

---

---

# Analysis of Top To Random Shuffles

---

PERSI DIACONIS<sup>†</sup>, JAMES ALLEN FILL<sup>‡</sup> and JIM PITMAN<sup>†§</sup>

<sup>†</sup> Dept. of Mathematics, Harvard University, Cambridge, MA 02138

<sup>‡</sup> Dept. of Mathematical Sciences, The Johns Hopkins University, Baltimore, MD 21218-2689

<sup>†§</sup> Dept. of Statistics, University of California, Berkeley, CA 94720

*Received 14 January 1992*

A deck of  $n$  cards is shuffled by repeatedly taking off the top  $m$  cards and inserting them in random positions. We give a closed form expression for the distribution after any number of steps. This is used to give the asymptotics of the approach to stationarity: for  $m$  fixed and  $n$  large, it takes  $\frac{n}{m}(\log n + c)$  shuffles to get close to random. The formulae lead to new subalgebras in the group algebra of the symmetric group.

## 1. Introduction

Aldous and Diaconis [1] studied the top to random shuffle defined as follows: given a deck of  $n$  cards, remove the top card and put it back in the deck at random. More precisely, let

$$(1.1) \quad Q(\pi) := \begin{cases} \frac{1}{n} & \text{if } \pi \text{ is the identity or one of the cycles } (1\ 2 \cdots i), \quad 1 < i \leq n \\ 0 & \text{otherwise} \end{cases}$$

Let  $U(\pi) = \frac{1}{n!}$  denote the uniform distribution. Aldous and Diaconis showed it takes order  $n \log n$  shuffles for convergence to uniformity in the sense that for  $k = n \log n + cn$

$$(1.2a) \quad \|Q^{*k} - U\| \leq e^{-c} \quad \text{for } c > 0$$

$$(1.2b) \quad \|Q^{*k} - U\| \rightarrow 1 \quad \text{as } n \rightarrow \infty, \quad c = c_n \rightarrow -\infty,$$

where total variation is defined as

$$(1.3) \quad \|P - Q\| = \max_{A \subset S_n} |P(A) - Q(A)|.$$

The present paper refines (1.2) by giving a closed form expression for  $Q^{*k}(\pi)$  (Corollary 2.1). This is used to derive the following limit for total variation.

**Theorem 1.1.** For  $Q$  defined by (1.1), let  $k = \lfloor n \log n + cn \rfloor$  for  $c$  fixed in  $\mathbb{R}$ . Then, for large  $n$ ,

$$(1.4) \quad \|Q^{*k} - U\| = f(c) + o(1)$$

with

$$(1.5a) \quad f(c) = \frac{1}{2}(1 - e^{-e^{-c}}(1 + e^{-c})) \quad \text{for } c \geq 0,$$

$$(1.5b) \quad f(c) = 1 - e^{-e^{-c}} \sum_{u=0}^{\ell^*} e^{-uc} \left( \frac{1}{u!} - \frac{1}{(\ell^* + 1)!} \right) - \frac{1}{(\ell^* + 1)!} \quad \text{for } c < 0$$

with

$$\ell^* = \ell^*(c) = \lfloor \frac{\log(e^{e^{|c|}}(e^{|c|} - 1) + 1)}{|c|} \rfloor - 1.$$

Theorem 1.1 follows from our analysis of the top  $m$  to random shuffle. Here the top  $m$  cards are removed from a deck of  $n$  cards and inserted at random into the remaining  $n - m$  cards one at a time. We find closed form expressions for repeated convolutions of such measures where  $m$  is allowed to vary from shuffle to shuffle. This is the content of Section 2.

Section 3 derives the asymptotics of the total variation distance. It presents a number of families where the cutoff phenomenon observed by Aldous and Diaconis [1] can be rigorously determined.

The closed form expressions of Section 2 give rise to some new semisimple algebras. Briefly, say a permutation has a descent at  $i$  if  $\pi(i + 1) < \pi(i)$ . Let  $F(\pi)$  be the location of the first descent in  $\pi$  with  $F(id) = n$ , where  $id$  is the identity. Let  $F_i = \sum_{F(\pi)=i} \pi$  be the formal sum of permutations with first descent at  $i$ . In Section 4 we show that the  $F_i$ ,  $1 \leq i \leq n$ , span a commutative semisimple subalgebra of the group algebra of the symmetric group. Thus  $F_i F_j = F_j F_i = \sum f_{ij}^k F_k$ . A closely related result is the determination of the eigenvalues of the Markov chain underlying the random walk (1.1). These are the numbers  $0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-2}{n}, 1$ , where  $i/n$  occurs with multiplicity the number of permutations  $\pi$  in  $S_n$  with exactly  $i$  fixed points,  $0 \leq i \leq n$ .

There is another natural generalization of the top to random shuffle. Suppose the top  $m$  cards are removed and inserted into random positions, preserving their relative order. This shuffle is analyzed in Section 5. The argument gives a probabilistic proof that Solomon's descent algebra is an algebra. Section 6 treats related examples.

The arguments in this paper are in close parallel to the arguments of Bayer and Diaconis [2] on repeated riffle shuffles. They too find closed form expressions and derive (different) new algebras. Some of the parallels are explained in Sections 3 and 5.

The inverse of top to random removes a random card and puts it at the top. This is a special case of a dynamic list management algorithm actively studied in computer science. Phatarfod [15] gives a recent survey. Our results give the first rate of convergence for such a problem.

We conclude this introduction with one crucial piece of notation for permutations. Places in the deck are numbered 1 through  $n$  from top to bottom. Initially, cards labelled 1 through  $n$  are in perfect order. After shuffling, the state of the deck is the permutation  $\pi$  with  $\pi(i)$  the label of the card at position  $i$ . Thus  $\pi(1) = 5$  means that the card labelled 5 is on top of the deck and  $\pi^{-1}(6) = 3$  means that the card labelled 6 is third from the top. Reversing the usual convention, we multiply permutations left to right: for example, following top to third ( $\pi = \text{cycle } (1\ 2\ 3)$ ) by top to fifth ( $\sigma = \text{cycle } (1\ 2\ 3\ 4\ 5)$ ) results in the cycle  $(1\ 3\ 4\ 5\ 2)$ , which is the composition  $\sigma$  followed by  $\pi$  and which we denote as  $\sigma\pi$ . A *shuffle* is a probability on the symmetric group  $S_n$ . The result of successive shuffles is given by convolution:  $(Q * P)(\sigma) = \sum_{\pi} Q(\sigma\pi^{-1})P(\pi)$  is the result of doing  $P$  followed by  $Q$ .

### 2. Top $m$ to random shuffle

Fix an integer  $m$ ,  $0 \leq m \leq n$ . A top  $m$  to random shuffle can be performed by taking the top  $m$  cards from a deck of  $n$  cards and inserting these one at a time at random into the remaining  $n - m$  cards. Of course, if  $m = 0$ , the identity permutation results, while if  $m = n - 1$  or  $n$ , the resulting distribution is uniform. Elementary considerations show there are  $n!/(n - m)!$  possible arrangements following a top  $m$  to random shuffle, each of these being equally likely.

The possible arrangements have a simple description in terms of  $L(\pi)$ , the length of the rising sequence containing  $n$ : for  $\pi \in S_n$ ,  $L(\pi) = \ell$  provided

$$(2.1) \quad \pi^{-1}(n) > \pi^{-1}(n - 1) > \dots > \pi^{-1}(n - \ell + 1) < \pi^{-1}(n - \ell).$$

Thus  $L(\pi) = \ell$  means that the original bottom  $\ell$  cards appear in the shuffled deck in their original order, but the bottom  $\ell + 1$  cards do not. Clearly  $1 \leq L(\pi) \leq n$ . It is elementary that under the uniform distribution  $P\{L(\pi) \geq \ell\} = \frac{1}{\ell!}$ ,  $1 \leq \ell \leq n$ . A top  $m$  to random shuffle  $Q_m$  is uniform over all  $\pi \in S_n$  with  $L(\pi) \geq n - m$ :

$$(2.2) \quad Q_m(\pi) = \begin{cases} 0 & \text{if } L(\pi) < n - m \\ \frac{(n-m)!}{n!} & \text{if } L(\pi) \geq n - m. \end{cases}$$

For further developments, it is useful to consider mixtures of the  $Q_m$ : for  $\mu$  a probability on  $\{0, 1, 2, \dots, n\}$ , define

$$(2.3) \quad Q_\mu(\pi) = \sum_{m=0}^n \mu(m)Q_m(\pi).$$

It is clear that under  $Q_\mu$  the conditional distribution of  $\pi$ , given  $L(\pi) = \ell$ , is uniform over  $\{\pi : L(\pi) = \ell\}$ .

The main result of this section says that the convolution of  $Q_{\mu_i}$  shuffles is again of the same form. To state the result, we need to introduce a product on the probabilities on  $\{0, 1, \dots, n\}$ . For  $\mu$  and  $\nu$  probabilities on  $\{0, 1, \dots, n\}$ , define  $\nu\#\mu$  as the distribution of the number of occupied cells when integers  $i$  and  $j$  are chosen from  $\mu$  and  $\nu$ , and  $i$  balls are dropped randomly into distinct cells (chosen from  $n$ ), and then  $j$  balls are independently

dropped randomly into distinct cells (chosen from the same  $n$ ). Thus, define

$$(2.4) \quad v\#\mu(k) = \sum_{i,j} \mu(i)v(j) \binom{n-i}{k-i} \binom{i}{j-(k-i)} / \binom{n}{j}.$$

It is evident that this forms an associative, commutative product on probabilities. With this notation we can state:

**Theorem 2.1.** *Let  $\mu_1, \mu_2, \dots, \mu_k$  be probabilities on  $\{0, 1, \dots, n\}$ . Then*

$$Q_{\mu_k} * \dots * Q_{\mu_1} = Q_{\mu}$$

with  $\mu = \mu_k\#\mu_{k-1}\#\dots\#\mu_1$ .

**Remark.** The map taking  $\mu \mapsto Q_{\mu}$  is an isomorphism of the convex set of probabilities on  $\{0, 1, \dots, n\}$  under  $\#$  into the convex set of probabilities on  $S_n$  under convolution.

**Proof.** Theorem 2.1 is proved by a marking argument. Consider the first shuffle to be performed,  $Q_{\mu_1}$ . Let  $M_1$  be distributed as  $\mu_1$ . Thus, to begin,  $M_1$  cards are cut off. As each card is inserted, it is marked; we picture this as a large check-mark on its back. This results in marks on cards labeled  $1, 2, \dots, M_1$ . It is clear that these cards are in random locations. Next let  $M_2$  be distributed as  $\mu_2$ . As each of the top  $M_2$  cards is inserted, any unmarked cards are marked. The newly marked cards are those labeled  $M_1 + 1$  up to what may be called  $M_2\#M_1$ , this having a hypergeometric distribution

$$P\{M_2\#M_1 = k | M_1, M_2\} = \frac{\binom{n-M_1}{k-M_1} \binom{M_1}{M_2-(k-M_1)}}{\binom{n}{M_2}}.$$

Since the newly marked cards are inserted into random positions and such insertion doesn't change the fact that previously labeled cards are in random positions, the whole results in a  $\mu_2\#\mu_1$  shuffle as defined in (2.3)-(2.4) above. Continuing inductively completes the proof. □

**Corollary 2.1.** *After  $k$  independent top to random shuffles, the probability of the permutation  $\pi$  is*

$$(2.5) \quad Q_1^{*k}(\pi) = \frac{1}{n!} \sum_{u=0}^{L(\pi)} u! P_k(u)$$

with  $P_k(u)$  the chance that there are  $u$  unoccupied cells after  $k$  balls are dropped into  $n$  boxes. Thus (see Feller [6], p. 102)

$$(2.6) \quad P_k(u) = \sum_{v=u}^n (-1)^{v-u} \binom{n}{v} \binom{v}{u} \left(1 - \frac{v}{n}\right)^k.$$

Further, under  $Q_1^{*k}$  the distribution of  $L(\pi)$ , the length of the rising sequence containing  $n$  (see (2.1)), is

$$(2.7) \quad Q_1^{*k}(L(\pi) \geq \ell) = \sum_{u=0}^{\ell-1} P_k(u) \frac{u!}{\ell!} + \sum_{u \geq \ell} P_k(u).$$

Finally, under  $Q_1^{*k}$  the conditional distribution given  $L(\pi) = \ell$  is uniform over  $\{\pi : L(\pi) = \ell\}$ .

**Remark.** Similar closed form expressions are available for more complex schemes. For any choice  $m_1, m_2, \dots, m_k$

$$Q_{m_k} * Q_{m_{k-1}} * \dots * Q_{m_1}(\pi) = \frac{1}{n!} \sum_{u=0}^{L(\pi)} u! P_{m_1, \dots, m_k}(u)$$

with

$$(2.8) \quad P_{m_1, \dots, m_k}(u) = \sum_{v=u}^n (-1)^{v-u} \binom{n}{v} \binom{v}{u} \prod_{j=1}^k \frac{\binom{n-v}{m_j}}{\binom{n}{m_j}}.$$

A derivation of (2.8) is given by Holst [13], who cites many earlier sources.

Formulas (2.5) and (2.8) show how to convolve point masses for the convolution  $\mu \# \nu$  defined at (2.4). Another example yielding a closed form formula is the following, used in Section 3:

**Lemma 2.1.** For the convolution  $\#$  defined at (2.4),

$$(2.9) \quad \text{binomial}(n, p_1) \# \text{binomial}(n, p_2) = \text{binomial}(n, 1 - (1 - p_1)(1 - p_2)).$$

**Proof.** Let  $R$  and  $S$  be independent binomial variates with parameters  $(n, p_1)$  and  $(n, p_2)$ , respectively. To begin with,  $R$  balls are dropped into distinct boxes randomly chosen from  $n$ . The sequence of indicators of occupied boxes is therefore a sequence of  $n$  independent trials with probability  $p_1$ . A similar description holds for the second stage. The chance that a given box is still unoccupied after both stages is thus  $(1 - p_1)(1 - p_2)$ . The result follows because different boxes are independent.  $\square$

### 3. Total variation calculations

Since the convolution of top to random measures is of the same form, it suffices to bound the total variation distance of one of these measures to the stationary distribution. The measure  $Q_{\mu_n}$  will be close to uniform when the measure  $\mu_n$  concentrates near  $n$ . For repeated convolutions,  $\mu_n(n - \cdot)$  tends to be approximately Poisson. Consider the following three examples:

**Example 1. Top to random.** The convolution of  $k$  top to random shuffles is a  $Q_{\mu_n}$  shuffle, where  $\mu_n$  is the law of the number of occupied cells when  $k$  balls are dropped at random into  $n$  boxes. This is the classical coupon collector's problem. As shown in Feller [6], p. 105, when  $k = \lfloor n \log n + cn \rfloor$ , the number of unoccupied cells has an approximate Poisson distribution with parameter  $e^{-c}$ . Using this in Proposition 3.1 below, gives a proof of Theorem 1.1.

**Example 2.** The convolution of  $k$  top  $M$  to random shuffles, where  $M$  is  $\text{binomial}(n, \frac{1}{n})$ , is of the same form with  $M$  having a  $\text{binomial}(n, 1 - (1 - \frac{1}{n})^k)$  distribution. Take  $k =$

$[n \log n + cn]$ . The Poisson approximation to the binomial shows that the number of unoccupied cells has an approximate  $\text{Poisson}(e^{-c})$  distribution, as in Example 1. Using this in Proposition 3.1 below gives the same conclusion as Theorem 1.1 for this measure.

**Example 3.** If  $M$  has a  $\text{binomial}(n, \frac{1}{2})$  distribution, the convolution of  $k$  top  $M$  to random shuffles is of the same form, with  $M$  having a  $\text{binomial}(n, 1 - (\frac{1}{2})^k)$  distribution. If  $k = \log_2 n + c$ , the number of unoccupied cells has an approximate  $\text{Poisson}(2^{-c})$  distribution.

These examples motivate the following result:

**Proposition 3.1.** *Let  $\mu_n$  be a probability on  $\{0, 1, \dots, n\}$  such that  $\mu_n(n-j) \rightarrow e^{-\lambda} \lambda^j / j!$  for each fixed  $j$  as  $n \rightarrow \infty$ . Let  $Q_{\mu_n}$  be defined by (2.3). Then, as  $n \rightarrow \infty$ ,*

$$(3.1) \quad \|Q_{\mu_n} - U\| = 1 - e^{-\lambda} \sum_{u=0}^{\ell^*} \lambda^u \left( \frac{1}{u!} - \frac{1}{(\ell^* + 1)!} \right) - \frac{1}{(\ell^* + 1)!} + o(1)$$

with

$$\ell^* = \begin{cases} 1 & \text{for } 0 < \lambda \leq 1 \\ \lfloor \frac{\log(e^\lambda(\lambda-1)+1)}{\log \lambda} \rfloor - 1 & \text{for } \lambda > 1. \end{cases}$$

**Proof.** Let  $L(\pi)$  be the length of the rising sequence containing  $n$ . Under both  $Q_{\mu_n}$  and  $U$ , the conditional distribution of  $\pi$ , given  $L(\pi) = \ell$ , is uniform. It follows that the variation distance between  $Q_{\mu_n}$  and  $U$  equals the variation distance between the induced laws of  $L$ . As in Section 2,

$$(3.2) \quad U\{L \geq \ell\} = \frac{1}{\ell!},$$

$$(3.3) \quad \begin{aligned} Q_{\mu_n}\{L \geq \ell\} &= \sum_{u=0}^{\ell-1} \mu_n(n-u) \frac{u!}{\ell!} + \sum_{u \geq \ell} \mu_n(n-u) \\ &= 1 - \sum_{u=0}^{\ell-1} \mu_n(n-u) \left(1 - \frac{u!}{\ell!}\right). \end{aligned}$$

Taking differences and rearranging,

$$\begin{aligned} U\{L = \ell\} &= \frac{\ell}{(\ell+1)!} \quad \text{for } \ell = 1, 2, \dots, n-1, \quad U\{L = n\} = \frac{1}{n!}, \\ Q_{\mu_n}\{L = \ell\} &= U\{L = \ell\} \sum_{u=0}^{\ell} \mu_n(n-u) u! \quad \text{for } \ell = 1, 2, \dots, n. \end{aligned}$$

Since evidently  $Q_{\mu_n}\{L = \ell\} / U\{L = \ell\}$  increases with  $\ell$ , it follows that

$$(3.4) \quad \begin{aligned} \|Q_{\mu_n} - U\| &= \sum_{\ell=0}^{\ell_n^*} [U\{L = \ell\} - Q_{\mu_n}\{L = \ell\}] = Q_{\mu_n}\{L \geq \ell_n^* + 1\} - U\{L \geq \ell_n^* + 1\} \\ &= 1 - \sum_{u=0}^{\ell_n^*} \mu_n(n-u) \left(1 - \frac{u!}{(\ell_n^* + 1)!}\right) - \frac{1}{(\ell_n^* + 1)!}, \end{aligned}$$

where  $\ell_n^*$  is the largest integer  $\ell$  for which  $\sum_{u=0}^{\ell} \mu_n(n-u)u! \leq 1$ . It follows easily that (3.1) holds, where  $\ell^*$  is the largest  $\ell$  for which

$$(3.5) \quad \sum_{u=0}^{\ell} \lambda^u \leq e^\lambda.$$

Now  $1 + \lambda \leq e^\lambda$  for all  $\lambda \in [0, \infty)$ , so  $\ell^*(\lambda) \geq 1$  for all  $\lambda$ . A little calculus shows that

$$1 + \lambda + \lambda^2 > e^\lambda \quad \text{for } 0 < \lambda \leq 1,$$

in which case  $\ell^*(\lambda) = 1$ . For fixed  $\lambda > 1$ , (3.5) says  $\ell \leq \frac{\log(e^\lambda(\lambda-1)+1)}{\log \lambda} - 1$ , and so  $\ell^*$  is given as in the statement of the proposition. This proves (3.1).  $\square$

**Remarks**

1. It is straightforward to see that for fixed  $\ell \geq 2$ , the equation  $1 + \lambda + \dots + \lambda^\ell = e^\lambda$  has exactly one strictly positive root. Call this  $\lambda_\ell$ . Then  $0 =: \lambda_1 < \lambda_2 < \lambda_3 < \dots \uparrow \infty$ . Now  $\ell^*(\lambda)$  starts at 1 at  $\lambda = 0$  and is a step function increasing by 1 at each  $\lambda_\ell$ , for  $\ell > 1$ . A table of approximate values of  $\lambda_\ell$  is

$\ell$	1	2	3	4	5	6	7
$\lambda_\ell$	0	1.8	5.1	8.8	12.8	17.1	21.5

Thus, for  $\lambda \in (5.1, 8.8)$ , the sum in Proposition 3.1 has four terms.

2. For  $\lambda < 1$ , the limiting (in  $n$ ) variation distance  $v(\lambda)$  equals

$$\frac{1}{2}(1 - e^{-\lambda}(1 + \lambda)) = \frac{1}{2} \sum_{j=2}^{\infty} (-1)^j (j-1) \frac{\lambda^j}{j!}.$$

In particular,  $v(\lambda) = \frac{1}{4}\lambda^2(1 + O(\lambda)) = o(1)$  as  $\lambda \downarrow 0$ . Thus, in Theorem 1.1 we have the exponential decay  $f(c) \sim \frac{1}{4}e^{-2c}$  as  $c \rightarrow \infty$ .

3. Asymptotics as  $\lambda \rightarrow \infty$  are more involved. To see how to proceed, observe first that

$$\log(e^\lambda(\lambda-1)+1) = \lambda + \log \lambda - (1 + o(1)) \frac{1}{\lambda} \quad \text{as } \lambda \rightarrow \infty;$$

hence,  $\ell^* = \lfloor \frac{\lambda - (1+o(1))\lambda^{-1}}{\log \lambda} \rfloor = \lceil \frac{\lambda}{\log \lambda} \rceil - 1$  for all large  $\lambda$ . Next, writing  $p(u)$  for the Poisson( $\lambda$ ) probability mass  $e^{-\lambda}\lambda^u/u!$  and  $F(\ell)$  for the cumulative  $\sum_{u \leq \ell} p(u)$ , we have the simple bounds

$$\begin{aligned} 1 &\leq \frac{F(\ell)}{p(\ell)} \leq \sum_{u \leq \ell} \frac{\ell(\ell-1) \cdots (\ell-u+1)}{\lambda^{\ell-u}} \\ &\leq \sum_{u \leq \ell} \left(\frac{\ell}{\lambda}\right)^{\ell-u} = \frac{1 - (\ell/\lambda)^{\ell+1}}{1 - \ell/\lambda} = 1 + o(1), \end{aligned}$$

where the last estimate holds provided  $\ell/\lambda = o(1)$ . In particular,

$$(3.6) \quad \begin{aligned} v(\lambda) &= 1 - (1 - o(1))F(\ell^*) - \frac{1}{(\ell^* + 1)!} = 1 - (1 + o(1))p(\ell^*) - \frac{1}{(\ell^* + 1)!} \\ &= 1 - \frac{1}{(\ell^* + 1)!} [(1 + o(1))e^{-\lambda} \lambda^{\ell^*} \ell^* + 1] \end{aligned}$$

as  $\lambda \rightarrow \infty$ ; the first equality here follows from (3.1) since  $0 < u! / (\ell^* + 1)! \leq 1 / (\ell^* + 1) = o(1)$  for  $0 \leq u \leq \ell^*$ .

Neither of the terms in brackets in (3.6) can be omitted. Indeed, consideration of the fractional part of  $\lambda / \log \lambda$  reveals

$$\liminf_{\lambda \rightarrow \infty} e^{-\lambda} \lambda^{\ell^*} \ell^* / (1 / \log \lambda) = 1 \quad \text{while} \quad \limsup_{\lambda \rightarrow \infty} e^{-\lambda} \lambda^{\ell^*} \ell^* / (\lambda / \log \lambda) = 1.$$

To gauge the size of  $v(\lambda)$  relative to the single parameter  $\lambda$ , recall  $\ell^* + 1 = \lceil \lambda / \log \lambda \rceil$  and apply Stirling's formula to  $(\ell^* + 1)!$ . The result, with  $L \equiv \log$ , is

$$\begin{aligned} 1 - v(\lambda) &= (1 + o(1)) \frac{1}{\sqrt{2\pi}} \exp\left\{-\lambda + \frac{\lambda LL\lambda}{L\lambda} + \frac{\lambda}{L\lambda} - \frac{1}{2}(L\lambda - LL\lambda)\right\} \\ &\quad \times \left[ \frac{1}{L\lambda} + \exp\left\{-\left(\left\lceil \frac{\lambda}{L\lambda} \right\rceil - \frac{\lambda}{L\lambda}\right)(L\lambda - LL\lambda)\right\} \right], \end{aligned}$$

which oscillates between  $\frac{1}{\sqrt{2\pi}} \exp\left\{-\lambda + \frac{\lambda LL\lambda}{L\lambda} + \frac{\lambda}{L\lambda} - \frac{1}{2}L\lambda \pm \frac{1}{2}LL\lambda\right\}$ . In any case,

$$1 - v(\lambda) = \exp\left\{-\lambda + \frac{\lambda LL\lambda}{L\lambda} + \frac{\lambda}{L\lambda} - \frac{1}{2}L\lambda + O(LL\lambda)\right\} = \exp(-(1 + o(1))\lambda) = o(1),$$

and  $f(c)$  of Theorem 1.1 converges to 1 at a doubly exponential rate as  $c \rightarrow -\infty$ .

4. A standard alternative measure of discrepancy from stationarity, the so-called separation

$$\text{sep}(Q_{\mu_n}, U) = \max_{\pi \in \mathcal{S}_n} (1 - Q_{\mu_n}(\pi) / U(\pi)),$$

is easier to handle than total variation. Indeed, by (2.2)–(2.3) the least likely permutation under  $Q_{\mu_n}$  is rev, the reverse of natural order. So, for  $\lambda \in (0, \infty)$

$$\begin{aligned} \text{sep}(Q_{\mu_n}, U) &= 1 - n! Q_{\mu_n}(\text{rev}) = 1 - \mu_n(n) - \mu_n(n-1) \\ &= 1 - e^{-\lambda}(1 + \lambda) + o(1) \end{aligned}$$

in the setting of Proposition 1 above. In Theorem 1.1 we thus have that  $\lim_{n \rightarrow \infty} \text{sep}(Q_1^{*k}, U)$  is twice the expression  $f(c)$  of (1.5a), whether  $c \geq 0$  or  $c < 0$ . The convergence rate is again exponential as  $c \rightarrow \infty$ , and doubly exponential as  $c \rightarrow -\infty$ .

5. It is instructive to compare Example 3 of this section with the results of Bayer and Diaconis [2]. In their shuffle, a binomial( $n, \frac{1}{2}$ ) number of cards is cut off and riffled into the remaining cards. In the shuffle of Example 3, a binomial( $n, \frac{1}{2}$ ) number of cards is cut off, thoroughly shuffled, and then riffled into the remaining cards. Bayer and Diaconis showed it takes  $\frac{3}{2} \log_2 n + c$  shuffles to mix up  $n$  cards. The results of Example 3, coupled with Proposition 3.1, show it takes  $\log_2 n + c$  top binomial( $n, \frac{1}{2}$ ) to random shuffles to mix up  $n$  cards. It is intuitively clear that this second method is faster. The calculation shows

it is not much faster. In the Bayer-Diaconis result, the total variation has a Gaussian limiting shape as compared to the somewhat complex behavior of Proposition 3.1.

The final result of this section derives the limiting total variation for a class of measures that includes top  $m$  to random for any fixed  $m$ . The result gives a class of probabilities where an explicit cutoff occurs.

**Proposition 3.2.** Fix  $b$ . Let  $\mu$  be a probability on  $\{0, 1, 2, \dots, b\}$  with positive mean  $\bar{\mu}$ . Let  $Q_\mu$  be defined on  $S_n$  by (2.2)–(2.3). Suppose

$$(3.6) \quad k = \frac{n}{\bar{\mu}}(\log n + c) + o(n).$$

Then

$$\|Q_\mu^{*k} - U\| = f(c) + o(1)$$

with  $f(c)$  given in (1.5).

**Proof.** The convolution is a top  $M_k$  to random shuffle where  $M_k$  has the same distribution as the number of occupied boxes when, first,  $X_1$  distinct boxes are picked at random from  $n$  boxes and a ball placed in each, and this is then repeated for  $X_2, \dots, X_k$ , where the  $X_i$  are independent and identically distributed according to  $\mu$ . We will argue that if  $n$  and  $k$  tend to infinity as at (3.6), the number of unoccupied boxes has a limiting Poisson( $e^{-c}$ ) distribution. The result then follows from Proposition 3.1.

It is convenient to represent  $M_k$  as follows. Place balls in the  $n$  boxes ball by ball independently at random. Let  $N(t)$  be the number of occupied boxes after  $t$  balls have been placed; so  $N(t)$  evolves as a Markov chain on states  $\{0, 1, \dots, n\}$  with transition matrix  $P(i, i + 1) = 1 - i/n, P(i, i) = i/n$ , starting from  $N(0) = 0$ . Let  $W_1$  be the number of steps to get  $X_1$  distinct boxes. Thus  $W_1 = \inf\{t : N(t) = X_1\}$ . At time  $W_1$  there are  $X_1$  non-empty boxes, and given  $X_1 = j$ , these boxes are equally likely to form any of the  $\binom{n}{j}$  possible subsets of  $j$  boxes. Let  $W_2$  be the number of further steps to get  $X_2$  distinct boxes (distinct from each other, but perhaps overlapping with the first  $X_1$  boxes already filled). Similarly, define  $W_3, \dots, W_k$ . Clearly

$$M_k \text{ has the same distribution as } N(W_1 + \dots + W_k).$$

The  $W_i$  are independent and identically distributed like  $W$ , where, given  $X = x$ ,  $W$  has the same distribution as

$$T_1 + T_2 + \dots + T_x.$$

Here  $T_i$  are independent geometric( $\frac{n-i+1}{n}$ ) variables and  $X$  has distribution  $\mu$ . Now

$$E(W|X = x) = 1 + \frac{n}{n-1} + \dots + \frac{n}{n-x+1} = x + O(\frac{1}{n}) \text{ as } n \rightarrow \infty$$

$$\text{Var}(W|X = x) = 0 + \frac{n}{(n-1)^2} + \dots + \frac{(x-1)n}{(n-x+1)^2} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Since  $X$  is bounded, as  $n \rightarrow \infty$

$$E(W) = \bar{\mu} + O(\frac{1}{n});$$

$$\text{Var}(W) = E \text{Var}(W|X) + \text{Var}(E(W|X)) \rightarrow \text{Var}(X).$$

Now Chebychev's inequality implies

$$W_1 + \dots + W_k = n \log n + (c + \epsilon_n)n,$$

where  $\epsilon_n$  converges in probability to zero as  $n \rightarrow \infty$ .

Finally, the coupon collector's result gives  $n - N(n \log n + cn) \rightarrow \text{Poisson}(e^{-c})$  in distribution for  $n$  large. For each fixed  $n$ ,  $n - N(t)$  is decreasing in  $t$ . It follows that  $n - M_k$ , which has the same distribution as  $n - N(W_1 + \dots + W_k)$ , converges in distribution to  $\text{Poisson}(e^{-c})$ .  $\square$

**Remark.** The boundedness assumptions on  $X_i$  can be weakened to moment assumptions. Further, it is straightforward to allow different distributions for the  $X_i$ . Error bounds could be derived by proving the Poisson convergence by Stein's method.

#### 4. Spectral analysis and algebras

One classical route for analyzing a Markov chain is through the spectral decomposition of its transition matrix. The main result of this section determines this decomposition for the top to random shuffle. This analysis is used to prove that the formal sums  $\sum_{F(\pi)=i} \pi$  of permutations with first descent at  $i$  span a commutative semisimple algebra.

##### 4.1. Spectral analysis

**Theorem 4.1.** *For the top to random shuffle  $Q$  defined at (1.1), the associated Markov transition matrix  $\mathbf{P}$  is diagonalizable, with spectral decomposition  $\mathbf{P} = \sum_i \lambda_i \mathbf{E}_i$ . The distinct eigenvalues  $\lambda_i$  are  $0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-2}{n}, 1$ . The eigenvalue  $i/n$  has multiplicity equal to the number of permutations in  $S_n$  with exactly  $i$  fixed points, and this equals the rank of the principal idempotent*

$$(4.1) \quad \mathbf{E}_i = \sum_{j=i}^n (-1)^{j-i} \binom{j}{i} \binom{n}{j} \mathbf{P}_j.$$

Here  $\mathbf{P}_j$  is the transition matrix associated with the top  $j$  to random shuffle  $Q_j$  defined at (2.2).

Some discussion is needed before beginning the proof.

(a) On any finite group  $G$ , the map taking the algebra  $L(G)$  of functions  $f$  with convolution to that of group circulant matrices  $\mathbf{M}(\pi, \sigma) = f(\sigma\pi^{-1})$  (with  $f_2 * f_1$  mapped to  $\mathbf{M}_1 \mathbf{M}_2$ ) is a faithful representation of  $L(G)$ , i.e., an isomorphism of algebras. In particular, a probability  $Q$  is mapped to a stochastic matrix  $\mathbf{P}$ , and this  $\mathbf{P}$  is the transition matrix for the random walk on  $G$  taking steps according to  $Q$ .

(b) Now suppose that  $\mathbf{P}$  is any square matrix for which a formula

$$(4.2) \quad \mathbf{P}^k = \sum_i \lambda_i^k \mathbf{E}_i \quad \text{for all } k \geq 0$$

is known, where the  $\lambda_i$  are distinct and no  $\mathbf{E}_i$  is the zero matrix. (Here we use the convention  $0^0 = 1$ .) Then, it follows from basic linear algebra that  $\mathbf{P}$  is diagonalizable,

with spectral decomposition  $\mathbb{P} = \sum_i \lambda_i \mathbb{E}_i$ . (Briefly put, were there a nondiagonal Jordan block in the canonical form for  $\mathbb{P}$ , it would show up in the formula as terms involving  $\binom{k}{j} \lambda_i^k$  with  $j \neq 0$ . We omit the details.) In particular, from (4.2) it follows immediately that the multiplicity of  $\lambda_i$  is the trace of  $\mathbb{E}_i$ . Note that the projectors  $\mathbb{E}_i$  satisfy  $\mathbb{E}_i \mathbb{E}_{i'} = \delta_{i,i'} \mathbb{E}_i$ .

**Proof.** After some rearrangement, one finds from (2.5) and (2.6) that (4.2) holds with  $\lambda_i$  and  $\mathbb{E}_i$  as stated, where  $\mathbb{P}^k(\pi, \sigma) = Q_1^{*k}(\sigma\pi^{-1})$ . The sum is over all  $i = 0, \dots, n$ , but note that  $\mathbb{E}_{n-1} = 0$ ; thus  $\frac{n-1}{n}$  is not an eigenvalue. On the other hand, for  $i \neq n-1$  the multiplicity of  $\lambda_i = \frac{i}{n}$  equals

$$\begin{aligned} \text{tr}(\mathbb{E}_i) &= \sum_{j=i}^n (-1)^{j-i} \binom{j}{i} \binom{n}{j} \text{tr}(\mathbb{P}_j) = \sum_{j=i}^n (-1)^{j-i} \binom{j}{i} \binom{n}{j} (n-j)! \\ &= \frac{n!}{i!} \sum_{u=0}^{n-i} \frac{(-1)^u}{u!} = \binom{n}{i} D_{n-i} = \text{as claimed,} \end{aligned}$$

where  $D_n$  is the number of derangements (permutations with no fixed points) in  $S_n$ . □

**Remarks**

1. Generalizing Theorem 4.1, for the top  $m$  to random shuffle  $Q_m$  of (2.2), the spectral decomposition  $\mathbb{P}_m = \sum_i \lambda_{mi} \mathbb{E}_{mi}$  has eigenvalues  $\lambda_{mi} = \binom{i}{m} / \binom{n}{m}$ ,  $i = m-1, m, \dots, n-2, n$ , with multiplicities

$$\mu_{mi} = \begin{cases} \sum_{j=0}^{m-1} \mu_{1j} & \text{if } i = m-1 \\ \mu_{1i} & \text{otherwise} \end{cases}$$

and  $\mathbb{E}_{mi} = \sum_{j=0}^{m-1} \mathbb{E}_{1j}$  or  $\mathbb{E}_{1i}$  accordingly.

2. The top to random chain is not reversible. It is curious that we do not know how to use the eigenvalues (and their multiplicities) alone to say anything quantitative about the rate of convergence; one seems to require some knowledge about the corresponding idempotents to get rates of convergence.

3. Phatarfod [15] derives all the eigenvalues for the inverse chain *with weights*. That is, each card has an associated weight  $w_i$ , and at each unit of time, a card is removed, with probability proportional to its weight, and put on top. The stationary distribution is not hard to write down, and the eigenvalues are partial sums of the weights. To derive the eigenvalues in the unweighted case, as listed in Theorem 4.1, requires passage to a limit as the distinct weights all tend to a constant value. This is slightly tricky, and the above proof provides both an independent check and explicit determination of the idempotents.

In independent work, both Fill and also Diaconis, Phil Hanlon, and Dan Rockmore [5] have extended Phatarfod's results for weighted schemes. Fill has produced an explicit formula extending (2.5) above for the probability of any order of the deck after any number of shuffles. He has also determined the stationary distribution for a certain weighted  $m$  to top generalization, and Diaconis, Hanlon, and Rockmore have determined eigenvalues for various weighted  $m$  to top variations.

4. The multiplicities in Theorem 4.1 can also be derived from results of Calderbank et al. [3] or Hanlon [12].

4.2. **Algebras.** The main result of this section gives an algebraic interpretation to the previous considerations. The result is expressed in terms of another standard isomorphic image of  $L(G)$  with convolution, namely, the group algebra  $\mathcal{A}(G)$  of formal linear combinations  $\sum_{\pi \in G} f(\pi)\pi$ . Define

$$(4.3) \quad A_i = \sum_{L(\pi)=n-i} \pi, \quad i = 0, \dots, n-1,$$

$$(4.4) \quad B_j = \sum_{i=0}^j A_i = \sum_{L(\pi) \geq n-j} \pi, \quad j = 0, \dots, n-1.$$

**Theorem 4.2.** Let  $B_n := B_{n-1}$ .

(a)  $B_1$  generates an  $n$ -dimensional commutative semisimple subalgebra  $\mathcal{B}$  of  $\mathcal{A}(S_n)$  over the field  $\mathbb{Q}$  of rational numbers.

(b) The primitive idempotents of  $\mathcal{B}$  are  $E_0, E_1, E_2, \dots, E_{n-2}, E_n$ , with

$$(4.5) \quad E_i := \sum_{j=i}^n (-1)^{j-i} \binom{j}{i} \frac{B_j}{j!};$$

moreover,

$$(4.6) \quad B_1^k = \sum_{i \neq n-1} i^k E_i, \quad k \geq 0.$$

(c) The  $B_j$  of (4.4) form a basis of  $\mathcal{B}$ ; moreover,

$$(4.7) \quad B_i B_j = \sum_{k=0}^n b_{ij}^k B_k, \quad i, j = 0, \dots, n-1,$$

where

$$(4.8) \quad b_{ij}^k := \begin{cases} \frac{i!j!}{(k-i)!(k-j)!(i+j-k)!} & \text{if } \max(i, j) \leq k \leq i+j \\ 0 & \text{otherwise.} \end{cases}$$

(d) The  $A_i$  of (4.3) also form a basis of  $\mathcal{B}$ .

**Proof.** Any square matrix  $\mathbf{M}$  generates the commutative matrix algebra of polynomials in  $\mathbf{M}$ . According to Exercise XVII.10 in Lang [14], this algebra is semisimple if and only if  $\mathbf{M}$  is diagonalizable. The element  $\frac{(n-j)!}{n!} B_j$  of  $\mathcal{A}(S_n)$  is just the isomorphic image of  $\mathbb{P}_j$  for  $j = 0, \dots, n-1$ . In particular, it follows from Theorem 4.1 that  $B_1$  generates an  $n$ -dimensional commutative semisimple algebra over  $\mathbb{Q}$  spanned by  $B_0, \dots, B_{n-1}$ . Clearly both (4.3) and (4.4) provide bases for  $\mathcal{B}$ . We have established (a), (c), and (d), with the exception of (4.8). But (4.5)–(4.6) and (4.7)–(4.8) are obtained by rescaling the isomorphic reexpressions of Theorem 4.1 above and Theorem 2.1 (with  $k = 2$ ,  $\mu_1 = \delta_i, \mu_2 = \delta_j$ ), respectively.  $\square$

**Remarks**

1. By considering inverse shuffles, it is clear that  $\mathcal{B}$  is isomorphic to the algebra generated by the sums of permutations with last descent at  $i$ ,  $i = 0, \dots, n - 1$ . Then, by considering the reverse labelling of the cards,  $\mathcal{B}$  is also isomorphic to the algebra generated by the sums with first descent at  $j$ ,  $j = 1, \dots, n$ .
2. For  $m = 0, \dots, n - 1$ ,  $A_m$  of (4.3) corresponds to the shuffle giving equal probability to each of the  $\frac{n-m}{(n-m+1)!} n!$  permutations  $\pi$  with  $L(\pi) = n - m$ . This shuffle can be realized physically by removing the top  $m$  cards from the deck, inserting the bottommost of these cards at random into one of the  $n - m$  possible positions determined by the remaining  $n - m$  cards when the top position is disallowed, and then inserting the other  $m - 1$  cards one at a time completely at random into the deck.
3. Remark 1 following Theorem 4.1 can also be reformulated in terms of the group algebra  $\mathcal{A}(S_n)$ . In particular, each  $B_m$  alone ( $1 \leq m \leq n - 1$ ) generates an  $(n - m + 1)$ -dimensional commutative semisimple subalgebra  $\mathcal{B}_m$  over  $\mathbb{Q}$ ; both  $\{B_0 = \text{id}, B_m, B_{m+1}, \dots, B_{n-1}\}$  and  $\{A_0 = \text{id}, \sum_{i=1}^m A_i, A_{m+1}, \dots, A_{n-1}\}$  provide bases. This gives a chain

$$\mathcal{B} = \mathcal{B}_1 \supseteq \mathcal{B}_2 \supseteq \dots \supseteq \mathcal{B}_{n-1} \supseteq \{\text{multiples of id}\}$$

of  $n - 1$  new subalgebras of  $\mathcal{A}(S_n)$ . The fact that  $B_0, B_m, B_{m+1}, \dots, B_{n-1}$  span a commutative algebra is also clear from (4.7), since  $b_{ij}^k = 0$  unless  $k \geq \max(i, j)$ .

**5. Shuffles and descents**

The previous sections have analyzed a generalization of the top to random shuffle. There is a different generalization which involves cutting off the top  $m$  cards and inserting them randomly among the remaining  $n - m$  cards, keeping both packets in the same relative order. This intermixing can be achieved by using the Gilbert-Shannon-Reeds measure described in Aldous and Diaconis [1] or Bayer and Diaconis [2]. The shuffle amounts to a riffle shuffle in which exactly  $m$  cards are cut off.

This section gives an analysis of these shuffles, and several variants, by means of marking schemes. It gives a clear connection to the large literature on descents.

The first step is to work with inverse shuffles. The inverse shuffle chooses a subset of size  $m$  with uniform probability  $1/\binom{n}{m}$ . The cards in these positions are removed and placed on top, keeping the  $m$  cards in their same relative order. The rate of convergence of repeated shuffles to the uniform distribution doesn't change under inverses; indeed, if  $\tilde{Q}(\pi) = Q(\pi^{-1})$  is inverse to  $Q$ , then  $\tilde{Q}^{*k}(\pi) = Q^{*k}(\pi^{-1})$  for every  $k$ .

The natural generalization can now be stated. Let  $\nu = (\nu_1, \dots, \nu_n)$  be a composition of  $n$ . A  $\nu$ -shuffle begins by choosing  $\nu_1$  cards at random and labeling them with 1's; then  $\nu_2$  of the remaining cards are chosen and labeled with 2's; and so on. Following this labeling, the cards are rearranged by moving all cards labeled 1 to the top, followed by all cards labeled 2, and so on; all cards with the same labels are kept in their initial relative order.

For example, if  $n = 10$  and  $v = (2, 5, 3)$ , the labeling

2	2	1	2	1	2	3	3	3	2
1	2	3	4	5	6	7	8	9	10

results in the permutation

$$3 \ 5 \ 1 \ 2 \ 4 \ 6 \ 10 \ 7 \ 8 \ 9.$$

The inverse shuffles described in the introduction to this section are thus  $(m, n - m)$ -shuffles.

Recall that a permutation  $\pi$  in  $S_n$  has a *descent at  $i$*  if  $\pi(i + 1) < \pi(i)$ . The descent set of  $\pi$  is the subset  $D(\pi) = \{i : 1 \leq i \leq n - 1, \pi(i + 1) < \pi(i)\}$ . Subsets  $D \subset \{1, 2, \dots, n - 1\}$  can be placed into 1 - 1 correspondence with compositions of  $n$  by

$$\{d_1 < d_2 < \dots < d_k\} \mapsto (d_1 - d_0, d_2 - d_1, \dots, d_{k+1} - d_k)$$

where  $d_0 = 0, d_{k+1} = n$ . Let  $D(v)$  be the subset corresponding to the composition  $v$ .

The following result is clear from the definitions.

**Lemma 5.1.** *Let  $v$  be a composition of  $n$ . A  $v$ -shuffle can be described as*

$$(5.1) \quad Q_v(\pi) = \begin{cases} 1/\binom{n}{v_1, \dots, v_n} & \text{if } D(\pi) \subset D(v) \\ 0 & \text{otherwise.} \end{cases}$$

*Thus a  $v$ -shuffle is uniform over all permutations with descent set contained in  $D(v)$ .*

If  $v^1, v^2, \dots, v^k$  are compositions of  $n$ , the shuffle  $Q_{v^k} * \dots * Q_{v^1}$  can be carried out as follows. Prepare an  $n \times k$  array by filling the columns independently, the  $i^{\text{th}}$  column having  $v_j^i$   $j$ -symbols in random positions,  $1 \leq j \leq n$ . Note that here, and in the following,  $v_j^i$  is allowed to be zero for some values of  $j$ . A deck of cards is then labeled by copying the  $\ell^{\text{th}}$  row of the array onto the  $\ell^{\text{th}}$  card. The shuffle is carried out in  $k$  stages. First permute the cards according to the marks in the leftmost place of their labels; this performs a  $v^1$ -shuffle. Then permute the cards according to the second place of their labels, and so on.

As an example, consider  $n = 6$ , and three successive  $(2,4)$ -shuffles. The following shows the original array together with successive rearrangements.

$$(5.2) \quad \begin{array}{ccccccc} & A & & & & & \\ & 1 & 1\ 2\ 2 & & 1\ 2\ 2 & & 1\ 1\ 1 & & 1\ 1\ 1 \\ & 2 & 2\ 2\ 1 & \text{shuffle} & 1\ 1\ 1 & \text{shuffle} & 2\ 1\ 2 & \text{shuffle} & 2\ 2\ 1 \\ & 3 & 2\ 1\ 2 & 1 & 2\ 2\ 1 & 2 & 1\ 2\ 2 & 3 & 2\ 1\ 2 \\ & 4 & 1\ 1\ 1 & \longrightarrow & 2\ 1\ 2 & \longrightarrow & 2\ 2\ 1 & \longrightarrow & 1\ 2\ 2 \\ & 5 & 2\ 2\ 2 & & 2\ 2\ 2 & & 2\ 2\ 2 & & 2\ 2\ 2 \\ & 6 & 2\ 2\ 2 & & 2\ 2\ 2 & & 2\ 2\ 2 & & 2\ 2\ 2 \end{array}$$

**Note.** The final arrangement has rows sorted in lexicographic order from right to left.

**Notation.** Let  $A$  be an  $n \times k$  array containing  $v_j^i$  symbols  $j$  in the  $i^{\text{th}}$  column. Let  $v(A) = (v_1(A), \dots, v_n(A))$  be the composition corresponding to successively repeated rows when the rows of  $A$  are sorted lexicographically from right to left.

For the example (5.2) above,  $v(A) = (1, 1, 1, 1, 2)$ . The main result of this section can now be stated:

**Theorem 5.1.** *Let  $v^1, v^2, \dots, v^k$  be compositions of  $n$ . Then*

$$(5.3) \quad Q_{v^k} * \dots * Q_{v^1} = \sum_v c(v; v^1, \dots, v^k) Q_v$$

where  $c(v; v^1, \dots, v^k)$  is the probability that a random array  $A$  formed by independently placing  $v_j^i$   $j$ -symbols into column  $i$  has  $v(A) = v$ .

**Proof.** Let  $(A_{ij}, 1 \leq i \leq n, 1 \leq j \leq k)$  be the random array corresponding to a shuffle drawn from  $Q_{v^k} * \dots * Q_{v^1}$ . By construction, the distribution of  $(A_{\pi(i)j})$  is the same as the distribution of  $(A_{ij})$ , for any  $\pi \in S_n$ . Thus, given  $v(A)$ , any subset of  $v_1(A)$  cards are equally likely to be on top. Any subset of  $v_2(A)$  of the remaining cards are equally likely to be next, and so on. This is just the definition of a  $v(A)$  shuffle.  $\square$

**Remarks**

1. Let  $A^k$  denote the random array following  $Q_{v^1}, \dots, Q_{v^k}$ . Then  $v(A) \equiv (v(A^k), k \geq 0)$  is a Markov chain:  $v(A^k) = (n)$  when  $k = 0$ , and

$$c(v; v^1, \dots, v^k) = \sum_{v'} c(v'; v^1, \dots, v^{k-1}) c(v; v', v^k)$$

for  $k \geq 1$ . In particular, for the case  $v^1 = \dots = v^k$  of repeated shuffles,  $v(A)$  is a time homogeneous Markov chain. This chain is dual to the random walk  $X$  on  $S_n$  generated by  $Q_{v^1}$  in the sense of Diaconis and Fill [4]. Indeed, regard a composition  $v$  of  $n$  as the subset  $v^* := \{\pi \in S_n : D(\pi) \subset D(v)\}$  of  $S_n$ . Then  $(v(A^k))^*$  is a set-valued strong stationary dual of  $X$ , with state space  $\mathcal{S}^*$  consisting of all  $v^*$ , initial distribution equal to point mass at  $(n)^* = \{\text{id}\}$ , and transition matrix  $P^*((v')^*, v^*) = c(v; v', v^1)$ .

2. By the preceding remark, to calculate  $c(v; v^1, \dots, v^k)$  in (5.3) it is (in principle) enough to handle the case  $k = 2$ . A combinatorial interpretation of  $c(v; v^1, v^2)$  is provided by Corollary 5.1 below: in the notation there,

$$c(v^3; v^1, v^2) = \frac{\binom{n}{v_1^3, \dots, v_n^3}}{\binom{n}{v_1^1, \dots, v_n^1} \binom{n}{v_1^2, \dots, v_n^2}} C_{12}^3.$$

We draw two consequences from Theorem 5.1, and more are given in the next section. The first gives a proof of a theorem of Louis Solomon [16], relating to what is now called the descent algebra. Let  $D \subset \{1, 2, \dots, n-1\}$ . Let

$$A_D := \sum_{D(\pi) \subset D} \pi.$$

This is a formal linear combination in the group algebra of  $S_n$ .

**Corollary 5.1.** (Garsia and Remmel [8])

$$A_{D_2} A_{D_1} = \sum_{D_3} C_{12}^3 A_{D_3},$$

where  $C_{12}^3$  is the number of arrays with nonnegative integer entries having row sums  $v(D_1)$ , column sums  $v(D_2)$ , and content  $v(D_3)$ .

**Proof.** Let  $D_1$  and  $D_2$  correspond to compositions  $v^1$  and  $v^2$ . Multiplying the words  $A_{D_2}$  and  $A_{D_1}$  corresponds (modulo scaling) to doing the shuffle  $Q_{v^2} * Q_{v^1}$ . Say  $v^1 = (v_1^1, \dots, v_r^1)$ ,  $v^2 = (v_1^2, \dots, v_s^2)$ . In the shuffling interpretation, an  $n \times 2$  array  $A$  is formed. Let this correspond to an  $r \times s$  matrix  $M$  by setting  $M_{ij}$  to be the number of rows in  $A$  with first entry  $i$  and second entry  $j$ . Clearly  $M$  has row sums  $v_i^1$  and column sums  $v_j^2$ . The composition  $v(A)$  can be read off  $M$  by taking the entries in the first column (top to bottom) followed by the entries in the second column top to bottom, and so on. This composition is what is meant by the content of  $M$  (zero parts can be omitted).  $\square$

**Remarks**

1. Corollary 5.1 shows that the  $A_D$  form an algebra over the ring  $\mathbb{Z}$ . This is Louis Solomon's [16] famous descent algebra, which has been the subject of intense study in recent years. Garsia [7], Garsia and Reutenauer [9] or Gessel and Reutenauer [11] are central references with pointers to the rest of this literature.

2. As discussed in the next section, the development of the previous sections corresponds to the special case of shuffles of the form  $v = (1, 1, \dots, 1, n - m)$ .

The second corollary gives a probabilistic analysis of these shuffles. It uses the notion of strong stationary time introduced by Aldous and Diaconis [1] and developed by Diaconis and Fill [4]. Recall the definition of separation:  $\text{sep}(Q, U) = \max_{\pi} (1 - \frac{Q(\pi)}{U(\pi)})$ .

**Corollary 5.2.** Let  $v^1, v^2, \dots, v^k$  be compositions of  $n$  with  $v^i = (v_1^i, \dots, v_{r_i}^i)$ . Then

$$(5.4) \quad \|Q_{v^k} * \dots * Q_{v^1} - U\| \leq \text{sep}(Q_{v^k} * \dots * Q_{v^1}, U) \leq \binom{n}{2} \prod_{i=1}^k \sum_{j=1}^{r_i} \frac{v_j^i (v_j^i - 1)}{n(n-1)}.$$

**Proof.** Let  $A^k$  denote the random matrix after  $Q^k := Q_{v^k} * \dots * Q_{v^1}$  has been performed. Let  $T$  be the first time  $k$  that the rows of  $A_k$  are distinct. Then given  $T = k$  (for any fixed  $k$ ), the order of the deck has distribution  $Q_{(1, \dots, 1)} = U$ ; thus (by definition)  $T$  is a strong stationary time. As shown by Aldous and Diaconis [1], one always has

$$(5.5) \quad \|Q^k - U\| \leq \text{sep}(Q^k, U) \leq P(T > k).$$

Fix  $k$  and let  $X_{ij}$ ,  $1 \leq i < j \leq n$ , be indicator random variables taking values 1 or 0 as the  $i$ th and  $j$ th rows of  $A^k$  agree or not. Clearly the chance that any given  $X_{ij}$  is 1 is given by the product in (5.4), and  $P\{T > k\} \leq \Sigma P\{X_{ij} = 1\}$ .  $\square$

**Remark.** The stopping time  $T$  in the preceding proof is actually a time to stationarity,

i.e., a fastest strong stationary time. This is because the reversal permutation cannot be achieved prior to time  $T$ . (In the terminology of Diaconis and Fill [4], the dual chain  $(v(A^k))$  is sharp.) Thus, the second inequality in (5.5) is actually equality, and the only source of inequality in the separation bound of (5.4) is the subadditive estimate of the tails of  $T$ .

**Examples.** Suppose  $v^i = (m, n - m)$  for all  $i$ . The bound becomes

$$\|Q_{m,n-m}^{*k} - U\| \leq \binom{n}{2} \left(1 - \frac{2m(n-m)}{n(n-1)}\right)^k.$$

If  $m$  is fixed, this shows that  $k = \frac{n}{m}(\log n + c)$  steps are enough. Note that this is the same requirement as the apparently faster top  $m$  to random shuffles analyzed in previous sections.

Suppose next that  $n$  is even and  $m = n/2$ . This corresponds to repeated riffle shuffles with cuts exactly in half. Now  $2(\log_2 n + c)$  shuffles are enough. The disparity with the Bayer-Diaconis result of  $\frac{3}{2}(\log_2 n + c)$  comes because of the first inequality in (5.4). Indeed, the results of Bayer and Diaconis imply that it takes  $2(\log_2 n + c)$  riffle shuffles to make separation small.

**Remark.** Carl Dou has provided sharper bounds using Stein's method. For example, for  $m$  fixed, the limiting (as  $n \rightarrow \infty$ ) separation after  $k = \frac{n}{m}(\log n + c)$  steps is  $1 - e^{-\frac{1}{2}e^{-2c}}$  for  $(m, n - m)$  shuffles, which is, for any  $c \in \mathbb{R}$ , strictly larger than the corresponding expression  $1 - e^{-e^{-c}}(1 + e^{-c})$  for  $(1^m, n - m)$  shuffles (in the notation described at the beginning of the next section).

### 6. Some Special Cases

Theorem 5.1 reduces the analysis of repeated shuffling on  $S_n$  (dimension  $n!$ ) to that of the dual Markov chain on compositions of  $n$  (dimension  $2^{n-1}$ ), for a fairly broad class of shuffles. In this section we use the framework it provides to consider several shuffles in detail. Often the dimension is made effectively quite small, as the first example shows.

**Notation.**  $v = (v_1^{m_1}, \dots, v_\ell^{m_\ell})$  is shorthand for  $v = (v_1, \dots, v_1, \dots, v_\ell, \dots, v_\ell)$  with  $m_i$   $v_i$ 's,  $1 \leq i \leq \ell$ .

**Example 1. Top  $m$  to random,** as in Sections 1-4. Here, shuffles  $v$  of the form  $v = (1^m, n - m)$  are convolved. Accordingly, let  $v^1 = (1^{m_1}, n - m_1)$  and  $v^2 = (1^{m_2}, n - m_2)$ , and consider  $Q_{v^2} * Q_{v^1}$ . It is an instructive exercise to use Corollary 5.1 and Remark 2 following Theorem 5.1 to compute

$$c(v; v^1, v^2) = \begin{cases} (\delta_{m_2} \# \delta_{m_1})(m) & \text{if } v \text{ is of the form } (1^m, n - m) \\ 0 & \text{otherwise} \end{cases}$$

in the notation (2.4). The results of Sections 1-4 (e.g., Theorem 2.1) now follow by induction and mixing.

**Example 2. Top  $m$  to random**, as in Section 5. Repeated  $(m, n - m)$  shuffles were analyzed in Section 5. It is worth noting that the dual composition-chain always takes values of the form  $\nu = (\nu_1, \dots, \nu_{r-1}, \nu_r)$ , where each  $\nu_i$ , with the possible exception of  $\nu_r$ , belongs to  $\{1, \dots, m\}$ . We do not know an exact expression for the distribution of this chain at time  $n$ , even when  $m = 2$ , and we do not know whether the algebra generated is semisimple. From (5.3), to determine the eigenvalues and their algebraic multiplicities for the shuffle-chain, it would be enough to know the joint distribution of  $\omega^k = (\omega_1^k, \omega_2^k, \dots, \omega_m^k)$ , where  $\omega_j^k$  records the number of parts that equal  $j$  for the composition at time  $k$ .

From the work of Diaconis, Hanlon, and Rockmore, we know that the  $n!$  eigenvalues for repeated  $(m, n - m)$  shuffles are all nonnegative and are given as follows. Briefly, all eigenvalues are of the form  $i / \binom{n}{m}$ , and such an eigenvalue has multiplicity equal to the number of permutations in  $S_n$  that fix exactly  $i$   $m$ -sets. This can be made more explicit, as follows. For each  $n$ -tuple  $c = (c_1, c_2, \dots, c_n)$  of nonnegative integers with  $\sum_i ic_i = n$ , there is an eigenvalue

$$(6.1) \quad \lambda_c = \frac{1}{\binom{n}{m}} \sum_j \prod_{i=1}^n \binom{c_i}{j_i},$$

where the sum is over  $n$ -tuples  $j = (j_1, j_2, \dots, j_n)$  of nonnegative integers with  $\sum_i ij_i = m$ , with algebraic multiplicity

$$(6.2) \quad \mu_c = \frac{n!}{\prod_{i=1}^n (c_i! i^{c_i}).}$$

The number  $\mu_c$  is just the number of permutations  $\pi$  in  $S_n$  having cycle index  $c$  (i.e.,  $c_1$  fixed points,  $c_2$  transpositions,  $c_3$  3-cycles, etc.), and the sum in (6.1) is the number of sets of size  $m$  that are fixed (setwise) by any such  $\pi$ . Note that the  $\lambda_c$  of (6.1) are not all distinct. The largest eigenvalue is of course  $1 = \lambda_{(n,0,\dots,0)}$ , and the second largest is  $\lambda_{(n-2,1,0,\dots,0)} = \left[ \binom{n-2}{m} + \binom{n-2}{m-2} \right] / \binom{n}{m}$ , with (total) algebraic multiplicity  $\binom{n}{2}$ . Recall that for repeated  $(1^m, n - m)$  shuffles, the second largest eigenvalue is the somewhat smaller  $\binom{n-2}{m} / \binom{n}{m}$ , with multiplicity  $\binom{n}{2}$ .

**Example 3. Top  $m$  and bottom  $p$  to random shuffles.** The top  $m$  and bottom  $p$  cards are removed from the deck and inserted one at a time at random. The (inverse) shuffle then corresponds to the composition  $\nu = (1^m, n - m - p, 1^p)$ . Specializing (5.1),

$$(6.3) \quad Q_\nu(\pi) = \begin{cases} (n - m - p)! / n! & \text{if } D(\pi) \subset \{1, \dots, m\} \cup \{n - p, \dots, n - 1\} \\ 0 & \text{otherwise.} \end{cases}$$

Unlike Example 1, there is no low-dimensional sufficient statistic.

*These shuffles generate a noncommutative algebra.* To see this, use the results of Section 5, or argue directly as in Section 2, that if  $\nu^i = (1^{m_i}, n - m_i - p_i, 1^{p_i})$  for  $i = 1, 2$ , then  $c(\nu; \nu^1, \nu^2) = 0$  unless  $\nu$  is of the form  $(1^m, n - m - p, 1^p)$ , in which case

$$c(\nu; \nu^1, \nu^2) = \frac{\binom{n-m_2-p_2}{m-m_2, p-p_2, n-m-p}}{\binom{n}{m_1, n-m_1-p_1, p_1}} \frac{\binom{m_2+p_2}{m_1+m_2-m, p_1+p_2-p, (m-m_1)+(p-p_1)}}{\binom{n}{m_1, n-m_1-p_1, p_1}}.$$

So the shuffles generate an algebra. But if, for example,  $v^1 = (1, n - 1)$  and  $v^2 = (n - 1, 1)$ , then the  $n$ -cycle  $(1\ 2\ \cdots\ n)$  is assigned mass  $2/n^2$  by  $Q_{v^2} * Q_{v^1}$ , but mass  $1/n^2$  by  $Q_{v^1} * Q_{v^2}$ . So the algebra is noncommutative. Various subalgebras are considered in Examples 4–6.

**Example 4. Repeated top or bottom to random.** The top or bottom card is chosen at random and inserted at random into the deck. This is a special case of Example 3 where each  $v^i$  is randomly (and independently)  $(1, n - 1)$  or  $(n - 1, 1)$ . Here the dual chain is 2-dimensional (indexed by  $m$  and  $p$ ), and, as in Example 2, we do not have an exact expression for dual distributions. However, one sees from (6.3) that, for computing either separation (using  $\pi = \text{rev}$ ) or eigenvalues with their algebraic multiplicities (using  $\pi = \text{id}$ ), one need only know the distributions for the univariate evolution of  $\ell := m + p$ . But these distributions are clearly the same as for the number of parts equal to 1 for repeated  $(1, n - 1)$ -shuffles. Thus separation and eigenvalues are the same; the results are contained in Remark 4 of Section 3 and Theorem 4.1, respectively.

**Example 5. Repeated top and bottom to random.** The top and bottom card are both removed from the deck and inserted one at a time at random, and this basic shuffle is repeated. Here one sees that separation and eigenvalues are the same as for the repeated top 2 to random shuffles studied in Sections 1–4 (for which  $v^i = (1, 1, n - 2)$ ).

**Example 6. Trinomial top and bottom to random.** We present an analogue of Lemma 2.1. Suppose we let  $v^i = (1^{m_i}, n - m_i - p_i, 1^{p_i})$  and put a trinomial  $(n; a_i, b_i, c_i)$  distribution on  $(m_i, n - m_i - p_i, p_i)$ ,  $i = 1, 2$ . If we perform a  $v^1$ -shuffle and then a  $v^2$ -shuffle, we get a  $v$ -shuffle, where  $v = (1^m, n - m - p, 1^p)$  and  $(m, n - m - p, p)$  has the

$$\text{trinomial}(n; a_1 b_2 + a_2, b_1 b_2, c_1 b_2 + c_2)$$

distribution. (Thus the shuffles commute if and only if  $a_1 c_2 = a_2 c_1$ .) By induction, after  $k$  such shuffles we obtain the distribution

$$\text{trinomial}(n; \sum_{i=1}^k a_i \prod_{j=i+1}^k b_j, \prod_{j=1}^k b_j, \sum_{i=1}^k c_i \prod_{j=i+1}^k b_j).$$

Again it is easy to relate separation to earlier results.

As a special case, suppose  $a_i = c_i = \frac{1}{2}(1 - b_i)$  for all  $i$ . Then after  $k$  shuffles we get the distribution

$$\text{trinomial}(n; \frac{1}{2}(1 - \prod_{j=1}^k b_j), \prod_{j=1}^k b_j, \frac{1}{2}(1 - \prod_{j=1}^k b_j)).$$

In particular, such trinomial  $(n; \frac{1}{2}(1 - b), b, \frac{1}{2}(1 - b))$  shuffles generate a commutative algebra.

**Example 7. Biased cut riffle shuffles.** We consider repeated riffle shuffling: A binomial  $(n, p)$  number of cards is cut off the top, and the resulting packets are intermixed at random. The case  $p = 1/2$  is treated in depth by Bayer and Diaconis [2].

More generally, we consider the family of  $v$ -shuffles, where  $v = (v_1, \dots, v_r)$  is given the multinomial  $(n; p_1, \dots, p_r)$  distribution. Both  $p = (p_1, \dots, p_r)$  and  $r$  are allowed to vary

from shuffle to shuffle. For a single multinomial( $n; p$ ) (inverse) shuffle  $Q$ , we have

$$(6.4) \quad Q(\pi) = \sum_{\nu: D(\pi) \subset D(\nu)}^* \prod_{j=1}^r p_j^{\nu_j},$$

where the sum is restricted to compositions  $\nu$  having exactly  $r$  nonnegative parts. (In particular, if  $\pi$  has  $r$  or more descents, then  $Q(\pi) = 0$ .) In the case of equal weights  $p_j \equiv 1/r$ , one recaptures the Bayer–Diaconis result

$$Q(\pi) = \binom{n+r-|D(\pi)|-1}{n} / r^n.$$

We do not know how to simplify (6.4) for any other examples.

If  $\nu^i = (\nu_1^i, \dots, \nu_{r_1}^i) \sim \text{multinomial}(n; p_1^i, \dots, p_{r_1}^i)$  for  $i = 1, 2$ , then  $Q_{\nu^2} * Q_{\nu^1} = Q_\nu$ , where  $\nu \sim \text{multinomial}(n; p)$  with  $p$  the tensor product  $p = (p_1^2 p_1^1, p_2^2 p_2^1, \dots, p_{r_1}^2 p_{r_1}^1; \dots; p_{r_2}^2 p_{r_2}^1, \dots, p_{r_2}^2 p_{r_2}^1)$  of the category probabilities. Thus *the family of biased multinomial cut riffle shuffles forms a noncommutative algebra*.

As a special case, consider now the case of repeated binomial( $n, p$ ) shuffles, as in the introduction to this example. Fix  $p = 1 - q \in (0, 1)$  and let  $n \rightarrow \infty$ . Then it can be shown that the separation after  $k = \lfloor (2 \log n - \log 2 + c) / |\log(1 - 2pq)| \rfloor$  steps tends to  $1 - e^{-e^{-c}}$  for any fixed  $c \in \mathbb{R}$ . Since  $1/|\log(1 - 2pq)|$  increases with  $|p - 1/2|$ , the unbiased case  $p = 1/2$  provides the fastest shuffle (in this sense at least).

### Acknowledgements

Carl Dou, Phil Hanlon, and Dan Rockmore all provided help in developing this paper.

### References

- [1] Aldous, D. and Diaconis, P. (1986) Shuffling cards and stopping times. *Amer. Math. Monthly* **93**, 333–348.
- [2] Bayer, D. and Diaconis, P. (1991) Trailing the dovetail shuffle to its lair. *Ann. Appl. Prob.* **2**, 294–313.
- [3] Calderbank, A. R., Hanlon, P., Sundaram, S. and Wales, D. (1990) *Representations of the symmetric group in deformations of the free Lie algebra*. Technical report, Dept. of Math., University of Michigan.
- [4] Diaconis, P. and Fill, J. (1990) Strong stationary times via a new form of duality. *Ann. Prob.* **18**, 1483–1522.
- [5] Diaconis, P., Hanlon, P. and Rockmore, D. (1991) *The eigenvalues of Markov chains on permutations*. Unpublished manuscript.
- [6] Feller, W. (1968) *An Introduction to Probability Theory and its Applications*. Vol. 1, 3rd ed., Wiley, New York.
- [7] Garsia, A. (1990) Combinatorics of the free Lie algebra and the symmetric group. In P.R. Rabinowitz, E. Zehnder (eds.) *Analysis, etc.* Academic Press, Boston, 309–382.
- [8] Garsia, A. and Remmel, J. (1985) Shuffles of permutations and the Kronecker product. *Graphs Combin.* **1**, 217–263.
- [9] Garsia, A. and Reutenauer, C. (1989) A decomposition of Solomon’s descent algebra. *Adv. Math.* **77**, 189–262.

- [10] Gerstenhaber, M. and Schack, S. (1987) A Hodge-type decomposition for commutative algebra cohomology. *J. Pure Appl. Algebra* **48**, 229–247.
- [11] Gessel, I. and Reutenauer, C. (1991) *Counting permutations with given cycle structure and descent set*. Technical Report, Dept. of Mathematics, Brandeis University.
- [12] Hanlon, P. (1990) The action of  $S_n$  on the components of the Hodge decomposition of Hochschild homology. *Michigan Math. J.* **37**, 105–124.
- [13] Holst, L. (1980) On matrix occupancy, committee, and capture-recapture problems. *Scand. J. Statist.* **7**, 139–146.
- [14] Lang, S. (1984) *Algebra*. Addison-Wesley, Menlo Park, California.
- [15] Phatarfod, R. M. (1991) On the matrix occurring in a linear search problem. *Jour. Appl. Prob.* **28**, 336–346.
- [16] Solomon, L. (1976) A Mackey formula in the group ring of a Coxeter group. *J. Algebra* **41**, 255–264.

