Can. J. Math.Vol. **45** (3), 1993 pp. 638–649

ON A CLASS OF GENERALIZED BAKER'S TRANSFORMATIONS

M. RAHE

ABSTRACT Let f define a baker's transformation (T_f, P_f) We find necessary and sufficient conditions on f for (T_f, P_f) to be an $N(\omega)$-step random Markov chain Using this result, we give a simplified proof of Bose's results on Holder continuous baker's transformations where f is bounded away from zero and one We extend Bose's results to show that, for the class of baker's transformations which are random Markov chains where N has finite expectation, a sufficient condition for weak Bernoullicity is that the Lebesgue measure $\lambda\{x \; f(x) = 0 \text{ or } f(x) = 1\} = 0$ We also examine random Markov chains satisfying a strictly weaker condition, those for which the differences between the entropy of the process and the conditional entropy given the past to time n form a summable sequence, and we show that a similar result holds A condition is given on f which is weaker than Holder continuity, but which implies that the entropy difference sequence is summable Finally, a particular baker's transformation is exhibited as an easy example of a weakly Bernoulli transformation on which the supremum of the measure of atoms indexed by n-strings decays only as the reciprocal of n

1. **Introduction.** In this article we derive some results about asymptotic independence properties of generalized baker's transformations. These objects, defined by Bose in his dissertation, can be used in two ways. First, given an arbitrary non-atomic ergodic transformation of entropy at most $\log 2$, there is an isomorphic representation of that transformation as a generalized baker's tranformation. Facts about the representation can then be carried back to the original via the isomorphism. Second, each function $f\colon [0,1) \longmapsto [0,1]$ determines a generalized baker's transformation. By proper choice of f, one has the potential of constructing simple examples of transformations with interesting properties.

More particularly, denote by $\mathbf{N}$ the set of nonnegative integers and by $\mathbf{Z}$ the set of all integers. Let $\hat{\mu}$ be a non-atomic probability measure on the set $X = \{0,1\}^{\mathbf{N}}$ of one-sided infinite strings of zeros and ones, endowed with the σ-algebra generated by the finite cylinder sets. We may embed a distribution copy of $(X, \hat{\mu})$ into the unit interval $[0,1)$ in a natural manner by inductively defining subintervals on $[0,1)$ with measures corresponding to the measures of the finite cylinder sets. In particular, if s is a finite string, let A_s denote the cylinder set of all sequences in X which begin with s. We begin by subdividing the interval $[0,1)$ into two left-closed, right-open intervals I_0 and I_1, having lengths $\hat{\mu}(A_0)$ and $\hat{\mu}(A_1)$, respectively, with I_0 preceeding I_1. Then, at the n-th stage, given

Received by the editors July 18, 1991

AMS subject classification Primary 28A65

Key words and phrases generalized baker's transformation, random Markov chains, uniform martingales, weak Bernoulli

an interval I_s indexed by a string s of length n, we may subdivide it into two half-open subintervals I_{s0} and I_{s1}, having relative lengths $\hat{\mu}(A_{s0}|A_s)$ and $\hat{\mu}(A_{s1}|A_s)$, respectively, with I_{s0} preceding I_{s1}.

Since $\hat{\mu}$ is non-atomic, there is a measure preserving map Φ: $[0, 1) \longmapsto X$ which associates to each number $r \in [0, 1)$ the unique sequence s whose initial segments index the nested intervals containing r.

Let $\mathbf{G}$ denote the σ-algebra on X generated by the coordinate maps with indices greater than zero. For any version of $\hat{\mu}(A_0|\mathbf{G})$, there is a corresponding function f: $[0, 1) \longmapsto [0, 1]$ defined by $f(r) = \hat{\mu}(A_0|\mathbf{G})\big(\Phi(r)\big)$. The function f then determines a generalized baker's transformation, which will represent the original transformation.

Alternatively, we may begin, not with a given transformation, but with a particular f: $[0, 1) \longmapsto [0, 1]$. This choice of f determines a measure preserving transformation. By careful choice of f, we can provide examples of measure preserving transformations of various types.

In his dissertation, C. Bose showed that if the function f is Hölder continuous and bounded away from zero and one, then $\hat{\mu}$ exhibits an extremely strong form of asymptotic independence. He then derived precise estimates of the rate at which such independence is obtained, based on the constant and exponent of Hölder continuity.

In this note we show some related results. We begin by deriving necessary and sufficient conditions on f for (T_f, P_f) to be an $N(\omega)$-step random Markov chain. In essence, f must be continuous except for possible jump discontinuities at a countable number of points, at which we must have right and left limits, and then only if the points of discontinuity are properly placed and the jump sizes are controlled. We give an example of a baker's transformation with f being a step function of two levels, where (T_f, P_f) is not a random Markov chain.

Using the random Markov chain analysis, we then give a simplified proof of Bose's results on Hölder continuous baker's transformations where f is bounded away from zero and one. In particular, we examine the rate at which asymptotic independence obtains and its relationship to the random step size.

We extend Bose's results to show that, for the class of baker's transformations which are random Markov chains where N has finite expectation, a sufficient condition for weak Bernoullicity is that the Lebesgue measure $\lambda\{x : f(x) = 0 \text{ or } f(x) = 1\} = 0$. An example is given of a baker's transformation which has $f(x) = 0$ on a set of positive measure and which is not ergodic.

We then explore an alternative to the condition of having finite expected step size, a weaker requirement involving summability of certain conditional entropy differences, deriving similar results. A uniformity condition is given on f which is weaker than Hölder continuity but which implies that the process has this summability property.

Finally, we apply the generalization of Bose's results to the case where $f = 1 - x$, for which the resulting baker's transformation is a weakly Bernoulli transformation on which the supremum of the measure of atoms indexed by n-strings decays only as the reciprocal of n. This gives an easy example of the degree to which the Shannon McMillan Breiman theorem fails to give uniform convergence, even in the presence of strong asymptotic independence.

2. **Notation.** Let X be a Lebesgue space, B a σ-algebra of subsets of X, and μ a probability measure on B.

For $\mathcal{P}$ and $\mathcal{Q}$ partitions of X, we write $\mathcal{P} \perp_\epsilon \mathcal{Q}$ if

$$\sum_{A \in \mathcal{P},\ B \in \mathcal{Q}} |\mu(A \cap B) - \mu(A)\mu(B)| < \epsilon.$$

Denote by P_n^m the partition $\vee_n^m T^i P$.

DEFINITION. By a *process* we mean a pair (T, P) consisting of P, a finite partition of subsets of B, and $T \colon X \longmapsto X$, a μ-measure preserving transformation.

DEFINITION. We say that a process (T, P) is a *K-automorphism* if given ϵ and k, there exists n such that $P_{-\infty}^0 \perp_\epsilon P_n^{n+k}$.

DEFINITION. We say that a process (T, P) is *weak Bernoulli* if given ϵ, there exists n such that for all j and k we have $P_{-j}^0 \perp_\epsilon P_n^{n+k}$.

After Bose, we define a generalized baker's transformation. Let λ be Lebesgue measure on the Lebesgue subsets B of $[0, 1)$. Let $\mu = \lambda \times \lambda$ be 2-dimensional Lebesgue measure on the Lebesgue subsets G of the unit square $S = \{(x, y) \mid x \in [0, 1), y \in [0, 1)\}$, and let $f \colon [0, 1) \longmapsto [0, 1]$ be a B-measurable function. Define two mappings $\psi, \phi \colon [0, 1) \longmapsto [0, 1]$ by the formulae

$$\psi_f(x) = \int_0^x f(t)\, dt$$
$$\phi_f(x) = 1 - \int_x^1 1 - f(t)\, dt.$$

For each $(x, y) \in S$ we set

$$T_f(x, y) = \begin{cases} (\psi_f(x), \frac{y}{f(x)}), & \text{if } 0 \le y \le f(x) \\ (\phi_f(x), 1 - \frac{1-y}{1-f(x)}), & \text{if } f(x) \le y < 1. \end{cases}$$

It is easy to show, that $T_f \colon S \longmapsto S$ is measurable, invertible, and preserves μ.

Let P_f be the partition of S into two sets

$$P_0 = \{(x, y) \mid x \in [0, 1), f(x) \le y < 1\}$$

and $P_1 = S - P_0$. We may consider the process (T_f, P_f). Note that for each positive integer $j > 1$, an atom of the partition P_1^j consists of $I \times [0, 1)$, where I is a left closed, right open interval.

https://doi.org/10.4153/CJM-1993-035-0 Published online by Cambridge University Press

3. **The baker's transformation representation.** We begin by establishing an isomorphism.

THEOREM. *Let μ be a non-atomic measure on the space $X = \{0,1\}^{\mathbf{Z}}$ which is invariant and ergodic under the shift transformation T. Let $P = \{P_0, P_1\}$ be the time-zero partition on X. Suppose that the entropy $h(T,P)$ is some constant $h > 0$. Then the embedding function Φ: $[0,1) \longmapsto X$ establishes an isomorphism between (T,P) and (T_f, P_f), where $f(r) = \mu(P_0|P_1^\infty)\big(\Phi(r)\big)$.*

PROOF. The function Φ is clearly measure preserving, hence we need only show that the partition P_f will separate points under T_f. Since $h(T,P) > 0$, then $\mu(P|P_1^\infty)$ will be bounded away from zero and one by some value c with $0 < c < \frac{1}{4}$ on some set $A \in P_1^\infty$ with $\mu(A) > 0$. Hence f will also be bounded away on $\Phi^{-1}(A)$.

For each positive integer n, the atoms of $(P_f)_1^n$ are rectangles $I \times [0,1)$. As n increases, the width of the interval goes to zero since μ is non-atomic. Thus, if (x,y) and $(\hat{x},\hat{y})$ are points in S which lie in the same atom of $(P_f)_1^\infty$, then $x = \hat{x}$. Now since μ is ergodic, it follows that the T^{-1}-orbit of points in $\Phi(x)$ will enter the set A infinitely often, with μ-probability one. Hence $y = \hat{y}$, since two points (x,y) and $(x,\hat{y})$ in an atom of $\vee_{-n}^{0} T_f^i P_f$ have second coordinates that differ by at most $(1-2c)^m$, where m is the number of times that $T^{-j}\big(\Phi(x)\big)$ has entered A for $1 \le j \le n$.

It follows as an obvious corollary that if we show that (T_f, P_f) is ergodic, mixing, a K-automorphism, or Bernoulli, the original (T,P) must also be.

By a well known theorem of Krieger, if (T,P) is ergodic with entropy at most $\log 2$, one can construct a two-set generating partition. The above approach will then yield an isomorphic representation of (T,P) as a generalized baker's transformation.

Of course, the resulting (T_f, P_f) will also be ergodic in this case, since we are copying an existing ergodic transformation. However, if one begins instead only with an arbitrary uniformly continuous f, and constructs a transformation T_f in the obvious way, it need not be the case that P_f generates under T_f, nor can we rule out the existence of T_f-invariant sets G with $0 < \mu(G) < 1$.

As an example, consider

$$f(x) = \begin{cases} 0, & x \in [0, \frac{1}{3}), \\ 3x - 1, & x \in [\frac{1}{3}, \frac{2}{3}), \\ 1, & x \in [\frac{2}{3}, 1). \end{cases}$$

The set $\{[\frac{1}{6}, \frac{1}{3}) \cup [\frac{2}{3}, \frac{5}{6})\}$ is clearly invariant and P_f does not separate points, since T_f merely flips pairs of corresponding points in the two rectangles.

The above example has a continuously differentiable f, but is not ergodic since there is a nontrivial invariant set $I \times [0,1)$ with $I \subset \{x : f(x) = 0 \text{ or } f(x) = 1\}$. Our goal is to show that the only way that ergodicity can fail to hold for functions f which display sufficient regularity is to have a set of positive Lebesgue measure where f is zero or one.

4. **Random Markov processes.** In [2], Kalikow defines a particular kind of skew product over an independent process which takes non-negative integer values.

DEFINITION. Let F be a finite set. Let $\{a_\iota, N_\iota\}$ be a stationary process, where each $a_\iota \in F$, each $N_\iota \in \mathbf{N}$, N_0 is independent of $\{a_\iota, N_\iota\}_{\iota<0}$ and for each j

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

Then $\{a_\iota, N_\iota\}_{\iota\in\mathbf{Z}}$ is a *complete random Markov process.*

DEFINITION. A *random Markov process* is the projection on the first coordinate of a complete random Markov process, *i.e.*, if $\{a_\iota, N_\iota\}$ is a complete random Markov process, then $\{a_\iota\}$ is a random Markov process.

It is easy to see that if the values of the random variable N are bounded by some fixed integer n, then a random Markov process is actually an n-step Markov process.

Kalikow also defined the concept of a uniform martingale.

DEFINITION. Let F be a finite set, and let $\{a_\iota\}_{\iota\in\mathbf{Z}}$ be a stationary process, with all $a_\iota \in F$. If, for all $\epsilon > 0$, there exists $N_\iota \in \mathbf{N}$ such that for all $M > N_\iota$ and all $\{F_\iota\}_{\iota=0}^{\infty}$ with all $F_\iota \in F$,

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

then $\{a_\iota\}$ is a uniform martingale.

He then establishes the following fact.

THEOREM. *A process is a uniform martingale iff it is a random Markov process.*

5. **Baker's transformations which are random Markov chains.** In this section we examine necessary and sufficient conditions on f for (T_f, P_f) to be a random Markov chain.

DEFINITION. A function $f\colon [0,1) \longrightarrow [0,1]$ is *regular* if given $\epsilon > 0$ there exists a finite sequence $d_0 = 0, d_1, \ldots, d_n = 1$ partitioning $[0,1)$, such that for each $j = 0, \ldots, n-1$ we have $|\sup_{(d_j,d_{j+1})} f - \inf_{(d_j,d_{j+1})} f| < \epsilon$.

DEFINITION. We say that a function $f\colon [0,1) \longrightarrow [0,1]$ is *adapted* if given $\epsilon > 0$ there exists n such that all jump discontinuities of size at least ϵ occur at endpoints of intervals which are projections of the vertical σ-algebra $(P_f)_1^n$.

We remark that regularity is implied by either piecewise uniform continuity (with finitely many intervals) or by bounded variation, but is strictly weaker than either of these conditions.

Fix a given regular function f. We will consider a sequence $\epsilon_n = 2^{-n}$ and denote by $\mathbf{D}$ the countable union of all d's given by the definition of regularity. For all $x \notin \mathbf{D}$, we have continuity at x. Moreover, even for $x \in \mathbf{D}$, both $\lim_{y\to x^+} f$ and $\lim_{y\to x^-} f$ must exist.

By Kalikow's result, a process is a random Markov process iff it is a uniform martingale. For generalized baker's transformations, this reduces to saying that f is the uniform limit (outside a countable set) of the conditional expectations $f_n = E\big(f \mid \pi(P_f)_1^n\big)$. Each f_n is a step function, whose jumps occur at endpoints of intervals which are projections of the vertical σ-algebra $(P_f)_1^n$.

Clearly, any uniform limit of step functions is regular, hence regularity is a necessary condition on f for (T_f, P_f) to be a random Markov chain. On the other hand, it is easy to see that regularity is not sufficient. As an example, consider $f(x) = 2/3$ on $[0, c)$ and $f(x) = 1/3$ on $[c, 1)$, for some c. Since the stationary distribution is determined by transition probabilities, then by avoiding at most a countable collection of values, we can choose c so that (T_f, P_f) is not n-step Markov for any $n \in \mathbf{N}$. Moreover, (T_f, P_f) is not even random Markov, since uniform approximation fails for any $\epsilon < 1/3$ in the interval containing the discontinuity. In other words, regularity fails to be sufficient precisely because f must also be adapted.

THEOREM. *The generalized baker's transformation (T_f, P_f) is a random Markov chain iff f is regular and adapted.*

PROOF. The above discussion has shown that regularity and adaptation are necessary. The proof that they are sufficient is also easy. Given $\epsilon > 0$, we apply the adaptation condition for $\epsilon/3$ to get N_1 such that all jump discontinuities of size $\epsilon/3$ or greater are contained in endpoints of projections of atoms in $(P_f)_1^{N_1}$. Then, applying regularity for $\epsilon/3$, we get a finite sequence $d_0, \ldots, d_n$ for which $|\sup_{(d_j, d_{j+1})} f - \inf_{(d_j, d_{j+1})} f| < \epsilon/3$, $j = 1, \ldots, n-1$.

If two distinct values of d are never separated by $\pi(P_f)_1^\infty$, then by Bose's result, f must be identically zero or identically one on the entire atom. Otherwise we may take $N_2 > N_1$ so large that the projection of each atom contains at most one of the d_j in its interior. Either situation will suffice to find a step function $\hat{g}$, whose intervals of constancy are measurable with respect to the projection of $(P_f)_1^{N_2}$, whose value on each interval is the infimum of f on that interval, and which is within ϵ of f, except possibly at the finite collection of endpoints of intervals.

We can use this fact to write f as an infinite sum. Set $f_0 = f$, so that $f_0 \leq 2^0$; and given $\epsilon = 2^{-1}$, find $\hat{g}_1$, measurable with respect to $\pi(P_f)_1^{n_1}$, for some n_1. Let g_1 be $\hat{g}_1$ truncated at 2^{-1}. Then by induction, given f and $g_1, g_2, \ldots, g_l$, with $g_j \leq 2^{-j}$, for $j = 1, 2, \ldots, l$, let $f_l = f - \sum_{j=1}^{l} g_j$. Then $f_l \leq 2^{-l}$, and given $\epsilon_{l+1} = 2^{-l-1}$, find $\hat{g}_{l+1}$ measurable on $\pi(P_f)_1^{n_l}$. Let g_{l+1} be $\hat{g}_{l+1}$ truncated at 2^{-l-1}.

Thus, given a sequence $\epsilon_k = 2^{-k}$, we can write f as an infinite sum of step functions g_k, where $0 \leq g_k \leq 2^{-k}$ and g_k is measurable with respect to the projection of $(P_f)_1^{n_k}$ for some strictly increasing sequence $n_k \longrightarrow \infty$. This construction results naturally in the required skew product, where the values $P(x_0 = 0 \mid x_1 \cdots x_{n_k}) = 2^k g$ determine the transition probabilities of an n_k-step Markov process on a skewing set of measure 2^{-k}.

Of equal importance, statements about the random step size $N(\omega)$, which is a random variable in the skew product but not on the unit square, are interpretable as statements about the summands g_k which make up f.

In particular, every uniformly continuous function f: $[0, 1) \longrightarrow [0, 1]$ gives a random Markov chain. Moreover, Bose's result that Hölder continuous baker's transformations with f bounded away from zero and one are weakly Bernoulli is a statement about a particular class of random Markov chains. This observation will allow us not only to simplify some of Bose's proofs, but also to extend the results to a larger class where f need not be bounded away.

6. **Hölder continuous baker's transformations with f bounded away.** We remark that it is an easy calculation that the Hölder condition $|f(x) - f(y)| < M|x - y|^\alpha$ and the existence of $\delta > 0$ for which $\delta < f < 1 - \delta$, imply that the expected step size $E[N]$ is finite. It suffices to take

$$n_k = \left[\frac{-k\log 2 - \log M}{\alpha \log(1-\delta)}\right] + 1,$$

(where [ ] indicates the greatest integer function), since the maximum size of an atom in $(P_f)_1^{n_k}$ is $(1-\delta)^{n_k}$, and we then have

$$|f(x) - f(y)| \le M|x - y|^\alpha \le M(1-\delta)^{\alpha n_k} \le 2^{-k},$$

as required. Thus $E[N] = \sum n_k 2^{-k} < \infty$.

Moreover, Bose's result that $(P_f)_{-\infty}^0 \perp^{\hat{M}\theta^{\sqrt{R}}} (P_f)_R^\infty$ also becomes a more transparent and tractable computation. One can demonstrate that the distributions of $(P_f)_R^\infty$ on two atoms of $(P_f)_{-m}^0$ are ϵ-independent in the following manner: use the finite expectation of $N(\omega)$ to find $\hat{N}$ sufficiently large that, except for possibly a set of measure $\epsilon/2$, we have $N(\omega) \le \hat{N}$, $N(T\omega) \le \hat{N} + 1, \ldots$, *i.e.*, the random Markov process never looks back farther than a certain fixed past of length $\hat{N}$. Then, given $A, B \in (P_f)_{-m}^0$, form a joining of the (nonstationary) distributions conditioned on A and B, which is product measure until the first time that both strings agree in a block of $\hat{N}$ consecutive times, starting at a time which is a multiple of $\hat{N}$. After that point, the distribution remains identical in both processes.

It will require at most $R = \hat{N}L$ outputs to get such an agreement on $\hat{N}$ consecutive places, on a set of relative measure at least $1 - \epsilon/2$, where L is the smallest value for which $(1 - \delta^{2\hat{N}})^L < \epsilon/2$.

But one can easily compute $\hat{N}$. By the expression for n_k given above, $n_k = O(-k\log 2)$. On the other hand, for $\hat{N} = n_k$,

$$\begin{aligned}\sum_{j=0}^\infty P\big(N(T^j\omega) > \hat{N} + j\big) &\le \sum_{j=0}^\infty P\big(N(T^j\omega) > n_{k+j}\big) \\ &= \sum_{j=0}^\infty 2^{-k-j} = 2^{-k-1}.\end{aligned}$$

Thus, it suffices to take $2^{-k-1} < \epsilon/2$, or $-k = O(\log \epsilon)$, so $\hat{N} = O(\log \epsilon)$. But we also have $L = O(\log \epsilon)$. Hence $R = \hat{N}L = O(\log \epsilon \log \epsilon)$, as desired.

7. **Baker's transformations with finite expected step size.** As shown in our previous example, not every baker's transformation need be ergodic, even if f satisfies a Lipschitz condition: there must also be some restriction on the set where f is zero or one. On the other hand, we need not be content with Bose's stringent requirement that $\delta < f < 1-\delta$. Instead, it is sufficient to require that the set of zeros and ones be negligible in the sense of Lebesgue.

THEOREM. *If (T_f, P_f) is a random Markov chain with finite expected step size and $\lambda\{x : f(x) = 0 \text{ or } 1\} = 0$, then (T_f, P_f) is weakly Bernoulli. Moreover, the rate at which $(P_f)_{-\infty}^{0}$ becomes ϵ-independent of $(P_f)_R^{\infty}$ is easily computed in terms of the rates at which $P(N(\omega) > \hat{N}) \longrightarrow 0$ and $\lambda(f > \delta) \longrightarrow 1$.*

PROOF. In the previous section, we reduced the problem of establishing weak Bernoullicity to easier problems: finding $\hat{N}$, the existence of which followed from a finite expected step size, and a separate problem of finding L, which resulted from a lower bound δ on f. In the present situation, given ϵ and $\hat{N}$, we take $\delta = \delta(\epsilon, \hat{N})$ to be so small that $\lambda\{x : \delta < f\} > 1 - (\epsilon/8\hat{N})^8$. Thus, all except $\epsilon/4$ in measure of the atoms in $(P_f)_{-m}^{0}$, have all except $\epsilon/4$ in relative measure of their mass in a set on which the frequency of blocks of length $\hat{N}$, beginning at times which are multiples of $\hat{N}$, on which $f > \delta$ at each time index is at least $1 - \epsilon/4$. Then the previous argument goes through.

8. **Summable convergence of conditional entropy.** Kalikow has shown that the class of random Markov chains with finite expected step size is closed under the taking of inverses, *i.e.*, if one runs a random Markov chain backwards, one gets another random Markov chain with finite expected step size. In this section, we consider another condition which implies this same property.

DEFINITION. We say that a transformation (T, P) has *summably convergent conditional entropy* if

$$\sum_{k=1}^{\infty} h(P \mid P^{-1} \cdots P^{-k}) - h(P \mid P^{-1} \cdots) < \infty.$$

This concept was originally defined in [3], where it was shown that if (T, P) is mixing and has summably convergent conditional entropy, then (T, P) is weakly Bernoulli. In the context of generalized baker's transformations, we can replace the mixing condition with $\lambda\{x : f(x) = 0 \text{ or } 1\} = 0$.

REMARK. It is perhaps worthwhile to mention that Bose has an example in his dissertation of a baker's transformation (T_f, P_f) where $\frac{1}{3} \leq f(x) \leq \frac{2}{3}$, yet T^2 is not ergodic. Hence, the condition $\lambda\{x : f(x) = 0 \text{ or } 1\} = 0$ does not imply mixing.

THEOREM. *If (T_f, P_f) has summably convergent conditional entropy, and $\lambda\{x : f(x) = 0 \text{ or } 1\} = 0$, then (T_f, P_f) is weakly Bernoulli.*

PROOF. In the proof of the corresponding theorem for $E[N]$ finite, we established a value $\hat{N}$ such that $(P_f)_1^m \perp_\epsilon (P_f)_{-l}^{-\hat{N}}$ on each atom $A \in (P_f)_{-\hat{N}}^{0}$, for all m, l. We now do exactly the same thing, using the entropy condition instead of $E[N]$. (This is essentially the same argument as [3], repeated here for completeness.) given $\epsilon > 0$, there exists $\delta > 0$ such that for any partitions $\mathcal{P}$ and $\mathcal{Q}$, $h(\mathcal{P}) - h(\mathcal{P} \mid \mathcal{Q}) < \delta$ implies $\mathcal{P} \perp_\epsilon \mathcal{Q}$. By the finite sum hypothesis, given δ, we can choose $\hat{N}$ such that for all positive integers m and l, we have

$$\begin{aligned} 0 &\leq h(P_1^m | P^0_{-\hat{N}}) - h(P_1^m | P^0_{-\hat{N}-k}) \\ &\leq h(P_1^m | P^0_{-\hat{N}}) - h(P_1^m | P^0_{-\infty}) \\ &= \sum_{j=1}^m [h(P^j | P^{j-1}_{-\hat{N}}) - h(P^j | P^{j-1}_{-\infty})] \\ &= \sum_{j=1}^m [h(P | P^{-1}_{-\hat{N}}) - h(P | P^{-1}_{-\infty})] < \delta. \end{aligned}$$

Then we have the requisite ϵ-independence conditioned on atoms of $P^{-1}_{-\hat{N}}$. The rest of the argument follows as before.

It is interesting to note that the baker's transformation with two-step f where (T_f, P_f) failed to be a random Markov chain also gives an example of a process which has conditionally convergent entropy, yet which is not a random Markov chain. We do not require *uniform* convergence of the martingale: instead, sufficiently rapid convergence in an L^1 sense will suffice.

We also note that by stationarity, the class of processes for which we have summably convergent conditional entropy is closed under taking inverses.

In section 10, we give an example of a random Markov process which does not have finite expected step size, but which has summably convergent conditional entropy, showing that for random Markov chains, $E[N] < \infty$ is not a necessary condition for weak Bernoullicity. It would be interesting to know if the entropy condition is necessary.

We can show that finite expected step size implies the entropy condition. We begin the analysis by constructing another sum. Except for countably many $x \in [0, 1)$, there is a natural sequence of $A_k \in (P_f)_1^k$ which is nested decreasing to $x \times [0, 1]$. Let $f_x(k)$ be the fluctuation of f on $\pi(A_k)$, *i.e.*, $f_x(k) = |\sup_{x \in \pi(A_k)} f(x) - \inf_{x \in \pi(A_k)} f(x)|$. Let $S_x = \sum_{k=1}^\infty f_x(k)$. Then

$$S = \int S_x \, d\lambda = \sum_{k=1}^\infty \int f_x(k)$$

is the desired sum.

Now by definition $E[N] = \sum_k n_k 2^{-k}$, and this sum will be finite iff we have $\sum_k \Delta n_k 2^{-k} < \infty$, where $\Delta n_k = n_k - n_{k-1}$. In fact, since

$$\sum f_x(k) \leq 2 \sum_k \Delta n_k 2^{-k} < \infty,$$

we have a uniform bound on S_x. On the other hand, the entropy condition requires no uniform bound on S_x, as we utilize in section 10.

https://doi.org/10.4153/CJM-1993-035-0 Published online by Cambridge University Press

For $t \leq 2^{-1}$, we have $-t\log t - (1-t)\log(1-t) \leq 2(-t\log t)$. Hence to show that $\sum_j h(P|P_1^j) - h(P|P_1^\infty) < \infty$, it suffices to consider the following sum

$$\mathcal{S} = \sum_{k=1}^{\infty} \sum_{j=n_k}^{n_{k+1}-1} \sum_{\cup A_l \in P_1^j} \int_{\pi(\cup A_l)} \nu_1\left(\frac{1}{\lambda(\pi(\cup A_l))} \int_{\pi(\cup A_l)} f(t)\, d\lambda(t)\right) - \frac{1}{\lambda(\pi(\cup A_l))} \int_{\pi(\cup A_l)} \nu_1\big(f(t)\big)\, d\lambda(t)\, d\lambda(t),$$

where the A_l are atoms in $P_1^{n_{k+1}-1}$ and $\nu_1(t) = -t\log t$. Note that for fixed k, on each $\cup A_l$, we have that $\max_{\pi(\cup A_l)} f(t) - \min_{\pi(\cup A_l)} f(t) \leq 2^{-k+1}$. Hence, by the concavity of ν_1, it suffices to assume that $0 \leq f(t) \leq 2^{-k+1}$.

Fixing $\cup = \pi(\cup A_l)$, we can rewrite

$$\int_{\cup}\left[\nu_1\left(\frac{1}{\lambda(\cup)}\int_{\cup} f(t)\, d\lambda\right) - \frac{1}{\lambda(\cup)}\int_{\cup}\nu_1\big(f(t)\big)\, d\lambda(t)\right] d\lambda(t) = \lambda(\cup)\int_{[0,R]} \nu_1\big(E_\mu(s)\big) - E_\mu\big(\nu_1(s)\big)\, d\mu(s),$$

where $R = 2^{-k+1}$ is the maximum range of f, and μ is the measure on $[0,R]$ induced by f, *i.e.*,

$$\mu(S) = \frac{\lambda(f^{-1}(S))}{\lambda(\cap)}.$$

We use the following technical lemma.

LEMMA. *There exists a constant $M > 0$ such that for each $x > 0$ we have*

$$0 \leq \sup_{\mu \in Prob\,[0,x]} \nu_1\big(E_\mu(s)\big) - E_\mu\big(\nu_1(s)\big) \leq Mx,$$

where E_μ is the expected value with respect to the probability measure μ on $[0,x]$.

PROOF. For $x = 1$, $\nu_1\big(E_{\hat\mu}(t)\big) - E_{\hat\mu}\big(\nu_1(t)\big)$ is a continuous function of $\hat\mu$. Hence by the compactness of the unit ball, it assumes a maximum value, which we denote by M. For $x < 1$, we consider the map $\xi_x\colon [0,1] \longrightarrow [0,x]$ defined by $\xi_x(t) = tx$. Any probability measure μ on $[0,x]$ then induces a probability measure $\hat\mu = \mu \circ \xi^{-1}$ on $[0,1]$, so that

$$\nu_1\big(E_\mu(s)\big) - E_\mu\big(\nu_1(s)\big) = \nu_1\big(E_{\hat\mu}(tx)\big) - E_{\hat\mu}\big(\nu_1(tx)\big).$$

On the other hand, $\nu_1(\alpha\beta) = \beta\nu_1(\alpha) + \alpha\nu_1(\beta)$, so

$$\begin{aligned}\nu_1\big(E_{\hat\mu}(tx)\big) - E_{\hat\mu}\big(\nu_1(tx)\big) &= \nu_1\big(xE_{\hat\mu}(t)\big) - E_{\hat\mu}\big(x\nu_1(t) + t\nu_1(x)\big)\\ &= E_{\hat\mu}(t)\nu_1(x) + x\nu_1\big(E_{\hat\mu}(t)\big) - xE_{\hat\mu}\nu_1(t) - \nu_1(x)E_{\hat\mu}(t)\\ &= x[\nu_1(E_{\hat\mu}(t) - E_{\hat\mu}\nu_1(t)].\end{aligned}$$

Thus the sum $\mathcal{S}$ is majorized by

$$\sum_{k=1}^{\infty}\sum_{j=n_k}^{n_{k+1}-1} M2^{-k+1} \leq \sum_{k=1}^{\infty} M\Delta n_{k+1} 2^{-k+1} < \infty.$$

We can summarize these results in the following theorem.

THEOREM. *If a random Markov chain has finite expected step size, then it has summably convergent conditional entropy. Moreover, the integrand in the entropy condition is uniformly bounded.*

9. A sufficient condition for having summably convergent conditional entropy.

DEFINITION. Let $g\colon [0,1] \longrightarrow [0,\infty)$ denote a monotone nondecreasing function satisfying $\int_0^1 -x^{-1}g(x)\,dx < \infty$. We say that f satisfies a *g-uniformity* condition if $|f(x_1) - f(x_2)| < g(|x_1 - x_2|)$, for all $x_1, x_2 \in [0,1)$.

THEOREM. *Suppose that f satisfies a g-uniformity for some g. Then if there exists $c > 0$ such that $c < f(x) < 1-c$, we have that (T_f, P_f) has summably convergent conditional entropy.*

PROOF. Let $\pi\colon [0,1)\times[0,1] \longrightarrow [0,1)$ be the projection $\pi(x,y) = x$. Let $A \in (P_f)_1^n$. Then $\pi(A)$ is a left closed, right open interval in $[0,1)$. Since $c < f(x) < 1-c$, it follows that $\lambda\big(\pi(A)\big) < (1-c)^n$.

Hence, if $x_1, x_2 \in \pi(A)$, then

$$|f(x_1) - f(x_2)| \le g\big((1-c)^n\big).$$

Thus

$$\frac{1}{\lambda(\pi(A))}\int_{\pi(A)} f\,d\lambda - f(x) = \mathrm{E}\big(f|(P_f)_1^n\big)(x) - f(x) \le g\big((1-c)^n\big),$$

for each $x \in \pi(A)$, where $\mathrm{E}\big(f|\pi(P_f)_1^n\big)(x) = f_n(x)$ denotes the conditional expectation. Define

$$\nu(x) = \begin{cases} -x\log x - (1-x)\log(1-x), & 0 < x < 1 \\ 0, & \text{otherwise.} \end{cases}$$

It follows that

$$\sum_{n=1}^{\infty} h\big(P_f|(P_f)_1^n\big) - h\big(P_f|(P_f)_1^\infty\big) \le \sum_{n=1}^{\infty}\int_0^1 \big|\nu\big(f_n(x)\big) - \nu\big(f(x)\big)\big| < \infty,$$

since by the intermediate value theorem

$$\big|\nu\big(f_n(x)\big) - \nu\big(f(x)\big)\big| \le \log\Big(\frac{1-c}{c}\Big) g\big((1-c)^n\big),$$

and

$$\sum_{n=1}^{\infty} g\big((1-c)^n\big) \le \frac{-1}{\log((1-c)}\int_0^1 \frac{g(x)}{x}\,dx$$

by the standard change of variables argument.

https://doi.org/10.4153/CJM-1993-035-0 Published online by Cambridge University Press

10. **A weak Bernoulli example.** The condition that $c \leq f(x) \leq 1 - c$ allowed us to guarantee that $\sup_{A \in (P_f)_1^n} \mu(A)$ decreases exponentially in n. Although the Shannon McMillan Breiman Theorem assures us that $\mu(A)$ decays exponentially for most atoms, it is not true in general that the supremum will decrease this fast, even if (T_f, P_f) is weakly Bernoulli. As an example, we consider the generalized Baker's transformation (T_f, P_f) with $f(x) = 1 - x$.

Denote by A_j^n, $j = 1, \ldots, 2^n$, the atoms of $(P_f)_1^n$, from left to right. By symmetry, one observes that $\lambda\big(\pi(A_j^n)\big) = \lambda\big(\pi(A_{2^n-j+1}^n)\big)$ for subscripts $j = 1, \ldots, 2^{n-1}$. Again by symmetry, it suffices to consider iteration under the map $\psi_f(x) = \int_0^x f(x)\,dt = x - \frac{x^2}{2}$. The map is monotone, and one can see by induction that

$$\frac{1}{n+1} \leq \lambda\big(\pi(A_1^n)\big) \leq \frac{2}{n+1}$$

and

$$\lambda\big(\pi(A_2^n)\big) = \frac{\lambda(\pi(A_1^{n-1}))^2}{2} \leq \frac{2}{n^2}.$$

By monotonicity of f, $\lambda\big(\pi(A_1^n)\big) \geq \lambda\big(\pi(A_j^n)\big)$ for $j = 1, 2, \ldots, 2^n$. By symmetry, $\lambda\big(\pi(A_1^n)\big) = \lambda\big(\pi(A_{2^n}^n)\big)$. Moreover, a computation reveals that

$$\lambda\big(\pi(A_2^n)\big) = \lambda\Big(\pi\Big(\psi(A_{2^n-1}^{n-1})\Big)\Big) = \lambda\big(\pi(A_{2^n-1}^n)\big),$$

so by induction $\lambda\big(\pi(A_2^n)\big) \geq \lambda\big(\pi(A_j^n)\big), j = 2, \ldots, 2^n - 1$, for all positive integers n.

Thus, there exists a constant $M > 0$ such that

$$|\nu f_n - \nu f| \leq \nu(|f_n - f|) \leq -|f_n - f|(\log|f_n - f| + 1) \leq \Big(\frac{M \log n}{n^2}\Big)$$

on $\bigcup_{j=2}^{2^n-1} A_j^n$ and $|\nu f_n - \nu f| \leq \frac{M}{n}$ on A_1^n and $A_{2^n}^n$. Hence $\sum_{n=1}^{\infty} \int_0^1 \nu f_n - \nu f \, d\lambda < \infty$. Thus (T_f, P_f) has summably convergent conditional entropy, hence by the results of section 8, the process is weakly Bernoulli.

On the other hand, as noted above, $\sup_j \lambda\big(\pi(A_j^n)\big) \geq \frac{1}{n+1}$. Thus it follows that the expected step size is infinite.

REFERENCES

1. C J Bose, *Generalized baker's transformations*, Ergodic Theory and Dynamical Systems **9**(1989), 1–17
2. S Kalikow, *Random Markov Processes and Uniform Martingales*, Isreal Journal of Mathematics **71**(1990), 33–54
3. A del Junco and M Rahe, *Finitary codings and weak Bernoulli partitions*, Proceedings of the American Mathematical Society **75**(1979), 259–264

Texas A & M University
College Station, Texas
U S A