Information loss in card shuffling

Dudley Stark, A. Ganesh, Neil O'Connell
BRIMS, Hewlett-Packard Laboratories
Filton Road, Stoke Gifford
BS34 8QZ, UK

August 31, 1999

Abstract

We find limits for the relative entropy (to stationarity) of a commonly used model for riffle-shuffling a deck of n cards m times. We establish the somewhat surprising fact, which was predicted by Trefethen and Trefethen in a recent numerical study, that for $m < \log_2 n$ the relative entropy decreases linearly and for $m > \log_2 n$ it decreases geometrically. Thus there is a kind of secondary phase transition, which may be to some extent general in nature. The deck becomes random in relative entropy after $m = \frac{3}{2}\log_2 n$ shuffles.

1 Introduction

The riffle-shuffle model of Gilbert, Shannon and Reeds [5] is a random walk on the symmetric group S_n. Statistical experiments presented in [3] show that it is a realistic description of the way people really shuffle cards. Aldous [1] and Bayer and Diaconis [2] have shown that, under the GSR model, it takes $m = \frac{3}{2}\log_2 n$ shuffles to randimise a deck of n cards. This is to say, they show that this is the point at which there is a sharp cutoff (from 1 to 0) in the total variation distance between the law of the deck configuration and the uniform distribution, assuming n is large. Bayer and Diaconis obtain more precise asymptotics in the region of the cutoff. The use of total variation distance to measure rates of convergence to equilibrium for Markov chains (and random walks on groups) is largely a matter of convention, although often bounds are obtained on the l_2 distance which in turn provide bounds on the total variation distance. The relative entropy (which typically lies somewhere in between the two) has not been considered to the same extent, particularly in the context of cutoff phenomena. There is a chapter in the thesis of Su [6] which is devoted to this topic, where some elementary inequalities relating the relative entropy to total variation distance and l_2 distance, as well as detailed asymptotics for the random walk on a hypercube, are presented. Similar inequalities are discussed by Diaconis and Saloff-Coste [4].

In a recent numerical study, Trefethen and Trefethen [7] observed that for the GSR model:

(a) the threshold at $m = \frac{3}{2}\log_2 n$ shuffles is also observed in relative entropy;

(b) after exactly $m = \frac{3}{2}\log_2 n$ shuffles, 0.0601 bits of information remain;

(c) the relative entropy decays *linearly* for $m < \log_2 n$, and *geometrically* for $m > \log_2 n$.

The purpose of this note is to make these observations precise, by careful analysis of the GSR model.

The observation (a) is not surprising, and can be expected to hold quite generally. However, the observation (c) is rather puzzling and it is not clear in what sense it might be a general phenomena. Consider, for example, the (continuous time) random walk on a hypercube, where the relative entropy between the position of the walker after time t and the uniform distribution on an n-dimensional cube is given by

$$\frac{n}{2}\left[(1-e^{-2t/n})\log_2(1-e^{-2t/n}) + (1+e^{-2t/n})\log_2(1+e^{-2t/n})\right].$$

For this model, there is a sharp threshold in total variation distance at $t = \frac{1}{4}n\ln n$, and the same threshold is observed in relative entropy. There is also a similar kind of secondary phase transition at $t = n$: for $t << n$, the relative entropy is

$$n + t\log_2\frac{2t}{n} + O(t)$$

and for $t >> n$, the relative entropy is

$$\frac{n}{2}e^{-4t/n} + O\left(ne^{-6t/n}\right).$$

The rate of decay before the transition is not linear, but we do see a sudden switch to geometric decay. If one were to speculate on what might be a general rule, perhaps it would be that there is a kind of 'pre-threshold', at some point earlier than the absolute cutoff, at which point the relative entropy begins to decay geometrically.

In the next two sections we state and prove the main results of the paper, on the relative entropy to equilibrium in the GSR model.

2 The main results

Let m be the number of times a deck of n cards is shuffled under the GSR model and let D_i, $i = 1, 2, \ldots, n$ be i.i.d. variables taking values uniformly at random on the the set $\{0, 1, 2, \ldots, 2^m - 1\}$. The D_i determine positions to which cards in various 'packets' are taken.

Let us illustrate the GSR model with an example. Suppose $n = 8$, $m = 2$ and the values of the D_i are $D_1 = 3$, $D_2 = 1$, $D_3 = 0$, $D_4 = 0$, $D_5 = 2$, $D_6 = 3$, $D_7 = 0$, $D_8 = 2$. The end configuration is represented by $(3, 1, 0, 0, 2, 3, 0, 2)$ and the initial configuration is represented by $(0, 0, 0, 1, 2, 2, 3, 3)$. The ith occurrence of a given value in the initial configuration is mapped to the ith occurrence of that value in the end configuration. The corresponding permutation is

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 4 & 7 & 2 & 5 & 8 & 1 & 6 \end{pmatrix}.$$

In this way, the D_i are mapped in a many-to-one fashion to S_n. We denote the distribution on S_n induced by the GSR model with m shuffles and initial mass concentrated on the identity e of S_n by $P_{n,m}$.

A rising sequence of a permutation $\pi \in S_n$ is a maximal subset of an arrangement of cards consisting of successive values displayed in order. In the example, the rising sequences are (1), (2), (3, 4, 5, 6) and (7, 8). The number of rising sequences is denoted by r. It is shown in [2] that the value of $P_{n,m}(\pi)$ only depends on the number of rising sequences $r = r(\pi)$ and

$$P_{n,m}(\pi) = \binom{2^m + n - r}{n} 2^{-mn}.$$

Let U be the uniform distribution on S_n, so that $U(\pi) = 1/n!$ for all $\pi \in S_n$. It is well known that the GSR model has U as a stationary distribution.

The relative entropy to stationarity $H(P_{n,m}|U)$ is

$$H(P_{n,m}|U) = \sum_{\pi \in S_n} P_{n,m}(\pi) \log_2 \left(n! P_{n,m}(\pi)\right).$$

We get explicit limits for $H(P_{n,m}|U)$, which reveal the surprising linear/geometric transition in the decrease of $H(P_{n,m}|U)$ at $m = \log_2 n$.

Theorem 1 *Let $\omega(n)$ be an arbitrary function such that $\omega(n) \to \infty$ and let $\alpha = \log_2 e \doteq 1.44270$. Then,*

$$H(P_{n,m}|U) = \log_2 n! - mn + o(n) \quad \text{for} \quad m = \log_2 n - \omega(n)$$

and

$$H(P_{n,m}|U) = \frac{\alpha}{24} n^3 2^{-2m} + O(n^4 2^{-3m}) + O(n^2 2^{-2m}) \quad \text{for} \quad m = \log_2 n + \omega(n).$$

Moreover, if $m \in [\log_2 n - O(1), \log_2 n - 2]$, then

$$(2-\alpha)n + O(n^{1/2}) \leq H(P_{n,m}|U) = O(n).$$

The first $O(\cdot)$ in the expression for $H(P_{n,m}|U)$ for $m = \log_2 n + \omega(n)$ dominates the second $O(\cdot)$ when $m = 2\log_2 n - \omega(n)$, while the second dominates the first when $m = 2\log_2 n + \omega(n)$.

We display the linear/geometric transition more explicitly in our first corollary.

Corollary 1 *If $m = \log_2 n - \omega(n)$, $\omega(n) \to \infty$, then*

$$H(P_{n,m+1}|U) - H(P_{n,m}|U) = -n + o(n).$$

If $m = \log_2 n + \omega(n)$, $\omega(n) \to \infty$, then

$$\frac{H(P_{n,m+1}|U)}{H(P_{n,m}|U)} = \frac{1}{4} + o(1).$$

Proof The first statement follows automatically from Theorem 1. The second statement results from $n^4 2^{-3m}/(n^3 2^{-2m}) = n 2^{-m} = o(1)$ and $n^2 2^{-2m}/(n^3 2^{-2m}) = n^{-1} = o(1)$. ∎

Our second corollary demonstrates that $m = \frac{3}{2}\log_2 n$ is the threshold to mix a deck of cards in relative entropy.

Corollary 2 *If $m = \frac{3}{2}\log_2 n + c$, c a real constant, then*

$$H(P_{n,m}|U) = \frac{\alpha}{24}4^{-c}.$$

As we remwrked earlier, it was noted in [7] that $H(P_{n,m}|U) \doteq 0.06011$ when $m = \log_2 n$, agreeing with our limit $H(P_{n,m}|U) \to \alpha/24 \doteq 0.060112293$.

Corollary 2 may also be proven in the manner that [2] proves limits at threshold for total variation distance: by using the normality of the Eulerian numbers.

The proof of Theorem 1 for $m = \log_2 - \omega(n)$ uses estimates which are uniform over all $r \in [1, 2^m]$. The proof for $m = \log_2 n + \omega(n)$ uses an expansion of the logarithm of the ratio of probability densities. Key tools are estimates on the expectation and variance of the number of rising sequences R under the measure $P_{n,m}$.

The way we use estimate $E(R)$ is the through the identity (6), which basically corrects equation (4.23) of [1]. The number of ascents of a permutation π is

$$A(\pi) = |\{i \in [1, n-1] : \pi(i) < \pi(i+1)\}|.$$

An estimate on the expectation of A after m reverse shuffles is given in [1] to help prove the existence of a threshold in total variation distance at $m = \frac{3}{2}\log_2 n$. The number of descents is $D(\pi) = n - 1 - A(\pi)$ and a permutation has r rising sequences if and only if its inverse has $r-1$ descents. It therefore follows from Lemma 3 that the expected number of ascents after m reverse shuffles is

$$E_{\text{reverse}}(A) = n-1-(E(R)-1) = \frac{n}{2} + \frac{1}{12}n^2 2^{-m} - \frac{1}{2} + O(n^3 2^{-2m}) + O(n2^{-m}).$$

The proof of (4.19) in [1] is valid with this modification.

3 Proofs

Lemma 1, Lemma 2 and Lemma 4 handle the cases $m = \log_2 n - \omega(n)$, $m = \log_2 n + O(1)$, and $m = \log_2 n + \omega(n)$, respectively.

Lemma 1 *If $m = \log_2 n - \omega(n)$, $\omega(n) \to \infty$, then*

$$H(P_{n,m}|U) = \log_2 n! - mn + o(n).$$

Proof For $r \leq 2^m$, define $A(r)$ to be

$$A(r) := \binom{n+2^m-r}{n} \leq \binom{n+2^m}{n}, \tag{1}$$

so that

$$n!P_{n,m}(\pi) = n!2^{-mn}A(r). \tag{2}$$

We use Stirling's formula:

$$\begin{aligned} \log_2 n! &= \alpha \ln n! \\ &= \alpha n \ln n - \alpha n + O\left(n^{1/2}\right) \\ &= n \log_2 n - \alpha n + O\left(n^{1/2}\right). \end{aligned} \tag{3}$$

It follows from (3) and $2^m = n2^{-\omega(n)}$ that

$$\begin{aligned} \log_2 \binom{n+2^m}{n} &= \log_2(n+2^m)! - \log_2 n! - \log_2(2^m)! \\ &= (n+2^m)\log_2(n+2^m) - n\log_2 n - 2^m \log_2 2^m + O\left(n^{1/2}\right) \\ &= o(n). \end{aligned} \tag{4}$$

Combining (2), (1), and (4) completes the proof. ■

Lemma 2 *If* $m \in [\log_2 n - O(1), \log_2 n - 2]$, *then*

$$(2-\alpha)n + O(n^{1/2}) \leq H(P_{n,m}|U) = O(n).$$

Proof The lower bound is given by

$$\begin{aligned} H(P_{n,m}|U) &\geq \log_2(n!) - mn \\ &= n\log_2 n - \alpha n - mn + O(n^{1/2}) \\ &\geq 2n - \alpha n + O(n^{1/2}). \end{aligned} \tag{5}$$

The upper bound follows from (5) and

$$\begin{aligned} \log_2 \binom{n+2^m}{n} &\leq \log_2 \binom{2n}{n} \\ &= \log_2(2n)! - 2\log_2(n!) \\ &= 2n\log_2(2n) - 2(n\log_2 n) + O(n^{1/2}) \\ &= 2n + O(n^{1/2}). \end{aligned}$$

■

Lemma 3 *Let R be the random variable defined on the measure space of permutations counting the number of rising sequences. If $m = \log_2 n + \omega(n)$, $\omega(n) \to \infty$, then, under $P_{n,m}$,*

$$E(R) = \frac{n}{2} - \frac{1}{12}n^2 2^{-m} + \frac{1}{2} + O(n^3 2^{-2m}) + O(n 2^{-m})$$

and

$$\mathrm{Var}(R) = O(n).$$

Proof Suppose that a realization of the D_i takes on L distinct values between 0 and $2^m - 1$, so that L is a random variable, and list the values in order $0 \leq c_1 < c_2 < \cdots < c_{L-1} < c_L \leq 2^m - 1$. Let $N_j = \sum_{k=1}^{n} I\left[D_k = c_j\right]$, $j \in [1, L]$, count the number of times the value c_j is taken on. Consider the distribution of shuffles conditioned on $\mathcal{F}$, the σ-field generated by L, $(c_1, c_2, \ldots, c_L)$, and $(N_1, N_2, \ldots, N_L)$. The values of the permutation generated by the D_i that are taken to $\{i : D_i = c_1\}$ always begin a rising sequence. The values of the permutation generated by the D_i that are taken to $\{i : D_i = c_{j+1}\}$ for a fixed $j \in [1, L-1]$ will begin a new rising sequence unless $\inf\{i : D_i = c_{j+1}\} > \sup\{i : D_i = c_j\}$. The random variable counting the number of rising sequences of the permutation generated by the D_i's is therefore

$$R = L - \sum_{j=1}^{L-1} I_j, \tag{6}$$

where I_j are the indicator variables

$$I_j = I\left[\inf\{i : D_i = c_{j+1}\} > \sup\{i : D_i = c_j\}\right].$$

Indicators I_j and I_k are independent whenever $|j - k| > 1$ and they have means

$$E(I_j) = \binom{N_j + N_{j+1}}{N_j}^{-1}.$$

Let J_{l_1,l_2}, $1 \leq l_1, l_2 \leq n$, be the random variables defined by

$$J_{l_1,l_2} = \sum_{j=1}^{L-1} I\left[N_j = l_1, N_{j+1} = l_2\right].$$

The conditional expectation of R is

$$\begin{aligned}
E(R|\mathcal{F}) &= L - \sum_{j=1}^{L-1} \binom{N_j + N_{j+1}}{N_j}^{-1} \\
&= L - \sum_{j=1}^{L-1} \sum_{l_1,l_2} I\left[N_j = l_1, N_{j+1} = l_2\right] \binom{l_1 + l_2}{l_2}^{-1} \\
&= L - \sum_{l_1,l_2} \sum_{j=1}^{L-1} I\left[N_j = l_1, N_{j+1} = l_2\right] \binom{l_1 + l_2}{l_2}^{-1} \\
&= L - \sum_{l_1,l_2} J_{l_1,l_2} \binom{l_1 + l_2}{l_2}^{-1}.
\end{aligned}$$

The expected value of R is

$$E(R) = E(E(R|\mathcal{F})) = E(L) - \sum_{l_1,l_2} E(J_{l_1,l_2}) \binom{l_1 + l_2}{l_2}^{-1}. \tag{7}$$

We calculate

$$\begin{aligned}
E(L) &= 2^m \left(1 - \left(1 - 2^{-m}\right)^n\right) \\
&= 2^m \left(n2^{-m} - \binom{n}{2} 2^{-2m} + O\left(n^3 2^{-3m}\right)\right) \\
&= n - \binom{n}{2} 2^{-m} + O\left(n^3 2^{-2m}\right). \qquad (8)
\end{aligned}$$

Estimating the $E(J_{l_1,l_2})$ takes a substantial part of the rest of the proof. Let $K_l = \sum_{j=1}^{n} I[N_j = l]$ and note that

$$\begin{aligned}
K_l &= \sum_{l_2} J_{l,l_2} + I[N_L = l] \qquad (9) \\
&= \sum_{l_1} J_{l_1,l} + I[N_1 = l]. \qquad (10)
\end{aligned}$$

Taking $l = 2$ gives in (9) gives

$$K_2 = J_{2,1} + J_{2,2} + \cdots + J_{2,n} + I[N_L = 2]. \tag{11}$$

Since $J_{l,l_2} \le K_l$ and

$$\begin{aligned}
E(K_l) &= 2^m \binom{n}{l} 2^{-lm} \left(1 - 2^{-m}\right)^{n-l} \qquad (12) \\
&\le 2^m \frac{(n2^{-m})^l}{l!}
\end{aligned}$$

we have

$$
\begin{aligned}
E(J_{2,3}) + E(J_{2,4}) + \cdots + E(J_{2,n}) &\leq E(K_3) + E(K_4) + \cdots + E(K_n) \\
&\leq \sum_{l=3}^{\infty} 2^m \frac{(n2^{-m})^l}{l!} \\
&= O(n^3 2^{-2m}). \qquad (13)
\end{aligned}
$$

The expected value of $J_{2,2}$ is

$$
\begin{aligned}
E(J_{2,2}) &= \sum_{1 \leq s < t \leq 2^m} \binom{n}{2}\binom{n-2}{2} 2^{-4m} \left(1 - \frac{1+t-s}{2^m}\right)^{n-4} \\
&\leq n^4 2^{-4m} \sum_{1 \leq s < t \leq 2^m} \left(1 - \frac{1+t-s}{2^m}\right)^{n-4},
\end{aligned}
$$

where s and t represent c_j with $N_j = 2$, the binomial coefficients are the number of ways of choosing two i such that $D_i = s$ and another two i such that $D_i = t$, 2^{-4m} is the probability that the chosen D_i are s and t, and the other factor is the probability that none of the other D_i's equal s or t or a value in between them. The sum is bounded by

$$
\begin{aligned}
\sum_{1 \leq s < t \leq 2^m} \left(1 - \frac{1+t-s}{2^m}\right)^{n-4} &\leq n \sum_{k=2}^{\infty} \left(1 - \frac{k}{2^m}\right)^{n-4} \\
&\leq n \sum_{k=2}^{\infty} \exp\left(-\frac{k(n-4)}{2^m}\right) \\
&= n \frac{\exp\left(-\frac{2(n-4)}{2^m}\right)}{1 - \exp\left(-\frac{(n-4)}{2^m}\right)} \\
&= O\left(2^m\right),
\end{aligned}
$$

hence

$$
E(J_{2,2}) = O(n^4 2^{-3m}). \qquad (14)
$$

The expectation of the indicator in (11) is

$$
P(N_L = 2) = \sum_{k=1}^{2^m} \binom{n}{2} 2^{-2m} \left(1 - \frac{k}{2^m}\right)^{n-2},
$$

where $n-k+1$ represents c_L, so that

$$\begin{aligned} P(N_L=2) &\le \sum_{k=1}^{2^m} \frac{n^2}{2} 2^{-2m} \exp\left(-\frac{k(n-2)}{2^m}\right) \\ &\le \frac{n^2}{2} 2^{-2m} \frac{\exp\left(-\frac{2(n-2)}{2^m}\right)}{1-\exp\left(-\frac{(n-2)}{2^m}\right)} \\ &= O(n2^{-m}). \end{aligned} \tag{15}$$

Putting (13), (14), and (15) into the expected value of both sides of (11) gives

$$E(J_{2,1}) = E(K_2) + O\left(n^3 2^{-2m}\right) + O\left(n2^{-m}\right). \tag{16}$$

A similar argument using (10) produces

$$E(J_{1,2}) = E(K_2) + O\left(n^3 2^{-2m}\right) + O\left(n2^{-m}\right). \tag{17}$$

Letting $l=1$ in (10) gives

$$K_1 = J_{1,1} + J_{2,1} + \cdots + J_{n,1} + I[N_1 = 1]. \tag{18}$$

If one takes expectations on both sides of (18), then by (16) and the argument showing (13) one has

$$E(K_1) = E(J_{1,1}) + E(K_2) + P(N_1=1) + O(n^3 2^{-2m}) + O\left(n2^{-m}\right). \tag{19}$$

By definition,

$$n = K_1 + 2K_2 + 3K_3 + \cdots + nK_n,$$

An argument like the one showing (13) implies

$$3E(K_3) + \cdots + nE(K_n) = O(n^3 2^{-2m}),$$

so that

$$E(K_1) = n - 2E(K_2) + O(n^3 2^{-2m}). \tag{20}$$

Note that

$$\begin{aligned} P(N_1>1) &= \sum_{k=2}^{n} P(N_1=k) \\ &= \sum_{k=2}^{n}\sum_{i=1}^{2^m} \binom{n}{k} 2^{-km}\left(1-\frac{i}{2^m}\right)^{n-k} \\ &= \sum_{k=2}^{\lceil n/2\rceil}\sum_{i=1}^{2^m} \binom{n}{k} 2^{-km}\left(1-\frac{i}{2^m}\right)^{n-k} + \sum_{k=\lceil n/2\rceil+1}^{n}\sum_{i=1}^{2^m} \binom{n}{k} 2^{-km}\left(1-\frac{i}{2^m}\right)^{n-k}. \end{aligned}$$

The first sum is bounded by

$$
\begin{aligned}
\sum_{k=2}^{\lceil n/2\rceil}\sum_{i=1}^{2^m}\binom{n}{k}2^{-km}\Big(1-\frac{i}{2^m}\Big)^{n-k} &\le \sum_{k=2}^{\lceil n/2\rceil}\sum_{i=1}^{2^m}\frac{n^k}{k!}2^{-km}\exp\left(-\frac{i(n-k)}{2^m}\right)\\
&\le \sum_{k=2}^{\lceil n/2\rceil}\sum_{i=1}^{2^m}\frac{n^k}{k!}2^{-km}\exp\left(-\frac{in}{2^{m+1}}\right)\\
&\le \sum_{k=2}^{\lceil n/2\rceil}\frac{n^k}{k!}2^{-km}\frac{\exp\left(n2^{-m-1}\right)}{1-\exp\left(n2^{-m-1}\right)}\\
&= O\left(n^2 2^{-2m}\cdot 2^m n^{-1}\right)\\
&= O(n2^{-m}).
\end{aligned}
$$

The second sum is bounded by

$$
\begin{aligned}
\sum_{k=\lceil n/2\rceil+1}^{n}\sum_{i=1}^{2^m}\binom{n}{k}2^{-km}\Big(1-\frac{i}{2^m}\Big)^{n-k} &\le 2^{-(\lceil n/2\rceil+1)m}\sum_{k=\lceil n/2\rceil+1}^{n}\sum_{i=1}^{2^m}\binom{n}{k}\\
&\le 2^{-\lceil n/2\rceil m+n}
\end{aligned}
$$

Hence, $P(N_1 > 1) = O(n2^{-m})$ and

$$P(N_1 = 1) = 1 - O(n2^{-m}). \tag{21}$$

It follows from (19), (20), and (21) that

$$E(J_{1,1}) = n - 3E(K_2) - 1 + O(n^3 2^{-2m}) + O(n2^{-m}). \tag{22}$$

The sum in (7) for $l_1, l_2 \ge 3$ is bounded by

$$\sum_{l_1\ge 3}\left(\sum_{l_2\ge 3}E(J_{l_1,l_2})\right) \le \sum_{l_1\ge 3}E(K_{l_1}) = O(n^3 2^{-2m}). \tag{23}$$

Substituting (16), (17), (22), and (23) into (7), we have

$$
\begin{aligned}
E(R) &= E(L) - \frac{1}{2}(n - 3E(K_2) - 1) - \frac{2}{3}E(K_2) + O(n^3 2^{-2m}) + O(n2^{-m})\\
&= E(L) - \frac{n}{2} + \frac{5}{6}E(K_2) + \frac{1}{2} + O(n^3 2^{-2m}) + O(n2^{-m}). \qquad (24)
\end{aligned}
$$

If we let $l = 2$ in (12), we obtain

$$E(K_2) = \binom{n}{2} 2^{-m} + O(n^3 2^{-2m}). \tag{25}$$

Using (8) and then (25) in (24) gives us the first statement of the lemma,

$$\begin{aligned} E(R) &= \frac{n}{2} - \frac{1}{6}\binom{n}{2} 2^{-m} + \frac{1}{2} + O(n^3 2^{-2m}) + O(n2^{-m}) \\ &= \frac{n}{2} - \frac{1}{12} n^2 2^{-m} + \frac{1}{2} + O(n^3 2^{-2m}) + O(n2^{-m}). \end{aligned}$$

In bounding the variance of R, it is helpful to define I'_k, $k = 1, 2, \ldots, n$ by $I'_k = I_k$ for $k < L$ and $I'_k = 1$ for $k \geq L$. We may then write (6) as

$$R = 1 - \sum_{j=1}^{L-1} (I_j - 1) = 1 - \sum_{k=1}^{n} (I'_k - 1),$$

the variance of which expression is

$$\begin{aligned} \mathrm{Var}(R) &= \mathrm{Var}\left(\sum_{k=1}^{n} (I'_k - 1) \right) \\ &= \sum_{k=1}^{n} \mathrm{Var}(I'_k - 1) + 2 \sum_{1 \leq k < j \leq n} \mathrm{Cov}(I'_k - 1, I'_j - 1). \end{aligned}$$

The covariances are

$$\begin{aligned} \mathrm{Cov}(I'_k - 1, I'_j - 1) &= E\left((I'_k - E(I'_k))(I'_j - E(I'_j)) \right) \\ &= E\left(E\left((I'_k - E(I'_k))(I'_j - E(I'_j)) | \mathcal{F} \right) \right). \end{aligned} \tag{26}$$

If $j > k+1$ and $j \geq L$, then I'_j is identically 1 and the conditional expectation in (26) is 0; while if $j > k + 1$ and $j < L$, then $I'_k = I_k$ and $I'_j = I_j$ are independent and again it is 0. Therefore,

$$\begin{aligned} \mathrm{Var}(R) &= \sum_{k=1}^{n} \mathrm{Var}(I'_k - 1) + 2 \sum_{1 \leq k \leq n-1} \mathrm{Cov}(I'_k - 1, I'_{k+1} - 1) \\ &\leq n + 2(n-1) < 3n. \end{aligned}$$

∎

Lemma 4 *If $m = \log_2 n + \omega(n)$, $\omega(n) \to \infty$, then*

$$H(P_{n,m}|U) = \frac{\alpha}{24} n^3 2^{-2m} + O\left(n^4 2^{-3m}\right) + O\left(n^2 2^{-2m}\right).$$

Proof The relative entropy is

$$H(P_{n,m}|U) = \sum_{\pi \in S_n} P_{n,m}(\pi) \log_2(n! P_{n,m}(\pi)) = \alpha \sum_{\pi \in S_n} P_{n,m}(\pi) \ln(n! P_{n,m}(\pi)). \tag{27}$$

We have

$$\begin{aligned}
\ln(n! P_{n,m}(\pi)) &= \ln\left((2^m + n - r)_n 2^{-mn}\right) \\
&= \ln\left(\prod_{k=0}^{n-1} \frac{2^m + n - k - r}{2^m}\right) \\
&= \sum_{k=0}^{n-1} \ln\left(1 + \frac{n-k-r}{2^m}\right) \\
&= \sum_{k=0}^{n-1} \frac{n-k-r}{2^m} - \frac{1}{2} \sum_{k=0}^{n-1} \left(\frac{n-k-r}{2^m}\right)^2 + O\left(n^4 2^{-3m}\right) \\
&= \left(\frac{n(n+1)}{2} - rn\right) 2^{-m} - \frac{1}{2}\left(\frac{n^3}{3} - rn(n+1) + r^2 n\right) 2^{-2m} \\
&\quad + O\left(n^4 2^{-3m}\right) + O\left(n^2 2^{-2m}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sum_{\pi \in S_n} P_{n,m}(\pi) \ln(n! P_{n,m}(\pi)) &= \left(\frac{n(n+1)}{2} - E(R)n\right) 2^{-m} \\
&\quad - \frac{1}{2}\left(\frac{n^3}{3} - E(R)n(n+1) + \left(\mathrm{Var}(R) + (E(R))^2\right) n\right) 2^{-2m} \\
&\quad + O\left(n^4 2^{-3m}\right) + O\left(n^2 2^{-2m}\right).
\end{aligned}$$

Lemma 3 and (27) give

$$\begin{aligned}
H(P_{n,m}|U) &= \left(\frac{1}{12} - \frac{1}{6} + \frac{1}{4} - \frac{1}{8}\right) \alpha n^3 2^{-2m} + O\left(n^4 2^{-3m}\right) + O(n^2 2^{-2m}) \\
&= \frac{\alpha}{24} n^3 2^{-2m} + O\left(n^4 2^{-3m}\right) + O\left(n^2 2^{-2m}\right).
\end{aligned}$$

■

References

[1] Aldous, D. (1983) Random walks on finite groups and rapidly mixing Markov chains. *Séminaire de Probabilités XVII. Lecture Notes in Math.*, **986**, 243 –297. Springer, New York.

[2] Bayer, D., and Diaconis, P. (1992) Trailing the dovetail shuffle to its lair. *Ann. Appl. Prob.*, **2**, 294 –313.

[3] Diaconis, P. (1988) *Group Representations in Probability and Statistics.* IMS, Hayward, Calif.

[4] Diaconis, P. and Saloff-Coste, L. (1996) Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Prob.*, **6**, 695 – 750.

[5] Gilbert, E. N. (1955) Theory of shuffling. Bell Telephone Laboratories memorandum.

[6] F.E. Su. PhD Thesis, Harvard, 1995.

[7] Trefethen, L. N. and Trefethen, L. M. (1998) How many shuffles to randomize a deck of cards? Oxford University Computing Laboratory, Technical Report Number 98/09.