ISRAEL JOURNAL OF MATHEMATICS, Vol. 71, No. 1, 1990

RANDOM MARKOV PROCESSES AND UNIFORM MARTINGALES

BY

STEVE KALIKOW†

Department of Mathematics, University of Southern California, Los Angeles, CA 90089, USA

ABSTRACT

It is shown here that a certain generalization of an n-step Markov chain is equivalent to the uniform convergence of the martingale $\{P(X_0 | X_{-1} X_{-2} \cdots X_{-n})\}_{n=1}^{\infty}$. Ergodic and probabilistic properties of this process are explored.

1. Statement of results with discussion (references are given in the next section)

DEFINITION 1. Complete Random Markov Process (abbreviated C.R.M.). Let F be a finite set. Let $\{a_i, N_i\}$ be a stationary process where each $a_i \in F$, each $N_i \in \mathbf{N}$, N_0 is independent of $\{a_i, N_i\}_{i<0}$, and for each j

$$P(a_0 = k \mid a_{-1} a_{-2} \cdots a_{-j} \wedge N_0 = j) = P(a_0 = k \mid \{a_i\}_{i<0} \wedge N_0 = j).$$

Then $\{a_i, N_i\}_{i \in \mathbf{Z}}$ is a C.R.M.

DEFINITION 2. Random Markov Process (abbreviated R.M.). A Random Markov Process is the first coordinate of a C.R.M., i.e., if $\{a_i, N_i\}$ is a C.R.M. then $\{a_i\}$ is a R.M.

NOTE. If the $\{N_i\}$ process in the C.R.M. is bounded above by n, then the R.M. is an n-step Markov process. Thus, in general, an R.M. is a generalization of an n-step Markov chain.

DEFINITION 3. Uniform Martingale (abbreviated U.M). Let F be a finite set and let $\{a_i\}_{i \in \mathbf{Z}}$ be a stationary process, all $a_i \in F$. If, for all $\epsilon > 0$, there exists $N_\epsilon \in \mathbf{N}$ such that for all $M > N_\epsilon$ and all $\{F_i\}_{i=0}^{\infty}$ with all $F_i \in F$,

†This work was initiated at SUNY/Albany.

Received April 16, 1989

$$|P(a_0 = F_0 | a_{-1} = F_1, a_{-2} = F_2, \ldots, a_{-m} = F_m) - P(a_0 = F_0 | a_{-i} = F_i \text{ for all } i)| < \epsilon,$$

then a_i is a U.M.

The notion of an U.M. turns out to be exactly the same as Mike Keane's notion of a "continuous" g-function [4] (see References section).

NOTE. To say that a process is a U.M. is merely to say that the martingale convergence theorem holds uniformly on the martingale

$$\{P(a_0 = F_0 | a_{-1} = F_1, \ldots, a_{-m} = F_m)\}_{m=1}^{\infty}.$$

THEOREM 4. *U.M. = R.M.*

COMMENT. This theorem provides an easy way to check whether or not a process has a R.M. representation. All one must check is the U.M. condition, which is relatively easy to check.

THEOREM 5. *Any zero entropy* (see References section) *process which is not merely a finite state rotation cannot be a U.M.*

COMMENT. In a future paper by Y. Katznelson, B. Weiss and me, we show that every zero entropy process can be extended to a R.M.

EXAMPLE 6. There exists a transformation which is weak Bernoulli and not R.M.

COMMENT. We will give two examples of this: one simple example constructed precisely for the purpose of establishing Example 6 and also Example 16 provides another example.

THEOREM 7. *In a C.R.M., if $E(N_0) < \infty$, and some minimum extra condition such as weak mixing or $P(X_0 = s/\text{past})$ bounded below for some state s, then it is weak Bernoulli* (see References section).

COROLLARY 8. *Let X_n be a U.M. and, for each $\epsilon > 0$, let N_ϵ be as in the definition of U.M. If there exists a sequence ϵ_n which decreases geometrically but such that N_{ϵ_n} increases slower than geometrically (and the same minimal extra condition as Theorem 7) then X_n is weak Bernoulli.*

This corollary may follow from the work of Petit [5] (see References section).

DEFINITION 9. Consider a C.R.M. For simplicity, consider a two-valued C.R.M. $\{0,1\}$. The values of $P(X_0 = 1 | X_{-1}, X_{-2}, \ldots, X_{-n} \wedge N_0 = n)$ are called the *table*. The table together with the values of $P(N_0 = n)$ for every n is called a *R.M. representation* of the canonical R.M. factor. We also say that this canonical R.M. factor *supports* the R.M. representation.

COMMENT. The R.M. representation together with the canonical R.M. factor determines the C.R.M.

DEFINITION 10. If there is a R.M. factor which can support a given R.M. representation, then we say that the given R.M. representation satisfies *existence*. If there do not exist two R.M. factors which can support a given R.M. representation, then we say that the given R.M. representation satisfies *uniqueness*.

THEOREM 11. *All R.M. representations satisfy existence.*

COMMENT. This is proved by Keane [4].

EXAMPLE 12. Some R.M. representations satisfy uniqueness and some don't.

CONJECTURE. Consider the following R.M. representation. Choose a rapidly increasing sequence of positive odd integers a_n. Let the support of N_0 be on the values $\{a_n\}_{n=0}^{\infty}$. Let $\{0,1\}$ be the support of the $\{X_i\}_{i=-\infty}^{\infty}$. Let $P(N_0 = a_i) = 1/2^i$ and let the table be defined as follows:

$$P(X_0 = 1 \mid X_{-1}, X_{-2}, \ldots, X_{-a_n} \wedge N_0 = a_n)$$
$$= \begin{cases} 0.9 & \text{if a majority of } X_{-1}, \ldots, X_{-a_n} \text{ are } 1, \\ 0.1 & \text{else.} \end{cases}$$

Weiss and I conjecture that this R.M. representation does not satisfy uniqueness.

COMMENT. We have considerable evidence for our conjecture but we do not have a rigorous proof. If the conjecture is valid it shoots down a reasonable hope for a theorem implying uniqueness, namely that when the table probabilities are bounded away from zero and one, you get uniqueness. It should be mentioned that even if this conjecture cannot be established, Mike Keane claims that he has established another example shooting down that possibility.

EXAMPLE 13. There exists a R.M. that is K and not Bernoulli (see References section).

COMMENT. This raises hopes that maybe all K transformations can be extended to a R.M. which is K.

DEFINITION 14. *B.U.M.* A U.M. with $P(X_0 = s|\text{past})$ bounded away from 0 and 1, for each state s, is called a B.U.M. Here, "past" means "$X_{-1}, X_{-2}, \ldots$".

THEOREM 15. *Let $\{X_i\}$ be a B.U.M. and, in particular, suppose for each s, $P(X_0 = s|\text{past})$ lies between ϵ and $1-\epsilon$. Then for any positive $\delta < \epsilon$, $\{X_i\}$ can be extended to a C.R.M. with table between δ and $1-\delta$.*

EXAMPLE 16. The inverse of a R.M. with $E(N_0) < \infty$ need not be a U.M.

COMMENT. Example 16 gives another example for Example 6 because $E(N_0) < \infty$ with minimal extra condition (which is satisfied here) implies weak Bernoulli, and weak Bernoulli is closed under inverse.

EXAMPLE 17. The inverse of a B.U.M. need not be a U.M.

THEOREM 18. *The inverse of a B.U.M. which has a R.M. representation with $E(N_0) < \infty$ must be a B.U.M.*

EXAMPLE 19. There exists a B.U.M. with a R.M. representation with $E(N_0) < \infty$ such that its inverse has no R.M. representation with $E(N_0) < \infty$.

COMMENT. This is frustrating. I can't seem to find any class of U.M.'s which is closed under inversion. This is annoying, because R.M.'s are a generalization of n-step Markov chains, and n-step Markov chains are closed under inversion.

CONJECTURE. $E(N_0^n) < \infty$ and B.U.M. implies that the inverse has a representation with $E(N_0^{n-1}) < \infty$.

CONJECTURE. There exists, for each n, an example with $E(N_0^n) < \infty$ and B.U.M. but where the inverse has no R.M. representation with $E(N_0^n) < \infty$.

2. References section

A stationary process $\{x_i\}$ is called *weak Bernoulli* [6] if, for every $\epsilon > 0$, there is a joint process $\{y_i, z_i\}$ such that

(1) the $\{y_i\}$ process has the same distribution as the $\{x_i\}$ process,

(2) the $\{z_i\}$ process has the same distribution as the $\{x_i\}$ process,

(3) $\{y_i\}_{i<0}$ is independent of $\{z_i\}_{i<0}$,

(4) with probability $1-\epsilon$, there exists N such that $\forall n > N$, $y_n = z_n$.

A process has *zero entropy* if the past determines the present. A general discussion of entropy can be found in [7].

A stationary process is called K [8] (for Kolmogorov) if its tailfield is trivial (equivalently if $P(x_n = a_0, x_{n+1} = a_1, \ldots, x_{n+m} = a_m \mid past)$ approaches $P(x_0 = a_0, \ldots, x_m = a_m)$ as $n \to \infty$ for all pasts, $m, a_0, a_1, \ldots, a_m$.)

A stationary process $\{x_i\}$ is called Bernoulli [9] if there is an i.i.d. process $\{y_i\}$ which is isomorphic to $\{x_i\}$. This means that if P, Q are the sets of values that x_0, y_0 take on respectively, and if $T: P^{\mathbf{Z}} \to P^{\mathbf{Z}}$ is defined by $(t\alpha)_i = \alpha_{i+1}$ and $\hat{T}: Q^{\mathbf{Z}} \to Q^{\mathbf{Z}}$ defined by $(\hat{T}\beta)_i = \beta_{i+1}$, then there exists $\phi: P^{\mathbf{Z}} \to Q^{\mathbf{Z}}$ such that

(1) ϕ is a measurable bijection up to measure 0,

(2) ϕ and ϕ^{-1} preserves measure (where the x_i process is regarded as a measure on $P^{\mathbf{Z}}$ and the y_i process on $Q^{\mathbf{Z}}$),

(3) $\hat{T} \circ \phi = \phi \circ T$.

Example 13 says that there is a R.M. which is K and not Bernoulli. This is obtained by referring to the TT^{-1} process, which is known to be K and not Bernoulli. This is proved in [3].

Theorem 11 and Example 12 discuss existence and uniqueness of R.M. representations. A R.M. representation gives you the probability of the present given the past. The general notion of starting with the probability of the present given the past and asking the questions of existence and uniqueness is already well known.

The notion of probability of present given past is known in the literature as a g-function. Berbee [1] defines a g-function as follows: A non-negative function g on $I \times \prod_{n \le 1} (I)$ is called a g-function if $\sum_{i_0 \in I} g(i_0 \mid i_{-1}) = 1$ for $i_- \in I_- = \prod_{n \le -i} I$.

These processes were introduced by Doeblin and Fortet [2]. Keane [4] proved an existence result which is my Theorem 11. He introduced the notion of a "continuous" g-function and demonstrated that all continuous g-functions satisfy existence. The notion is precisely the same thing as the notion of a U.M. He also had uniqueness results where under certain conditions g-functions have unique measures which were mixing. Berbee [1] developed more general uniqueness results. Petit [5] extended Keane's work where Keane's "continuous" notion was replaced by "differentiable" and he got that all differentiable g-functions that are bounded away from zero and one have unique measures which are weak Bernoulli. It is quite likely that my Corollary 8 follows from this result.

3. Proofs of theorems and examples demonstrated

Proof of Theorem 4. We wish to prove U.M. = R.M. Obviously R.M. → U.M. where N_ϵ is chosen so that $P(N_0 > N_\epsilon) < \epsilon$. We now show U.M. → R.M. For simplicity assume X_0 takes on only two values, 0 and 1 (the proof can be carried out in the event that X_0 takes on more than two values, but the extra complication would confuse the reader).

Case 1. $P(X_0 = 1 \mid \text{past})$ is bounded below by a bound that does not depend on the past.

We will now construct a R.M. Let $\hat{P}$ be the probability law of the R.M. that we construct. Let P be the probability law of the U.M. When one is given an U.M. and wants to show it to be a R.M., one has to merely exhibit a table and a lookback time for the process so that the value $\hat{P}(\text{present}\mid\text{past})$ for the table and the lookback time equals $P(\text{present}\mid\text{past})$ for the U.M. We do not actually have to check that $\hat{P}$ has the same cylinder set probabilities as P because the table and lookback time only give $\hat{P}(\text{present}\mid\text{past})$ so once $\hat{P}(\text{present}\mid\text{past}) = P(\text{present}\mid\text{past})$ we can simply *define* the cylinder probabilities of $\hat{P}$ to be the same as those of P. Thus if we can construct the R.M. so that for every past

(a) $P(X_0 = 1 \mid \text{past}) = \hat{P}(X_0 = 1 \mid \text{past})$,

then we are done. We choose a rapidly increasing sequence of positive integers $\hat{N}_0, \hat{N}_1, \hat{N}_2, \ldots$ and let them be the support of N_0 with

$$\hat{P}(N_0 = \hat{N}_i) = \frac{1}{2^i}.$$

Now all that is left is to define the table, i.e. we need to define

$$T = \hat{P}(X_0 = 1 \mid X_{-1} = a_{-1}, X_{-2} = a_{-2}, \ldots, X_{-\hat{N}_i} = a_{-\hat{N}_i} \wedge N_0 = \hat{N}_i).$$

Let b range over all sequences $b_1, b_2, \ldots$.

We will define T by induction on i so that, at each stage of the induction, we can ensure that

(b) $\sup_b P(X_0 = 1 \mid X_{-1} = a_{-1}, \ldots, X_{-N_i} = a_{-\hat{N}_i}, X_{-(\hat{N}_i+1)} = b_1, X_{-(\hat{N}_i+2)} = b_2, \ldots)$
$= \hat{P}(X_0 = 1 \mid X_{-1} = a_{-1}, X_{-2} = a_{-2}, \ldots, X_{-\hat{N}_i} = a_{-\hat{N}_i}, N_0 \le \hat{N}_{i-1})$.

Clearly, if this holds for all i, we will achieve our desired goal, (a), for all pasts. By the induction hypothesis we have (again, let b range over all sequences $b_1, b_2, \ldots$)

(c) $\sup_b P(X_0 = 1 \mid X_{-1} = a_{-1}, \ldots, X_{-\hat{N}_{i-1}} = a_{-\hat{N}_{i-1}}, X_{-(N_{i-1}+1)} = b_1, X_{-(\hat{N}_{i-1}+2)} = b_2, \ldots)$
$= \hat{P}(X_0 = 1 \mid X_{-1} = a_{-1}, \ldots, X_{-\hat{N}_{i-1}} = a_{-\hat{N}_{i-1}}, N_0 \le \hat{N}_{i-1})$.

We are assuming (c) and trying to establish (b). As will be shown, (b) and (c) together determine T. This T, together with (c), implies (b), All that needs to be shown is that $0 \le T \le 1$. Therefore we assume both (b) and (c). Let l_1 and r_1 be the left and right sides of (b) respectively. Let l_2 and r_2 be the left and right sides of (c) respectively. It is clear that

$$\left(r_2\left(1-\frac{1}{2^{i-1}}\right)+T\left(\frac{1}{2^i}\right)\right)\Big/\left(1-\frac{1}{2^i}\right)=r_1;$$

this expresses r_1 as a weighted average of r_2 and T which we denote by $w(r_2,T)=r_1$. Since we are assuming $l_1=r_1$ and $l_2=r_2$ we have

(d) $w(l_2,T)=l_1$.

This equation solves for T. All that is necessary is to show $0\le T\le 1$. Clearly $l_2\ge l_1$ so it follows that $T\le l_1\le 1$.

By choosing n_{i-1} large enough we can insist on l_1 and l_2 being close to each other, and since they are bounded below it follows that $T>0$ and we are done. □

Case 2.

We drop the assumption of case 1. We assume nothing.

Let $\hat{N}_0$ be chosen so that $\hat{N}_0>N_{1/6}$. This means that for a given sequence a_{-1}, $\dots,a_{-\hat{N}_i}$, if

$$P(X_0=1\,|\,X_{-1}=a_{-1}\dots X_{-\hat{N}_0}=a_{-\hat{N}_0})\ge\frac{1}{2}$$

then $P(X_0=1\,|\,X_{-1}=a_{-1},\dots,X_{-\hat{N}_i}=a_{-\hat{N}_i},X_{-(\hat{N}_i+1)}=b_1,X_{-(\hat{N}_i+2)}=b_2,\dots)$ is bounded below by $\frac{1}{3}$ independent of i, and $b_1,b_2,\dots$. Similarly, if

$$P(X_0=1\,|\,X_{-1}=a_{-1},\dots,X_{-\hat{N}_0}=a_{-\hat{N}_0})<\frac{1}{2}$$

then $P(X_0=1\,|\,X_{-1}=a_{-1},\dots,X_{-\hat{N}_i}=a_{-\hat{N}_i},X_{-(\hat{N}_i+1)}=b_1,\dots)$ is bounded above by $\frac{2}{3}$. Choose the table to inductively make sure that

$$\hat{P}(X_0=1\,|\,X_{-1}=a_{-1},\dots,X_{-\hat{N}_i}=a_{-\hat{N}_i},N_0\le\hat{N}_i)$$

$$=\begin{cases}\sup_{b_1,b_2,\dots}P(X_0=1\,|\,X_{-1}=a_{-1},\dots,X_{-\hat{N}_1}=a_{-\hat{N}_i},X_{-(\hat{N}_i+1)}=b_1,\dots)\\ \qquad\text{if } P(X_0=1\,|\,X_{-1}=a_{-1},\dots,X_{-\hat{N}_0}=a_{-\hat{N}_0})\ge\frac{1}{2},\\ \inf_{b_1,b_2,\dots}P(X_0=1\,|\,X_{-1}=a_{-1},\dots,X_{-\hat{N}_i}=a_{-\hat{N}_i},X_{-(\hat{N}_i+1)}=b_1,\dots)\\ \qquad\text{if } P(X_0=1\,|\,X_{-1}=a_{-1},\dots,X_{-\hat{N}_0}=a_{-\hat{N}_0})<\frac{1}{2},\end{cases}$$

we are essentially in case 1. □

PROOF OF THEOREM 5. Since the process is not a finite state rotation, there must exist, for all n, a sequence $a,b_1,b_2,b_3,\dots,b_n$ such that $P(X_0=a\,|\,x_1=b_1,\dots,X_n=b_n)$ is neither 0 nor 1. Since the process has 0 entropy, the finite sequence $b_1,b_2,\dots,b_n$ can be extended to two distinct pasts, past_1 and past_2, such that

$$P(X_0=a\,|\,\text{past}_1)=1\quad\text{and}\quad P(X_0=a\,|\,\text{past}_2)=0.$$

EXAMPLE 6. Consider a Bernoulli $\frac{1}{2}$, $\frac{1}{2}$ sequence of 0's and 1's. Every time you see a zero, then 10^n ones ($n \geq 1$, $n \in \mathbf{N}$), then a zero, cross out the 10^n ones. Consider the process made up of the remaining 0's and 1's. It may be objected that this process is not well defined because one can't tell where the origin is after the ones have been crossed out. However, a process is well defined if one can explicitly define the cylinder set probabilities, and for this partially crossed out process, it is clear what the cylinder set probabilities are.

It is weak Bernoulli because whenever there is a zero on the origin, conditioned on that zero, the past is independent of the future.

I will make the argument of the previous paragraph more rigorous. Join two copies of the process so that the pasts are independent. Let the futures be joined independently also until there is a zero on both coordinates. The two conditional measures from that point on are the same, so we can couple them to be identical.

However, it is not a U.M. because $P(X_0 = 1 \mid \text{past } 5(10^n) \text{ terms all one})$ is close to $\frac{1}{2}$, for large n, but $P(X_0 = 1 \mid \text{past } 10^{n+1} \text{ terms all one and } 10^{n+1} + 1 \text{ term } 0) = 1$.

PROOF OF THEOREM 7. We join two copies of the process together as follows. Join the pasts independently.

Join their futures independently until the R.M. factors agree on a long stretch (which will eventually happen by the minimal extra condition). After this long stretch (which I will refer to as a gap) I will join the two processes to be identical until they look before the gap, i.e. we make both look just as far back and if they don't look before the gap, the probability of the next term is the same for both processes, so we can make the next term identical for both processes; the condition $E(N_0) < \infty$ precisely says that it is unlikely that they will ever look before the gap, if the gap is long enough. □

PROOF OF COROLLARY 8. In the proof that U.M. → R.M. we select a sequence N_i and let $P(N_0 = N_i) = 1/2^i$. The N_i's have to increase rapidly enough so that ϵ_{N_i} is small in comparison to $1/2^i$. The conditions of this corollary guarantee that we can do this while choosing the N_i's to grow slower than geometrically, thereby making $E(N_0) < \infty$. In particular, we can let $N_n = N_{\epsilon_{kn}}$ for large fixed k, where $N_{\epsilon_{kn}}$ is as in the definition of U.M. and ϵ_{kn} is defined in the statement of the corollary. □

PROOF OF THEOREM 11. Keane [4] proves this result; for the sake of the reader I include the proof here.

Choose the past arbitrarily. Once the past is chosen, the R.M. representation allows you to run the future (i.e. the R.M. representation gives $P(\text{time } 0 \mid \text{past})$,

P(time 1|time 0 and past) etc.). Thus run the process into the future. When we are finished we have a randomly chosen doubly infinite word $\{a_i\}_{i\in\mathbf{Z}}$ where $\{a_i\}_{i<0}$ is determined.

Define a measure u_n on words of length n by

$$u_n(w) = \frac{1}{2^{2^n}}\,\#\{i : 0 \le i < 2^{2^n} \text{ and } w = a_i, a_{i+1}, \ldots, a_{i+n-1}\}$$

for any word w of length n. Let u be a weak limit of the measures u_n. Then u defines a stationary process. We now show that u supports the given R.M. representation. This precisely means that

(a) $u(X_0 = 1 \mid \text{past})$ is the value $P(X_0 = 1 \mid \text{part})$ given by the R.M.

We show (a) by showing

(b) $\left| u\big(X_0 = 1 \mid X_{-1} = b_1, X_{-2} = b_2, \ldots, X_{-(n-1)} = b_{n-1}\big) - P\big(X_0 = 1 \mid X_{-1} = b_1, X_{-2} = b_2, \ldots, X_{-(n-1)} = b_{n-1} \wedge N_0 \le n-1\big)\right|$ goes to zero as $n \to \infty$.

Note: I must write "$N_0 \le n - 1$" in (b) because $P(X_0 = 1 \mid X_{-1} = b_1, X_{-2} = b_2, \ldots, X_{-n-1} = b_{n-1})$ is not defined.

Let L be a large number, $L \gg n$. We will show (b) by showing, for any fixed n, that for sufficiently large L,

(c) $|u_L(X_0 = 1 \mid X_{-1} = b_1, \ldots, X_{-(n-1)} = b_{(n-1)}) - P(X_0 = 1 \mid X_{-1} = b_1, \ldots, X_{-(n-1)} = b_{n-1} \wedge N_0 \le n-1) \to 0$ as n approaches infinity.

(c) is clear from the strong law of large numbers and the way a_i is chosen once $a_{i-1} = b_1$, $a_{i-2} = b_2, \ldots, a_{i-(n-1)}$ is known. To make it easy for the reader, I will remind him how a_i is chosen. Start with $a_{i-1} = b_1, a_{i-2} = b_2, \ldots, a_{i-(n-1)} = b_{n-1}$. Now independent of this information choose N_i (which usually turns out to be less than $n-1$). Then, given $N_i = k$ (which we presume to be less than $n - 1$) we choose a_i to be 1 with probability $P\big(X_0 = 1 \mid X_{-1} = b_1, \ldots, X_{-(n-1)} = b_{n-1}, N_0 = k\big)$. (c) follows easily from the strong law of large numbers. □

Example 12. $N_0 = 1$ always, and $X_0 = X_{-1}$ always, admits two measures, the measures all $X_i = 0$, and all $X_i = 1$. $N_0 = 0$ always, and $X_0 = 0$ with probability $\frac{1}{2}$ and $X_0 = 1$ with probability $\frac{1}{2}$ always, admits only i.i.d. $\frac{1}{2}, \frac{1}{2}$ product measure.

Example 13. The T, T^{-1} transformation is a transformation which is K and not Bernoulli (see the References section). We will define the T, T^{-1} transformation here for the benefit of the reader. Here we extend the transformation to a R.M. which is still K. It remains non-Bernoulli because the property non-Bernoulli is closed under extension.

The T, T^{-1} transformation is defined as follows. We will describe the T, T^{-1}

transformation as a process rather than as a transformation. The T, T^{-1} process has a four-letter alphabet, $\binom{H}{L}$, $\binom{H}{R}$, $\binom{T}{L}$, $\binom{T}{R}$. Here the H and T stand for head and tail, and L and R stand for left and right. First select a random doubly infinite sequence $\{a_i\}_{i\in\mathbf{Z}}$, where each $a_i \in \{H, T\}$. The a_i's are chosen with $\frac{1}{2}, \frac{1}{2}$ product measure. Similarly, a random sequence $\{b_i\}_{i\in\mathbf{Z}}$, each $b_i \in \{R, L\}$, is chosen with $\frac{1}{2}, \frac{1}{2}$ product measure. The sequence $\{a_i\}$ is called the scenery, and $\{b_i\}$ is called the path. The sequence $\{c_i\}_{i\in\mathbf{Z}}$ is defined by

$$c_i = \begin{bmatrix} a_j \\ b_i \end{bmatrix} \qquad \text{where } j = \begin{cases} 0 & \text{if } i = 0, \\ \#\{k : 0 \le k < i \wedge b_k = R\} - & \\ \#\{k : 0 \le k < i \wedge b_k = L\} & \text{if } i > 0, \\ \#\{k : i \le k < 0 \wedge b_k = L\} - & \\ \#\{k : i \le k < 0 \wedge b_k = R\} & \text{if } i < 0. \end{cases}$$

The process $\{c_i\}_{i\in\mathbf{Z}}$ is called the TT^{-1} process. The way that this is supposed to be thought about is that every time you see an L, i.e. whenever $b_i = L$, the entire scenery shifts to the right (i.e. the origin shifts to the left). Every time you see an R, the entire scenery shifts to the left (i.e. the origin shifts to the right). a_j is just the 0 coordinate of the shifted scenery.

Definition of "*consistent*". A doubly infinite word made from the alphabet $\binom{H}{L}$, $\binom{H}{R}$, $\binom{T}{L}$, $\binom{T}{R}$ is said to be *consistent* if it can be obtained from a scenery and a path, as in the above definition of the T, T^{-1} process.

Definition. Let $\binom{a_i}{b_i}_{i\in\mathbf{Z}}$ be a doubly infinite sequence from the alphabet $\binom{H}{L}$, $\binom{H}{R}$, $\binom{T}{L}$, $\binom{T}{R}$. Let $i_1 < i_2$. We say that i_1 and i_2 see the *same piece of scenery* if

$$\#\{i; i_1 \le i < i_2 \wedge b_i = L\} = \#\{i : i_1 \le i < i_2 \wedge b_i = R\}.$$

Proposition. *A doubly infinite word* $\binom{a_i}{b_i}_{i\in\mathbf{Z}}$ *made from the alphabet* $\binom{H}{L}$, $\binom{H}{R}$, $\binom{T}{L}$, $\binom{T}{R}$ *is consistent iff* $a_{i_1} = a_{i_2}$ *whenever* i_1 *and* i_2 *see the same piece of scenery.*

The proof is left to the reader.

Definition. *A* T, T^{-1}*-question mark path* is a doubly infinite sequence $\binom{a_i}{b_i}$ from the six-letter alphabet $\binom{H}{L}$, $\binom{H}{R}$, $\binom{T}{L}$, $\binom{T}{R}$, $\binom{?}{L}$, $\binom{?}{R}$.

Definition. A T, T^{-1}-question mark path is said to be *?-consistent* if there is a way to replace each "?" with either a "H" or a "T" in such a way that the resulting path is consistent.

PROPOSITION. *A T, T^{-1}-question mark process is consistent if there does not exist i_1 and i_2 which see the same piece of scenery with $a_{i1} = H$ and $a_{i2} = T$.*

The proof is left to the reader.

DEFINITION. Choose $\hat{N}_0$ to be a huge positive number, and choose $\hat{N}_{i+1}$ so that $\hat{N}_{i+1} \gg \hat{N}_i$. *The canonical process* is a doubly infinite i.i.d. sequence of random variables $\{N_i\}_{i \in \mathbf{Z}}$ where, for any i and j, $P(N_i = \hat{N}_j) = 1/2^j$.

DEFINITION. "0 order process". If you cross the T, T^{-1} process with the canonical process you get the 0-order process.

DEFINITION. "1 order process". Start with the 0 order process. Redefine any $\left(\binom{a_i}{b_i}, N_i\right)$ to be $\left(\binom{?}{b_i}, N_i\right)$ if there is no j, $i - N_i \le j < i$, such that j sees the same piece of scenery as i. The new process is called the 1-order process.

DEFINITION. "2 order process". Start with the 1-order process. Redefine any $\left(\binom{a_i}{b_i}, N_i\right)$ to be $\left(\binom{?}{b_i}, N_i\right)$ if there does not exist j, $i - N_i \le j \le i$, such that j sees the same piece of scenery as i and $a_j \ne$ "?". The resulting process is called the *2-order process.*

DEFINITION. The *n-order process* is defined the same as the 2-order process except that instead of starting with the 1-order process, you start with the $n - 1$ order process.

DEFINITION. An *n-order question mark* is an a_i which is a question mark for the n-order process but not for the $(n - 1)$-order process.

DEFINITION. The *final process* is obtained by starting with the 0-order process and changing each a_i to a "?" if it is an n-order question mark for some n.

DEFINITION. The *final factor* is the factor of the final process obtained by removing all the N_i, $i \in \mathbf{Z}$.

The final factor is the desired example (Example 15). To prove this we need to show that the final factor

(1) is an R.M.,

(2) is K,

(3) is an extension of the TT^{-1} process.

PROOF OF 1. The final process is actually a C.R.M. At time zero you look back N_0 and see if there is any i, $-N_0 \le i < 0$ such that $a_i \ne$ "?" and i sees the same piece of scenery as zero. If there is such an i then $a_0 = a_i$. Otherwise $a_0 =$ "?".

Thus, the final process is a C.R.M. and the final factor is its canonical R.M. factor.

PROOF OF 2. The final factor is a factor of the final process, which in turn is a factor of the 0-order process which is the product of the T, T^{-1} process with an independent process, both of which are K.

PROOF OF 3. Definition. "son". In this definition it does not matter whether we consider the 0-order process or the final process because we only talk about $\{b_i\}$ and $\{N_i\}$, which are the same for the 0-order process and the final process. Let $i \in \mathbf{Z}$, $N_i = \hat{N}_k$, and if there is a number j, $i - N_i \leq j < i$ such that j sees the same piece of scenery as i and $N_j = \hat{N}_{k+1}$, then j is a *son* of i.

LEMMA. *If there is a sequence of integers $\{k_j\}_{j \in \mathbf{N}}$ such that $k_0 = i$ and for all j, k_{j+1} is a son of k_j, then a_i is not a "?".*

PROOF. None of the a_{k_j} is a first order "?". Therefore, none of them is a second order "?" etc.

COROLLARY. *If $N_0 = \hat{N}_k$, then $P(a_0 = \text{"?"}) < 1/2^k$.*

PROOF. Keep in mind that if j is the son of i and $N_i = \hat{N}_L$, then $N_j = \hat{N}_{L+1}$. The result follows from the lemma if the $\hat{N}_i$ grow fast enough.

COROLLARY. You can recover all in the 0-order process from the final process (i.e. the final process is an extension of the 0-order process).

PROOF. Fix $\epsilon > 0$ and $i \in \mathbf{Z}$. I will prove you can recover a_i by proving P(you cannot recover a_i) $< \epsilon$. Choose k so that $1/2^k < \epsilon$. Since random walk is recurrent there are infinitely any $j < i$ such that j sees the same piece of scenery as i. Therefore there must exist such a j where $N_j = \hat{N}_k$. The result follows from the previous corollary.

All that is left to do is to show that the final factor is an extension of the T, T^{-1}.

DEFINITION. A *descendent* of i of a $j < i$ which sees the same piece of scenery as i.

When we proved the final process to be an extension of T, T^{-1}, all we needed was to prove "Each i has a descendent which is not a question mark". This statement does not refer to lookback times. It continues to hold for the final factor.

PROOF OF THEOREM 15. We define the table values to inductively force the following equation to hold for all k;

$$P(X_0 = 1 \mid X_1 X_2 \ldots X_{n_k}) = \hat{P}(X_0 = 1 \mid X_0 X_1 \ldots X_{n_k} \wedge N_0 \leq n_k).$$

By letting $P(N_0 = n_k) = 1/2^k$ and then letting n_k grow rapidly enough, our result is obtained (see proof of Theorem 4).

EXAMPLE 16. We define a R.M. representation. The alphabet is $\{0,1\}$. Let $P(N_0 = n + 20) = 1/2^n$. The table is defined as follows. Look back N_0. If $X_{-1} = 0$ and if there is no $i \leq N_0 - 2$ such that

(a) $X_{-i} = 1,\ X_{-(i+1)} = 0,\ X_{-(i+2)} = 1$

or

(b) $X_{-i} = 1,\ X_{-(i+1)} = 0,\ X_{-(i+2)} = 0$

then let $X_0 = 1$ with probability 1. Otherwise (still considering the case where $X_{-1} = 0$) choose the smallest $i \leq N_0 - 2$ where (a) or (b) holds. If $X_{-(i+2)} = 1$, then let $X_0 = 1$ with probability 0.9 and if $X_{-(i+2)} = 0$ then let $X_0 = 0$ with probability 0.9. If $X_{-1} = 1$ let $P(X_0 = 1) = \frac{1}{2}$ independent of N_0. From here on, we let P be the probability measure of a C.R.M. with the above R.M. representation. Let $\hat{P}$ be the probability measure of the inverse of the canonical R.M. factor of the C.R.M. Our purpose is to show that $\hat{P}$ is not a U.M.

We do this by proving that A and B differ substantially where

$$A = P(X_0 = 0 \mid X_1 = 0, X_2 = X_3 = \cdots = X_{1+m} = 1, X_{m+2} = \cdots = X_{1+m+n} = 0),$$

and

$$B = P(X_0 = 0 \mid X_1 = 0, X_2 = X_3 = \cdots = X_{1+m} = 1),$$

m chosen large, n chosen much larger. We do this by proving that B does not depend on m and $B \neq 1$, but that A approaches 1 as $n \to \infty$.

Proof that B does not depend on m and B ≠ 1. Let "$X_0 = 0$" $= C$, "$X_1 = 0$, $X_2 = X_3 = \cdots = X_{1+m} = 1$" $= D$. Looking at our R.M. table we see that

$$P(X_3 = X_4 = \cdots = X_{1+m} = 1 \mid X_0 = 0, X_1 = 0, X_2 = 1) = \left(\frac{1}{2}\right)^{m-2},$$

$$P(X_3 = X_4 = \cdots = X_{1+m} = 1 \mid X_1 = 0, X_2 = 1) = \left(\frac{1}{2}\right)^{m-2}.$$

Therefore $B = P(C \mid D) = P(C \wedge D)/P(D) = P(X_0 = 0, X_1 = 0, X_2 = 1)\left(\frac{1}{2}\right)^{m-2} \div P(X_1 = 0, X_2 = 1)\left(\frac{1}{2}\right)^{m-2} = P(X_0 = 0, X_1 = 0,\ X_2 = 1) \div P(X_1 = 0, X_2 = 1)$ which does not depend on m.

All that is necessary is to prove $P(X_0=0,X_1=0,X_2=1)/P(X_1=0,X_2=1)\neq 1$, which is equivalent to proving $P(X_0=1,X_1=0,X_2=1)>0$. Given any past at all, and given any value for N_0, $P(X_0=1)\geq \frac{1}{10}$. Given $X_0=1$ and any past, $P(X_1=0)=\frac{1}{2}$. Given any past, any value for X_1 and X_2, and any N_2, $P(X_2=1)\geq \frac{1}{10}$ so

$$P(X_0=1,X_1=0,X_2=1)\geq \left(\frac{1}{10}\right)\left(\frac{1}{2}\right)\left(\frac{1}{10}\right)=\frac{1}{200}>0.$$

Proof that A approaches 1 as n approaches ∞. Let "$X_0=0$" $=C$, "$X_0=1$" $=F$ and, as above, "$X_1=0,X_2=X_3=\cdots=X_{1+m}=1$" $=D$ and let "$X_1=0,X_2=\cdots=X_{1+m}=1,X_{m+2}=\cdots=X_{1+m+n}=0$" $=E$. $A=P(C\mid E)=P(C\cap E)/P(E)=P(C\cap E)/(P(C\cap E)+P(F\cap E))$. To say that A is close to one is equivalent to saying that

(a) $P(C\cap E)$ is much bigger than $P(F\cap E)$.

Select some i, $m+2<i<1+m+n$. Let $E_i=$ "$X_1=0,X_2=X_3=\cdots=X_{m+1}=1,X_{m+2}=X_{m+3}=\cdots=X_i=0$".

We will now compare $P(E_i\cap C)\div P(E_{i-1}\cap C)$ with $P(E_i\cap F)\div P(E_{i-1}\cap F)$.

$$P(E_i\cap C)\div P(E_{i-1}\cap C)=P(X_i=0\mid E_{i-1}\cap C).$$

We can compute this using the R.M. rule. If $N_i<i$, and E_{i-1}, then X_i must be 1. If $N_i\geq i$, and $E_{i-1}\cap C$, then $X_i=0$ with probability 0.9. Thus $P(E_i\cap C)\div P(E_{i-1}\cap C)=P(X_i=0\mid E_{i-1}\cap C)=0.9\,P(N_i\geq i)$.

Similarly, $P(E_i\cap F)\div P(E_{i-1}\cap F)=0.1(P(N_i\geq i))$. Hence,

$$[P(E_i\cap C)\div P(E_{i-1}\cap C)]\div[P(E_i\cap F)\div P(E_{i-1}\cap F)]=9.$$

It follows that $[P(C\cap E)\div P(C\cap D\cap X_{m+2}=0)]\div[P(F\cap E)\div P(F\cap D\cap X_{m+2}=0)]=9^{n-2}$. Since $P(C\cap D\cap(X_{m+2}=0))$ and $P(F\cap D\cap X_{m+2})$ are fixed non-zero constants that don't depend on n, (a) is proved; we are done. (To see that they are non-zero, just note that they are non-zero for any fixed past except the all zero past. Because 1 has at least $\frac{1}{10}$ probability given any past, it follows that the all zero past has zero probability.)

Example 17. This example is almost identical to the previous one with minor modifications. If we want we can explicitly describe a distribution on N_0, such as $P(N_0=2^n)=1/2^n$, but really all we use is that $E(N_0)=\infty$. The table is identical to that of the table of the previous example except that when the previous table says $P(X_0=1\mid N_0$ and past$)=1$, we instead have $P(X_0=1\mid N_0$ and past$)=\frac{1}{2}$.

Define C, E, and F as in the previous example. As in that example we need only show that $P(C \cap E)$ is much larger than $P(F \cap E)$. The condition $E(N_0) = \infty$ precisely says that if n is chosen large (m and n defined as in the previous example) then there will usually be many i, $m + 2 < i < 1 + m + n$, with $N_i \geq i$ (because altogether there are infinitely many i with $N_i \geq i$). Condition on $N_0, N_1, N_2, \ldots$. For any i, $m + 2 < i < 1 + m + n$ in which $N_i \geq i$,

$$[P(E_i \cap C) \div P(E_{i-1} \cap C)] \div [P(E_i \cap F) \div P(E_{i-1} \cap F)] = 9$$

(all terms defined as in the previous example). The result follows by the argument of the previous example.

PROOF OF THEOREM 18. We start out by making some general arguments about sets on a probability space. Let θ_1, θ_2, θ_3, and θ_4 be four sets in a probability space. Fix a small ϵ and suppose it is our goal to show $|P(\theta_3 | \theta_2) - P(\theta_3)| < \epsilon$. It suffices to show (a), (b), and (c) below.

(a) $P(\theta_1)$ is almost 1 (the meaning of "almost" is chosen after ϵ is chosen).

(b) Almost all of θ_2 is in θ_1 (i.e. $P(\theta_2 \cap \theta_1)/P(\theta_2)$ is almost 1).

(c) θ_2 and θ_3 are independent given θ_1.

Hence, if it is our goal to show

(d) $|P(\theta_3 | \theta_2 \cap \theta_4) - P(\theta_3 | \theta_4)| < \epsilon$,

it suffices to show (e), (f) and (g) below,

(e) $P(\theta_1 | \theta_4)$ is almost 1 ("almost" chosen after ϵ).

(f) Given θ_4, almost all θ_2 is in θ_1.

(g) θ_2 and θ_3 are independent given θ_1 and θ_4.

We now consider the problem of showing that the inverse process is a U.M. We wish to show that there exists $\{\epsilon_m\}_{m=1}^{\infty}$ such that $\lim_{m \to \infty}(\epsilon_m) = 0$ and the following holds:

$$|P(X_0 = a_0)| X_1 = a_1, \ldots, X_m = a_m, X_{m+1} = a_{m+1}, \ldots, X_{m+n} = a_{m+n})$$
$$- P(X_0 = a_0 | X_1 = a_1, \ldots, X_m = a_m)| < \epsilon_m.$$

The original process has lookback distances N_i. Let l_i be the event $N_i \geq i$. The condition $E(N_0) < \infty$ precisely says that $\{P(l_i)\}_{i=1}^{\infty}$ is summable. Let θ_1 be the event that l_i fails for all $i > m$. Let θ_2 be the event

$$\text{"}X_{m+1} = a_{m+1}, X_{m+2} = a_{m+2}, \ldots, X_{m+n} = a_{m+n}\text{"}.$$

Let θ_3 be the event "$X_0 = a_0$" and let θ_4 be the event "$X_1 = a_1, X_2 = a_2, \ldots, X_m = a_m$". The statement that the inverse process is a U.M. is precisely (d) with ϵ replaced by ϵ_m. (d) will be established once we establish (e), (f), and (g).

PROOF OF (e). θ_1 is independent of θ_4 so we need only show $P(\theta_1)$ is almost 1. This follows from the fact that $\sum_{m+1}^{m+n} P(l_i)$ can be made arbitrarily small just by choosing m large.

PROOF OF (f). Throughout this proof, the reader is expected to remember that everything is conditioned on θ_4. I will not keep repeating that.

We now compare $P(\theta_2 \wedge l_i)$ with $P(\theta_2)$, where $m + 1 \leq i \leq m + n$. We let $\theta_{2,1} =$ "$X_{m+1} = a_{m+1}, \dots, X_{i-1} = a_{i-1}$". Let $\theta_{2,2} =$ "$X_i = a_i$". Let $\theta_{2,3} =$ "$X_{i+1} = a_{i+1}, \dots, X_{m+n} = a_{m+n}$". Then

$$P(\theta_2) = P(\theta_{2,1} \wedge \theta_{2,2} \wedge \theta_{2,3}) = P(\theta_{2,1})P(\theta_{2,2} | \theta_{2,1})P(\theta_{2,3} | \theta_{2,1} \wedge \theta_{2,2}).$$

On the other hand $P(\theta_2 \wedge l_i) = P(\theta_{2,1})P(l_i)P(\theta_{2,2} | \theta_{2,1} \wedge l_i)P(\theta_{2,3} | \theta_{22} \wedge \theta_{21})$. [*Note*: the last term does not have l_i in it because, conditioned on $\theta_{22} \sim \theta_{21}$, l_i is independent of $\theta_{2,3}$.]

We now have $P(l_i | \theta_2) = P(l_i \wedge \theta_2) \div P(\theta_2) = P(l_i)P(\theta_{22} | \theta_{21} \wedge l_i) \div P(\theta_{22} | \theta_{2,1}) \leq KP(l_i)$ for some fixed constant K, because the process we are inverting has a table which is bounded away from 0 and 1.

$$P(\theta_1 | \theta_2) \geq 1 - \sum_{i=n+1}^{n+m} P(l_i | \theta_2) \geq 1 - \sum_{i=m+1}^{m+n} KP(l_i)$$

which can be made arbitrarily close to 1 by choosing m sufficiently large.

PROOF OF (g). Obvious.

This concludes the proof that the inverse process is a U.M. We must still prove that it is a B.U.M. This means that there exists a number $\epsilon > 0$ such that for all n, and values $a_0, a_1, \dots, a_n$,

(h) $P(X_0 = a_0 | X_1 = a_1, \dots, X_n = a_n) > \epsilon$.

If (h) is true for large values of n, then it is true for all n so it suffices to prove for sufficiently large n.

In the proof of (f) we proved the existence of a fixed K such that $P(l_i | \theta_2) \leq KP(l_i)$. Using the exact same proof we can prove the existence of a K such that for any i, n,

(i) $P(l_i | X_1 = a_1, X_2 = a_2 \dots X_n = a_n)) \leq KP(l_i)$.

We recall that $\sum_{i=1}^{\infty} P(l_i) < \infty$ so there exists m such that

(j) *Definition of* m: $\sum_{i=m}^{\infty} KP(l_i) < \frac{1}{2}$.

(k) *Definition of* D: $D =$ "l_i fails for all $i \geq m$".

(i), (j) and (k) imply

(l) $P(D | X_1 = a_1, X_2 = a_2, \dots, X_n = a_n) > \frac{1}{2}$.

Select $n > m$. Note that "$X_{m+1} = a_{m+1}, X_{m+2} = a_{m+2}, \dots, X_n = a_n$" is independent of "$X_0 = a_0$" given "$D$ and $X_1 = a_1, X_2 = a_2, \dots, X_m = a_m$" so

(m) $P(X_0 = a_0 | X_1 = a_1, X_2 = a_2, \dots, X_n = a_n \wedge D) = P(X_0 = a_0 | X_1 = a_1, X_2 = a_2, \dots, X_m = a_m \wedge D)$.

D is independent of $X_0, X_1, \dots, X_m$ so

(n) $P(X_0 = a_0 | X_1 = a_1, X_2 = a_2, \dots, X_m = a_m \wedge D)$
$= P(X_0 = a_0 | X_1 = a_1, X_2 = a_2, \dots, X_m = a_m)$.

Because the forward process is bounded,

(o) There exists $\delta > 0$ such that for any $i, k, b_0, b_1, b_2, \dots, b_{k-1}$, $\delta^k \leq P(X_0 = b_0, X_1 = b_1, \dots, X_{k-1} = b_{k-1}) \leq (1-\delta)^k$.

By (l), (m), (n), and (o) we have

$$\begin{aligned}
&P(X_0 = a_0 | X_1 = a_1, X_2 = a_2, \dots, X_n = a_n) \\
&\geq P((X_0 = a_0) \wedge D | X_1 = a_1, X_2 = a_2, \dots, X_n = a_n) \\
&= P(X_0 = a_0 | X_1 = a_1, X_2 = a_2, \dots, X_n = a_n \wedge D) \\
&\quad \times P(D | X_1 = a_1, X_2 = a_2, \dots, X_n = a_n) \\
&\geq P(X_0 = a_0 | X_1 = a_1, X_2 = a_2, \dots, X_m = a_m) \tfrac{1}{2} \\
&= \tfrac{1}{2} P(X_0 = a_0, X_1 = a_1, \dots, X_m = a_m)/P(X_1 = a_1, \dots, X_m = a_m) \\
&\geq \tfrac{1}{2}\delta^m/(1-\delta)^m
\end{aligned}$$

so we have proved (h) for all $n > m$, with $\epsilon > \frac{1}{2}\delta^m/(1-\delta)^m$. □

EXAMPLE 19. One of the problems that must be overcome is to find some method of guaranteeing that a given U.M. has no R.M. representation with $E(N) < \infty$. Suppose such a representation does exist. Then, for that particular representation,

(a) $\sum_{i=1}^{\infty} \sum_{j=i}^{\infty} P(N = j) < \infty$.

Let $a_i = \sum_{j=i}^{\infty} P(N = j)$. Then (a) becomes

(b) $\sum_{i=1}^{\infty} a_i < \infty$.

a_i, of course, depends on the particular R.M. representation of the U.M. However, I will exhibit, for all i, a value $\hat{a}_i$, dependent only on the U.M. itself, and not on any particular R.M. representation of it, such that

(c) $\hat{a}_i \leq a_i$

no matter which R.M. representation is used to define a_i. Then, if we can establish that

(d) $\sum_{i=1}^{\infty} \hat{a}_i = \infty$,

(b) becomes impossible for any R.M. representation. We now define $\hat{a}_i$.

Let $\dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots$ be a stationary process. For specific values

$b_0, b_{-1}, b_{-2}, \ldots$, the assignment $X_0 = b_0, X_{-1} = b_{-1}, X_{-2} = b_{-2}, \ldots$ is called a *past*. For specific i, and specific values $b_1, b_2, \ldots, b_{i-1}$, the assignment $X_1 = b_1, X_2 = b_2, \ldots, X_{i-1} = b_{i-1}$ is called a *middle$_i$*.

We define $\hat{a}_i$ by

$$\hat{a}_i = \sup_{\substack{past_1 \\ past_2 \\ middle_i}} \left[P\left[X_i = 0 \,\middle|\, \begin{matrix} past_1 \\ and \\ middle_i \end{matrix} \right] - P\left[X_i = 0 \,\middle|\, \begin{matrix} past_2 \\ and \\ middle_i \end{matrix} \right] \right].$$

Suppose we have a realization of the process as an R.M. Then

$$\left[P\left[X_i = 0 \,\middle|\, \begin{matrix} past_1 \\ and \\ middle_i \end{matrix} \right] \right] = \sum_{j=1}^{\infty} P\left[X_i = 0 \wedge N_i = j \,\middle|\, \begin{matrix} past_1 \\ and \\ middle_i \end{matrix} \right].$$

Thus

$$P\left[X_i = 0 \,\middle|\, \begin{matrix} past_1 \\ and \\ middle_i \end{matrix} \right] - P\left[X_i = 0 \,\middle|\, \begin{matrix} past_2 \\ and \\ middle_i \end{matrix} \right]$$

$$= \sum_{j=1}^{\infty} \left[P\left[(X_i = 0) \wedge (N_i = j) \,\middle|\, \begin{matrix} past_1 \\ and \\ middle_i \end{matrix} \right] - P\left[(X_i = 0) \wedge (N_i = j) \,\middle|\, \begin{matrix} past_2 \\ and \\ middle_i \end{matrix} \right] \right]$$

$$\leq \sum_{j=1}^{i-1} (0) + \sum_{j=\infty}^{\infty} (N_i = j) = a_i$$

thus establishing (c).

We now conclude by exhibiting a random Markov process with all its table values in the interval $[\frac{1}{15}, \frac{3}{5}]$ such that $E(N) < \infty$ and the $\hat{a}_i$, computed for the inverse process, satisfies $\hat{a}_i = (1/i)$ [i.e. there exists $h > 0$ such that $\hat{a}_i > h/i$ for all sufficiently large i].

The process has three states 0, 1, and n (for neutral). The table assigns each past word three values that sum to one, i.e. $T(\text{word}) = (P(X_0 = 0), P(X_0 = 1), P(X_0 = n))$.

We now describe the table. We fix a word of the past. Call it "word". In the word we seek out any two successive letter subsequence of the form $0, n$ or $1, n$. If no such two letter subsequence exists we let $T(\text{word}) = (P(X_0 = 0) = P(X_0 = 1) = P(X_0 = n) = \frac{1}{3})$. Otherwise, choose the last such two-letter subsequence of "word" ("last" means largest j such that $X_j X_{j+1}$ is such a two-letter subsequence). If it is $0, n$ let $T(\text{word}) = (P(X_0 = 0) = \frac{3}{5}, P(X_0 = 1) = \frac{1}{15}, P(X_0 = n) = \frac{1}{3})$. If it is $1, n$ let $T(\text{word}) = (P(X_0 = 0) = \frac{1}{15}, P(X_0 = 1) = \frac{3}{5}, P(X_0 = n) = \frac{1}{3})$.

We define the distribution of lookback time, N, by $P(N = i) = \gamma/i^3$, where $\gamma = \sum_{i=1}^{\infty} (1/i^3)$. Clearly $E(N) < \infty$.

By Theorem 11 there exists a stationary measure with this table and lookback time endow $\{X_i\}$ with any such stationary distribution. Fix i_0 and let $Y_i = X_{i_0-i}$. Then $\{Y_i\}_{i=-\infty}^{\infty}$ has the distribution of the inverse process of $\{X_i\}_{i=0}^{\infty}$. From here on $\hat{a}_{i_0}$ means the $\hat{a}_{i_0}$ calculated from the process $\{Y_i\}$.

We wish to show that $\hat{a}_{i_0} \geq 0(1/i_0)$. In the Y_i process let

$$\text{“middle”} = \text{“}Y_1 = Y_2 = \dots Y_{i_0-1} = n\text{”} = \text{“}X_1 = X_2 \dots X_{i_0-1} = n\text{”},$$

$$\text{“past}_0\text{”} = \text{“}Y_0 = Y_{-1}, \dots, Y_{-m} = 0\text{”} = \text{“}X_{i_0} = X_{i_0+1} = \dots = X_{i_0+m} = 0\text{”}, \quad m \text{ chosen huge},$$

$$\text{“past}_1\text{”} = \text{“}Y_0 = Y_{-1} = \dots Y_{-m} = 1\text{”} = \text{“}X_{i_0} = X_{i_0+1} = \dots X_{i_0+m} = 1\text{”}.$$

Let middle $= B$, past$_0 = C$, past$_1 = D$, “$Y_{i_0} = 0$” = “$X_0 = 0$” $= A$. From the definition of $\hat{a}_{i_0}$,

(e) $\hat{a}_{i_0} \geq P(A \mid B \wedge C) - P(A \mid B \wedge D)$.

From here on ignore the Y_i process and consider only the X_i process.

$$\text{(f)}\ \frac{\hat{a}_{i_0}}{P(A)} \geq \frac{P(A \mid B \wedge C)}{P(A)} - \frac{P(A \mid B \wedge D)}{P(A)} = \frac{P(B \wedge C \mid A)}{P(B \wedge C)} - \frac{P(B \wedge D \mid A)}{P(B \wedge D)}$$
$$= \frac{P(B \mid A) P(C \mid B \wedge A)}{P(B) P(C \mid B)} - \frac{P(B \mid A) P(D \mid B \wedge A)}{P(B) P(D \mid B)}$$
$$= \frac{P(C \mid B \wedge A)}{P(C \mid B)} - \frac{P(D \mid B \wedge A)}{P(D \mid B)}.$$

The last equality holds because B is independent of A. This is true because the middle is the all neutral state and neutral has probability $\frac{1}{3}$ no matter what the past.

We will show that there are h_1, h_2, h_3, h_4 all greater than zero, such that

(g) $h_1 > h_2$,

(h) $h_3 > h_4$,

(i) $P(C \mid B \wedge A) > (\frac{1}{3})^{m+1}(1 + h_1/i_0)$,

(j) $P(C \mid B) < \frac{1}{3}^{m+1}(1 + h_2/i_0)$,

(k) $P(D \mid B \wedge A) < \frac{1}{3}^{m+1}(1 - h_3/i_0)$,

(l) $P(D \mid D) > \frac{1}{3}^{m+1}(1 - h_4/i_0)$.

If we can establish (g), (h), (i), (j), (k) and (l), then by (f) we have, for sufficiently large i_0,

$$\frac{\hat{a}_{i_0}}{P(A)} \geq \frac{1+h_1/i_0}{1+h_2/i_0} - \frac{1-h_3/i_0}{1-h_4/i_0}$$

$$= \left[1 + \frac{(h_1-h_2)/i_0}{1+h_2/i_0}\right] - \left[1 - \frac{(h_3-h_4)/i_0}{1-h_4/i_0}\right]$$

$$> \frac{h_1-h_2}{2i_0} + \frac{h_3-h_4}{i_0}.$$

Therefore

$$\hat{a}_{i_0} \geq \left(\frac{h_1-h_2}{2} + h_3 - h_4\right)P(A)/i_0 \geq \left(\frac{h_1-h_2}{2} + h_3 - h_4\right)\frac{1}{15}\Big/ i_0$$

and we will be done. The last inequality holds because $P(A \mid N \wedge X_{-1} \wedge X_{-2} \cdots)$ is bounded below by $\frac{1}{15}$.

We now conclude this paper by defining h_1, h_2, h_3, h_4 and establishing (g), (h), (i), (j), (k), (l). First we bound $P(C \mid B \wedge A)$ and $P(D \mid B \wedge A)$.

We now expand this using the definition of this random Markov process.

$$P(C \mid B \wedge A) = \prod_{i=i_0}^{i_0+m} \sum_{j=0}^{\infty} P((N_i = j) \wedge X_i = 0 \mid X_{i-1} = X_{i-2} = \ldots = X_{i_0} = 0 \wedge B \wedge A)$$

$$= \prod_{i=i_0}^{i_0+m} \sum_{j=0}^{\infty} P(N_1 = j)\left(\left\{\frac{1}{3} \text{ if } j < i,\ \frac{3}{5} \text{ if } j \geq i\right)\right.$$

$$= \sum_{i=i_0}^{i_0+m}\left(\frac{1}{3}\left(1 - \sum_{j\geq i}\frac{\gamma}{j^3}\right) + \frac{3}{5}\left(\sum_{j\geq i}\frac{\gamma}{j^3}\right)\right)$$

$$= \frac{1}{3}^{1+m}\left(\left(\prod_{i=i_0}^{i_0+m}\left(1 - \frac{\gamma}{2i^2} + \frac{9}{5}\left(\frac{\gamma}{2i^2}\right)\right)\right) + \text{error}_1\right)$$

$$= \frac{1}{3}^{1+m}\left(\left(\prod_{i=i_0}^{i_0+m}\left(1 + \frac{2}{5}\frac{\gamma}{i^2}\right)\right) + \text{error}_1\right)$$

$$= \frac{1}{3}^{m+1}\left(1 + \frac{2}{5}\frac{\gamma}{i_0} + \text{error}_1 + \text{error}_2\right)$$

where error_1 is of order $1/i_0^3$ and error_2 is of order $1/i_0^2$ given that m is sufficiently large. Thus, for any small number, say 10^{-6}, for sufficiently large i_0, and then m chosen large after i_0 is chosen,

$$\frac{1}{3}^{m+1}\left(1 + \frac{2\gamma/5 - 10^{-6}}{i_0}\right) < P(C \mid B \wedge A) < \frac{1}{3}^{m+1}\left(1 + \frac{2\gamma/5 + 10^{-6}}{i_0}\right).$$

We now bound $P(D|B \wedge A)$ using exactly the same reasoning we used to bound $P(C|B \wedge A)$.

$$P(D|B \wedge A) = \prod_{i=i_0}^{i_0+m} \left(\frac{1}{3}\left(1 - \sum_{j \geq i} \frac{\gamma}{j^3}\right) + \frac{1}{15}\left(\sum_{j \geq i} \frac{\gamma}{j^3}\right)\right)$$

$$= \frac{1}{3}^{m+1}\left(1 - \frac{2}{5}\frac{\gamma}{i_0} + \text{error}_1 + \text{error}_2\right),$$

$$(*) \quad \frac{1}{3}^{m+1}\left(1 - \frac{(2\gamma/5 + 10^{-6})}{i_0}\right) < P(D|B \wedge A) < \frac{1}{3}^{m+1}\left(1 - \frac{(2/5\gamma - 10^{-6})}{i_0}\right).$$

We now have to bound $P(C|B)$ and $P(D|B)$.

$$P(C|B) = (P(C|B \wedge (X_0 = 0)))P(X_0 = 0)$$
$$+ P(C|B \wedge (X_0 = n))P(X_0 = n) + P(C|B \wedge (X_0 = 1))P(X_0 = 1).$$

Clearly, the way this random Markov process is defined implies

$$P(C|B \wedge (X_0 = 1)) \leq P(C|B \wedge (X_0 = n)) \leq P(C|B \wedge (X_0 = 0)).$$

Recall that "$X_0 = 0$" $= A$. We have

$$P(C|B) \leq P(C|B \wedge A)(1 - P(X_0 = 1)) + (P(C|B \wedge (X_0 = 1)))P(X_0 = 1)$$
$$\leq \frac{14}{15}\left(P(C|B \wedge A) + \frac{1}{15} P(C|B \wedge (X_0 = 1)\right).$$

The last inequality holds because $P(X_0 = 1 | N_0 \wedge X_{-1}, X_{-2}, \ldots)$ is bounded below by $\frac{1}{15}$ so $P(X_0 = 1) \geq \frac{1}{15}$. Note that $P(C|B \wedge (X_0 = 1))$ computed term by term is precisely the same thing as $P(D|B \wedge A)$. Hence we have

$$P(C|B) \leq \frac{14}{15} P(C|B \wedge A) + \frac{1}{15} P(D|B \wedge A)$$

$$< \frac{1}{3^{(m+1)}}\left(\frac{14}{15}\left(1 + \left(\frac{2}{5}\gamma + 10^{-6}\right)\Big/ i_0\right) + \frac{1}{15}\left(1 - \left(\frac{2}{5}\gamma - 10^{-6}\right)\Big/ i_0\right)\right)$$

$$= \frac{1}{3^{m+1}}\left(1 + \left(\frac{26}{75}\gamma + 10^{-6}\right)\Big/ i_0\right),$$

$$(*) \qquad P(C|B) < \frac{1}{3^{m+1}}\left(1 + \left(\frac{26}{75}\gamma - 10^{-6}\right)\Big/ i_0\right).$$

We use the same reasoning to bound $P(D|B)$ that we used to bound $P(C|B)$.

$$
\begin{aligned}
P(D|B) &\geq \frac{14}{15} P(D|B \wedge A) + \frac{1}{15} P(D|B \wedge \text{“}X_0 = 1\text{”}) \\
&= \frac{14}{15} P(D|B \wedge A) + \frac{1}{15} P(C|B \wedge A) \\
&> \frac{1}{3^{m+1}} \Bigg(\frac{14}{15} \Bigg(1 - \Bigg(\frac{2}{5}\gamma + 10^{-6} \Bigg) \Big/ i_0 \Bigg) + \frac{1}{15} \Bigg(1 + \Bigg(\frac{2}{5} - 10^{-6} \Bigg) \Big/ i_0 \\
&= \frac{1}{3^{m+1}} \Bigg(1 - \Bigg(\Bigg(\frac{26}{75}\gamma + 10^{-6} \Bigg) \Big/ i_0,
\end{aligned}
$$

$$
(*) \qquad P(D|B) > \frac{1}{3^{m+1}} \left(1 - \left(\frac{26}{75}\gamma + 10^{-6} \right) \Big/ i_0 \right).
$$

We are done. Let $h_1 = \frac{2}{5}\gamma - 10^{-6}, h_2 = \frac{26}{75}\gamma + 10^{-6}, h_3 = \frac{2}{5}\gamma - 10^{-6}, h_4 = \frac{26}{75}\gamma + 10^{-6}$ and (g), (h), (i), (j), (k) and (l) all hold. This can easily be seen using the asterisk equations. □

Acknowledgement

Prior to my coming to the Hebrew University, I established the subject of this paper and proved some theorems. However, during the entire year I spent in the Hebrew University, work on this paper has been done jointly with Benjamin Weiss. Consequently, Weiss is responsible for a significant fraction of this paper.

References

1. H. Berbee, *Chains with infinite connections: Uniqueness and Markov representation*, Probability Theory and Related Fields **76** (1987), 243–253.

2. W. Doeblin and R. Fortet, *Sur les chains a liaisons completes*, Bull. Soc. Math. Fr. **65** (1937), 132–148.

3. S. Kalikow, *T, T^{-1} transformation is not loosely Bernoulli*, Ann. of Math. **115** (1982), 393–409.

4. M. Keane, *Strongly mixing g-measures*, Invent. Math. **16** (1972), 309–324.

5. B. Petit, *Schemas de Bernoulli et g-measures*, C.R. Acad. Sci. Paris, Ser. A **280** (1975), 17–20.

6. [Weak Bernoulli] P. Shields, *The Theory of Bernoulli Shifts*, University of Chicago Press, 1973, p. 89.

7. [Entropy] P. Shields, *The Theory of Bernoulli Shifts*, University of Chicago Press, 1973, p. 48 and p. 56.

8. [K] P. Shields, *The Theory of Bernoulli Shifts*, University of Chicago Press, 1973, p. 103.

9. [Bernoulli] P. Shields, *The Theory of Bernoulli Shifts*, University of Chicago Press, 1973, p. 110.