FPSAC 2012, Nagoya, Japan DMTCS proc. **AR**, 2012, 445–456

Riffle shuffles with biased cuts

Sami Assaf[1] Persi Diaconis[2] Kannan Soundararajan[3]

[1] *Department of Mathematics, University of Southern California, Los Angeles, CA, USA*
[2] *Department of Statistics, Stanford University, Stanford, CA, USA*
[3] *Department of Mathematics, Stanford University, Stanford, CA, USA*

Abstract. The well-known Gilbert-Shannon-Reeds model for riffle shuffles assumes that the cards are initially cut 'about in half' and then riffled together. We analyze a natural variant where the initial cut is biased. Extending results of Fulman (1998), we show a sharp cutoff in separation and L-infinity distances. This analysis is possible due to the close connection between shuffling and quasisymmetric functions along with some complex analysis of a generating function.

Résumé. Le modèle de Gilbert-Shannon-Reeds pour melange de cartes suppose que les cartes sont d'abord coupés énviron de moitié', puis intescaler ensemble. Nous analysons une variante naturelle, où la coupe initiale est biaisé. Nous propons une extension des résultats de Fulman (1998), nous montrent une forte coupure dans les distances de séparation et L-infinity. Cette analyse est possible grâce à l'étroite relation entre brassage et fonctions quasisymmetric.

Keywords: card shuffling, biased cuts, quasisymmetric functions

1 Introduction

We analyze a natural one-parameter model for riffle shuffling a deck of n cards. Roughly, the deck is cut into two piles with a binomial (n, θ) distribution. Then the piles are riffled together sequentially according to the following rule: if the left pile has A cards and the right pile has B cards, then drop the next card from the bottom of the left pile with probability $A/(A+B)$. Continue until all cards are dropped. Starting at the identity, let $P_\theta(w)$ be the probability of the permutation w after one such θ-shuffle. Define convolution by

$$P_\theta^{*k}(w) = \sum_v P_\theta(v) P_\theta^{*(k-1)}(v^{-1}w), \tag{1.1}$$

and define the uniform distribution by $U(w) = 1/n!$.

When $\theta = 1/2$, this is the widely studied Gilbert-Shannon-Reeds model. The natural version with biased cuts was studied by [Diaconis et al.(1992)], [Lalley(1996), Lalley(2000)] and most thoroughly by [Fulman(1998)]. A literature review is in Section 2 below. Here we study the rate of convergence in the separation and ℓ_∞ metrics:

$$\mathrm{SEP}(k) = \max_w \left(1 - \frac{P^{*k}(w)}{U(w)}\right) \tag{1.2}$$

$$\ell_\infty(k) = \max_w \left|1 - \frac{P^{*k}(w)}{U(w)}\right|. \tag{1.3}$$

Note that $\mathrm{SEP}(k)$ is bounded above by 1, and $\ell_\infty(k)$ can be as large as $n!-1$. Further, both $\mathrm{SEP}(k)$ and $\ell_\infty(k)$ are upper bounds for the total variation metric:

$$\|P^{*k}-U\|_{TV}=\frac{1}{2}\sum_w |P^{*k}(w)-U(w)| \le \mathrm{SEP}(k) \le \ell_\infty(k).$$

A main result of this note gives closed form expressions

$$\mathrm{SEP}(k) = 1-\sum_{w\in\mathfrak{S}_n} \mathrm{sgn}(w)\prod_{i=1}^{n}\left(\theta^i+(1-\theta)^i\right)^{kn_i(w)} \tag{1.4}$$

$$\ell_\infty(k) = \sum_{w\in\mathfrak{S}_n}\prod_{i=1}^{n}\left(\theta^i+(1-\theta)^i\right)^{kn_i(w)}-1, \tag{1.5}$$

where $n_i(w)$ is the number of i-cycles in the permutation w. Using these formulae we prove the following.

Theorem 1 *For the θ-biased riffle shuffle measure on $\mathfrak{S}_n$, let*

$$k=\left\lfloor \frac{2\log n-\log 2+c}{-\log(\theta^2+(1-\theta)^2)}\right\rfloor. \tag{1.6}$$

Then

$$\mathrm{SEP}(k) \sim \exp(e^{-c})-1 \tag{1.7}$$

$$\ell_\infty(k) \sim 1-\exp(-e^{-c}) \tag{1.8}$$

for any fixed real c as n tends to ∞. Here $0<\theta<1$ is fixed.

An upper bound on separation of this form is given in [Fulman(1998)]. Theorem 1 shows this bound is tight, holds also for ℓ_∞, and establishes the cutoff phenomenon. Note that, as a function of θ, k as defined in (1.6) above is smallest when $\theta=1/2$, so unbiased cuts lead to fastest mixing.

Background on Markov chains and shuffling is given in Section 2. There is an intimate connection between these biased shuffles and quasisymmetric functions explained in Section 3 where we prove (1.4) and (1.5). The upper bound in [Fulman(1998)] is derived using a strong stationary time. This is shown to be exact and equivalent to (1.4) in Section 4. The proof of Theorem 1, which has extensions to allow θ to depend on n (e.g. $\theta=1/n$), is in Section 5.

2 Riffle Shuffling

A superb introduction to Markov chains which treats riffle shuffling and stationary times is the book by [Levin et al.(2009)]. The analysis of riffle shuffling has connections to algebra, geometry and combinatorics; a detailed survey is in [Diaconis(2003)]. The results and references in [Assaf et al.(2011)] and [Conger and Howald(2010)] bring this up to date.

For present purposes, the following extension is needed. Let $1\le a\le\infty$, and let $\boldsymbol{\theta}=(\theta_1,\theta_2,\dots,\theta_a)$, with $0\le\theta_i\le 1$ and $\theta_1+\cdots+\theta_a=1$, be fixed. A $\boldsymbol{\theta}$-shuffle of a deck of n cards proceeds as follows: Choose $\{N_i\}_{i=1}^a$ from the multinomial$(n,\boldsymbol{\theta})$ distribution, that is, with the distribution of n balls being dropped into a boxes independently according to $\boldsymbol{\theta}$. Cut the deck into a packets of sizes $N_1, N_2, \ldots, N_a$ (some of the packets may be empty). Now sequentially drop cards from the bottom of each packet, choosing to drop from pack i with probability proportional to its current packet size. Continue until all cards have been dropped into a single pile. Let $P_{\boldsymbol{\theta}}$ denote the associated measure on $\mathfrak{S}_n$. Note that several more detailed descriptions of $P_{\boldsymbol{\theta}}$ appear in [Fulman(1998)].

When $a = 2$ and $\theta_1 = \theta_2 = 1/2$, this is the basic Gilbert-Shannon-Reeds measure. When $a = 2$ and $\theta_1 = \theta$, $\theta_2 = 1 - \theta$, this is the θ-biased shuffle studied in the present paper. The measures $P_{\boldsymbol{\theta}}$ were studied by [Diaconis et al.(1992)] who prove that they convolve nicely: if $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_a)$ and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_b)$, then set $\boldsymbol{\theta} * \boldsymbol{\eta} = (\theta_1\eta_1, \ldots, \theta_1\eta_b, \theta_2\eta_1, \ldots, \theta_a\eta_b)$, a vector of length ab.

Proposition 2 ([Diaconis et al.(1992)]) *On $\mathfrak{S}_n$, we have*

$$P_{\boldsymbol{\theta}} * P_{\boldsymbol{\eta}} = P_{\boldsymbol{\theta} * \boldsymbol{\eta}}.$$

Thus $P_{\boldsymbol{\theta}}^{*k} = P_{\boldsymbol{\theta}^{*k}}$, and the combinatorics of $P_{\boldsymbol{\theta}}$ determines the convolution powers. [Fulman(1998)] works out many properties of these measures giving closed formulae and asymptotics for the distribution of cycle structure, inversions and descents.

When $\theta = 1/2$, a sharp analysis of the rate of convergence for the Gilbert-Shannon-Reeds measure in total variation distance appears in [Bayer and Diaconis(1992)]. It is an open problem to give a similarly sharp analysis for the measures $P_{\boldsymbol{\theta}}^{*k}$.

Hyperplane walks

Our θ-shuffles may be studied from other points of view as well. They are a special case of hyperplane walks introduced in [Bidigare et al.(1999)] and further studied in [Brown and Diaconis(1998)] and more recently in [Athanasiadis and Diaconis(2010)] and [Diaconis et al.(2011)]. Further, they fall into the class of "Hopf-square" walks studied in [Diaconis et al.(2011)]. Each of these perspectives adds to our picture. A brief commentary follows.

The braid arrangement is based on the $\binom{n}{2}$ hyperplanes $H_{i,j} = \{x \in \mathbb{R}^n \mid x_i = x_j\}$, $1 \le i < j \le n$. This divides $\mathbb{R}^n$ into chambers and faces. As shown in [Bidigare et al.(1999)], the chambers are indexed by permutations and the faces are indexed by block ordered set partitions. There is a simple projection operator which, given a chamber C and a face F, returns the chamber $C * F$ that is adjacent to F and closest to C (in the sense of crossing the fewest number of hyperplanes). Details are in [Bidigare et al.(1999) , Brown and Diaconis(1998)]. It is shown there that projection operates as a kind of inverse riffle shuffle. Put a probability measure on faces of form S, S^c, with $S \subset [n]$, giving probability $\theta^{|S|}(1-\theta)^{n-|S|}$ to each (S may be empty). The resulting hyperplane walk may be explained as follows: Picture a deck of n cards in order. For each card, flip an independent θ-coin. Remove all cards where the coin comes up heads, keeping their relative order fixed, and move them to the top of the deck. This is precisely an inverse θ-shuffle.

The theory of [Bidigare et al.(1999) , Brown and Diaconis(1998)] gives useful expressions for the eigenvalues of any hyperplane walk. Specialized to θ-shuffles, they show there is one eigenvalue β_w for each permutation $w \in \mathfrak{S}_n$. Further, [Diaconis et al.(2011)] gives a description of the left eigen vectors. These give right eigen vectors and values of the "forward" θ-shuffles.

As one example, [Brown and Diaconis(1998)] gives a rate of convergence after k-steps. In the present case, this reads

$$\|K^k - U\|_{TV} \leq \sum_{1\leq i<j\leq n} \beta_{i,j}^k \tag{2.1}$$

with $\beta_{i,j} = \sum_{F\subseteq H_{i,j}} w(F)$. By symmetry, $\beta_{i,j} = \beta_{1,2}$ is constant in i, j. The sum is over all set partitions S, S^c where either $\{1,2\} \subseteq S$ or $\{1,2\} \subseteq S^c$. So $\{1,2\}$ contributes $\sum_{A\subseteq[n-1]} \theta^2\theta^{|A|}(1-\theta)^{n-2-|A|} = p^2$, the compliment contributes $(1-\theta)^2$, and so $\beta_{i,j} = \theta^2 + (1-\theta)^2$. The bound above becomes

$$\|K^k - U\|_{TV} \leq \binom{n}{2} \left(\theta^2 + (1-\theta)^2\right)^k. \tag{2.2}$$

This is exactly the birthday bound derived differently below. Of course, these are just upper bounds, and it is of interest to know if they can be improved. The theory developed below shows that

$$\|K^k - U\|_{TV} \leq \mathrm{SEP}(k) \leq \binom{n}{2} \left(\theta^2 + (1-\theta)^2\right)^k. \tag{2.3}$$

for fixed θ in $(0,1)$. Theorem 1 shows that $\mathrm{SEP}(k) \sim \binom{n}{2}\left(\theta^2 + (1-\theta)^2\right)^k$, so the bound is best possible.

Recall that any $w \in \mathfrak{S}_n$ has a unique factorization as a product of decreasing Lyndon words: $w = \ell_1\ell_2\cdots\ell_k$. Here ℓ_i is Lyndon if it is lexicographically least among all cyclic rearrangements (so 132 is Lyndon but 213 is not). For example $236415 = 236 \cdot 4 \cdot 15$. The theorem in [Diaconis et al.(2011)] shows

$$\beta_w = \prod_{i=1}^{k} \left(\theta^{|\ell_i|} + (1-\theta)^{|\ell_i|}\right) \tag{2.4}$$

where $|\ell_i|$ is the length of the Lyndon word ℓ_i. If w is the reverse of the identity, then all $|\ell_i| = 1$ and $\beta_w = 1$. The second eigenvalue is $\theta^2 + (1-\theta)^2$ with multiplicity $\binom{n}{2}$, so the bound (2.3) uses precisely these eigen values. More generally, the eigen values are $\prod_{i=1}^{n}\left(\theta^i + (1-\theta)^i\right)^{a_i}$ for any $0 \leq a_i \leq n$ with $\sum ia_i = n$, each with multiplicity $n!/(\prod_i i^{a_i}a_i!)$.

3 Quasisymmetric Functions

Background on symmetric function theory is in [Macdonald(1995)] with [Stanley(1999)] developing the extension to quasisymmetric functions. We work with infinitely many variables $X = \{x_i\}_{i=1}^{\infty}$. The space of quasisymmetric functions homogeneous of degree n has dimension 2^{n-1}. A basis for this space is indexed by subsets of $[n-1] = \{1, 2, \ldots, n-1\}$ or, equivalently, by compositions of n. We use the following bijection between subsets $D = \{D_1 < D_2 < \cdots < D_{a-1}\}$ of $[n-1]$ and compositions $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_a)$ of n to identify subsets and compositions, which we denote by $\alpha \leftrightarrow D(\alpha)$:

$$\begin{array}{rcl} (\alpha_1, \alpha_2, \ldots, \alpha_a) & \longmapsto & \{\alpha_1, \alpha_1+\alpha_2, \ldots, \alpha_1+\cdots+\alpha_{a-1}\}, \\ (D_1, D_2 - D_1, \ldots, n - D_{a-1}) & \longleftarrow & \{D_1 < D_2 < \ldots < D_{a-1}\}. \end{array}$$

The *monomial* quasisymmetric function basis is defined by

$$M_\alpha(X) = \sum_{i_1<i_2<\cdots<i_a} x_{i_1}^{\alpha_1} x_{i_2}^{\alpha_2} \cdots x_{i_a}^{\alpha_a}. \tag{3.1}$$

For example, $M_{(1,2,1)}(X) = \sum_{i_1<i_2<i_3} x_{i_1}x_{i_2}^2x_{i_3}$.

The *fundamental* quasisymmetric function basis of [Gessel(1984)] is defined by

$$Q_D(X) = \sum_{\substack{i_1\leq\cdots\leq i_n \\ i_j=i_{j+1}\Rightarrow j\notin D}} x_{i_1}\cdots x_{i_n}. \tag{3.2}$$

For example, for $n = 4$, $Q_{\{1\}}(X) = \sum_{i_1<i_2\leq i_3\leq i_4} x_{i_1}x_{i_2}x_{i_3}x_{i_4}$. Expressed in terms of monomial quasisymmetric functions, $Q_{\{1\}}(X) = M_{(1,3)}(X) + M_{(1,2,1)}(X) + M_{(1,1,2)}(X) + M_{(1,1,1,1)}(X)$. In general, the fundamental basis is related to the monomial basis by

$$Q_{D(\beta)}(X) = \sum_{\alpha \text{ refines } \beta} M_\alpha(X), \tag{3.3}$$

where a composition α of length a refines the composition β of length b if there exist indices $0 = i_0, i_1, i_2, \ldots, i_b = a$ such that $\alpha_{i_{j-1}+1} + \cdots + \alpha_{i_j} = \beta_j$. For example, both $(1,2,1)$ and $(1,1,2)$ refine $(1,3)$ but $(2,1,1)$ does not.

[Stanley(2001)], based on results in [Fulman(1998)], established a sharp connection between $\boldsymbol{\theta}$-shuffling and quasisymmetric functions.

Theorem 3 ([Stanley(2001)](Theorem 2.1)) *Let $w \in \mathfrak{S}_n$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_a)$ be given. Then*

$$P_{\boldsymbol{\theta}}(w) = Q_{\mathrm{iDes}(w)}(\boldsymbol{\theta}),$$

where $\mathrm{iDes}(w) = \mathrm{Des}(w^{-1})$ *is the inverse descent set of* w.

This identification together with (3.3) gives a useful inequality which shows that separation and ℓ_∞ are achieved at the reversal and the identity permutations, respectively.

Proposition 4 *For permutations w and u, if* $\mathrm{iDes}(w)$ *contains* $\mathrm{iDes}(u)$*, then* $\mathrm{Prob}(w) \leq \mathrm{Prob}(u)$ *with equality if and only if* $\mathrm{iDes}(w) = \mathrm{iDes}(u)$.

Proof: First note that α refines β if and only if $D(\alpha)$ contains $D(\beta)$. Let α and β be such that $D(\alpha) = \mathrm{iDes}(w)$ and $D(\beta) = \mathrm{iDes}(u)$. From (3.3) and the transitivity of refinement, we have

$$\begin{aligned} Q_{D(\beta)}(X) &= \sum_{\gamma \text{ refines } \beta} M_\gamma(X) \\ &= \sum_{\gamma \text{ refines } \alpha} M_\gamma(X) + \sum_{\substack{\gamma\prime \text{ refines } \beta \\ \gamma\prime \text{ not refine } \alpha}} M_{\gamma\prime}(X) \\ &= Q_{D(\alpha)}(X) + \sum_{\substack{\gamma\prime \text{ refines } \beta \\ \gamma\prime \text{ not refine } \alpha}} M_{\gamma\prime}(X). \end{aligned}$$

Furthermore, $\alpha \neq \beta$ if and only if β does not refine α, in which case the summand contains the term M_β. Since the x_i are probabilities, they are all nonnegative, thus making $Q_{D(\alpha)}$ strictly less than $Q_{D(\beta)}$. □

In the partial order on subsets or, equivalently, composition, $[n-1] = D(1^n)$ is the unique minimal element and $\varnothing = D(n)$ is the unique maximal element. Therefore Proposition 4 has the following consequence.

Corollary 5 *For any $\boldsymbol{\theta}$, we have*

$$\begin{aligned}\mathrm{SEP}(P_{\boldsymbol{\theta}}) &= 1 - n! \cdot Q_{[n-1]}(\boldsymbol{\theta}) \\ \ell_\infty(P_{\boldsymbol{\theta}}) &= \max(1 - n! \cdot Q_{[n-1]}(\boldsymbol{\theta}), n! \cdot Q_{\varnothing}(\boldsymbol{\theta}) - 1).\end{aligned}$$

Remark 6 *When $\boldsymbol{\theta} = (\theta, 1-\theta)^{*k}$, we show below that the maximum is taken on at the second argument. This is not always the case. On the cyclic group $\mathcal{C}_3$, with $\mu(1) = \mu(-1) = \frac{1}{2}, \mu(0) = 0$, we have $3\mu(1) - 1 = \frac{1}{2}$ and $1 - 3\mu(0) = 1$.*

For some permutations, the associated quasisymmetric functions are easy to write down. This happens in particular if the quasisymmetric function is symmetric. Below we need the elementary symmetric functions $e_n(X)$, the complete homogeneous symmetric functions $h_n(X)$, and the power sum symmetric functions $p_n(X)$. For λ a partition with $n_i = n_i(\lambda)$ parts equal to i, define $e_\lambda = \prod_i e_i^{n_i}$, $h_\lambda = \prod_i h_i^{n_i}$, $p_\lambda = \prod_i p_i^{n_i}$. As λ ranges over partitions of n, these are the familiar bases for the homogeneous symmetric functions of degree n.

Note that

$$e_n(X) = Q_{[n-1]}(X), \qquad h_n(X) = Q_{\varnothing}(X), \qquad p_n(X^{*k}) = (p_n(X))^k. \tag{3.4}$$

Theorem 7 *For any $\boldsymbol{\theta}$, with $\mathrm{id} = 1, 2, \ldots, n$ and $\mathrm{rev} = n, n-1, \ldots, 1$, we have*

$$P_{\boldsymbol{\theta}}^{*k}(\mathrm{rev}) = \sum_{\lambda \vdash n} (-1)^{n-\ell(\lambda)} z_\lambda^{-1} \prod_{i=1}^{n} p_i(\boldsymbol{\theta})^{kn_i(\lambda)}, \tag{3.5}$$

$$P_{\boldsymbol{\theta}}^{*k}(\mathrm{id}) = \sum_{\lambda \vdash n} z_\lambda^{-1} \prod_{i=1}^{n} p_i(\boldsymbol{\theta})^{kn_i(\lambda)}, \tag{3.6}$$

where $\ell(\lambda)$ is the number of parts of λ and $z_\lambda = \prod_i i^{n_i(\lambda)} n_i(\lambda)!$.

Proof: The result follows from Theorem 3, (3.4) and the standard expansions [Macdonald(1995)]

$$e_n = \sum_\lambda \epsilon_\lambda z_\lambda^{-1} p_\lambda \qquad \text{and} \qquad h_n = \sum_\lambda z_\lambda^{-1} p_\lambda. \tag{3.7}$$

□

Remark 8 *For both (3.5) and (3.6) in the theorem, when $\lambda = (1^n)$, $z_\lambda^{-1} = 1/n!$ and $\prod_i p_i(\boldsymbol{\theta})^{kn_i(\lambda)} = 1$. Thus the lead term is $1/n!$ and all other terms are strictly less than 1. As k tends to ∞, these terms tend to 0 and $P_{\boldsymbol{\theta}}^{*k}(\mathrm{id}) \sim P_{\boldsymbol{\theta}}^{*k}(\mathrm{rev}) \sim \frac{1}{n!}$. Of course, our work is to quantify this convergence.*

Corollary 9 *For any $\boldsymbol{\theta}$ and all $k \geq 0$, we have*

$$\begin{aligned}\mathrm{SEP}(P_{\boldsymbol{\theta}}^{*k}) &= 1 - n! P_{\boldsymbol{\theta}}^{*k}(\mathrm{rev}), \\ \ell_\infty(P_{\boldsymbol{\theta}}^{*k}) &= n! P_{\boldsymbol{\theta}}^{*k}(\mathrm{id}) - 1.\end{aligned}$$

Proof: The first equality follows from the definition. For the second inequality,

$$\ell_\infty(P_{\boldsymbol{\theta}}^{*k}) = \max(1 - n! P_{\boldsymbol{\theta}}^{*k}(\mathrm{rev}), n! P_{\boldsymbol{\theta}}^{*k}(\mathrm{id}) - 1).$$

In comparing terms, the 1 cancels in both, and the second term is a sum of positive terms while the first has the same terms, some with negative signs. □

Specializing to θ-biased shuffles, Corollary 9 and Theorem 7 imply (1.4) and (1.5).

4 Strong Stationary Times

Repeated shuffling from any of the measures in Section 2 forms a Markov chain $id = W_0, W_1, W_2, \ldots$ taking values in $\mathfrak{S}_n$. A *strong stationary time* (SST) T is a stopping time (meaning $P\{T > k\}$ only depends on $W_0, W_1, \ldots, W_k$) such that for all $k \geq 0$ and all $w \in \mathfrak{S}_n$,

$$P\{W_k = w \mid T \leq k\} = U(w). \tag{4.1}$$

We will build an SST for the Markov chain induced by $P_{\boldsymbol{\theta}}^{*k}$. A basic proposition of this theory [Levin et al.(2009)][Lemma 6.1] is

$$\mathrm{SEP}(k) \leq P\{T > k\} \quad \text{for all } k \geq 0. \tag{4.2}$$

Further, [Aldous and Diaconis(1987)] shows that there always exists a fastest SST T^* satisfying (4.2) with equality for all k.

Background on stationary times is in [Diaconis and Fill(1990)]. In this section, we build a fastest SST (following [Aldous and Diaconis(1987)] and [Fulman(1998)]) involving a birthday problem to bound the right hand side of (4.2). Solving this birthday problem by inclusion-exclusion gives a probabilistic proof of (1.4), Theorem 7 and even the expression for the elementary symmetric function e_n in terms of the power sums (3.7).

Constructing an SST for $P_{\boldsymbol{\theta}}^{*k}$

Consider the inverse process in which cards are labeled i with probability θ_i independently. Then all the cards labeled 1 are removed, keeping them in their same relative order, followed by all cards labeled 2, and so on. This is one inverse $\boldsymbol{\theta}$-shuffle. Repetitions may be realized by labeling each card with a vector with coordinates chosen independently from $\boldsymbol{\theta}$. The first shuffle is read off the first coordinate of each card, the second shuffle off the second coordinate, and so on. Conceptually, each card may be labeled with a vector of infinite length.

Consider the first time T that the first T coordinates of the n cards are distinct. Repeated inverse shuffling sorts the vectors lexicographically, leaving the card with the smallest vector on top, the next smallest second, and so on. By symmetry, at time T, the deck is uniformly distributed, even conditional on $T = k$. This is (4.1). Further, this T is fastest. To see this, note that the reversal permutation $\mathrm{rev} = n, n-1, \ldots, 1$ is a *halting state*: $P\{T \leq k\} \leq P\{W_T = \mathrm{rev}\}$. Indeed, if $W_T = \mathrm{rev}$, then every pair of cards must have a distinct label. Existence of a halting state implies that T is fastest ([Diaconis and Fill(1990)][Remark 2.39] and [Levin et al.(2009)][Remark 6.12]), separation is achieved at rev, and

$$\mathrm{SEP}(k) = P\{T > k\} \quad \text{for all } k \geq 0. \tag{4.3}$$

To work with the right hand side of (4.3), let $A_{i,j}$ be the event that the first k coordinates of the labels on the cards i and j are equal. Thus $P\{A_{i,j}\} = \left(\sum_a \theta_a^2\right)^k$ and

$$\{T > k\} = \bigcup_{1 \le i < j \le n} \{A_{i,j}\}. \tag{4.4}$$

Bounding the probability of the union by the sum of the probabilities yields

$$\mathrm{SEP}(k) \le \binom{n}{2} \left(\sum_a \theta_a^2\right)^k. \tag{4.5}$$

This bound is also derived in [Fulman(1998)]. The asymptotics of Section 5 show it is quite accurate.

Inclusion-Exclusion and the Birthday Problem

Consider this version of the birthday problem: n balls are dropped independently into B boxes with the chance of box i being η_i. If $B_{i,j}$ is the event that balls i and j both wind up in the same box, the chance of success (having two or more balls in the same box) is

$$P(\text{success}) = P\left(\bigcup_{1 \le i < j \le n} B_{i,j}\right). \tag{4.6}$$

Elementary considerations show that the chance of failure (all balls in distinct boxes) is expressible using elementary symmetric functions e_n as $1 - P(\text{success}) = n! e_n(\eta_1, \dots, \eta_B)$. Using the expression for e_n in terms of the power sums (3.7) gives

$$P(\text{success}) = 1 - \sum_{w \in \mathfrak{S}_n} \mathrm{sgn}(w) p_{\lambda(w)}(\boldsymbol{\eta}) = 1 - n! \sum_{\lambda \vdash n} (-1)^{n-\ell(\lambda)} z_\lambda^{-1} p_\lambda(\boldsymbol{\eta}) \tag{4.7}$$

The inclusion-exclusion expansion of (4.6) gives a sum of polynomials which must match the neat expressions in (4.7). This may be seen explicitly using the inclusion-exclusion formula for the chromatic polynomial in [Stanley(1995)]. For example,

$$P\{B_{1,2} \cup B_{1,3} \cup B_{2,3}\} = 3P(B_{1,2}) - 3P(B_{1,2} \cap B_{2,3}) + P(B_{1,2} \cap B_{1,3} \cap B_{2,3}) = 3(\sum p_j^2) - 2(\sum p_j^3),$$

while (4.7) gives $6(-\frac{1}{2} p_{(2,1)}(\boldsymbol{\eta}) + \frac{1}{3} p_3(\boldsymbol{\eta}))$ matching (4.6).

Remark 10 *Since separation is achieved (uniquely) at the reversal permutation,* (4.3)*,* (4.4)*,* (4.6)*,* (4.7) *give a probabilistic proof of Theorem 7.*

Remark 11 *This connection between inclusion–exclusion, birthday problems and symmetric functions seems generally useful. See, for example, [Montgomery and Soundararajan(2004)][pg. 604–605].*

5 Main Result

This section derives the asymptotic results of Theorem 1 and some extensions. Without loss of generality, suppose $1/2 \leq \theta \leq 1$. To bound the ℓ_∞ distance, using Corollary 9 together with (1.5) and (1.4), we are interested in

$$\ell(k,n) = \sum_{w \in S_n} \prod_j \theta_j^{k n_j(w)}, \tag{5.1}$$

where $\theta_j = \theta^j + (1-\theta)^j$ and $n_j(w)$ denotes the number of j cycles in the permutation w. If

$$f_n(x_1, \ldots, x_n) = \sum_{w \in S_n} \prod_j x_j^{n_j(w)}$$

then we have the identity

$$\sum_{n=0}^{\infty} \frac{z^n}{n!} f_n(x_1, \ldots, x_n) = \exp\Big(\sum_{j=1}^{\infty} \frac{z^j}{j} x_j \Big). \tag{5.2}$$

Therefore we have that

$$\sum_{n=0}^{\infty} \frac{z^n}{n!} \ell(k,n) = \exp\left(\sum_{j=1}^{\infty} \frac{z^j}{j} \theta_j^k \right). \tag{5.3}$$

Theorem 12 *Define*

$$M = M(k,n) = \sum_{j=2}^{\infty} n^j \theta_j^k.$$

If $M \leq \sqrt{n}/(10 \log n)$, then we have

$$\ell(k,n) = \exp\left(\sum_{j=2}^{\infty} \frac{n^j}{j} \theta_j^k \right) \left(1 + O\left(\frac{1+M}{\sqrt{n}} \right) \right).$$

Proof: Set $F_k(z) = \sum_{j=1}^{\infty} \frac{z^j}{j} \theta_j^k$. By the residue theorem we have

$$\ell(k,n) = \frac{n!}{2\pi i} \int_{|z|=n} \exp(F_k(z)) z^{-n} \frac{dz}{z} = \frac{n!}{2\pi n^n} \int_{-\pi}^{\pi} \exp(F_k(ne^{ix}) - inx) dx.$$

We divide the integral into the ranges when $|x| \leq (\log n)/\sqrt{n}$ which gives the main contribution, and $\pi \geq |x| > (\log n)/\sqrt{n}$.

Consider first the range $|x| \leq (\log n)/\sqrt{n}$. Here we have

$$F_k(ne^{ix}) = ne^{ix} + \sum_{j=2}^{\infty} \frac{n^j}{j} \theta_j^k e^{ijx} = ne^{ix} + \sum_{j=2}^{\infty} \frac{n^j}{j} \theta_j^k + O(|x|M),$$

since $e^{ijx} = 1 + O(j|x|)$. Therefore, using $ne^{ix} = n + inx - nx^2/2 + O(|x|^3 n)$ and Stirling's formula, the integral over this region is

$$\frac{n!}{2\pi n^n} \int_{|x| \le (\log n)/\sqrt{n}} \exp\Big(n - \frac{nx^2}{2} + \sum_{j=2}^{\infty} \frac{n^j}{j} \theta_j^k + O\Big(|x|^3 n + |x| M\Big)\Big) dx = \exp\Big(\sum_{j=2}^{\infty} \frac{n^j}{j} \theta_j^k\Big)\Big(1 + O\Big(\frac{1+M}{\sqrt{n}}\Big)\Big).$$

Now consider the range $\pi \ge |x| > (\log n)/\sqrt{n}$. Here we have

$$Re(F_k(ne^{ix})) \le F_k(n) - n(1 - \cos(x)) \le F_k(n) - c(\log n)^2,$$

for some positive constant c. Using Stirling's formula again, the contribution of this segment of the integral is therefore

$$\ll \frac{n!}{n^n} \exp(F_k(n) - c(\log n)^2) \ll \sqrt{n} \exp\Big(\sum_{j=2}^{\infty} \frac{n^j}{j} \theta_j^k - c(\log n)^2\Big),$$

which may be absorbed into our error term. $\square$

From this Theorem we can read off the behavior of the ℓ^∞ distance after k biased shuffles. First consider the case when $(1-\theta)\log n$ is large. In this range put

$$k = \Big\lfloor \frac{2\log n - \log 2 + c}{-\log(\theta^2 + (1-\theta)^2)} \Big\rfloor. \tag{5.4}$$

We find that the contribution to $M(k,n)$ arises mainly from $j = 2$ and so $M(k,n) \ll e^{-c}$, and we have

$$\ell(k,n) \sim \exp(e^{-c}), \tag{5.5}$$

so that the ℓ^∞ distance behaves like $\exp(e^{-c}) - 1$, and similarly the separation distance behaves like $1 - \exp(-e^{-c})$, in agreement with [Diaconis et al.(1992)].

Next consider the case when $(1-\theta)\log n = \kappa \in [0, \infty)$. Keep the notation above for k, here we find that $n^2\theta_2^k = 2e^{-c}$, as before, and for $j \ge 3$,

$$n^j \theta_j^k \sim \exp(\frac{j}{2}(-\kappa + \log 2 - c)\Big). \tag{5.6}$$

Therefore, if $c > \log 2 - \kappa$, then $M(k,n)$ is small, and Theorem 12 applies. Moreover in this case we have

$$\ell(k,n) \sim \exp\Big(e^{-c} + \sum_{j=3}^{\infty} \frac{1}{j} \exp(\frac{j}{2}(-\kappa + \log 2 - c))\Big). \tag{5.7}$$

Finally, consider the extreme case $\theta = 1 - 1/n$. It is convenient here to define $k = n\log n + cn$. Then $n^j\theta_j^k \sim e^{-jc}$ for $j \ge 2$, and $M(k,n)$ is small provided $c > 0$. In that case we have

$$\ell(k,n) \sim \exp\Big(\sum_{j=2}^{\infty} \frac{e^{-jc}}{j}\Big) = \frac{e^{-e^{-c}}}{1 - e^{-c}}. \tag{5.8}$$

Compare with Theorem 1.1 of [Diaconis et al.(1992)].

Acknowledgements

The authors thank Amy Pang for helpful conversations about the hyperplane perspective, and Jason Fulman for careful comments and references.

References

[Aldous and Diaconis(1987)] D. Aldous and P. Diaconis. Strong uniform times and finite random walks. *Adv. in Appl. Math.*, 8(1):69–97, 1987.

[Assaf et al.(2011)] S. Assaf, P. Diaconis, and K. Soundararajan. A rule of thumb for riffle shuffling. *Ann. Appl. Probab.*, 21(3):843–875, 2011.

[Athanasiadis and Diaconis(2010)] C. A. Athanasiadis and P. Diaconis. Functions of random walks on hyperplane arrangements. *Adv. in Appl. Math.*, 45(3):410–437, 2010.

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

[Bidigare et al.(1999)] P. Bidigare, P. Hanlon, and D. Rockmore. A combinatorial description of the spectrum for the Tsetlin library and its generalization to hyperplane arrangements. *Duke Math. J.*, 99(1):135–174, 1999.

[Brown and Diaconis(1998)] K. S. Brown and P. Diaconis. Random walks and hyperplane arrangements. *Ann. Probab.*, 26(4):1813–1854, 1998.

[Conger and Howald(2010)] M. A. Conger and J. Howald. A better way to deal the cards. *Amer. Math. Monthly*, 117(8):686–700, 2010.

[Diaconis(2003)] 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.

[Diaconis and Fill(1990)] P. Diaconis and J. A. Fill. Strong stationary times via a new form of duality. *Ann. Probab.*, 18(4):1483–1522, 1990.

[Diaconis et al.(1992)] P. Diaconis, J. A. Fill, and J. Pitman. Analysis of top to random shuffles. *Combin. Probab. Comput.*, 1(2):135–155, 1992.

[Diaconis et al.(2011)] P. Diaconis, A. Pang, and A. Ram. Hopf algebras and Markov chains: Two examples and a theory, 2011. Preprint.

[Fulman(1998)] J. Fulman. The combinatorics of biased riffle shuffles. *Combinatorica*, 18(2):173–184, 1998.

[Gessel(1984)] I. M. Gessel. Multipartite P-partitions and inner products of skew Schur functions. In *Combinatorics and algebra (Boulder, Colo., 1983)*, volume 34 of *Contemp. Math.*, pages 289–317. Amer. Math. Soc., Providence, RI, 1984.

[Lalley(1996)] S. P. Lalley. Cycle structure of riffle shuffles. *Ann. Probab.*, 24(1):49–73, 1996.

[Lalley(2000)] S. P. Lalley. On the rate of mixing for p-shuffles. *Ann. Appl. Probab.*, 10(4):1302–1321, 2000.

[Levin et al.(2009)] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.

[Macdonald(1995)] I. G. Macdonald. *Symmetric functions and Hall polynomials*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, second edition, 1995. With contributions by A. Zelevinsky, Oxford Science Publications.

[Montgomery and Soundararajan(2004)] H. L. Montgomery and K. Soundararajan. Primes in short intervals. *Comm. Math. Phys.*, 252(1-3):589–617, 2004.

[Stanley(1999)] R. P. Stanley. *Enumerative combinatorics. Vol. 2*, volume 62 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999.

[Stanley(1995)] R. P. Stanley. A symmetric function generalization of the chromatic polynomial of a graph. *Adv. Math.*, 111(1):166–194, 1995.

[Stanley(2001)] R. P. Stanley. Generalized riffle shuffles and quasisymmetric functions. *Ann. Comb.*, 5 (3-4):479–491, 2001. Dedicated to the memory of Gian-Carlo Rota (Tianjin, 1999).