Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

S. Kalikow

Random Markov processes and uniform martingales

Department of Operations Research, Statistics, and System Theory Report BS-R8908 May

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Research (N.W.O.).

Random Markov Processes and Uniform Martingales

Steven Kalikow
Delft University of Technology, Faculty of Mathematics and Informatics,
Julianalaan 132, 2628 BL Delft, The Netherlands

1. **Abstract:** It is shown here that a certain generalization of an n-step Markov Chain is equivalent to the uniform convergence of the martingale

$\{P(X_0|X_{-1}X_{-2}...X_{-n})\}_{n=1}^{\infty}$.

Ergodic and probablistic properties of this process are explored.

AMS 1980 Subject Classification: 60J10, 60G48, 60G10, 28D05.
Keywords: random Markov process, uniform martingale, weak Bernoulli.
Note: the author visited the CWI some months at the end of 1988.

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

3. Statement of results with discussion

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 $N_i \in \mathbb{N}, N_0$ is independent of $\{a_i, N_i\}_{i<0}$ and for each j

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

Then $\{a_i, N_i\}_{i\in\mathbb{Z}}$ 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 $\{a_i\}_{i\in\mathbb{Z}}$ be a stationary process, all $a_i \in F$. If, for all $\epsilon > 0$, there exists $N_\epsilon \in \mathbb{N}$ such that for all $M > N_\epsilon$ and all $\{F_i\}_{i=0}^{\infty}$ with all $F_i \in F$,

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

then a_i is a U.M. ///

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

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

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 process which is not meerly a finite state rotation cannot be an U.M. ///

Report BS-R8908
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

Comment: In a future paper between Y. Katznelson, B. Weiss and me, we show that every zero entropy process can be extended to a R.M. (I am not sure our result is that strong. We may only be able to extend to a C.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 18 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 $P(X_0 = s/past)$ bounded below for some state s, then it is Weak Bernoulli. ///

Corollary 8: Let X_n be an 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 ///

Comment: I so not know how to prove the corrolary without first proving the theorem. This demonstrates the value of the R.M. representation. ///

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 cannonical R.M. Factor. We also say that this cannonical R.M. Factor *supports* the R.M. representation. ///

Comment: The R.M. representation, together with the cannonical 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 *existance*. If there does 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 existance. ///

Examples 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 $P(N_0 = a_i) = \frac{1}{2^i}$ and let the table be defined as follows:

$$P(X_0 = 1|X_{-1}, X_{-2} \ldots X_{-a_n} \wedge N_0 = a_n) = \begin{cases} .9 & \text{if a majority of } X_{-1} \ldots X_{-a_n} \text{ are 1} \\ .1 & \text{else} \end{cases}$$

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. ///

Theorem 13: Any R.M. representation supporting a K transformation satisfies uniqueness. ///

Corollary 14: A R.M. representation satisfying $E(N_0) < \infty$ and the same minimal extra condition as theorem 7 satisfies uniqueness. ///

Example 15: There exists a R.M. that is K and not Bernoulli. ///

Comment: This rises hopes that maybe all K transformations can be extended to a R.M. which is K. ///

Definition 16: B.U.M. An 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 17: 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 18: The inverse of a R.M. with $E(N_0) < \infty$ need not be an U.M. ///

Comment: Example 18 gives another example for example 6 because $E(N_0) < \infty$ with minimal extra condition (which is satisfies here) implies weak Bernoulli, and weak Bernoulli is closed under inverse. ///

Example 19: The inverse of a B.U.M. need not be an U.M. ///

Theorem 20: 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 21: 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 $E(N_0^n) < \infty$.

4. Proofs of theorems and Examples demonstrated

Proof of theorem 4:

We wish to prove U.M.=R.M. Obviously R.M. $\to$ U.M. where N_ϵ is chosen so that $P(N_0 > N_\epsilon) < \epsilon$. We now show U.M. $\to$ 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: $P(X_0 = 1|\text{past})$ 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. If we can construct the R.M. so that for every past

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

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

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 insure that

b) $\sup_b P(X_0 = 1|X_{-1} = a_{-1}, \ldots X_{-N_i} = a_{-\hat{N}_i}, X_{-(\hat{N}_i+1)} = b_1, X_{-(\hat{N}_i+2)} = b_2 \ldots)$
$= \hat{P}(X_0 = 1|X_{-1} = a_{-1}, X_{-2} = a_{-2} \ldots X_{-\hat{N}_i} = a_{-\hat{N}_i}, N_0 \leq \hat{N}_{i-1})$.
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) $\sup_b P(X_0 = 1|X_{-1} = a_{-1} \ldots X_{-\hat{N}_{i-1}} = a_{-\hat{N}_{i-1}}, X_{-(N_{i-1}+1)} = b_1$
$X_{-(\hat{N}_{i-1}+2)} = b_2 \ldots) = \hat{P}(X_0 = 1|X_{-1} = a_{-1} \ldots X_{-\hat{N}_{i-1}} = a_{-\hat{N}_{i-1}}, N_0 \leq \hat{N}_{i-1})$
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 $0 \leq T \leq 1$.
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 side of (c) respectively. It is clear that

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

this expresses r_1 as a wheighted average of r_2 and T which we denote by $w(r_2, T) = r_1$. Since we are assuming $\ell_1 = r_1$ and $\ell_2 = r_2$ we have

d) $w(\ell_2, T) = \ell_1$
This equation solves for T. All that is necessary is to show $0 \leq T \leq 1$. Clearly $\ell_2 \geq \ell_1$ so it follows that $T \leq \ell_1 \leq 1$.
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. □

<u>Case 2</u> 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 $a_{-1} \ldots a_{-\hat{N}_i}$, if

$P(X_0 = 1|X_{-1} = a_{-1} \dots X_{-\hat{N}_0} = a_{-\hat{N}_0}) \geq \frac{1}{2}$ 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
$\hat{P}(X_0 = 1|X_{-1} = a_{-1} \dots X_{-\hat{N}_i} = a_{-\hat{N}_i}, N_0 \leq \hat{N}_i)$

$$= \begin{cases} \sup_{b_1,b_2,\dots} & P(X_0 = 1|X_{-1} = a_{-1}, \dots X_{-\hat{N}_1} = a_{-\hat{N}_i}, X_{-(\hat{N}_i+1)} = b_1 \dots) \\ & \text{if } P(X_0 = 1|X_{-1} = a_{-1} \dots X_{-\hat{N}_0} = a_{-\hat{N}_0}) \geq \frac{1}{2} \\ \inf_{b_1,b_2,\dots} & PX_0 = 1|X_{-1} = a_{-1}, \dots X_{-\hat{N}_i} = a_{-\hat{N}_i}, X_{-(\hat{N}_i+1)} = b_1 \dots \\ & \text{if } P(X_0 = 1|X_{-1} = a_{-1} \dots X_{-\hat{N}_0} = a_{-\hat{N}_0}) < \frac{1}{2} \end{cases}$$

we are essentially in case $\overline{1}$. □

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

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

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 $(n \geq 1, n \in \mathbb{N})$ 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 cylendar set probabilities, and for this partially crossed out process, it is clear what the cylendar 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 identicle.

However, it is not an U.M. because $P(X_0 = 1|\text{past } 5(10^n) \text{ terms all one})$ is close to $\frac{1}{2}$, for large n, but $P(X_0 = 1|\text{past}10^{n+1} \text{ terms all one and } 10^{n+1} + 1 \text{ term } 0)$=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 strech (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 $P(N_0 = N_i) = \frac{1}{2^i}$. The $N_i^!s$ have to increase rapidly enough so that ϵ_{N_i} is small in comparison to $\frac{1}{2^i}$. 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 corrolary. □

Proof of theorem 11:

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(time 0|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 $\{a_i\}_{i\in\mathbb{Z}}$ where $\{a_i\}_{i<0}$ is determined.

Define a measure u_n on words of length n by $u_n(w) = \dfrac{1}{2^{2^n}}\#\{i : 0 \leq i < 2^{2^n}$ 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) $u(X_0 = 1|\text{past})$ is the value $P(X_0 = 1|\text{ past})$ given by the R.M.
We show (a) by showing

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

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

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

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

⌣ 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 $P(X_0 = 1|X_{-1} = b_1, \dots X_{-(n-1)} = b_{n-1}, N_0 = k)$. 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 $X_i = 0$ always, and all $X_i = 1$ always. $N_0 = 0$ always, and $X_0 = \{0,1\}$ with $\frac{1}{2}, \frac{1}{2}$ product measure always, admits only one measure.

Proof of Theorem 13

Let P be a R.M. representation. Then P provides sufficient information to compute the probability law of X_0 given the past. Continuing this reasoning, we can compute the probability law of $X_n, X_{n+1}, \dots X_{n+k}$ given the past for every $n > 0$ and $k > 0$. Suppose P supports a given K-process $\hat{P}$. Then fixing k and letting n approach ∞, the probability law of $X_n, X_{n+1} \dots X_{n+k}$ given the past approaches the fixed Probability law $\hat{P}$ on $X_0, X_1 \dots X_k$. This forces $\hat{P}$ to be the only probability law supported by P.

Proof of corollary 14

Theorem 7 implies weak Bernoulli which in turn implies K.

Example 15

The T, T^{-1} transformation is a transformation which is K and not Bernoulli. 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} transforamtion is defined as follows. We will describe the 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 $\{a_i\}_{i \in \mathbb{Z}}$, 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 $\{b_i\}_{i \in \mathbb{Z}}$, 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 $\{c_i\}_{i \in \mathbb{Z}}$ define by

$$c_i = \left\{ \begin{array}{c} a_j \\ b_i \end{array} \right\} \quad \text{where } j = \left\{ \begin{array}{ll} 0 \ \text{ if } i \ \text{ is } 0 & \\ \hline \#\{k : 0 \le k < i \wedge b_k = R\} - & \\ \#\{k : 0 \le k < i \wedge b_k = L\} & \text{if } i > 0 \\ \hline \#\{k : i \le k < 0 \wedge b_k = L\} - & \\ \#\{k : i \le k < 0 \wedge b_k = R\} & \text{if } i < 0 \end{array} \right.$$

The process $\{c_i\}_{i \in \mathbb{Z}}$ 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 origen shifts to the left). Every time you see an R, the entire scenery shifts to the left (i.e. the origen shifts to the right). a_j is just the 0 coordinate of the shifted scenery.

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

Definition: Let $\binom{a_i}{b_i}_{i\in\mathbb{Z}}$ 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 $\binom{a_i}{b_i}_{i\in\mathbb{Z}}$ 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

- proof left to reader

Definition: A T, T^{-1}-question mark path is a doubly infinite sequence $\binom{a_i}{b_i}$ from the 6 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 $a_{i_1} = H$ and $a_{i_2} = T$

-proof left to reader

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

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

Definition "1 order process" start with the 0 order process. Redefine any $(\binom{a_i}{b_i}, N_i)$ to be $(\binom{?}{b_i}, N_i)$ 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 $(\binom{a_i}{b_i}, N_i)$ to be $(\binom{?}{b_i}, N_i)$ 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 $N_i \quad i \in \mathbb{Z}$.

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 $i, -N_0 \leq i < 0$ such that $a_i \neq$ "?" 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 cannonical 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 $i \in \mathbb{Z}, N_i = \hat{N}_k$, 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 $\{k_j\}_{j \in \mathbb{N}}$ 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.

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

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.

Corrolary: 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 $i \in \mathbb{Z}$. I will prove you can recover a_i by proving $P(\text{you cannot recover } a_i) < \epsilon$. Choose k so that $\frac{1}{2^k} < \epsilon$. Since random walk is recurrent there are infinitely many $j < i$ 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 corrolary.

All that is left to do is to show that the final factor is an extension of the T, T^{-1}. Definition A Decendent of i is 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 17:

We define the table values to inductively force the following equation to hold for all k.

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

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

Example 18: We define a R.M. representation. The alphabet is $\{0, 1\}$. Let $P(N_0 = n + 20) = \frac{1}{2^n}$. 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 $i \le N_0 - 2$ where (a) or (b) holds. If $X_{-(i+2)} = 1$ then let $X_0 = 1$ with probability .9 and if $X_{-(i+2)} = 0$ then let $X_0 = 0$ with probability .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 cannonical R.M. factor of the C.R.M. Our purpose is to show that $\hat{P}$ is not an U.M.

We do this by proving that A and B differ substancially where $A = P(X_0 = 0|X_1 = 0, X_2 = X_3 = \dots X_{1+m} = 1, X_{m+2} = \dots X_{1+m+n} = 0)$, and $B = P(X_0 = 0|X_1 = 0, X_2 = X_3 = \dots X_{1+m} = 1)$, 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 $n \to \infty$.

<u>Proof that B does not depend on m and $B \neq 1$.</u>

Let "$X_0 = 0$" $= C$, "$X_1 = 0, X_2 = X_3 = \dots X_{1+m} = 1$" $= D$. Looking at our R.M. table we see that $P(X_3 = X_4 = \dots X_{1+m} = 1|X_0 = 0, X_1 = 0, X_2 = 1) = \frac{1}{2}^{m-2}$.
$P(X_3 = X_4 = \dots X_{1+m} = 1|X_1 = 0, X_2 = 1) = \frac{1}{2}^{m-2}$.
Therefore $B = P(C|D) = P(C \wedge D)/P(D) = P(X_0 = 0, X_1 = 0, X_2 = 1)\frac{1}{2}^{m-2} \div P(X_1 = 0, X_2 = 1)\frac{1}{2}^{m-2} = P(X_0 = 0, X_1 = 0, X_2 = 1) \div P(X_1 = 0, X_2 = 1)$ 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 $N_0, P(X_0 = 1) \ge \frac{1}{10}$. 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 $N_2, P(X_2 = 1) \ge \frac{1}{10}$ so $P(X_0 = 1, X_1 = 0, X_2 = 1) \ge (\frac{1}{10})(\frac{1}{2})(\frac{1}{10}) = \frac{1}{200} > 0$.

<u>Proof that A approaches 1 as n approaches ∞</u>

Let "$X_0 = 0$" $= C$, "$X_0 = 1$" $= F$ and, as above, $X_1 = 0, X_2 = X_3 = \dots = X_{1+m} = 1$" $= D$ and let "$X_1 = 0, X_2 = \dots = X_{1+m} = 1, X_{m+2} = \dots = X_{1+m+n} = 0$" $= E$. $A = P(C|E) = P(C \cap E)/P(E) = P(C \cap E)/(P(C \cap E) + P(F \cap E))$. 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 $E_i =$ "$X_1 = 0, X_2 = X_3 = \dots X_{m+1} = 1, X_{m+2} = X_{m+3} = \dots X_i = 0$".

We will now compare $P(E_i \cap C) \div P(E_{i-1} \cap C)$ with $P(E_i \cap F) \div P(E_{i-1} \cap F)$.
$P(E_i \cap C) \div P(E_{i-1} \cap C) = P(X_i = 0|E_{i-1} \cap C)$.

We can compute this using the R.M. rule. If $N_i < i$, and E_{i-1}, then X_i must be 1. If $N_i \ge i$, and $E_{i-1} \cap C$, then $X_i = 0$ with probability .9. Thus $P(E_i \cap C) \div P(E_{i-1} \cap C) = P(X_i = 0|E_{i-1} \cap C) = .9\, P(N_i \ge i)$.

Similarly, $P(E_i \cap F) \div P(E_{i-1} \cap F) = .1(P(N_i \ge i))$. 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 19:

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 $P(N_0 = 2^n) = \frac{1}{2^n}$, but really all we use is that $E(N_0) = \infty$. The table is identicle to that of the table of the previous example except that when the previous table says $P(X_0 = 1|N_0$ and past)=1, we instead have $P(X_0 = 1|N_0$ 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 previous example). The result follows by argument of the previous example.

Proof of theorem 20:

We start out by making some general arguments about sets on a probability space. Let $\theta_1, \theta_2, \theta_3$, 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 an 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 $\{P(\ell_i)\}_{i=1}^{\infty}$ is summable. Let θ_1 be the event that ℓ_i fails for all $i > m$. Let θ_2 be the event "$X_{m+1} = a_{m+1}, X_{m+2} = a_{m+2} \ldots X_{m+n} = a_{m+n}$". 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 an 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 $\sum_{m+1}^{m+n} P(\ell_i)$ 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 $P(\theta_2 \wedge \ell_i)$ with $P(\theta_2)$, where $m+1 \leq i \leq m+n$. We let $\theta_{2,1}$ = "$X_{m+1} = a_{m+1}, \ldots X_{i-1} = a_{i-1}$". Let $\theta_{2,2}$ = "$X_i = a_i''$. Let $\theta_{2,3}$ = "$X_{i+1} = a_{i+1}, \ldots X_{m+n} = a_{m+n}$". 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 $P(\theta_2 \wedge \ell_i) = P(\theta_{2,1})P(\ell_i)P(\theta_{2,2}|\theta_{2,1} \wedge \ell_i)P(\theta_{2,3}|\theta_{22} \wedge \theta_{21})$ [*Note* the last term does not have ℓ_i in it because conditioned on $\theta_{22} \wedge \theta_{21}$, ℓ_i is independent of $\theta_{2,3}$].

We now have $P(\ell_i|\theta_2) = P(\ell_i \wedge \theta_2) \div P(\theta_2) = P(\ell_i)P(\theta_{22}|\theta_{21} \wedge \ell_i) \div P(\theta_{22}|\theta_{2,1}) \leq KP(\ell_i)$ for some fixed constant K, because the process we are inverting has a table which is bounded away from 0 and 1. $P(\theta_1|\theta_2) \geq 1 - \sum_{i=n+1}^{n+m} P(\ell_i|\theta_2) \geq 1 - \sum_{i=m+1}^{m+n} KP(\ell_i)$ 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 an U.M. We still must prove that it is a B.U.M. This means that there exists a number $\epsilon > 0$ such that for all n, and values $a_0, a_1, \ldots a_n$,

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

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 existance of a fixed K such that $P(\ell_i|\theta_2) \leq KP(\ell_i)$. Using the exact same proof we can prove the existance of a K such that for any i, n, i) $P(\ell_i|X_1 = a_1, X_2 = a_2 \ldots X_n = a_n)) \leq KP(\ell_i)$.

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

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

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

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

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

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

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

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

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

Because the forward process is bounded,

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

By (l), (m), (n), and (o) we have $P(X_0 = a_0|X_1 = a_1, X_2 = a_2 \ldots X_n = a_n) \geq P((X_0 = a_0) \wedge D|X_1 = a_1, X_2 = a_2 \ldots X_n = a_n) =$

$P(X_0 = a_0 | X_1 = a_1, X_2 = a_2 \ldots X_n = a_n \wedge D) P(D | X_1 = a_1, X_2 = a_2 \ldots X_n = a_n) \geq P(X_0 = a_0 | X_1 = a_1, X_2 = a_2, \ldots X_m = a_m) \frac{1}{2} = \frac{1}{2} P(X_0 = a_0, X_1 = a_1, \ldots X_m = a_m) / P(X_1 = a_1 \ldots X_m = a_m)$
$\geq \frac{1}{2} \delta^m / (1-\delta)^m$ so we have proved (h) for all $n > m$, with $\epsilon > \frac{1}{2} \frac{\delta^m}{(1-\delta)^m}$. □

Example 21

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 vthe U.M. However, I will exhibit, for all i, a value $\hat{a}_i$, dependent only on the U.M. itself, and not an any particular R.M. representation of it, such that
c) $\hat{a}_i \leq a_i$
no matter what 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 $\ldots X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots$ 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 asignment $X_1 = b_1, X_2 = b_2 \ldots X_{i-1} = b_{i-1}$ 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

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

Thus

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

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

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

thus establishing (c).

We now conclude by exhibitting 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 $\hat{a}_i = 0(\frac{1}{i})$ [i.e. there exists $h > 0$ such that $\hat{a}_i > \frac{h}{i}$ 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 seak 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_jX_{j+1} is such a 2 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 $T(\text{word}) = (P(X_0 = 0) = \frac{1}{15}, P(X_0 = 1) = \frac{3}{15}, P(X_0 = n) = \frac{1}{3})$.

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

We will denote this process by $\{X_i\}_{i=-\infty}^{\infty}$. 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 $\hat{a}_{i_0} \geq 0(\frac{1}{i_0})$. In the Y_i process
Let "middle" $=$ "$Y_1 = Y_2 = \ldots Y_{i_0-1} = n$" $=$ "$X_1 = X_2 \ldots X_{i_0-1} = n$"
"past$_0$" $=$ "$Y_0 = Y_{-1} \ldots Y_{-m} = 0$" $=$ "$X_{i_0} = X_{i_0+1} \cdots = X_{i_0+m} = 0$",
m chosen huge.
"past$_1$" $=$ "$Y_0 = Y_{-1} \ldots Y_{-m} = 1$" $=$ "$X_{i_0} = X_{i_0+1} \ldots X_{i_0+m} = 1$"

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) $\hat{a}_{i_0} \geq P(A|B \wedge C) - P(A|B \wedge D)$
From here on ignor the Y_i process and consider only the X_i process.

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

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

We will show that there is a 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) $P(C|B \wedge A) > (\frac{1}{3})^{m+1}(1 + \frac{h_1}{i_0})$
j) $P(C|B) < \frac{1}{3}^{m+1}(1 + \frac{h_2}{i_0})$
k) $P(D|B \wedge A) < \frac{1}{3}^{m+1}(1 - \frac{h_3}{i_0})$
l) $P(D|B) > \frac{1}{3}^{m+1}(1 - \frac{h_4}{i_0})$

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

$$\frac{\hat{a}_{i_0}}{P(A)} \geq \frac{1+\frac{h_1}{i_0}}{1+\frac{h_2}{i_0}} - \frac{1-\frac{h_3}{i_0}}{1-\frac{h_4}{i_0}} = [1+\frac{(h_1-h_2)/i_0}{1+\frac{h_2}{i_0}}] - [1-\frac{(h_3-h_4)/i_0}{1-\frac{h_4}{i_0}}] > \frac{h_1-h_2}{2i_0} + \frac{h_3-h_4}{i_0}$$

. Therefore

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

and we will be done. The last inequality holds because $P(A|N \wedge X_{-1} \wedge X_{-2} \ldots)$ 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 $P(C|B \wedge A)$ and $P(D|B \wedge A)$

We now expand this using the definition of this Random Markov Chain.

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

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

$$\prod_{i=i_0}^{i_0+m} (\frac{1}{3}(1 - \sum_{j \geq i} \frac{\gamma}{j^3}) + \frac{3}{5}(\sum_{j \geq i} \frac{\gamma}{j^3})) = \frac{1}{3}^{1+m}((\prod_{i=i_0}^{i_0+m}(1 - \frac{\gamma}{2i^2} + \frac{9}{5}(\frac{\gamma}{2i^2}))) + error_1) =$$

$$\frac{1}{3}^{1+m}((\prod_{i=i_0}^{i_0+m}(1 + \frac{2}{5}\frac{\gamma}{i^2})) + error_1) = \frac{1}{3}^{m+1}(1 + \frac{2}{5}\frac{\gamma}{i_0} + error_1 + error_2)$$

where $error_1$ is of order $\frac{1}{i_0^3}$ and $error_2$ is of order $\frac{1}{i_0^2}$ 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

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

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

$$P(D|B \wedge A) = \prod_{i=i_0}^{i_0+m} (\frac{1}{3}(1 - \sum_{j \geq i} \frac{\gamma}{j^3}) + \frac{1}{15}(\sum_{j \geq i} \frac{\gamma}{j^3})) = \frac{1}{3}^{m+1}(1 - \frac{2}{5}\frac{\gamma}{i_0} + error_1 + error_2).$$

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

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

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 $P(X_0 = 1) \leq \frac{1}{15}$. 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

$$P(C|B) \leq \frac{14}{15}P(C|B \wedge A) + \frac{1}{15}P(D|B \wedge A) < \frac{1}{3^{(m+1)}}(\frac{14}{15}(1 + (\frac{2}{5}\gamma + 10^{-6})/i_0) + \frac{1}{15}(1 - (\frac{2}{5}\gamma - 10^{-6})/i_0)) = \frac{1}{3^{m+1}}(1 + (\frac{26}{75}\gamma + 10^{-6})/i_0)$$

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

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

$$P(D|B) \geq \frac{14}{15}P(D|B \wedge A) + \frac{1}{15}P(D|B \wedge \text{“}X_0 = 1\text{''}) = \frac{14}{15}P(D|B \wedge A) + \frac{1}{15}P(C|B \wedge A) > \frac{1}{3^{m+1}}(\frac{14}{15}(1 - (\frac{2}{5}\gamma + 10^{-6})/i_0) + \frac{1}{15}(1 + (\frac{2}{5} - 10^{-6})/i_0 = \frac{1}{3^{m+1}}(1 - ((\frac{26}{75}\gamma + 10^{-6})/i_0.$$

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

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 asterick equations. □