180

COLLOQUIA MATHEMATICA SOCIETATIS JÁNOS BOLYAI
27. RANDOM FIELDS, ESZTERGOM (HUNGARY), 1979.

STATIONARY MARKOV CHAINS

S. GOLDSTEIN* — J.L. LEBOWITZ** — E. PRESUTTI**

0. PREFACE

We discuss some of the aspects of stationary Markov chains which we have found to be useful in some physical problems. We review and somewhat extend those properties connected with the asymptotic behavior of the chain and the ergodic properties of the associated dynamical system.

We place special emphasis on the relationship between the deterministic and the stochastic aspect of the motion. In particular we emphasize stationary Markov chains which arise from a certain kind of stochastic perturbation of a deterministic motion, namely the (B, μ, T, π) scheme, which will be discussed in Section 1. There we also describe in greater detail the problems we will be concerned with.

* Supported in part by NSF Grant No. PHY 78-03816.
**Supported in part by NSF Grant No. PHY 78-15920.

1. INTRODUCTION

Given an abstract dynamical system [1] (B, μ, T) and a measurable partition [2] π of B, we define a Markov chain (B, μ, T, π) as follows: a point $x \in B$ moves to Tx and then is randomly scattered with "uniform density" with respect to μ in the fiber $\pi(Tx)$ of π in which it is situated. The transition probability of the Markov chain (B, μ, T, π) is thus

$$(1.1) \qquad P(x, dy) = \mu(dy \mid \pi(Tx))$$

where $\mu(dy \mid \pi(Tx))$ is the conditional probability of μ with respect to π evaluated at Tx, i.e. the conditional probability given $\pi(Tx)$. P leads to "local equilibrium" in the sense that any initial probability measure ν is carried by our Markov chain to

$$(1.2) \qquad \nu P(dy) := \int_B \nu(dx) P(x, dy)$$

which agrees with the "equilibrium measure" μ in each fiber of π: the conditional probabilities with respect to π of μ and νP are the same.

From the fact that T preserves μ it follows immediately that $\mu P = \mu$; μ is stationary for the Markov chain as well. We wish to investigate the relationship between the ergodic properties of (B, μ, T) and those of (B, μ, T, π).

(In order to insure that $P(x, dy)$ is everywhere well defined, we make the following assumptions:

(i) μ is defined on the completion of a Standard Borel space (B, Σ) a measurable space isomorphic to the unit interval with Borel sets.

(ii) T and T^{-1} are everywhere defined and Σ-mesurable.

(iii) $\mu(dy \mid \pi(x))$ is a version of the conditional probability given π for which $\mu(\cdot \mid \pi(x))$ is a probability measure on Σ supported on $\pi(x)$ for each $x \in B$; and $\mu(A \mid \pi(\cdot))$ is Σ-measurable for each $A \in \Sigma$.

If these assumptions are not originally satisfied, they will be after removal of an appropriate set of measure zero. We also remark that if we had chosen different versions of T, π, Σ and $\mu(\cdot \mid \pi(x))$, the transition probability obtained would be identical mod 0 with $P(x, dy)$ [2]).

The language used above and the motivation for the introduction of (B, μ, T, π) are derived from the following

Example 1.1. Consider a system of n interacting particles moving in a region Λ according to Newton's equations of motion. The specification of the deterministic motion is completed by imposing elastic reflections on $\partial\Lambda$ [3], the boundary of Λ. Interactions with an external thermal reservoir are represented by random reflections from the walls such that the outgoing (after collisions with $\partial\Lambda$) velocities have "Maxwellian" distributions at the temperature of the reservoir. By representing the deterministic motion as a special flow [5] over a basis B, the phase points for which a particle is at $\partial\Lambda$, we obtain a (discrete) dynamical system (B, μ, T): T is the return mapping on B and μ is the projection onto B of the canonical ensemble at the temperature of the reservoir. Let π be the partition into sets of phase points which differ only by the velocities of the particles at $\partial\Lambda$. Then the conditional probability with respect to π gives the Maxwellian law for reflections from the walls. A detailed study of this system is given in [6] in which we also examine the case of non constant boundary temperature, so that heat transport properties can be studied. (In the latter case we do not however have any a priori stationary μ, and the problem is therefore more difficult.)

Another class of examples of (B, μ, T, π) schemes is provided by "random walk" type processes:

Example 1.2. We may associate with the dynamical system $(\bar{B}, \bar{\mu}, \bar{T})$ the "random walk" $x \to \bar{T}x$ or $\bar{T}^{-1}x$ with probabilities $\frac{1}{2}, \frac{1}{2}$. This Markov chain is a factor of the chain (B, μ, T, π) (see the definition given after this example), where

$$B := \bar{B} \times \{-1, 1\} \qquad \mu := \bar{\mu} \times \lambda, \qquad \lambda(-1) = \lambda(1) = \frac{1}{2}$$

$$T(x, \sigma) := (\bar{T}^{\sigma} x, \sigma), \qquad \sigma = \pm 1$$

and π is the partition of B according to the first component:

$$\pi(x, \sigma) = \{(x, 1), (x, -1)\}, \quad x \in \bar{B}, \ \sigma = \pm 1.$$

In particular, if $(\bar{B}, \bar{\mu}, \bar{T})$ is a Bernoulli shift, then (B, μ, T, π) is essentially the "T, T^{-1} process", a K-system which fails to be Bernoulli [18].

Example 1.2 suggests that many interesting processes arise as factors of (B, μ, T, π) Markov chains. By a *factor* of a Markov chain (B, μ, P) with state space B, transition probability P and stationary probability measure μ, we mean the following: Let α be any measurable partition of B. Then the factor space $\frac{B}{\alpha}$ is the set of all fibers $\alpha(x)$, $x \in B$, of α. Let H_α be the natural homomorphism $B \to \frac{B}{\alpha}$: $H_\alpha(x) = \alpha(x)$. The natural σ-algebra Σ_α on $\frac{B}{\alpha}$ consists of sets A with $H_\alpha^{-1}(A) \in \Sigma$. Let the sequence X_i, $i \in Z$, of B-valued random variables be a realization of our Markov chain with stationary one dimensional distribution μ:

$$\text{Prob}\{X_i \in dx\} = \mu(dx)$$

$$\text{Prob}\{X_{i+1} \in dy \mid X_i = x\} = P(x, dy).$$

We say that the stochastic process Y_i, $i \in Z$, is a *factor* of our stationary process if there exists a measurable partition α of B such that the process Y_i is isomorphic to the process $X_i^{(\alpha)} := H_\alpha(X_i)$. ($Y_i$ is the *factor* with respect to α of X_i). Note that the factor $X_i^{(\alpha)}$ has stationary one dimensional distribution $\mu_\alpha = \mu \circ H_\alpha^{-1}$ on Σ_α.

(B, μ, T, π) Markov chains are closely connected with Markov approximations:

Example 1.3. Suppose the sequence Y_i, $i \in Z$, of $\bar{B}$-valued random variables forms a stationary stochastic process. By the *Markov approximation* X_i of Y_i we mean the $\bar{B}$-valued Markov chain with the same one dimensional distribution and one-step transition probability:

$$\text{Prob}\{X_i \in dx\} = \text{Prob}\{Y_i \in dx\} = \bar{\mu}(dx)$$

$$\text{Prob}\{X_{i+1} \in dy \mid X_i = x\} = \text{Prob}\{Y_{i+1} \in dy \mid Y_i = x\}.$$

Let μ be the probability measure on $B = \bar{B}^Z$ induced by Y_i, $i \in Z$.

For $x \in B$, $x = (\xi_i)$, $\xi_i \in \bar{B}$, let T be the shift on B: $Tx = x' = (\xi_i')$, $\xi_i' = \xi_{i+1}$. Let π be the partition of B according to ξ_0, the time zero partition. Then the Markov approximation of Y_i is the factor with respect to π of the (B, μ, T, π) Markov chain.

Example 1.3 shows that the problem of the relationship between the ergodic properties of a process and those of its Markov approximation is equivalent to the corresponding problem concerning the relationship between (B, μ, T) and (B, μ, T, π). The example also shows that all Markov chains are in fact factors of some (B, μ, T, π) Markov chain; just let Y_i be a Markov chain.

Example 1.4 (master equation) [7]. If we treat the reservoir in Example 1.1 honestly we obtain a dynamical system (B, μ, T) where $B = S \times R$ (system $\times$ reservoir), μ is the equilibrium measure at the temperature of the reservoir, and T is the unit time mapping induced by the equations of motion. Let π be the partition according to the system coordinates (so that $\frac{B}{\alpha} \cong S$). If we observe only the system we obtain an S-valued process; the master equation describes the Markov approximation of this process. It corresponds to putting the reservoir into equilibrium relative to the system at every "instant". It is thus the factor relative to π of the (B, μ, T, π) Markov chain. (Note that the choice of time unit is important. Different units would lead to essentially different master equations.)

A word about notation. We will frequently denote a measurable partition and its corresponding σ-algebra by the same symbol. To the extent that the distinction matters, it should be clear from the context what is intended. We will also frequently use operator notation. We will, e.g., write Pf and $\mu(f)$ for $\int P(x, dy) f(y)$ and $\int \mu(dx) f(x)$ respectively. The n-step transition probability $P^n(x, A)$ is given by $(P^n I_A)(x)$, where I_A is the indicator function of the set $A \in \Sigma$. For a (B, μ, T, π) Markov chain $P = UK$, where $(Uf)(x) := f(Tx)$ and $K(x, dy) := \mu(dy \mid \pi(x))$.

Asymptotic behavior

The asymptotic behavior of a Markov chain (B, μ, P) may be studied from several points of view. One may, for example, study the ergodic properties of the dynamical system (B^Z, P_μ, S) corresponding to the Markov chain. Here P_μ is the Markov probability measure on B^Z (equipped with the product σ-algebra $\mathscr{J}$) with stationary distribution μ and one step transition probability P. Namely, let us write $x = (x_i)$, $x_i \in B$, $i \in Z$, for $x \in B^Z$, and let $X_i(x) = x_i$ be the state at time i. Then $P_\mu\{X_i \in dx\} = \mu(dx)$

$$P_\mu\{X_{i+1} \in dy \mid X_i = x, X_{i-1} = x_{i-1}, X_{i-2} = x_{i-2} \ldots\} =$$
$$= P(x, dy).$$

S is the shift on B^Z. We will say that (B, μ, P) is ergodic, mixing, ... according to whether (B^Z, P_μ, S) is ergodic, mixing, ...

Another approach is to investigate directly the action of P on B, i.e., to study the asymptotic behavior of $P^n(f)$, $n \to \infty$. This is the approach of e.g. F o g u e l [9]. Note that the operator adjoint P^+ of P ($\int d\mu g(Pf) = \int d\mu f(P^+ g)$ for $f \in L^\infty(B, \mu)$, $g \in L^1(B, \mu)$, i.e., for $\nu << \mu$, $P^+\left(\frac{d\nu}{d\mu}\right) = \frac{d(\nu P)}{d\mu}$) is the transition probability for the time reversed Markov chain, since

$$\int d\mu g(Pf) = \int dP_\mu g(X_{-1}) f(X_0)$$

and

$$\int d\mu f(P^+ g) = \int dP_\mu f(X_0)(P^+ g)(X_0).$$

(P and P^+ act on $L^1(B, \mu)$, and in particular take null functions to null functions, because μ is stationary.) Thus if (B, μ, P) is the (B, μ, T, π)-Markov chain, then (B, μ, P^+) is the (B, μ, T^{-1}, π_-) Markov chain, where $\pi_- = T^{-1}\pi$.

Needless to say, the two approaches are closely related. For example, various types of convergence properties for P^n can be formulated in terms of conditions on certain σ-algebras on B^Z:

Let $\mathscr{I}_i = \sigma(X_i)$, the σ-algebra generated by X_i. Let

$$\underset{(\geq m)}{\mathscr{I}_{\leq m}} = \bigcup_{\substack{i \leq m \\ (i \geq m)}} \mathscr{I}_i$$

the σ-algebra of events observable prior to (after) time m. Then the tail fields $\mathscr{I}_{-\infty}$ and $\mathscr{I}_{\infty}$ are given by

$$\mathscr{I}_{-\infty} = \bigcap_m \mathscr{I}_{\leq m}$$

$$\mathscr{I}_{\infty} = \bigcap_m \mathscr{I}_{\geq m}.$$

$\mathscr{I}_{-\infty}$ ($\mathscr{I}_{\infty}$) is the σ-algebra of events observable in the arbitrarily distant past (future). We will also consider the σ-algebra

$$\mathscr{I}_{|\infty|} := \mathscr{I}_{\infty} \cap \mathscr{I}_{-\infty}$$

and the σ-algebra $\mathscr{I}$ of invariant events: $A \in \mathscr{I} \Longleftrightarrow SA = A$. Note that $\mathscr{I} \subset \mathscr{I}_{|\infty|}$. $\mathscr{I}_{\infty}$ and $\mathscr{I}_{-\infty}$ are closely connected with the convergence of P^n to equilibrium μ:

(1) $\mathscr{I}_{-\infty}$ is trivial $\Longleftrightarrow P^n f \to \mu(f)$ in L^1 for all $f \in L^2(B, \mu)$.

(2) $\mathscr{I}_{\infty}$ is trivial $\Longleftrightarrow \nu P^n \to \mu$ in variation norm for all probability measures $\nu << \mu$ (the variation norm of a signed measure ν is

$$\|\nu\| = \sup_{\|f\|_{\infty} \leq 1} |\nu(f)|).$$

(1) follows from the martingale convergence theorem and the stationarity of μ. (2) follows from (1) upon taking adjoints. A condition, "stability", providing information about $\mathscr{I}_{-\infty}$ will be investigated in Section 6.

If we strengthen the convergence in (2) by requiring that it hold for all probability measures ν, i.e., by requiring that

$$P^n(x, dy) \to \mu(dy) \quad \text{in variation norm}$$

we obtain an ergodic, aperiodic Harris process, various formulation of which will be studied in Section 5. There it will be explained that when the (B, μ, T, π) process is Harris and (B, μ, T) is mixing then (B, μ, T, π)

is Bernoulli. $\mathscr{I}_{\infty}$, $\mathscr{I}_{-\infty}$ and $\mathscr{I}_{|\infty|}$ will also be identified for Harris processes, they are in fact the same. This is an important case since many chains arising in physical applications are Harris, e.g. Example 1.1, [6].

If on the other hand we weaken the convergence in (2) and require only that $\nu P^n \to \mu$ weakly, i.e. that $\nu P^n(f) \to \mu(f)$ for all $f \in L^\infty(B, \mu)$ we obtain a condition equivalent to mixing. A sufficient condition for mixing is that $\mathscr{I}_{|\infty|}$ be trivial (see Section 4).

It turns out that $\mathscr{I}_{|\infty|}$ may be identified with a measurable partition w of B which defines the largest deterministic factor of (B, μ, P). We call this factor the deterministic part of (B, μ, P). If w is trivial, we say that the Markov chain is purely stochastic. That purely stochastic chains are mixing is a well known result in the theory of Markov chains [8], [9]. More generally, a stationary Markov chain is mixing (ergodic) if and only if its deterministic part is mixing (ergodic). These matters are discussed in Sections 2, 3 and 4 where various characterizations of w and $\mathscr{I}$ are given, some in terms of σ-algebras on B^Z, some directly in terms of P, and some, for (B, μ, T, π) schemes, in terms of π and T.

If a stationary Markov chain is not purely stochastic, i.e. if w is not trivial, the convergence to equilibrium as in (1) or (2) cannot hold since $P^n(x, dy)$ will be supported by $w(T^n x)$. (Note also that in this case $\mathscr{I}_{\infty} \cap \mathscr{I}_{-\infty}$ is not trivial.) However one can still ask the following: Do points in the same fiber of w have the same asymptotic behavior (local approach to equilibrium)? Equivalently, does $\mathscr{I}_{-\infty} = w$, i.e. does knowledge of the distant past only tell us which fiber of w the system is presently in? In Section 6 we give examples showing that this is not always true.

However, for Harris chains it is true (Section 5), and a similar but weaker result holds for stable chains (Section 6). In fact if the chain is Harris w has countably many atoms and

$$(1.2) \qquad \lim_{n \to \infty} \| P^n(x, dy) - \mu(dy \mid w(T^n x) \| = 0.$$

(1.2) remains true if the Harris condition is "relativized" to w, though now w may have uncountably many fibers. Relativized Harris chains are investigated in Section 7 where it is also shown that w in (B^Z, P_μ, S) splits off: (B^Z, P_μ, S) can be written as a direct product of the deterministic part of (B, μ, P) and a Bernoulli shift. Thus relativized Harris chains (B, μ, T, π) are Bernoulli provided (B, μ, T) is Bernoulli, since the deterministic part of (B, μ, T, π) is a factor of (B, μ, T) and factors of Bernoulli systems are Bernoulli [4].

2. ERGODICITY

We first give various characterizations of the σ-algebra $\mathscr{I}$ of the invariant sets of (B^Z, P_μ, S), a stationary Markov chain which may or may not be a (B, μ, T, π) scheme. As the proof is classical [8], [16] we just sketch it.

Lemma 2.1. $\mathscr{I}_{|\infty|} \subset \mathscr{I}_0$ *and therefore* $\mathscr{I} \subset \mathscr{I}_0$.

Proof. Let g be a bounded $\mathscr{I}_{|\infty|}$ measurable function. Then $g \in \mathscr{I}_\infty$, $g \in \mathscr{I}_{-\infty}$ and by the Markov property

$$P_\mu(|g|^2 \mid \mathscr{I}_0) = P_\mu(g \mid \mathscr{I}_0)^2$$

proving the lemma.

Thus $g \in \mathscr{I}$ is of the form $g = \hat{g}(X_0)$, $\hat{g}$ measurable on B. Moreover

Lemma 2.2. *Let* $g \in \mathscr{I}$ *and* $g \in L^1(B^Z, P_\mu)$. *Then*

$$P(\hat{g}) = P^+(\hat{g}) = \hat{g}$$

where $g = \hat{g}(X_0)$.

Proof. $(P\hat{g})(X_0) = P_\mu(\hat{g}(X_1) \mid \mathscr{I}_0) = P_\mu(\hat{g}(X_0) \mid \mathscr{I}_0) = \hat{g}(X_0)$ and similarly for P^+.

The converse of Lemma 2.2 is also true:

Lemma 2.3. *Let* $g \in L^1(B, \mu)$ *and let either* $g = Pg$ *or* $g = P^+g$; *then* $g(X_0) \in \mathscr{I}$.

Proof. Suppose $g = Pg$. Then $g_n = g(X_n)$ is a uniformly integrable martingale [17] with respect to $\mathscr{I}_{\leq n}$. Therefore $g_\infty = \lim_{n \to \infty} g_n$ is P_μ a.e. defined, $g_\infty \in L^1(B^Z, P_\mu) \cap \mathscr{I}$, and $P_\mu(g_\infty \mid X_n) = g_n$. Therefore by Lemma 2.1,

$$g_\infty = P_\mu(g_\infty \mid \mathscr{I}_0) = g(X_0) \in \mathscr{I}.$$

In case $g = P^+ g$, the proof is similar.

We collect these results and others in the

Theorem 2.1.

(i) $f = Pf \Longleftrightarrow f = P^+ f$, $f \in L^1(B, \mu)$ *(f is then called harmonic).*

(ii) $\mathscr{I} \subset \mathscr{I}_0$. *Thus* $\mathscr{I}$ *may be regarded as coming from a partition of* B, *which we denote by* τ.

(iii) $\tau = \{A \in \Sigma \mid PI_A = I_A\}$.

(iv) *For* $f \in L^1(B, \mu)$, $f \in \tau \Longleftrightarrow f$ *is harmonic.*

(v) *Let* ν *be stationary for* P, $\nu P = \nu$, *and absolutely continuous with respect to* μ. *Then* $\frac{d\nu}{d\mu}$ *is harmonic, the invariant measure* $P_\nu << P_\mu$ *and* $\frac{dP_\nu}{dP_\mu} = \frac{d\nu}{d\mu}(X_0)$. *Conversely, if* Q *is an invariant measure under* S *and* $Q << P_\mu$ *then* $\frac{dQ}{dP_\mu} \in \mathscr{I}$, $\widehat{\frac{dQ}{dP_\mu}}\,\mu$ *is stationary for* P, *where* $\frac{dQ}{dP_\mu} = \widehat{\frac{dQ}{dP_\mu}}(X_0)$.

(vi) *For a* (B, μ, T, π) *Markov chain* τ *is the* σ*-algebra of* T *invariant* π *measurable sets.*

(vii) *The following "definitions" of ergodicity are equivalent:*

(a) $\mathscr{I}$ *is trivial.*

(b) τ *is trivial.*

(c) *Other than constant multiples of* μ, *no measures stationary for* P *and absolutely continuous with respect to* μ *exist.*

Proof. All but (vi) are straightforward consequences of Lemmas 2.1, 2.2, 2.3. (vi) follows easily from (iii):

$$PI_A := \dot{U}KI_A = I_A \Leftrightarrow KI_A = I_A \quad \text{and} \quad UI_A = I_A \Leftrightarrow$$
$$\Leftrightarrow A \in \pi \quad \text{and} \quad TA = A,$$

since U is unitary and K is an orthogonal projection on $L^2(B, \mu)$.

3. THE DETERMINISTIC PART OF A PROCESS

We wish to investigate the extent to which stochasticity improves convergence to "equilibrium" over that which occurs in deterministic processes, for which δ measures evolve into δ measures. To study the effect of stochasticity we isolate the deterministic part of a process and study convergence to equilibrium within the fibers of the deterministic factor.

We say that a stationary process Y_i, $i \in Z$, is *deterministic* if it is isomorphic to $\bar{Y}_i(x) = T^i x$, $i \in Z$, $x \in B$, (B, μ, T) an (invertible) dynamical system. If the factor with respect to α is deterministic we say that α is *deterministic*.

By the deterministic part of a stationary process X_i or of the stationary Markov chain (B, μ, P) with realization X_i, we mean the largest deterministic factor of X_i. This is the factor with respect to $w = \sup \alpha$ (mod 0), where α runs over all deterministic measurable partitions of B. (It follows from Theorem 3.1 that $\sup \alpha$ is deterministic.)

Any sub-σ-algebra on B, in particular w, may be regarded as a sub-σ-algebra of $\mathcal{J}_0$, and hence a σ-algebra on B^Z. Theorem 3.1 should be read with this in mind.

w has many different characterizations:

Theorem 3.1. *For a stationary Markov chain* (B, μ, P), *the following are identical* (mod 0)

(i) *the partition giving rise to the deterministic part of* (B, μ, P),

(ii) *the* σ*-algebra* $\{A \in \Sigma \mid P^n I_A = I_{B_n},\ P^{+n} I_A = I_{C_n},\ n \geq 1\}$.

(iii) *the* σ*-algebra* $\{A \in \Sigma \mid \|P^n I_A\|_2 = \|P^{+n} I_A\|_2 = \|I_A\|_2\}$ *whe*
$\|\cdot\|$ *denotes the norm of* $L^2(B, \mu)$,

(iv) *the* σ*-algebra* w *such that* $f \in L^2(B, \mu)$ *is* w*-masurable*
$\Longleftrightarrow \|P^n f\|_2 = \|P^{+n} f\|_2 = \|f\|_2 \quad (n \geq 1)$,

(v) $\mathscr{I}_{|\infty|} \equiv \mathscr{I}_{-\infty} \cap \mathscr{I}_{\infty}$,

(vi) $\bigcap_{i=-\infty}^{+\infty} \mathscr{I}_i$,

For (B, μ, T, π) *Markov chains:*

(vii) *the finest measurable partition* w *satisfying*

(a) $w \leqslant \pi$,

(b) $Tw = w$ *(the finest* *T-invariant measurable partition coa*
than π*),*

(viii) $\bigcap_{n=-\infty}^{+\infty} T^n \pi$,

(ix) *the* σ*-algebra* w *such that* $f \in L^1(B, \mu)$ *is* w*-measurable*
$\Longleftrightarrow P^n f = U^n f,\ P^{+n} f = U^{-n} f,\ n \geq 1$.

Proof. Most of the proof is straightforward, and we omit de

(ii), (iii) and (iv) are studied by F o g u e l [9]. To see that (i) and are equivalent, note that the σ-algebra $\hat{w}$ described in (ii) is inva under P and P^+, i.e.

$$A \in \hat{w} \Longleftrightarrow PI_A = I_D,\ D \in \hat{w} \ \text{ and } \ P^+ I_A = I_{D^+},\ D^+ \in \hat{w}$$

(using $PI_A = I_D \Longleftrightarrow I_A = P^+ I_D$.)

Note also that for A and $D \in \hat{w}$, $(PI_A)(PI_D) = P(I_A I_D)$. It fo that the mapping $\Sigma \to \Sigma$ induced by P comes from an automorp on (B, μ). That this automorphism is maximal is obvious.

Let $Ef = \mu(f \mid w)$. Note that for a (B, μ, T, π) Markov ch $EU = UE$ and $EK = KE = E$. More generally $EP = PE = U_w E$, wh U_w acts on w-measurable functions by

$$U_w f(w(x)) = f(T_w(w(x))).$$

Set

$$(f, g) = \int d\mu f(x) g(x).$$

Theorem 4.1. *Let* $f \in L^\infty(B, \mu)$, $g \in L^1(B, \mu)$ *and let* $f^\perp = f -$ $g^\perp = g - Eg$. *Then*

$$\lim_{n \to \infty} (g, P^n f^\perp) = \lim_{n \to \infty} (g^\perp, P^n f) = 0. \tag{4.1}$$

(4.1) *can be restricted as:*

$$\text{weak} \lim_{n \to \infty} (P^n - U_w^n E) = 0.$$

Corollary. (B^Z, P_μ, S) *is mixing (ergodic) if and only if the terministic factor* $\left(\frac{B}{w}, \mu_w, T_w\right)$ *is mixing (ergodic).*

Proof. The mixing statement follows from the theorem and ergodicity statement from the observation $\tau \leqslant w$.

Note that it follows from Theorem 4.1 (or from the Corollary si the trivial dynamical system is mixing) that purely stochastic chains mixing. We would like to know how strong the ergodic properties purely stochastic chains must be. More generally, we would like to kn the extent to which we have convergence to "equilibrium" in the fib of w, namely

$$P^n(x, dy) - \mu(dy \mid w(T^n x)) \to 0.$$

Some questions:

(1) Are purely stochastic chains Bernoulli (K-systems)?

(2) If $\left(\frac{B}{w}, \mu_w, T_w\right)$ is Bernoulli (K) is (B^Z, P_μ, S) Bernoulli (

With regard to (iii) note that $(n \geq 1)$

$$\| P^n I_A \|_2 = \| I_A \|_2 \Leftrightarrow P_\mu (I_A(X_n) \mid X_0) = I_A(X_n) \Leftrightarrow$$
$$\Leftrightarrow I_A(X_n) = I_D(X_0) \mod 0 \Leftrightarrow P^n I_A = I_D .$$

For (vi), note that

$$A \in \bigcap_i \mathcal{I}_i \Leftrightarrow I_A(X_n) = I_D(X_0), \quad n \in Z.$$

As far as (v) is concerned, it is clear that $\bigcap_{i=-\infty}^{+\infty} \mathcal{I}_i \subset \mathcal{I}_{|\infty|}$. On the other hand, it follows from Lemma 2.1 that $\mathcal{I}_{|\infty|} \subset \mathcal{I}_i$ for all i, i.e. that $\mathcal{I}_{|\infty|} \subset \bigcap_{i=-\infty}^{+\infty} \mathcal{I}_i$.

For (vii) note that from (iv) we have that for $f \in L^2$, $f \in w$, $Kf = f \Rightarrow f \in \pi$ so that $w \leq \pi$. Since for π measurable A, $PI_A = UI_A$, it follows from the invariance of w under P and P^+ that $Tw = w$.

(viii) and (ix) are easily seen to be equivalent to (vii).

We denote by T_w the automorphism of $\left(\frac{B}{w}, \mu_w\right)$ giving rise to the deterministic part of our process. We will also say that $\left(\frac{B}{w}, \mu_w, T_w\right)$ is the deterministic part of our process. Note that for a (B, μ, T, π) Markov chain, T_w is induced by T: $T_w(w(x)) = w(Tx)$. Note also that $\tau \leq w$.

The ergodic properties of (B, μ, P) are limited by those of $\left(\frac{B}{w}, \mu_w, T_w\right)$. We will say that our process (B, μ, P) is *purely stochastic* if it has no deterministic part, i.e., if w is trivial. Since $\tau \leq w$, purely stochastic chains are ergodic. It follows from the theorem of the next section that purely stochastic processes are mixing.

4. MIXING

The following theorem is proven in the literature, see e.g. [9]. In Appendix A we will prove a stronger version, see Lemma A.5, which will be needed in Section 7.

(3) Note that $w = \mathscr{I}_{-\infty} \cap \mathscr{I}_{\infty} \subset \mathscr{I}_{\infty}, \mathscr{I}_{-\infty}$. Must $w = \mathscr{I}_{\infty} = \mathscr{I}_{-\infty}$?

We note that if the answer to (2) were "yes", then the (B, μ, T, π) process would be Bernoulli (K) provided (B, μ, T) were, because factors of Bernoulli systems are Bernoulli. However, as far as Bernoulliness is concerned in general the answer is not "yes"

We conclude this section with an example of a purely stochastic chain which is not Bernoulli (see also Example 1.2). In Section 6 we give an example of a purely stochastic chain with nontrivial $\mathscr{I}_{-\infty}$

Example 4.1. Let $(\bar{B}, \bar{\mu}, \bar{T})$ be an ergodic system which is not loosely Bernoulli [10]. Consider the Markov chain $(\bar{B}, \bar{\mu}, P)$ arising from $(\bar{B}, \bar{\mu}, \bar{T})$ through the transitions:

$$x \longmapsto \bar{T}x \quad \text{or} \quad x \longmapsto x$$

with probabilities $\frac{1}{2}, \frac{1}{2}$. (As in Example 1.2 this is a factor of a (B, μ, T, π) chain, where $B = \bar{B} \times \{0, 1\}$.)

Using the ergodic theorem, one may show that $\mathscr{I}_{-\infty}$ is trivial for this Markov chain. Thus w is trivial. Since $\bar{T}$ is a factor of a system induced from T, T is not loosely Bernoulli, hence not Bernoulli (see Feldman [10]).

5. HARRIS CHAINS

The answers to the questions (1), (2) and (3) of Section 4 are affirmative for Harris chains.

Definition 5.1. A stationary Markov chain (B, μ, P) is called a *Harris chain* if for μ-a.e. $x \in B$ there exists an $n = n(x)$ such that $P^n(x, dy)$ has a component absolutely continuous with respect to μ.

(A stationary Markov chain is a conservative Markov process. If (B, μ, P) is not stationary but otherwise satisfies the conditions of Definition 5.1, it is called a Harris chain provided it is conservative as well, see Foguel [9].)

Let $P^n = Q_n + R_n$ be the decomposition of P^n into its absolutely continuous part (Q_n) and its singular part. Q_n and R_n are measurable, i.e. $Q_n f$ and $R_n f$ are measurable for $f \in L^\infty(B, \mu)$.

To see this note that $Q_n(x, dy) = q_n(x, y)\mu(dy)$ for μ-a.e. x, where $q_n(x,y) = \dfrac{d(\mu \times P)_{\text{a.c.}}}{d(\mu \times \mu)}$. Here

$$(\mu \times P)(dxdy) = \mu(dx)P(x, dy) \quad \text{on} \quad B \times B,$$

and $(\mu \times P)_{\text{a.c.}}$ is the component absolutely continuous with respect to $\mu \times \mu$.

A key observation is the following:

(5.1) $\qquad P^m Q_n \leqslant Q_{m+n}$

(5.2) $\qquad Q_m P^n \leqslant Q_{m+n}$

(5.3) $\qquad R_n R_m \geqslant R_{m+n}.$

Some basic preliminary facts about Harris chains are collected in the following

Theorem 5.1.

(i) (B, μ, P) *is Harris* $\Longleftrightarrow \sum_{n=1}^{\infty} Q_n f \not\equiv 0$ *for all* $0 \leqslant f \not\equiv 0$, $f \in$ $\in L^\infty(B, \mu)$.

(ii) *If* (B, μ, P) *is Harris, then* $R_n(x, B) \searrow 0$ *for* μ*-a.e.* x.

(iii) (B, μ, P) *is Harris* $\Longleftrightarrow (B, \mu, P^+)$ *is Harris.*

(iv) *Suppose* (B, μ, P) *is ergodic. Then it is Harris* $\Longleftrightarrow$ *there is a* μ *full set* $B' \subset B$ *such that if* $\mu(A) > 0$ *then*

$$P_x(\mathscr{A})(:= P_{\delta_x}(\mathscr{A})) = 1 \quad \textit{for} \quad x \in B'.$$

where

$$\mathscr{A} = \{X_n \in A,\ n \geqslant 0,\ \textit{infinitely often}\}$$

(Orey's condition).

Proof. See e.g. [8], [9]. For ease of reference we give a sketch.

(i) Without loss of generality we may assume that (B, μ, P) is ergodic. By the ergodic theorem,

$$\left(\sum_{n=0}^{\infty} P^n f\right)(x) = P_x\left(\sum_{n=0}^{\infty} f(X_n)\right) = \infty$$

for μ-a.e. x. Thus using (5.2)

$$\sum_{n=0}^{\infty} Q_n f \geqslant \sum_{n=m}^{\infty} Q_n f \geqslant Q_m \sum_{n=0}^{\infty} P^n f \geqslant Q_m 1.$$

Since Definition 5.1. says that $\sum_{n=0}^{\infty} Q_m 1 > 0$, for μ-a.e. x, "$\Rightarrow$" follows "$\Leftarrow$" follows by ergodicity, from the observation that the set $\left\{\sum_{n=1}^{\infty} Q_n 1 = 0\right\}$ is invariant (use (5.1)).

(ii) $R_n(x, B) := R_n 1$ is decreasing by (5.3). Let $g = \lim_{n \to \infty} R_n 1$. By (5.3), $R_n g \geqslant g$. Thus $P^n g \geqslant g$. It follows that $P^n g = g$, since $\mu(P^n g - g) = \mu(g) - \mu(g) = 0$. Thus $Q_n g = 0$ for all n, i.e. by (i), $g = 0$.

(iii) follows from (i) by noticing that Q_n^+, the absolutely continuous part of P_n^+, has the kernel $q_n^+(x, y) = q_n(y, x)$. (See the proof of Theorem 7.3 (iii).)

(iv)

The "$\Leftarrow$" part. Assume that for $x \in B'$, $\mu(B') = 1$, $P_x(\mathscr{A}) = 1$ whenever $\mu(A) > 0$. Suppose $P^n(x, dy)$ is singular for all $n \geqslant 1$. Then there exists a set $C(x)$, $\mu(C(x)) = 0$, such that $P^n(x, C(x)) = 1$ for all $n \geqslant 1$. Let $A = B - C(x)$. Then $\mu(A) = 1$ and $P_x(\mathscr{A}) = 0$, so that $x \notin B'$.

The "$\Rightarrow$" part. Suppose (B, μ, P) is Harris. By the ergodic theorem $P_\mu(\mathscr{A}) = 1$, and $P_y(\mathscr{A}) = 1$ for μ a.e. y. Let

$$B' = \{x \in B \mid R_n(x, B) \searrow 0\}.$$

Then for $x \in B'$

$$P_x(\mathscr{A}) = \int P^n(x, dy) P_y(\mathscr{A}) \geqslant \int Q_n(x, dy) P_y(\mathscr{A}) =$$
$$= Q_n(x, B) = 1 - R_n(x, B) \to 1$$

by (ii).

The condition in Definition 5.1 seems quite natural physically, and in particular is satisfied in Example 1.1.

Harris chains have very strong ergodic properties [8], [9].

Theorem 5.2. *Let* (B, μ, P) *be Harris. Then*

(i) w *is atomic, i.e. it has countably many fibers, each of positive measure* (mod 0). *If* (B, μ, P) *is ergodic* w *is finite* (mod 0).

(ii) *For* μ*-a.e.* $x \in a$, *an atom of* w*:*

(5.4) $$\lim_{n \to \infty} \| \delta_\kappa P^n - \mu(\cdot \mid T_w^n a) \| = 0.$$

(iii) $\mathscr{I}_{-\infty} = w = \mathscr{I}_\infty$ $(P_\mu$ mod 0). *In particular, if the Harris chain* (B, μ, P) *is purely stochastic,* $\mathscr{I}_{-\infty}$ *and* $\mathscr{I}_\infty$ *are trivial and for* μ*-a.e.* x,

(5.5) $$\delta_\kappa P^n = P^n(x, \cdot) \to \mu$$

in variation norm. Therefore (B^Z, P_μ, S) *is Bernoulli* [4] *in this case.*

Remark. It follows from (5.4) that there exists a set $\hat{B} \subset B$, $\mu(\hat{B}) = 1$, such that if ν is a probability measure supported by $a \cap \hat{B}$, then

(5.6) $$\lim_{h \to \infty} \| \nu P^n - \mu(\cdot \mid T_w^n a) \| = 0.$$

Proof. The proof of (i) is straightforward. (ii) is proven by first applying Theorem 4.1 and Theorem 5.1 (ii) to P^+ to obtain (5.6) for $\nu << \mu$ and then using Theorem 5.1 (ii) again to obtain (5.4). For details, see Foguel [9], or Section 7. (iii) follows from (ii), see e.g. Section 6.

(B^Z, P_μ, S) is Bernoulli because every finite measurable partition $\leqslant \mathscr{I}_0$ is a weak Bernoulli partition, which easily follows from (5.5) [11], [4].

6. THE ENDOMORPHIC PART OF A PROCESS; STABLE PROCESSES

In Theorem 3.1 it is shown that $\mathscr{J}_{-\infty} = w$ for Harris chains. That $\mathscr{J}_{-\infty} = w$ is not true generally follows from the following

Example 6.1. Let $B = \{-1, 1\}^Z$ and let (B, μ, T) be the $\frac{1}{2}, \frac{1}{2}$ Bernoulli shift. For $x \in B$ we write $x = (\xi_i)$, $i \in Z$, $\xi_i = \pm 1$. Let

$$\hat{\mathscr{J}}_{\geq m} = \sigma(\xi_j \mid j \geq m).$$

Let $\pi = \hat{\mathscr{J}}_{\geq 0}$ and consider the (B, μ, T, π) process. Note that $T^{-m}\pi = \hat{\mathscr{J}}_{\geq m}$. Therefore $w = \bigcap\limits_n T^n \pi = \bigcap\limits_m \hat{\mathscr{J}}_{\geq m}$ is trivial. However, though $\mathscr{J}_{+\infty}$ is also trivial, $\mathscr{J}_{-\infty} \geq \pi$. (Recall that the $\mathscr{J}$'s refer to σ-algebras on B^Z, while $\hat{\mathscr{J}}$'s refer to σ-algebras on B). In fact, it follows from the fact $T\pi \geq \pi$ that T acts as an endomorphism T_π on $\frac{B}{\pi}$ and the factor with respect to π in the purely stochastic (B, μ, T, π) chain is an endomorphic process, i.e. it is induced by the (non-invertible) measure preserving transformation T_π on $\frac{B}{\pi}$. (T_π on B_π is just the one sided shift). It follows that $\mathscr{J}_{-\infty}$ may be identified with the doubly infinite trajectories of this endomorphic process, i.e.

$$\mathscr{J}_{-\infty} = \bigcup_i \pi_i, \qquad \pi_i = \sigma(H_\pi(x_i))$$

and

$$P_\mu(f(X_i) \mid \mathscr{J}_{-\infty}) = \mu(f \mid \pi(X_i)).$$

Example 6.1 suggests that we study the relationship between $\mathscr{J}_{-\infty}$ and the *endomorphic part* of a stationary Markov chain (B, μ, P). Analogously to the deterministic part of a process, the *endomorphic part* of (B, μ, P) is defined to be its largest endomorphic factor.

We will say that a measurable partition α of B is *endomorphic* if the factor with respect to α of the (B, μ, P) process is endomorphic. For any measurable partition, let

$$\alpha_i = \sigma(H_q(X_i))$$

be the partition according to α at time "i". Note that α is endomorphic iff $\alpha_{i+1} \leqslant \alpha_i$, $i \in Z$. Note also that for a (B, μ, T, π) process α is endomorphic iff $\alpha \leqslant \pi$, $\alpha \leqslant T\alpha$.

The proof of the following theorem is similar to that of Theorem 3.1.

Theorem 6.1. *For a stationary Markov chain* (B, μ, P) *the following are identical* (mod 0)

(i) *the partition giving rise to the endomorphic part of* (B, μ, P), *i.e. the largest endomorphic partition of* B,

(ii) *the* σ*-algebra*

$$\{A \in \Sigma \mid P^n I_A = I_{B_n},\ n \geqslant 1\},$$

(iii) *the* σ*-algebra*

$$\{A \in \Sigma \mid \|P^n I_A\|_2 = \|I_A\|_2\},$$

(iv) *the* σ*-algebra* w^+ *such that* $f \in L^2(B, \mu)$ *is* w^+ *measurable iff* $\|P^n f\|_2 = \|f\|_2$, $n \geqslant 1$,

(v) $\bigcap_{n \leqslant 0} \mathscr{I}_n$.

For a (B, μ, T, π) *Markov chain:*

(vi) *the finest measurable partition* w^+ *satisfying*

(a) $w^+ \leqslant \pi$

(b) $Tw^+ \geqslant w^+$,

(vii) $\bigcap_{n \geqslant 0} T^n \pi$,

(viii) *the* σ*-algebra* w^+ *such that* $f \in L^1(B, \mu)$ *is* w^+ *measurable iff* $P^n f = U^n f$, $n \geqslant 1$.

We denote by w^+ the measurable partition described in Theorem 6.1. We write T_+ for the endomorphism of $\frac{B}{w^+}$ which induces the endomorphic part of (B,μ,P). Note that for a (B,μ,T,π) scheme, $T_+(w^+(x)) = w^+(Tx)$.

Notice that the counterpart of Theorem 3.1 (v), which would say that $\mathscr{I}_{-\infty} = \bigcup_i w_i^+$ is missing. At the end of this section we give an example (Example 6.3) in which w^+ and hence $\bigcup_i w_i^+$ is trivial, though $\mathscr{I}_{-\infty}$ is not trivial, i.e. a stationary Markov process in which the tail $\mathscr{I}_{-\infty}$ contains more than merely what is required by its endomorphic part. (Note that $w \leqslant w^+ \leqslant \bigcup_i w_i^+ \leqslant \mathscr{I}_{-\infty}$ is always true, and that for Harris chains, $\mathscr{I}_{-\infty} = w = w^+$.)

We first give a condition under which the counterpart of Theorem 4.1 (v) is true.

Let $\widetilde{\mathcal{O}} \subset \Sigma$ be a countable algebra which σ-generates Σ, and define the family $\mathcal{O}$ of functions $0 \leqslant f \leqslant 1$ on B by

$$\mathcal{O} = \{f \mid f(x) = P_x(X_0 \in A_0, \ldots, X_n \in A_n,\ n \geqslant 0,$$
$$A_0, \ldots, A_n \in \widetilde{\mathcal{O}}\}.$$

We say that an endomorphic partition σ of B is *stable* if (mod 0)

$$x \overset{\sigma}{\sim} y \Longleftrightarrow \lim_{n\to\infty} (P^n(y,A)) = 0 \qquad \text{for all} \qquad A \in \mathcal{O}.$$

($x \overset{\sigma}{\sim} y$ means that x and y are in the same fiber of σ.) We say that (B,μ,P) is stable if w^+ is stable. Note that if (B,μ,P) has any stable partition σ then w^+ is stable, since $\sigma \leqslant w^+$.

We will discuss the structure of the tail field $\mathscr{I}_{-\infty}$ in stable systems. First we give an argument showing that any (B,μ,T,π) chain is "almost" stable. However, at the end of this section, in Example 6.3, we describe a stationary Markov chain which is not stable (and the canonical (B,μ,T,π) chain in which this chain is the factor with respect to π (see Example 1.3) is therefore not stable either).

Let λ denote the partition of B defined by

$$x \overset{\lambda}{\sim} y \Longleftrightarrow \lim_{n\to\infty} [P^n(x,A) - P^n(y,A)] = 0 \qquad \text{for all} \qquad A \in \mathcal{O}.$$

In Appendix A we prove

Theorem 6.2. *For a* (B, μ, T, π) *process* $\lambda \leq T^n \pi \pmod 0$ *for all* $n \in Z$.

Note that it seems to follow from $\lambda \leq T^n \pi$, $n = 0, 1, 2, \ldots$ that

$$\lambda \leq \bigcap_{n=0}^{\infty} T^n \pi = w^+$$

and hence that every (B, μ, T, π) process is stable. The catch is that λ need not be measurable. (Recall that

$$\bigcap_{n=0}^{\infty} T^n \pi$$

is the finest *measurable* partition $\leq T^n \pi$, $n \geq 0$. See in this regard Rohlin [2], No. 3.)

We give a characterization of $\mathscr{I}_{-\infty}$ for stable systems in Theorem 6.3 below. The following lemma follows from the Lebesgue dominated convergence theorem.

Lemma 6.1. *Let* σ *be a stable partition of* (B, μ, P). *Then for* μ*-e.e.* x

$$\lim_{n \to \infty} (P^n(x, A) - \varphi_n(x, A)) = 0 \quad \textit{for all} \quad A \in \mathscr{O}$$

where

$$\varphi_n(x, A) = \int \mu(dy \mid \sigma(x)) P^n(y, A).$$

Theorem 6.3. *Let* σ *be a stable partition for* (B, μ, P). *Then*

$$\mathscr{I}_{-\infty} = \bigcup_{i=0}^{-\infty} \sigma_i = \bigcap_{i=0}^{-\infty} \bigcup_{j=i}^{-\infty} \sigma_j. \tag{6.1}$$

Proof. The last equality in (6.1) is a consequence to the fact that $\sigma_{i+1} \leq \sigma_i$. Applying Doob's martingale convergence theorem to $\sigma_{-n} \nearrow \nearrow \bigcup_{i=0}^{-\infty} \sigma_i$ and to $\mathscr{I}_{\leq -n} \searrow \mathscr{I}_{-\infty}$ we obtain for P_μ-a.e. x and every $i \in Z$

$$P_\mu\Big(A(X_i) \mid \Big(\bigcup_{i=0}^{-\infty} \sigma_i\Big)(x)\Big) = \lim_{n\to\infty} P_\mu(A(X_i) \mid \sigma_{-n}(x)) =$$

$$(6.2) \qquad = \lim_{n\to\infty} \int \mu(dy \mid \sigma(x_{-n}))P^{n+i}(y, A) =$$

$$= \lim_{n\to\infty} \varphi_{n+i}(X_{-n}, A), \qquad A \in \mathcal{O},$$

and

$$P_\mu(A(X_i) \mid \mathcal{J}_{-\infty}(x)) = \lim_{n\to\infty} P_\mu(A(X_i) \mid \mathcal{J}_{\leqslant -n}(x)) =$$

$$(6.3) \qquad = \lim_{n\to\infty} P_\mu(A(X_i) \mid \mathcal{J}_{-n}(x)) = \lim_{n\to\infty} P^{n+i}(X_{-n}, A).$$

To complete the proof it will suffice to show that (6.2) and (6.3) agree P_μ-a.e. Let $g_n(x) = |\varphi_{n+i}(x, A) - P^{n+i}(x, A)|$. By Lemma 6.1 $g_n(x) \xrightarrow[n\to\infty]{} 0$ and by (6.2) and (6.3) the limit $g_n(X_{-n}) \xrightarrow[n\to\infty]{} g(x) \geqslant 0$ exists. But

$$\int P_\mu(dx)g(x) = \lim_{n\to\infty} \int P_\mu(dx)g_n(X_{-n}) =$$

$$= \lim_{n\to\infty} \int \mu(dx)g(x) = 0.$$

Thus $g(x) = 0$ and the proof is complete.

If a condition stronger than stability is satisfied, a more explicit formula for the conditional probabilities $P_\mu(dx \mid \mathcal{J}_{-\infty})$ can be obtained.

For any partition $\sigma \leqslant w^+$ and any fiber $a \in w^+$, let $\sigma(a)$ denote the fiber of σ containing a. Let us write $\sigma(T^n x)$ for $\sigma(T_+^n w^+(x))$. Note that for a (B, μ, T, π) process this notation is consistent.

We say that a stable partition x is uniformly stable if for every $A \in \mathcal{O}$ there exists a σ-measurable function $\varphi(x, A)$ $(\varphi(x, A) = \varphi(\sigma(x), A))$ such that

$$(6.4) \qquad \lim_{n\to\infty} (\varphi(\sigma(T^n x), A) - P^n(x, A)) = 0, \qquad \mu\text{-a.e.}$$

Note that if (B,μ,P) has a uniformly stable partition then w^+ is uniformly stable. We say that (B,μ,P) is uniformly stable if w^+ is uniformly stable.

Theorem 6.4. *Let* σ *be uniformly stable. Let* $f \in L^\infty(B,\mu)$. *Then*

$$(6.5) \qquad P_\mu(f(X_i) \mid \mathscr{I}_{-\infty}(x)) = \mu(f \mid \sigma(X_i)), \quad P_\mu\text{-}a.e.$$

and

$$(6.6) \qquad \mu(Pf \mid \sigma(x)) = \mu(f \mid \sigma(Tx)), \quad \mu\text{-}a.e.$$

i.e.

$$\mu(\cdot \mid \sigma(x))P = (\cdot \mid \sigma(Tx)), \quad \mu\text{-}a.e.$$

Proof. Note first that $\sigma(T^{n+i}X_{-n}) = \sigma(X_i)$, P_μ-a.e. Thus for $A \in \mathscr{O}$

$$\lim_{n\to\infty} \varphi(\sigma(T^{n+i}X_{-n}), A) = \lim_{n\to\infty} \varphi(\sigma(X_i), A) = \varphi(\sigma(X_i), A).$$

Therefore by the same argument as at the end of the proof of Theorem 6.3

$$(6.7) \qquad \begin{aligned} g(x) := P_\mu(A(X_i) \mid \mathscr{I}_{-\infty}(x)) &= \lim_{n\to\infty} P^{n+i}(X_{-n}, A) = \\ &= \lim_{n\to\infty} \varphi(\sigma(T^{n+i}X_{-n}), A) = \varphi(\sigma(X_i), A), \quad P_\mu\text{-a.e.} \end{aligned}$$

Thus $g(x)$ is σ_i measurable, and, since $\sigma_i \leqslant \mathscr{I}_{-\infty}$, we have that

$$P_\mu(A(X_i) \mid \mathscr{I}_{-\infty}(x)) = P_\mu(A(X_i) \mid \sigma_i(x)) = \mu(A \mid \sigma(X_i)), \quad P_\mu\text{-a.e.}$$

This proves (6.5).

Let $f \in L^\infty(B,\mu)$. Since $\mathscr{I}_{\leqslant 0} \geqslant \mathscr{I}_{-\infty}$,

$$(6.9) \qquad P_\mu(f(X_1) \mid \mathscr{I}_{-\infty}) = P_\mu(Pf(X_0) \mid \mathscr{I}_{-\infty}), \quad P_\mu\text{-a.e.}$$

by the Markov property. By (6.5) the left hand side of (6.9) equals

$$\mu(f \mid \sigma(X_1)) = \mu(f \mid \sigma(TX_0)), \quad P_\mu\text{-a.e.}$$

and the right hand side of (6.9) equals $\mu(Pf \mid \sigma(X_0))$, P_μ-a.e., which proves (6.6).

From Theorem 6.4 (and its proof) we have the following

Corollary. *Suppose* σ *is uniformly stable. Then* (mod 0)

(i) *For* $A \in \mathcal{O}$

$$\varphi_n(x, A) = \varphi(\sigma(T^n x), A) = \mu(A \mid \sigma(T^n x)),$$

(ii) $\sigma = w^+$.

There exist stable systems which are not uniformly stable:

Example 6.2. Let (B, μ, P) be the stationary process (B, μ, T, π) described in Example 6.1. Let $\boldsymbol{\epsilon} = (\epsilon_i)$, $0 \leqslant \epsilon_i \leqslant 1$, $i \in Z$, and let

$$I_{\boldsymbol{\epsilon}}(x, \cdot) = \mathop{\times}_{i=-\infty}^{+\infty} (\epsilon_i \delta_{-\xi_i} + (1 - \epsilon_i)\delta_{\xi_i}), \qquad x = (\xi_i).$$

$I_{\boldsymbol{\epsilon}}$ represents an independent spin flip $\xi_i \to -\xi_i$ at each "site i", with probability ϵ_i, $i \in Z$.

Let $P_{\boldsymbol{\epsilon}} = I_{\boldsymbol{\epsilon}}(Tx, \cdot)$. $P_{\boldsymbol{\epsilon}}$ represents a shift followed by independent spin flips. Let

$$\epsilon_i = \begin{cases} \frac{1}{2} & i < 0 \\ \frac{1}{4} & i = 0 \\ 0 & i > 0 \end{cases}.$$

Then $P_{\boldsymbol{\epsilon}}$, $\boldsymbol{\epsilon} = (\epsilon_i)$, is P followed by an independent flip at $i = 0$ with probability $\frac{1}{4}$. For the process $(B, \mu, P_{\boldsymbol{\epsilon}})$ $w^+ = \hat{\mathscr{I}}_{\geqslant 1}$ (see Example 6.1 for the notation). This process is clearly stable but not uniformly stable.

For a different choice of $\boldsymbol{\epsilon}$ in the above example we obtain a process which is not stable.

Example 6.3. Let $(B, \mu, P_{\boldsymbol{\epsilon}})$ be as in Example 6.2, now with $\boldsymbol{\epsilon} = (\epsilon_i)$ satisfying $\epsilon_i > 0$ for all $i \in Z$ and $\sum_{i=-\infty}^{+\infty} \epsilon_i < \infty$. (This process may be found in Rosenblatt [12] Ch. IV, Sec. 4.) $(B, \mu, P_{\boldsymbol{\epsilon}})$ is not stable.

In fact w^+ is trivial and $\mathscr{I}_{-\infty}$ is nontrivial. $\mathscr{I}_{-\infty}$ is nontrivial because, e.g., the functions

$$g_n^{(i)}(x) = \xi_{i+n}(-n),$$

where

$$x = (x_j), \quad x_j = (\xi_k(j)),$$

($g_n^{(i)}(x)$ is the value of the spin at site $i + n$ at time $-n$) are independent of n for n sufficiently large. Thus their limits $g_\infty^{(i)}(x) \in \mathscr{I}_{-\infty}$ are nontrivial. w^+ is trivial because, in fact, $\mathscr{I}_0 \cap \mathscr{I}_{-1}$ is trivial. This is so because for $f \in L^\infty(B, \mu)$, $f(X_0) \in \mathscr{I}_0 \cap \mathscr{I}_{-1} \Rightarrow f \in \hat{\mathscr{I}}_{\leqslant -m} \cup \hat{\mathscr{I}}_{\geqslant m}$ for all $m \geqslant 0$, so by the $0 - 1$ law for independent random variables, f is constant μ-a.e.

7. RELATIVIZED HARRIS CHAINS

In this section we study relativized Harris chains, which are defined by the condition that μ-a.e. $P^n(x, dy)$ has a component absolutely continuous with respect to $\mu(dy \mid w(T^n x))$ for some n. (We will frequently write T instead of T_w and we will write $w(T^n x)$ for $T_w^n(w(x))$, even if (B, μ, P) is not a (B, μ, T, π) chain.) This condition is quite a bit weaker than the Harris condition (see Section 5) and, in particular w may have uncountably many fibers. However the results for Harris chains, Theorem 5.2 (ii) and (iii), essentially extend and (B^Z, P_μ, S), if ergodic, can be factored into a direct product of $\left(\frac{B}{w}, \mu_w, T_w\right)$ and a Bernoulli scheme.

From now on we will write B_w for $\frac{B}{w}$.

We say (B, μ, P) factors if there exists a Lebesgue space $(B_\perp, \mu_\perp)$ and an isomorphism (mod 0) from $(\widetilde{B}, \widetilde{\mu}) := (B_w, \mu_w) \times (B_\perp, \mu_\perp)$ onto (B, μ) which carries the partition according to the first component of $\widetilde{B}$ ("vertical" fibers) to w.

Theorem 7.1. (B, μ, T, π) *factors if it is ergodic.*

$$P_{n+m}(a) = P_n(a)P_m(T^n a)$$

$$P^+_{n+m}(a) = P^+_a(a)P^+_m(T^{-n}a).$$

The basic properties of P_n and P^+_n and the relationship between them is given by

Theorem 7.2. *For* μ_w*-a.e.* $a \in w$,

$$\begin{aligned} &\mu(\cdot \mid a)P^n = \mu(\cdot \mid T^n a) \\ (7.2) \qquad & \\ &\mu(\cdot \mid a)P^{+n} = \mu(\cdot \mid T^{-n}a). \end{aligned}$$

Suppose (B, μ, P) *factors. Then for* μ_w*-a.e.* $a \in w$,

$$\begin{aligned} &\mu_\perp P_n(a) = \mu_\perp \\ (7.3) \qquad & \\ &\mu_\perp P^+_n(a) = \mu_\perp \end{aligned}$$

so that $P_n(a)$ *and* $P^+_n(a)$ *are contractions on* L^p, $1 \leqslant p \leqslant \infty$, *leaving* $\mu_\perp$ *invariant. Moreover* P_n *and* P^+_n *are adjoint:*

$$(7.4) \qquad \int d\mu_\perp g P_n(a) f = \int d\mu_\perp f P^+_n(T^n a) g, \quad \mu_w\text{-a.e.}$$

for $f \in L^\infty(B_\perp, \mu_\perp)$ *and* $g \in L^1(B_\perp, \mu_\perp)$, *i.e.*,

$$(P_n(a))^+ = P^+_n(T^n a).$$

Proof. (7.2) follows easily from the stationary of μ. (7.4) follows from the fact that P^n and P^{+n} are adjoint. Note that for $f \in L^\infty(B, \mu)$, $f = f(a, \xi)$, $P^n = P_n U^n_w$, where

$$(P_n f)(a, \xi) := (P_n(a) f(a, \cdot))(\xi)$$

$$(U^n_w f)(a, \xi) = f(T^n a, \xi).$$

Similarly

$$P^{+n} = P^+_n U^{-n}_w,$$

where

$$(P^+_n f)(a, \xi) := (P^+_n(a) f(a, \cdot))(\xi).$$

Proof. It follows from Rohlin [2] Sec. 4, No. 1, that if each of the Lebesgue spaces $(w(x), \mu(\cdot \mid w(x)))$, $x \in B$, has (mod 0) the same isomorphism class, then (B, μ, P) factors. There Rohlin also shows that the isomorphism class of $(w(x), \mu(\cdot \mid w(x)))$ depends measurably on x. For a (B, μ, T, π) scheme, T induces an isomorphism between $(w(x), \mu(\cdot \mid w(x)))$ and $(w(Tx), \mu(\cdot \mid w(Tx)))$, so that in this case the isomorphism class is a constant of the motion. Thus if (B, μ, T, π) is ergodic, the isomorphism class is the same for all x and so the theorem is proven.

Suppose (B, μ, P) factors. Then we may identify (B, μ) with $(\widetilde{B}, \widetilde{\mu})$ and drop the "$\sim$". Thus in this case $x \in B$ is of the form $x = (a, \xi)$ $(a \in B_w,\ \xi \in B_\perp)$, $w(x) = (a, B_\perp)$, and we may set $\mu(\cdot \mid w(x)) = \delta_a \times \mu_\perp$. Moreover the system (B, μ, P) may now be factored into a "skew product" whose first component is (B_w, μ_w, T_w). The transition probability $P(x, \cdot)$ defined for $\widetilde{\mu}$-a.e. x, induces a family of transition probabilities $P_n(\xi, \cdot\,; a)$ on $(B_\perp, \mu_\perp)$ defined, for μ-a.e. $x = (a, \xi)$, by

$$P^n(x, \cdot) = \delta_{T^n a} \times P_n(\xi, \cdot\,; a), \qquad \mu\text{-a.e.}$$

Similarly we define $P_n^+(\xi, \cdot\,; a)$ by

$$P^{+n}(x, \cdot) = \delta_{T^{-n} a} \times P_n^+(\xi, \cdot\,; a), \qquad \mu\text{-a.e.}$$

We may assume, by removing, if necessary, an appropriate set of μ_w measure zero from B_w, that for all $a \in B_w$, $P_n(\xi, \cdot\,; a)$ and $P_n^+(\xi, \cdot\,; a)$ are defined for $\mu_\perp$-a.e. ξ, as measures on $\Sigma_\perp$. (Here $\Sigma_\perp$ is the "Borel" σ-algebra on $B_\perp$. We may assume that $\Sigma = \Sigma_w \times \Sigma_\perp$, where (B_w, Σ_w) and $(B_\perp, \Sigma_\perp)$ are standard Borel spaces). Moreover, for $A \in \Sigma_\perp$ and $a \in B_w$,

(7.1) $$P_{n+m}(\xi, A; a) = \int P_n(\xi, d\eta; a) P_m(\eta, A; T^n a), \qquad \mu_\perp\text{-a.e.},$$

(7.2) $$P_{n+m}^+(\xi, A; a) = \int P_n^+(\xi, d\eta; a) P_m^+(\eta, A; T^{-n} a), \qquad \mu_\perp\text{-a.e.}$$

For $f \in L^\infty(B_\perp, \mu_\perp)$ we write $(P_n(a)f)(\xi)$ for $P_n(\xi, f; a)$ and similarly for P_n^+. Thus (7.1) may be written as

Moreover, the convergence in (7.7) *and* (7.8) *is extendable.*

Proof. (7.5) and (7.6) are proven in Appendix A. Suppose $f(\xi)$ and $g(\xi)$ are as in the theorem. Then from the L^1 convergence of (7.5), one can conclude that there exists a subsequence $\{n_k^{(1)}\}$ for which (7.7) holds, and is extendable. Moreover, given $\bar{f}(\xi)$ and $\bar{g}(\xi)$ there exists a subsequence $\{n_k^{(2)}\}$ of $\{n_k^{(1)}\}$ for which (7.7) also holds for $\bar{f}$ and $\bar{g}$ and is still extendable. By a diagonalization argument one may obtain a subsequence $\{n_k\}$ for which (7.7) holds, extendably, for all f and g chosen from a countable subset of $L^\infty(B_\perp, \mu_\perp)$ dense in $L^1(B_\perp, \mu_\perp)$. By a density argument the convergence can be extended, first to all $g \in L^\infty(B_\perp, \mu_\perp)$, then to all $f \in L^1(B_\perp, \mu_\perp)$. The proof of (7.8) is similar.

Definition 7.1. We say that the stationary Markov chain (B, μ, P) is a *relativized Harris chain* if for μ-a.e. $x \in B$ there exists an $n = n(x)$ such that $P^n(x, \cdot)$ has a component absolutely continuous with respect to $\mu(\cdot \mid w(T^n x))$.

Note that if (B, μ, P) factors, it forms a relativized Harris chain if and only if for μ-a.e. $x = (a, \xi)$ there exists an $n = n(x)$ such that $P_n(\xi, \cdot\,; a)$ has a component absolutely continuous with respect to $\mu_\perp$.

Until we indicate otherwise we will assume that (B, μ, P) factors. We will follow as far as convenient the development in Section 5.

Let $P_n(a) = Q_n(a) + R_n(a)$ be the decomposition of $P_n(a)$ into its absolutely continuous part, $Q_n(a)$, (with respect to $\mu_\perp$) and its singular part. We write $\hat{P}_n(x, \cdot)$, $\hat{Q}_n(x, \cdot)$ and $\hat{R}_n(x, \cdot)$ for $P_n(\xi, \cdot\,; a)$ $Q_n(\xi, \cdot\,; a)$ and $R_n(\xi, \cdot\,; a)$ respectively, $x = (a, \xi)$. $\hat{Q}_n$ and $\hat{R}_n$ are measurable, i.e., $\hat{Q}_n f$ and $\hat{R}_n f$ are measurable in B for $f \in L^\infty(B_\perp, \mu_\perp)$. To see this note that

$$\hat{Q}_n(x, dy) = \hat{q}_n(x, y)\mu_\perp(dy),$$

where

$$\hat{q}_n(x, y) = \frac{d(\mu \times \hat{P}_n)_{\text{a.c.}}}{d(\mu \times \mu_\perp)}.$$

We write $q_n(\xi, \eta; a)$ for $\hat{q}_n(x, y)$, $x = (a, \xi)$, so that

Thus

$$P_n^+ U_w^{-n} = P^{+n} = (P^n)^+ = (P_n U_w^n)^+ = U_w^{-n}(P_n)^+ =$$
$$= (U_w^{-n}(P_n)^+ U_w^n) U_w^{-n},$$

and

$$P_n^+ = U_w^{-n}(P_n)^+ U_w^n,$$

i.e.

$$P_n^+(a) = P_n(T^{-n}a)^+,$$

since

$$(U_w^{-n}(P_n)^+ U_w^n f)(a, \xi) = ((P_n(T^{-n}a))^+ f(a, \cdot))(\xi).$$

In the next lemma we strengthen the convergence given by Theorem 4.1, which will later be further strengthened using the "relativized Harris condition". Suppose $\{a_n\}$ is a sequence of numbers and suppose $a_{n_k} \xrightarrow[k \to \infty]{} a$. We say the convergence $a_{n_k} \xrightarrow[k \to \infty]{} a$ is *extendable* if $a_{m_k} \xrightarrow[k \to \infty]{} a$, where $\{m_k\}$ is obtained from $\{n_k\}$ by inserting the next k integers before and after each n_k.

Lemma 7.1. *Let* $f \in L^1(B, \mu)$, $g \in L^\infty(B, \mu)$ *and* $g^\circ = g - Eg$ *where* E *is the conditional expectation with respect to* w. *Then*

$$\lim_{n \to \infty} \int \mu(dx) \left| \int d\mu(\cdot \mid w(x)) f P^n g^\circ \right| = 0 \tag{7.5}$$

$$\lim_{n \to \infty} \int \mu(dx) \left| \int d\mu(\cdot \mid w(x)) f P^{+n} g^\circ \right| = 0. \tag{7.6}$$

Furthermore if (B, μ, P) *factors, there exists a subsequence* $\{n_k\}$ *such that for every* $f \in L^1(B_1, \mu_1)$, $g \in L^\infty(B_1, \mu_1)$, $g^\circ = g - Eg$,

$$\lim_{n \to \infty} \int d\mu_1 f P_{n_k}(a) g^\circ = 0, \quad \mu_w\text{-a.e.} \tag{7.7}$$

and

$$\lim_{n \to \infty} \int d\mu_1 f P_{n_k}^+(a) g^\circ = 0, \quad \mu_w\text{-a.e.} \tag{7.8}$$

(ii) *If* (B, μ, P) *is r-Harris, then* $R_n 1 \searrow 0$, *μ-a.e.*

(iii) (B, μ, P) *is r-Harris* $\Longleftrightarrow (B, \mu, P^+)$ *is r-Harris.*

Proof. We will use the following notation for (B, μ, P^+). We write $P_n^+(a) = Q_n^*(a) + R_n^*(a)$ for the decomposition of $P_n^+(a)$ into its absolutely continuous and singular parts;

$$Q_n^*(\xi, d\eta; a) := q_n^*(\xi, \eta; a)\mu_\perp(d\eta).$$

Corresponding to (7.12) we have the decomposition

$$P^{+n} = Q_n^* + R_n^*.$$

The proofs of (i) and (ii) are identical to the proofs of Theorem 5.1 (i) and (ii). The proof of (iii) is similar to that of Theorem 5.1 (iii): Note that it follows from

$$(P_n(a))^+ = P_n^+(T^n a)$$

that

$$(Q_n(a))^+ = Q_n^*(T^n a). \tag{7.16}$$

In fact if $P^{(1)}$ and $P^{(2)}$ are transition probabilities on $(B_\perp, \mu_\perp)$, and if $Q^{(1)}$ and $Q^{(2)}$ are the absolutely continuous parts of $P^{(1)}$ and $P^{(2)}$ respectively, then

$$P^{(2)} = P^{(1)+} \Rightarrow Q^{(2)} = Q^{(1)+}:$$

$$P^{(1)}(\xi, d\eta) = q^{(1)}(\xi, \eta)\mu_\perp(d\eta) + R^{(1)}(\xi, d\eta) \Rightarrow$$

$$\Rightarrow P^{(1)+}(\xi, d\eta) = q^{(1)}(\eta, \xi)\mu_\perp(d\eta) + R^{(1)+}(\xi, d\eta).$$

Thus $q^{(2)}(\xi, \eta) \geqslant q^{(1)}(\eta, \xi)$ i.e. $Q^{(2)} \geqslant Q^{(1)+}$. Similarly, $Q^{(1)} \geqslant Q^{(2)+}$. Thus $Q^{(1)+} = Q^{(2)}$. It follows from (7.16) that

$$Q_n^+ = Q_n^* \tag{7.17}$$

(see e.g. the proof of Theorem 7.2).

Thus for $f \in L^\infty(B, \mu)$, $0 \leqslant f \not\equiv 0$,

$$Q_n(\xi, d\eta; a) = q_n(\xi, \eta; a)\mu_1(d\eta).$$

Just as in Section 5,

$$(7.9) \qquad P_m(a)Q_n(T^m a) \leqslant Q_{m+n}(a)$$

$$(7.10) \qquad Q_m(a)P_n(T^m a) \leqslant Q_{m+n}(a)$$

$$(7.11) \qquad R_m(a)R_n(T^m a) \geqslant R_{m+n}(a).$$

We find it convenient to use the operators $Q_n(x, dy)$ and $R_n(x, dy)$: for $f \in L^\infty(B, \mu)$, $f = f(a, \xi)$,

$$(Q_n f)(a, \xi) = (Q_n(a) f(T^n a, \cdot))(\xi),$$

$$(R_n f)(a, \xi) = (R_n(a) f(T^n a, \cdot))(\xi).$$

Q_n and R_n provide a natural decomposition of P^n (recall that

$$(P^n f)(a, \xi) = (P_n(a) f(T^n a, \cdot))(\xi))$$

$$(7.12) \qquad P^n = Q_n + R_n.$$

Note that (B, μ, P) is a relativized Harris chain iff μ-a.e.

$$\sum_{n=1}^\infty Q_n 1 > 0.$$

From (7.9), (7.10) and (7.11) we obtain:

$$(7.13) \qquad P^m Q_n \leqslant Q_{m+n}$$

$$(7.14) \qquad Q_m P^n \leqslant Q_{m+n}$$

$$(7.15) \qquad R_m R_n \geqslant R_{m+n}.$$

Some basic preliminary facts about relativized Harris chains are collected in the following: (we will say r-Harris for relativized Harris)

Theorem 7.3. *Suppose* (B, μ, P) *factors.*

(i) (B, μ, P) *is r-Harris* $\Longleftrightarrow \sum_{n=1}^\infty Q_n f \not\equiv 0$ *for all* $0 \leqslant f \not\equiv 0$, $f \in L^\infty(B, \mu)$.

Lemma 7.3. *Suppose* (B, μ, P) *factors and is* r*-Harris. Let* $g \in L^\infty(B_\perp, \mu_\perp)$. *Then there exists a subsequence* $\{n'_k\}$ *of* $\{n_k\}$ *such that* (mod 0)

$$\lim_{k \to \infty} P^+_{n'_k}(T^{n'_k} a) g^\circ = 0.$$

Proof. Let $f_n(x) = (P^+_n(a) g^\circ)(\xi)$, $x = (a, \xi)$. By Lemma 7.2 $f_{n_k}(x) \to 0$, μ-a.e. Thus

$$\| U_w^{n_k} f_{n_k} \|_1 = \| f_{n_k} \|_1 \to 0.$$

Therefore there exists a subsequence $\{n'_k\}$ of $\{n_k\}$ for which $U_w^{n'_k} f_{n'_k} = f_{n'_k}(T^{n'_k} a, \xi) = (P^+_{n'_k}(T^{n'_k} a) g^\circ)(\xi) \to 0$, μ-a.e.

Lemma 7.4. *Suppose* (B, μ, P) *factors and is* r*-Harris. Let* $g \in L^1(B_\perp, \mu_\perp)$ *and let* $d\mu_{g^\circ} = g^\circ d\mu_\perp$. *Then for* μ_w*-a.e.* a,

$$\mu_{g^\circ} P_n(a) \to 0 \quad \textit{in variation norm}$$

(i.e. $\lim\limits_{n \to \infty} \sup\limits_{\|f\|_\infty \leq 1} | \int d\mu_\perp g^\circ P_n(a) f | = 0$*).*

Proof. Using Lemma 7.3,

$$\| P^+_{n'_k}(T^{n'_k} a) g^\circ \|_1 \to 0$$

for g bounded. Since the $P^+_n(a)$'s are contractions on $L^1(B_\perp, \mu_\perp)$, this convergence can be extended to all $g \in L^1(B_\perp, \mu_\perp)$ with the restriction to the subsequence $\{n'_k\}$ removed, i.e., $\| P^+_n(T^n a) g^\circ \|_1 \to 0$ for all $g \in L^1(B_\perp, \mu_\perp)$. Since $P^+_n(T^n a) = P_n(a)^+$ this is the same as the conclusion of the lemma.

Theorem 7.4. *Suppose* (B, μ, P) *factors and is* r*-Harris. Then for* μ*-a.e.* $x = (a, \xi)$

$$P_n(\xi, \cdot\,; a) \to \mu_\perp \quad (n \to \infty) \quad \textit{in variation norm.}$$

Proof. It follows from Lemma 7.4 that for $g \in L^1(B_\perp, \mu_\perp)$, for μ_w-a.e. a

$$\sum_{n=1}^{\infty} Q_n f \not\equiv 0 \Rightarrow \sum_{n=1}^{\infty} (\mu Q_n)(f) \neq 0 \Rightarrow$$

$$\Rightarrow \sum_{n=1}^{\infty} (f, Q_n^* 1) \equiv \left(f, \sum_{n=1}^{\infty} Q_n^* 1\right) \neq 0$$

from which (iii) follows easily.

Corollary 7.1. *Suppose* (B, μ, P) *factors and is* r*-Harris. Then*

(7.18) $\quad \lim_{n \to \infty} R_n(\xi, B_\perp; a) = 0, \quad \text{mod } 0,$

and

(7.19) $\quad \lim_{n \to \infty} R_n^*(\xi, B_\perp; a) = 0, \quad \text{mod } 0.$

Our goal now is to use Lemma 7.1 and the above corollary to prove the counterpart of Theorem 5.2 for r-Harris chains. The basic idea is very similar to that of Theorem 5.2.

Lemma 7.2. *Suppose* (B, μ, P) *factors and is* r*-Harris. Let* $\{n_k\}$ *be the subsequence described in Lemma* 7.1. *Let* $g \in L^\infty(B_\perp, \mu_\perp)$. *Then* (mod 0)

(7.20) $\quad \lim_{n \to \infty} P_{n_k}^+(a) g^\circ = 0.$

Proof. Using (7.1) we have

$$P_n^+(a) g^\circ = P_m^+(a) P_{n-m}^+(T^{-m} a) g^\circ =$$
$$= Q_m^*(a) P_{n-m}^+(T^{-m} a) g^\circ + R_m^*(a) P_{n-m}^+(T^{-m} a) g^\circ .$$

Thus

$$\begin{aligned} |P_n^+(a) g^\circ| \leqslant {} & \left| \int \mu_\perp(d\eta) q_m^*(\cdot, \eta; a)(P_{n-m}^+(T^{-m} a) g^\circ)(\eta) \right| + \\ & + 2\|g\|_\infty R_m^*(\cdot, B_\perp; a). \end{aligned} \tag{7.21}$$

Using Corollary 7.1 we can fix m so large that the second term on the right hand side of (7.21) is smaller than ϵ. Then, by Lemma 7.1, the first term on the right hand side, with $n = n_k$, will also be smaller than ϵ for k sufficiently large.

rise to a partition Z_0 of B^Z (the time 0 partition) which is very weak Bernoulli relative to w [13] and hence [13], [14] is relatively finitely determined with respect to w. Thus the Thouvenot's relative theory [15], [14] the factor of (B^Z, P_μ, S) induced by $Z \cup w$, i.e. the dynamical system $\left(B^Z, w \cup \bigcup_{n=-\infty}^{+\infty} S^n Z_0, P_\mu, S\right)$, is a direct product of (B_w, μ_w, T_w) with a Bernoulli shift. Choosing Z finer and finer, we see that (B^Z, P_μ, S) is the increasing limit of factors which are direct products of (B_w, μ_w, T_w) with Bernoulli shifts and hence is itself a direct product of (B_w, μ_w, T_w) with a generalized Bernoulli shift [14].

(iv) follows immediately from (iii).

Corollary.

(i) *Suppose* (B, μ, T) *is Bernoulli. Then* (B, μ, T, π) *is Bernoulli if it is r-Harris.*

(ii) *Suppose the stationary stochastic process* $\{Y_t\}$ *is Bernoulli. Then its Markov approximation is Bernoulli if it is r-Harris.*

ACKNOWLEDGEMENTS

It is a pleasure to acknowledge very useful discussions with M. Aizenmann, H. Spohn and D. Ornstein.

APPENDIX A

We consider a (B, μ, T, π) scheme. Recall that:

$$Uf = f(Tx), \qquad K(x, dy) = \mu(dy \mid \pi(x)), \qquad P = UK.$$

Let $\|\cdot\| = \|\cdot\|_{L^2(B,\mu)}$. Suppose $f \in L^2(B, \mu)$, $g \in L^2(B, \mu)$.

Lemma A.1.

(i) $\|P^n f\|$ *is non-increasing.*

(ii) $\sum_{n=0}^{\infty} \|(1-K)P^n f\|^2 < \infty.$

(7.21) $\mu_g P_n(a) \to \left(\int g d\mu_\perp\right) \mu_\perp$ in variation norm.

Since

$$P_n(\xi, \cdot; a) = \int \mu_\perp(d\eta) q_m(\xi, \eta; a) P_{n-m}(\eta, \cdot; T^m a) +$$

$$+ R_m(\xi, d\eta; a) P_{n-m}(\eta, \cdot; T^m a)$$

the theorem follows from (7.21) and (7.18).

The assumption that (B, μ, P) factors can be removed.

Theorem 7.5. *Suppose* (B, μ, P) *is r-Harris. Then*

(i) $\lim_{n \to \infty} \|P^n(x, \cdot) - \mu(\cdot \mid w(T^n x))\| = 0$

where "$\|\cdot\|$" *is the variation norm.*

(ii) (B, μ, P) *is uniformly stable and* $w^+ = w$ *(see Section 6). Thus*

(7.22) $w = \mathscr{I}_{-\infty} = \mathscr{I}_\infty$.

(iii) *If* (B, μ, P) *is also ergodic,* (B^Z, P_μ, S) *is isomorphic to the direct product of* (B_w, μ_w, T_w) *with a (generalized) Bernoulli shift* [4] *(and under this isomorphism the first component* (B_w, μ_w, T_w) *of the product corresponds to itself).*

(iv) *If the deterministic part* (B_w, μ_w, T_w) *of* (B, μ, P) *is Bernoulli, then so is* (B^Z, P_μ, S).

Proof.

(i) By Theorems 7.1 and 7.4, (i) will hold for any ergodic (B, μ, T, π) chain. By decomposition into ergodic components (i) holds for any (B, μ, T, π) chain $(w \geqslant \tau)$. Since any stationary Markov chain (B, μ, P) is a factor of a $(\bar{B}, \bar{\mu}, T, \pi)$ chain (see Example 1.3) such that w for $(\bar{B}, \bar{\mu}, T, \pi)$ agrees with w for (B, μ, P), (i) follows.

(ii) follows easily from (i). ((7.22) follows e.g. from Theorem 7.3 and Theorem 6.4.)

(iii) It follows from (i) that every finite partition Z of B gives

Lemma A.3. *For* μ*-a.e.* $x \in B$

$$\text{(A.1)} \qquad \lim_{n \to \infty} \left[P^n(x,f) - \int \mu(dy \mid \pi_m(x)) P^n(y,f) \right] = 0$$

where $\pi_m = T^m \pi, \; m \in Z$.

Proof. Case $m = 0$. The function in the square bracket on the left hand side of (A.1) is $(1-K)P^n f$. By (ii) of Lemma A.1 $(1-K)P^n f$ converges μ-a.e. to 0.

Case $m \neq 0$. The operator $U^m K U^{-m}$ is the conditional expectation with respect to π_m. Use of Lemma A.2 (iii) gives then the result.

Proof of Theorem 6.2. It is obtained from Lemma A.3 by letting f vary in $\mathcal{O}$.

We put

$$\mathcal{H}_{\pm p} = \{ f \in L^2(B, \mu) \mid U^{\pm p} f = (U^{\pm p} K)^p f \}, \qquad p \geqslant 0$$

$$\mathcal{H}^p = \mathcal{H}_p \cap \mathcal{H}_{-p}, \qquad p \geqslant 0$$

$$\mathcal{H} = \bigcap_{n \geqslant 0} \mathcal{H}^n .$$

$\mathcal{H}_p$, $p \geqslant 0$, is the subspace of functions f such that $U^i f$ is π measurable for $0 \leqslant i \leqslant p-1$, i.e. of functions f which are

$$\bigcap_{i=0}^{p-1} T^i \pi$$

measurable. $\mathcal{H}_{-p}$, $p \geqslant 0$ has an analogous interpretation. Let

$$\pi^n = \bigcap_{i=-n+1}^{n-1} T^i \pi, \qquad n > 0.$$

Then $\mathcal{H}^n = L^2(B, \pi^n, \mu)$; $\mathcal{H} = L^2(B, w, \mu)$; E_n, the orthogonal projection on $\mathcal{H}^n$, is the conditional expectation with respect to π^n; and E, the conditional expectation with respect to w, is the orthogonal projection on $\mathcal{H}$, and is therefore the strong limit of the E_n.

Let $\mathcal{H}^{\perp}_{\pm p}$, $p \geqslant 0$, be the orthogonal complement of $\mathcal{H}_{\pm p}$. Note that

Proof. Statements (i) and (ii) are consequence of the following

$$\|P^n f\|^2 = \|KP^n f\|^2 + \|(1-K)P^n f\|^2 =$$
$$= \|UKP^n f\|^2 + \|(1-K)P^n f\|^2 =$$
$$= \|P^{n+1} f\|^2 + \|(1-K)P^n f\|^2$$

because U is a unitary operator in L^2 and K is an orthogonal projection.

Lemma A.2. *For any* $q = 1, 2, \ldots$

(i) $\displaystyle\sum_{n=0}^{\infty} \|((UK)^q - U^q)P^n f\|^2 < \infty.$

(ii) $\displaystyle\sum_{n=0}^{\infty} \|((U^{-1}K)^q - U^{-q})P^n f\|^2 < \infty.$

(iii) $\displaystyle\sum_{n=0}^{\infty} \|(U^{\pm q} K U^{\mp q} - 1)P^n f\|^2 < \infty.$

Proof. The proof makes repeated use of (ii) of Lemma A.1. For instance when we have a term like $U^q P^n f$ we can write it as follows:

$$U^q P^n f = U^{q-1} U((1-K) + K)P^n f =$$
$$= U^{q-1} P^{n+1} f + U^q (1-K)P^n f =$$
$$= \sum_{i=0}^{q-1} U^{q-i}(1-K)P^{n+i} + P^{n+q} f.$$

Therefore

$$\|(UK)^q - U^q)P^n f\| \leqslant \sum_{i=0}^{q-1} \|(1-K)P^{n+i} f\| \leqslant$$
$$\leqslant q^{\frac{1}{2}} \Big(\sum_{i=0}^{q-1} \|(1-K)P^{n+i} f\|^2 \Big)^{\frac{1}{2}};$$

$$\sum_{n=0}^{\infty} \|(UK)^q - U^q)P^n f\|^2 \leqslant q \sum_{n=0}^{\infty} \sum_{i=0}^{q-1} \|(1-K)P^{n+i} f\|^2,$$

which proves (i). The other statements are proven similarly.

$$\mathscr{H}^{\perp}_{\pm p} = \text{the closure of } \{g \in L^2\ g = [\ U^{\mp p} f - (KU^{\mp 1})^p f],$$
$$f \in L^2(B, \mu)\}$$

and that

$$\mathscr{H}'^p = (\mathscr{H}'^{\perp}_p + \mathscr{H}'^{\perp}_{-p})^{\perp}.$$

Lemma A.4. *Let* $\psi \in L^\infty(B, w, \mu)$. *Then*

$$E(\psi f) = \psi E f$$

$$K(\psi f) = \psi K f$$

$$U(\psi f) = (U\psi)(Uf).$$

Proof. Since E and K are conditional expectations and $EK = KE = E$ the first two equalities follow. The third one is a consequence of the fact that U is generated by the point transformation T.

Lemma A.5.

$$\text{(A.2)} \qquad \lim_{n \to \infty} \sup_{\substack{\|\psi\|_\infty \leq 1 \\ \psi \in w}} |((1 - E)g\psi, P^n f)| = 0.$$

Proof. Fix $\epsilon > 0$. Then there is p such that

$$\|(E - E_p)g\|_2 < \epsilon.$$

Then for every $n > 0$ we have by Lemma A.4:

$$\text{(A.3)} \qquad |((1 - E)g\psi, P^n f)| \leq |(\psi(1 - E_p)g, P^n f)| + \epsilon \|f\|_2$$

because $\|\psi\|_\infty \leq 1$. Since

$$(1 - E_p)g = (U^{-p} - (KU^{-1})^p)g_p + (U^p - (KU)^p)g_{-p} + \widetilde{g}$$

where $\|\widetilde{g}\| < \epsilon$, from (A.3) we get

$$|((1 - E)g\psi, P^n f)| \leq 2\epsilon\|f\| + \|g_p\|\,\|(U^p - (UK)^p)P^n f\| +$$
$$+ \|g_{-p}\|\,\|(U^{-p} - (U^{-1}K)^p P^n f\| \leq 2\epsilon\|f\| + \epsilon$$

for n large enough, because of Lemma A.2 (i) and (ii).

Proof of (7.5) and (7.6). For a (B, μ, T, π) scheme (7.5) is obtained from Lemma A.5 because of the arbitrariness of ψ. Because every stationary Markov chain (B, μ, P) is a factor of a $(\bar{B}, \bar{\mu}, T, \pi)$ chain, see Example 1.3, such that w for $(\bar{B}, \bar{\mu}, T, \pi)$ agrees with w for (B, μ, P), we obtain (7.5) and (7.6) for any (B, μ, P).

REFERENCES

[1] V.I. Arnold – A. Avez, *Ergodic problems of classical mechanics,* Benjamin, New York, 1968.

[2] V.A. Rohlin, On the fundamental ideas of measure theory, *Transl. Am. Math. Soc.*, (1) 10 (1962), 1-54.

[3] C. Marchioro – A. Pellegrinotti – E. Presutti – M. Pulvirenti, On the dynamics of particles in a bounded region: a measure theoretical approach, *J. Math. Phys.*, 17 (1976), 647-652.

[4] D.S. Ornstein, *Ergodic theory, randomness and dynamical systems,* Yale University Press, New Haven and London, 1974.

[5] W. Ambrose, *Ann. Math.*, 42 (1941), 3.
W. Ambrose – S. Kakutani, *Duke Math. J.*, 9 (1942), 25.

[6] S. Goldstein J.L. Lebowitz – E. Presutti, Mechanical systems with stochastic boundaries, (in this volume).

[7] R. Zwanzig, Contribution in Lectures in theoretical physics, Boulder, University of Colorado, 1960.

[8] D. Revuz, *Markov chains,* North-Holland, Amsterdam and Oxford, 1975.

[9] S.R. Foguel, *The ergodic theory of Markov processes,* Van Nostrand, New York, 1969.

[10] J. Feldman, New K-automorphisms and a problem of Kakutani, *Israel J. Math.*, 24 (1976), 1.

[11] D.S. Ornstein – P.C. Shields, Mixing Markov shifts of kernel type are Bernoulli, *Adv. Math.*, 10 (1973), 143-146.

[12] M. Rosenblatt, *Markov processes. Structure and asymptotic behavior,* Springer-Verlag, Berlin – Heidelberg – New York, 1971.

[13] J. Thouvenot, *Israel J. Math.*, 21 (1975), 215.

[14] D.A. Lind, *Israel J. Math.*, 28 (1977), 205.

[15] J. Thouvenot, *Israel J. Math.*, 21 (1975), 177.

[16] M. Kern – R. Nagel – G. Palm, Dilations of positive operators: construction and ergodic theory, *Math. Z.*, 156 (1977), 262-272.

[17] L. Breiman, *Probability,* Addison-Wesley, London, 1968.

[18] S. Kalikow, (to appear).

S. Goldstein – J.L. Lebowitz

Department of Mathematics, Rutgers University, New Brunswick, N.J. 08903, USA.

E. Presutti

Istituto Matematico, Università di Roma, 00100 Roma, Italy.