*The Annals of Probability*
1986, Vol. 14, No. 4, 1262–1271

# REGENERATIVE REPRESENTATION FOR ONE-DIMENSIONAL GIBBS STATES[1]

BY S. P. LALLEY

*Columbia University*

It is shown that one-dimensional Gibbs states may be represented as concatenations of infinite lists of iid "words." It follows that Gibbs states inherit many properties of recurrent Markov chains on denumerable state spaces.

**1. Introduction.** The purpose of this paper is to exhibit a regenerative representation for the class of stationary processes determined by the so-called *Gibbs states* (also known as *DLR states*) in one dimension. These processes have proved to be of central importance in topological dynamics and ergodic theory: see Bowen (1975) or Ruelle (1978) for a full account. Roughly, the connection is as follows. For many dynamical systems $(M, T, \mu)$, where $\mu$ is an invariant probability measure for the transformation $T$: $M \to M$, it is possible to find a "smooth" partition $\bigcup_{i=1}^m K_i$ of $M$ so that the distribution of the process

$$Y_n = \sum_{i=1}^{m} i1\{T^n x \in K_i\}$$

is a Gibbs state when $x$ has distribution $\mu$. This partition may also be chosen in such a way that almost every $x \in M$ corresponds to a unique orbit in a certain closed subset of $\{1, 2, \dots, m\}^{\mathbf{Z}}$, and vice versa. The class of dynamical systems for which such a partition exists includes many Anosov and Axiom A diffeomorphisms $T$: $M \to M$, ergodic automorphisms of compact abelian groups, and expansive maps of the unit interval, among others.

One of the celebrated results of ergodic theory has it that the dynamical system determined by a Gibbs state and the forward shift on $\{1, 2, \dots, m\}^{\mathbf{Z}}$ is isomorphic to a Bernoulli shift of the same entropy. The existence of such an isomorphism is commonly interpreted as a manifestation of the randomness or chaotic behavior inherent in such systems. (However, it should be kept in mind that many probabilistic properties, e.g., the central limit theorem and law of the iterated logarithm for additive functionals, are *not* invariant under such isomorphisms.)

The isomorphism theorem may be thought of as follows. A system $(M, T, \mu)$ which is isomorphic to a Bernoulli shift may be simulated by observing the output of a double-ended sequence of coin tosses and then coding this output to obtain a point of $M$. Unfortunately, one must generally know the entire sequence of coin tosses to determine the value of a corresponding point of $M$. In many

Received August 1984; revised July 1985.
[1]Partially supported by NSF grant DMS 82-01723.
*AMS* 1980 *subject classifications.* 60G10, 60K35.
*Key words and phrases.* Gibbs state, chain with complete connections, regenerative representation.

circumstances there do exist finitary codes, for which the value of a point in $M$ is determined by a finite segment of the coin-tossing experiment. However, it has been shown by Parry (1979) that typically there is no code with finite *expected* coding time whose inverse also has finite expected coding time; and recent results of Krieger (1983) suggest that typically there may not exist a code with finite expected coding time.

The main result of this paper states that a Gibbs process may always be simulated by stringing together iid "words" of symbols from the underlying alphabet, with the word length random variable having finite *exponential* moments. Equivalently, a Gibbs process may always be realized as a function of a recurrent (countable state) Markov chain whose recurrence times have exponentially decaying tails. This representation will make it apparent that many properties of recurrent Markov chains are inherited by Gibbs processes.

The possibility of obtaining such a representation was suggested to me by a construction of Athreya and Ney (1978) and Nummelin (1978) for Harris recurrent Markov chains. The details are quite different in this case, however, because of the infinite dependence.

**2. Statement of principal results.** A *chain with complete connections* is a stationary process $\{Y_n\}_{n \in \mathbf{Z}}$ taking values in a finite state space $\mathscr{Y}$ such that

$$P(Y_1 = \xi_1, \ldots, Y_n = \xi_n) > 0 \qquad \forall \xi_1, \ldots, \xi_n \in \mathscr{Y}, \tag{2.1}$$

$$\begin{aligned} &\lim_{m \to \infty} P(Y_0 = \xi_0 | Y_n = \xi_n, -m \le n \le -1) \\ &\quad = P(Y_0 = \xi_0 | Y_n = \xi_n, n \le -1) \text{ exists for all } \xi_0, \xi_{-1}, \xi_{-2}, \ldots, \in \mathscr{Y}; \end{aligned} \tag{2.2}$$

and

$$\gamma_m \downarrow 0, \tag{2.3}$$

where

$$\begin{aligned} \gamma_m \triangleq \sup\Bigg\{ &\left| \frac{P(Y_n = \xi_n, 0 \le n \le r | Y_n = \xi_n, n \le -1)}{P(Y_n = \xi_n, 0 \le n \le r | Y_n = \xi_n^*, n \le -1)} - 1 \right| : \\ &r < \infty;\ \xi_j, \xi_j^* \in \mathscr{Y}; \text{ and } \xi_n = \xi_n^* \ \forall n, -m \le n \le -1 \Bigg\}. \end{aligned} \tag{2.4}$$

Such processes were first studied by Onicescu and Mihoc (1935), Doeblin and Fortet (1937), and Harris (1955). This definition is somewhat different than that of Doeblin and Fortet (1937), but more suitable for our purposes. Observe that the conditional probabilities in (2.4) are defined for all sequences $\xi_n, \xi_n^*$ from $\mathscr{Y}$ by (2.2) and the stationarity of $\{Y_n\}_{n \in \mathbf{Z}}$. Notice also that $k$-step Markov dependence is equivalent to $\gamma_m = 0$ for all $m \ge k$. It was observed by Ledrappier (1976) (and is quite easy to prove) that the class of chains with complete connections for which $\gamma_m$ decays exponentially (i.e., $\liminf m^{-1} \log \gamma_m^{-1} > 0$) coincides with the class of stationary processes induced by the one-dimensional *Gibbs states* of Bowen (1975). (*Note*: Bowen's processes need not satisfy (2.1), but by changing the state space from $\mathscr{Y}$ to $\mathscr{Y}^k$ for some large $k$ and "blocking" the observations from Bowen's processes into blocks of size $k$ one may always obtain a Gibbs process for which (2.1) does hold. Thus there is no real loss of generality in assuming (2.1).)

The regenerative representation of chains with complete connections involves a special class of Markov chains which I will call *list processes*. A list process is a positive recurrent Markov chain $\{X_n\}$ with state space $\bigcup_{k\geq 1}\mathscr{Y}^k$ and stationary probability measure $\nu$ which satisfies the following transition rules:

$$P(X_{n+1} = (\xi_1, \xi_2, \dots, \xi_m) | X_n = (\zeta_1, \zeta_2, \dots, \zeta_k)) = 0 \tag{2.5}$$

unless either $m = 1$ or $m = k + 1$ and $\zeta_i = \xi_i$ for each $i = 1, 2, \dots, k$; and

$$P(X_{n+1} = (\xi_1) | X_{n+1} \in \mathscr{Y}^1; X_n) = \nu((\xi_1))/\nu(\mathscr{Y}^1) \tag{2.6}$$

for all $\xi_1 \in \mathscr{Y}$. Observe that (2.6) implies that the successive excursions between successive visits to $\mathscr{Y}^1$ are iid. Thus the list process evolves by concatenation of one symbol from $\mathscr{Y}$ at a time, except that at certain times the old list is destroyed and a new one begun from scratch, independently of the past.

Let $\pi$: $\bigcup_{k\geq 1}\mathscr{Y}^k \to \mathscr{Y}$ be the projection onto the *last* coordinate, i.e., $\pi(\xi_1, \dots, \xi_k) = \xi_k$. The main result is

THEOREM 1. *Suppose $\{Y_n\}_{n\in\mathbf{Z}}$ is a chain with complete connections for which the sequence $\{\gamma_m\}_{m\geq 1}$ decays exponentially. Then there is a stationary list process $\{X_n\}_{n\in\mathbf{Z}}$ such that*

$$\mathscr{D}(\{\pi(X_n)\}_{n\in\mathbf{Z}}) = \mathscr{D}(\{Y_n\}_{n\in\mathbf{Z}}). \tag{2.7}$$

*Moreover, the list process $\{X_n\}$ may be constructed so that for some integer $r \geq 1$ and some $\delta > 0$*

$$P(X_{n+1} \in \mathscr{Y}^1 | X_n = (\xi_1, \xi_2, \dots, \xi_{kr})) \equiv \delta \tag{2.8}$$

*for all $k = 1, 2, \dots$ and all $(\xi_1, \xi_2, \dots, \xi_{kr}) \in \mathscr{Y}^{kr}$.*

Notice that (2.8) implies that the recurrence times of $\{X_n\}$ have finite exponential moments; in fact, if $T = \inf\{n > 0\colon X_n \in \mathscr{Y}^1\}$ then for some $\beta > 1$, $E(\beta^T | X_0)$ is uniformly bounded.

In crude terms Theorem 1 states that a chain with complete connections for which $Y_m$ decays exponentially may always be realized by stringing together iid lists of symbols from $\mathscr{Y}$, with the list length random variable having finite exponential moments. It is apparent from Theorem 1 that many properties of recurrent Markov chains are inherited by chains with complete connections: for example, the classical limit theorems for additive functionals (the central limit theorem, law of the iterated logarithm, Berry–Esseen bounds, renewal theorem, etc.) follow trivially from the corresponding results for Markov chains.

It is natural to inquire whether the exponential decay of $\{\gamma_m\}$ is necessary for the existence of a representation like that provided by Theorem 1. The answer is no: one may have (2.7) and (2.8) hold for a chain with complete connections whose $\{\gamma_m\}$ sequence decays arbitrarily slowly. A simple example is the stationary renewal process for which the density $\{f_n\}_{n\geq 1}$ of the interoccurrence times is given by $f_n = \beta^{n-1}a_{n-1} - \beta^n a_n$, where $0 < \beta < 1$ and $\{a_n\}$ is a sequence of constants such that $a_n \to 1$, $a_n > \beta a_{n+1}$, and $a_0 = 1$. It is easily verified that for the stationary renewal process

$$\gamma_m \geq \sup_{n>m} \left| \frac{1 - \beta a_{n+1} a_n^{-1}}{1 - \beta a_{m+1} a_m^{-1}} - 1 \right|,$$

which may converge to zero quite slowly, depending on the rate at which $a_n \to 1$.

However, it seems that the rate of decay of the sequence $\{\gamma_m\}$ does have something to do with the rate of regeneration in a chain with complete connections. The following results indicate to some extent the nature of the connection.

THEOREM 2. *Suppose* $\{Y_n\}_{n\in\mathbf{Z}}$ *is a chain with complete connections for which* $\gamma_m = O(m^{-\beta})$. *Then for every* $\alpha < \beta$ *there exists a stationary list process* $\{X_n\}_{n\in\mathbf{Z}}$ *satisfying*

$$\mathscr{D}(\{\pi(X_n)\}_{n\in\mathbf{Z}}) = \mathscr{D}(\{Y_n\}_{n\in\mathbf{Z}}) \tag{2.9}$$

*and*

$$E(T^\alpha | X_0 = (\xi_1)) < \infty \tag{2.10}$$

*for each* $(\xi_1) \in \mathscr{Y}^1$, *where* $T = \inf\{n > 0\colon X_n \in \mathscr{Y}^1\}$.

THEOREM 3. *For each* $\beta > 1$ *there is a chain with complete connections* $\{Y_n\}_{n\in\mathbf{Z}}$ *for which* $\gamma_m = O(m^{-\beta+1})$ *and having the following property. If* $\{X_n\}_{n\in\mathbf{Z}}$ *is a stationary Markov chain on a countable state space* $\mathscr{X}$, *and if* $\pi\colon \mathscr{X} \to \mathscr{Y}$ *is a map for which*

$$\mathscr{D}(\{\pi(X_n)\}_{n\in\mathbf{Z}}) = \mathscr{D}(\{Y_n\}_{n\in\mathbf{Z}}), \tag{2.11}$$

*then for every* $x \in \mathscr{X}$

$$E(T_x^{\beta+1} | X_0 = x) = \infty, \tag{2.12}$$

*where* $T_x = \inf\{n > 0\colon X_n = x\}$.

Theorem 3 says, in essence, that Theorem 2 cannot be much improved: in particular, the minimal rate of decay of $\{\gamma_m\}$ sufficient to guarantee the existence of a Markov chain satisfying (2.9) with recurrence times $T$ having infinite $\beta$th moments is somewhere between $O(m^{-\beta+2})$ and $o(m^{-\beta})$. I have not succeeded in determining what the minimal rate of decay is.

The proofs of Theorems 2 and 3 will not be given in this paper, as they are considerably more complicated than that of the main result. They will (perhaps) appear in a subsequent publication. The class of counterexamples guaranteed by Theorem 3 is easy to describe: They are the stationary processes valued in $\mathscr{Y} = \{0, 1\}$ specified by the transition laws

$$P(Y_0 = 1 | Y_{-n} = \xi_n, n \geq 1) = \tfrac{1}{3} + ((1 - 2^{-\beta})/3) \sum_{k=0}^{\infty} \xi_{2^k} 2^{-k\beta}.$$

**3. Proof of Theorem 1.** Throughout this section it will be assumed that $\{Y_n\}_{n \in \mathbb{Z}}$ is a chain with complete connections defined on a probability space $(\Omega, \mathscr{F}, P)$ for which the sequence $\{Y_m\}_{m \geq 1}$ satisfies

$$\gamma_m \leq 2^{-4m}. \tag{3.1}$$

There is no real loss of generality in assuming (3.1), for if it is not satisfied the sequence $\{Y_n\}_{n \in \mathbb{Z}}$ may be replaced by the sequence $\{Z_n\}_{n \in \mathbb{Z}}$, where

$$Z_n = (Y_{nr+1}, Y_{nr+2}, \dots, Y_{nr+r})$$

for some large value of $r$. It is clear that $\{Z_n\}$ is a chain with complete connections, and that if $r$ is chosen sufficiently large (3.1) will hold.

The problem of constructing a list process for which (2.7) holds is essentially equivalent to the problem of constructing a single regeneration point for $\{Y_n\}$. This is the content of

Lemma 1. *To prove Theorem* 1 *it suffices to show that on some probability space a version of* $\{Y_n\}_{n \geq 0}$ *and a random variable* $T$ *valued in* $\mathbb{Z}^+$ *are defined such that for all* $k$, $m \in \mathbb{Z}^+$ *and* $\xi_j, \zeta_i \in \mathscr{Y}$

$$\begin{aligned} &P(Y_{m+n} = \xi_n \ \forall 0 \leq n \leq k \mid T = m;\ Y_n = \zeta_n \ \forall 0 \leq n < m) \\ &\quad = P(Y_n = \xi_n \ \forall 0 \leq n \leq k) \end{aligned} \tag{3.2}$$

*and*

$$P(T = m \mid T > m - 1;\ Y_n = \zeta_n \ \forall 0 \leq n < m) = \delta. \tag{3.3}$$

Proof. Define transition laws for a Markov chain $\{X_n\}$ valued in $\bigcup_{k \geq 1} \mathscr{Y}^k$ as follows:

$$\begin{aligned} &P(X_{n+1} = (\xi_1, \xi_2, \dots, \xi_{m+1}) \mid X_n = (\xi_1, \xi_2, \dots, \xi_m)) \\ &\quad = P(Y_m = \xi_{m+1}; T > m \mid T \geq m;\ Y_{j-1} = \xi_j \ \forall 0 \leq j < m) \\ &\quad = (1 - \delta) P(Y_m = \xi_{m+1} \mid T > m;\ Y_{j-1} = \xi_j \ \forall 0 \leq j < m) \end{aligned} \tag{3.4}$$

and

$$\begin{aligned} &P(X_{n+1} = (\zeta_1) \mid X_n = (\xi_1, \xi_2, \dots, \xi_m)) \\ &\quad = \delta P(Y_0 = \zeta_1). \end{aligned} \tag{3.5}$$

All other transitions must have probability zero, so (2.5) holds. Clearly (2.6) holds for $\nu((\zeta_1)) = \delta P(Y_0 = \zeta_1)$, for the ergodic theorem and (3.5) imply that the proportion of transitions to $(\zeta_1)$ is $\delta P(Y_0 = \zeta_1)$. Thus (3.4) and (3.5) are the transition laws for a list process. To prove Theorem 1 it must only be verified that (2.7) holds (since (2.8) holds trivially with $r = 1$, by (3.5)).

Assume that random variables $\{X_n\}_{n \geq 0}$ satisfying (3.4) and (3.5) and with initial distribution

$$P(X_0 = (\zeta_1)) = P(Y_0 = \zeta_1) \tag{3.6}$$

are defined on the same probability space as $\{Y_n\}_{n \geq 0}$ in such a way that the sequences $\{X_n\}_{n\geq 0}$ and $\{Y_n\}_{n\geq 0}$ are independent. I will argue that the sequences $\{\pi(X_n)\}_{n\geq 0}$ and $\{Y_n\}_{n\geq 0}$ have the same law.

Let $0 = T_0 < T_1 < T_2 < \cdots$ be the instants of successive visits to $\mathscr{Y}^1$ by $X_n$. It is apparent from (3.4)–(3.6) that the successive excursions $(X_0,\dots,X_{T_1-1})$, $(X_{T_1},\dots,X_{T_2-1}),\dots$ are iid. Now (3.2), (3.4), and (3.6) imply that the sequence $\pi(X_0),\pi(X_1),\dots,\pi(X_{T_1-1}),Y_0,Y_1,\dots$ has the same law as the sequence $Y_0,Y_1,\dots$. Consequently it follows by induction on $k$ and the iid property of the successive excursions of $\{X_n\}$ that the sequence $\pi(X_0),\pi(X_1),\dots,\pi(X_{T_k-1}),Y_0,Y_1,\dots$ has the same law as $Y_0,Y_1,\dots$ for any $k\geq 1$. Letting $k\to\infty$ one sees that $\pi(X_0),\pi(X_1),\dots$ must have the same law as $Y_0,Y_1,\dots$.

Now the process $\{Y_n\}_{n\geq 0}$ is stationary. Thus $Y_0,Y_1,\dots$ must have the same law as $\pi(X_m),\pi(X_{m+1}),\dots$ for any $m\geq 0$. It is clear from (3.5) that $\{X_n\}_{n\geq 0}$ is an aperiodic, irreducible, positive recurrent Markov chain, so for any initial distribution (including (3.6)) the distribution of $X_m$ for large $m$ approaches the unique stationary distribution. Therefore it follows that if $X_0$ had been given the stationary distribution instead of (3.6), it would still be the case that the sequences $\pi(X_0),\pi(X_1),\dots$ and $Y_0,Y_1,\dots$ have the same law. This proves (2.7). □

Notational conventions. For the remainder of this section I will use certain shorthand notations and conventions. The sequence $\{Y_n\}_{n\in\mathbf{Z}}$ will always consist of $\mathscr{Y}$-valued random variables; under the probability measure $P$ $\{Y_n\}$ will always be a chain with complete connections satisfying (3.1), but under a probability measure labelled $Q$ $\{Y_n\}$ may have another distribution. For any subset $A$ of the integers $\xi(A)$ will denote the event $\{Y_n=\xi_n\ \forall n\in A\}$; thus

$$Q(\xi(A)) = Q(Y_n=\xi_n\ \forall n\in A),$$

$$P(\xi(A)|\xi(B)) = P(Y_n=\xi_n\ \forall n\in A\,|\,Y_n=\xi_n\ \forall n\in B),$$

etc. Finally, the interval notations [ , ], ( , ], etc., will be used to denote intervals of *integers*, e.g., $[m,n]=\{m,m+1,\dots,n\}$ and $(m,n]=\{m+1,m+2,\dots,n\}$.

Lemma 2. *Suppose $\{Y_n\}_{n\in\mathbf{Z}}$ is a chain with complete connections defined on $(\Omega,\mathscr{F},P)$. Then there exists $\delta_0>0$ such that for every $A\subset(-\infty,-1]$ and finite $B\subset[0,\infty)$, and all values $\xi_n\in\mathscr{Y}$,*

$$P(\xi(B)|\xi(A))\geq\delta_0 P(\xi(B)). \tag{3.7}$$

Proof. It clearly suffices to consider only the cases $A=(-\infty,-1]$ and $B=[0,m]$. According to (2.3) there exist $\delta_1>0$ and an integer $1\leq r<\infty$ such that for all choices of $\xi_n\in\mathscr{Y}$ and all $m<\infty$

$$P(\xi[0,m]|\xi(-\infty,-1])\geq\delta_1 P(\xi[0,m]|\xi[-r,-1]), \tag{3.8}$$

and consequently, for all $k\geq r$,

$$P(\xi[0,m]|\xi[-k,-1])\geq\delta_1 P(\xi[0,m]|\xi[-r,-1]). \tag{3.9}$$

Now since there are only finitely many configurations (choices of $\xi_n$) on

$[-r, r-1]$ it follows from (2.1) that there exists $\delta_2 > 0$ such that for all $\xi_n \in \mathscr{Y}$

$$P(\xi[0, r-1] | \xi[-r, -1]) \geq \delta_2 P(\xi[0, r-1]). \tag{3.10}$$

Therefore for all $\xi_n \in \mathscr{Y}$ and $m > r-1$,

$$\begin{aligned} &P(\xi[0, m] | \xi[-r, -1]) \\ &\quad = P(\xi[r, m] | \xi[-r, r-1]) P(\xi[0, r-1] | \xi[-r, -1]) \\ &\quad \geq \delta_1 P(\xi[r, m] | \xi[0, r-1]) \delta_2 P(\xi[0, r-1]) \\ &\quad = \delta_1 \delta_2 P(\xi[0, m]). \end{aligned} \tag{3.11}$$

(Here we have used (3.9) and (3.10), together with the stationarity of $\{Y_n\}$, which guarantees that the conditional probabilities are translation invariant.) Combining (3.8) and (3.11), we conclude that for all $\xi_n \in \mathscr{Y}$ and all $m < \infty$

$$P(\xi[0, m] | \xi(-\infty, -1]) \geq \delta_1^2 \delta_2 P(\xi[0, m]). \qquad \square$$

Suppose $(\Omega, \mathscr{F}, P)$ is the probability space supporting the chain with complete connections $\{Y_n\}$. For each subset $A$ of the integers let $\mathscr{F}_A$ be the $\sigma$-algebra generated by $Y_n$, $n \in A$. Define a probability measure $Q_0$ on $(\Omega, \mathscr{F}_{[0,\infty)})$ by

$$Q_0(\xi[0, m]) = P(\xi[0, m]) \tag{3.12}$$

for all $m \geq 0$ and all $\xi_0, \xi_1, \dots, \xi_m \in \mathscr{Y}$. Thus $Q_0$ is just the restriction of $P$ to $\mathscr{F}_{[0,\infty)}$. Fix $\delta > 0$; for each $k \geq 0$ and each choice of $\xi_0, \xi_1, \dots, \xi_k \in \mathscr{Y}$ let $Q_{k+1}^{\xi[0,k]}$ be the probability measure on $(\Omega, \mathscr{F}_{[k+1,\infty)})$ specified by

$$\begin{aligned} &Q_{k+1}^{\xi[0,k]}(\xi[k+1, k+m]) \\ &\quad = (1-\delta)^{-1} \{ Q_k^{\xi[0,k-1]}(\xi[k+1, k+m] | \xi_k) \\ &\qquad - \delta P(\xi[k+1, k+m]) \}. \end{aligned} \tag{3.13}$$

(*Note*: When $k = 0$, $[0, k-1] = \varnothing$, so $Q_k^{\xi[0,k-1]}$ is just $Q_0$ in this case.)

LEMMA 3. *If $\delta < \delta_0/4$, where $\delta_0 > 0$ is the constant provided by Lemma* 2, *then* (3.13) *is a valid recursive definition*: *in particular*,

$$Q_k^{\xi[0,k-1]}(\cdot | \xi_k) \geq \delta P(\cdot) \tag{3.14}$$

*for all $k \geq 0$ and $\xi_0, \xi_1, \dots, \xi_k \in \mathscr{Y}$.*

PROOF. The argument is by induction on $k$. I will show (inductively) that for all $k \geq 0$, $m \geq 1$, and $r \geq 0$, and all choices of $\xi_n \in \mathscr{Y}$,

$$\begin{aligned} &\left| \frac{Q_k^{\xi[0,k-1]}(\xi[k+m, k+m+r] | \xi[k, k+m-1])}{P(\xi[k+m, k+m+r] | \xi[k, k+m-1])} - 1 \right| \\ &\quad \leq 4(16)^{-m}. \end{aligned} \tag{3.15}$$

Since $4(16)^{-m} \leq \frac{1}{2}$, it follows immediately from (3.7), (3.15), and the fact that $\delta < \delta_0/4 \leq \frac{1}{4}$ that

$$Q_k^{\xi[0,k-1]}(\xi[k+1, k+n] | \xi_k) \geq \tfrac{1}{2}\delta_0 P(\xi[k+1, k+n]) \tag{3.16}$$

for all $k \geq 0$, $n \geq 1$, and all choices of $\xi_j \in \mathscr{Y}$. This clearly implies (3.14).

Notice that (3.15), and (3.16), are trivial for $k = 0$ by (3.12). We now assume that (3.15) and (3.16) are true for some indeterminate value of $k$, and proceed to show that (3.15), and hence (3.16), must also hold for $k + 1$.

Write

$$Q_{k+1}^{\xi[0,k]}(\xi[k+1+m, k+1+m+r] \| \xi[k+1, k+m]) = \frac{Q_{k+1}^{\xi[0,k]}(\xi[k+1, k+1+m+r])}{Q_{k+1}^{\xi[0,k]}(\xi[k+1, k+m])}. \tag{3.17}$$

Now apply (3.13) to both numerator and denominator of the r.h.s. of (3.17) and then divide by $P(\xi[k+1+m, k+1+m+r] \| \xi[k+1, k+m])$ to obtain

$$\begin{aligned}
&\left| \frac{Q_{k+1}^{\xi[0,k]}(\xi[k+1+m, k+1+m+r] \| \xi[k+1, k+m])}{P(\xi[k+1+m, k+1+m+r] \| \xi[k+1, k+m])} - 1 \right| \\
&\quad = \left| \frac{Q_{k}^{\xi[0,k-1]}(\xi[k+1+m, k+1+m+r] \| \xi[k, k+m])}{P(\xi[k+1+m, k+1+m+r] \| \xi[k+1, k+m])} - 1 \right| \\
&\qquad \times \left| 1 - \delta \frac{P(\xi[k+1, k+m])}{Q_{k}^{\xi[0,k-1]}(\xi[k+1, k+m] \| \xi_k)} \right|^{-1} \\
&\quad \le \left(1 - 2\delta_0^{-1}\delta\right)^{-1} \\
&\qquad \times \left| \frac{Q_{k}^{\xi[0,k-1]}(\xi[k+1+m, k+1+m+r] \| \xi[k, k+m])}{P(\xi[k+1+m, k+1+m+r] \| \xi[k+1, k+m])} - 1 \right|.
\end{aligned} \tag{3.18}$$

The last inequality follows from (3.16), which holds by virtue of the induction hypothesis. Now since $\delta < \delta_0/4$,

$$\left(1 - 2\delta_0^{-1}\delta\right)^{-1} \le 2,$$

so to complete the proof it suffices to show that the $|\cdot|$ factor on the r.h.s. of (3.18) is no larger than $2^{-4m+1}$. But

$$\begin{aligned}
&\frac{Q_{k}^{\xi[0,k-1]}(\xi[k+1+m, k+1+m+r] \| \xi[k, k+m])}{P(\xi[k+1+m, k+1+m+r] \| \xi[k+1, k+m])} \\
&\quad = \left[ \frac{Q_{k}^{\xi[0,k-1]}(\xi[k+1+m, k+1+m+r] \| \xi[k, k+m])}{P(\xi[k+m, k+m+r] \| \xi[k, k+m])} \right] \\
&\qquad \times \left[ \frac{P(\xi[k+m, k+m+r] \| \xi[k, k+m])}{P(\xi[k+m, k+m+r] \| \xi[k+1, k+m])} \right] \\
&\quad = \left[1 \pm 4(16)^{-m-1}\right]\left[1 \pm \gamma_m\right]
\end{aligned} \tag{3.19}$$

by the induction hypothesis (3.15) and the definition (2.4) of $\gamma_m$. Consequently the $|\cdot|$ factor on the r.h.s. of (3.18) is no larger than

$$\left(\tfrac{1}{4}\right)(16)^{-m} + \gamma_m + \left(\tfrac{1}{4}\right)(16)^{-m}\gamma_m \le 2(16)^{-m},$$

by (3.1). □

PROOF OF THEOREM 1. By Lemma 1 it suffices to construct a version of $\{Y_n\}$ and a random variable $T$ on some probability space in such a way that (3.2) and (3.3) hold. Let $(\bar{\Omega}, \bar{\mathscr{F}}, Q)$ be a probability space on which are defined random variables $\{Y_n^A\}_{n\ge 0}$, $\{Y_n^B\}_{n\ge 0}$ (all valued in $\mathscr{Y}$), and $T$ (valued in $\{1,2,\dots\}$) such that

$$\begin{aligned}(3.20)\qquad & Q\left(Y_n^A = \xi_n, n \in \Lambda_1; Y_n^B = \zeta_n, n \in \Lambda_2; T = k\right)\\ &\quad = Q\left(Y_n^A = \xi_n, n \in \Lambda_1\right)Q\left(Y_n^B = \zeta_n, n \in \Lambda_2\right)Q(T = k)\end{aligned}$$

for all $\xi_n$, $\zeta_n \in \mathscr{Y}$, and $k \in \mathbb{Z}^+$, and all finite subsets $\Lambda_1$, $\Lambda_2 \subset \mathbb{Z}$; and

$$\begin{aligned} & Q(T = k) = \delta(1-\delta)^{k-1}, \qquad k = 1,2,\dots,\\ (3.21)\qquad & Q\left(Y_n^B = \xi_n, n \in \Lambda\right) = P(\xi(\Lambda)),\\ & Q\left(Y_n^A = \xi_n, 0 \le n \le k\right) = Q_0(\xi_0)\prod_{j=1}^{k} Q_j^{\xi[0,j-1]}(\xi_j)\end{aligned}$$

for all $\xi_n \in \mathscr{Y}$, $\Lambda \subset \mathbb{Z}$, and $k \ge 1$. Define new random variables $Y_n^*$, $n \ge 0$, on $(\bar{\Omega}, \bar{\mathscr{F}}, Q)$ by

$$\begin{aligned}(3.22)\qquad Y_n^* &= Y_n^A, \qquad n < T\\ &= Y_n^B, \qquad n \ge T.\end{aligned}$$

I will argue that $\{Y_n^*\}_{n\ge 0}$ has the same distribution as the original process $\{Y_n\}_{n\ge 0}$. It is clear from the construction that for all $\xi_j, \zeta_j \in \mathscr{Y}$, $k, m \in \mathbb{Z}^+$,

$$\begin{aligned} & Q\left(Y_{m+n}^* = \xi_n, 0 \le n \le k \mid T = m; Y_j = \zeta_j, 0 \le j \le m\right)\\ &\quad = Q\left(Y_{m+n}^B = \xi_n, 0 \le n \le k\right)\\ &\quad = P(\xi[0,k]),\end{aligned}$$

so by Lemma 1 showing $\{Y_n\}_{n\ge 0} =_{\mathscr{D}} \{Y_n^*\}_{n\ge 0}$ will suffice to complete the proof of Theorem 1.

To see that $\{Y_n\}_{n\ge 0} =_{\mathscr{D}} \{Y_n^*\}_{n\ge 0}$, use (3.20), (3.21), and (3.22) to write the finite-dimensional distributions of $\{Y_n^*\}_{n\ge 0}$ as

$$\begin{aligned} & Q\left(Y_n^* = \xi_n, 0 \le n \le k\right)\\ &\quad = \sum_{m=1}^{k} Q(T = m)Q\left(Y_n^A = \xi_n, 0 \le n < m\right)Q\left(Y_n^B = \xi_n, m \le n \le k\right)\\ (3.23)\qquad &\qquad + Q(T > k)Q\left(Y_n^A = \xi_n, 0 \le n \le k\right)\\ &\quad = \sum_{m=1}^{k} \delta(1-\delta)^{m-1}\left[\prod_{j=0}^{m-1} Q_j^{\xi[0,j-1]}(\xi_j)\right]P(\xi[m,k])\\ &\qquad + (1-\delta)^k \prod_{j=0}^{k} Q_j^{\xi[0,j-1]}(\xi_j).\end{aligned}$$

Now use the relation (3.13) successively for $Q_k^{\xi[0,k-1]}$, then $Q_{k-1}^{\xi[0,k-2]}$, etc., to get

$$\begin{aligned}
\text{r.h.s. } (3.23) &= \sum_{m=1}^{k-1} \delta(1-\delta)^{m-1}\left[\prod_{j=0}^{m-1} Q_j^{\xi[0,j-1]}(\xi_j)\right] P(\xi[m,k]) \\
&\quad + (1-\delta)^{k-1}\left[\prod_{j=0}^{k-1} Q_j^{\xi[0,j-1]}(\xi_j)\right] Q_{k-1}^{\xi[0,k-2]}(\xi_k|\xi_{k-1}) \\
&= \sum_{m=1}^{k-2} \delta(1-\delta)^{m-1}\left[\prod_{j=0}^{m-1} Q_j^{\xi[0,j-1]}(\xi_j)\right] P(\xi[m,k]) \\
&\quad + (1-\delta)^{k-2}\left[\prod_{j=0}^{k-2} Q_j^{\xi[0,j-1]}(\xi_j)\right] Q_{k-2}^{\xi[0,k-3]}(\xi[k-1,k]|\xi_{k-2}) \\
&= \cdots \\
&= Q_0(\xi_0)Q_0(\xi[1,k]|\xi_0) \\
&= Q_0(\xi[0,k]) \\
&= P(\xi[0,k]). \qquad \square
\end{aligned}$$

## REFERENCES


ATHREYA, K. and NEY, P. (1978). A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.* **245** 493–501.

BOWEN, R. (1975). *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms. Lecture Notes in Math.* **470**. Springer, Berlin.

DOBRUSHIN, R. (1968). The description of a random field by means of conditional probabilities and conditions of its regularity. *Theory Probab. Appl.* **13** 197–224.

DOEBLIN, W. and FORTET, R. (1937). Sur les chaînes à liaisons complètes. *Bull. Soc. Math. France* **65** 132–148.

HARRIS, T. E. (1955). On chains of infinite order. *Pacific J. Math.* **5** 707–724.

KRIEGER, W. (1983). On the finitary isomorphisms of Markov shifts that have finite expected coding time. *Z. Wahrsch. verw. Gebiete* **65** 323–328.

LANFORD, O. E. and RUELLE, D. (1969). Observables at infinity and states with short range correlations in statistical mechanics. *Comm. Math. Phys.* **13** 194–215.

LEDRAPPIER, F. (1976). Sur la condition de Bernoulli faible et ses applications. *Théorie Ergodique. Lecture Notes in Math.* **532** 152–159. Springer, Berlin.

NUMMELIN, E. (1978). A splitting technique for Harris recurrent Markov chains. *Z. Wahrsch. verw. Gebiete* **43** 309–318.

ONICESCU, O. and MIHOC, G. (1935). Sur les chaînes statistiques. *C. R. Acad. Sci. Paris* **200** 511–512.

PARRY, W. (1979). Finitary isomorphisms with finite expected code lengths. *Bull. Lond. Math. Soc.* **11** 170–176.

RUELLE, D. (1978). *Thermodynamic Formalism*. Addison-Wesley, Reading, Mass.



DEPARTMENT OF STATISTICS
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027