Probab. Th. Rel. Fields 76, 243–253 (1987)

Probability Theory and Related Fields



# Chains with Infinite Connections: Uniqueness and Markov Representation

Henry Berbee

Centre for Mathematics and Computer Science, P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

**Summary.** If for a process $(\xi_n)_{n=-\infty}^{\infty}$ the conditional distribution of $\xi_n$ given the past does not depend on $n$ for e.g. $n \geqq 0$, then the process may be called a chain with infinite connections. Under a well-known continuity condition on this conditional distribution the process is shown to be distributed as an instantaneous function of a countable state Markov chain. Also under a certain weaker continuity condition uniqueness of the distributions of the stationary chains is obtained.

## 1. Introduction and Results

Let $I$ be a finite or countable set. A non-negative function $g$ on $I \times \prod_{n \leqq -1} I$ is called a $g$-function if

$$\sum_{i_0 \in I} g(i_0 \mid i_-) = 1 \qquad \text{for } i_- \in I_- := \prod_{n \leqq -1} I.$$

With a $g$-function one associates $I$-valued processes as follows. Let us say that if for a process $(\xi_n)$

$$P(\xi_n = i_0 \mid (\xi_k)_{k<n}) = g(i_0 \mid (\xi_k)_{k<n}) \text{ a.s.} \tag{1.1}$$

for $n \geqq 0$ (or for all $n$) then it *develops* according to $g$ for these $n$. Suppose that the distribution of $(\xi_k)_{k<0}$ is known. Then by (1.1) for $n=0, 1, \ldots$ one determines successively the distribution of $(\xi_k)_{k \leqq n}$ and then by Kolmogorov's extension theorem one finds the unique distribution of the entire sequence $(\xi_k)$. These processes were introduced by Doeblin and Fortet [4] under the name chain with infinite connections. The distributions of the stationary processes of this form were called $g$-measures by Keane [9].

We measure the continuity of $g$ using $r_n$, $n \geqq 0$, defined by requiring

$$e^{-r_n} = \inf \frac{g(i_0 \mid i_-)}{g(j_0 \mid j_-)}$$

where the infimum is taken over all $i.$ and $j.$ such that $i_0=j_0, \ldots, i_{-n}=j_{-n}$. We will assume that g is bounded from below by a positive constant, so then $r_1 \geqq r_2 \geqq \ldots$ are finite.

The following result discusses uniqueness of $g$-measures.

**Theorem 1.1.** *If* $g$ *satisfies*

$$\sum_{n\geqq 1} \exp(-r_1 - \ldots - r_n) = \infty \tag{1.2}$$

*then there is a unique shift invariant g-measure* $\mu$. *Under this measure the shift is a Bernoulli shift and moreover*

$$P(\xi_n = i_0, \ldots, \xi_{n+k} = i_k \mid \xi_- = i_-) \to P(\bar{\xi}_0 = i_0, \ldots, \bar{\xi}_k = i_k)$$

*as* $n \to \infty$ *uniformly in* $i_-$. *Here* $\xi_n$ *is any process developing according to* $g$ *for* $n \geqq 0$ *and* $(\bar{\xi}_n)$ *has the aforementioned g-measure as distribution.*

This answers a question in Ledrappier [11] which considers the traditional condition

$$\sum_{n\geqq 1} r_n < \infty, \tag{1.3}$$

that was discussed already by Doeblin and Fortet [4]. Our proof uses also coupling. Let us mention that under (1.3) the coupling contact can be made "lasting" while this may perhaps not always be true under the weaker condition (1.2).

In a chain with infinite connections the distribution of $\xi_n$ given the past may depend on the entire past $(\xi_k)_{k<n}$, which is an infinite sequence. Below we succeed in "simplifying" the description of this process at the cost of a randomization. We construct a Markov representation, i.e. a Markov chain $(X_n)_{n\geqq 0}$ with a countable state space $S$ and a function $f: S \to I$ such that

$$(f(X_n))_{n\geqq 0} \stackrel{d}{=} (\xi_n)_{n\geqq 0}. \tag{1.4}$$

The next result applies also in the non-stationary case.

**Theorem 1.2.** *If* $g$ *satisfies* (1.3) *and* $(\xi_n)$ *is a chain with infinite connections such that* (1.1) *holds for* $n \geqq 0$, *then there exists a Markov representation* (1.4) *for* $(\xi_n)$.

Let us mention that the Markov chain in the representation happens to be quite simple: it has the form $X_n = (\xi_{n-j})_{0 \leqq j \leqq \tau_n}$ where $\tau_n$ is a.s. finite. It was already known that in the stationary case a chain with infinite connections has very strong mixing properties. However Berbee and Bradley [1] have shown by examples that existence of a Markov representation is only weakly related to mixing and so our result gives definite new information. Recently, Lalley [10] obtained by methods different from ours a similar result for the important subcase where $r_n \to 0$ at exponential rate. A remark at the end of Sect. 3 indicates a problem for this case that is still open. Much earlier Harris [6] obtained a uniqueness result using a condition related to (1.2), that is even weaker in case $|I|=2$. The corresponding limit result of this paper uses however a condition related to (1.3) instead of to (1.2). Kaijser [8] discusses the literature further.

Theorem 1.2 can be applied also to one-dimensional Ising systems where the continuity of the g-function can be investigated using Lemma 1 in Gallavotti [5]. We do not try here to get a generalization of Theorem 1.1 for Ising models. Let us just mention that the form of condition (1.2) seems to be pointing in the right direction by an example in Hofbauer [7].

## 2. A Setting for Markov Representation

Suppose $g$ is a given $g$-function. Let $(\xi_n)$ be stationary such that (1.1) holds for all $n$. Assume there is a Markov representation as follows: there is a stationary Markov chain with a transition probability $Q$ from a countable state $S$ to itself and there is a function $f: S \to I$ such that $(f(X_n))=(\xi_n)$. For the ease of the exposition we assume $\xi_n = f(X_n)$.

To the pair $f$, $Q$ describing the representation there is associated in a natural way an entrance law $\tilde{Q}$ from $I_-$ to $S$ calculated as

$$\tilde{Q}_{i_-,x} := \lim_{k\to\infty} P(X_n = x \mid \xi_{n-1} = i_{-1}, \ldots, \xi_{n-k} = i_{-k})\, P((\xi_k)_{k<n} \in di_-)-\text{a.s.} \tag{2.1}$$

The a.s.-existence of this limit is a consequence of the backward martingale theorem and by stationarity $\tilde{Q}$ does not depend on $n$. Because of (1.1) we have outside a null set

$$\sum_{x\in f^{-1}(i_0)} \tilde{Q}_{i_-,x} = g(i_0 \mid i_-). \tag{2.2}$$

By the Markov property

$$P(X_n = y \mid (\xi_k)_{k<n}) = E(P(X_n = y \mid X_{n-1}) \mid (\xi_k)_{k<n}). \tag{2.3}$$

From this we can calculate $P(\xi_{n-1} = i_0, X_n = y \mid (\xi_k)_{k<n-1} = i_-)$ in two ways and we get the equality

$$g(i_0 \mid i_-)\, \tilde{Q}_{i_-i_0,y} = \sum_{x\in f^{-1}(i_0)} \tilde{Q}_{i_-,x}\, Q_{xy} \tag{2.4}$$

valid outside a $P(\xi_- \in di_-)$-null set for a suitable version of $\tilde{Q}$.

The relation (2.4) is crucial for the Markov representation. It reflects that $\tilde{Q}$ in (2.1) does not depend on $n$ and that (2.2) holds. In Sect. 3 where we assume (1.3) we find a representation with $\tilde{Q}$ defined everywhere and also the relations above are valid everywhere. We mention that in case $g$ is discontinuous, which may occur for nice processes one cannot always find these relations to be valid everywhere.

The following converse is easily proved by an inductive calculation of (2.5) as described above.

**Lemma 2.1.** *Let $g$ be a $g$-function for $I$. Suppose $f: S \to I$ and assume $\tilde{Q}$ and $Q$ are transition probabilities from $I_-$ to $S$ and from $S$ to $S$ respectively such that (2.4) holds for all $i_- \in I_-$, $i_0 \in I$ and $x, y \in S$. Let*

$$\xi_-, X_0, X_1, \ldots$$

*describe a Markov chain with arbitrary initial distribution on $I_-$ and transition probabilities $\tilde{Q}, Q, Q, \ldots$. Then writing $\xi_n := f(X_n)$ for $n \geqq 0$ we have*

$$P(X_n = x \mid (\xi_k)_{k<n}) = \tilde{Q}_{(\xi_k)_{k<n}, x} \text{ a.s.} \tag{2.5}$$

*for $n = 0, 1, \ldots$ and because of (2.2) the process $\xi_n$ develops as $g$ for $n \geqq 0$.*

*Remark.* Suppose $Q$ has only one invariant distribution $\pi$. If in Lemma 2.1 the initial distribution is chosen such that the full sequence $(\xi_n)$ is stationary then $X_0$ has distribution $\pi$ and there is only one $g$-measure.

## 3. The Markov Representation

To prove Theorem 1.2 we construct an entrance law $\tilde{Q}$ from $I_-$ to the state space $S$ consisting of finite strings of $I$-elements. We construct $\tilde{Q}_{i_-,.}$ uniquely for all $i_- \in I_-$, using continuity of $g$. Basic in our use of continuity below and in Sect. 4 are the functions

$$g(i_0 \mid i_{-1} \ldots i_{-n}) := \inf_{(i_{-j})_{j>n}} g(i_0 \mid i_-). \tag{3.1}$$

Clearly $g(i_0 \mid i_{-1} \ldots i_{-n}) \uparrow g(i_0 \mid i_-)$ because $r_n \downarrow 0$ and we may decompose $g$ as

$$g(i_0 \mid i_-) = \sum_{n \geqq 0} \Delta g(i_0 \mid i_{-1} \ldots i_{-n}) \tag{3.2}$$

which is a sum of the non-negative terms

$$\begin{aligned} \Delta g(i_0 \mid i_{-1} \ldots i_{-n}) &:= g(i_0 \mid i_{-1} \ldots i_{-n}) - g(i_0 \mid i_{-1} \ldots i_{-n+1}) \quad &&\text{for } n \geqq 1, \\ &:= g(i_0) &&\text{for } n = 0. \end{aligned}$$

The split-up (3.2) suggests the construction of a Markov triple as follows. We construct a random vector $(\xi_0, \nu_0) \in I \times \{0, 1, 2, \ldots\}$ such that

$$P(\xi_0 = i_0, \nu_0 = n_0 \mid \xi_- = i_-) = \Delta g(i_0 \mid i_{-1} \ldots i_{-n_0}). \tag{3.3}$$

Note that $\xi_0$ will have by (3.2) the right marginal conditional distribution. Also introducing $\nu_0$ as above needs a randomization because $\nu_0$ is not given deterministically in terms of $\xi$-values.

Let us mention that as follows one can also construct $\nu_0$ in steps, conditionally given $\xi_- = i_-$. Generate the event $\{\nu_0 = 0\}$ such that $P(\xi_0 = i_0, \nu_0 = 0 \mid \xi_- = i_-) = g(i_0)$. Subsequently for $n = 1, 2, \ldots$ generate $\{\nu_0 = n\} = \{\nu_0 \leqq n\} \setminus \{\nu_0 \leqq n - 1\}$ such that

$$P(\xi_0 = i_0, \nu_0 \leqq n \mid \xi_- = i) = g(i_0 \mid i_{-1} \ldots i_{-n}). \tag{3.4}$$

This can be done consistently because the right hand side increases in $n$. Also the right hand side does not depend on $i_j$ for $j<-n$ and one checks that

$$\xi_-,\ \xi_{-\nu_0}^{-1},\ \xi_0$$

forms a Markov triple. Here one writes $\xi_m^n:=(\xi_m, \xi_{m+1}, \dots, \xi_n)$, $m\leqq n$ and one says that $X, Y, Z$ forms a *Markov triple* if $X$ and $Z$ are independent given $Y$.

Loosely speaking we may say that the random truncation $\xi_{-\nu_0}^{-1}$ of $\xi_-$ contains the relevant information from the past $\xi_-$ to "generate" $\xi_0$. However it may be untrue that we can truncate subsequently $(\xi_{-\nu_0}^{-1}, \xi_0)$ so that this random vector contains the relevant information to generate $\xi_1$. Related to this difficulty is that to the entrance law

$$\tilde{Q}'_{i_-,x}:=P(X'_0=x\,|\,\xi_-=i_-) \qquad \text{where } X'_0:=\xi_{-\nu_0}^0$$

one may not be able to find $Q'$ linked with $\tilde{Q}'$ such that (2.4) holds. Note however that

$$\tilde{Q}'_{i_-,x}=\Delta g(i_0\,|\,i_{-1}, \dots, i_{-n_0})$$

for $x=(i_{-n_0}, \dots, i_0)$ has the property that it does not depend on $i_j, j<-n$. This nice property also holds for the "right" entrance law $\tilde{Q}$ that we define below and makes it quite simple to come to our Markov representation.

We constructed $(\xi_0, \nu_0)$ given $\xi_-$. Construct also $(\xi_1, \nu_1), (\xi_2, \nu_2), \dots$ successively in the same way such that $\nu_k$ is independent of all other variables, given $(\xi_j)_{j\leqq k}$. To this end one requires

$$\begin{aligned} P(\xi_0=i_0, \nu_0=n_0, \dots, \xi_N=i_N, \nu_N=n_N\,|\,\xi_-=i_-)\\ =\Delta g(i_0\,|\,i_{-1}\dots i_{-n_0})\dots\Delta g(i_N\,|\,i_{N-1}, \dots, i_{N-n_N}). \end{aligned} \tag{3.5}$$

Above we noted that $(\xi_-, \xi_{-\nu_0}^{-1}, \xi_0)$ is a Markov triple and similarly that $(\xi_{-\infty}^{k-1}, \xi_{k-\nu_k}^{k-1}, \xi_k)$ is a Markov triple. We now want to form a Markov triple of the form $(\xi_{-\infty}^{-1}, \xi_{-\tau_0}^{-1}, (\xi_0, \xi_1, \dots))$. Let $F_k$ be the random set $F_k:=\{k-\nu_k, \dots, k-1\}$, $k=0, 1, 2, \dots$ and take $\tau_0$ such that

$$\{-\tau_0, \dots, -1\}=[F_0\cup F_1\cup\dots]\cap\{\dots, -2, -1\}.$$

Lemma 3.3 will imply that $\tau_0$ is finite a.s. Define similarly $\tau_k:=\sup_{j\geqq 0}\{\nu_k, \dots, \nu_{k+j}-j, \dots\}$. Then as we will see below

$$\xi_-,\ X_0:=\xi_{-\tau_0}^0, \dots, X_k:=\xi_{k-\tau_k}^k, \dots \tag{3.6}$$

describes the Markov chain for our Markov representation. From (3.5) one notes that the entrance law

$$\tilde{Q}_{i_-,x}=P(X_0=x\,|\,\xi_-=i_-)$$

is defined uniquely for all $i_-\in I_-$ as in lemma 2.1 and one takes $f(x)$ as the last element in the string $x$ of $I$-elements.

**Lemma 3.1.** $\xi_-, X_0, ((\xi_1, \nu_1), (\xi_2, \nu_2), \ldots)$ *is a Markov triple.*

We prove this lemma below. Using the following trivial technical lemma

**Lemma 3.2.** *If* $X, Y, Z$ *is a Markov triple and* $\tilde{X}$ *is* $(X, Y)$*-measurable and* $\tilde{Z}$ *is* $(Y, Z)$*-measurable then* $\tilde{X}, Y, \tilde{Z}$ *is a Markov triple,*
it follows from Lemma 3.1 that $\xi_-, X_0, X_1$ is also a Markov triple and one easily verifies (2.4). Thus Lemma 2.1 gives us the Markov representation. Below we prove moreover that (3.6) gives us a representation, which needs some more arguments.

To this end we use Lemma 3.1 again to show that (3.6) is Markov chain. Fix $k \geqq 0$ and write $\bar{\xi}_j = \xi_{k+j}, \bar{X}_j = X_{k+j}$, etc. Note that the distribution of $(\xi_j, \nu_j)_{j \geqq 0}$ given $\xi_- = i_-$ is the same as the distribution of $(\bar{\xi}_j, \bar{\nu}_j)_{j \geqq 0}$ given $\bar{\xi}_- = i_-$. Thus by Lemma 3.1 $\bar{\xi}_-, \bar{X}_0, (\bar{\xi}_j, \bar{\nu}_j)_{j \geqq 1}$ is a Markov triple. Because of the conditional independence of $(\nu_j)_{0 \leqq j < k}$ and the other variables of that triple, given $\bar{\xi}_- = i_-$ it follows that

$$(\xi_-, (\xi_0, \nu_0), \ldots, (\xi_{k-1}, \nu_{k-1})), X_k, ((\xi_{k+1}, \nu_{k+1}), \ldots)$$

is a Markov triple. By Lemma 3.2 it follows also that $(\xi_-, X_0, \ldots, X_{k-1}), X_k, X_{k-1}$ is a Markov triple for any $k$ and thus (3.6) describes a Markov chain. Because the conditional distribution given $\bar{\xi}_- = i_-$ of $(\bar{\xi}_j, \bar{\nu}_j)_{j \geqq 0}$ and so of $X_k, X_{k+1}, \ldots$ does not depend on $k$ the Markov chain (3.6) has stationary transition probabilities.

*Proof of Lemma 3.1.* Let $X_0^{(N)} := \xi^0_{(-\nu_0) \wedge \ldots \wedge (N - \nu_N)}$. We will show for any $N$ that

$$\xi_-, X_0^{(N)}, ((\xi_0, \nu_0), \ldots, (\xi_N, \nu_N)) \tag{3.7}$$

is a Markov triple. To get the assertion observe that then for each $n \leqq N$ also $\xi_-, X_0^{(N)}, (\xi_j, \nu_j)_{j=0}^n$ forms a Markov triple. For fixed $n$ let $N \to \infty$. Then $X_0^{(N)}$ is a vector increasing in length to $X_0$ a.s., which has finite length by Lemma 3.3. The assertion follows.

To prove (3.7) we have to investigate the ratio of

$$P(X_0^{(N)} = x, \xi_0 = i_0, \nu_0 = n_0 \ldots \xi_N = i_N, \nu_N = n_N \mid \xi_- = i_-) \tag{3.8'}$$

and

$$P(X_0^{(N)} = x \mid \xi_- = i_-), \tag{3.8''}$$

and we want to show that this ratio depends only on $x = (i_0, \ldots, i_{-t_0})$ and $(i_0, n_0, \ldots, i_N, n_N)$. Rewrite (3.8') as the product

$$= \Delta g(i_0 \mid i_{-1} \ldots i_{-n_0}) \ldots \Delta g(i_N \mid i_{N-1}, \ldots, i_{N-n_N}) \tag{3.9'}$$

in case $t_0 = n_0 \vee (n_1 - 1) \vee \ldots \vee (n_N - N)$ and as

$$= 0 \quad \text{otherwise}. \tag{3.9'}$$

Note that $x$ determines $t_0$. To find a similar expression for (3.8'') we have to sum (3.9') over all $(i_0, n_0, \ldots, i_N, n_N)$ for which

$$x = (i_{(-n_0) \wedge (1-n_1) \wedge \ldots \wedge (N-n_N)}, \ldots, i_{-1}, i_0).$$

Now note that in these expressions no $i_{-j}$ occurs with $j > t_0 = n_0 \vee \ldots \vee (n_N - N)$. This proves the assertion about the ratio and thus proves the lemma.

**Lemma 3.3.** *If $\Sigma r_n < \infty$ then $\tau_0 = \sup_{j \geq 0} (v_j - j)$ is finite a.s.*

*Proof.* Write

$$P(\tau_0 \leqq n \mid \xi_- = i_-) = \lim_{N \to \infty} P(v_0 \leqq n, \ldots, v_N \leqq n + N \mid \xi_- = i_-).$$

By (3.4) and (3.5) the expression in the limit can be written as

$$\sum_{i_0 \ldots i_N} g(i_0 \mid i_{-1} \ldots i_{-n}) \, g(i_1 \mid i_0 \ldots i_{-n}) \ldots g(i_N \mid i_{N-1} \ldots i_{-n}). \tag{3.10}$$

Note that as $N$ increases this descends, say to $q_n(i_-)$ for $N \to \infty$. Clearly $P(\tau_0 < \infty \mid \xi_- = i_-) = 1$ if and only if

$$q_n(i_-) \uparrow 1 \quad \text{as} \quad n \to \infty. \tag{3.11}$$

We prove this using the condition $\Sigma r_n < \infty$. We bound $q_n(i_-)$ from below. By the definition of $r_n$ we have

$$g(i_0 \mid i_{-1} \ldots i_{-n}) \geqq e^{-r_n} g(i_0 \mid i_-).$$

Hence (3.10) is bounded from below by

$$e^{-r_n - \cdots - r_{n+N}} \sum_{i_0 \ldots i_N} g(i_0 \mid i_-) \, g(i_1 \mid (i_0 \, i_-)) \ldots g(i_N \mid \ldots).$$

Because $g$ is a g-function the sum above equals 1. Hence

$$q_n(i_-) \geqq \lim_{N \to \infty} e^{-r_n - \cdots - r_{n+N}} = e^{-r_n - r_{n+1} - \cdots}.$$

So if $\Sigma r_n < \infty$ we indeed have (3.11). □

*Note.* The condition $\Sigma r_n < \infty$ is a smooth uniform continuity condition on $g$. By the proof above it can be related to (3.11), retaining a.s. finiteness of $\tau_0$.

*Remark.* We mention an open problem. Let $(\xi_n)$ be a chain with infinite connections for which $r_n \to 0$ exponentially. For this case Bowen [3] in the proof of 1.25 verifies the $\psi$-mixing (or *-mixing) rate to be exponential. Blum et al. [2] introduce this mixing condition and prove that a $\psi$-mixing Markov chain is exponentially $\psi$-mixing. The question arises whether if $r_n \to 0$ exponentially there is a Markov representation based on a $\psi$-mixing Markov chain. The Markov chain constructed above is obtained as a truncation of the past and does not help to answer this, loosely speaking because too much detailed information of the past may be preserved.

## 4. Uniqueness of g-Measures

We prove Theorem 1.1. We study measures $\mu$ on $I^{\mathbb{Z}}$ (provided with the product $\sigma$-field) and we do not yet assume shift invariance of $\mu$. Let $\xi_n$ be the projection on the $n$th coordinate. Assume for $n=0, 1, \ldots$ that $\xi_n$ develops as $g$, i.e. (1.1) holds. Then as we noted in the introduction $\mu$ is determined uniquely by the distribution $\mu_-$ of $\xi_- := (\xi_k)_{k<0}$ under $\mu$. In case $\xi_- = i_-$ $\mu$-a.s. where $i_- \in I_-$ we write $\mu \equiv \mu_{i_-}$.

We investigate the dependence of $\mu_{i_-}$ on $i_-$ and do this by formalizing the classical coupling of Doeblin and Fortet [4]. Suppose $\xi_{-1} = i_{-1}, \ldots, \xi_{-n} = i_{-n}$ is given but $\xi_{-n-1}$ is "unknown under $\mu$". Then $\{\xi_0 = i_0\}$ under $\mu$ has at least mass (3.1). So despite our lack of knowledge concerning $\xi_{-n-1}$ this gives some information concerning the distribution of $\xi_0$. We will try to make similar statements below (e.g. (4.2)). We will express absence of knowledge concerning $\xi_{-n-1}$ by $\xi_{-n-1} = \partial$ where $\partial$ is some point outside $I$. Write $I_\partial := I \cup \{\partial\}$. We extend the $g$-function $g$ defined with respect to $I$ to a $g$-function $g_\partial$ with respect to $I_\partial$ such that

$$g_\partial(j_0 \mid j_{-1}, j_{-2}, \ldots) := g(i_0 \mid i_{-1} \ldots i_{-n}) \tag{4.1}$$

in case $j_{-n-1} = \partial$, and $i_0 = j_0, \ldots, i_{-n} = j_{-n}$ are in $I$. The g-function normalization determines $g_\partial(\partial \mid .)$. In this definition $n = \infty$ is allowed and we may consider $g_\partial$ as a continuous extension of $g$. Let $\mu_\partial$ on $I_\partial^{\mathbb{Z}}$ be such that $\xi_n = \partial$ for $n<0$ while $\xi_n$ develops according to $g_\partial$ for $n \geqq 0$. We prove uniformly for the measures $\mu$ in the first paragraph that

$$\mu_\partial(\xi_{n_0} = i_0, \ldots, \xi_{n_k} = i_k) \leqq \mu(\xi_{n_0} = i_0, \ldots, \xi_{n_k} = i_k) \tag{4.2}$$

for $0 \leqq n_0 < n_1 < \ldots$ and all $i_j \in I$. We also show that (1.2) implies the important property $\mu_\partial(\xi_n = \partial) \to 0$ as $n \to \infty$, so under $\mu_\partial$ we have for $n<0$ that $\xi_n$ equals $\partial$, so is "unknown" while it becomes "known" for large $n$. This will imply our results.

To get (4.2) we construct step by step a coupling. Order $I_\partial$ partially by letting $i \leqq j$ if $j = \partial$ or $i = j$. We construct a probability $P$ on $(I_\partial \times I_\partial)^{\mathbb{Z}}$. Let $(\xi_n, \xi_n^\partial)$ on this product space be the projection on the $n$th coordinate. We assume that $(\xi_n, \xi_n^\partial)$, $n<0$, under $P$ has an arbitrary distribution subject to the condition that $\xi_n \leqq \xi_n^\partial$, $n<0$, a.s. We want to construct $P$ such that this inequality holds for all $n$ and such that marginally both $\xi_n$ and $\xi_n^\partial$ develop according to $g_\partial$, for $n \geqq 0$. Let us specify for $n \geqq 0$

$$P(\xi_n = i_0, \xi_n^\partial = j_0 \mid \xi_{n-1} = i_{-1}, \xi_{n-1}^\partial = j_{-1}, \xi_{n-2} = i_{-2}, \ldots) \tag{4.3}$$

where $i_{-1} \leqq j_{-1}, i_{-2} \leqq j_{-2}, \ldots$. In case $i_0 \in I$ and $j_0 = i_0$ define (4.3) as

$$g_\partial(i_0 \mid j_{-1}, j_{-2}, \ldots)$$

and if $i_0 \in I$ and $j_0 = \partial$ as

$$g_\partial(i_0 \mid i_{-1}, i_{-2}, \ldots) - g_\partial(i_0 \mid j_{-1}, j_{-2}, \ldots)$$

which is easily verified to be non-negative because $i_- \leqq j_-$. Make (4.3) a $g$-function for $I_\partial \times I_\partial$, so a conditional probability, by assigning the remaining mass $g_\partial(\partial \mid i-_1, i_{-2}, \ldots)$ to (4.3) for $i_0 = j_0 = \partial$. Now let $(\xi_n, \xi_n^\partial)$ develop according to (4.3) for $n \geqq 0$. This determines $P$ and describes a "coupling".

We get (4.2) if we specify $\xi_n^\partial = \partial$, $n<0$, and let $(\xi_n)_{n<0}$ be distributed under $P$ as under $\mu$. Then marginally $\xi_.$ and $\xi_.^\partial$ have distribution $\mu$ and $\mu_\partial$ and using that a.s. $\xi_n \leqq \xi_n^\partial$ (so $\xi_n = \xi_n^\partial$ as soon as $\xi_n^\partial \in I$) we easily obtain (4.2).

We can also get more information about $\mu_\partial$ in this way. Replace $\mu$ in the argument in the last paragraph by $T^{-1}\mu_\partial$, where $T$ is the shift on sequence space. Because now also $\xi_n \leqq \xi_n^\partial$ a.s. for $n<0$ and consequently for all $n$, we have

$$\mu_\partial(\xi_{n_0}=i_0, \ldots, \xi_{n_k}=i_k) \leqq T^{-1}\mu_\partial(\xi_{n_0}=i_0, \ldots, \xi_{n_k}=i_k)$$

where all $i_j \in I$. Hence $\mu_\partial(\xi_{n_0+m}=i_0, \ldots, \xi_{n_k+m}=i_k)$ is increasing in $m$ and then $T^{-m}\mu_\partial$ converges weakly to a shift invariant measure $\mu_{\partial*}$ on $(I_\partial)^{\mathbb{Z}}$. If now $\mu$ on $I^{\mathbb{Z}}$ in (4.2) is required to be translation invariant then we can improve (4.2) to

$$\mu_{\partial*}(\xi_{n_0}=i_0, \ldots, \xi_{n_k}=i_k) \leqq \mu(\xi_{n_0}=i_0, \ldots, \xi_{n_k}=i_k). \tag{4.4}$$

The follows because $T^{-m}\mu=\mu$ and by replacing $n_j$ by $n_j+m$, $m\to\infty$ in (4.2).

By Lemma 4.1 below (1.2) implies that $\mu_\partial(\xi_n=\partial)$ descends to 0. Then $\mu_{\partial*}(\xi_0=\partial)=0$ and $\mu_{\partial*}$ is concentrated on $I^{\mathbb{Z}}$. Then we should have equality in (4.4) and so there is only one translation invariant measure $\mu \equiv \mu_{\partial*}$ on $I^{\mathbb{Z}}$ for which (1.1) holds for $n \geqq 0$.

The limit assertion of Theorem 1.1 is now easily seen. Using (4.2)

$$\mu_{i-}(\xi_m=i_0, \ldots, \xi_{m+k}=i_k) \geqq T^{-m}\mu_\partial(\xi_0=i_0, \ldots, \xi_k=i_k)$$
$$\uparrow \mu_{\partial*}(\xi_0=i_0, \ldots, \xi_k=i_k) \quad \text{as} \quad m\to\infty.$$

Because $\mu_{\partial*}$ is concentrated on $I^{\mathbb{Z}}$ it is easy to see from this that we have the asserted convergence, uniformly in $i_-$.

We claim also that $\xi_n$ develops according to $g$ under $\mu_{\partial*}$. To this end note that $\xi_n$ develops according to $g_\partial$ under $\mu_\partial$ for $n \geqq 0$, and by Lemma 4.1 takes the value $\partial$ increasingly less often. Moreover $g_\partial$ is a (continuous) extension of $g$. The claim follows now easily by evaluating and estimating $\mu_{\partial*}(\xi_0=i_0 \mid \xi_{-1}=i_{-1} \ldots \xi_{-n}=i_{-n})$ using that $\mu_{\partial*}$ is a weak limit of $T^{-m}\mu_\partial$.

Let us now proceed to show that the shift $T$ under this unique $\mu_{\partial*}$ is a Bernoulli shift. Let us verify the very weak Bernoulli condition (see Shields [13] and Schwarz [12]).

Consider now $P \equiv P_*$ as above such that $\xi_.$ and $\xi_.^\partial$ are distributed as $\mu_{\partial*}$ and $\mu_\partial$ respectively. Similarly we can define $P=P_-$ such that these marginal distributions are $\mu_{i-}$ and $\mu_\partial$ respectively. Now we construct a new probability space with processes $\xi_.^*$, $\xi_.^-$ and $\xi_.^\partial$ such that

(i) $(\xi_.^*, \xi_.^\partial)$ has distribution $P_*$

(ii) $(\xi_.^-, \xi_.^\partial)$ has distribution $P_-$.

This could be done e.g. by letting $\xi_.^*$ and $\xi_.^-$ be independent given $\xi_.^\partial$ such that (i) and (ii) hold. On this new probability space we have clearly that as soon as $\xi_n^\partial \neq \partial$ then $\xi_n^* = \xi_n^- = \xi_n^\partial$ a.s. and so

$$\frac{1}{n+1}\sum_{k=0}^{n} 1_{\xi_k^* \neq \xi_k^-} \leqq \frac{1}{n+1}\sum_{k=0}^{n} 1_{\xi_k^\partial=\partial}.$$

So the $\bar{d}$-distance of $(\xi_0, \ldots, \xi_n)$ and $(\xi_0, \ldots, \xi_n)|(\xi_j = i_j)_{j<0}$ (see Shields [13]) is at most

$$\frac{1}{n+1} \sum_{k=0}^{n} P(\xi_k^\partial = \partial)$$

and tends to 0 for $n \to \infty$. Note that this holds even uniformly in $i_- = (i_j)_{j<0}$. Thus the coupling has led us very easily to the verification of the very weak Bernoulli condition and so the shift is a Bernoulli shift under $\mu = \mu_{\partial *}$.

Using the notation of the proof above we have the following comparison lemma.

**Lemma 4.1.** *Suppose $\xi_n = \partial$, $n < 0$, and let $\xi_n$ for $n \geqq 0$ develop according to $g_\partial$, the extension of g determined by* (4.1). *If* (1.2) *holds for g then*

$$\lim_{n \to \infty} P(\xi_n = \partial) = 0.$$

*Proof.* The process $(\xi_n)$ has distribution $\mu_\partial$. We want to compare $\varepsilon_n := 1_{\{\xi_n = \partial\}}$, $n \in \mathbb{Z}$, with a simpler process. By (4.1) for $n \geqq 0$

$$P(\varepsilon_n = 0 \mid \xi_{n-1}, \xi_{n-2}, \ldots) \geqq p_m := \inf_{i_{-1} \ldots i_{-m}} \sum_{i_0} g(i_0 \mid i_{-1} \ldots i_{-m})$$

on the set $\{\lambda_n = m\}$ where $\lambda_n$ is the smallest $m \geqq 0$ for which $\xi_{n-m-1} = \partial$. Hence taking conditional expectations we have on $\{\lambda_n = m\} = \{\varepsilon_{n-1} = \ldots = \varepsilon_{n-m} = 0, \varepsilon_{n-m-1} = 1\}$

$$P(\varepsilon_n = 0 \mid \varepsilon_{n-1}, \varepsilon_{n-2}, \ldots) \geqq p_m, \qquad m, n \geqq 0. \tag{4.5}$$

If equality would hold above then $\varepsilon$ would be a renewal process. We construct $\tilde{\varepsilon}_n \geqq \varepsilon_n$ satisfying this property. Let $\tilde{\varepsilon}_n := 1$, $n < 0$. We prescribe for $n \geqq 0$ a $g$-function

$$P(\varepsilon_n = i_0, \tilde{\varepsilon}_n = j_0 \mid (\varepsilon_{n-1}, \tilde{\varepsilon}_{n-1}) = (i_{-1}, j_{-1}), (\varepsilon_{n-2}, \tilde{\varepsilon}_{n-2}) = (i_{-2}, j_{-2}), \ldots) \tag{4.6}$$

where $i_{-k} \leqq j_{-k}$ for all $k$. Let $m$ and $\tilde{m} \geqq 0$ be the smallest integers for which $i_{-m-1} = 1$ and $j_{-\tilde{m}-1} = 1$. Clearly $\tilde{m} \leqq m$. For $i_0 = j_0 = 0$ let (4.6) be $p_{\tilde{m}}$ and for $i_0 = 0, j_0 = 1$

$$P(\varepsilon_n = 0 \mid \varepsilon_{n-1} = i_{-1}, \varepsilon_{n-2} = i_{-2}, \ldots) - p_{\tilde{m}}.$$

This is nonnegative because of (4.5) and $p_m \geqq p_{\tilde{m}}$. Thus $\varepsilon_n$ has the right conditional marginal distribution. Because we want $\varepsilon_n \leqq \tilde{\varepsilon}_n$ the remaining mass has to be assigned to $\{i_0 = j_0 = 1\}$ to make (4.6) a probability. So $\tilde{\varepsilon}_n$ is a renewal process, satisfying for $n \geqq 0$

$$P(\tilde{\varepsilon}_n = 0 \mid \tilde{\varepsilon}_{n-1}, \tilde{\varepsilon}_{n-2}, \ldots) = p_{\tilde{m}} \quad \text{on} \quad \{\tilde{\varepsilon}_{n-1} = \ldots = \tilde{\varepsilon}_{n-\tilde{m}} = 0, \tilde{\varepsilon}_{n-\tilde{m}-1} = 1\}$$

if we let $(\varepsilon_n, \tilde{\varepsilon}_n)$ develop according to the g-function (4.6). Because $\{\varepsilon_n = 1\} \subset \{\tilde{\varepsilon}_n = 1\}$ it is sufficient to prove

$$P(\xi_n = \partial) \leqq P(\tilde{\varepsilon}_n = 1) \to 0 \quad \text{as} \quad n \to \infty. \tag{4.7}$$

The renewal process $\tilde{\varepsilon}$ has the following property. Observe that for any $n \geqq -1$ on the set $\{\tilde{\varepsilon}_n = 1\}$ the distance $\eta := \inf\{k \geqq 0: \tilde{\varepsilon}_{n+k} = 1\}$ to the next renewal has a conditional distribution $F$, given by

$$P(\eta > m \mid \tilde{\varepsilon}_n, \tilde{\varepsilon}_{n-1}, \ldots) = p_0 \, p_1 \ldots p_m$$

$F$ determines the distribution of $\tilde{\varepsilon}$. Its mean is $\mu := \sum_{m \geqq 0} p_0 \ldots p_m$. $F$ may be defective so $\lim_{m \to \infty} p_0 \ldots p_m > 0$. Then $\mu = \infty$ and $\#\{n \geqq 0: \varepsilon_n = 1\}$ is finite a.s. implying (4.7). Otherwise by the renewal theorem

$$\lim_{n \to \infty} P(\tilde{\varepsilon}_n = 1) = \frac{1}{\mu}.$$

So in either case $\mu = \infty$ implies (4.7). Because (1.2) implies $\mu = \infty$ by the definition of $p_m$ this completes the proof. □

## References


1. Berbee, H.C.P., Bradley, R.C.: A limitation of Markov representation for stationary processes. Stochastic Processes Appl. **18**, 33–45 (1984)
2. Blum, J.R., Hanson, D.L., Koopmans, L.H.: On the strong law of large numbers for a class of stochastic processes, Z. Wahrscheinlichkeitstheor. Verw. Gebiete **2**, 1–11 (1963)
3. Bowen, R.: Equilibrium states and the ergodic theory of Anosov diffeomorphisms, Lect. Notes Math. **470**. Berlin Heidelberg New York: Springer (1975)
4. Doeblin, W., Fortet, R.: Sur les chaînes à liaisons complètes. Bull. Soc. Math. France **65**, 132–148 (1937)
5. Gallavotti, G.: Ising model and Bernoulli schemes in one dimension. Comm. Math. Phys. **32**, 183–190 (1973)
6. Harris, T.E.: On chains of infinite order. Pacific J. Math. **5**, 707–724 (1955)
7. Hofbauer, F.: Examples for the uniqueness of the equilibrium state. Trans. Am. Math. Soc. **228**, 223–241 (1971)
8. Kaijser, T.: On a new contraction condition for random systems with complete connections. Rev. Roumaine Mat. Pures Appl. **26**, 1075–1117 (1981)
9. Keane, M.: Strongly mixing $g$-measures. Invent. Math. **16**, 309–324 (1972)
10. Lalley, S.P.: Regenerative representation for one-dimensional Gibbs states. Ann. Probab. **14**, 1262–1271 (1986)
11. Ledrappier, F.: Principe variationnel et systèmes dynamiques symboliques. Z. Wahrscheinlichkeitstheor. Verw. Gebiete **30**, 185–202 (1974)
12. Schwarz, G.: Finitely determined processes – an indiscrete approach. J. Math. Anal. Appl. **76**, 146–158 (1980)
13. Shields, P.: The theory of Bernoulli shifts. Chicago London: University of Chicago Press (1973)


Received December 17, 1985