

Notes On Card Shuffling

Nathanaël Berestycki

March 1, 2007

Take a deck of $n = 52$ cards and shuffle it. It is intuitive that if you shuffle your deck sufficiently many times, the deck will be in an approximately random order. But how many is sufficiently many ?

These notes will try to explain the amazing phenomenon that for the most popular shuffling method, the “riffle shuffle”, 7 shuffles are necessary and sufficient to bring the deck close to randomness, and that for a general deck of n cards, 7 becomes $(3/2) \log_2 n$. This is the content of two papers by Aldous [1] and by Bayer and Diaconis [4]. We refer the interested reader to the excellent monographs of Diaconis [6] and Saloff-Coste [9] for more about mixing time of random walks and Markov chains. See also the book in preparation [3] (available on the web).

Many thanks to Laurent Saloff-Coste, who clarified several points for me while I was reading these papers.

1 Introduction

A deck of cards is viewed as a permutation $\sigma \in \mathcal{S}_n$ the symmetric group (where in practice $n = 52$). To make this identification, we use the convention

$$\sigma(i) = \text{label of card in position } i \text{ for } 1 \leq i \leq n \quad (1)$$

E.g., the deck

1 7 2 8 9 3 10 4 5 11 6 12 13

is represented by the permutation that maps 1 into 1, 2 into 7, 3 into 2, and so on. Note that $\sigma^{-1}(i)$ is the position of the card with label i in the deck.

1.1 Model.

Our basic framework is the Gilbert-Shannon-Reeds model for card shuffling, which is defined as follows. We first cut the deck in two piles of size k and $n - k$, where the position k of the cut follows a Binomial $(n, 1/2)$ distribution. Then, if we imagine that we hold the two piles in our left and right hand, drop the next card from the left or right pile with probability proportional to the size of the pile. That is, if there are a cards in the left hand and b cards in the right hand, drop from the left with proba $a/(a + b)$ and from the right with proba $b/(a + b)$. This gives you a new deck which is the result of one shuffle. This shuffle is then repeated many times.

1.2 Shuffling as a Random Walk.

It is easy to see that repeated shuffling can be viewed as a random walk on \mathcal{S}_n in the following sense. Let G be a group, and let S be a generating set of G . (That is, any element of G can be written as a finite product of elements in S). Let μ be a probability measure on S .

Definition 1. *The random walk on G is the process $(X_m, m \geq 0)$ such that*

$$X_m = g_1 \cdot \dots \cdot g_m$$

where the g_i 's are an *i.i.d.* sequence distributed according to μ .

If $G = \mathbf{Z}^d$ and S is the set of all $2d$ nearest-neighbours of the origin, and μ is the uniform measure on S , this leads to the usual simple random walk of our undergraduate classes. Here, $G = \mathcal{S}_n$ and S is the set of permutation that can be reached in one shuffle (later, this will be called R , and is the set of permutation with two rising sequences, plus the identity). μ is the probability measure assigned to a particular ordering of the deck under one riffle shuffle.

1.3 Exponential convergence.

Thus X_m is a random walk on the finite-state space \mathcal{S}_n . What could be simpler? In particular, it is a Markov chain, and it is easy to check that it is irreducible aperiodic. Denoting by P the transition matrix, elementary theory tells us that X_m has an invariant measure U which turns out to be simply the uniform measure on \mathcal{S}_n . Invariant refers to the fact that if you start with distribution U and apply a riffle shuffle then the state is still uniform. (At first it does not seem easy to see why is the uniform distribution invariant, but we will see why later). Moreover the

Perron-Frobenius theorem tells us that if the second eigenvalue of P is λ (always < 1), then

$$\sup_x \|P^m x - U\| \leq C\lambda^m$$

where the norm $\|\cdot\|$ is your favourite norm. At first sight this seems to be giving all the information we care about: we converge to stationarity exponentially fast. In fact this is very misleading: the Perron-Frobenius says nothing quantitative. The only information is that the *eventual* rate of decay is exponential with a rate determined by the second largest eigenvalue of the chain.

1.4 Total Variation Distance.

We will thus formulate the question more precisely. To start with, we need a way to measure how far away from stationarity we are. This is done by introducing the total variation between two probability measure on the state-space of the chain. Let μ and ν be two probability measures on a space X . (For us, X is the symmetric group \mathcal{S}_n).

Definition 2. *The total variation distance between μ and ν is*

$$d_{TV}(\mu, \nu) = \|\mu - \nu\| = \sup_{A \in X} |\mu(A) - \nu(A)| \quad (2)$$

An easy exercise shows that if X is discrete,

$$\|\mu - \nu\| = \frac{1}{2} \sum_{x \in X} (\mu(x) - \nu(x))^+$$

and as a consequence

$$0 \leq d_{TV}(\mu, \nu) \leq 1$$

That is, the maximal value that the total variation distance can take is 1.

1.5 The Cutoff Phenomenon.

Fact: for many random walks, “as $n \rightarrow \infty$ ”, if

$$d(m) = \max_{x \in G} d_{TV}(P_x(X_m \in \cdot), U)$$

where P_x denotes the law of the process started at $x \in G$, then $d(m)$ has the following approximate behaviour. Let $\tau = \inf\{m \geq 1 : d(m) < 1/e\}$ (where the constant $1/2$) is chosen for convenicen, but could be replaces by anything < 1).

Then $d(m)$ is approximately equal to 1 for all $m < \tau$, drops sharply at about τ , and is approximately 0 for $m > \tau$. I.e.,

$$d(t\tau) \rightarrow \mathbf{1}_{\{0 \leq t < 1\}} \quad (3)$$

in probability. If (3) holds for a certain Markov chain, we say that it experiences a cutoff at time τ , and this is referred to as the **cutoff phenomenon**. The results that we will describe in these notes prove that the Gilbert-Shannon-Reeds riffle shuffle experiences the cutoff at time $\tau = (3/2) \log_2 n$, and give in fact a very detailed study of what happens near the cutoff.

A simple but useful observation is that $d(m)$ satisfies the two following properties.

- (i) $d(m)$ is non-increasing.
- (ii) Even stronger, $d(m)$ is a *submultiplicative* function of m . That is,

$$d(m + m') \leq d(m)d(m')$$

In view of (ii), if at some time m the distance becomes $\leq c < 1$, then from then on it will decrease exponentially fast: $d(km) \leq c^k$. This partly explains why the cutoff phenomenon is believed to be widely observable.

2 Statement of the Results

We are now ready to state the main theorems. Let X_m be the state of the deck after m riffle shuffles.

Theorem 1. (Aldous 1983 [1])

- 1. If $m = (1 - \varepsilon)(3/2) \log_2 n$ then $d(m) \rightarrow_p 1$
- 2. If $m = (1 + \varepsilon)(3/2) \log_2 n$ then $d(m) \rightarrow_p 0$.

This says that X has a cutoff at time $(3/2) \log_2 n$. The following results of Bayer and Diaconis analyze this in an exact and much sharper way. The first amazing result is an exact formula for the probability distribution of the walk after m steps.

Theorem 2. (Bayer-Diaconis 1992 [4]) *After m shuffles,*

$$P(X_m = \sigma) = \frac{1}{2^{mn}} \binom{2^m + n - R(\sigma)}{n}$$

where $R(\sigma)$ is the number of rising sequences of σ , defined below.

Using this exact formula, Bayer and Diaconis were able to study in great detail what happens near the cutoff point, after of order $(3/2) \log_2 n$ shuffles have been performed.

Theorem 3. (Bayer-Diaconis 1992 [4]) *Let $m = \log_2(n^{3/2}c)$. Then*

$$d(m) = 1 - 2\Phi\left(-\frac{1}{4\sqrt{3}c}\right) + O(n^{-1/4})$$

where $\Phi(x)$ is the cumulative distribution function of a standard normal random variable:

$$\Phi(x) = \int_{-\infty}^x e^{-u^2/2} \frac{du}{\sqrt{2\pi u}}$$

We now comment on the numerical values of those constants for $n = 52$. First, note that in this case,

$$(3/2) \log_2 n = 8.55 \dots$$

which indicates that of order 8 or 9 shuffles are necessary and sufficient.

However, based on the Bayer-Diaconis formula and an exact expression for the number of permutation with a given number of rising sequences (an *Eulerian number*, discussed later), we obtain the exact value for $d(m)$

m	5	6	7	8	9
$d(m)$	0.92	0.614	0.33	0.167	0.085

As we see from this table, it is clear that convergence to equilibrium occurs after no less than 7 shuffles. The total variation distance decreases by 2 after each successive shuffle following the transition point.

Remark. It is interesting to note that while 7 is a very small number compared to the size of the state-space ($52!$ which has about 60 digits), this is a rather large number in practice. Nobody ever shuffles a deck of card more than 3 or 4 times. It is easy to take advantage of this in magic tricks (and in casinos, apparently). Bayer and Diaconis describe some very pleasant tricks which exploit the non-randomness of the deck at this stage, which are based on the analysis of the riffle shuffle and in particular of the rising sequences. The reading of the original paper [4] is wholeheartedly recommended !

3 Proof of Aldous' result

We will present the key ideas that lead to the proof of Theorem 1. As we will see, many of the ideas that were used by Bayer and Diaconis were already present in that paper, which appeared about 10 years before.

Before we do anything, we need to define what are the rising sequences of a permutation σ , as the analysis essentially concentrates on the description of their evolution under the shuffle.

Definition 3. *Let $\sigma \in \mathcal{S}_n$. The rising sequences of the arrangement of cards σ are the maximal subsets of successive card labels such that these cards are in increasing order.*

This definition is a little hard to digest at first but a picture illustrates the idea, which is very simple. For instance, if $n = 13$ and the deck consists of the following arrangement:

1 7 2 8 9 3 10 4 5 11 6 12 13

then there are two rising sequences:

$$\begin{array}{cccccc} 1 & & 2 & & 3 & & 4 & 5 & & 6 \\ & 7 & & 8 & 9 & & 10 & & 11 & & 12 & 13 \end{array}$$

The number of rising sequences of σ is denoted by $R(\sigma)$. Note that rising sequences form a partition of the card labels $1, \dots, n$.

The reason why rising sequences are so essential to the analysis is because when we perform a shuffle, we can only double $R(\sigma)$. The above example illustrates well this idea. The two rising sequences identify the two piles that have resulted from cutting the deck and that have been used to generate the permutation σ in one shuffle. This leads to the following equivalent description of the Gilbert-Shannon-Reeds riffle shuffle measure μ .

Description 2. μ is uniform on the set R of permutation with exactly two rising sequences, and gives mass $(n + 1)2^{-n}$ to the identity.

To see this, fix a permutation $\sigma \in R$. The two rising sequences of σ have length L and $n - L$, say. Then as explained above, they identify the cut of the two piles that have resulted from cutting the deck. The probability of having made exactly this cut is $\binom{n}{L}2^{-n}$. We then need to drop the cards from the two piles in the correct order. This corresponds to the product of terms of the form $a/(a + b)$, where a and b are the packet sizes. If we focus on the denominators first, note that this will always be the number of cards remaining in our hands, hence it will be $n, n - 1, \dots, 2, 1$. As for the numerators, cards dropping from the left hand will give us the terms $L, L - 1, \dots, 2, 1$ and terms from the right hand will give us $n - L, n - L - 1, \dots, 2, 1$. It follows that the probability of obtaining σ

$$\mu(\sigma) = \frac{\binom{n}{L}}{2^n} \frac{1}{n!} L!(n - L)! = 2^{-n}$$

Note that a riffle is entirely specified by saying which card comes from the left pile and which from the right pile. Thus, we associate to each card c a binary digit $D(c) = 0$ or 1 , where 0 indicates left and 1 indicates right. By the above description, the resulting deck can be described by sequence of n bits which is uniformly distributed over all possible sequences of n binary digits. (Check that this works with the identity as well). This leads to the following description. Let $\mu'(\sigma) = \mu(\sigma^{-1})$ be the measure associated with the reverse move.

Description 3. The reverse shuffle (i.e., the shuffle associated with the measure μ'), can be described as assigning i.i.d. 0-1 digits to every card c , with $P(D(c) = 1) = 1/2$ and $P(D(c) = 0) = 1/2$. The set of cards c such that $D(c) = 0$ is then put on top of the set of cards with $D(c) = 1$.

The beautiful idea of Aldous is to notice that this reverse description (the backward shuffle) is a lot easier to analyze. Let $(X'_m, m \geq 0)$ be the random walk associated with the shuffling method μ' . Since

$$X'_m = g'_1 \dots g'_m = g_1^{-1} \dots g_m^{-1} = (g_m \dots g_1)^{-1}$$

we see that

$$X'_m =_d X_m^{-1}$$

and it follows easily that the mixing time of the forward shuffle X is the same as the mixing time of the backward shuffle X' . In fact if d' is the total variation function for the walk X' we have

$$d(m) = d'(m)$$

We are thus going to analyze X' and show that it takes exactly $3/2 \log_2 n$ steps to reach equilibrium with this walk.

A very beautiful idea. To describe the state of the deck after m backward shuffles, we successively assign i.i.d. binary digits 0 or 1 to indicate (respectively) top or bottom pile. E.g., after 2 shuffles:

deck	1st shuffle	2d shuffle
—	1	0
—	0	0
—	0	1
—	1	1
—	1	0
—	0	0

Reading right to left, it is easy to see that the deck consists of the cards with labels 00, then 01, then 10, then 11. This generalizes as follows. For any card c ,

we attach m binary digits 0 and 1 which tell us if the card is going to the top or the bottom pile in m successive backward shuffles. We may interpret this sequence by reading from right to left as the binary expansion of a number $D_m(c)$. Then the fundamental properties of the deck are:

- (a) The deck is ordered by increasing values of $D_m(c)$.
- (b) If two cards c and c' have the same value of $D_m(c) = D_m(c')$ then they retain their initial ordering.

Note that the numbers $(D_m(c), 1 \leq c \leq n)$ are i.i.d. for different cards, with a distribution that uniform on $\{0, \dots, 2^m - 1\}$.

3.1 A first upper-bound

One immediate consequence of properties (a) and (b) is that if T = the first time at which all labels $D_m(c)$ are distinct, then the deck is exactly uniformly distributed. We use this remark to get a first upper-bound on the time it takes to get close to stationarity.

Lemma 1. *If $m \gg 2 \log_2 n$ then with high probability all labels $D_m(c)$ are distinct.*

Proof. The proof is elementary, and is a reformulation of the Birthday problem. We view the $M = 2^m$ possible values of $D_m(c)$ as M urns and we are throwing independently n balls into them at random. The probability that they all fall in distinct urns is

$$\begin{aligned} P(\text{all labels distinct}) &= 1 \left(1 - \frac{1}{M}\right) \left(1 - \frac{2}{M}\right) \dots \left(1 - \frac{n-1}{M}\right) \\ &= \exp\left(\sum_{j=0}^{n-1} \ln\left(1 - \frac{j}{M}\right)\right) \\ &\approx \exp\left(-\sum_{j=0}^{n-1} \frac{j}{M}\right) \approx \exp(-n^2/2M) \end{aligned}$$

It follows that if $M \ll n^2$ then some cards will have the same label, but if $M \gg n^2$ then with high probability all cards will have distinct labels. But $M = n^2$ is equivalent to $m = 2 \log_2 n$. \square

To rigorously use Lemma 1 to conclude that the distance function at time $(1+\varepsilon) \log_2 n$ is small, we need a Lemma which expresses the total variation distance as the probability that the two distributions can be successfully coupled. If μ, ν are probability measures, a coupling of μ and ν is a pair of random variables (X, Y) defined on the same probability space such that $X =_d \mu$ and $Y =_d \nu$.

Lemma 2. *The total variation distance is equal to*

$$d_{TV}(\mu, \nu) = \inf_{\text{couplings } (X, Y)} P(X \neq Y)$$

Since $X'_T =_d U$ is uniform, the above Lemma tells us that

$$d(m) \leq P(T > m)$$

and $P(T > m) \rightarrow 0$ if $m = (2 + \varepsilon) \log_2 n$. This is not the $(3/2) \log_2 n$ we were hoping for, but building on these ideas we will do better a bit later.

3.2 A first lower-bound

In the forward shuffle, the essential concept is that of rising sequences. In the backward shuffle, the equivalent notion is that of descents of a permutation. We say that σ has a descent at j (where $1 \leq j \leq n - 1$) if $\sigma(j) > \sigma(j + 1)$. Let

$$\text{Des}(\sigma) = \#\text{descents of } \sigma = \sum_j a_j \tag{4}$$

where a_j is the indicator of the event that σ has a descent at j . It is trivial to observe that

$$R(\sigma) = \text{Des}(\sigma^{-1}) - 1$$

In this lower-bound, we will show that for $m < \log_2 n$, the number of descents of X'_m is not close to the number of descents of a uniform permutation. This will show that the distance is approximately 1.

Lemma 3. *Let $\sigma =_d U$. Then*

$$E(\text{Des}(\sigma)) = (n - 1)/2 \text{ and } \text{var } \text{Des}(\sigma) \sim n/12 \tag{5}$$

The expectation is very easy to compute. In a random permutation each j has probability $1/2$ of being a descent. Moreover there is a lot of independence between the a_j , so it is not surprising that the variance is of order n . In fact, as we will mention later, $\text{Des}(\sigma)$ is approximately normally distributed with this mean and variance.

Now, consider our urn representation of the deck X'_m . Each of the 2^m urns corresponds to a possible value of $D_m(c)$, and those cards which fall in the same urn retain their initial order. It is *obvious* that each urn can create at most one descent when we put piles on top of each other (wince within each urn, the order is the same as initially). It follows that

$$\text{Des}(X'_m) \leq 2^m - 1$$

If $m = (1 - \varepsilon) \log_2 n$ then $\text{Des}(X'_m) \leq n^{1-\varepsilon}$ and thus this is incompatible with (5). The two distributions (X'_m and U) concentrate on permutations with very different number of descents, hence the total variation is close to 1.

3.3 A true lower-bound

Here we we push a bit further the lower-bound of the previous section. We will show that for $m = \alpha \log_2 n$ and $\alpha < 3/2$, then

$$E(\text{Des}(X'_m)) = \frac{n-1}{2} - n^\beta \quad (6)$$

with $\beta > 1/2$, while the variance of $\text{Des}X'_m$ stays $O(n)$. This will imply again that the total variation distance is approximately 1 in this regime. Indeed, (5) implies that for a uniform permutation, the number of descents is $n/2 \pm \sqrt{n}$. Here, (6) implies that the number of descents is $n/2 - n^\beta \pm \sqrt{n}$. Since $\beta > 1/2$, this implies that the two distributions concentrate on permutations with a distinct number of descents.

We need the following lemma, which is a simple modification of the Birthday problem.

Lemma 4. *Throw n balls in M urns, and suppose $M \sim n^\alpha$. Let*

$$U_n = \#\{j \leq n : \text{ball } j \text{ and ball } i \text{ fall in same urn, for some } i < j\}$$

Then

$$E(U_n) \sim \frac{1}{2}n^{2-\alpha} \text{ and } \text{var}(U_n) \sim \frac{1}{2}n^{2-\alpha} \quad (7)$$

There surely is a central limit theorem, too.

To prove (6), consider the set J_m of positions j in the resulting deck such that the card in position j and in position $j+1$ have the same value of D_m . Then note that this j can not be a descent for X'_m . On the other hand, note that the random variables a_j are almost iid outside of J_m . More precisely, conditionally on J_m , the random variables $(a_j, j \text{ odd and } j \notin J_m)$ are independent, and each has expectation $1/2$ (and similarly with even values of J). From this we deduce:

$$E(\text{Des}(X'_m)|J_m) = \frac{n-1}{2} - \#J_m$$

(each integer gives us probability $1/2$ of being a descent, except those who are in J_m). Also,

$$\text{var Des}(X'_m) = O(n)$$

Now, to conclude, remark that $\#J_m =_d U_n$ in equation (7) and thus

$$E(\#J_m) \sim \frac{1}{2}n^{2-\alpha}.$$

Since $\beta = 2 - \alpha > 1/2$, the lower-bound is proved.

3.4 Guessing the true upper-bound

We now wish to prove that after $m = (3/2 + \varepsilon) \log_2 n$, the deck is well-mixed. Aldous [1] has a calculation that looks pretty simple but that I haven't managed to clarify completely. Instead I propose the following intuitive explanation.

After $\alpha \log_2 n$ shuffles and $\alpha > 3/2$, the number of descents can still be written as

$$\frac{n-1}{2} - n^{2-\alpha} + \text{standard deviation term}$$

What happens is that $n^{2-\alpha}$ becomes $o(n^{1/2})$ and hence the variance term takes over. It is in fact not hard to believe that at this stage, $\text{Des}X'_m$ is in fact approximately normally distributed with mean $n/2 + o(n^{1/2})$ and variance cn for some $c > 0$. This is almost the same thing as for a uniform permutation, except that the constant for the variance may be different.

Lemma 5. *Let X and Y have two normal distribution with mean 0 and variance σ_1^2 and σ_2^2 . Then*

$$d_{TV}(X, Y) = f(\sigma_1/\sigma_2).$$

f satisfies $0 < f(x) < 1$ for all $x \neq 1$

Lemma 5 and the above comment thus imply that the total variation distance between the law of $\text{Des}X'_m$ and $\text{Des}\sigma$ (where σ is uniform) is at most a constant < 1 .

While that seems pretty far away from our desired conclusion (the total variation distance between X'_m and σ is also < 1), we can in fact get there by using in anticipation the Bayer-Diaconis formula. That formula shows that the number of rising sequences of X_m is a sufficient statistics for X_m . (Here, sufficient statistics refers to the fact that knowing $R(\sigma)$ is enough to know the chance of σ - the meaning may be different in statistics...). Thus, $\text{Des}(X'_m)$ is a sufficient statistics for X'_m , and it is obviously so for a uniform permutation as well. On the other hand,

Lemma 6. *The total variation distance between X and Y is equal to the total variation distance between any two sufficient statistics of X and Y .*

For a proof of this lemma, Bayer and Diaconis refer to an early work of Diaconis and Zebell (1982) [7]. This is a pretty intuitive fact, and from there the upper-bound follows easily !

4 Proof of the results of Bayer and Diaconis

All the foundations are now laid down, and the Bayer-Diaconis formula will follow instantly from the following description of the forward riffle shuffle. (It is a consequence of the urns and balls description of Aldous, but can be proved by other elementary means).

Description 4. X_m is uniform over all ways of splitting the deck into 2^m piles and then riffing the piles together.

4.1 A proof of the Bayer-Diaconis formula.

We now prove the Bayer-Diaconis formula:

$$P(X_m = \sigma) = \frac{1}{2^{mn}} \binom{2^m + n - R(\sigma)}{n}$$

Let $a = 2^m$. There are a^n shuffles in total. Hence it suffices to prove that the number of ways to obtain the permutation σ is $\binom{2^m + n - R(\sigma)}{n}$.

Note that after the a piles are riffled together, the relative order of the cards within a pile remains constant. Hence this gives at most a rising sequences. Let $r = R(\sigma)$, and consider the partition of σ induced by the rising sequences. These r blocks must correspond to r cuts of the deck. The remaining $a - r$ cuts may be placed anywhere in the deck. To count how many ways there are of doing this, we use what Bayer and Diaconis call the “stars and bars” argument. Increase the deck size to $n + a - r$. Now, we must choose $a - r$ positions to put our $a - r$ cuts. There are

$$\binom{n + a - r}{a - r} = \binom{n + a - r}{n}$$

of doing so. Hence the result !

4.2 A proof of the convergence to equilibrium.

Using the above formula we can be very explicit about the total variation distance function. Note that

$$d(m) = \sum_{\pi \in \mathcal{S}_n} \left(P^m(\pi) - \frac{1}{n!} \right)^+ = \sum_{\pi \in \mathcal{S}_n} \frac{1}{n!} (n! P^m(\pi) - 1)^+ \quad (8)$$

Let $m = \log_2(n^{3/2}c)$.

$$\begin{aligned} n!P^m(\pi) &= n! \frac{1}{2^{mn}} \frac{(2^m + n - r) \dots (2^m - r + 1)}{n!} \\ &= \frac{2^m + n - r}{2^m} \dots \frac{2^m - r}{2^m} \\ &= \exp\left(\sum_{i=0}^{n-1} \ln\left(1 + \frac{n - r - i}{2^m}\right)\right) \end{aligned}$$

After an exciting expansion of the log up to the 4th order, and replacing $2^m = n^{3/2}c$ and writing $r = n/2 + h$ (where h may range from $-n/2 + 1$ to $n/2$, we get

$$n!P^m(\pi) = f_n(h) := \exp\left(\frac{-h}{c\sqrt{n}} - \frac{1}{24c^2} - \frac{1}{2}\left(\frac{h}{cn}\right)^2 + O(1/n) + O(h/n)\right) \quad (9)$$

Let h^* be defined by

$$h \leq h^* \iff P^m(\pi) \geq 1/n!$$

This h^* tells us what are the nonzero terms in (8). Now, by setting the exponent equal to 0 in (9), we obtain

$$h^* = -\frac{\sqrt{n}}{24c} + \frac{1}{12c^3} + B + O(1/\sqrt{n}) \quad (10)$$

It follows that

$$d(m) = \sum_{-n/2 \leq h \leq h^*} \frac{R_{nh}}{n!} (f_n(h) - 1)$$

where R_{nh} is the number of permutations with $n/2 + h$ rising sequences. This number is well-known to combinatorists. The number of permutations with j rising sequences is called the Eulerian number a_{nj} , see, e.g. Stanley [10]. Tanny and Stanley show the remarkable formula that if X_1, \dots, X_n are iid uniform on $(0, 1)$

$$\frac{a_{nj}}{n!} = P(j \leq X_1 + \dots + X_n \leq j + 1) \quad (11)$$

This implies in particular the normal approximation for the descents (or the rising sequences) of a uniform random permutation, with variance equal to $n/12$ as claimed in (5).

From then on, it is essentially a game of algebraic manipulations to obtain Theorem 3. We refer the interested reader to p. 308 of [4] for details.

5 A few complements

These results have had a huge impact and have spurred research in many directions. Here are just a couple of them outside of the obvious field of mixing times.

5.1 Strong stationary times

The wonderful stopping time T where all the values of $D_m(c)$ are distinct for the first time, (at which point the deck is *exactly uniform*) is what Aldous and Diaconis [2] have called **strong stationary time**. These are stopping times T such that

- $X_T =_d \pi$ the stationary distribution
- X_T is independent of T .

Here, T is moreover *optimal* in the sense that some configurations can not be reached before T . (Think about the deck $n \ n - 1 \ \dots \ 2 \ 1$. Since before T some two random variables have the same label, their initial ordering is preserved and the deck cannot be in the above configuration). This means that T is the smallest strong stationary time. The theory developed by Aldous and Diaconis [2] shows that if

$$\text{sep}(m) = \sup_{\pi \in \mathcal{S}_n} (n!P^m(\pi) - 1)$$

is the *separation distance* between P^m and the uniform distribution, then

$$\text{sep}(m) = \inf_{T \text{ s.s.t.}} P(T > m)$$

and hence the separation distance has a cutoff at time exactly $2 \log_2 m$. In general the cutoff for the separation distance occurs later than the cutoff for the total variation distance, but never more than twice as long.

5.2 Connection to Dynamical Systems and a Geometric representation

Diaconis [6] mentions the following beautiful geometric representation of the riffle shuffle.

Description 5. Drop n points uniformly on the unit interval, and let $x_1 < \dots < x_n$ be their ordered statistics. To each x_i , apply the transform

$$T : x \mapsto 2x \pmod{1}$$

The resulting order of the x_i 's after applying the transform T gives the deck after one forward riffle shuffle. To obtain the state of the deck after multiple forward shuffles, it now suffices to apply the transform T iteratively !

The evolution of the deck can thus be seen as a dynamical system on the torus with random initial conditions.

5.3 The intermediary phase transition

Let $d_{RS}(\sigma)$ be the minimal number of riffle shuffles needed to build a permutation σ using riffle shuffles. Consider the evolution of this distance as a function of the number of shuffles performed, that is, let $D(m) = d_{RS}(X_m)$. Rick Durrett and I have proved in [5] that

$$\frac{D(t \log_2 n)}{\log_2 n} \rightarrow_p \min(t, 1)$$

In other words, the random walk escapes with a speed equal to 1 as long as $t < 1$ but then the random walk gets abruptly stuck at distance $\log_2 n$, which is the average diameter of the graph.

This seems intuitively related to an older result of Stark, Ganesh, and O'Connell [11], which says that the entropy of the riffle shuffle decays linearly up to time $\log_2 n$ but exponentially fast after that. The relation remains mysterious, although I conjecture that such a connection holds pretty generally (eg, random transpositions, and many similar random walks).

Finally, Fulman [8] has analyzed in quite a lot of detail what happens close to the transition point $t = 1$ of this intermediary phase transition. His results can be stated as follows. Fix an $\alpha > 0$ and consider the state of the Gilbert-Shannon-Reeds riffle shuffle after $m = \lfloor \log_2(\alpha n) \rfloor$ shuffles. Part of Fulman's result may be reformulated in the following way. Let $R(X_m)$ be the number of rising sequences of X_m .

Theorem (Fulman [8]). Suppose $\alpha > 1/(2\pi)$.

$$\frac{1}{n} E(R(X_m)) \rightarrow \alpha - \frac{1}{e^{1/\alpha} - 1} \quad (12)$$

This results leaves open the intriguing question of what happens for smaller values of α . The study of fluctuations is also a very interesting question, see [8] and [5] for a discussion.

References

- [1] D. Aldous (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] D. Aldous and P. Diaconis (1987). Strong uniform times and finite random walks. *Adv. Appl. Math.* 8, 69–97.
- [3] D. Aldous and J. Fill. *Reversible Markov chains and Random Walks on Graphs*. Monograph in preparation, available at <http://www.stat.berkeley.edu/~aldous/RWG/book.html>.
- [4] D. Bayer and P. Diaconis (1992). Trailing the dovetail shuffle to its lair. *Ann. Probab.*, 2, 294-313.
- [5] N. Berestycki and R. Durrett. (2007). Limiting behavior for the distance of a random walk. In preparation.
- [6] P. Diaconis (1988). *Group representation in Probability and Statistics*, Institute of Mathematical Statistics Lecture Notes, Vol. 11.
- [7] P. Diaconis and S. Zabel (1982) Updating subjective probability. *J. Amer. Statist. Assoc.* 77, 822–830.
- [8] J. Fulman (2005). Stein’s method and minimum parsimony distance after shuffles. *Electr. J. Probab.* 10, 901–924.
- [9] L. Saloff-Coste (2003). Random Walks on Finite Groups. In: H. Kesten, ed. *Probability on Discrete Structures*, Encyclopaedia of Mathematical Sciences (110), Springer.
- [10] R. Stanley (1977). Eulerian partitions of a unit hypercube. In: *Higher Combinatorics* (M. Aigner, ed.) Reidel, Dordecht.
- [11] D. Stark, A. Ganesh, and N. O’Connell (2002). Information loss in riffle shuffling, *Combin., Probab., and Computing* 11, 79–95