

Riffle Shuffles and Dynamical Systems on the Unit Interval

Steven P. Lalley*
Purdue University

June 6, 1994

1 Introduction

The preferred method of randomizing a deck of N cards is the *riffle shuffle*—cut the deck into two stacks, then riffle the two stacks together. A number of mathematical models have been proposed for this process. The model which has received the most attention, and about which the most is known, is the *GSR* (for Gilbert, Shannon, and Reeds) shuffle. In this model, all permutations with exactly one or two rising sequences are equally likely (a *rising sequence* for a permutation π is a maximal sequence of consecutive integers $i, i + 1, \dots, j$ such that $\pi(i) < \pi(i + 1) < \pi(i + 2) < \dots < \pi(j)$). An alternative description of the shuffle is as follows: break the deck at an integer k chosen from the Binomial $(N, \frac{1}{2})$ distribution, then perform an “unbiased” riffle of the two stacks. (In an *unbiased riffle* of two stacks of sizes A, B , cards are dropped one at a time from the bottoms of the stacks; at any step, if the two stacks have A', B' cards, respectively, then the probability that the next card is dropped from the stack with A' cards is $A'/(A' + B')$.)

Bayer and Diaconis (1992) gave an intriguing “dynamical” description of the GSR shuffling process. In this representation, the individual cards are represented by points of the unit interval; these points are gotten by taking a sample of size N from the uniform distribution. The assignment of points to cards is such that the original order of the N cards in the deck agrees with the natural order of the corresponding points in the unit interval. The evolution of the deck is determined by the motion of the N points under the action of the doubling map $T : x \rightarrow 2x \bmod 1$: the order of the cards in the deck after n shuffles is determined by the order of the (marked) points in $[0,1]$ after T has been applied n times. Using this representation, Bayer and Diaconis obtained the following exact expression for the probability that the deck is in state π after n shuffles:

$$\binom{2^n + N - r}{N}, \tag{1}$$

where r is the number of rising sequences in the permutation π . From this they deduced an improvement of Aldous’ (1983) theorem on the rate of mixing in the GSR shuffle, which

*Supported by National Science Foundation Grant DMS-9307855

states that for any $\varepsilon > 0$,

$$\lim_{N \rightarrow \infty} d_N((1 - \varepsilon)\frac{3}{2} \log_2 N) = 1 \text{ and } \lim_{N \rightarrow \infty} d_N((1 + \varepsilon)\frac{3}{2} \log_2 N) = 0 \quad (2)$$

where $d_N(n)$ is the total variation distance between $\mathcal{D}(\mathcal{X}_n)$ and the uniform distribution, and X_n is the state of the deck after n shuffles.

The purpose of this paper is to show that there is an analogous dynamical description for a large class of models for riffle shuffles, and to show how the associated dynamical system constrains the rate of convergence to uniformity. The GSR shuffle is but one of a multitude of interesting models for riffle shuffles. It may be generalized in a number of obvious ways, for instance: (1) Instead of breaking the deck at k with a Binomial $(N, \frac{1}{2})$ distribution, one might break it at k with a Binomial (N, p) or a uniform distribution, or at a nonrandom k . (2) Instead of using an “unbiased” riffle, in which all riffles with given stack sizes are equally likely, one might use a “biased” riffle, in which cards from the left stack (say) are more likely to drop than cards from the right stack. For instance, one might consider the (u, v) -weighted riffle, in which the chance that the next card is dropped from the left stack is $uA/(uA+vB)$, with A, B being the sizes of the left and right stacks, respectively. (3) Instead of riffling all the stacks, one might use a more complicated rule for combining stacks. For example, keep two stacks, a “top” t and a “bottom” b . On each shuffle, break the bottom stack into two substacks, b_T and b_B ; riffle t and b_B to obtain a replacement for b , and replace t by b_T . Each of these is an interesting model for card-shuffling.

The associated dynamical system for a riffle shuffle may be roughly described as follows (see sec.2 for a more precise description). In any riffle shuffle there are finitely many “stacks”, which may be combined in various ways. Let \mathcal{A} be a finite alphabet indexing the set of available stacks. If the shuffle is repeated infinitely many times, each card will have an “orbit”, an \mathcal{A} -valued sequence describing which stack the card visits at each time. The join of the N orbits is the associated dynamical system; the time evolution is described by the unilateral shift operator. The orbit of a single randomly chosen card will be called the associated marginal process. We will show (sec. 3) that for many interesting shuffling models, the associated marginal process has (at least asymptotically as the deck size $N \rightarrow \infty$) a simple form, which in a number of cases may be described as the motion of a randomly chosen point of the unit interval under a deterministic map. But we will also show (sec. 4) that in general the orbits of the different cards are *not* independent — the GSR shuffles are exceptional in this regard. Finally, we will show (sec. 5) that the limiting form of the associated marginal process constrains the rate of convergence to uniformity in a rather simple way: namely, for any $\varepsilon > 0$,

$$d_N((1 - \varepsilon)h^{-1} \log N) \rightarrow 2$$

as $N \rightarrow \infty$, where h is the “fiber entropy” of the limiting dynamical system. Aldous’ result for the GSR shuffle suggests that for many of the riffle shuffles considered here there is a rapid transition to uniformity at about $C \log N$ repetitions, with C depending on the details of the shuffling process. Our result shows that if this in fact true, then h^{-1} is a lower bound for C . The GSR shuffle, where $C = \frac{3}{2}h^{-1}$, shows that in general $C > h^{-1}$; but in general the cutoff phenomenon seems rather difficult to establish, and it appears that the constant C may depend on the parameters of the shuffling process in a complicated and mysterious way.

2 The Associated Dynamical System

2.1 The Canonical Dynamical System

Every stochastic process X_n taking values in the permutation group \mathcal{S}_N has a natural representation as a discrete-time particle system on the unit interval. Represent the “cards” $\{1, 2, \dots, N\}$ by “particles” situated at the points $\{0, \frac{1}{N-1}, \frac{2}{N-1}, \dots, 1\}$; at time 0 the particle representing card i is situated at the point $\frac{i-1}{N-1}$. At any time n thereafter, the particle representing card i is located at $\frac{X_n(i)-1}{N-1}$. Clearly, X_n is completely determined by the history of this particle system. We will refer to this particle system as the *canonical associated dynamical system*.

The canonical dynamical system is not the only or even the most natural dynamical system associated with the shuffle. For riffle shuffles, a more useful dynamical system is gotten by “symbolic dynamics”, in which one marks the “orbit” of a card by the sequence of “stacks” it visits. The remainder of this section is devoted to a description of this system, which we shall call the *associated symbolic dynamical system*.

2.2 The GSR Shuffle

Consider first a riffle shuffle that cuts the deck into just two stacks and then riffles them together, e.g., the GSR shuffle. Suppose that the shuffle is repeated infinitely many times; then each card will have an “orbit”, specifically, the sequence of zeroes and ones recording which of the two stacks (top = 0, bottom = 1) the card visits at each step. What characterizes a *riffle* shuffle is that the original order of the cards in the deck agrees with the lexicographic ordering of their orbits. This is because if card b lies below card a originally, then card b will continue to lie below card a for as long as their orbits agree; and thus, at the first time their orbits *disagree* (if there is such a time), card b will be in the bottom stack and card a in the top. If the shuffling process is ergodic, then the orbits of cards a and b cannot agree forever, so the pairing of cards and infinite 0-1 sequences is injective.

Hence, for an ergodic shuffling process, the history of the deck is completely determined by the N infinite 0-1 sequences giving the orbits of the different cards. To determine where card i is at time n , mark sequence x^i ; apply the unilateral shift σ to each of the N orbits n times; then determine where $\sigma^n x^i$ lies in the lexicographic ordering of $\sigma^n x^1, \sigma^n x^2, \dots, \sigma^n x^N$. Observe that if the shuffling process is invariant under relabellings of the cards (as for the GSR shuffle and its natural generalizations) then the process

$$Y_n \triangleq \{\sigma^n x^1, \sigma^n x^2, \dots, \sigma^n x^N\}$$

is stationary.

For the GSR shuffle, the join of the N orbits has an appealingly simple interpretation. Identify each orbit with a point of the unit interval $[0, 1]$ by binary expansion. Then the joint distribution of the set of N points so obtained is identical to that of a sample (unordered) of N i.i.d. random variables, each uniformly distributed. This was apparently first noticed by Diaconis (1988) although the interpretation of the time-reversed orbits as Bernoulli processes was known to Reeds (1981) and further exploited by Aldous (1983). Thus, the GSR shuffle admits the following description: Start with a sample of N uniformly distributed

points, and attach indices $1, 2, \dots, N$ to the points in accordance with their relative orders. Multiply by 2 and reduce modulo 1; then the GSR shuffle is the random permutation induced by the reordering of the points.

2.3 Riffle Shuffles in General

We shall consider in this paper riffle shuffles in which the deck is cut into a finite set of stacks, labelled by a finite alphabet $\mathcal{A} = \{\lrcorner_\infty, \lrcorner_\epsilon, \dots, \lrcorner_{\mathcal{L}}\}$. The alphabet \mathcal{A} is ordered $a_1 < a_2 < \dots < a_L$; this order determines how the stacks are put back together after the shuffle has been executed a finite number of times — stack a_1 is on top, stack a_2 is just below a_1 , etc. Recombination of stacks is governed by a *routing matrix*, an *irreducible* 0-1 matrix R on $\mathcal{A} \times \mathcal{A}$: on each shuffle, the stacks that are “riffled” to obtain the next stack $b \in \mathcal{A}$ are precisely those stacks $a \in \mathcal{A}$ such that $R(a, b) = 1$. The orbit of any card, i.e., the sequence of stacks that the card visits, is an element of the sequence space

$$\Sigma = \Sigma_R = \{x_1 x_2 \dots \in \mathcal{A}_+^{\mathbb{Z}} : \mathcal{R}(\mathfrak{s}, \mathfrak{s})_{+\infty} = \infty \forall\}$$

The main law governing the riffles is as follows:

Assumption 1 *In each riffle, the relative order of the cards in each stack $a \in \mathcal{A}$ is preserved.*

By the same reasoning as used above for the GSR shuffle, Assumption 1 implies that the order of cards in the deck is the same as the lexicographic ordering of their orbits. Consequently, if the deck starts in the totally ordered configuration, then the state of the deck at any time n is completely determined by the (unordered) set Y_n of orbits

$$Y_n = Y_n^N \triangleq \{\sigma^n x^1, \sigma^n x^2, \dots, \sigma^n x^N\}. \quad (3)$$

We shall call $\{Y_n\}_{n \geq 0}$ the *associated symbolic dynamical system*. The orbit $\{\xi_n\}_{n \geq 0}$ of a randomly selected card will be called the *associated marginal process*. In general, the N sequences in Σ representing the orbits of the N cards need not be independent. For the GSR model, and a number of others, the orbits *are* independent, but as we will show later, this is exceptional.

Let Z_n denote the vector recording the composition of the stacks after n shuffles (the composition of any particular stack is the *ordered* list of cards in the stack). We will henceforth refer to $\{Z_n\}_{n \geq 0}$ as the *shuffling process*. Clearly, $\{Z_n\}_{n \geq 0}$ completely determines $\{Y_n\}_{n \geq 0}$ and hence also the state X_n of the deck at each time $n \geq 0$. (Note that Z_n cannot in general be recovered from Y_n , because Y_0 contains no information about the initial ordering of the deck. The sequence $\{Z_n\}_{n \geq 0}$ can, however, be recovered from (Z_0, Y_0) .)

Assumption 2 *The process $\{Z_n\}_{n \geq 0}$ is an ergodic Markov chain whose law is invariant under relabellings of the cards.*

Here the term “ergodicity” means that the transition probability matrix for the process Z_n is aperiodic and irreducible. Since the state space is finite, it follows that there is a unique invariant probability distribution ν , and that regardless of the law of Z_0 , $\mathcal{D}(Z_n) \rightarrow \nu$ as

$n \rightarrow \infty$. Since the law of the process is invariant under relabellings of the cards, it follows that the invariant distribution ν also has this property; thus, ν is determined by the induced invariant measure ν^* on the vector of stack *sizes*. In section 3.1 below, we will provide an example of a shuffling process (the perfect shuffle) which satisfies all of Assumptions 1 and 2 but the ergodicity requirement.

If $\mathcal{D}(Z_t) = \nu$, then $\{Z_n\}_{n \geq 0}$ and all induced processes (including the associated symbolic dynamical system and the associated marginal process) are stationary and mixing. In fact, since the associated symbolic dynamical system contains no information about the original ordering of the cards in the deck, it will be stationary provided the distribution of Z_0 is such that the distribution of the stack *sizes* agrees with ν^* , e.g., if the deck starts in a completely ordered state, but the initial division into stacks is done according to ν^* . Unless otherwise indicated, assume that this is the case: thus, the associated symbolic dynamical system and the associated marginal process are stationary, and the deck is initially in a totally ordered state.

The associated symbolic dynamical system Y_n induces a random process X_n on the permutation group \mathcal{S}_N , X_n being the permutation describing the state of the deck after n shuffles. In detail, $\{X_n\}_{n \geq 0}$ is determined by $\{Y_n\}_{n \geq 0}$ as follows: $X_0 = \text{identity}$, and for each $n > 0$, $X_n(i)$ is the rank of the n -shifted orbit $\sigma^n x^i$ in the set of all n -shifted orbits $\{\sigma^n x^1, \sigma^n x^2, \dots, \sigma^n x^N\}$, where $Y_0 = \{x^1, x^2, \dots, x^N\}$ and x^1, x^2, \dots are listed in lexicographic order. If Y_n is assumed to be a stationary process, then so is the sequence of “increments” $\chi_1 = X_1, \chi_2 = X_2 X_1^{-1}, \chi_3 = X_3 X_2^{-1}, \dots$. The stochastic process X_n on \mathcal{S}_N is not necessarily a random walk (i.e., the increments χ_1, χ_2, \dots , although stationary, are not necessarily independent). An example will be given in section 3.4 below. Note, however, that the uniform distribution on \mathcal{S}_N is an invariant measure for the process X_n , since the law of $\{Z_n\}_{n \geq 0}$ is invariant under relabellings of the cards. The standing hypothesis that the deck is in a totally ordered state is equivalent to the statement $X_0 = \text{identity}$.

Note that the sequence $\{Z_n\}_{n \geq 0}$ cannot in general be recovered from $\{X_n\}_{n \geq 0}$, because several different stack configurations might represent the same permutation (e.g., in a shuffling process with just 2 stacks, the configurations (a) stack a_2 empty and stack $a_1 = (1, 2, 3, \dots, N)$ and (b) stack a_1 empty and stack $a_2 = (1, 2, 3, \dots, N)$ both represent the identity permutation). In fact, the induced process $\{X_n\}_{n \geq 0}$ does not uniquely determine either the shuffling process $\{Z_n\}_{n \geq 0}$ or the associated symbolic dynamical system $\{Y_n\}_{n \geq 0}$. Consider, for example, the GSR shuffle. The natural shuffling process Z_n associated with the GSR shuffle maintains 2 stacks, a top stack t and a bottom stack b , as described earlier. But a model that maintains 3 stacks, t_t, t_b, b may also be given: this model is derived from the 2-stack model by the device of artificially dividing the top stack t into 2 substacks, t_t and t_b (e.g., by choosing the top $\text{Binomial}(|t|, \frac{1}{2})$ cards of t for substack t_b), but then ignoring this subdivision in the next riffle (thus, before the next riffle, t_t and t_b are recombined in their original order, and the reconstructed stack t is then riffled with b as in the natural model). It is clear that the induced process $\{X_n\}_n$ is not affected. However, the state space for Z_n is different, and the alphabet \mathcal{A} for the associated symbolic dynamical system has 3 letters instead of 2.

2.4 Dynamical Systems on $[0,1]$

The associated symbolic dynamical system has an equivalent description as an N -particle system on the unit interval. (This is *not* the same system as the canonical dynamical system.) This representation is gotten by mapping the sequence space Σ_R onto the unit interval in the natural way:

$$x_1x_2\dots \xrightarrow{\pi} P\{\xi_1\xi_2\dots \leq x_1x_2\dots\}$$

where the implied ordering of sequences is the lexicographic order. The mapping π is clearly measure-preserving and order-preserving. Assumption 2 implies that π is continuous, because ergodicity of the shuffling process guarantees that the distribution of ξ has no atoms. In general, π will not be one-to-one; but the set of “intervals” $[x, x'] \subset \Sigma_R$ that are mapped to points by π is countable, and their union has measure 0, so with probability 1 sequences y from these “bad spots” will not occur as orbits in the associated marginal process. Thus, orbits $x \in \Sigma_R$ are effectively identified with points of $[0,1]$ in a one-to-one manner, and hence the associated symbolic dynamical system may be viewed as a particle system on $[0,1]$. We will not always distinguish between the associated symbolic dynamical system and its π -projection to the unit interval.

It will sometimes (but not always) be the case that the motion of particles in the π -projection of the associated symbolic dynamical system will be via a deterministic map, as for the GSR shuffle (where particles move by the $x \rightarrow 2x \bmod 1$ map). Interesting examples where this is not the case are the Borel and GSR- F shuffles discussed in section 3. If this *is* the case, then the mapping $T : [0, 1] \rightarrow [0, 1]$ must be a measure-preserving mapping; and its restriction to each of the intervals $J(a_i) \stackrel{\Delta}{=} \pi\Sigma_R(a_i)$ must be nondecreasing and continuous. (Here $\Sigma_R(a_i)$ is the set of all sequences in Σ_R with first entry a_i . T is monotone on $J(a_i)$ because σ is monotone on $\Sigma_R(a_i)$, by Assumption 1, and T is (semi-)conjugate to σ .) It follows that T is absolutely continuous and a.e. differentiable. A necessary and sufficient condition that T be measure-preserving is that for a.e. $y \in (0, 1)$,

$$\sum_{x \in T^{-1}(y)} 1/T'(x) = 1. \quad (4)$$

The entropy of the measure-preserving system $([0, 1], T, \text{Lebesgue})$ is given by

$$h = \int_0^1 \log |T'(x)| dx \quad (5)$$

(See, e.g., Mane [1987]).

What makes the associated symbolic dynamical system and the associated marginal process useful constructs is that for many interesting classes of shuffles the marginal process ξ_n^N has a limiting form as the deck size $N \rightarrow \infty$, and this limiting form itself has a simple description as the orbit of a randomly chosen point of $[0, 1]$ under the iterates of a measure-preserving map. The simplest instance of this is the GSR shuffle: in this case the marginal process ξ_n^N has the same law for all N , and the associated symbolic dynamical system Y_n is the join of independent copies of ξ_n^N . In the next section, we shall discuss a number of other examples.

3 Examples

3.1 Perfect Shuffles

A *perfect shuffle* of a deck of $2N$ cards breaks the deck into two stacks of size N , then perfectly interlaces the cards in the top stack with those of the bottom stack. This shuffle is completely nonrandom, so the “motion” of the deck under repeated applications of the shuffle is periodic. The orbits of individual cards may be described as follows. Represent the cards by the $2N$ points $0, \frac{1}{2N-1}, \frac{2}{2N-1}, \dots, 1$, with the natural ordering of the unit interval determining the order of the cards in the deck. Then the perfect shuffle is nothing more than the $2x \bmod 1$ map. The orbit of any particular card i may be deduced from the motion of the corresponding point as follows: the n^{th} entry of the sequence is either 0 or 1, depending on whether the corresponding point has moved into $[0, \frac{1}{2})$ or $[\frac{1}{2}, 1)$ after n applications of the doubling map. This sequence is, of course, nothing other than the binary expansion of the point $\frac{i-1}{N-1}$ initially representing card i . Therefore, the period of the shuffle (the number of perfect shuffles needed to restore the deck to its initial configuration) is precisely the period of the binary expansion of $\frac{1}{2N-1}$. See Diaconis and Graham (1985) for more.

Observe that the associated symbolic dynamical system is an ergodic transformation of a *finite* measure space, hence has entropy 0 and fails to be mixing. It is interesting to note, however, (and this is our main reason for discussing this particular model) that as $N \rightarrow \infty$, the orbit of a randomly chosen card converges in distribution to a Bernoulli- $\frac{1}{2}$ sequence, and in fact the joint distribution of the orbits of any k randomly chosen cards converges to that of k independent Bernoulli- $\frac{1}{2}$ sequences. In this sense the perfect shuffles are “asymptotically indistinguishable” from the GSR shuffles. This should serve notice that the mixing properties of a sequence of riffle shuffles are not determined *solely* by the limiting form of the associated marginal process.

3.2 Borel’s Shuffle

In this shuffle, a packet of cards is randomly selected from the middle of the deck, then moved to the top. It is usually assumed that the packet is selected by choosing two of the “spaces” between cards at random, i.e., as a sample of size 2 from the uniform distribution on $N + 1$ spaces, and letting the cards between these two spaces be the packet. Thus, the shuffling process maintains 3 stacks (top, middle, bottom), with the relative order of each preserved in each repetition of the shuffle. The associated symbolic dynamical system admits an interpretation as a “random interval map” as follows. The cards are represented by the points $0, \frac{1}{N}, \frac{2}{N}, \dots, \frac{N-1}{N}$. Two independent uniform $(0,1)$ r.v.s U_1, U_2 are selected and ordered: $U_{(1)} < U_{(2)}$. The unit interval is then broken at the points $\frac{i}{N}, \frac{j}{N}$ to the immediate left of $U_{(1)}$ and $U_{(2)}$, respectively. The middle stick and the left stick are then interchanged. The orbit of a particular card is just the sequence of 1s, 2s, and 3s recording which of the random intervals $(0, U_{(1)}), (U_{(1)}, U_{(2)}), (U_{(2)}, 1)$ the corresponding point falls into at each time n . (The author is indebted to T. SELLKE for the details of this representation.)

The associated marginal process clearly converges as $N \rightarrow \infty$ to the orbit of a randomly chosen point of the unit interval under the random interval map in which the unit interval is broken at two uniforms and the left and middle sticks are interchanged. Iteration of this procedure results in an ergodic, mixing, but *zero entropy* motion of the unit interval.

Entropy here is *fiber entropy*, i.e., conditional entropy given the σ -algebra generated by the sequence of uniform rvs determining the locations of the breaks. The rationale for this particular entropy will become clear in section 5. The formal definition is as follows. Let \mathcal{P}_n be the partition of the unit interval induced by the first n iterations of the map (thus, there are $2n + 1$ atoms, each a [possibly degenerate] interval). Then

$$h \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} \mathcal{H}(\mathcal{P}_n)$$

where $\mathcal{H}(\mathcal{P}_n) = \sum_{\mathcal{A} \in \mathcal{P}_n} -|\mathcal{A}| \log |\mathcal{A}|$ with $|\cdot|$ denoting Lebesgue measure. This limit is zero because the partition \mathcal{P}_n has only $2n + 1$ atoms. (Unconditionally, there are 2^n possible orbits of length n ; however, conditioning on the sequence of break points limits the set of possible orbits to $2n + 1$.)

It is known (see Diaconis and Saloff-Coste [1993]) that order $N \log N$ Borel shuffles are needed to randomize a deck of N cards. This is consistent with the randomization principle discussed below (section 5), that at least on the order of $\log N/h$ shuffles are needed to randomize a deck of N cards.

3.3 Modified GSR Shuffles

The GSR shuffle may be modified by changing the distribution of the break point from Binomial $(n, \frac{1}{2})$ to Binomial (n, p) or to a mixture of Binomial (n, p) .

The GSR- p Shuffle If the break point distribution is Binomial (n, p) then the associated symbolic dynamical system has a simple structure— it consists of N independent sequences of iid Bernoulli- $q = 1 - p$ random variables. This system has an alternative geometric description. Start with a sample of N iid uniform- $(0, 1)$ random variables, each representing a card; split the unit interval into the subintervals $[0, p]$, $(p, 1]$, and map each onto $[0, 1]$ by a monotone increasing affine transformation. The (random) permutation induced by the reordering of the “cards” is one repetition of the shuffle. The entropy of the measure-preserving transformation T is $h = H(p) = -p \log p - q \log q$ (Shannon’s entropy function).

The GSR-Uniform Shuffle Next consider the modified GSR shuffle where the break point is chosen from the uniform distribution on $\{0, 1, 2, \dots, N\}$. Since the uniform distribution on $\{0, 1, \dots, N\}$ is the mixture (over p) of the Binomial (N, p) distributions with mixing measure uniform on $[0, 1]$, the associated symbolic dynamical system consists of N sequences that may be generated as follows: Choose N iid uniform rvs U_1, U_2, \dots, U_N on $[0, 1]$ —these will represent the individual cards. The N points undergo random transformations determined by an independent sequence V_1, V_2, \dots of iid uniform rvs on $[0, 1]$. At time n the unit interval is broken into the two subintervals $[0, V_n]$ and $[V_n, 1]$, and each of these is mapped linearly onto $[0, 1]$. The orbit of each of the N points U_1, U_2, \dots, U_N determines a 0-1 sequence— the n^{th} entry just records whether the point has moved into the left interval $[0, V_n]$ or the right interval $[V_n, 1]$ at time n .

Although the points of the unit interval representing the cards (in their original positions) are iid uniforms, the 0-1 sequences making up the associated symbolic dynamical system are *not* independent. In fact, it is easily verified that the associated *marginal* process

is unconditionally a Bernoulli- $\frac{1}{2}$ sequence. If the different sequences were independent, then the shuffling process would have the same law as the GSR shuffle, which is obviously not the case. What is true in this case is that *conditional* on the sequence V_1, V_2, \dots the orbits of different cards are iid. The appropriate entropy for this example (again, see section 5) is the conditional entropy given the σ -algebra generated by V_1, V_2, \dots . By formula (6) below, it is given by

$$h = \int_0^1 H(p) dp = \frac{1}{2}.$$

The GSR- F Shuffle The general case, where the distribution of the break point is

$$\int \text{Binomial-}(q) F(dp)$$

for some mixing distribution F , is similar. The associated symbolic dynamical system has the following description. Let R_1, R_2, \dots be a sequence of iid random variables with distribution F . Conditional on the values of R_1, R_2, \dots , generate N independent sequences of Bernoulli random variables, with the n^{th} entry of each sequence having the Bernoulli- R_n distribution. Then the superposition of these N sequences is the associated symbolic dynamical system for the modified GSR shuffle with “parameter” F . Observe that the law of the associated marginal process is the same as that of the modified GSR shuffle with break point distributed as Binomial ($N, \int q F(dp)$). The fiber entropy (conditional entropy given the σ -algebra \mathcal{R} generated by R_1, R_2, \dots) is given by the formula

$$h = \int_0^1 H(p) F(dp), \tag{6}$$

where $H(p)$ is the Shannon entropy function. To see this, note that n -orbits (sequences of 0s and 1s of length n) are in one-to-one and measure-preserving correspondence with the atoms of the partition \mathcal{P}_\setminus of $[0,1]$ defined inductively by (1) splitting $[0,1]$ at R_1 ; and (2) for $n > 1$, splitting each atom of $\mathcal{P}_{\setminus-\infty}$ into intervals of relative lengths R_n and $1 - R_n$. The entropy of the partition \mathcal{P}_\setminus is by definition $\sum_{\mathcal{P}_\setminus} |A| \log |A|$, which by an easy calculation equals $\sum_{i=1}^n H(R_i)$. Dividing by n , letting $n \rightarrow \infty$, and using the SLLN, one obtains the advertised formula for h .

Multistack GSR Shuffles One may also consider generalizations of these models in which more than 2 stacks are maintained. For example, let $\mathbf{p} = (p_1, p_2, \dots, p_L)$ be a probability distribution in the L -simplex with all entries positive. Consider the shuffle in which there are L stacks; on each shuffle, each of the L stacks a_i is broken into L substacks a_{ij} according to the multinomial $\mathcal{M}(\mathbf{p})$ distribution, and then corresponding substacks (those with the same second index j) are riffled to form the next stack j . The riffles are unbiased, i.e., given the substack cardinalities $|a_{ij}|$, all possible (order-preserving) riffles are equally likely. Call this model the “GSR- \mathbf{p} ” shuffle. For this shuffle, the associated symbolic dynamical system is the superposition of N independent multinomial $\mathcal{M}(\mathbf{p})$ sequences. The associated marginal process may be modelled as the orbit of a randomly chosen point under the action of the piecewise linear mapping whose restriction to each of the intervals (t_i, t_{i+1}) is an

increasing linear homeomorphism onto $(0, 1)$, where $t_j = p_1 + p_2 + \dots + p_j$. For future reference (sec. 4 below), note that a shuffling process that maintains L stacks and whose associated symbolic dynamical system is the superposition of N independent multinomial $\mathcal{M}(\mathbf{p})$ sequences is a GSR- \mathbf{p} shuffle.

Formal proofs of the representations are easy to give. The two essential ingredients are: (1) In a random sample of size n from the uniform distribution on $[0,1]$, the distribution of the number of points to the left of q is Binomial (N, q) ; and (2) the superposition of distinguishable independent uniform samples of sizes N_1 and N_2 gives a “marked” sample of size $N_1 + N_2$ in which the distribution of the marks is the same as that obtained in doing an unbiased riffle of two card stacks of sizes N_1 and N_2 .

3.4 The Fibonacci Shuffle

This is the shuffle mentioned in the introduction. The rationale for the term “Fibonacci shuffle” will become apparent shortly. At each time n , there are two stacks t, b . Stack b is broken into top and bottom stacks b_T, b_B by selecting the top k cards of b for b_T , where k has the Binomial $(|b|, p)$ distribution and $|b|$ is the number of cards in b at time n . Then stacks t and b_B are riffled to obtain the next b , while b_T becomes the next t . Assume that the riffle is *unbiased*, i.e., that all riffle permutations of stack t and stack b_B are equally likely. The alphabet \mathcal{A} for this shuffle is $\{0, 1\}$ ($0=t, 1=b$), and the routing matrix is

$$R = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Thus, orbits of cards are restricted to those sequences of 0s and 1s such that every 0 is followed by a 1. (There is a close connection between the matrix R and the sequence of Fibonacci numbers: for instance, the entries of R^{n+2} are the n^{th} , $(n+1)^{\text{st}}$, and $(n+2)^{\text{nd}}$ Fibonacci numbers. Hence the term “Fibonacci shuffle”.)

Proposition 1 *The associated symbolic dynamical system consists of N iid copies of the stationary Markov chain on the state space \mathcal{A} with transition probability matrix*

$$\begin{pmatrix} 0 & 1 \\ p & 1-p \end{pmatrix}. \tag{7}$$

Proof: Consider the superposition of N iid copies of the Markov chain, the realization of each copy an infinite sequence of 0s and 1s. Identify these sequences with cards, with the order of the N sequences determining the labels of the corresponding cards at time 0. The order of the cards at any subsequent time n is the relative order of the n -shifted sequences.

At any time n , the top and bottom stacks will be determined by the n^{th} entries of the sequences representing the cards. For each sequence with a 1 in the n^{th} coordinate, the $(n+1)^{\text{st}}$ coordinate is Bernoulli- q , with the Bernoulli variables independent of each other and conditionally independent of the past, by the Markov property. For each sequence with a 0 in the n^{th} coordinate, the $(n+1)^{\text{st}}$ coordinate is necessarily 1. Thus, the “top” and “bottom” stacks are broken according to the same rules as for the Fibonacci shuffle.

Now consider the “riffle”. The cards that make up the bottom stack at time $n + 1$ may come from either the top or the bottom stack at time n . By the previous paragraph, all cards from the time- n top stack go to the bottom stack at time $n + 1$, while cards from the time- n bottom stack are selected for the time- $(n + 1)$ bottom stack by independent Bernoulli- p variables. All cards chosen for the time- $(n + 1)$ bottom stack have a 1 in the $(n + 1)^{\text{st}}$ coordinate, however. Hence, by the Markov property, the futures of these sequences (after time $n + 1$) are exchangeable. It follows that all possible arrangements of the cards from the time- n top stack and cards from the time- n bottom stack that preserve the relative orders in these two stacks are equally likely. Moreover, the choice of an arrangement is uninfluenced by the histories of the sequences up to time $n - 1$. Thus, the riffles are unbiased. ///

The stationary distribution of the two-state Markov chain with transition probability matrix given in the preceding proposition is $\nu^* = (p(1 + p)^{-1}, (1 + p)^{-1})$, and the entropy is

$$h = \frac{H(p)}{1 + p}$$

where $H(p) = p \log p + q \log q$. The chain may be represented a measure-preserving transformation F of Lebesgue space (the unit interval with the uniform distribution) as follows:

$$\begin{aligned} F(x) &= x/p, & 0 \leq x \leq p/(2 - p); \\ &= x + (1 - p)/(2 - p), & p/(2 - p) \leq x \leq 1/(2 - p); \\ &= (x - 1/(2 - p))(p/(1 - p)), & 1/(2 - p) < x \leq 1. \end{aligned}$$

Observe that F is piecewise linear. The identification of sequences with points is made by examining orbits: the n^{th} coordinate of the sequence is 0 or 1 according as the point is in $[0, (2 - p)^{-1}]$ or $[(2 - p)^{-1}, 1]$ at time n .

Proposition 2 *For the Fibonacci shuffle with $p = \frac{1}{2}$, the induced process X_n on the permutation group \mathcal{S}_N is not a random walk.*

Note: It is assumed here that the shuffling process $\{Z_n\}_{n \geq 0}$ is such that (a) the initial state X_0 of the induced process X_n is $X_0 = \text{identity}$, and (b) the distribution of the stack sizes $|t|, |b|$ at time 0 is

$$\nu^* = \left(\frac{1}{2}\left(1 + \frac{1}{2}\right)^{-1}, \left(1 + \frac{1}{2}\right)^{-1}\right)$$

(the steady-state distribution of stack sizes).

Proof: We shall only prove this for large N .

Consider first the following two procedures for constructing a random permutation in \mathcal{S}_N : (1) Break the deck into a top stack t and a bottom stack b in such a way that the distribution of $|t|$ is F ; then break the bottom stack b at Binomial $(|b|, \frac{1}{2})$ into a top substack b_T and a bottom substack b_B ; put b_T on top; riffle t and b_T and put the resulting stack on the bottom of the deck, below b_T . Call the resulting permutation Π . (2) Same as (1), but in the initial division of the deck, the distribution of $|t|$ is $G \neq F$. Call the resulting

permutation Π' . Assume that F, G are such that $P\{b_T = \emptyset \text{ or } t = \emptyset\} \rightarrow 0$ as $N \rightarrow \infty$. We claim that, at least for large N , the distribution of Π' is different from that of Π . To see this, condition on the event A that the top card of the deck *after* Π (or Π') is not the same as the original top card 1. Since $b_T \neq \emptyset, t \neq \emptyset$ implies that A occurs, $P(A) \rightarrow 1$ as $N \rightarrow \infty$. If A occurs, the “rising sequence” in Π (or Π') beginning with card 1 (i.e., the maximal sequence $1, 2, 3, \dots, k$ of cards that remain in their original relative order after Π or Π') is just the top stack t ; thus, on the event A , the composition of the top stack t is observable from Π (or Π'). (NOTE: In general, the stacks t and b cannot be reconstructed from the permutations Π, Π' .) But the distributions of $|t|$ are different in the constructions of Π, Π' . Since $\mathcal{D}(|\square|)$ and $\mathcal{D}(|\square||A)$ differ only slightly for large N for both constructions (recall that $P(A) \rightarrow 1$), it follows that Π and Π' cannot have the same distribution.

Now consider the permutation π that moves the top card to the bottom of the deck and leaves the relative order of the remaining $N - 1$ cards unchanged (thus, all of the cards below the top card move up 1 notch). Under the Fibonacci model with $p = \frac{1}{2}$, the event $X_1 = \pi$ has positive probability, but occurs *only if* the initial composition $Z_0 = (t, b)$ of the stacks is $t = (1), b = (2, 3, \dots, N)$ (because if the top stack t had more than the top card 1, card 1 could not be moved to the bottom of the deck by X_1). Thus, given that $X_1 = \pi$, the composition $Z_1 = (t', b')$ of the stacks after the first shuffle must be such that the cardinality $|t'|$ of the top stack has the Binomial $(N - 1, \frac{1}{2})$ distribution. But this implies that the conditional distribution of $X_2 X_1^{-1}$ given the event $X_1 = \pi$ is not the same as the (unconditional) distribution of X_1 (by the previous paragraph); hence, the sequence X_n does not have iid “increments”, and therefore is not a random walk. ///

3.5 The General Unbiased Riffle Shuffle

The two preceding examples may be generalized as follows. Maintain a finite set \mathcal{A} of stacks. Break each stack a (from top to bottom) into $|\mathcal{A}|$ substacks of sizes $K_a(b), b \in \mathcal{A}$, where the vector $\{K_a(b)\}_{b \in \mathcal{A}}$ has the multinomial distribution $\mathcal{M}(|a|; \{\mathcal{P}_{\cdot|}\}_{|\in \mathcal{A}})$. Here $|a|$ denotes the number of cards in a , and $(P_{ab})_{a, b \in \mathcal{A}}$ is a stochastic matrix chosen from a mixing distribution G on the space of all stochastic matrices on $\mathcal{A} \times \mathcal{A}$. Once the stacks a have been broken into substacks ab , take all substacks with second index b and riffle them to form the next stack b (these riffles should be unbiased, i.e., all possible order-preserving permutations should be equally likely).

The associated symbolic dynamical system again has a simple description. Conditional on the sequence P_n of stochastic matrices determining the breaks, it is the superposition of N iid time-inhomogeneous Markov chains with time- n transition probability matrix P_n . The fiber entropy h is determined as follows. Let Π_n be a stationary probability vector for the product stochastic matrix $P_1 P_2 \dots P_n$; then by a theorem of Furstenberg (see Bougerol [1985]), as $n \rightarrow \infty$ the sequence Π_n of random vectors converges in distribution to a random vector Π , *provided* the mixing distribution G is sufficiently diffuse. Let P_0 be a copy of P_i independent of Π . Then

$$h = E \left(\sum_{a, b \in \mathcal{A}} \Pi_a P_0(a, b) \log P_0(a, b) \right). \quad (8)$$

In the special case where $P_n \equiv P$ is nonrandom, the Markov chain is time-homogeneous with transition probability matrix P , and the entropy h is just the usual entropy for a Markov chain. In this case the associated marginal process has a representation as a piecewise-linear measure-preserving transformation of Lebesgue space.

3.6 The (u, v) -Weighted Riffle

In this shuffle the deck is divided into two stacks which are then riffled; however, this riffle is biased so that cards from one of the stacks are more likely to drop than cards from the other. Specifically, if at some stage of the riffle the top stack has A cards remaining and the bottom stack has B cards remaining, then the probability that the next card dropped comes from the top stack is $uA/(uA + vB)$. Assume for the sake of discussion that the division of the deck into two stacks is such that the number of cards in the top stack has the Binomial (N, p) distribution.

The associated symbolic dynamical system for this shuffle seems not to have a simple structure—it is not, for instance, the superposition of N iid sequences, nor (apparently) is the law of the associated marginal process independent of N . However,

Proposition 3 *As $N \rightarrow \infty$, the associated symbolic dynamical system converges in distribution to the superposition of the orbits of countably many independent uniform-(0,1) random variables under iteration of the map*

$$G_p(x) = H_p \circ M_{1/v} \circ L\left(\frac{p-x}{p}\right) \quad \text{if } x \in [0, p]; \quad (9)$$

$$G_p(x) = H_p \circ M_{1/u} \circ L\left(\frac{1-x}{1-p}\right) \quad \text{if } x \in (p, 1] \quad (10)$$

where

$$L(y) = -\log y, \quad (11)$$

$$M_\alpha(y) = \alpha y, \quad (11)$$

$$H_\beta(y) = 1 - \beta e^{-vy} - (1 - \beta)e^{-uy}. \quad (12)$$

Note: (1) Denote by Y^∞ the infinite particle system with particles moving deterministically under the action of $G = G_p$ and with particles located at iid uniform-(0,1) random points at time 0. Convergence in law to Y^∞ means that for any $k \geq 1$, the joint distribution of the orbits of k randomly selected cards converges to that of the orbits of k iid uniform-(0,1) random variables under the mapping T .

(2) As usual, the orbit of a point under T determines a 0-1 sequence, the n^{th} coordinate recording whether the point is in $[0, p]$ or $(p, 1]$ at time n . Note that for all but countably many points of $[0, 1]$ (namely, those whose orbits hit one of the points $0, p, 1$), the orbit is *uniquely* determined. (The assignment of orbit to point is the inverse of the natural projection $\pi : \Sigma \rightarrow [0, 1]$ introduced in section 2.4, and π [semi-]conjugates $\sigma : \Sigma \rightarrow \Sigma$ with $T : [0, 1] \rightarrow [0, 1]$.) Consequently, to prove the proposition, it suffices to show that the *canonical* associated dynamical system converges in law to Y^∞ .

As usual, the orbit of a point under this map determines a 0-1 sequence, the n^{th} coordinate recording whether the point is in $[0, p]$ or $(p, 1]$ at time n . Convergence in distribution means that for any $k \geq 1$, the joint distribution of the orbits of k randomly selected cards converges to that of the orbits of k iid uniform-(0,1) random variables under the mapping T . The entropy of the mapping T may be computed from (5) above.

Proof: Here is a useful model of the (u, v) -weighted riffle. Given stack sizes A, B , choose A points from the exponential distribution with mean $1/u$, and mark them $1, 2, \dots, A$ in accordance with their relative order in the sample. Similarly, choose B points from the exponential distribution with mean $1/v$, and mark them $A + 1, A + 2, \dots, A + B$, again in accordance with their relative order. Now superpose the two samples.

Claim: The random permutation that maps i to the number of points in the combined sample to the left of or equal to the point with mark i has the same distribution as the permutation determined by the (u, v) -weighted riffle shuffle.

Proof of Claim:: This is an easy consequence of the memoryless property of the exponential distribution. The distribution of the leftmost point in the combined sample is exponential with mean $1/(Au + Bv)$, and the probability that the leftmost point comes from the sample of A exponential- u^{-1} rvs is $uA/(uA + vB)$. Conditioning on this point and the “stack” (top or bottom) from which it came, and using the memoryless property reduces the problem to the similar problem for stacks of sizes $A - 1$ and B or A and $B - 1$, to which the same argument may be applied. This process may be continued until one of the “stacks” is empty. ///

The proposition follows routinely from this model, using the WLLN and the (weak) Glivenko-Cantelli theorem. In the weighted- (u, v) riffle shuffle, the cardinality $|t|$ of the top stack t has the Binomial- (N, p) distribution; by the WLLN, as $N \rightarrow \infty$, $|t|/N \rightarrow p$. By the Glivenko-Cantelli theorem, the empirical distributions of the two exponential samples converge to the exponential distributions with means u^{-1}, v^{-1} , respectively. Therefore, by an easy calculation, for any $x \in [0, 1]$ bounded away from p the position of the particle representing card $i = [xN]$ in the canonical associated dynamical system *after* the shuffle converges to $G_p(x)$ as $N \rightarrow \infty$. It follows immediately that the *canonical* associated dynamical system converges in law to the process Y^∞ . ///

Remark: Here is a related model for the shuffle. Start with a random sample of N iid uniform-(0,1) rvs, labelled $1, 2, \dots, N$ in accordance with their relative order in the unit interval. Split $[0, 1]$ into $[0, p]$ and $(p, 1]$; this selects Binomial (N, p) for the top stack 0. Map each of these two intervals onto the half-line $(0, \infty)$ by the appropriate logarithmic transformation: $-\log((p - x)/p)$ for $[0, p]$ and $-\log((1 - x)/(1 - p))$ for $(p, 1]$. (Note that these maps are monotone increasing, so the relative order of “cards” in each of the two piles is preserved.) This creates two independent samples of exponential-1 rvs, one corresponding to the top stack, the other to the bottom. Multiply the first by $1/u$ and the second by $1/v$, then superpose the two. This has the effect of “riffling” the two stacks in the desired fashion.

The superposition of the two samples is a random sample of N iid rvs from the p, q mixture of the exponential distributions with means $1/u, 1/v$, respectively. Applying *any* monotone increasing transformation will preserve the relative ordering of the sample, resulting in a (u, v) -weighted riffle. The appropriate transformation for our purposes maps the halfline $[0, \infty)$ monotonically onto the unit interval $[0, 1]$ in such a way that *conditional on the riffle* the resulting set of points in $[0, 1]$ has the same law as a sample of N iid uniform- $[0, 1]$ rvs. Repetition of the whole procedure will then result in *independent* (u, v) -weighted riffle shuffles. The monotone transformation $T : [0, \infty) \rightarrow [0, 1]$ with the advertised property is easily described. Condition on a given riffle, e.g., 01011. Then the sample of $N = 5$ points constitutes a point process in $[0, \infty)$ with (predictable) intensity λ_{01011} (see, e.g., Bremaud (197) for the definition). The transformation $T = T_{01011}$ is that which maps this intensity onto the intensity for a uniform sample of size N on the unit interval.

Although the transformation T depends in general on the riffle, for large N it converges in probability to the nonrandom transformation H_p . (This follows from the weak law of large numbers.) This proves again that the orbit of a randomly chosen card will have distribution close to that of a randomly chosen point of the unit interval under the mapping G_p . It also shows that the associated symbolic dynamical system is tantalizingly close to the superposition of N independent copies of the limiting marginal process ($([0, 1], G_p, \text{Lebesgue})$). But not *exactly*—the conditional intensities given the riffles are not identical with the unconditional intensities.

3.7 f -Shuffles

Given a measure-preserving, expanding map f of the unit interval onto itself, one may construct (a sequence of) riffle shuffles for which the associated marginal processes converge (as the deck size $N \rightarrow \infty$) to f . The models introduced above for the (u, v) -weighted riffle shuffles suggest how this may be done. Rather than attempt to formulate a general theorem to this effect, we shall limit our attention to 2-1 mappings f with the following properties: (1) There exists $p \in (0, 1)$ such that the restriction of f to either of the intervals $[0, p)$ or $[p, 1)$ is an increasing, C^2 homeomorphism onto $[0, 1)$ with bounded derivatives; and (2) For every $y \in (0, 1)$, $f'(x_1)^{-1} + f'(x_2)^{-1} = 1$, where $x_1 = x - 1(y)$ and $x_2 = x_2(y)$ are the two points of $(0, 1)$ that T maps to y (this guarantees that f is measure-preserving).

The random permutation determining the shuffle is constructed as follows. Drop N points at random into the unit interval (according to the uniform distribution) and mark the points $1, 2, \dots, N$ in accordance with their order in $[0, 1]$. Then apply the transformation f to each of the N points—this results in a new ordering. The random permutation π determining the “shuffle” is the permutation defined by the new ordering of the points after the transformation f is applied (thus, $\pi(i)$ is the relative position in the post- f sample of the point with mark i , i.e., the number of points in the post- f sample to the left of or equal to the point with mark i).

Proposition 4 *As the deck size $N \rightarrow \infty$, the associated marginal process converges in distribution to the orbit of a randomly chosen point of the unit interval under iterates of the transformation f .*

The proof is similar to that given for the similar result concerning the (u, v) -weighted riffle shuffle, using the WLLN and the weak form of the Glivenko-Cantelli Theorem, but is even simpler, and is therefore omitted.

Several points are worth noting in this construction. (1) The random permutation π has either one or two rising sequences: the points in $(0, p)$ (the “top stack”) have their relative order preserved by f , as do also the points in $(p, 1)$. Moreover, the size of the top stack has the Binomial (N, p) distribution. (2) The pre- f sample consists of N iid uniformly distributed random variables, and so does the post- f sample, since f is a measure-preserving map. However, *conditional on* π , the post- f sample is *not* an iid uniform sample unless f is linear on each of the intervals $(0, p)$ and $(p, 1)$ — see section 4 below.

The entropy h of the measure-preserving transformation f is given by (5) above, with $T = f$.

4 Independent Superposition Shuffles

We have discussed several models for which the associated symbolic dynamical system is the superposition of N independent copies of the associated marginal process. Diaconis (1988), discussing the dynamical model for the GSR shuffle, suggests that it might be possible to “take other measure-preserving systems . . . and convert them to other shuffles”. The next result suggests, however, that the possibilities are quite limited, if one insists that the orbits of cards are independent *and* that the induced process $\{X_n\}_{n \geq 0}$ on the permutation group is a random walk. For simplicity, we will consider only riffle shuffles for which the routing matrix R has all entries 1, and such that all possible transitions have positive probability. More precisely,

Hypothesis 1 *For every finite sequence $x_1 x_2 \dots x_n$ with entries in the stack alphabet \mathcal{A} ,*

$$P\{\xi : \xi_1 \xi_2 \dots \xi_n = x_1 x_2 \dots x_n\} > 0.$$

Recall that ξ is the associated marginal process.

Proposition 5 *Let $\{X_n\}_{n \geq 0}$ be an ergodic random walk on \mathcal{S}_N induced by a riffle shuffle satisfying Hypothesis 1, with $N \geq L$. Assume that the associated symbolic dynamical system is the superposition of N independent copies of the orbit of a randomly selected point of the unit interval under iterates of a measure-preserving map $T : [0, 1] \rightarrow [0, 1]$. Then T is piecewise linear; more precisely, there exist $0 = t_0 < t_1 < t_2 < \dots < t_{L-1} < t_L = 1$ such that the restriction of T to each of the intervals (t_{i-1}, t_i) is an increasing, linear homeomorphism onto the unit interval $(0, 1)$.*

Remark: This implies that the shuffle is a GSR- \mathbf{p} shuffle, where $p_i = t_i - t_{i-1}$.

Proof: (1) First we will show that under Hypothesis 1, there exist $0 = t_0 < t_1 < t_2 < \dots < t_{L-1} < t_L = 1$ such that the restriction of T to any of the intervals $J_i = (t_{i-1}, t_i)$ is an increasing homeomorphism onto the unit interval $(0, 1)$. The hypothesis that the associated symbolic dynamical system is the superposition of N independent copies of T implies that

the associated marginal process is just T ; more precisely, there is an order-preserving, measure-preserving mapping $\alpha : \Sigma \rightarrow [0, 1]$ such that T is (semi-)conjugate to the shift by α , i.e., $T \circ \alpha = \alpha \circ \sigma$. Consequently, there exist $0 = t_0 \leq t_1 \leq t_2 \leq \dots \leq t_{L-1} \leq t_L = 1$ such that for each i the interval $[t_{i-1}, t_i]$ is the image of the set $\Sigma(a_i)$ of sequences with first entry a_i . (NOTE: The endpoints are included because $\Sigma(a_i)$ contains minimal and maximal points.) Hypothesis 1 implies that each of these intervals has *positive* measure. Since both α and $\sigma|_{\Sigma(a_i)}$ are order-preserving, $T|_{J_i}$ is nondecreasing. Moreover, since $\sigma|_{\Sigma(a_i)}$ is a homeomorphism onto Σ , and since by Hypothesis 1 all cylinder sets have positive measure, $T|_{J_i}$ must also be a *continuous* mapping of J_i onto the whole unit interval $(0, 1)$. (If T were not continuous, there would be a jump discontinuity in one of the intervals J_i , which by the conjugacy $T \circ \alpha = \alpha \circ \sigma$ would imply that the image of $\sigma|_{\Sigma(a_i)}$ omits a cylinder set, namely one whose α -image is contained in the interval excluded by $T|_{J_i}$.) Finally, since T is measure-preserving, its restriction to each of the intervals J_i must be *strictly* increasing (otherwise, there would be a point $y \in [0, 1]$ whose inverse image $T^{-1}(y)$ would contain an interval of positive length, contradicting the assumption that T is measure-preserving).

NOTE: Since the restriction of T to each of the intervals is an increasing homeomorphism, T is a.e. differentiable.

(2) Next, let $\{\xi^1, \xi^2, \dots, \xi^N\}$ be the orbits of the N cards, and let $\{\zeta^1, \zeta^2, \dots, \zeta^N\}$ be the corresponding points of the unit interval. By hypothesis, $\{\zeta^1, \zeta^2, \dots, \zeta^N\}$ is a sample (unordered) from the uniform distribution. We claim that for each $k \geq 1$,

$$\{T^k \zeta^1, T^k \zeta^2, \dots, T^k \zeta^N\}$$

is, *conditional on* X_1, X_2, \dots, X_k , distributed as an iid sample from the uniform distribution. By hypothesis, $\{X_n\}_{n \geq 0}$ is a random walk on the group \mathcal{S}_N ; this implies that, *conditional on* X_1, X_2, \dots, X_k , the “future” $X_{k+1}X_k^{-1}, X_{k+2}X_k^{-1}, \dots$ is an independent replica of the process X_1, X_2, \dots . It follows that the post- k future Y_k of the associated symbolic dynamical system is, conditional on X_1, X_2, \dots, X_k , a replica of Y_0 .

To see this, observe that one may construct a version of the process $\{X_n\}_{n \geq 0}$ as follows. Construct a version $\{Z'_n\}_{n \geq 0}$ of $\{Z_n\}_{n \geq 0}$ and an independent copy of $\{\tilde{X}_n\}_{0 \leq n \leq k}$, with $\tilde{X}_0 = \text{identity}$ and Z'_0 such that (a) the induced order of the deck is the identity, and (b) the distribution of the set of initial stack sizes is ν^* . Each variable Z'_n is an assignment of ordered sets of cards to the members of the stack alphabet \mathcal{A} . Create a new sequence $\{Z''_n\}_{n \geq 0}$ by relabelling all the cards in $\{Z'_n\}_{n \geq 0}$ according to the permutation \tilde{X}_k (i.e., conjugate with \tilde{X}_k). Let $\tilde{X}_k, \tilde{X}_{k+1}, \dots$ be the permutations induced by the stack compositions Z''_0, Z''_1, \dots . That the whole sequence $\tilde{X}_n, n \geq 0$ we have just defined is a version of the original process $X_n, n \geq 0$ follows from the hypothesis that X_n is a random walk on \mathcal{S}_N , together with Assumption 2. Now the post- k future Y_k of the associated symbolic dynamical system is a function only of $\{Z'_n\}_{n \geq 0}$ (the ordering of cards in the deck after k shuffles does not affect Y_k , since Y_k is the *unordered* set of orbits); consequently, by construction, Y_k is independent of $\{\tilde{X}_n\}_{0 \leq n \leq k}$. Finally, $\{T^k \zeta^1, T^k \zeta^2, \dots, T^k \zeta^N\}$ is a function of Y_k , so it too is independent of $\{\tilde{X}_n\}_{0 \leq n \leq k}$, so our claim is proved.

NOTE: It may not be clear why the hypothesis that X_n is a random walk is needed in the above construction, since Assumption 2 states that the future $Z_n, n \geq k$ is conditionally independent of the past $Z_n, n < k$ given the present Z_k . The subtlety is that $Z_n, n \geq k$ need

not be conditionally independent of X_n , $n < k$ given X_k . Moreover, one cannot construct a version of $\{Z_n\}$ by welding a copy $\{Z'_n\}$ onto an initial segment $\{Z_n\}_{0 \leq n \leq k}$, because the stack cardinalities in Z_k and Z'_0 may not be the same; one can only hope to weld a copy $\{Z'_n\}$ of $\{Z_n\}$ onto an initial segment of the X_n process, and this requires the random walk assumption.

(3) Consider the permutation π satisfying $\pi(i) = i + L - 1$ for $1 \leq i \leq N - L + 1$ and $\pi(N - i) = i + 1$ for $0 \leq i \leq L - 2$. This describes the shuffle in which the top $N - L + 1$ cards are moved to the bottom of the deck, and the bottom $L - 1$ cards are arranged in reverse order at the top of the deck. In order that $X_1 = \pi$, the points $\zeta^1, \zeta^2, \dots, \zeta^{N-L+1}$ must be arranged in increasing order in the interval J_1 ; the points $\zeta^{N-L+2}, \dots, \zeta^N$ must be arranged in order in J_2, J_3, \dots, J_L , one to an interval; and the arrangement must be such that

$$T(\zeta^N) < T(\zeta^{N-1}) < \dots < T(\zeta^{N-L+2}) < T(\zeta^1) < T(\zeta^2) < \dots < T(\zeta^{N-L+1}).$$

Conversely, if these conditions on the arrangement of $\zeta^1, \zeta^2, \dots, \zeta^N$ are met, then $X_1 = \pi$. Thus, conditioning on the event $X_1 = \pi$ is equivalent to conditioning on the particular arrangement of points in the unit interval just described: call this event A .

NOTE: This is the part of the argument where the hypothesis that $N \geq L$ is used: $N \geq L$ guarantees that each of the L intervals J_i has at least one point.

Set $z^i = T\zeta^i$, $i = 1, 2, \dots, N$. By (2) above, *conditional on A* the distribution of $\{z^1, z^2, \dots, z^N\}$ is that of a uniform sample. But the conditional distribution of the set $\{\zeta^1, \zeta^2, \dots, \zeta^N\}$ given the event A is uniform on the subset of $(0, 1)^N$ consisting of points satisfying the restrictions described above, and the mapping $T \times T \times \dots \times T$ is one-to-one on this set, so the Jacobian transformation formula implies that the joint density of z^1, z^2, \dots, z^N given A is

$$\prod_{i=1}^N T'(\zeta^i).$$

This must be constant, since the conditional joint distribution is uniform. Clearly, it follows that T' must be constant on each of the intervals J_i . ///

5 Fiber Entropy

Henceforth, we will only consider shuffles for which the associated symbolic dynamical system Y_n^N is mixing, as we are primarily interested in convergence to uniformity. Because we are interested in asymptotic theory as the deck size $N \rightarrow \infty$, we must consider a sequence of shuffling processes Z_n^N , one on each permutation group \mathcal{S}_N . Let Y_n^N be the associated symbolic dynamical system: thus, Y_n^N consists of N sequences of stationary 0-1 sequences, labelled $\xi_*^1, \xi_*^2, \dots, \xi_*^N$, and these N component processes are exchangeable. A minimal precondition for an asymptotic theory is

Assumption 3 *The sequence Y_*^N converges in distribution as $N \rightarrow \infty$ to a process Y_*^∞ such that (i) Y_n^∞ is stationary and mixing; and (ii) the component sequences in Y_*^∞ are exchangeable.*

The implied topology is that of coordinatewise convergence. Thus, we are assuming that $\forall k, n < \infty$ the joint distribution of the rvs ξ_j^i , $1 \leq i \leq k$ and $0 \leq j \leq n$, converges as $N \rightarrow \infty$ to the joint distribution of the corresponding rvs for the process Y_*^∞ .

This assumption is not *sufficient* to guarantee a satisfactory asymptotic theory: for instance, the associated symbolic dynamical system for the perfect shuffle on a deck of size N converges in distribution as $N \rightarrow \infty$ to the same limiting process as the GSR shuffle. It is easy to modify the perfect shuffles (e.g., allowing an “imperfect” shuffle with probability $a(N)$, where $a(N) \rightarrow 0$ extremely rapidly) so that they become mixing, but nevertheless have an arbitrarily slow rate of convergence to uniformity.

The limiting process Y_n^∞ consists of an exchangeable sequence of stationary \mathcal{A} -valued sequences, with $Y_n^\infty = \sigma^n Y_0^\infty$. By DeFinetti’s theorem, these are conditionally iid given the σ -algebra \mathcal{U} of exchangeable events.

Lemma 1 *Let μ_N be the empirical distribution of the N orbits $\xi_*^1, \xi_*^2, \dots, \xi_*^N$, considered as a random measure on the space of all \mathcal{A} -valued sequences. As $N \rightarrow \infty$,*

$$\mathcal{D}(\mu_N) \longrightarrow \mathcal{D}(\xi_* | \mathcal{U}).$$

Proof: By Assumption 3, the joint distribution of the first k sequences of Y_*^N converges as $N \rightarrow \infty$ to that of the first k sequences of Y_*^∞ . But the distribution of the first k sequences of Y_*^N is the same as the distribution of *any* set of k sequences chosen without replacement from Y_*^N , by exchangeability. As $N \rightarrow \infty$, the difference between sampling without replacement and sampling with replacement disappears; hence, the k -fold tensor product of μ_N converges in distribution to the joint law of the first k sequences of Y_*^∞ . In particular, for any k continuous real-valued functions g_i on sequence space Σ_R ,

$$\begin{aligned} E \left(\prod_{i=1}^k \int g_i d\mu_N \right) &\longrightarrow E \left(\prod_{i=1}^k g_i(\xi_*^{(i)}) \right) \\ &= E \left(E \left(\prod_{i=1}^k g_i(\xi_*^{(i)}) \mid \mathcal{U} \right) \right) \\ &= E \left(\prod_{i=1}^k E(g_i(\xi_*) \mid \mathcal{U}) \right), \end{aligned}$$

the second equality by DeFinetti’s theorem. The result is easily deduced from this by the method of moments. ///

Define the *fiber entropy* (of the limiting marginal process ξ_*) by

$$h = \text{entropy}(\xi_* | \mathcal{U}), \tag{13}$$

i.e.,

$$h = \lim_{n \rightarrow \infty} \frac{1}{n} \mathcal{H}(\mathcal{P}_\setminus | \mathcal{U}) \tag{14}$$

where \mathcal{P}_\setminus is the partition of sequence space $\mathcal{A}_+^{\mathbb{Z}}$ determined by the first n coordinate variables. That h is well-defined and constant follows from the assumption that Y_n^∞ is mixing

(and therefore ergodic). See, e.g., Walters (1982) for a comprehensive discussion of conditional entropy. It is easily verified that this definition of h is consistent with each of the entropies introduced in the examples of section 3. (Note that for the perfect shuffles the entropy of the associated marginal process is zero for each finite deck size N , but that the limiting process consists of a superposition of independent Bernoulli- $\frac{1}{2}$ sequences, so that $h = \log 2$.)

The main result of this section is

Theorem 1 *If the fiber entropy h of the limiting process is positive then for every $\varepsilon > 0$,*

$$\lim_{N \rightarrow \infty} d_N \left((1 - \varepsilon) \frac{\log N}{h} \right) = 1. \quad (15)$$

If the fiber entropy of the limiting process is 0, then for any $C > 0$,

$$\lim_{N \rightarrow \infty} d_N(C \log N) = 1. \quad (16)$$

The proof depends on lemmas which show that fiber entropy is the quantity that controls the rate at which the orbits of distinct cards “separate”. Say that (the orbits of) two cards are *separated* by time n if at some time before time $n + 1$ the two cards are in different stacks, i.e., if the sequences describing their orbits differ in a coordinate between the 0th and the n th. Say that a card is *isolated* by time n if it is separated from every other card by time n , and that it is *k -clumped* if there are at least $k - 1$ other cards from which it is *not* separated. For any $\varepsilon > 0$ and any integers $N, n > 1$, let $A_{N,n}^\varepsilon$ be the event that at least $N(1 - 2\varepsilon)$ cards are k -clumped at time n .

Proposition 6 *Suppose that the fiber entropy h of the limiting process is positive. Then for each $k = 2, 3, 4, \dots$ and any $\varepsilon > 0$, if $n = \lceil (1 - \varepsilon) \frac{\log N}{h} \rceil$, then*

$$\liminf_{N \rightarrow \infty} P(A_{N,n}^\varepsilon) \geq 1 - \varepsilon. \quad (17)$$

Proof: (1) Here is a simple heuristic argument for the case where \mathcal{U} is 0-1, so that h is just the ordinary (unconditional) entropy of the limiting marginal process ξ_n . Entropy determines the growth of the number of high probability orbits—specifically, for each $\varepsilon > 0$ there exist $m \geq 1$ and a set \mathcal{O} of m -orbits (sequences of length m with values in \mathcal{A}) of probability at least $1 - \varepsilon$ with cardinality on the order of e^{hm} . But if $m \leq (1 - \varepsilon) \log N/h$ then $e^{hm} \leq N^{1-\varepsilon}$, so there are not enough available orbits to accommodate N different cards without “crowding”.

This is not quite rigorous even in the case where \mathcal{U} is 0-1, because the orbits of individual cards in a deck of finite size N are *not* assumed to have exactly the same distribution as the component sequences of the limit process Y_n^∞ , but only to converge in distribution. Thus, we are not justified in letting $m \rightarrow \infty$ as $N \rightarrow \infty$.

(2) We now give a rigorous argument for the special case where the limiting process Y_*^∞ is such that the component sequences are independent, i.e., the σ -algebra \mathcal{U} is 0-1. In this case h is just the entropy of the limiting marginal process ξ_n . Consequently, for any $\delta > 0$ there is an integer $m \geq 1$ and a set \mathcal{O} of m -orbits such that (i) the probability that

$\xi_1 \xi_2 \dots \xi_m \in \mathcal{O}$ is greater than $1 - \delta^2$; and (ii) the cardinality of \mathcal{O} is less than $\exp((h + \delta)m)$. By Assumption 3, for all sufficiently large N the orbit $\xi_1^i \xi_2^i \dots$ of a randomly chosen card satisfies $P\{\xi_1 \xi_2 \dots \xi_m \in \mathcal{O}\} \geq \infty - \delta^\epsilon$. Fix the integer m and the set \mathcal{O} .

Let the deck size satisfy $N > e^{hm}$ and set $n = \lceil (1 - \epsilon) \log N/h \rceil$. The n -orbit of any card may be broken into n/m blocks of length m . Call a block *good* if it is an element of \mathcal{O} and *bad* otherwise. By (i) the probability that any given block of ξ_*^1 is bad is less than δ^2 , provided the deck size N is sufficiently large; hence, by the Markov inequality and stationarity, the probability that the fraction of bad blocks among the first n/m blocks exceeds δ is no greater than δ . Thus, with probability approaching 1 as $N \rightarrow \infty$, the proportion of the N orbits included in Y_*^N for which the first n entries are such that the fraction of bad blocks does not exceed δ is less than δ , by Lemma 1. The number of such n -orbits may be estimated as follows: For each good block, there are at most $\exp((h + \delta)m)$ possibilities; for each bad block, there are at most $|\mathcal{A}|^\uparrow$ possibilities; and the number of ways to intersperse bad blocks and good blocks is at most on the order of $\exp(nH(\delta))$, where $H(\cdot)$ is the Shannon entropy function. Hence, the total number of possibilities is at most on the order of $\exp((h + \delta)n) |\mathcal{A}|^\uparrow \exp(nH(\delta))$. By choosing $\delta < \epsilon$ small, this can be made smaller than $N^{1-\epsilon}$. Thus, the number of cards that are not k -clumped is at most $N\epsilon + kN^{1-\epsilon}$.

(3) Finally, consider the general case. By definition of the fiber entropy h , for any $\delta > 0$ there exists $m \geq 1$ such that with probability at least $(1 - \delta)$ there is a (possibly random, but \mathcal{U} -measurable) set \mathcal{O} of m -orbits satisfying (i) the conditional probability given \mathcal{U} that $\xi_1 \xi_2 \dots \xi_m \in \mathcal{O}$ is at least $(1 - \delta^2)$; and (ii) the cardinality of \mathcal{O} is less than $\exp((1 - \delta)hm)$.

Now consider the processes Y_*^N . By Assumption 3 the empirical distribution of the component sequences $\xi_*^1, \xi_*^2, \dots, \xi_*^N$ converges in law to the random measure $\mathcal{D}(\xi_* | \mathcal{U})$ where ξ_* is the first component sequence of the limiting process Y_*^∞ . Moreover, the same is true for the empirical distribution of the component sequences of the km -shifted processes $\sigma^{km} Y_*^N$, by stationarity. Consequently, for each sufficiently large N there exists a sequence $\mathcal{O}_\infty, \mathcal{O}_\epsilon, \dots$ of random subsets of \mathcal{A}^\uparrow such that (i) for every $k \geq 0$ the proportion of the km -shifted component sequences $\sigma^{km} \xi_*^1, \sigma^{km} \xi_*^2, \dots, \sigma^{km} \xi_*^N$ lying in \mathcal{O}_\parallel is at least $(1 - \delta^2)$, and (ii) the limiting proportion of those members of the sequence $\mathcal{O}_\infty, \mathcal{O}_\epsilon, \dots$ with cardinality $\geq \exp((1 - \delta)hm)$ is less than δ . (NOTE: To obtain (i) we have used the ergodicity of the associated symbolic dynamical system Y_n^N .)

Fix the deck size $N > e^{Cm}$ and set $n = \lceil C \log N \rceil$, as above. For each card in the deck, break the orbit $\xi_1^i \xi_2^i \dots \xi_n^i$ into n/m blocks of length m ; for card i call the j^{th} block *good* if that block lies in \mathcal{O}_\parallel and *bad* otherwise. Call a card *bad* if its fraction of bad blocks exceeds δ . Then by (i) of the preceding paragraph and the Markov inequality, there are no more than $N\delta$ bad cards in the deck. Those cards which are not bad have restricted orbits: the fraction of bad blocks is less than δ . The number of such orbits may be estimated exactly in the same manner as in the proof of the special case considered earlier—specifically, it is no more than order $e^{nH(\delta)} \exp((1 - \delta)hn) |\mathcal{A}|^\uparrow$. The result now follows as before by taking δ sufficiently small and using the pigeonhole principle. ///

The same argument as used in the proof of Proposition 6 gives the following:

Proposition 7 *Suppose that the fiber entropy of the limiting process is 0. Then for each*

$C > 0$, $k = 2, 3, 4, \dots$ and any $\varepsilon > 0$,

$$\liminf_{N \rightarrow \infty} P\{\text{at least } N(1 - 2\varepsilon) \text{ cards are } k\text{-clumped at time } n = \lceil C \log N \rceil\} \geq 1 - \varepsilon. \quad (18)$$

Proof of Theorem 1: Consider first the case where $h > 0$. By the proposition 6, with probability in excess of $1 - \varepsilon$ at least 99% of the cards are 100-clumped at time $n = \lceil (1 - \varepsilon) \log N/h \rceil$. Now the cards in any k -clump retain their original order in the deck. Consequently, at least 95% of the adjacent pairs of cards in the deck are in their original order at time n . But in a random permutation of a large deck N , the fraction of adjacent pairs that retain their original order is $\approx \frac{1}{2}$.

A similar argument applies in the case $h = 0$. ///

There is another well known entropy bound for the speed of convergence to uniformity by random walks on finite groups. If μ is a probability distribution on a finite group \mathcal{G} , then its entropy $H(\mu)$ is defined by $H(\mu) = \sum_{x \in \mathcal{G}} -\mu(x) \log \mu(x)$. The entropy of the n^{th} convolution power μ^n is bounded above by $nH(\mu)$. Since the entropy of the uniform distribution on the symmetric group \mathcal{S}_N is about $N \log N$, it follows that for a random walk whose “increments” have distribution μ , at least on the order of $(N \log N)/H(\mu)$ steps are needed in order that the total variation distance to the uniform distribution be less than $1 - \varepsilon$.

It is interesting to compare this bound to that provided by Theorem 1. For certain of the shuffles we have described in section 2, the two entropies are in fact the same, e.g., the GSR shuffle and the modified GSR shuffle with breaks chosen from the Binomial- (N, p) distribution. It should be noted, however, that even in cases where this is true, it may be quite difficult to show, as the case of the modified GSR shuffle makes clear. In others, for example, the (u, v) -weighted riffle shuffle, it appears to be a formidable task to calculate the entropy of the distribution μ of the “increments” at all, but the entropy of the limiting marginal process is easy to obtain. Thus, Theorem 1 often provides a bound much easier to calculate than the bound described in the preceding paragraph. Finally, there are some interesting shuffles that are *not* random walks on \mathcal{S}_N , for instance, the Fibonacci shuffle, for which the associated marginal process is a Markov chain. For these shuffles, Theorem 1 gives bounds on the rate of convergence where none can be gotten by considering convolutions.

Acknowledgment This research had its genesis in a number of enlightening discussions with PERSI DIACONIS at the Institute for Mathematics and its Applications in the autumn of 1993. The idea (or hope) of representing shuffles other than the GSR shuffle by dynamical models is due to him.

References

- [1] Aldous, D.(1983). Random Walk on finite groups and rapidly mixing Markov chains. *Sem. de Prob. XVII. Lecture Notes in Math.* **986** 243-297. Springer, NY.
- [2] Bougerol, P. (1985). *Products of Random Matrices with Applications to Schrodinger Operators*. Birkhauser, Boston.

- [3] Diaconis, P. (1988). *Group Representations in Probability and Statistics*. IMS, Hayward CA.
- [4] Bayer, D. and Diaconis, P. (1992) Trailing the dovetail shuffle to its lair. *Annals of Appl. Prob.* **2** 294-313.
- [5] Diaconis, P. and Graham, R. (1985). The mathematics of perfect shuffles.
- [6] Diaconis, P. and Saloff-Coste, L. (1993) Comparison techniques for random walks on finite groups. Preprint.
- [7] Mane, R. (1987) *Ergodic Theory and Differentiable Dynamics*. Springer, NY.
- [8] Reeds, J. (1981) Unpublished manuscript.
- [9] Walters, P. (1982) *An Introduction to Ergodic Theory*. Springer, NY.