now consider $A = \sup\{n(\sigma_b^n)^2 : b \in \mathbf{R}, n \in \mathbf{N}\}$. Assume that A is infinite. By suitable choices of $b \in \mathbf{R}$ and $m, n \in \mathbf{N}$, we can then simultaneously make m/n arbitrarily small and $m(\sigma_b^n)^2$ arbitrarily large. Thus, by (2.18) and (2.19),

$$|q_{m/n}^{(n)}(a, \{y : y \leq x\}) - \Phi((x - m\alpha_b^n)/(\sigma_b^n \sqrt{m}))| < \epsilon$$

$$\text{for all } a \in \mathbf{R} \text{ and } x \in \mathbf{R},$$

provided m/n is small enough and $\sigma_b^n \sqrt{m}$ is large enough. But this implies that, for any given $\epsilon > 0$ and bounded interval I, we can choose $m, n \in \mathbf{N}$ so that

$$q_{m/n}^{(n)}(a, I) < \epsilon \qquad \text{for all } a \in \mathbf{R}.$$

Obviously this contradicts the assumption (iii) and so A must be finite.

By Chebyshev's inequality,

$$\mathbf{P}[|S_a^{(n,m)} - m\alpha_a^n| > \epsilon] \leq mA/n\epsilon^2 \tag{2.20}$$

If $B = \sup\{n(\alpha_a^n - a_0^n) : a \in \mathbf{R}, n \in \mathbf{N}\}$ were infinite, we would be able to choose subsequences $\{a'\} \subseteq \mathbf{R}$ and $\{m'\}, \{n'\} \subseteq \mathbf{N}$ such that $m'/n' \to 0$ and

$$S_{a'}^{(n',m')} - m'\alpha_{a'}^{n'} \xrightarrow{\text{i.p.}} 0, \qquad S_0^{(n',m')} - m'\alpha_{a'}^{n'} \xrightarrow{\text{i.p.}} c \neq 0.$$

However, this contradicts (2.17). Thus, B must be finite and $\{(\mu_a^{(n)} - \alpha_0^n)^{n*}\}_{a \in \mathbf{R}, n \in \mathbf{N}}$ is tight. By Theorem 2.2, the random Markov lines $Z_n(t) = Y_n(t) - [nt]\alpha_0^n$ form a tight sequence in D[0, 1]. By the assumption (iii), this is possible only if $C = \sup\{|n\alpha_0^n| : n \in \mathbf{N}\}$ is finite. Hence $\{Y_n\}_n$ is also tight in D[0, 1].

Remark. In Theorem 2.5 it was proved that $\{Y_n\}_n$ is tight in D[0,1] if $\{(\mu_a^{(n)})^{n*}\}_{a \in \mathbf{R}, n \in \mathbf{N}}$ is tight. The converse is not true in general. However, we can see from the proof of Theorem 2.5 that tightness of $\{Y_n\}_n$ implies tightness of $\{(\mu_a^{(n)})^{n*}\}_{a \in \mathbf{R}, n \in \mathbf{N}}$ if we make the additional assumptions (i) and (ii).

3. The Markovian character of the limit process

Assume that we have established the convergence

$$Y_n^\varepsilon \to Y^\varepsilon \qquad \text{as } n \to \infty$$

in D[0, 1]. For any continuous functional h on D[0, 1] we then get

$$h(Y_n^\varepsilon) \overset{w}{\to} h(Y^\varepsilon) \qquad \text{as } n \to \infty \,, \tag{3.1}$$

As we pointed out in the Introduction, (3.1) is useful mainly in the case when we can compute the distribution of $h(Y^\varepsilon)$. Therefore, it is interesting to see if, under general and simple conditions, we can show that the limit process $\{Y(t);\ t \in [0,1]\}$ must be of a particularly simple type. In this section we shall give sufficient conditions for the limit process to be a Markov process. In our theorems all the approximating Markov chains are assumed to be stochastically monotone. In a remark at the end of this section we will indicate how the general case can be treated. The notation is the same as in the previous sections.

We start by giving some measure-theoretical facts.

Proposition 3.1. *Let F_a and F_b be any two probability distributions with finite mean-values α_a and α_b, respectively. Assume that*

$$F_a(x) \geq F_b(x) \qquad \textit{for all } x \in \mathbf{R}\,.$$

Then

$$0 \leq \int u(x)\,F_b(\mathrm{d}x) - \int u(x)\,F_a(\mathrm{d}x)$$

$$\leq (\alpha_b - \alpha_a)\,\sup\{(u(y)-u(x))/(y-x)\colon\ x,y \in \mathbf{R},\ x \neq y\}$$

for all increasing functions u such that $\sup\{(u(y)-u(x))/(y-x)$: $x,y \in \mathbf{R}, x \neq y\}$ *is finite.*

Proof. Let $c = \sup\{(u(y)-u(x))/(y-x)\colon x,y \in \mathbf{R}, x \neq y\}$ and put $v(x) = c\,x$. For the two functions $u(x)$ and $v(x)$ it then holds that $u(y) - u(x) \leq v(y) - v(x)$ for all $x \leq y$. Observing that

$$(\alpha_b - \alpha_a)\,\sup\{(u(y)-u(x))/(y-x)\colon\ x,y \in \mathbf{R},\ x \neq y\} =$$

$$= \int v(x)\,F_b(\mathrm{d}x) - \int v(x)\,F_a(\mathrm{d}x)\,,$$

and approximating the integrals by sums we can easily complete the proof.

Proposition 3.2. *Assume that*

(i) *there is a constant K such that*

$$(b-a)^{-1} \left(\int x\, p^{(n)}(b, \mathrm{d}x) - \int x\, p^{(n)}(a, \mathrm{d}x) \right) \leq 1 + K/n$$

for all $n \in \mathbf{N}$ and $a < b$;

(ii) *$p^{(n)}(a, \{x; x \leq y\})$ is non-increasing in $a \in \mathbf{R}$ for fixed $n \in \mathbf{N}$ and $y \in \mathbf{R}$.*

Then we have

$$0 \leq (b-a)^{-1} \left(\int u(x)\, q_\Delta^{(n)}(b, \mathrm{d}x) - \int u(x)\, q_\Delta^{(n)}(a, \mathrm{d}x) \right)$$

$$\leq \mathrm{e}^{K\Delta} \sup\{(u(y)-u(x))/(y-x) : x, y \in \mathbf{R}, x \neq y\}$$

for all increasing functions u.

Proof. Apply Proposition 3.1.

In Section 4 we shall also need the following result.

Proposition 3.3. *Let $\{p^{(n)}(a, \cdot)\}_{a \in \mathbf{R}, n \in \mathbf{N}}$ be a family of transition probabilities satisfying conditions (i) and (ii) of Proposition 3.2, and assume that, for some $\epsilon > 0$ and $a \in \mathbf{R}$,*

$$q_\Delta^{(n)}(a, (a+\epsilon, \infty)) \geq \epsilon .$$

Then there exists $\delta > 0$, depending only on ϵ and K, such that

$$q_\Delta^{(n)}(b, (b + \tfrac{1}{2}\epsilon^2, \infty)) \geq \tfrac{1}{4}\epsilon$$

for all $b \in [a-\delta, a+\delta]$. Similarly,

$$q_\Delta^{(n)}(a, (-\infty, a-\epsilon)) \geq \epsilon$$

implies that

$$q_\Delta^{(n)}(b, (-\infty, b - \tfrac{1}{2}\epsilon^2)) \geq \tfrac{1}{4}\epsilon$$

for all $b \in [a-\delta, a+\delta]$, where δ depends only on ϵ and K.

Proof. It is enough to prove the first assertion. Applying Proposition 3.2 to the function

$$u(x) = \begin{cases} 1 & \text{for } x \geq a+\epsilon\,, \\ (x-a)/\epsilon & \text{for } a \leq x \leq a+\epsilon, \\ 0 & \text{for } x \leq a\,, \end{cases}$$

we can easily choose $\delta > 0$ depending only on ϵ and K, such that $\int q_\Delta^{(n)}(b, \mathrm{d}x)u(x) > \frac{3}{4}\epsilon$ for all $b \in [a-\delta, a]$. Some simple estimations then show that

$$q_\Delta^{(n)}(b, (b+\tfrac{1}{2}\epsilon^2, \infty)) > \tfrac{1}{4}\epsilon$$

for all $b \in [a-\delta, a]$. On the other hand, for $b \in [a, a+\frac{1}{2}\epsilon]$, it follows from the stochastical monotonicity that

$$q_\Delta^{(n)}(b, (b+\tfrac{1}{2}\epsilon^2, \infty)) \geq q_\Delta^{(n)}(b, (a+\epsilon, \infty)) > \epsilon\ .$$

Remark. For simplicity, we will always assume that the transition probabilities $p^{(n)}(a, \cdot)$ have finite mean values. Actually, if there is a constant K' such that

$$\sup\{p^{(n)}(a, (a-K', a+K')^c) \colon a \in \mathbf{R}\} = o(1/n)\ ,$$

neither the convergence nor the limit is changed if we let $p^{(n)}(a, \cdot)$ give all its mass to the finite interval $(a-K', a+K')$.

Under conditions (i) and (ii) of Proposition 3.2, it is possible to prove that the weak limit in $D[0,1]$ of a sequence $\{Y_n\}_n$ of random Markov lines must be a stochastically continuous Markov process. However, the proof of the stochastical continuity is quite technical unless we make the additional assumption that for every $\epsilon > 0$,

$$q_\Delta^{(n)}(a, (a-\epsilon, a+\epsilon)^c) \to 0 \qquad \text{as } \Delta \to 0\,, \tag{3.3}$$

uniformly in $n \in \mathbf{N}$ and $a \in C$, for each compact set C. In view of Theorem 2.1, the assumption (3.3) is rather natural. It will also permit us to give a direct construction of the semi-group corresponding to the limit process.

Theorem 3.4. *Let $\{Y_n\}_n$ be a tight sequence of random Markov lines. Assume that the transition probabilities satisfy condition (3.2) above as well as*

(i) *there exists a constant K such that*

$$(b-a)^{-1}\left[\int x\, p^{(n)}(b,\mathrm{d}x) - \int x\, p^{(n)}(a,\mathrm{d}x)\right] \le 1 + K/n$$

for all $n \in \mathbf{N}$ and $a < b$,

(ii) *$p^{(n)}(a,\{x; x \le y\})$ is non-increasing in $a \in \mathbf{R}$ for fixed $n \in \mathbf{N}$ and $y \in \mathbf{R}$.*

Then every limit process of $\{Y_n\}_n$ is a stochastically continuous Markov process having a Feller semi-group.

Proof. By Prokhorov's theorem, $\{Y_n\}_n$ is relatively compact in $D[0,1]$, so we can without restriction assume that $\{Y_n\}_n$ converges weakly to some random element Y in $D[0,1]$. From condition (3.2) above, it is easily deduced that $\{Y(t); t \in [0,1]\}$ is stochastically continuous. Let us then consider the projection π_t taking $x(\cdot) \in D[0,1]$ into $x(t) \in \mathbf{R}$. π_t is continuous at $x(\cdot) \in D[0,1]$ if and only if $x(\cdot)$ is continuous at t. Thus, the stochastical continuity of $\{Y(t); t \in [0,1]\}$ implies that

$$Y_n(t) \overset{w}{\to} Y(t) \qquad \text{as } n \to \infty \tag{3.3}$$

for all $t \in [0,1]$. In terms of the transition probabilities, (3.3) can be written

$$q_t^{(n)}(0,\cdot) \overset{w}{\to} Y(t) \qquad \text{as } n \to \infty .$$

Applying Proposition 3.2, we can then immediately show that

$$\{q_t^{(n)}(a,\cdot)\}_{n\in\mathbf{N}} \text{ is tight for every } a \in \mathbf{R} \text{ and } t \in [0,1] .$$

By taking a subsequence $\{n'\} \subseteq \mathbf{N}$ if necessary, we get:

$$\{q_t^{(n)}(a,\cdot)\}_n \text{ is weakly convergent for all } a \in \mathbf{Q} \text{ and all } t \in \mathbf{Q} \cap [0,1] , \tag{3.4}$$

where $\mathbf{Q}$ denotes the rational numbers. Using exactly the same method as in the proof of Proposition 3.3, we can show that the convergence in (3.4) must hold for all $a \in \mathbf{R}$. By the Markov property and the assumption (3.2), we can even show that

$$\{q_t^{(n')}(a,\cdot)\}_{n'} \text{ is weakly convergent for all } a \in \mathbf{R} \text{ and } t \in [0,1]. \quad (3.5)$$

Let $\mathcal{C}'$ denote the class of all bounded increasing functions f such that $\sup\{(f(y)-f(x))/(y-x): x, y \in \mathbf{R}, x \neq y\}$ is finite. We can then define a linear mapping of $\mathcal{C}'$ into itself by the relation

$$H_t f(a) = \lim_{n' \to \infty} \int q_t^{(n')}(a, \mathrm{d}x) f(x) .$$

Applying the Markov property and the assumption (3.2) once again, it is a routine to prove that

$$H_{s+t} f(a) = H_s H_t f(a)$$

for all $a \in \mathbf{R}$, $f \in \mathcal{C}'$ and $s, t, s+t \in [0,1]$. Furthermore,

$$H_t f_1 \leq H_t f_2 \qquad \text{if } f_1 \leq f_2,\ f_1, f_2 \in \mathcal{C}' ,$$

$$H_t 1 \equiv 1 ,$$

$$H_t f(Y(s)) = \mathbf{E}\{f(Y(s+t)) \mid Y(s)\} \quad \text{a.s.}, \quad f \in \mathcal{C}' .$$

Until now we have only defined $H_t f(a)$ for $f \in \mathcal{C}'$. But we can immediately extend the definition to the linear space $\mathcal{D}$ consisting of all differences of functions in $\mathcal{C}'$. Then H_t becomes a positive linear operator on $\mathcal{D}$. Since $\mathcal{D}$ is dense in the space $\mathcal{C}_1$ consisting of all bounded continuous functions with limits at $+\infty$ and $-\infty$, we can extend H_t uniquely to a positive linear operator on $\mathcal{C}_1$ such that $H_t 1 \equiv 1$. By Riesz' representation theorem, there exists, for every fixed a and t, a unique probability measure $p_t(a, \cdot)$ such that

$$H_t f(a) = \int p_t(a, \mathrm{d}x) f(x) \quad (3.6)$$

for all $f \in \mathcal{C}_1$. It is a routine to prove that the $p_t(a, \cdot)$ form a family of transition probabilities generating our limit process $\{Y(t); t \in [0,1]\}$. From (3.5) and the stochastical monotonicity it follows immediately that the family $\{p_t(a, \cdot)\}_{a \in I}$ of probability measures is tight for every bounded interval I. Hence (3.6) defines a Feller semi-group, i.e., $H_t f$ is bounded and continuous for all bounded and continuous functions f. This completes the proof of Theorem 3.4.

Remark. In this section we have assumed that the approximating Markov chains all have transition probabilities satisfying condition (i) and (ii) of Proposition 3.2. These two conditions can be replaced by the weaker condition

(i′) there exists a constant K such that

$$\sup_{\substack{a,b\in\mathbf{R}\\ a\neq b}} \left\{ |b-a|^{-1} \left| \int p^{(n)}(b, \mathrm{d}x) f(x) - \int p^{(n)}(a, \mathrm{d}x) f(x) \right| \right\} \leq$$

$$\leq (1+K/n) \sup\{|f(b)-f(a)|/|b-a|: a,b\in\mathbf{R}, a\neq b\}$$

for all bounded continuous functions f and all $n\in\mathbf{N}$.

A repetition of the arguments in this section shows that Theorem 3.4 will continue to hold true.

4. Tightness conditions for sequences of stochastically monotone random Markov lines

In this section the random Markov lines $\{Y_n(t);\, t\in[0,1]\}$ will always be constructed from a sequence $\{X_{n,0}, X_{n,1}, X_{n,2}, \ldots, X_{n,n}\}_n$ of stochastically monotone Markov chains. If $\{Y_n\}_n$ is $\mathcal{D}$-convergent with limit $\{Y(t);\, t\in[0,1]\}$, we know that the one-dimensional projections $Y_n(t)$ converge weakly to $Y(t)$ with a possible exception for a countable set of time-points. It is our intention to find out to what extent we can argue in the opposite direction. Actually we shall show that, under rather general conditions, it is possible to deduce tightness of $\{Y_n\}_n$ directly from the properties of the projections $Y_n(t)$ and $Y(t)$. The main tools will be some well-known theorems on martingales. Therefore, we shall start by proving a lemma on the convergence of a sequence of martingales that might be useful even if we can not establish tightness of the corresponding sequence of random elements in D[0, 1]. The notation is the same as in the previous sections.

Lemma 4.1. *Assume that*

(i) $\sup\{E\{|Y_n(1)|\}: n\in\mathbf{N}\} = K < \infty$;

(ii) $\int x\, p^{(n)}(a, \mathrm{d}x) = a$.

Then every subsequence of $\{Y_n\}_n$ *contains a further subsequence* $\{Y_{n'}\}_{n'}$ *such that*

(a) $\{Y_{n'}(t)\}_{n'}$ *is weakly convergent for all* $t\in[0,1]$.

(b) *The limit distributions* F_t *of* $\{Y_{n'}(t)\}_{n'}$ *define a function from*

[0, 1] into the space $\mathscr{P}$ of one-dimensional probability distributions, which is continuous at all $t \in (0, 1]$, if $\mathscr{P}$ has the Lévy metric.

Proof. Conditions (i) and (ii) show that $\{X_{n,k}\}_{k=0}^{n}$ is a martingale. Thus, for all $\lambda > 0$,

$$\mathbf{P}[\sup\{|Y_n(t)|: t \in [0,1]\} > \lambda] = \mathbf{P}\{\sup\{|X_{n,k}|: k = 0, 1, 2, \ldots, n\} > \lambda\}$$

$$\leq 2\mathbf{E}\{|X_{n,n}|\}/\lambda = 2\mathbf{E}\{|Y_n(1)|\}/\lambda$$

$$\leq 2K/\lambda . \tag{4.1}$$

The family $\{F_t^{(n)}\}_{n\in\mathbf{N}, t\in[0,1]}$ of distributions corresponding to the $Y_n(t)$ is therefore tight and we can immediately find a subsequence $\{Y_{n'}\}_{n'}$ such that $\{Y_{n'}(r)\}_{n'}$ is weakly convergent to some distribution F_r for all $r \in \mathbf{Q} \cap [0, 1]$. Here $\mathbf{Q}$ denotes the set of rational numbers.

Let us now assume that we can find a time-point $t \in (0, 1]$ such that w-lim$\{F_r: r \uparrow t, r \in \mathbf{Q}\}$ does not exist. Then, for some $\delta > 0$, there is a sequence $\{r_k\}_k \subseteq \mathbf{Q}$ which is increasing to t and such that the Lévy distance $\rho(F_{r_k}, F_{r_{k+1}})$ exceeds δ for all k. For every fixed m we obtain

$$\rho(F_{r_k}^{(n')}, F_{r_{k+1}}^{(n')}) > \delta , \qquad k = 1, 2, \ldots, m , \tag{4.2}$$

for all sufficiently large $n' \in \mathbf{N}$.

Because $\{F_t^{(n)}\}_{n\in\mathbf{N}, t\in[0,1]}$ is tight, we can find a constant K_0 such that

$$F_r^{(n)}(K_0) - F_r^{(n)}(-K_0) > 1 - \delta \tag{4.3}$$

for all $r \in \mathbf{Q} \cap [0, 1]$ and $n \in \mathbf{N}$. By (4.2) and the definition of the Lévy metric we can, for $k = 1, 2, \ldots, m$ and all sufficiently large $n' \in \mathbf{N}$, choose $x_{n',k}$ so that either

$$F_{r_k}^{(n')}(x_{n',k}) > F_{r_{k+1}}^{(n')}(x_{n',k} + \delta) + \delta \tag{4.4}$$

or

$$F_{r_k}^{(n')}(x_{n',k}) < F_{r_{k+1}}^{(n')}(x_{n',k} - \delta) - \delta . \tag{4.5}$$

By (4.3), $x_{n',k} \in [-K_0, K_0]$. Let us now cover the interval $[-K_0 - \delta, K_0 + \delta]$ with finitely many intervals $I_j = [a_j, b_j]$, each one having a length between $\frac{1}{4}\delta$ and $\frac{1}{2}\delta$. Denote by $U_I^{(n)}$ and $D_I^{(n)}$, respectively, the number of up-crossings and down-crossings by the function $\{Y_n(t); t \in [0, 1]\}$ of the interval I. Then by (4.4) and (4.5) we obtain

$$E\{\Sigma_i U_i^{(n')} + \Sigma_j D_j^{(n')}\} \geq m\delta \tag{4.6}$$

for all sufficiently large n'. But (4.6) contradicts the fact that

$$E\{U_j^{(n')}\} \leq \frac{E\{(Y_{n'}(1) - a_j)^+\}}{b_j - a_j} \leq \frac{k + k_0 + \delta}{\frac{1}{2}\delta},$$

$$E\{D_j^{(n')}\} \leq \frac{E\{(Y_{n'}(1) - b_j)^+\}}{b_j - a_j} \leq \frac{k + k_0 + \delta}{\frac{1}{2}\delta}.$$

So we have proved that w-lim$\{F_r : r \uparrow t, r \in \mathbf{Q}\}$ exists for all $t \in (0, 1]$. Similarly w-lim$\{F_r : r \downarrow t, r \in \mathbf{Q}\}$ exists for all $t \in [0, 1)$. Denote these limits by G_t and H_t, respectively.

We shall now show that

$$F_t^{(n')} \xrightarrow{w} G_t \qquad \text{as } n' \to \infty \tag{4.7}$$

for all $t \in (0, 1]$. We let $\mathfrak{M}(F)$ denote the support of the distribution F. Assume that, for some $a \in \mathfrak{M}(G_t)$, there exists a strictly positive $\varepsilon = \varepsilon(a)$ such that, for arbitrarily small Δ, the inequality

$$q_\Delta^{(n')}(a, (a - \varepsilon, a + \varepsilon)^c) > \varepsilon \tag{4.8}$$

holds for some $n' \in \mathbf{N}$. By Proposition 3.3, we can choose $\delta_1 = \delta_1(\varepsilon) < \frac{1}{32}\varepsilon^2$ so that

$$q_\Delta^{(n')}(b, (b - \tfrac{1}{4}\varepsilon^2, b + \tfrac{1}{4}\varepsilon^2)^c) > \tfrac{1}{2}\varepsilon \tag{4.9}$$

for all $b \in [a - \delta_1, a + \delta_1]$. Since w-lim$\{F_r : r \uparrow t, r \in \mathbf{Q}\} = G_t$ and $a \in \mathfrak{M}(G_t)$, we can also determine $\Delta_0 = \Delta_0(\delta_1, \varepsilon, a)$ so that

$$\liminf_{n'} F_r^{n'}((a - \delta_1, a + \delta_1)) \geq F_r((a - \delta_1, a + \delta_1)) > \tfrac{1}{2} G_t((a - \delta_1, a + \delta_1)) > 0 \tag{4.10}$$

for all $r \in \mathbf{Q} \cap [t - \Delta_0, t)$. For every fixed integer m, we now choose m rational points, $r_1 < r_2 < \ldots < r_m$, in the interval $(t - \Delta_0, t)$. Keeping (4.9) and (4.10) in mind, it is not very difficult to see that from each interval $[r_k, r_{k+1}]$ we get a contribution exceeding $\frac{1}{2} G_t((a - \delta_1, a + \delta_1)) \cdot \frac{1}{2}\varepsilon$ to the value of

$$E\{U^{(n')}_{[a + \varepsilon^2/32, a + 3\varepsilon^2/32]} + D^{(n')}_{[a - 3\varepsilon^2/32, a - \varepsilon^2/32]}\},$$

provided n' is suitably chosen. Consequently

$$\mathbf{E}\{U^{(n')}_{[a+\varepsilon^2/32,\, a+3\varepsilon^2/32]} + D^{(n')}_{[a-3\varepsilon^2/32,\, a-\varepsilon^2/32]}\} >$$
$$> (m-1)\cdot\tfrac{1}{8}\varepsilon\cdot\tfrac{1}{2}G_t((a-\delta_1, a+\delta_1)) \tag{4.11}$$

for some $n' \in \mathbf{N}$. Since m is arbitrary, (4.11) contradicts the general inequalities on the expected number of up-crossings and down-crossings of a martingale. Thus we have proved that, for every $\varepsilon > 0$,

$$q^{(n')}_{\Delta}(a, (a-\varepsilon, a+\varepsilon)^c) \to 0 \quad \text{as } \Delta \to 0,$$

uniformly in $n' \in \mathbf{N}$. Because of Proposition 3.3, we have even proved that

$$q^{(n')}_{\Delta}(b, (b-\varepsilon, b+\varepsilon)^c) \to 0 \quad \text{as } \Delta \to 0, \tag{4.12}$$

uniformly in $n' \in \mathbf{N}$ and $b \in C \cap S_{\delta_2}(\mathcal{M}(G_t))$, where C is compact and δ_2 depends only on ε.

For any given $\eta > 0$ we now choose K_η so large that

$$\liminf_{n'} F_r^{(n')}((-K_\eta, K_\eta)) \geq F_r((-K_\eta, K_\eta)) > 1 - \tfrac{1}{3}\eta \tag{4.13}$$

for all $r \in \mathbf{Q} \cap [0, 1]$. By (4.12), we can choose Δ_1 so small that

$$q^{(n')}_{\Delta}(b, (b-\eta, b+\eta)^c) < \tfrac{1}{3}\eta \tag{4.14}$$

for all $\Delta < \Delta_1$, $n' \in \mathbf{N}$ and $b \in S_{\delta_3}(\mathcal{M}(G_t)) \cap [-K_\eta, K_\eta]$. Here δ_3 depends only on η. Obviously, $S_{\delta_3}(\mathcal{M}(G_t))^c$ has probability zero with respect to the distribution G_t. Since w-lim$\{F_r : r \uparrow t, r \in \mathbf{Q}\} = G_t$, we can also assume that we have taken Δ_1 so small that

$$\limsup_{n' \to \infty} F^{(n')}_{t-\Delta}(S_{\delta_3}(\mathcal{M}(G_t))^c) \leq F_{t-\Delta}(\overline{S_{\delta_3}(\mathcal{M}(G_t))^c}) < \tfrac{1}{3}\eta \tag{4.15}$$

for all $t-\Delta \in \mathbf{Q} \cap (t-\Delta_1, t)$. It is then easy to show that (4.13)–(4.15) imply that

$$\limsup_{n' \to \infty} \mathbf{P}[|Y_{n'}(t) - Y_{n'}(t-\Delta)| \geq \eta] \leq \eta \tag{4.16}$$

for all $t-\Delta \in \mathbf{Q} \cap (t-\Delta_1, t)$. Finally, we get

$$\limsup_{n'\to\infty} \rho(F_t^{(n')}, F_{t-\Delta}) \le \limsup_{n'\to\infty} \rho(F_t^{(n')}, F_{t-\Delta}^{(n')}) + \limsup_{n'\to\infty} \rho(F_{t-\Delta}^{(n')}, F_{t-\Delta})$$
$$\le \eta \tag{4.17}$$

for all $t-\Delta \in \mathbf{Q} \cap (t-\Delta_1, t)$, which implies that

$$w\text{-}\lim_{n'\to\infty} F_t^{(n')} = G_t\,, \qquad t \in (0,1]\,.$$

Similar arguments show that

$$w\text{-}\lim_{n'\to\infty} F_t^{(n')} = H_t\,, \qquad t \in (0,1)\,,$$

and this completes the proof of Lemma 4.1.

Remark. In the proof of Lemma 4.1 we have used theorems on the number of up-crossings and down-crossings of a martingale. Similar theorems hold for supermartingales and submartingales. Under conditions (i) and (ii) of Theorem 4.2 below, there is a constant c such that $\{X_{n,k} - c \cdot k/n\}_{k=0}^{n}$ is a supermartingale (submartingale) for all $n \in \mathbf{N}$. Condition (iii) below is then sufficient for the lemma to remain true also in that case.

For sequences of stochastically monotone random Markov lines we can now prove

Theorem 4.2. *Assume that*

(i) $\sup\{\mathrm{E}\,|Y_n(1)|\colon n \in \mathbf{N}\} = K < \infty$;

(ii) $\sup\{(1/n)^{-1} \int (x-a)\, p^{(n)}(a; \mathrm{d}x)\colon a \in \mathbf{R};\, n \in \mathbf{N}\} < +\infty$ *or* $\inf\{(1/n)^{-1} \int (x-a)\, p^{(n)}(a; \mathrm{d}x)\colon a \in \mathbf{R};\, n \in \mathbf{N}\} > -\infty$;

(iii) *there exists a constant K_1 such that, for all $n \in \mathbf{N}$,*

$$\sup_{a,b \in \mathbf{R}} (b-a)^{-1} \left| \int x\, p^{(n)}(b; \mathrm{d}x) - \int x\, p^{(n)}(a; \mathrm{d}x) \right| \le 1 + K_1/n\,;$$

(iv) *$Y_n(1)$ converges weakly to some probability distribution with strictly increasing distribution function.*

Then the sequence $\{Y_n\}_n$ of random elements in $\mathrm{D}[0,1]$ is weakly conditionally compact.

Proof. Let $\{Y_{n'}\}_{n'}$ be an arbitrary subsequence of $\{Y_n\}_n$. We shall prove that $\{Y_{n'}\}_{n'}$ has a convergent subsequence $\{Y_{n''}\}_{n''}$. By the remark following Lemma 4.1, we can choose a subsequence $\{Y_{n''}\}_{n''}$ such that $Y_{n''}(t) \overset{w}{\to} F_t$ for all $t \in [0,1]$ and $w\text{-}\lim\{F_t : t \uparrow 1\} = F_1$. The assumption (iv) shows that $\mathcal{M}(F_1) = \mathbf{R}$. Repeating the arguments leading to (4.12) in the proof of Lemma 4.1, we can show that

$$q_\Delta^{(n'')}(a,(a-\epsilon,a+\epsilon)^c) \to 0 \qquad \text{as } \Delta \to 0\,,$$

uniformly in $n'' \in \mathbf{N}$ and $a \in C$, for every compact set C and every $\epsilon > 0$. Supermartingale (submartingale) inequalities show that

$$\sup_{n\in\mathbf{N}} \mathbf{P}[\sup\{|Y_n(t)| : t \in [0,1]\} > \lambda] \to 0 \qquad \text{as } \lambda \to \infty\,.$$

Thus, by Theorem 2.1, $\{Y_{n''}\}_{n''}$ is weakly conditionally compact and $\{Y_{n'}\}_{n'}$ must contain a convergent subsequence.

Application. Let $\{Z_n\}_n$ be a Galton–Watson branching process governed by certain fixed probabilities $\{p_k\}_k$. Here p_k is the probability that one individual in the j^{th} generation gives rise to k individuals in the $j+1^{\text{st}}$ generation. We will assume that $\{p_k\}$ considered as a probability distribution has mean 1 and finite strictly positive variance σ^2. Define, for each $n \in \mathbf{N}$, a continuous-time process

$$Y_n(t) = (Z_{[nt]} - a_n)/b_n \qquad \text{for } t \in [0,1]\,,$$

where $Z_0 = c_n$ is the number of individuals at time $t = 0$. $b_n > 0$ and $a_n \in \mathbf{R}$ are normalizing constants such that $a_n/b_n \to \infty$ as $n \to \infty$.

Let us now assume that there exists a stochastic process $\{X(t); t \in [0,1]\}$ such that $X(t)$ is non-degenerate for $t \in (0,1]$ and the finite-dimensional distributions of $\{Y_n(t); t \in [0,1]\}$ converge to those of $\{X(t); t \in [0,1]\}$. Then $\{X(t); t \in [0,1]\}$ is equivalent to the process $\{aB(t) + b; t \in [0,1]\}$, where $a > 0$ and $b \in \mathbf{R}$ are constants and $\{B(t); t \in [0,1]\}$ is a Brownian motion (see [1], pp. 228 and 239]). Still following Lamperti, we can also see that $(n c_n/b_n^2)_{n\in\mathbf{N}}$ is bounded and so

$$\sup_{n\in\mathbf{N}} \{\mathrm{Var}[Y_n(1)]\} = \sup_{n\in\mathbf{N}} \{n c_n \sigma^2/b_n^2\} < \infty\,.$$

Since $\mathbf{E}[Y_n(1)] = (c_n - a_n)/b_n \to b$ and $\int x\, p^{(n)}(a; dx) = a$, $\{Y_n\}_n$ is conditionally compact by Theorem 4.2. Thus, we get

$$Y_n \overset{\mathcal{D}}{\to} aB + b \qquad \text{as } n \to \infty\,. \tag{4.18}$$

This means that, for all functionals h on D[0, 1] which are continuous with respect to the Skorokhod topology, we have shown that

$$h(Y_n) \stackrel{w}{\to} h(aB+b) \qquad \text{as } n \to \infty .$$

The $\mathcal{D}$-convergence in (4.18) has been proved by other methods in [13]. In Section 5 we shall give a more detailed discussion of $\mathcal{D}$-convergence of sequences of branching processes.

Proceeding in the same spirit as in Lemma 4.1 and Theorem 4.2, we get:

Theorem 4.3. *Let $\{Y_n\}_n$ be a sequence of random Markov lines corresponding to a sequence of stochastically monotone Markov chains. Assume that*

(i) $\sup\{E\{|Y_n(1)|\}: n \in \mathbf{N}\} < \infty$;
(ii) $\sup\{(1/n)^{-1} \int (x-a)\, p^{(n)}(a, dx): a \in \mathbf{R}, n \in \mathbf{N}\} < +\infty$, *or* $\inf\{(1/n)^{-1} \int (x-a)\, p^{(n)}(a, dx): a \in \mathbf{R}, n \in \mathbf{N}\} > -\infty$;
(iii) *for every $t \in (0, 1]$, the projection $Y_n(t)$ converges weakly to some distribution F_t with strictly increasing distribution function.*

Then the sequence $\{Y_n\}_n$ of random elements in D[0, 1] *is weakly conditionally compact and every limit corresponds to a stochastically continuous process on the interval* [0, 1].

Proof. Using exactly the same method as in Lemma 4.1, we can immediately see that $\lim\{F_r: r \uparrow t, r \in \mathbf{Q}\}$ exists for all $t \in (0, 1]$. Denote this limit by G_t. We shall show that $G_t = F_t$ for at least one time-point t. Assume that the converse holds. Since the Lévy distance $\rho(F_t, G_t) > 0$ for uncountably many values of t, we can find a $\delta > 0$ such that $\rho(F_t, G_t) > \delta$ for infinitely many t. This means that, for any positive integer m, we can find $2m$ points, $r_1 < t_1 < r_2 < t_2 < \ldots < r_m < t_m$, such that

$$\rho(F_{r_k}, F_{t_k}) > \delta \,, \qquad k = 1, 2, \ldots, m \,.$$

As in Lemma 4.1, this contradicts the inequalities on the expected number of up-crossings and down-crossings of an interval. Thus, we can choose a point t_0 such that

$$\text{w-lim}\{F_r: r \uparrow t_0, r \in \mathbf{Q}\} = F_{t_0} \,,$$

where $\mathcal{M}(F_{t_0}) = \mathbf{R}$. Once again proceeding as in the proof of Lemma 4.1, we can show that, for every $\epsilon > 0$ and $a \in \mathbf{R}$,

$$q_\Delta^{(n)}(a,(a-\epsilon, a+\epsilon)^c) \to 0 \qquad \text{as } \Delta \to 0 \tag{4.19}$$

uniformly in $n \in \mathbf{N}$. Because the underlying Markov chains $\{X_{n,k}\}_{k=0}^n$ are stochastically monotone, the convergence in (4.19) must be uniform in $a \in C$ for all compact sets C. Applying Theorem 2.1 we can then complete the proof of Theorem 4.3.

We shall terminate this section by discussing the convergence of sequences of increasing and stochastically monotone Markov chains. In this case the conditions for conditional compactness become much simpler. In an important special case we can also show that every limit process $\{Y(t); t \in [0,1]\}$ is a Markov process if, for all $t \in (0,1]$, $Y(t)$ has a strictly positive density on $(0,\infty)$. These results are summarized in the following theorem.

Theorem 4.4. *Let $\{Y_n(t); t \in [0,1]\}_n$ be the random Markov lines corresponding to a sequence of increasing and stochastically monotone Markov chains. Then the sequence $\{Y_n\}_n$ is weakly conditionally compact provided*

(i) *for each $t \in (0,1]$, $Y_n(t)$ converges weakly to some distribution F_t with strictly positive density on $(0,\infty)$;*

(ii) *for every $\delta > 0$ and every compact set $C \subseteq [0,\infty)$,*

$$n\, p^{(n)}(a,(a-\delta, a+\delta)^c) \to 0 \qquad \text{as } n \to \infty\,,$$

uniformly in $a \in C$.

Moreover, every limit process is a Markov process and has continuous sample paths with probability 1.

Proof. The sequence $\{Y_n(1)\}_n$ is stochastically bounded. Therefore,

$$\sup_{n\in\mathbf{N}} \mathbf{P}[\sup_{t\in[0,1]} |Y_n(t)| > \lambda] = \sup_{n\in\mathbf{N}} \mathbf{P}[|Y_n(1)| > \lambda] \to 0 \qquad \text{as } \lambda \to \infty\,.$$

Obviously, $\{Y_n(t); t \in [0,1]\}$ has no down-crossings and at most one up-crossing of each interval. Thus, we can prove that $\{Y_n\}_n$ is conditionally compact exactly as in Theorem 4.3. Condition (ii) implies that every limit process has continuous sample paths with probability 1.

360 A. Grimvall, On transition from Markov chain to continuous time process

It remains to prove that every limit is a Markov process. According to the proof of Theorem 3.4, it is enough to show that, for any given $b \in \mathbb{R}$, $\Delta \in (0,1)$ and $\epsilon > 0$, there exist $\delta > 0$ and $n_0 \in \mathbb{N}$ such that

$$\rho(q_\Delta^{(n)}(b_1,\cdot),\, q_\Delta^{(n)}(b_2,\cdot)) \leq \epsilon \tag{4.20}$$

for all $b_1, b_2 \in [b-\delta, b+\delta]$ and all $n \geq n_0$. We shall assume the converse and derive a contradiction.

If 4.20 does not hold, there exist $b \in \mathbb{R}$, $\Delta \in (0,1)$ and $\epsilon > 0$ such that, for every neighbourhood U_b of b and every $n_0 \in \mathbb{N}$,

$$q_\Delta^{(n)}(b_1,(c-\epsilon,\infty)) + \epsilon < q_\Delta^{(n)}(b_2,(c,\infty)) \tag{4.21}$$

for some $b_1, b_2 \in U_b$, $n \geq n_0$ and $c = c(n,\Delta,\epsilon,b_1,b_2)$. Recalling that the underlying Markov chains are stochastically monotone we can show that it is no restriction to assume that $c \leq c_0$, where $c_0 = c_0(b,\Delta,\epsilon)$ is a constant depending only on b, Δ and ϵ. Using exactly the same method as in the proof of Lemma 4.1 we can choose $\Delta_1 \in (0,1)$ so small that

$$q_{\Delta_1}^{(n)}(a,[a,a+\tfrac{1}{2}\epsilon)^{\mathrm{c}}) < \tfrac{1}{2}\epsilon \tag{4.22}$$

for all $n \in \mathbb{N}$ and $a \in [0,c_0]$. By (4.21), (4.22) and Chapman–Kolmogorov's equation we then get

$$\begin{aligned} q_\Delta^{(n)}(b_1,(c-\epsilon,\infty)) + \tfrac{1}{2}\epsilon &< q_\Delta^{(n)}(b_2,(c,\infty)) - \tfrac{1}{2}\epsilon \\ &= \int q_{\Delta-\Delta_1}^{(n)}(b_2,\mathrm{d}y)\, q_{\Delta_1}^{(n)}(y,(c,\infty)) - \tfrac{1}{2}\epsilon \\ &\leq q_{\Delta-\Delta_1}^{(n)}(b_2,(c-\tfrac{1}{2}\epsilon,\infty)) + q_{\Delta_1}^{(n)}(c-\tfrac{1}{2}\epsilon,(c,\infty)) - \tfrac{1}{2}\epsilon \\ &< q_{\Delta-\Delta_1}^{(n)}(b_2,(c-\tfrac{1}{2}\epsilon,\infty))\,. \end{aligned} \tag{4.23}$$

But

$$\begin{aligned} &q_{\Delta-\Delta_1}^{(n)}(b_2,(c-\tfrac{1}{2}\epsilon,\infty)) - q_\Delta^{(n)}(b_1.(c-\epsilon,\infty)) = \\ &\quad = \int_{x \geq b_1} q_{\Delta_1}^{(n)}(b_1,\mathrm{d}x)\,[q_{\Delta-\Delta_1}^{(n)}(b_2,(c-\tfrac{1}{2}\epsilon,\infty)) - q_{\Delta-\Delta_1}^{(n)}(x,(c-\epsilon,\infty))] \\ &\quad \leq q_{\Delta_1}^{(n)}(b_1,[b_1,b_2))\,. \end{aligned}$$

Thus, by (4.23), we can find b_1 and b_2 arbitrarily close to b and an arbitrarily large $n \in \mathbf{N}$ such that

$$q_{\Delta_1}^{(n)}(b_1, [b_1, b_2)) > \tfrac{1}{2}\epsilon . \tag{4.24}$$

This means that there exist integers $m \geq 1$ with the following property:

P(m) For every $n_0 \in \mathbf{N}$ and every neighbourhood U_b of b we can find at least m disjoint intervals $[b_i, c_i] \subseteq U_b$, $i = 1, 2, \ldots, m$, such that, for some $n \geq n_0$,

$$q_{\Delta_1}^{(n)}(b_i, [b_i, c_i)) > \tfrac{1}{2}\epsilon \tag{4.25}$$

holds simultaneously for all $i \in \{1, 2, \ldots, m\}$.

Let, for all $a \in \mathbf{R}$, $\tau_a^{(n)}$ denote the hitting-time of $\{Y_n(t): t \in [0,1]\}$ for the interval $[a, \infty)$ and set $\tau_a^{(n)} = 2$ if $Y_n(t)$ is less than a for all $t \in [0, 1]$. Let us then consider $\tau_{c_1}^{(n)}, \tau_{c_2}^{(n)}, \ldots, \tau_{c_m}^{(n)}$. It is no restriction to assume that $b_1 < c_1 < b_2 < c_2 < \ldots < b_m < c_m$. Furthermore, by the assumption (ii),

$$\mathbf{P}[\tau_{c_i}^{(n)} \leq 1,\ i = 1, 2, \ldots, m] \cap \{\tau_{c_i}^{(n)} > b_{i+1} \text{ for some } i \leq m-1\} \to 0$$
$$\text{as } n \to \infty . \tag{4.26}$$

Using the (strong) Markov property, (4.25) and (4.26), we can then see that

$$\mathbf{P}[\tau_{b+1}^{(n)} \leq 1] = \mathbf{P}[\sup\{|Y_n(t)|: t \in [0,1]\} \geq b+1]$$

could be made arbitrarily small if m could be chosen arbitrarily large. But this contradicts the assumption (i). Thus there exists an integer $m_0 \geq 1$ such that $\mathbf{P}\{m_0\}$ holds true and $\mathbf{P}\{m\}$ does not hold for any $m > m_0$.

On the other hand, if there exists such a maximal m_0, we can find a fixed $\delta > 0$ and, for any $n_0 \in \mathbf{N}$ and any sphere $S_\gamma(b)$, m_0 disjoint intervals $[b_i, c_i] \subseteq S_\gamma(b)$, $i = 1, 2, \ldots, m_0$, such that the inequalities

$$q_{\Delta_1}^{(n)}(b_i, [b_i, c_i)) > \tfrac{1}{2}\epsilon , \qquad i = 1, 2, \ldots, m_0 , \tag{4.27}$$

$$q_{\Delta_1}^{(n)}(b - \delta, [b - \delta, b - 2\gamma)) \leq \tfrac{1}{2}\epsilon \tag{4.28}$$

hold simultaneously for some $n \geq n_0$. (Obviously it is no restriction to assume that $2\gamma < \frac{1}{2}\delta$.) From (4.27), it follows that

$$q_{\Delta}^{(n)}(a, [a, b+\gamma)) > \tfrac{1}{2}\epsilon \tag{4.29}$$

for all $\Delta \leq \Delta_1$ and all $a \leq b-\gamma$. Using the Markov property, (4.28) and (4.29), we get

$$\begin{aligned}
&P[Y_n(1-\Delta) \in (b-\delta; b-2\gamma);\\
&\quad Y_n(t) \in [b-2\gamma; b-\gamma) \text{ for some } t \in [1-\Delta; 1];\\
&\quad b-2\gamma \leq Y_n(1) \leq b+\gamma] \geq\\
&\geq \tfrac{1}{2}\epsilon\, P[Y_n(1-\Delta) \in (b-\delta; b-2\gamma);\\
&\qquad Y_n(t) \in [b-2\gamma; b-\gamma) \text{ for some } t \in [1-\Delta; 1]]\\
&\geq \tfrac{1}{2}\epsilon\, (P[Y_n(1-\Delta) \in (b-\delta; b-2\gamma); Y_n(1) \geq b-2\gamma]\\
&\qquad - P[Y_n(t) \text{ has a jump exceeding } \gamma])\\
&\geq \tfrac{1}{2}\epsilon\, ((1-\tfrac{1}{2}\epsilon) P[Y_n(1-\Delta) \in (b-\delta; b-2\gamma)]\\
&\qquad - P[Y_n(t) \text{ has a jump exceeding } \gamma]) .
\end{aligned}$$

Hence, for all $\gamma > 0$:

$$\begin{aligned}
&\limsup_{n\to\infty} P[Y_n(1) \in [b-2\gamma; b+\gamma)] \geq\\
&\quad \geq \tfrac{1}{2}\epsilon(1-\tfrac{1}{2}\epsilon) \liminf_{n\to\infty} P[Y_n(1-\Delta) \in (b-\delta; b-\tfrac{1}{2}\delta)] ;
\end{aligned}$$

which obviously contradicts the assumption (i). This completes the proof of Theorem 4.4.

5. Weak convergence of normalized Galton–Watson processes

In Section 4, we showed how our general results on the transition from a sequence of stochastically monotone Markov chains to a continuous-time process could be applied to sequences of normalized Galton–Watson processes. Here we will give a more detailed discussion of this topic.

For each fixed $n \in \mathbb{N}$, let $\{Z_j^{(n)}\}_j$ denote the random variables of a

Galton–Watson branching process governed by the probabilities $\{p_k^{(n)}\}_k$. Here $p_k^{(n)}$ denotes the probability that one individual in the j^{th} generation of the n^{th} branching process gives rise to k individuals in the $(j+1)^{\text{st}}$ generation of the same process. Let us then introduce the continuous-time processes

$$Y_n(t) = (Z_{[nt]}^{(n)} - a_n)/b_n\,; \qquad t \in [0,1]\,. \tag{5.1}$$

where $Z_0^{(n)} = c_n$. We shall always assume that c_n are positive integers $\to \infty$, while $b_n > 0$ and $a_n \in \mathbb{R}$ are normalizing constants. We shall also assume that there exist stochastic processes $\{Y(t);\, t \in [0,1]\}$ such that

$$(Y_n(t_1);\, Y_n(t_2);\, \ldots;\, Y_n(t_k)) \xrightarrow{w} (Y(t_1);\, Y(t_2);\, \ldots;\, Y(t_k)) \qquad \text{as } n \to \infty; \tag{5.2}$$

for all $0 \leq t_1 \leq t_2 \leq \ldots \leq t_k \leq 1$. Lamperti has then proved the following two theorems (see [12]):

Theorem 5.1. *Suppose that (5.2) holds with $a_n \equiv 0$, $c_n \to \infty$ and $\mathbb{P}[Y_t = 0] < 1$ for some $t \in (0,1]$. Then $\{Y(t);\, t \in [0,1]\}$ is a continuous state branching process, i.e., a Markov process with transition probabilities $p_t(a, E)$ satisfying:*

(i) *for fixed $t \in [0,1]$ and $x \geq 0$, $p_t(x, \cdot)$ is a probability measure on the class of Borel sets in $[0,\infty)$;*

(ii) *for every Borel set $E \subseteq [0,\infty)$, $p_t(x, E)$ is jointly measurable in $t \in [0,1]$ and $x \geq 0$;*

(iii) $\int_0^\infty p_t(x, du)\, p_s(u, E) = p_{t+s}(x, E)$;

(iv) *for all $x, y \geq 0$ and $t \in [0,1]$,*

$$p_t(x+y, \cdot) = p_t(x, \cdot) * p_t(y, \cdot)\,,$$

where $*$ denotes convolution;

(v) *there exist $t \in (0,1]$ and $x > 0$ such that $p_t(x, \{0\}) < 1$.*

Theorem 5.2. *Suppose again that (5.2) holds, where we now assume that $a_n/b_n \to \infty$. Then $\{Y(t);\, t \in [0,1]\}$ is a process with stationary independent increments. If $b_n \to \infty$, the canonical measure governing the distribution of the increments of $\{Y(t);\, t \in [0,1]\}$ has support contained in $[0,\infty)$. If $b_n \not\to \infty$, the canonical measure is supported on the set $\{n/b;\, n = -1, 0, 1, 2, \ldots\}$ for some positive b.*

We shall now examine the convergence in these two theorems and prove that convergence of the finite-dimensional distributions implies weak convergence in D[0, 1], provided the probability distribution $\{p_k^{(n)}\}_k$ has mean value 1 for all $n \in \mathrm{N}$. As we can see from the proof, it is easy to generalize this to the case when the mean value is of the form $1 + \alpha/n + o(1/n)$.

Let us start by examining the convergence in Theorem 5.1. Since $\{c_n/b_n\}_n$ evidently must converge and $\mathrm{E}\{|Y_n(1)|\} = \mathrm{E}\{|Y_n(0)|\} = c_n/b_n$, there exists a constant K such that

$$\mathrm{E}\{|Y_n(0)|\} + \mathrm{E}\{|Y_n(1)|\} \leq K \tag{5.3}$$

for all $n \in \mathrm{N}$. However, each $\{Y_n(t); t \in \{k/n: k = 0, 1, 2, ..., n\}\}$ is a martingale. Hence,

$$\mathrm{P}[\sup\{|Y_n(t)|: t \in [0,1]\} > \lambda] \leq 2K/\lambda\,, \tag{5.4}$$

for all $\lambda > 0$ and $n \in \mathrm{N}$, and the first condition in Theorem 2.1 is satisfied.

Before we start examining the transition probabilities $q_\Delta^{(n)}(a, E)$, let us consider the following:

Proposition 5.3. *Let $m \in \mathrm{N}$ and $\epsilon > 0$ be given numbers. Then we can find a $\delta = \delta(m, \epsilon) > 0$ such that, for any $k \in \mathrm{N}$ and any set $\{Y_1, Y_2, ..., Y_{km}\}$ of identically distributed independent random variables, it holds that*

$$\mathrm{P}[|\Sigma_{i=1}^k Y_i| > \delta] < \delta \Rightarrow \mathrm{P}[|\Sigma_{i=1}^j Y_i| > \epsilon] < \epsilon \quad \textit{for all } j \leq km\,.$$

Proof. We use the same kind of arguments as in the beginning of the proof of Theorem 2.2. Then we know that, for every t_0 and all η small enough,

$$|\varphi^k(t) - 1| \leq \tfrac{1}{4}\eta \qquad \text{for } |t| \leq t_0$$

implies that

$$|\varphi(t) - 1| \leq \eta/k \qquad \text{for } |t| \leq t_0\,.$$

And the last inequality implies that

$$|\varphi^j(t) - 1| \leq m\eta \qquad \text{for } |t| \leq t_0 \text{ and } j \leq km\,.$$

The rest of the proof is obvious.

We shall now prove that the transition probabilities $q_\Delta^{(n)}(a, \cdot)$ satisfy condition (ii) in Theorem 2.1. By assumption, there is a point $t_0 > 0$ such that

$$\mathbf{R}[Y(t_0) > 0] > 0 \,. \tag{5.5}$$

Proceeding as in the proof of Lemma 4.1, we get

$$w\text{-}\lim_{t \uparrow t_0} Y(t) = Y(t_0) \,. \tag{5.6}$$

From (5.5), (5.6) and the proof of Lemma 4.1, it follows that there exists a strictly positive real number b such that

$$q_\Delta^{(n)}(b, (b-\delta, b+\delta)^c) \to 0 \qquad \text{as } \Delta \to 0 \,, \tag{5.7}$$

uniformly in $n \in \mathbf{N}$, for every $\delta > 0$.

Since $\{Z_j^{(n)}\}_j$ is an integer-valued process, we can set

$$q_\Delta^{(n)}(a, \cdot) = q_\Delta^{(n)}(k/b_n, \cdot) \qquad \text{for } (k-1)/b_n < a \leq k/b_n \,.$$

But, for each Δ and n, we can find a set $\{\eta_i;\ i = 1, 2, 3, \ldots\}$ of independent and identically distributed random variables such that

$$q_\Delta^{(n)}(k/b_n, \mathrm{d}x) = \mathbf{P}[(\eta_1 + \eta_2 + \ldots + \eta_k)/b_n \in \mathrm{d}x]$$

for all $k \in \mathbf{N}$. By Proposition 5.3 and (5.7), we then get

$$q_\Delta^{(n)}(a, (a-\delta, a+\delta)^c) \to 0 \qquad \text{as } \Delta \to 0 \,,$$

uniformly in $a \in C$ and $n \in \mathbf{N}$, for every compact set $C \subseteq [0, \infty)$. Hence, by Theorem 2.1 and Prokhorov's theorem, $\{Y_n\}_n$ is conditionally compact in $D[0, 1]$. Let Y' and Y'' denote any two limit distributions of $\{Y_n\}_n$. We can then select subsequences $\{n'\}$ and $\{n''\} \in \mathbf{N}$ such that

$$(Y_{n'}(t_1), Y_{n'}(t_2), \ldots, Y_{n'}(t_k)) \xrightarrow{w} (Y'(t_1), Y'(t_2), \ldots, Y'(t_k)) \qquad \text{as } n' \to \infty \,;$$

$$(Y_{n''}(t_1), Y_{n''}(t_2), \ldots, Y_{n''}(t_k)) \xrightarrow{w} (Y''(t_1), Y''(t_2), \ldots, Y''(t_k)) \qquad \text{as } n'' \to \infty \,;$$

for all $t_1, t_2, \ldots, t_k \in T$, where $[0, 1] \setminus T$ is a countable set not containing 1 (see [1, p. 124]). By 5.2, the distributions of $(Y'(t_1), Y'(t_2), \ldots, Y'(t_k))$

and $(Y''(t_1), Y''(t_2), ..., Y''(t_k))$ coincide for all $t_1, t_2, ..., t_k \in T$. Applying [1, Theorem 14.5], we conclude that Y' and Y'' define the same probability distribution in D[0, 1]. But a conditionally compact sequence with only one limit point must be convergent. Thus,

$$Y_n \overset{\mathcal{D}}{\to} Y' \qquad \text{as } n \to \infty .$$

Since $\{Y'(t);\, t \in [0,1]\}$ is continuous in probability by Theorem 3.4,

$$(Y_n(t_1), Y_n(t_2), ..., Y_n(t_k)) \overset{w}{\to} (Y'(t_1), Y'(t_2), ..., Y'(t_k)) \qquad \text{as } n \to \infty ,$$

for all $0 \leq t_1 \leq t_2 \leq ... \leq t_k \leq 1$. But then the finite-dimensional distributions of Y' coincide with those of Y in (5.2), and this completes the discussion of Theorem 5.1.

We shall now turn to the discussion of the convergence in Theorem 5.2. We start by giving a result from the theory of triangular arrays.

Proposition 5.4. *Let* $\{Z_{n,1}, Z_{n,2}, ..., Z_{n,n}\}_n$ *be a triangular array of random variables such that*

(i) *for each* n *the variables* $Z_{n,1}, Z_{n,2}, ..., Z_{n,n}$ *are identically distributed and independent*;

(ii) $Z_{n,j} \geq -1,\ j = 1, 2, ..., n,\ n = 1, 2, 3, ...$;

(iii) $\mathrm{E}\{Z_{n,j}\} = 0,\ j = 1, 2, ..., n,\ n = 1, 2, 3, ...$

With $S_n = \sum_{j=1}^{n} Z_{n,j}$ *we then have* $\sup_{n\in\mathbb{N}} \mathrm{E}\{|S_n|\} < \infty$, *provided* $\{S_n\}_n$ *is stochastically bounded.*

Proof. Following [6, p. 308], we introduce the continuous truncation function

$$\tau_s(x) = \begin{cases} x & \text{for } |x| \leq s , \\ \pm s & \text{for } |x| \geq s . \end{cases}$$

Define $Z'_{n,j} = \tau_s(Z_{n,j})$ and $Z''_{n,j} = Z_{n,j} - Z'_{n,j}$. For $s \geq 1$, $Z''_{n,j}$ is non-negative. Moreover, for s sufficiently large, both $S'_n = \sum_{j=1}^{n} Z'_{n,j}$ and $S''_n = \sum_{j=1}^{n} Z''_{n,j}$ are stochastically bounded. Still following [6], we conclude that $\{\mathrm{E}\{(S'_n)^2\}\}_n$ is bounded and so $\{\mathrm{E}\{(S'_n)^-\}\}_n$ is bounded too. Since S''_n is non-negative and $E(S_n) = 0$, we obviously have

$$\sup_{n\in\mathbf{N}} \{\mathrm{E}\{|S_n|\}\}_n < +\infty \tag{5.8}$$

and the proposition is proved. There is no difficulty to generalize this result to triangular arrays $\{Z_{n,1}, Z_{n,2}, \ldots, Z_{n,k_n}\}_n$, where $\{k_n\}_n$ is an arbitrary sequence tending to infinity.

There is no loss of generality in assuming that $a_n = c_n$ in Theorem 5.2. If we exclude the case when the limit process is degenerate, we can also assume that $\inf\{b_n : n \in \mathbf{N}\} > 0$. Let us now consider the probability law of $Y_n(1)$. It coincides with the law of $\Sigma_{j=1}^{c_n}(X_j - 1)/b_n$, where the X_j are independent random variables representing the number of individuals in the n^{th} generation of $\{Z_j^{(n)}\}_j$ who are descended from each of the c_n original ancestors. Proposition 5.4, with $Z_{n,j} = (X_j - 1)/b_n$ shows that

$$\sup_{n\in\mathbf{N}} \mathrm{E}\{|Y_n(1)|\} < \infty .$$

Noticing that $\{Y_n(t); t \in \{k/n : k = 0, 1, 2, \ldots, n\}\}$ is a martingale, we have shown that the sequence $\{Y_n\}_n$ satisfies condition (i) of Theorem 2.1. Proceeding as in the discussion of the convergence in Theorem 5.1, we can prove that $\{Y_n\}_n$ is conditionally compact. Thus, we have the following.

Theorem 5.5. *Let $\{Z_j^{(n)}\}_j$, for each $n \in \mathbf{N}$, denote the random variables of a critical Galton–Watson branching process and define by* 5.1 *a sequence of continuous-time processes $\{Y_n(t); t \in [0,1]\}_n$. Assume that there is a non-degenerate stochastic process $\{Y(t); t \in [0,1]\}$ such that the finite-dimensional distributions of $\{Y_n(t); t \in [0,1]\}$ converge to those of $\{Y(t); t \in [0,1]\}$. Then there is a random element $\{Y'(t); t \in [0,1]\}$ in* D[0,1] *with the same finite-dimensional distributions as $\{Y(t); t \in [0,1]\}$ and such that*

$$Y_n \overset{\mathcal{D}}{\to} Y' \qquad \text{as } n \to \infty .$$

Remark. In this paper we have only considered Markov branching processes. In a coming paper, corresponding tightness and convergence results for age-dependent branching processes will be given.

References

[1] P. Billingsley, Convergence of probability measures (Wiley, New York, 1968).

[2] A.A. Borovkov, Convergence of weakly dependent processes to the Wiener process, Theor. Probability Appl. 12 (1967) 159–186.

[3] A.A. Borovkov, On the convergence to diffusion processes, Theor. Probability Appl. 12 (1967) 405–431.

[4] A.A. Borovkov, Theorems on the convergence to Markov diffusion processes, Z. Wahrscheinlichkeitstheorie verw. Geb. 16 (1970) 47–76.

[5] D.J. Daley, Stochastically monotone Markov chains, Z. Wahrscheinlichkeitstheorie verw. Geb. 10 (1968) 305–317.

[6] W. Feller, An introduction to probability theory and its applications, Vol. II (Wiley, New York, 1971).

[7] I.I. Gihkman, On the convergence to Markov processes, Ukr. Mat. Zh. 21 (1969) 316–324.

[8] I.I. Gihkman and A.V. Skorokhod, Introduction to the theory of random processes (Saunders, Philadelphia, 1969).

[9] A. Grimvall, A theorem on convergence to a Lévy process, Math. Scand. 30 (1972) 339–349.

[10] J. Lamperti, Probability (Benjamin, New York, 1966).

[11] J. Lamperti, Limiting distributions for branching processes, in: Proc. 5th Berkeley Symp. on Mathematical Statistics and Probability, Vol. 2 (2) (1967) 225–241.

[12] J. Lamperti, The limit of a sequence of branching processes, Z. Wahrscheinlichkeitstheorie verw. Geb. 7 (1967) 271–288.

[13] T. Lindvall, Convergence of critical Galton–Watson branching processes, J. Appl. Prob. 9 (1972) 445–450.

[14] Yu.V. Prokhorov, Convergence of random processes and limit theorems in probability theory, Theor. Probability Appl. 1 (1956) 157–214.

[15] A.V. Skorokhod, Studies in the theory of random processes (Addison-Wesley, Reading, Mass., 1965).

[16] A.V. Skorokhod, Limit theorems for stochastic processes, Theor. Probability Appl. 1 (1956) 261–290.

[17] A.V. Skorokhod, Limit theorems for stochastic processes with independent increments, Theor. Probability Appl. 2 (1957) 138–171.

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

[19] A.V. Skorokhod, A limit theorem for homogeneous Markov chains, Theor. Probability Appl. 8 (1963) 61–70.