A RULE OF THUMB FOR RIFFLE SHUFFLING

SAMI ASSAF, PERSI DIACONIS, AND K. SOUNDARARAJAN

ABSTRACT. We study how many riffle shuffles are required to mix n cards if only certain features of the deck are of interest, e.g. suits disregarded or only the colors of interest. For a wide variety of features, the number of shuffles drops from $\frac{3}{2}\log_2 n$ to $\log_2 n$. We derive closed formulae and an asymptotic 'rule of thumb' formula which is remarkably accurate.

1. INTRODUCTION

This paper studies the mixing properties of the Gilbert-Shannon-Reeds model for riffle shuffling n cards. Informally, the deck is cut into two piles by the binomial distribution, and the cards are riffled together according to the rule: if the left packet has A cards and the right has B cards, drop the next card from the left packet with probability $A/(A+B)$ (and from the right packet with probability $B/(A+B)$). Continue until all cards have been dropped. This defines a measure, denoted $Q_2(\sigma)$, on the symmetric group $\mathcal{S}_n$. Repeated shuffles are defined by *convolution powers*

$$Q_2^{*k}(\sigma) = \sum_{\tau\in\mathcal{S}_n} Q_2(\tau)Q_2^{*(k-1)}(\sigma\tau^{-1}). \tag{1}$$

The *uniform distribution* is $U(\sigma) = 1/n!$. There are several notions of the distance between Q_2^{*k} and U: the *total variation distance*

$$\|Q_2^{*k} - U\|_{TV} = \max_{A\subset\mathcal{S}_n} |Q_2^{*k}(A) - U(A)| = \frac{1}{2}\sum_{\sigma\in\mathcal{S}_n} |Q_2^{*k}(\sigma) - U(\sigma)|, \tag{2}$$

and the *separation* and l_∞ metrics

$$\mathrm{SEP}(k) = \max_\sigma 1 - \frac{Q_2^{*k}(\sigma)}{U(\sigma)}\ , \qquad l_\infty(k) = \max_\sigma \left|1 - \frac{Q_2^{*k}(\sigma)}{U(\sigma)}\right|\ . \tag{3}$$

In widely cited works, Aldous [2] and Bayer-Diaconis [3] show that $\frac{3}{2}\log_2(n) + c$ shuffles are necessary and sufficient to make the total variation distance small, while $2\log_2(n) + c$ shuffles are necessary and sufficient to make separation and l_∞ small.

The distances in (2) and (3) look at all aspects of a permutation. In many card games, only some aspects of the permutation matter. For example, in Black-Jack, suits are irrelevant; in Baccarat, suits are irrelevant and all 10's and picture cards are equivalent; ESP card guessing experiments use a Zener deck of 25 cards with each of 5 symbols repeated five times. It is natural to ask how many shuffles are required in these situations. These questions are studied by Conger and Viswanath [8, 9] who derive remarkable numerical procedures giving useful answers for cases of practical interest. Their work is reviewed at the end of this introduction.

In this paper, we develop formulae and asymptotics for a deck of n cards with D_1 cards labelled 1, D_2 cards labelled 2, ..., D_m cards labelled m. Most of the results are proved from the deck starting 'in order', i.e. with 1's on top through m's at the bottom. In Section 5, we show that initial order can change the conclusions.

Date: October 5, 2008.

In Section 2, we begin with $D_1 = 1$ and $D_2 = n-1$. The transition matrix for this case has interesting properties, rivaling the 'Amazing Matrix' in [20]. We show that $\log_2 n + c$ shuffles are necessary and sufficient for convergence in any of our metrics.

Section 3 studies $D_1 = R$, $D_2 = B$, with, for example, $R = B = 26$ modeling the red-black pattern for a standard 52 card deck. We derive a simple formula for $Q_2^{*k}(w)$ for any pattern w and use this to again show that $\log_2 n + c$ steps are necessary and sufficient for convergence to uniformity. We find this surprising as following a single card involves a state space of size n, reds and blacks involves a state space of size $\binom{n}{n/2}$, and yet the same number of shuffles are needed.

In Section 4, we treat the general case, deriving a formula which can be used for some limited calculations. We also reprove a result of Conger-Viswanath determining where the maximum for SEP and l_∞ are achieved. A main result is a unified formula, our *rule of thumb*:

Theorem 1.1. *Consider a deck of n cards with D_i cards of type i, $1 \le i \le m$ with $D_i \ge d \ge 3$, $n = D_1 + \cdots + D_m$. Then the separation distance after k shuffles is*

$$1 - (1+\eta)\frac{2^{k(m-1)}}{(n+1)\cdots(n+m-1)} \sum_{j=0}^{m-1} (-1)^j \binom{m-1}{j} \left(1 - \frac{j}{2^k}\right)^{n+m-1},$$

where η is a real number satisfying

$$|\eta| \le \left(1 + \frac{n^2}{3(d-2)(2^k - m + 1)^2}\right)^{m-1} - 1.$$

This result does not depend on the individual details of the D_i and shows that the same number of shuffles are necessary and sufficient for a variety of questions. For numerical approximation, we set $\eta = 0$ and simply compute the single sum. The bound on η gives explicit error estimates. We demonstrate that the rule of thumb is accurate for both single card and red-black problems studied in earlier sections. This also agrees with the extensive simulation results of Conger-Viswanath and allows results where exact computations and simulation seem out of reach. Some numerical results are summarized below.

TABLE 1. Rule of Thumb for the separation distance for k shuffles of 52 cards.

k	1	2	3	4	5	6	7	8	9	10	11	12
Bayer-Diaconis	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.995	.928	.729	.478	.278
blackjack	1.00	1.00	1.00	1.00	.999	.970	.834	.596	.366	.204	.108	.056
♣♢♡♠	1.00	1.00	.997	.976	.884	.683	.447	.260	.140	.073	.037	.019
redblack	.962	.925	.849	.708	.508	.317	.179	.095	.049	.025	.013	.006
○+≋□☆	1.00	1.00	.993	.943	.778	.536	.321	.177	.093	.048	.024	.012

Remarks on Table 1. The first row gives exact results from the Bayer-Diaconis formula for the full permutation group. The other numbers are from the rule of thumb. Roughly, the single card or red-black numbers suggest that half the usual number of shuffles suffice. The Black-Jack (equivalently Baccarat) numbers suggest a savings of two or three shuffles, and the suit numbers lie in between. The final row is the rule of thumb for the Zener deck with 25 cards, 5 cards for each of 5 suits.

In an appendix, we show that the processes studied below are quotient walks with respect to Young subgroups of $\mathcal{S}_n$. We show how representation theory can be used to derive results for features of the random transposition random walk.

Literature review of riffle shuffles. The basic shuffling model was introduced by Gilbert and Claude Shannon in an unpublished report [19]. The model was independently introduced and studied by Jim Reeds in unpublished work [21]. The first rigorous results are by Aldous [1] who showed that asymptotically $\frac{3}{2}\log_2(n)$ shuffles are correct for total variation. Separation distance is introduced in connection with stopping time arguments in Aldous and Diaconis [2]. They show that $2\log_2 n + c$ steps are necessary and sufficient for separation convergence. The cutoff phenomena is first noticed in this paper as well.

A generalization to a-shuffles is introduced by Bayer-Diaconis in [3]. Here the deck is cut into a packets by a multinomial distribution, and then cards are dropped from packets with probability proportional to packet size. Letting $Q_a(\sigma)$ denote this measure, they show

$$Q_a * Q_b = Q_{ab}. \tag{4}$$

Thus it is enough to study a single a-shuffle. The main result of their paper is the simple formula

$$Q_a(\sigma) = \frac{\binom{n+a-r}{n}}{a^n}, \tag{5}$$

where $r = r(\sigma)$ is the number of rising sequences in σ ($r(\sigma) = d(\sigma^{-1}) + 1$ with d the number of descents in σ). This allows simple closed form expressions for a variety of distances.

A number of extensions and variations have since developed. We will not survey these here (see [11] for a thorough treatment) but mention that features of permutations are shown to achieve the correct limiting distribution in fewer shuffles. For example, $\frac{5}{6}\log_2 n + c$ suffice for the longest increasing subsequences [17], $\log_2(n)$ for the descent structure [12], $k_n \to \infty$ arbitrarily slowly for the cycle structure [14] and a single shuffle suffices for the longest cycle [14]. A recent addition is the work of Chen and Saloff-Coste [6] studying random combinations of a-shuffles for randomly varying a.

Mark Conger and D. Viswanath study the same type of problems as we do. In [8], they lay out the basic problems, develop a formalism for calculations involving descent polynomials (a generalization of Eulerian polynomials), and use these to derive a closed formula for the chance of a given arrangement after an a-shuffle for decks labelled $\{1, 2, \ldots, h, x^n\}$. This includes both our single card case and the full deck case. They show that the probability of an arrangement is

$$\frac{1}{a^{n+h}} \sum_{m=r-1}^{a-1} \binom{m-r+h}{h-1} (a-m-1)^l (a-m)^{n-l}, \tag{6}$$

with r the number of cards labelled c, $1 \le c \le h$, that are not preceded by a card labelled $c-1$ and l the number of cards labeled x that precede the card labeled h. This elegant expression can be analyzed asymptotically using the analytic techniques of Sections 2-5 below. Their main results pertain to red-black decks where they derive equivalence relations on configurations that have the same probability. They point out that starting with the reds on top or reds alternating with blacks can lead to different conclusions.

In [9], the authors use their earlier work on descent polynomials to develop a fascinating Monte Carlo procedure for approximating the total variation distance. Our exact and asymptotic calculations overlap theirs in many places, and in every case we find their numbers spot on. This leads us to accept their estimates for problems of deck hands at bridge where we have not found a way to do exact calculations.

The results derived here add to the result of Conger-Viswanath in the following ways. First, we present some new formulae (e.g. the transition matrix for single card mixing or the red-black formula) which allow exact computations. Second, we derive asymptotic approximations for a variety of cases. There are no such computations in previous work. Third, we have made sense of this sea of formulae and approximations through our rule of thumb.

Finally, we mention the broad extensions of riffle shuffling to random walks on hyperplane arrangements due to Bidigare, Hanlon and Rockmore (see [11] for a survey). The process induced by observing which chamber of a sub-arrangement contains the present state of the original walk is still Markov. One might try to solve the problems of rates of convergence for selected features for any of these extensions.

2. Following a single card

Suppose one notices that the ace of spades is on the bottom of a deck of n cards. How many shuffles does it take until this one card is close to uniformly distributed on $\{1, 2, \dots, n\}$? As shown in an appendix, under repeated shuffles a single card moves according to a Markov chain. We begin by writing down the transition matrix.

Proposition 2.1. *Let $P_a(i, j)$ be the chance that the card at position i is moved to position j after an a-shuffle. Then for $1 \le i, j \le n$, we have*

$$P_a(i,j) = \frac{1}{a^n} \sum_{k=1}^{a} \sum_{r=l}^{u} \binom{j-1}{r} \binom{n-j}{i-r-1} k^r (a-k)^{j-1-r} (k-1)^{i-1-r} (a-k+1)^{(n-j)-(i-r-1)},$$

where the inner sum is from $l = \max(0, (i+j)-(n+1))$ to $u = \min(i-1, j-1)$.

Proof. We calculate $Q_a(j, i)$, the chance that an inverse a-shuffle brings the card at position j to position i. For this to occur, the card at position j may be labelled by k, $1 \le k \le a$. The r cards above this card may be labelled from 1 to k. All will appear before the card at position j in $\binom{j-1}{r}$ ways. The remaining cards above must labelled from $k+1$ to a. Here $0 \le r \le \min(j-1, i-1)$. Also if m cards below position j are labelled from 1 to $k-1$, then $m + r = i - 1, m < n - j$ and so $r \ge (i+j)-(n+1)$. Finally, $i-1-r$ cards below position j must be labelled from 1 to $k-1$ in $\binom{n-j}{i-r-1}$ ways, and the remaining cards must be labelled from $k+1$ to a. □

For example, the $n \times n$ transition matrices for $n = 2, 3$ are given below.

$$\frac{1}{2a} \begin{pmatrix} a+1 & a-1 \\ a-1 & a+1 \end{pmatrix} \qquad \frac{1}{6a^2} \begin{pmatrix} (a+1)(2a+1) & 2(a^2-1) & (a-1)(2a-1) \\ 2(a^2-1) & 2(a^2+2) & 2(a^2-1) \\ (a-1)(2a-1) & 2(a^2-1) & (a+1)(2a+1) \end{pmatrix}$$

Two other special cases to note are the extreme cases when $i = 1$ or $i = n$, which are given by

$$P_a(1,j) = \frac{1}{a^n} \sum_{k=1}^{a} (a-k)^{j-1} (a-k+1)^{n-j}, \qquad P_a(n,j) = \frac{1}{a^n} \sum_{k=1}^{a} k^{j-1} (k-1)^{n-j}.$$

These single card transition matrices are studied by Ciucu [7] who gives a closed form for all n when $a = 2$:

$$P_2(i,j) \;=\; \begin{cases} \frac{1}{2^n}\left(2^{i-1} + 2^{n-i}\right) & \text{if } i = j, \\ \frac{1}{2^{n-j+1}} \binom{n-j}{i-1} & \text{if } i > j, \\ P_2(n-i+1, n-j+1) & \text{if } i < j. \end{cases}$$

These matrices share many properties of the 'amazing matrix' developed by Holte [20]. The following Proposition is essentially due to Ciucu [7].

Proposition 2.2. *The transition matrices following a single card have the following properties:*

1. *they are* cross-symmetric*, i.e. $P_a(i, j) = P_a(n-i+1, n-j+1)$;*
2. *$P_a \cdot P_b = P_{ab}$;*
3. *the eigenvalues are $1, 1/a, 1/a^2, \dots, 1/a^{n-1}$;*
4. *the right eigen vectors are independent of a and have the simple form: $V_m(i) = (i-1)^{i-1}\binom{m-1}{i-1} + (-1)^{n-i+m}\binom{m-1}{n-i}$ for $1/a^m$, $m \ge 1$.*

Proof. The cross-symmetry (1) follows from Proposition 2.1, and the multiplicative property (2) follows from the shuffling interpretation and equation (4). Property (1) implies that the eigen structure is quite constrained; see [23]. Properties (3) and (4) follow from results of Cuicu [7]. □

Remark 2.3. We note that Holte's matrix arose from studying the 'carries process' of ordinary addition. Diaconis and Fulman [12] show that it is also the transition matrix for the number of descents in repeated a-shuffles. We have not been able to find a closer connection between the two matrices.

¿From Proposition 2.1 we obtain the following Corollary, which also follows as a special case of Theorem 2.2 in [8].

Corollary 2.4. *Consider a deck of n cards with the ace of spades starting at the bottom. Then the chance that the ace of spades is at position j from the top after an a-shuffle is*

$$Q_a(j) \ = P_a(n,j) \ = \ \frac{1}{a^n}\sum_{k=1}^{a}(k-1)^{n-j}k^{j-1}. \tag{7}$$

¿From the explicit formula, we are able to give exact numerical calculations and sharp asymptotics for any of the distances to uniformity. The results below show that $\log_2 n + c$ shuffles are necessary and sufficient for both separation and total variation (and there is a cutoff for these). This is surprising since, on the full permutation group, separation requires $2\log_2 n + c$ steps whereas total variation requires $\frac{3}{2}\log_2 n + c$. Of course, for any specific n, these asymptotic results are just indicative.

TABLE 2. Distance to uniformity for a deck of 52 distinct cards.

	1	2	3	4	5	6	7	8	9	10	11	12
TV	1.00	1.00	1.00	1.00	.924	.614	.334	.167	.085	.043	.021	.010
SEP	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.996	.931	.732	.479	.278
l_∞	9×10^{53}	1×10^{41}	3×10^{29}	7×10^{19}	3×10^{12}	2×10^{7}	1×10^{5}	128.5	11.3	2.57	.900	.380

TABLE 3. Distance to uniformity for a single card starting at the bottom of a 52 card deck.

	1	2	3	4	5	6	7	8	9	10	11	12
TV	.873	.752	.577	.367	.200	.103	.052	.026	.013	.007	.003	.002
SEP	1.00	1.00	.993	.875	.605	.353	.190	.098	.050	.025	.013	.006
l_∞	25.0	12.0	5.51	2.37	1.02	.460	.217	.105	.052	.026	.013	.006

Remarks on Table 3. We use Proposition 2.1 to give exact results when $n = 52$. For comparison, Table 2 gives exact results for the full deck using [3]. Tables 3 and 4 show that it takes about half as many or fewer shuffles to achieve a given degree of mixing for a card at the bottom of the deck. For example, the widely cited '7 shuffles' for total variation drops this distance to .334 for the full ordering, but this requires only 4 shuffles to achieve a similar degree of randomness for a single card at the bottom, and only 2 for a single card starting in the middle. Similar statements hold for the separation and l_∞ metrics.

TABLE 4. Distance to uniformity for a single card starting at the middle of a 52 card deck.

	1	2	3	4
TV	.494	.152	.001	.000
SEP	1.00	.487	.003	.000
l_∞	1.92	.487	.003	.000

For asymptotic results, we first derive an approximation to separation. Since separation is an upper bound for total variation, this gives an upper bound for total variation. Finally, we derive a matching lower bound for total variation.

Proposition 2.5. *After an a-shuffle, the probability that the bottom card is at position i satisfies*

$$\frac{1}{a}\frac{\alpha^{n-i+1}}{1-\alpha^n} \le Q_a(i) \le \frac{1}{a}\frac{\alpha^{n-i}}{1-\alpha^{n-1}},$$

where for brevity we have set $\alpha = 1-1/a$. In particular, the separation distance satisfies

$$1-\frac{n}{a}\frac{\alpha^n}{1-\alpha^n} \le \mathrm{SEP}(a) \le 1-\frac{n}{a}\frac{\alpha^{n-1}}{1-\alpha^{n-1}}.$$

Proof. Since $k/(k-1) \ge a/(a-1)$ for all $1 < k \le a$ we find that

$$\alpha^{n-i}Q_a(n) \ge Q_a(i) \ge \alpha^{-(i-1)}Q_a(1). \tag{8}$$

Therefore

$$1 = \sum_i Q_a(i) \ge Q_a(1)\sum_{i=1}^{n}\alpha^{-(i-1)} = Q_a(1)a\alpha^{1-n}(1-\alpha^n),$$

so that

$$Q_a(1) \le \frac{1}{a}\frac{\alpha^{n-1}}{1-\alpha^n} \le \frac{1}{a}\frac{\alpha^{n-1}}{1-\alpha^{n-1}}.$$

Since $Q_a(n) = Q_a(1) + 1/a$ it follows that $Q_a(n) \le \frac{1}{a}\frac{1}{1-\alpha^{n-1}}$. Using (8) the desired upper bound for $Q_a(i)$ follows.

Similarly,

$$1 = \sum_i Q_a(i) \le Q_a(n)\sum_{i=1}^{n}\alpha^{n-i} = Q_a(n)\frac{1-\alpha^n}{1-\alpha},$$

so that

$$Q_a(n) \ge \frac{1}{a}\frac{1}{1-\alpha^n}.$$

Since $Q_a(1) = Q_a(n) - 1/a$ it follows that $Q_a(1) \ge \frac{1}{a}\frac{\alpha^n}{1-\alpha^n}$, and from (8) the desired lower bound for $Q_a(i)$ follows. From (20) and the above estimates we obtain our bounds on $\mathrm{SEP}(a)$. □

If $a = 2^{\log_2(n)+c} = n2^c$, then our result shows that the $\mathrm{SEP}(a)$ is approximately

$$1-\frac{1}{2^c}\frac{e^{-2^{-c}}}{1-e^{-2^{-c}}},$$

and for large c this is $\approx 2^{-c-1}$. The fit to the data in Table 5 is excellent: for example after ten shuffles of a fifty-two card deck we have $2^{-c-1} = \frac{26}{1024}$ which is very nearly the observed separation distance of 0.025.

Remark 2.6. Proposition 2.5 gives a local limit for the probability that the original bottom card is at position j *from the bottom*. When the number of shuffles is $\log_2 n + c$, the density of this (with respect to the uniform measure) is asymptotically $z(c)e^{-j/2^c}$, with z a normalizing constant ($z(c) = 1/2^c(e^{j/2^c} - 1)$). The result is uniform in j for c fixed, n large.

Proposition 2.7. *Consider a deck of n cards with the ace of spades at the bottom. With $\alpha = 1-1/a$, the total variation distance for the mixing of the ace of spades after an a-shuffle is at most*

$$\frac{\alpha^{n+1}}{1-\alpha^n} - \frac{a\alpha^2(1-\alpha^{n-1})}{n(1-\alpha^n)} + \frac{1}{n\log(1/\alpha)}\log\left(\frac{a}{n}\frac{1-\alpha^n}{\alpha^{n+1}}\right),$$

and at least

$$\frac{\alpha^{n}}{1-\alpha^{n-1}} - \frac{a(1-\alpha^{n})}{n\alpha(1-\alpha^{n-1})} + \frac{1}{n\log(1/\alpha)}\log\Big(\frac{a}{n}\frac{1-\alpha^{n-1}}{\alpha^{n-1}}\Big).$$

Proof. Let $Q_a(i)$ denote the probability that the ace of spades is at position i from the top after an a shuffle. Note that $Q_a(i)$ is monotone increasing in i, and let i^* be such that $Q_a(i^*) < 1/n \leq Q_a(i^*+1)$. From Proposition 2.5 we find that i^* satisfies

$$\frac{\alpha^{n-i^*+1}}{a(1-\alpha^n)} < \frac{1}{n} \leq \frac{\alpha^{n-i^*-1}}{a(1-\alpha^{n-1})}, \tag{9}$$

so that

$$\log\left(\frac{a}{n}\frac{1-\alpha^{n-1}}{\alpha^{n-1}}\right) \leq i^*(\log 1/\alpha) \leq \log\left(\frac{a}{n}\frac{1-\alpha^{n}}{\alpha^{n+1}}\right) \tag{10}$$

From Proposition 2.5 we have that the desired total variation is

$$\sum_{i\leq i^*}\Big(\frac{1}{n} - Q_a(i)\Big) \leq \frac{i^*}{n} - \sum_{i\leq i^*}\frac{\alpha^{n-i+1}}{a(1-\alpha^n)} = \frac{i^*}{n} - \frac{\alpha^{n-i^*+1}}{1-\alpha^n}(1-\alpha^{i^*}),$$

and also

$$\sum_{i\leq i^*}\Big(\frac{1}{n} - Q_a(i)\Big) \geq \frac{i^*}{n} - \frac{\alpha^{n-i^*}}{1-\alpha^{n-1}}(1-\alpha^{i^*}).$$

Using (9) and (10) we obtain the Proposition. □

Remark 2.8. After $\log_2 n + c$ shuffles, that is when $a = 2^c n$, Proposition 2.7 shows that the total variation distance is approximately (with $C = 2^c$)

$$C\log\Big(C(e^{1/C}-1)\Big) + \frac{1 - C\log(e^{1/C}-1)}{(e^{1/C}-1)}.$$

Thus when c is 'large and negative,' the total variation is close to 1, and when c is large and positive, the total variation is close to 0. Thus total variation and separation converge at the same rate. This is an asymptotic result and, for example, Table 3 supports this.

Remark 2.9. From Proposition 4.1, the l_∞ distance is achieved for configurations with the ace of spades back on the bottom. Proposition 2.5 gives a formula for this and the arguments for Propositions 2.5 and 2.7 show that $\log_2 n + c$ shuffles are necessary and sufficient for convergence in l_∞.

Remark 2.10. Similar, but more demanding, calculations show that if the ace of spades starts at position i, and $\max(i/n, (n-i)/n) \geq A > 0$ for some fixed positive A, then $\frac{1}{2}\log_2 n$ shuffles suffice for convergence in any of the metrics. We omit further details.

3. A red-black deck

We focus now on riffle shuffles of a deck consisting of R red cards and B black cards. The purpose of this section is to give an explicit description of a-shuffles of the deck with initial configuration of red atop blacks. In Bayer-Diaconis [3], the formula describing when an a-shuffle of n distinct cards results in a particular permutation has the simple form

$$\frac{1}{a^n}\binom{a+n-r}{n},$$

where r is the number of rising sequences in the permutation. The analysis for the red-black deck is markedly different. One indication of this comes by noticing how likely the reverse deck is to occur. In the case of permutations, the reverse deck has n rising sequences, and so the Bayer-Diaconis formula dictates that this configuration cannot occur unless $a \geq n$. However, in the red-black case, the reverse deck (blacks atop reds) may occur after a single 2-shuffle no matter the deck size.

Theorem 3.1. *Consider a deck with R red cards on top of B black cards. The probability that an a-shuffle will result in the deck configuration w is*

$$Q_a(w) = \frac{1}{a^{R+B}} \sum_{k=1}^{a} \sum_{j=1}^{R} (k-1)^{R-j} k^{j-1} (a-k)^{b(j)} (a-k+1)^{B-b(j)} \tag{11}$$

where $b(j) = b_w(j)$ is the number of black cards above the jth red card in the deck w.

Proof. The general formula for the probability of w resulting from an a-shuffle is given by

$$\sum_{A_1+\cdots+A_a=R+B} \frac{1}{a^n}\binom{R+B}{A_1,\ldots,A_a}\mathrm{prob}(w|A), \tag{12}$$

where the sum is over all non-negative compositions $A = (A_1, A_2, \ldots, A_a)$ of $R+B$, i.e $A_i \geq 0$ and $A_1 + A_2 + \cdots + A_a = R+B$, and $\mathrm{prob}(w|A)$ denotes the probability that w results from successively dropping cards from the piles A_i. We break the sum into the following two cases: either there exists an integer k such that $A_1 + A_2 + \cdots + A_k = R$ or not.

Consider the case when the sum of the first k piles is exactly R. Then, the result of the subsequent riffle shuffle is equally likely to be any of the $\binom{R+B}{R}$ possible deck configurations. That is to say, given such a cut A, $\mathrm{prob}(w|A) = 1/\binom{R+B}{R}$ for every w. Therefore the contribution to $Q_a(w)$ from all such cuts is given by

$$\begin{aligned}
&\sum_{\substack{A_1+\cdots+A_a=R+B \\ \exists\ k \text{ s.t. } A_1+\cdots+A_k=R}} \frac{1}{a^{R+B}}\binom{R+B}{A_1,\ldots,A_a}\frac{1}{\binom{R+B}{R}} \;= \\
&\quad= \frac{1}{a^{R+B}} \sum_{k=1}^{a-1} \sum_{A_k=1}^{R} \sum_{\substack{A_{k+1}+\cdots+A_a=B \\ A_1+\cdots+A_{k-1}=R-A_k}} \binom{R}{A_k}\binom{R-A_k}{A_1,\ldots,A_{k-1}}\binom{B}{A_{k+1},\ldots,A_a} \\
&\quad= \frac{1}{a^{R+B}} \sum_{k=1}^{a-1} (a-k)^B \sum_{A_k=1}^{R} \binom{R}{A_k}(k-1)^{R-A_k} = \frac{1}{a^{R+B}} \sum_{k=1}^{a-1} (a-k)^B \left(k^R - (k-1)^R\right).
\end{aligned} \tag{13}$$

The choice to let k be the *first* index such that $A_1 + \cdots + A_k = R$ is necessary in order to avoid over counting compositions with many 0's. This choice seemingly breaks the symmetry between R and B in the final formulation. However, the symmetric version may be obtained by taking k to be the *last* index such that $A_1 + \cdots + A_k = R$. Finally, note that since $B \neq 0$, we may in fact take the sum over k to range from 1 to a.

Now consider the alternative case when there exists a pile (necessarily unique) containing both red and black cards. The assumption on A amounts to the existence of integers k, x, y, with

$1 \le k \le a$, $1 \le x \le R$, $1 \le y \le B$, such that $A_1 + \cdots + A_{k-1} = R - x$, $A_k = x + y$, and $A_{k+1} + \cdots + A_a = B - y$. Given such a cut A, $\text{prob}(w|A) = r_{x,y}(w)/\binom{R+B}{R-x,x+y,B-y}$, where $r_{x,y}(w)$ denotes the number of rising subsequences consisting of x red cards followed by y black cards. The resulting contribution to $Q_a(w)$ from all such cuts is given by

$$\begin{aligned}(14)\qquad &\sum_{\substack{A_1+\cdots+A_a=R+B \\ \exists\ k \text{ s.t. } A_1+\cdots+A_{k-1}<R \\ \text{and } A_{k+1}+\cdots+A_a<B}} \frac{1}{a^{R+B}}\binom{R+B}{A_1,\ldots,A_a}\text{prob}(w|A) \\ &= \frac{1}{a^{R+B}}\sum_{k=1}^{a}\sum_{x=1}^{R}\sum_{y=1}^{B} r_{x,y}(w) \sum_{\substack{A_1+\cdots+A_{k-1}=R-x \\ A_{k+1}+\cdots+A_a=B-y}} \binom{R-x}{A_1,\ldots,A_{k-1}}\binom{B-y}{A_{k+1},\ldots,A_a} \\ &= \frac{1}{a^{R+B}}\sum_{k=1}^{a}\sum_{x=1}^{R}\sum_{y=1}^{B} r_{x,y}(w)(k-1)^{R-x}(a-k)^{B-y}.\end{aligned}$$

For the final equation to make sense, we adopt the convention that $0^0 = 1$.

Let $b(j)$ denote the number of black cards above the jth red card in w. We may count rising subsequences of w by the last red card used in the subsequence, giving the equation

$$(15)\qquad r_{x,y}(w) = \sum_{j=1}^{R}\binom{j-1}{x-1}\binom{B-b(j)}{y}.$$

To see this, note that the first binomial coefficient counts the number choices of x red cards before the jth red card, and the second binomial coefficient counts the number of choices for y black cards after the jth red card. Inserting this into the x and y summations of (14) gives

$$\begin{aligned}(16)\qquad &\frac{1}{a^{R+B}}\sum_{k=1}^{a}\sum_{x=1}^{R}\sum_{y=1}^{B} r_{x,y}(w)(k-1)^{R-x}(a-k)^{B-y} \\ &= \frac{1}{a^{R+B}}\sum_{k=1}^{a}\sum_{j=1}^{R}\left(\sum_{x=0}^{R-1}\binom{j-1}{x}(k-1)^{R-x-1}\right)\left(\sum_{y=1}^{B}\binom{B-b(j)}{y}(a-k)^{B-y}\right) \\ &= \frac{1}{a^{R+B}}\sum_{k=1}^{a}\sum_{j=1}^{R}\left((k-1)^{R-j}k^{j-1}\right)\left((a-k)^{b(j)}\left((a-k+1)^{B-b(j)}-(a-k)^{B-b(j)}\right)\right).\end{aligned}$$

The probability $Q_a(w)$ is obtained by adding the expressions in (13) and (16). Since

$$\begin{aligned}\sum_{k=1}^{a}\sum_{j=1}^{R}(k-1)^{R-j}k^{j-1}(a-k)^{B} &= \sum_{k=1}^{a}k^{R-1}(a-k)^{B}\sum_{j=1}^{R}\left(\frac{k-1}{k}\right)^{R-j} \\ &= \sum_{k=1}^{a}k^{R-1}(a-k)^{B}\frac{1-(1-1/k)^{R}}{1-(1-1/k)} = \sum_{k=1}^{a}(a-k)^{B}(k^{R}-(k-1)^{R}),\end{aligned}$$

we obtain the desired expression. □

Given (15), Q_a gives a completely explicit description of a-shuffles, though this is difficult to evaluate for an arbitrary w. However, there are two special deck configurations for which Q_a simplifies nicely, namely reds atop blacks (where $r_{x,y}(w) = \binom{R}{x}\binom{B}{y}$) and blacks atop reds (where $r_{x,y}(w) = 0$). By Proposition 4.1, the formulae below can be used to give exact calculations for separation and l_∞.

Corollary 3.2. *The probability of an a-shuffle resulting in the original deck configuration of reds atop blacks is*

$$\frac{1}{a^{R+B}}\left(\sum_{k=1}^{a}\left(k^R-(k-1)^R\right)(a-k+1)^B\right).$$

The probability an a-shuffle resulting in the reverse deck configuration of blacks atop reds is

$$\frac{1}{a^{R+B}}\sum_{k=1}^{a-1}(a-k)^B\left(k^R-(k-1)^R\right).$$

Another special case to consider is tracking the position of a single card starting at the bottom of the deck. For this case, taking $B=1$ and $R=n-1$ in (11) we recover Corollary 2.4.

Note that if instead we consider a single red card, i.e. $R=1$ and $B=n-1$, starting at the top, then the distribution is the same. More precisely, let $\widetilde{Q}_a(i)$ denote the chance that, say, the 2 of hearts is at position i from the top of the deck after an a-shuffle. Then it is easy to verify that $Q_a(i)=\widetilde{Q}_a(n-i+1)$, which is just a special case of the cross-symmetry in Proposition 2.2.

Finally, consider the case of a single 2-shuffle for an arbitrary red-black deck. In this case, the left hand summand of (11) reduces to a single term evaluating to 1. For the right hand summand, note that $k=1$ forces $x=R$, and $k=a$ forces $y=B$.

Corollary 3.3. *The probability of a 2-shuffle resulting in a deck configuration w is*

$$Q_2(w)=\frac{1}{2^{R+B}}\left(2^{\mathrm{h}(w)}+2^{\mathrm{t}(w)}-1\right), \tag{17}$$

where $\mathrm{h}(w)$ denotes the number of red cards preceding the first black card in w, and $\mathrm{t}(w)$ denotes the number of black cards following the final red card of w.

Equation (17) can be used to give a simple formula for the total variation after a single 2-shuffle of a deck with n red cards and n black cards. Here note that any two configurations with the same number of red cards on top and black cards on bottom has the same likelihood of occurrence, so we get

$$\|Q_2-U\|_{TV}=\frac{1}{2}\left(\left(\frac{2^{n+1}-1}{2^{2n}}-\frac{1}{\binom{2n}{n}}\right)+\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\left|\frac{2^i+2^j-1}{2^{2n}}-\frac{1}{\binom{2n}{n}}\right|\binom{2n-(i+j+2)}{n-(i+1)}\right) \tag{18}$$

Using this formula, the total variation after a single 2-shuffle of a deck with 26 red and 26 black cards is 0.579, which agrees with the numerical approximations of Conger and Viswanath in [8]. We do not see how to compute total variation effectively after more shuffles.

Asymptotic results for the separation distance for red-black configurations appear in the following section.

4. Approach to uniformity in separation for general decks

In this section we work with general decks containing D_i cards labelled i, $1\le i\le m$. The following lemma shows that the separation distance is always achieved by reversing the initial deck configuration. Note this is equivalent to Theorem 2.1 from [8].

Proposition 4.1. *Let D be a deck as above. After an a-shuffle of the deck with 1's on top down to m's on bottom, the most likely deck configuration is this initial deck and the least likely configuration is the reverse deck w^* with m's on top down to 1's on the bottom. In particular, the separation distance is achieved for w^*.*

Proof. Note first that the initial configuration can result from any possible cut of the deck into a piles. Moreover, from any given cut of the deck, the identity is at least as likely to occur as any other configuration. The first assertion now follows. The only cuts of the initial deck which may result in w^* are those containing no pile with distinct letters. However, for all such cuts, each rearrangement of the deck is equally likely to occur. Therefore w^* minimizes $Q_a(w)$ and so maximizes $1 - Q_a(w)/U$. □

The explicit formula for $Q_a(w^*)$ given in Corollary 3.2 facilitates exact computations of SEP(a) for decks of practical interest. Similarly, we can compute $Q_a(w^*)$ for an arbitrary deck with D_i i's, $i = 1, \ldots, m$.

Theorem 4.2. *Consider a deck with n cards and D_i cards labeled i, $i = 1, \ldots, m$. Then the separation distance after an a-shuffle of the sorted deck (1's followed by 2's, etc) is given by*

$$\mathrm{SEP}(a) = 1 - \frac{1}{a^n}\binom{n}{D_1 \ldots D_m} \sum_{0=k_0<\cdots<k_{m-1}<a} (a-k_{m-1})^{D_m} \prod_{j=1}^{m-1} \left((k_j - k_{j-1})^{D_j} - (k_j - k_{j-1} - 1)^{D_j}\right).$$

Proof. From Proposition 4.1, w^* may only result from cuts with no pile containing distinct cards and any such cut is equally like to result in any deck. Therefore $Q_a(w^*)$ is given by

$$Q_a(w^*) = \sum_{\substack{A_1+\cdots+A_a=n \\ A \text{ refines } D}} \frac{1}{a^n}\binom{n}{A_1, \ldots, A_a} \frac{1}{\binom{n}{D_1,\ldots,D_m}},$$

where 'A refines D' means there exist indices $k_1, \ldots, k_{m-1}$ such that $A_1 + \cdots + A_{k_1} = D_1$ and, for $i = 2, \ldots, m-1$, $A_{k_{i-1}+1} + \cdots + A_{k_i} = D_i$. Just as in the proof of Theorem 3.1 we may take the k_i's to be minimal so that the expression for $Q_a(w^*)$ simplifies to give

$$Q_a(w^*) = \frac{1}{a^n} \sum_{0=k_0<k_1<\cdots<k_{m-1}<a} (a-k_{m-1})^{D_m} \prod_{j=1}^{m-1} \left((k_j - k_{j-1})^{D_j} - (k_j - k_{j-1} - 1)^{D_j}\right). \tag{19}$$

The result now follows from Proposition 4.1. □

TABLE 5. Separation distance for k shuffles of 52 cards.

k	1	2	3	4	5	6	7	8	9	10	11	12
Bayer-Diaconis	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.995	.928	.729	.478	.278
blackjack	1.00	1.00	1.00	1.00	.999	.970						
♣♢♡♠	1.00	.997	.997	.976	.884	.683	.447	.260	.140	.073		
A♠	1.00	1.00	.993	.875	.605	.353	.190	.098	.050	.025	.013	.006
redblack	.890	.890	.849	.708	.508	.317	.179	.095	.049	.025	.013	.006
○+≋□☆	1.00	1.00	.993	.943	.778	.536	.321	.177				

Remarks on Table 5. We calculate SEP after repeated 2-shuffles for various decks using Theorem 4.2: (blackjack) 9 ranks, say A23456789, with 4 cards each and another rank, say 10, with 16 cards; (♣♢♡♠) 4 distinct suits, say clubs, diamonds, hearts and spades, of 13 cards each; (A♠)the ace of spades and 51 other cards; (redblack) a two color deck with 26 red and 26 black cards; and (○+≋□☆) a deck with 5 cards in each of 5 suits. The missing entries in Table 5 highlight the limitations of exact calculations using Theorem 4.2.

Proposition 4.1 may be used with the Conger-Viswanath formula in (6) to give a simple expression for separation after an a-shuffle for a deck of size $h+n$ with cards labelled $1, 2, \dots, h$ and n cards labelled x:

$$\mathrm{SEP}(a) = 1 - \frac{(n+h)\cdots(n+1)}{a^{n+h}} \sum_{k=h-1}^{a-1} \binom{k}{h-1}(a-1-k)^n.$$

Now we derive a basic asymptotic tool, Proposition 4.3, which allows asymptotic approximations for general decks. As motivation, consider again the case of one card mixing, i.e. begin with n cards with the ace of spaces at the bottom of the initial deck. How many shuffles are required to randomize the ace of spades? Recall from Corollary 2.4 that the chance that the ace of spades is at position i from the top after an a-shuffle is given by

$$Q_a(i) \;=\; \frac{1}{a^n}\sum_{k=1}^{a}(k-1)^{n-i}k^{i-1},$$

with the convention $0^0 = 1$. Therefore from Proposition 4.1, we have

$$\mathrm{SEP}(a) = 1 - nQ_a(1) = 1 - \frac{n}{a^n}\sum_{k=1}^{a}(k-1)^{n-1}. \tag{20}$$

Exact calculations when $n = 52$ are given in Table 5.

Proposition 4.3. *Let a be a positive real number, and let r and s be natural numbers with $r,\ s \geq 2$. Let ξ be a real number in $[0,1]$. Then*

$$\begin{aligned} S(a,\xi;r,s) \;&:=\; \frac{1}{a^{r+s}} \sum_{0\leq k\leq a-\xi} (k+\xi)^r(a-k-\xi)^s \\ &=\; a\frac{r!s!}{(r+s+1)!} + \frac{\theta}{6a}\frac{r!s!}{(r+s-1)!}\Big(\frac{1}{r-1}+\frac{1}{s-1}\Big), \end{aligned}$$

where θ is a real number in $[-1,1]$.

Proof. Put $f(x) = x^r(1-x)^s$ for $x \in [0,1]$ and $f(x) = 0$ otherwise. The sum that we wish to evaluate is

$$\sum_{k\in\mathbb{Z}} f((k+\xi)/a) = a\sum_{\ell\in\mathbb{Z}} \hat{f}(a\ell)e(\ell\xi), \tag{21}$$

by the Poisson summation formula. Here, we write $e(x) = e^{2\pi i x}$ and $\hat{f}(y) = \int_{-\infty}^{\infty} f(x)e(-xy)dx$ denotes the Fourier transform.

Now note that

$$\hat{f}(0) = \int_0^1 x^r(1-x)^s dx = \frac{r!s!}{(r+s+1)!}. \tag{22}$$

Further

$$\hat{f}(y) = \int_0^1 x^r(1-x)^s e^{-2\pi i x y}dx = \frac{1}{2\pi i y}\int_0^1 f'(x)e^{-2\pi i x y}dx = \frac{1}{(2\pi i y)^2}\int_0^1 f''(x)e^{-2\pi i x y}dx,$$

upon integrating by parts twice, and since $r,\ s \geq 2$ we have $f(0) = f'(0) = f(1) = f'(1) = 0$. Therefore

$$|\hat{f}(y)| \leq \frac{1}{4\pi^2 y^2}\int_0^1 |f''(x)|dx.$$

Now

$$f''(x) = \Big(\frac{r}{x} - \frac{s}{1-x}\Big)^2 x^r(1-x)^s - \Big(\frac{r}{x^2} + \frac{s}{(1-x)^2}\Big)x^r(1-x)^s,$$

and so

$$\begin{aligned}\int_0^1 |f''(x)|dx &\leq \int_0^1 \Big(\frac{r}{x}-\frac{s}{1-x}\Big)^2 x^r(1-x)^s dx + \int_0^1 \Big(\frac{r}{x^2}+\frac{s}{(1-x)^2}\Big)x^r(1-x)^s dx \\ &= \frac{r!s!}{(r+s-1)!}\Big(\frac{2}{r-1}+\frac{2}{s-1}\Big).\end{aligned}$$

Combining the above estimates with (21) and (22) we conclude that our sum equals

$$a\frac{r!s!}{(r+s+1)!}+\frac{\theta}{2\pi^2 a}\frac{r!s!}{(r+s-1)!}\left(\frac{1}{r-1}+\frac{1}{s-1}\right)\sum_{\substack{\ell\in\mathbb{Z}\\ \ell\neq 0}}\frac{1}{\ell^2}$$

for some $\theta\in[-1,1]$. Since $\sum_{\ell=1}^{\infty}\ell^{-2}=\pi^2/6$ the Proposition follows. □

Now suppose we have n red cards and n black cards, so $2n$ cards altogether, with the red cards starting on top. In this case, the uniform distribution $U(w)=U=1/\binom{2n}{n}$. Again we use Proposition 4.1 this time with Corollary 3.2 to give a formula for the separation distance,

$$\mathrm{SEP}(a)=1-\binom{2n}{n}Q_a(w^*)=1-\frac{\binom{2n}{n}}{a^{2n}}\sum_{k=1}^{a-1}(a-k)^n\left(k^n-(k-1)^n\right) \tag{23}$$

For exact computations when $2n=52$, see Table 5. We now use Proposition 4.3 to calculate asymptotic expressions for this separation distance.

Corollary 4.4. *For $2n$ cards starting with n red cards on top, we have, with $\alpha=1-1/a$*

$$\mathrm{SEP}(a)=1-\frac{a}{2n+1}(1-\alpha^{2n+1})+\frac{2\theta}{3a}\frac{n}{(n-2)}(1-\alpha^{2n-1}),$$

for some real number $\theta\in[-1,1]$. In particular, for n large with $a=2^{\log_2(2n)+c}$,

$$\mathrm{SEP}(a)=1-2^c\left(1-e^{-2^{-c}}\right)+O\Big(\frac{1}{a}\Big).$$

Proof. Note that

$$\begin{aligned}\frac{1}{a^{2n}}\sum_{k=1}^{a}(a-k)^n(k^n-(k-1)^n) &= \frac{1}{a^{2n}}\sum_{k=1}^{a}(a-k)^n\int_0^1 n(k-1+\xi)^{n-1}d\xi \\ &= \frac{n}{a^{2n}}\int_0^1\sum_{k=0}^{a-1}(a-1+\xi-(k-1+\xi))^n(k-1+\xi)^{n-1}d\xi.\end{aligned}$$

Using Proposition 4.3 we see that the inner sum over k above equals

$$(a-1+\xi)^{2n}\frac{n!(n-1)!}{(2n)!}+(a-1+\xi)^{2n-2}\frac{\theta}{6}\frac{n!(n-1)!}{(2n-2)!}\Big(\frac{1}{n-1}+\frac{1}{n-2}\Big).$$

Using these observations in (23) we obtain that

$$\mathrm{SEP}(a)=1-\int_0^1\Big(\frac{a-1+\xi}{a}\Big)^{2n}d\xi+\frac{\theta}{6a^2}\frac{2n(2n-1)(2n-3)}{(n-1)(n-2)}\int_0^1\Big(\frac{a-1+\xi}{a}\Big)^{2n-2}d\xi.$$

With a little calculus the Corollary follows. □

The approximation

$$\binom{2n}{n}\sum_{k=1}^{a}(a-k)^n(k^n-(k-1)^n)\approx\frac{a^{2n+1}-(a-1)^{2n+1}}{2n+1} \tag{24}$$

which is the basis of our Corollary above is more accurate than suggested by the simple error bounds that we have given. For example, when $n=26$ and $a=16$, the actual separation distance (given in Table 5) differs from the approximation of the Corollary by about 7×10^{-12}. Put differently, note that the LHS and the RHS of (24) are both polynomials in a of degree $2n$, and in fact the coefficients of both polynomials match for all degrees between n and $2n$.

Before moving to general decks, we establish a generalization of Proposition 4.3.

Proposition 4.5. *Let $m\geq 2$ and a be natural numbers, let $\xi_1,\ldots,\xi_m$ be real numbers in $[0,1]$. Let $r_1,\ldots,r_m$ be natural numbers all at least $r\geq 2$. Let*

$$S_m(a;\underline{\xi},\underline{r})=\sum_{\substack{a_1,\ldots,a_m\geq 0\\ a_1+\ldots+a_m=a}}(a_1+\xi_1)^{r_1}\cdots(a_m+\xi_m)^{r_m}.$$

Then

$$\begin{aligned}\Big|S_m(a;\underline{\xi},\underline{r}) &- \frac{r_1!\cdots r_m!}{(r_1+\ldots+r_m+m-1)!}(a+\xi_1+\ldots+\xi_m)^{r_1+\ldots+r_m+m-1}\Big|\\ &\leq r_1!\cdots r_m!\sum_{j=1}^{m-1}\binom{m-1}{j}\Big(\frac{1}{3(r-1)}\Big)^j\frac{(a+\xi_1+\ldots+\xi_m)^{r_1+\ldots+r_m+m-1-2j}}{(r_1+\ldots+r_m+m-1-2j)!}.\end{aligned}$$

Proof. We establish this by induction on m. The case $m=2$ follows from Proposition 4.3, taking there a to be what we would now call $a+\xi_1+\xi_2$. Let now $m\geq 3$ and suppose the result has been established for $m-1$ variables. Now

$$S_m(a;\xi,r)=\sum_{a_1=1}^{a+\xi_2+\ldots+\xi_m-1}a_1^{r_1}S_{m-1}(a-a_1;\underline{\tilde{\xi}},\underline{\tilde{r}}) \tag{25}$$

with $\underline{\tilde{\xi}}=(\xi_2,\ldots,\xi_m)$ and $\underline{\tilde{r}}=(r_2,\ldots,r_m)$, and interpreting the terms with $a_1\geq a$ as being 0. Using the induction hypothesis we have that

$$\begin{aligned}&\Big|S_{m-1}(a-a_1;\underline{\tilde{\xi}},\underline{\tilde{r}})-\frac{r_2!\cdots r_m!}{(r_2+\ldots+r_m+m-2)!}(a-a_1+\xi_2+\ldots+\xi_m)^{r_2+\ldots+r_m+m-2}\Big| \\ &\leq r_2!\cdots r_m!\sum_{j=1}^{m-2}\binom{m-2}{j}\Big(\frac{1}{3(r-1)}\Big)^j\frac{(a-a_1+\xi_2+\ldots+\xi_m)^{r_2+\ldots+r_m+m-2-2j}}{(r_2+\ldots+r_m+m-2-2j)!}.\end{aligned} \tag{26}$$

Note that the above estimate is valid even if $a+\xi_2+\ldots+\xi_m-1\geq a_1\geq a$ since the RHS is larger than the main term that is being subtracted in the LHS. We use this estimate in (25), and then invoke Proposition 4.3 to handle each of the $m-1$ new sums that arise. Thus, the contribution of the main term in (26) is, for some $|\theta|\leq 1$,

$$\begin{aligned}\frac{r_1!\cdots r_m!}{(r_1+\ldots+r_m+m-1)!}&(a+\xi_1+\ldots+\xi_m)^{r_1+\ldots+r_m+m-1}\\ &+\frac{\theta}{3(r-1)}r_1!\cdots r_m!\frac{(a+\xi_1+\ldots+\xi_m)^{r_1+\ldots+r_m+m-3}}{(r_1+\ldots+r_m+m-3)!},\end{aligned}$$

while the j-th term on the RHS of (26) contributes

$$r_1!\cdots r_m!\binom{m-2}{j}\Big(\frac{1}{3(r-1)}\Big)^j \quad \Big(\frac{(a+\xi_1+\ldots+\xi_m)^{r_1+\ldots+r_m+m-1-2j}}{(r_1+\ldots+r_m+m-1-2j)!} + \frac{1}{3(r-1)}\frac{(a+\xi_1+\ldots+\xi_m)^{r_1+\ldots+r_m-1-2j-2}}{(r_1+\ldots+r_m+m-1-2j-2)!}\Big).$$

Using these in (26) and (25), and using the triangle inequality, and that $\binom{m-1}{j}=\binom{m-2}{j}+\binom{m-2}{j-1}$ we obtain the Proposition. □

Consider now a general deck of n cards with D_1 1's followed by D_2 2's and so on ending with D_m m's. Recall that the separation is maximum for the reverse configuration of the deck, and that probability is given in Theorem 4.2. We now use Proposition 4.5 to find asymptotics for that separation distance. The following is our 'rule of thumb.'

Theorem 4.6. *Consider a deck of n cards of m-types as above. Suppose that $D_i \geq d \geq 3$ for all $1 \leq i \leq m$. Then the separation distance is*

$$1-(1+\eta)\frac{a^{m-1}}{(n+1)\cdots(n+m-1)}\sum_{j=0}^{m-1}(-1)^j\binom{m-1}{j}\Big(1-\frac{j}{a}\Big)^{n+m-1},$$

where η is a real number satisfying

$$|\eta| \leq \Big(1+\frac{n^2}{3(d-2)(a-m+1)^2}\Big)^{m-1}-1.$$

Proof. Recall the expression for the separation distance given in Theorem 4.2. To evaluate this, we require an understanding of

$$\sum_{\substack{a_1+\ldots+a_m=a\\ a_j\geq 1}} a_m^{D_m}\prod_{j=1}^{m-1}(a_j^{D_j}-(a_j-1)^{D_j}) = \int_0^1\cdots\int_0^1\sum_{\substack{a_1+\ldots+a_m=a\\ a_j\geq 1}} a_m^{D_m}\prod_{j=1}^{m-1}\Big(D_j(a_j-1+\xi_j)^{D_j-1}d\xi_j\Big).$$

We now invoke Proposition 4.5. Thus the above equals for some $|\theta|\leq 1$

$$\prod_{j=1}^m D_j!\int_0^1\cdots\int_0^1\Big(\frac{(a-(m-1)+\xi_1+\ldots+\xi_{m-1})^n}{n!} +\theta\sum_{j=1}^{m-1}\binom{m-1}{j}\Big(\frac{1}{3(d-2)}\Big)^j\frac{(a-(m-1)+\xi_1+\ldots+\xi_{m-1})^{n-2j}}{(n-2j)!}\Big)d\xi_1\ \cdots\ d\xi_{m-1}.$$

We may simplify the above as

$$\Big(1+\theta\Big\{\Big(1+\frac{n^2}{3(d-2)(a-m+1)^2}\Big)^{m-1}-1\Big\}\Big)\frac{D_1!\cdots D_m!}{n!} \times \int_0^1\ldots\int_0^1(a-m+1+\xi_1+\ldots+\xi_{m-1})^n d\xi_1\cdots d\xi_{m-1},$$

and evaluating the integrals above this is

$$\Big(1+\theta\Big\{\Big(1+\frac{n^2}{3(d-2)(a-m+1)^2}\Big)^{m-1}-1\Big\}\Big)\frac{D_1!\cdots D_m!}{n!}\sum_{j=0}^{m-1}(-1)^j\binom{m-1}{j}(a-j)^{n-m+1}.$$

The Theorem follows. □

Remark 4.7. For simplicity we have restricted ourselves to the case when each pile has at least three cards. With more effort we could extend the analysis to include doubleton piles. The case of some singleton piles needs some modifications to our formula, but this variant can also be worked out.

Remark 4.8. From Theorem 4.6 one can show that for a general decks as above, one needs a of size about nm before the separation distance becomes small. We note that when a is of size about nm, the quantity η appearing in Theorem 4.6 is of size about $1/(m(d-2))$, so that the estimates furnished above represent a true asymptotic unless both m and d happen to be small. In other words, when we either have many piles, or a small number of thick piles, Theorem 4.6 gives a good asymptotic.

Remark 4.9. While asymptotic, Theorem 4.6 is astonishingly accurate for decks of practical interest. For example, comparing exact calculations in Table 5 with approximations using this rule of thumb in Table 1 shows that after only 3 shuffles, the numbers agree to the given precision. Moreover, the simplicity of the formula in Theorem 4.6 allows much further computations than are possible using the formula in Theorem 4.2.

We now give a heuristic for why our rule of thumb is numerically so accurate; this was hinted at previously in our remark following Corollary 4.4. Let $k \geq 0$ be an integer, and define

$$f_k(z) = \sum_{r=0}^{\infty} r^k z^r,$$

with the convention that $0^0 = 1$. Thus $f_0(z) = 1/(1-z)$, $f_1(z) = z/(1-z)^2$, and in general $f_k(z) = A_k(z)/(1-z)^{k+1}$ where $A_k(z)$ denotes the k-th Eulerian polynomial. The sum over $a_1, \ldots, a_m$ appearing in our proof of Theorem 4.6 is simply the coefficient of z^a in the generating function $(1-z)^{m-1}f_{D_1}(z)\cdots f_{D_m}(z)$. Our rule of thumb may be interpreted as saying that

$$(1-z)^{m-1}f_{D_1}(z)\cdots f_{D_m}(z) \approx \frac{D_1!\cdots D_m!}{(n+m-1)!}(1-z)^{m-1}f_{n+m-1}(z). \tag{27}$$

To explain the sense in which (27) holds, note that $f_k(z)$ extends meromorphically to the complex plane, and it has a pole of order $k+1$ at $z=1$. Moreover it is easy to see that $f_k(z) - k!/(1-z)^{k+1}$ has a pole of order at most k at $z=1$. Therefore, the LHS and RHS of (27) have poles of order $n+1$ at $z=1$, and their leading order contributions match. Therefore the difference between the RHS and LHS of (27) has a pole of order at most n at $z=1$. But in fact, this difference can have a pole of order at most $n-d$ at $z=1$, and thus the approximation in (27) is tighter than what may be expected *a priori*. To obtain our result on the order of the pole, we record that one can show

$$f_k(z) = \frac{k!}{(1-z)^{k+1}}\Big(\frac{(z-1)}{\log z}\Big)^{k+1} + \zeta(-k) + O(1-z).$$

5. Comparing 2-shuffles with different starting patterns

Conger and Viswanath note that the initial configuration can affect the speed of convergence to stationary. In this section, we investigate this for a deck with n red and n black cards. Consider first starting with reds on top. If the initial cut is at n (the most likely value) then the red-black

pattern is perfectly mixed after a single shuffle. More generally, by Corollary 3.3, the chance of the deck w resulting from a single 2-shuffle of a deck with n red cards atop n black cards is given by

$$Q_2(w) = \frac{1}{2^{2n}} \left(2^{\mathrm{h}(w)} + 2^{\mathrm{t}(w)} - 1\right).$$

Consider next the result of 2-shuffles on the *alternating deck* red-black-red-black-$\cdots$. As motivation, we recall a popular card trick: Begin with a deck of $2n$ cards arranged alternately red, black, red, black, etc. The deck may be cut any number of times. Have the deck turned face up and cut (with cuts completed) until one of the cuts results in the two piles having cards of opposite color uppermost. At this point, ask one of the participants to riffle shuffle the two piles together. The resulting arrangement has the top two cards containing one red and one black, the next two cards containing one red and one black, and so on throughout the deck. This trick is called the Gilbreath Principle after its inventor, the mathematician Norman Gilbreath. It is developed, with many variations, in Chapter 4 of [18].

¿From the trick we see that beginning with an alternating deck severely limits the possibilities. Which start mixes faster? The following developments both explain the trick and give a useful formula for analysis.

Lemma 5.1. *The number of deck patterns resulting from a cut with an odd number of cards in both piles followed by a riffle shuffle is 2^n. Similarly, the number of deck patterns resulting from a cut with both piles even followed by a riffle shuffle is 2^{n-1}.*

Proof. For the case of an odd cut, the last two cards after the riffle shuffle must be a red and a black card. No matter what piles these two cards fell from, the next two cards must also consist of one red and one black card. Continuing on, the possible resulting decks are exactly those where the ith and $i+1$st cards have different colors for $i = 1, 3, \ldots, 2n-1$. The number of such decks is exactly 2^n, since each of the order of each of the n pairs is independent.

For an even cut, we proceed by induction noting that the case when $n = 1, 2, 3$ are easily solved by inspection. In this case, the only resulting decks will necessarily begin with a red card and end with a black card. The number of decks beginning with two red cards or ending with two black cards is determined by the previous case since removing the top or bottom card from each pile results in piles with an odd number of cards, giving 2^{n-1} possibilities. However, we must discount the over counted case of decks beginning with two red cards and ending with two black cards, and, by induction since the piles are again both even, there are 2^{n-3} such decks. Finally, the remaining case must be decks beginning and ending with a red card followed by a black card. In this case, again, the piles remain even and by induction the number of such decks is 2^{n-3}. Therefore the total count for cuts with both piles even is $2^{n-1} - 2^{n-3} + 2^{n-3} = 2^{n-1}$. □

The proof of the lemma shows exactly why the card trick is a success: to have different colors on the top of the two piles, the cut must have been odd. Therefore the first two cards dropped consist of one red and one black, and the next two cards dropped consist of one red and one black, and so on. Also from the lemma, we see that the only deck that can result from either an odd cut or an even cut is the identity.

Proposition 5.2. *The chance of a 2-shuffle of the alternating deck resulting in a deck configuration w is given by*

$$2^{2n} \cdot Q_2(w) = \begin{cases} 2^{n-1} + 2^n & \text{if } w = w_0 \\ 2^{n-1} & \text{if } w \in O \setminus w_0, \\ 2^n & \text{if } w \in E \setminus w_0, \\ 0 & \text{otherwise}, \end{cases} \tag{28}$$

where w_0 is the initial alternating deck and O (respectively, E) is the set of decks that can result from riffling together the two piles from cutting the alternating deck when both piles have an odd (respectively, even) number of cards.

Proof. Let $w, u \in O$. Then the total number of ways w can result from any odd cut is equal to the total number of ways u can result from any odd cut. The same is true replacing O with E and "odd" with "even". From the binomial identity

$$\sum_{k=0}^{n}(-1)^k\binom{n}{k} = 0 \quad \rightsquigarrow \quad \sum_{k \text{ odd}}\binom{2n}{k} = \sum_{k \text{ even}}\binom{2n}{k},$$

we must have both the right-hand sums equal to 2^{2n-1}. Therefore, by Lemma 5.1, the total number of ways w can result from an odd cut (assuming it can) is $2^{2n-1}/2^n = 2^{n-1}$, and, similarly, the total number of ways w can result from an even cut (assuming it can) is $2^{2n-1}/2^{n-1} = 2^n$. □

It follows from (28) that the separation distance for a 2-shuffle is $\mathrm{SEP}(2) = 1$ when $n \geq 3$. Furthermore, since $\binom{2n}{n} \geq 2^n$, we can compute the total variation of a 2-shuffle to be

$$\|Q_2 - U\|_{TV} = \frac{1}{2}\left(1 - \frac{2^n + 2^{n-1} - 1}{\binom{2n}{n}}\right), \tag{29}$$

which goes to .5 exponentially fast as n goes to infinity. In contrast, starting with reds above blacks, asymptotic analysis of (18) shows that the total variation tends to 1 after a single shuffle when n is large. Thus an alternating start leads to faster mixing.

Acknowledgements

The authors thank Jason Fulman for careful comments and references. We also thank MSRI and the participants of the combinatorial representation theory program where this work began.

References

[1] D. Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on probability, XVII*, volume 986 of *Lecture Notes in Math.*, pages 243–297. Springer, Berlin, 1983.
[2] D. Aldous and P. Diaconis. Shuffling cards and stopping times. *Amer. Math. Monthly*, 93(5):333–348, 1986.
[3] D. Bayer and P. Diaconis. Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.*, 2(2):294–313, 1992.
[4] S. Boyd, P. Diaconis, P. Parrilo, and L. Xiao. Symmetry analysis of reversible Markov chains. *Internet Math.*, 2(1):31–71, 2005.
[5] T. Ceccherini-Silberstein, F. Scarabotti, and F. Tolli. *Harmonic analysis on finite groups*, volume 108 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2008. Representation theory, Gelfand pairs and Markov chains.
[6] G.-Y. Chen and L. Saloff-Coste. The cutoff phenomenon for randomized riffle shuffles. *Random Structures Algorithms*, 32(3):346–3745, 2008.
[7] M. Ciucu. No-feedback card guessing for dovetail shuffles. *Ann. Appl. Probab.*, 8(4):1251–1269, 1998.
[8] M. Conger and D. Viswanath. Riffle shuffles of decks with repeated cards. *Ann. Probab.*, 34(2):804–819, 2006.
[9] M. Conger and D. Viswanath. Normal approximations for descents and inversions of permutations of multisets. *J. Theoret. Probab.*, 20(2):309–325, 2007.
[10] P. Diaconis. *Group representations in probability and statistics*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988.
[11] P. Diaconis. Mathematical developments from the analysis of riffle shuffling. In *Groups, combinatorics & geometry (Durham, 2001)*, pages 73–97. World Sci. Publ., River Edge, NJ, 2003.
[12] P. Diaconis and J. Fulman. Carries, shuffling and an amazing matrix. preprint, 2008.
[13] P. Diaconis and S. P. Holmes. Random walks on trees and matchings. *Electron. J. Probab.*, 7:no. 6, 17 pp. (electronic), 2002.
[14] P. Diaconis, M. McGrath, and J. Pitman. Riffle shuffles, cycles, and descents. *Combinatorica*, 15(1):11–29, 1995.
[15] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57(2):159–179, 1981.

[16] A. Fässler and E. Stiefel. *Group theoretical methods and their applications*. Birkhäuser Boston Inc., Boston, MA, 1992. Translated from the German by Baoswan Dzung Wong.
[17] J. Fulman. Applications of symmetric functions to cycle and increasing subsequence structure after shuffles. *J. Algebraic Combin.*, 16(2):165–194, 2002.
[18] M. Gardner. *Martin Gardner's New Mathematical Diversions from Scientific American*. Simon & Schuster, New York, 1966.
[19] E. Gilbert. Theory of shuffling. Technical memorandum, Bell Laboratories, 1955.
[20] J. M. Holte. Carries, combinatorics, and an amazing matrix. *Amer. Math. Monthly*, 104(2):138–149, 1997.
[21] J. Reeds. Theory of shuffling. Unpublished manuscript, 1976.
[22] J.-P. Serre. *Linear representations of finite groups*. Springer-Verlag, New York, 1977. Translated from the second French edition by Leonard L. Scott, Graduate Texts in Mathematics, Vol. 42.
[23] J. R. Weaver. Centrosymmetric (cross-symmetric) matrices, their basic properties, eigenvalues, and eigenvectors. *Amer. Math. Monthly*, 92(10):711–717, 1985.

Appendix A. Random walks on groups

In this appendix, we reformulate shuffling in terms of random walks on the symmetric group $\mathcal{S}_n$, so that our investigation of particular properties of a deck becomes the quotient walk on Young subgroups of $\mathcal{S}_n$.

Let G be a finite group with $Q(g) \geq 0$, $\sum_{g\in G} Q(g) = 1$ a probability on G. The walk in (1) may be called the *left walk* since it consists of repeatedly picking elements independently with probability Q, say $g_1, g_2, g_3, \ldots$, and, starting at the identity 1_G, multiplying on the left by g_i. The generates a *random walk on* G,

$$1_G,\ g_1,\ g_2g_1,\ g_3g_2g_1,\ \ldots.$$

By inspection, the chance that the walk is at g after k steps is $Q^{*k}(g)$, where $Q^0(g) = \delta_{1_G,g}$.

An algebraic method of focusing on aspects of the walk is to use the *quotient walk*. Let $H \leq G$ be a subgroup of G, and set $X = G/H = \{xH\}$ to be the set of left cosets of H in G. The quotient walk is derived from the walk above by simply reporting to which coset the current position of the walk belongs. The quotient walk is a Markov chain on X with transition matrix given by

$$K(x, y) = Q(yHx^{-1}) = \sum_{h\in H} Q(yhx^{-1}). \tag{30}$$

Note that K is well-defined (i.e. independent of the choice of coset representatives) and that K is doubly stochastic. Thus the uniform distribution on X, $U(x) = |H|/|G|$, is a stationary distribution for K. The chain K is reversible if and only if Q is symmetric (i.e. $Q(g) = Q(g^{-1})$). Note that this is not the case for riffle shuffles. While intuitively obvious, the following shows the basic fact that powers of the matrix K correspond to convolving and taking cosets.

Proposition A.1. *For Q a probability distribution on a finite group G and K as defined in (30), we have*

$$K^l(x, y) = Q^{*l}(yHx^{-1}).$$

Proof. The result is immediate from the definitions for $l = 0, 1$. We prove the result for $l = 2$, the general case being similar. Note that

$$K^2(x, y) = \sum_z K(x, z)K(z, y) = \sum_z \sum_{h_1,h_2} Q(zh_1x^{-1})Q(yh_2z^{-1}).$$

Setting $h_2 = hh_1^{-1}$, noting that zh_1 runs over G as z runs over X and h_1 over H, and setting $g_1 = gx^{-1}$, we have

$$K^2(x, y) = \sum_h \sum_g Q(gx^{-1})Q(yhg^{-1}) = \sum_h \sum_{g_1} Q(g_1)Q(yhx^{-1}g_1^{-1}) = Q^2(yHx^{-1}).$$

□

We may identify permutations in $\mathcal{S}_n$ with arrangements of a deck of n cards by setting $\sigma(i)$ to be the label of the card at position i from the top. Thus the permutation 2 1 4 3 is associated with four cards where "2" is on top, followed by "1", followed by "4", and finally "3" is on the bottom. If we consider the cards labelled $1, 2, \ldots, k$ to be "red" cards, and the cards labelled $k+1, k+2, \ldots, n$ to be "black" cards, with all cards of the same color indistinguishable, the coset space

$$X = \mathcal{S}_n/(\mathcal{S}_k \times \mathcal{S}_{n-k})$$

is naturally associated with the $\binom{n}{k}$ arrangements of red and black *unlabeled* cards. Here, of course, we identify an element of $\mathcal{S}_k \times \mathcal{S}_{n-k} \leq \mathcal{S}_n$ as permuting the first k and last $n-k$ cards among themselves. Similar constructions work for suits or values. Thus Proposition A.1 shows that the processes studied in the body of this paper are Markov chains.

Appendix B. Shuffling by random transpositions

Let $L^2(X) = \{f : X \to \mathbb{C}\}$ be the set of complex-valued functions on X with inner product defined by

$$\langle f_1 | f_2 \rangle = \frac{1}{|X|} \sum_x f_1(x)\overline{f_2(x)}. \tag{31}$$

If K is symmetric, then real-valued functions may be used. The transition matrix K operates on L^2 via

$$Kf(x) = \sum_y K(x,y)f(y). \tag{32}$$

In the present case, $L^2(X) = \mathrm{Ind}_H^G(\mathbf{1})$, the usual permutation representation of G acting on left cosets $X = G/H$, with $T_g f(x) = f(g^{-1}x)$. By construction, the action of G commutes with K, i.e.

$$T_g(Kf) = K(T_g f) \tag{33}$$

for all $f \in L^2(X)$ and all $g \in G$. This implies that group representation theory can be used to reduce the operator K (or diagonalize K in the case when K is symmetric). This classical topic is well developed in Fässler-Steifel [16] and Boyd, et. al. [4].

Let $\widehat{G}$ denote the set of irreducible representations of the finite group G. For $\rho \in \widehat{G}$, the Fourier transform of $f \in L^2(G)$ at ρ is defined by

$$\widehat{f}(\rho) = \sum_{g \in G} f(g)\rho(g).$$

As usual, Fourier transform turns convolution into products, i.e.

$$\widehat{Q^{*k}}(\rho) = \widehat{Q}(\rho)^k.$$

Schur's lemma implies that the uniform distribution has zero transform

$$\widehat{U}(\rho) = \begin{cases} 1 & \text{if } \rho \text{ is trivial,} \\ 0 & \text{otherwise.} \end{cases}$$

The Fourier inversion theorem reconstructs f from $\{\widehat{f}(\rho)\}$ by

$$f(g) = \frac{1}{|G|} \sum_{\rho \in \widehat{G}} \dim_\rho \mathrm{Tr}\left(\widehat{f}(\rho)\rho(g^{-1})\right).$$

For background, see Serre [22], Diaconis [10] or Ceccherini, et. al [5] where many applications are given.

Suppose the induced representation $L^2(X)$ decomposes into irreducibles as

$$L^2(X) = \bigoplus_{\rho \in \widehat{G}} V_\rho^{\oplus a_\rho}. \tag{34}$$

Then since K commutes with G, K sends each of the spaces $V_\rho^{\oplus a_\rho}$ into itself. Further reductions may be possible if Q has suitable symmetries. The following widely studied special case is relevant.

Definition B.1. The pair $H \leq G$ is a *Gelfand pair* if $L^2(X)$ is multiplicity free, i.e. all a_ρ in (34) are either 0 or 1.

For example, when $1 \leq k \leq n/2$, $\mathcal{S}_k \times \mathcal{S}_{n-k} \leq \mathcal{S}_n$ is a Gelfand pair with

$$L^2(X) = \bigoplus_{i=0}^{k} S^{n-i,i}. \tag{35}$$

Recall that the irreducible representations of $\mathcal{S}_n$ are indexed by partitions λ of n. If S^λ denotes the λth representation (Specht modules), the sum in (35) runs over partitions into two parts with the smaller part at most k. For further background on Gelfand pairs, including examples and applications, see [10, 5].

Now we study a deck of red and black cards after repeated random transposition shuffles. Recall that Diaconis-Shahshahani [15] show that it takes $\frac{1}{2}n(\log(n) + c)$ shuffles to mix n distinct cards. To be precise, the measure on $\mathcal{S}_n$ that drives the walks is

$$Q(\sigma) = \begin{cases} 1/n & \text{if } \sigma = \text{id}, \\ 2/n^2 & \text{if } \sigma = (i,j), \\ 0 & \text{otherwise.} \end{cases}$$

Throughout the following, all walks begin at the identity permutation, and we use the convention that $\pi(i)$ is the label of the card at position i.

First, we follow the position of the top card; i.e. the two of hearts is the only red card followed by $n-1$ black cards. The transition matrix for this walk is given by

$$P(i,j) = \begin{cases} \dfrac{1}{n} + \dfrac{(n-2)(n-3)}{n^2} & \text{if } i = j, \\ \dfrac{2}{n^2} & \text{if } i \neq j. \end{cases} \tag{36}$$

Note that this is symmetric, with $\Pi(i) = 1/n$ as the stationary distribution.

Proposition B.2. *For the transition matrix $P(i,j)$ above and all $l \geq 0$, we have*

$$P^l(i,j) = \begin{cases} \dfrac{1}{n} + \left(1 - \dfrac{2}{n}\right)^l \left(1 - \dfrac{1}{n}\right) & \text{if } i = j, \\ \dfrac{1}{n} - \left(1 - \dfrac{2}{n}\right)^l \dfrac{1}{n} & \text{if } i \neq j. \end{cases} \tag{37}$$

From this it follows that

$$\text{SEP}(l) = \left(1 - \frac{2}{n}\right)^l \quad \text{and} \quad \|P - \Pi\|_{TV} = \left(1 - \frac{2}{n}\right)^l \left(1 - \frac{1}{n}\right).$$

Proof. The results for the separation and total variation distances follow from (37) and the definitions. It is possible to give a direct combinatorial argument for (37), but the following representation theoretic argument generalizes readily to find similar formula for j-tuples of cards.

The random transposition measure Q is constant on conjugacy classes of $\mathcal{S}_n$ and so acts on each irreducible representation as a constant times the identity. These constants are given explicitly by Diaconis-Shahshahani [15], involving characters and dimensions of the representation. Consider the operator $K(\sigma,\tau) = Q(\tau\sigma^{-1})$ on the regular representation. The function $f(\sigma) = \delta_{1,\sigma(i)} - 1/n$ lies in the $n-1$ copies of the $n-1$-dimensional representation corresponding to the partition $(n-1,1)$. The operator K acts on this space by multiplication by $1-2/n$. Thus

$$P_\sigma \begin{pmatrix} \text{card labelled 1} \\ \text{at position } i \\ \text{after } l \text{ shuffles} \end{pmatrix} - \frac{1}{n} = K^l f(\sigma) = \left(1-\frac{2}{n}\right)^l f(\sigma) = \left(1-\frac{2}{n}\right)^l \left(\delta_{1,\sigma(i)} - \frac{1}{n}\right).$$

Here σ is the starting arrangement. Evaluating the right-hand side gives (37). □

Next we consider the deck with $N = 2n$ cards where the (original) top n cards are red and the (original) bottom n cards are black. In this case, we think of the the random transposition operator acting on the quotient space $\mathcal{S}_N/\mathcal{S}_n \times \mathcal{S}_n$. For $x, y \in \mathcal{S}_N/\mathcal{S}_n \times \mathcal{S}_n$, the induced Markov chain is

$$K(x,y) = \begin{cases} \frac{1}{N^2} & \text{if } x \neq y \text{ differ by a transposition,} \\ \frac{1}{N} + \frac{(n(n-1))^2}{N^2} & \text{if } x = y, \\ 0 & \text{otherwise.} \end{cases} \tag{38}$$

This chain has uniform stationary distribution $\Pi(x) = 1/\binom{N}{n}$.

The chain K is invariant under $\mathcal{S}_N$, i.e. $K(x,y) = K(\sigma x, \sigma y)$, so the distance to stationary does not depend on the original configuration. As noted earlier, the pair $\mathcal{S}_n \times \mathcal{S}_n, \mathcal{S}_N$ is a Gelfand pair, so (35) allows an easy determination of the eigen values and rate of convergence.

Proposition B.3. *For the Markov chain K on $\mathcal{S}_N/\mathcal{S}_n \times \mathcal{S}_n$, the eigen values are*

$$\beta_0 = 1, \quad \beta_j = \frac{1}{N} + \frac{1}{N^2}\left((N-j)^2 - (N-j) + j^2 - 3j\right),$$

$j = 1, \ldots, n$. The multiplicity of β_j is $m_j = \binom{N-1}{j}$. Moreover, there is a universal constant A such that if $l = \frac{1}{4}N(\log N + C)$, then

$$\left\| K^l - \Pi \right\|_{TV} \leq Ae^{-c/2}.$$

Proof. The operator K acts on $L^2\left(\mathcal{S}_N/\mathcal{S}_n \times \mathcal{S}_n\right)$ as the element of the group algebra

$$\frac{1}{N}\mathrm{Id} + \frac{2}{N^2}\sum_{i<j}(i,j).$$

As shown in [13], this element acts on the irreducibles $\mathcal{S}^{n-j,j}$ as a constant times the identity, with the constant being β_j and the multiplicity being the dimension of $\mathcal{S}^{n-j,j}$. This proves the first part.

The remaining claims can be proved following the argument in [13]: bound the total variation distance by the L^2 norm, express this in terms of the eigen values and average over the starting state. This reduces the problem to bounding

$$\sum_{j=1}^{n} m_j \beta_j^{2l}.$$

The lead term in this is

$$(N-1)\left(1-\frac{2}{N}\right)^{2l} \leq e^{-c}.$$

For l of the form $\frac{1}{4}N(\log N + c)$, the other terms are smaller and sum in a reasonably standard fashion. The terms are the same as in [13], so we suppress further details. □

Remark B.4. It is easy to give a lower bound showing that after $l = \frac{1}{4}N(\log N + c)$ steps the distance to stationary is bounded away from 0 for large N. Further, in this case, the distance tends to 1 if $c = c_N$ tends to $-\infty$.

These results show that for red-black mixing, there is a total variation cutoff at $\frac{1}{4}N \log N$. Note that single card mixing does not have a cutoff, recalling that in Proposition B.2 the deck has size n and in Proposition B.3 the deck has size $N = 2n$.

Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139-4307
E-mail address: sassaf@math.mit.edu

Department of Statistics, Stanford University, 390 Serra Mall, Stanford, CA 94305-4065

Department of Mathematics, Stanford University, 450 Serra Mall, Building 380, Stanford, CA 94305-2125
E-mail address: ksound@math.stanford.edu