BIASED RIFFLE SHUFFLES, QUASISYMMETRIC FUNCTIONS, AND THE RSK ALGORITHM

YUFEI ZHAO

Abstract. This paper highlights the connections between riffle shuffles and symmetric functions. We analyze the effect of a riffle shuffle followed by the RSK algorithm on the resulting permutation. The theory of shuffling also allows us to deduce certain quasisymmetric function identities. Finally, we present a result about the eigenvalues of the Markov chain induced by riffle shuffles.

1. Introduction

Take a deck of cards, cut into two halves, and riffle them together. This is the *riffle shuffle*, one of the most common ways of shuffling cards.

Riffle shuffles have been extensively studied by many mathematicians. A realistic model of riffle shuffles, known as the Gilbert–Shannon–Reeds model, was given independently by Gilbert [12] and Reeds [13]. Later research studied the number of shuffles needed to evenly mix a deck of cards, as well permutation statistics from riffle shuffles. For instance, Bayer and Diaconis [4] showed that, roughly speaking, seven shuffles are needed to evenly mix a standard deck of 52 cards. This result of Bayer and Diaconis caused quite a bit of excitement in the mathematical community and beyond. According the MathSciNet review of [4], "rarely does a new mathematical result make both the New York Times and the front page of my local paper, and even more rarely is your reviewer asked to speak on commercial radio about a result, but such activity was caused by the preprint of this paper."

In this paper, we consider a different approach to studying riffle shuffles, by relating them to the theory of symmetric functions. This approach is credited to Stanley [16], who observed that a certain generalization of riffle shuffles, known as biased riffle shuffles, have probability distributions that can be easily described using quasisymmetric functions. Fulman [11] takes a similar approach and applies symmetric functions to analyze biased riffle shuffles, focusing primarily on the cycle structure and the increasing subsequence structure of the resulting permutations.

The purpose of this paper is to highlight the connections between riffle shuffles and symmetric functions. In Section 2, we describe biased riffle shuffles, which induce a probability distribution, known as the QS-distribution, on S_n. In Section 3, we show that the probability of obtaining $w \in S_n$ is equal to $L_{\mathrm{co}(w^{-1})}(x)$, where L denotes the fundamental quasisymmetric function. This result is key to the bridge between riffle shuffles and symmetric functions.

Date: April 2009.

In Section 4, we consider the effect of applying the RSK algorithm to $w \in S_n$ under the QS-distribution. This gives a shuffling interpretation of Schur functions $s_\lambda(x)$. Furthermore, we study the question of how close is a biased riffle shuffle to a truly random permutation if we are only allowed to observe the RSK shape of the permutation. We employ a technique by Diaconis and Fulman [8] to obtain a bound on the total variation distance of the two distributions on $\mathrm{Par}(n)$, the set of partitions of n. The proof makes use of the famous Cauchy identity.

In Section 5, we use the theory of biased riffle shuffles to deduce some identities involving quasisymmetric functions. In particular, we give a proof of the identity

$$\sum_{\lambda \vdash n} s_\lambda(x) s_\lambda(y) = \sum_{u \in S_n} L_{\mathrm{co}(u)}(x) L_{\mathrm{co}(u^{-1})}(y)$$

using biased riffle shuffles. Our technique can be generalized to more than two sets of variables. This approach of deducing quasisymmetric function identities may be original, as least to the author's knowledge.

In Section 6, we analyze the eigenvalues of the Markov chain on S_n induced by biased riffle shuffles. We explain how our setup is a special case of a more general setup studied by Bidigare, Hanlon, and Rockmore [5]. This allows us to the deduce that the eigenvalues are precisely the power sum symmetric functions $p_\lambda(x)$ with multiplicity equal to the number of elements in S_n with cycle type λ.

2. Riffle Shuffles and Generalization

In this section, we describe the Gilbert-Shannon-Reeds (GSR) model of shuffling cards, as well as a weighted generalization known as biased riffle shuffles.

2.1. **Description of riffle shuffles.**

The original GSR model can be described as follows [4]. Given a deck of n cards, first cut it into two piles, so that the first pile contains c cards with probability $\binom{n}{c}2^{-n}$. The two piles are then riffled together in by successively dropping cards from either pile with probability proportional to the number of cards in that pile. That is, if there are A cards remaining in the left pile and B cards remaining in the right pile, then the next card is dropped from the left pile with probability $A/(A+B)$. It is easy to show that this equivalent to saying that all $\binom{n}{c}$ ways of interleaving the two piles are equally likely.

The above riffle shuffle is known as a *2-shuffle*, because the original deck is cut into two piles. It can be generalized to *a-shuffles*, where the deck of n cards is cut into a piles, such that the sizes of the piles are $c_1, c_2, \ldots, c_a$ with probability $\binom{n}{c_1,\ldots,c_a}a^{-n}$. In the riffles, as before, we successively drop cards from one of the piles with probability proportional to the number of cards in that pile. Equivalently, all $\binom{c_1+c_2+\cdots+c_a}{c_1,c_2,\ldots,c_a}$ ways of interleaving the a piles are equally likely.

One further generalization, known as *biased riffle shuffles*, was introduced by Diaconis, Fill, and Pitman [7, p. 153]. Biased riffles differ from b-shuffle only in the cutting step, where, instead, we cut the deck of n cards into a piles, such that the sizes of the piles are

$c_1, \ldots, c_a$ with probability

$$\binom{n}{c_1, c_2, \ldots, c_a} x_1^{c_1} x_2^{c_2} \cdots x_a^{c_a}$$

where $x = (x_1, x_2, \ldots, x_a)$ is a vector of nonnegative real numbers with sum 1. We refer to this shuffle as an x-*biased shuffle*. Note that an a-shuffle is equivalent to a $(\frac{1}{a}, \frac{1}{a}, \ldots, \frac{1}{a})$-biased shuffle.

We will always read a deck of cards from top to bottom, and assume that the cards are faced down.

Example 1. Suppose that we would perform a (x, y, z)-biased riffle shuffle on a deck of three cards initially sorted as JQK. Table 1 shows the possible outcomes of the cuts, with the corresponding probabilities. On the last column is the set of possible arrangements of the three cards resulting from riffling the three piles together; conditioned on the cut, each arrangement is equally likely to occur. Table 2 shows the resulting probability distribution on the resulting deck. Note the presence of symmetric functions!

TABLE 1. Outcomes after cutting the deck.

Pile 1	Pile 2	Pile 3	Probability of cut	Riffle results
JQK	$\emptyset$	$\emptyset$	x^3	JQK
$\emptyset$	JQK	$\emptyset$	y^3	JQK
$\emptyset$	$\emptyset$	JQK	z^3	JQK
JQ	K	$\emptyset$	$3x^2y$	JQK, JKQ, KJQ
JQ	$\emptyset$	K	$3x^2z$	JQK, JKQ, KJQ
$\emptyset$	JQ	K	$3y^2z$	JQK, JKQ, KJQ
J	QK	$\emptyset$	$3xy^2$	JQK, QJK, QKJ
J	$\emptyset$	QK	$3xz^2$	JQK, QJK, QKJ
$\emptyset$	J	QK	$3yz^2$	JQK, QJK, QKJ
J	Q	K	$6xyz$	$JQK, JKQ, QJK, QKJ, KJQ, KQJ$

TABLE 2. Probability distribution of the shuffled deck.

Shuffled deck	Probability
JQK	$x^3 + y^3 + z^3 + x^2y + x^2z + y^2z + xy^2 + xz^2 + yz^2 + xyz$
JKQ	$x^2y + x^2z + y^2z + xyz$
KJQ	$x^2y + x^2z + y^2z + xyz$
QJK	$xy^2 + xz^2 + yz^2 + xyz$
QKJ	$xy^2 + xz^2 + yz^2 + xyz$
KQJ	xyz

It will be useful to allow the deck to be cut into piles which are indexed by some totally order set I, possibly of infinite size. Of course only finitely many such piles are nonempty for any given cut. Let $(x_i)_{i\in I}$ be a probability distribution on I. Then, the cut produces piles with sizes $c : I \to \mathbb{N}$ with probability

$$\Pr[\text{cut } c] = n! \prod_{i\in I} \frac{x_i^{c(i)}}{c(i)!}.$$

Note that all but finitely many factors are equal to 1.

The inverse of a biased shuffle has the following description [10]: start with an ordered deck of n cards. Starting from the bottom of the deck, each card is dropped into a pile (indexed by I) such that it is dropped to pile $i \in I$ with probability x_i, independently. After all the cards have been dropped, the piles are assembled from top to bottom in order of I.

2.2. **QS-distribution.** Given fixed $(x_i)_{i\in I}$ and $n \in \mathbb{P}$, the biased riffle shuffles induce a probability distribution on all possible orders on a deck of n cards. Equivalently, we get a probability distribution on the symmetric group S_n. Different authors may have different ways of corresponding shuffles to elements of S_n. We use the following interpretation. Start with a sorted deck of cards with face values $1, 2, \ldots, n$, and perform a biased riffle shuffle. Suppose that the shuffled deck reads $w(1), w(2), \ldots, w(n)$, for some $w \in S_n$, then we associate w to this shuffle.

We define the *QS-distribution* [16] as the probability distribution on the symmetric group induced by biased riffle shuffles, using the above correspondence. If we need to be explicit about the parameters $x = (x_i)_{i\in I}$, then we shall refer to the *QS_x-distribution.*

We give an alternate, perhaps more useful, description of the QS-distribution. For each $1 \le j \le n$, from the probability distribution $(x_i)_{i\in I}$ choose an independent random i_j. Then *standardize* the sequence $\boldsymbol{i} = i_1 i_2 \cdots i_n$ to obtain $w = w_1 w_2 \cdots w_n \in S_n$. In other words, sort the pairs $\{(i_j, j)\}_{j\in[n]}$ in lexicographic order, and let w_j be the position of (i_j, j). For instance, if $I = \mathbb{P}$, and $\boldsymbol{i} = 311431$, then $w = 412653$. This defines a probability distribution on S_n, which coincides with the QS-distribution. It is instructive to compare this description to the description of the inverse riffle shuffle.

The name for the QS-distribution is due to Stanley [16], and it comes from its connections with quasisymmetric functions, as we shall explain in Section 3.

Remark. The representation of card shuffles as permutations can be a rather confusing issue, and there are conflicting interpretations in literature. In our interpretation, elements of S_n are viewed as right actions on a deck of cards. Some authors, notably Diaconis, view permutations as left actions on the deck, where $\pi \in S_n$ is associated to the shuffle that sends the card in old position i to new position $\pi(i)$ (e.g., [1, 9, 7]). However, recent papers by Diaconis have switched to the right-action interpretation (e.g., [2, 3, 8]). The two different interpretations are in some sense inverses of each other. One should always check with the author's definition before interpreting any result on permutations induced shuffles. See the remark after Theorem 6 for an instance of such misinterpretation.

2.3. **Composition of shuffles.** Now we consider the effect of performing two biased riffle shuffles in succession. If $x = (x_i)_{i\in I}$ and $y = (y_j)_{j\in J}$ are two probability distributions, then we define the following probability distribution on $I \times J$

$$x \otimes y = (x_i y_j)_{(i,j)\in I\times J}$$

where we place the lexicographic order on $I \times J$, i.e., $(i,j) < (i',j')$ if $i < i'$, or $i = i'$ and $j < j'$.

Theorem 2. [16, Thm. 2.4] [10, Prop. 1] *An x-biased shuffle followed by a y-biased shuffle is equivalent to an $(x \otimes y)$-biased shuffle. Equivalently, we have the convolution of measures*

$$QS_x * QS_y = QS_{x \otimes y}.$$

Proof. We use the description of inverse biased shuffles. Consider the composition of an inverse y-biased shuffle followed by an inverse x-biased shuffle. Suppose that we start (or rather, end) with the deck with cards $1, 2, \ldots, n$. First, each card k is independently assigned to a random element $j_k \in J$ drawn from the distribution $(y_j)_{j \in J}$, and assigned the pair (j_k, k). The deck is then sorted according to the lexicographic order of the pairs. Then, each card k is independently assigned another random element $i_k \in I$ drawn from the distribution $(x_i)_{i \in I}$, and the pair is updated to $(i_k, (j_k, k))$. The deck is then resorted lexicographically using the new pairs. This gives the deck after the inverse shuffle. Note that this process is equivalent to assigning card k a random element $(i_k, j_k) \in I \times J$ drawn from the distribution $x \otimes y$, and then sort the cards lexicographically using $((i_k, j_k), k)$. The latter is also the description of the inverse $(x \otimes y)$-shuffle. Therefore, a x-biased shuffle followed by a y-biased shuffle is equivalent to an $(x \otimes y)$-biased shuffle.

The second statement follows from the observation that elements of S_n act on the deck from the right, so that $QS_x * QS_y$ is the distribution of an x-biased shuffle followed by a y-biased shuffle. □

The following unbiased version of Theorem 2 is due to Bayer and Diaconis [4].

Corollary 3. *An a-shuffle followed by a b-shuffle is equivalent to an ab-shuffle.*

Proof. Apply Theorem 2 to $x = (\frac{1}{a}, \ldots, \frac{1}{a})$ and $y = (\frac{1}{b}, \ldots, \frac{1}{b})$. □

3. Quasisymmetric Functions

In this section we connect riffle shuffles to quasisymmetric functions. We use [15, Ch. 7] as our reference for symmetric and quasisymmetric functions.

3.1. **Notation.** Recall that a *quasisymmetric function* is a formal power series $F(z)$ of bounded degree in variables $z = (z_i)_{i \in I}$, with the following property: if $i_1 < i_2 < \cdots < i_n$ and $j_1 < j_2 < \cdots < j_n$, where $i_k, j_k \in I$, and $a_1, \ldots, a_k \in \mathbb{P}$, then the coefficient of $z_{i_1}^{a_1} \cdots z_{i_n}^{a_n}$ is equal to the coefficient of $z_{j_1}^{a_1} \cdots z_{j_n}^{a_n}$ in $F(z)$. Let $\mathcal{Q}^n$ denote the $\mathbb{Q}$-vector space of all homogeneous quasisymmetric functions of degree n. If $|I| \geq n$, then $\dim \mathcal{Q}^n = 2^{n-1}$.

Let $\mathrm{Comp}(n)$ denote the set of all composition of n, i.e., sequences of positive integers that sum to n. If $\alpha = (\alpha_1, \ldots, \alpha_k) \in \mathrm{Comp}(n)$, the define

$$S_\alpha = \{\alpha_1, \alpha_1 + \alpha_2, \ldots, \alpha_1 + \cdots + \alpha_{k-1}\} \subset [n-1].$$

This defines a bijection between $\mathrm{Comp}(n)$ and subsets of $[n-1]$. Let $\mathrm{co} : 2^{[n-1]} \to \mathrm{Comp}(n)$ be the inverse map.

The *fundamental quasisymmetric function* $L_\alpha(z)$ is defined by

$$L_\alpha(z) = \sum_{\substack{i_1 \le \cdots \le i_n \\ i_j < i_{j+1} \text{ if } j \in S_\alpha}} z_{i_1} \cdots z_{i_n}.$$

When $|I| \ge n$, the set $\{L_\alpha : \alpha \in \mathrm{Comp}(n)\}$ is a $\mathbb{Q}$-basis for $\mathcal{Q}^n$ [15, Prop. 7.19.1].

If $w = w_1 w_2 \cdots w_n \in S_n$, let $D(w)$ denote the *descent set* of w, i.e.,

$$D(w) = \{i : w_i > w_{i+1}\}.$$

Also, write $\mathrm{co}(w)$ for $\mathrm{co}(D(w))$.

3.2. **Connecting biased riffle shuffles and quasisymmetric functions.** We view QS as a measure on S_n, and use $QS(w)$ to denote the probability of obtaining $w \in S_n$ from the given QS-distribution. In this section, we give a simple formula for $QS(w)$ in term of the fundamental quasisymmetric functions.

The following result due to Stanley [16] is the key for the connection between symmetric function theory and riffle shuffles.

Theorem 4. *Let $w \in S_n$ and $x = (x_i)_{i \in I}$ a probability distribution on some totally ordered set I. Then*

$$QS_x(w) = L_{\mathrm{co}(w^{-1})}(x).$$

The proof of Theorem 4 is perhaps best illustrated through an example. Suppose $w = 136245$. Then using our second description of the QS-distribution, we that w is the standardization of $i_1 i_2 i_3 i_4 i_5 i_6$ iff (in lexicographic order):

$$(i_1, 1) < (i_4, 4) < (i_2, 2) < (i_5, 5) < (i_6, 6) < (i_3, 3),$$

which is equivalent to $i_1 \le i_4 < i_2 \le i_5 \le i_6 < i_3$. The probability of obtaining $i_1 i_2 i_3 i_4 i_5 i_6$ is $x_{i_1} x_{i_2} x_{i_3} x_{i_4} x_{i_5} x_{i_6}$. Therefore,

$$QS_x(136245) = \sum_{i_1 \le i_4 < i_2 \le i_5 \le i_6 < i_3} x_{i_1} x_{i_2} x_{i_3} x_{i_4} x_{i_5} x_{i_6} = L_{\mathrm{co}\{2,5\}}$$

Note that the elements of $\{2, 5\}$ are precisely the k's where k appears to the right of $k+1$ in w, which are just the element of $D(w^{-1})$. It is easy to turn this example into a proof of the general case, so we omit the details.

If we use an uniform distribution, then Theorem 4 becomes the following well-known result due to Bayer and Diaconis [4, Thm. 3].

Theorem 5. *Let $w \in S_n$, and U_a the uniform distribution on a elements. If $r = \#D(w^{-1}) + 1$, then*

$$QS_{U_a}(\pi) = \frac{1}{a^n} \binom{a + n - r}{n}.$$

Indeed, we know that [15, p.364]

$$L_\alpha(1^a) = \binom{a - \#S_\alpha + n - 1}{n},$$

so that

$$QS_{U_a}(\pi) = L_{\text{co}(w^{-1})}\left(\frac{1}{a}, \dots, \frac{1}{a}\right) = \frac{1}{a^n} L_{\text{co}(w^{-1})}(1^a) = \frac{1}{a^n}\binom{a+n-r}{n}$$

The quantity r is referred as the number of *rising sequences* [4]. A rising sequence is a maximal subset of an arrangement of cards, consisting of successive face values displayed in order. For instance, the deck $1,5,2,3,6,7,4$ consists of two rising sequences $1,2,3,4$ and $5,6,7$ interleaved together. If the starting deck is $1,2,\dots,n$, it is easy to see that a 2-shuffle can create at most 2 rising sequences, and in general, and a-shuffle can create at most a rising sequences. This observation has been used in several card tricks, for instance, see [4, Sec. 2].

4. Schur Functions and the RSK Algorithm

In this section we relate the QS-distribution to the Robinson–Schensted–Knuth (RSK) algorithm [15, Sec. 7.11]. Recall that the RSK algorithm gives a bijection between elements w of S_n and pairs (T_i, T_r) of Standard Young Tableaux (SYT) of the same shape $\lambda \vdash n$. We denote $\lambda = \text{sh}(w)$ for the shape of the tableau, $\text{ins}(w) = T_i$ for the insertion tableau, $\text{rec}(w) = T_r$ for the recording tableau. We study the probability distribution on the shapes, which are partitions of n, induced by a biased riffle shuffle followed by the RSK algorithm.

4.1. **Schur function in biased shuffles.** Let s_λ denote the Schur function [15, Sec. 7.10] indexed by partition λ. Let $\text{SYT}(\lambda)$ denote the set of all SYTs of shape λ. The following theorem shows how to interpret s_λ in term of biased riffle shuffles.

Theorem 6. [16, Thm. 3.4] *Let $x = (x_i)_{i \in I}$ be a probability distribution on a totally ordered set I. Let n be a positive integer, and T_r a SYT of shape $\lambda \vdash n$. Then*

$$QS_x(w : \text{rec}(w) = T_r) = s_\lambda(x).$$

Proof. From [15, Thm. 7.19.7] we have

$$s_\lambda = \sum_{T \in \text{SYT}(\lambda)} L_{\text{co}(T)}$$

Here $\text{co}(T) = \text{co}(D(T))$ and where $D(T)$ is the descent set of T, consisting of all integers i such that $i+1$ appears in a lower row of T than i. Also, from [15, Lem. 7.23.1], we have $D(\text{ins}(w)) = D(w^{-1})$, so $\text{co}(\text{ins}(w)) = \text{co}(w^{-1})$.

Through the RSK correspondence, the set $\{w : \text{rec}(w) = T_r\}$ is in bijection with the set of SYTs of shape $\lambda = \text{sh}(T_r)$, where w is associated to $\text{ins}(w)$. Then, using Theorem 4, we get

$$\begin{aligned} QS_x(w : \text{rec}(w) = T_r) &= \sum_{w : \text{rec}(w) = T_r} L_{\text{co}(w^{-1})}(x) = \sum_{w : \text{rec}(w) = T_r} L_{\text{co}(\text{ins}(w))}(x) \\ &= \sum_{T \in \text{SYT}(\lambda)} L_{\text{co}(T)}(x) = s_\lambda(x) \end{aligned}$$

□

Remark. The above result was misinterpreted in [11], which claimed that $QS(w : \mathrm{ins}(w) = T) = s_\lambda(x)$ in addition to $QS(w : \mathrm{rec}(w) = T) = s_\lambda(x)$. The former claim is false, since $\mathrm{ins}(w) = T$ implies that $\mathrm{co}(w^{-1}) = \mathrm{co}(T)$, so

$$QS_x(w : \mathrm{ins}(w) = T) = \sum_{w:\mathrm{ins}(w)=T} L_{\mathrm{co}(w^{-1})}(x) = \sum_{w:\mathrm{ins}(w)=T} L_{\mathrm{co}(T)}(x) = f^\lambda L_{\mathrm{co}(T)}(x).$$

This misinterpretation is most likely due to the different interpretation of a permutation as a shuffle, as was mentioned in Section 2.2.

4.2. **Induced distribution on** $\mathrm{Par}(n)$ **and total variation distance.** The QS-distribution on S_n induces a probability distribution on $\mathrm{Par}(n)$, the set of partitions of n, by mapping each $w \in S_n$ to its shape $\mathrm{sh}(w) \in \mathrm{Par}(n)$. For the remainder of this section, let P denote this probability distribution on $\mathrm{Par}(n)$.

Corollary 7. [16, Cor. 3.5] *Let $\lambda \vdash n$, then*

$$P(\lambda) = f^\lambda s_\lambda(x),$$

where f^λ denotes the number of SYTs of shape λ.

Proof. Using Theorem 6, we have

$$P(\lambda) = \sum_{T\in\mathrm{SYT}(\lambda)} QS(w : \mathrm{rec}(w) = T) = \sum_{T\in\mathrm{SYT}(\lambda)} s_\lambda(x) = f^\lambda s_\lambda(x). \qquad \square$$

Let P_U denote the probability distribution on $\mathrm{Par}(n)$ induced by the uniform distribution on S_n. That is,

$$P_U(\lambda) = \frac{1}{n!}\#\{w \in S_n : \mathrm{sh}(w) = \lambda\} = \frac{(f^\lambda)^2}{n!},$$

where the last equality follows from the RSK correspondence. The probability measure P_U on S_n is known as the Plancherel measure.

Recall that the *total variation distance* between P and P_U is defined to be

$$\|P - P_U\|_{TV} = \max_{A\subset\mathrm{Par}(n)} |P(A) - P_U(A)| = \frac{1}{2}\sum_{\lambda\vdash n} |P(\lambda) - P_U(\lambda)|\,.$$

We would like to approximate this total variation distance. Roughly speaking, we are asking how close is a biased riffle shuffle to a truly random permutation if we are only allowed to observe the resulting RSK shape of the permutation.

The technique used in the following proof is based on a recent paper by Diaconis and Fulman [8, Thm. 3.2], where they demonstrate the connection between the distribution of "carries" in the addition of random integers and descent sets of a permutation induced by a riffle shuffle. We adapt their proof to analyze the distribution on $\mathrm{Par}(n)$.

Theorem 8. *Let c denote the coefficient of q^n in $\prod_{i,j\in I}(1 - qx_ix_j)^{-1}$. Then*

$$\|P - P_U\|_{TV} \le \frac{1}{2}\sqrt{n!c - 1}.$$

Proof. We have

$$\begin{aligned}\|P - P_U\|_{TV} &= \frac{1}{2}\sum_{\lambda\vdash n}|P(\lambda) - P_U(\lambda)| = \frac{1}{2}\sum_{\lambda\vdash n}\left|f^\lambda s_\lambda(x) - \frac{(f^\lambda)^2}{n!}\right| \\ &\leq \frac{1}{2}\sqrt{\sum_{\lambda\vdash n}\left(s_\lambda(x) - \frac{f^\lambda}{n!}\right)^2 \sum_{\lambda\vdash n}(f^\lambda)^2}\end{aligned}$$

where the last step follows from the Cauchy-Schwarz inequality. Since $P(\lambda) = f^\lambda s_\lambda(x)$ and $P_U(\lambda) = (f^\lambda)^2/n!$ are both probability distributions on $\mathrm{Par}(n)$, we have

$$\begin{aligned}\sum_{\lambda\vdash n}\left(s_\lambda(x) - \frac{f^\lambda}{n!}\right)^2\sum_{\lambda\vdash n}(f^\lambda)^2 &= \left(\sum_{\lambda\vdash n}s_\lambda(x)^2 - 2\sum_{\lambda\vdash n}\frac{f^\lambda s_\lambda(x)}{n!} + \sum_{\lambda\vdash n}\frac{(f^\lambda)^2}{n!^2}\right)n! \\ &= n!\sum_{\lambda\vdash n}s_\lambda(x)^2 - 2\sum_{\lambda\vdash n}P(\lambda) + \sum_{\lambda\vdash n}P_U(\lambda) \\ &= n!\sum_{\lambda\vdash n}s_\lambda(x)^2 - 1\end{aligned}$$

So

$$\|P - P_U\|_{TV} \leq \frac{1}{2}\sqrt{n!\sum_{\lambda\vdash n}s_\lambda(x)^2 - 1}.$$

Let $[q^n]f$ denote the coefficient of q^n in f. Then, using the Cauchy identity [15, Thm. 7.12.1], we have

$$\sum_{\lambda\vdash n}s_\lambda(x)^2 = [q^n]\sum_{\lambda\vdash n}s_\lambda(x)s_\lambda(qx) = [q^n]\prod_{i,j\in I}(1 - qx_ix_j)^{-1} = c.$$

The result follows. ☐

Let us give an estimate in the case of an a-shuffle. We have

$$c = [q^n]\left(1 - \frac{q}{a^2}\right)^{-a^2} = \frac{1}{a^{2n}}\binom{a^2+n-1}{n}.$$

Recall that $1 + x \leq e^x$ for all real x. So we have

$$n!c = \frac{n!}{a^{2n}}\binom{a^2+n-1}{n} = \prod_{i=0}^{n-1}\left(1 + \frac{i}{a^2}\right) \leq \prod_{i=0}^{n-1}\exp\left(\frac{i}{a^2}\right) = \exp\left(\frac{n(n-1)}{2a^2}\right).$$

Assume that $2a^2 \geq n(n-1)$. Then, since $e^x \leq 1 + 2x$ for all $0 \leq x \leq 1$, we get

$$\sqrt{n!c - 1} \leq \sqrt{\exp\left(\frac{n(n-1)}{2a^2}\right) - 1} \leq \sqrt{\frac{n(n-1)}{a^2}} \leq \frac{n}{a}.$$

Thus we obtain the following result.

Corollary 9. *Let n, a be positive integers such that $2a^2 \geq n(n-1)$. Let P_a be the probability distribution on* $\mathrm{Par}(n)$ *induced by an a-shuffle through the RSK algorithm. Then*

$$\|P_a - P_U\|_{TV} \leq \frac{n}{2a}.$$

In particular, if we perform 2-shuffles r times in succession, which is equivalent to a 2^r-shuffle, then

$$\|P_{2^r} - P_U\|_{TV} \leq \frac{n}{2^{r+1}}.$$

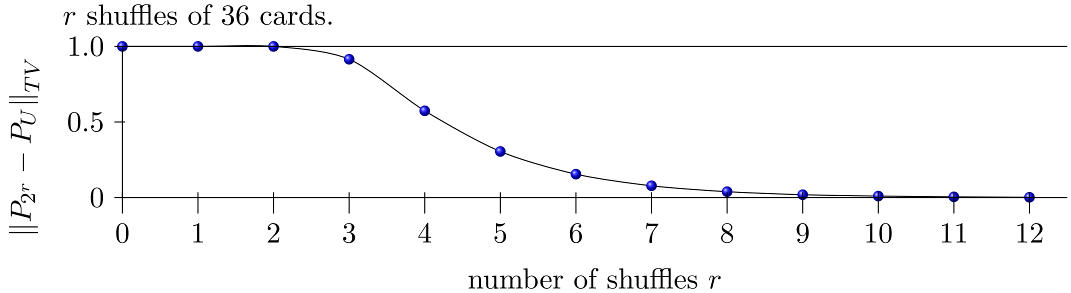
for $r \geq \log_2 n - \frac{1}{2}$. In particular, $\log_2 n + O(1)$ shuffles suffice to bring the total variation on the shapes down to a constant. In comparison, we know that $\frac{3}{2}\log_2 n + O(1)$ shuffles are needed for the underlying permutation to be close to uniform [4].

Using SAGE [14], I computed the total variation distance $\|P_{2^r} - P_U\|_{TV}$ for the case of 36 cards. Ideally, I would have like to do so for a deck of 52 cards in order to compare to the classic result of Bayer and Diaconis [4], however, the limits of computation prevented me from doing so. The code used for the computation is included in the appendix. The results are shown in Table 3 and plotted in Figure 1. Observe that after four shuffles, the total variation distance reduces by approximately half for each additional shuffle. This behavior agrees with the crude bounds that we proved just now. Such "cutoff phenomenon" is similar to what is exhibited by many Markov chains [6], in particular, the Markov chain induced by riffle shuffles on the underlying permutations in S_n [4].

TABLE 3. Total variation distance for the shapes induced by RSK under r shuffles of 36 cards.

r	1	2	3	4	5	6
$\|P_{2^r} - P_U\|_{TV}$	1.000	1.000	0.915	0.574	0.305	0.155
r	7	8	9	10	11	12
$\|P_{2^r} - P_U\|_{TV}$	0.078	0.039	0.019	0.010	0.005	0.002

FIGURE 1. Total variation distance for the shapes induced by RSK under r shuffles of 36 cards.



5. Applications to quasisymmetric function identities

In this section, we use biased riffle shuffles to derive some identities for quasisymmetric functions. While the results may not be new, the approach is certainly quite interesting.

Let us first warm up with an easy example. Since QS is a probability measure on S_n, $QS(S_n) = 1$, so

$$1 = \sum_{w \in S_n} QS(w) = \sum_{w \in S_n} L_{\mathrm{co}(w^{-1})}(x) = \sum_{w \in S_n} L_{\mathrm{co}(w)}(x)$$

as long as $\sum x_i = 1$. By normalizing, we get the following identity in $\mathcal{Q}^n$.

Proposition 10. *Let n be a positive integer. Then*

$$p_1^n = \sum_{w \in S_n} L_{\mathrm{co}(w)}.$$

Note that even though we only proved the result for real numbers $(x_i)_{i\in I}$ of finite sum, the argument can be extended to show that the identity holds formally for indeterminants x_i. For ease of exposition, we shall ignore this technicality.

Next, we use Theorem 2 to derive a few results about quasisymmetric functions in the variables $x\otimes y$. It is easy to see that a quasisymmetric function of degree n in the variables $x \otimes y$ is quasisymmetric in the x variables as well as quasisymmetric in the y variables. So $\mathcal{Q}^n(x \otimes y) \subset \mathcal{Q}^n(x) \otimes \mathcal{Q}^n(y)$. We would like to know how to express the fundamental quasisymmetric functions in $x \otimes y$ in terms those in x and y separately.

Proposition 11. *Let $\alpha \in \mathrm{Comp}(n)$. Fix a $w \in S_n$ such that $\mathrm{co}(w) = \alpha$. Then*

$$L_\alpha(x \otimes y) = \sum_{\substack{u,v\in S_n\\ vu=w}} L_{\mathrm{co}(u)}(x) L_{\mathrm{co}(v)}(y).$$

Proof. Theorem 2 implies that

$$L_{\mathrm{co}(w^{-1})}(x \otimes y) = QS_{x\otimes y}(w) = \sum_{\substack{u,v\in S_n\\ uv=w}} QS_x(u)QS_y(v) = \sum_{\substack{u,v\in S_n\\ uv=w}} L_{\mathrm{co}(u^{-1})}(x)L_{\mathrm{co}(v^{-1})}(y).$$

(Note the order of u and v in the product below the summation sign.) If we replace each of w, u, v by its inverse, then we get desired identity in the case of $\sum x_i = 1$ and $\sum y_i = 1$. Since the equation is homogeneous in both x and y, the result holds for arbitrary x and y. □

It would be interesting to find some formula for the coefficient of $L_\beta(x)L_\gamma(y)$ in $L_\alpha(x\otimes y)$, in the spirit of [15, Cor. 7.23.7]. Unfortunately, I was unable to find such a result.

Proposition 11 has the following obvious generalization to more sets of variables. The proof is analogous, so we omit it.

Proposition 12. *Let $x^{(1)}, \dots, x^{(k)}$ be disjoint sets of variables. Let $\alpha \in \mathrm{Comp}(n)$. Fix a $w \in S_n$ such that $\mathrm{co}(w) = \alpha$. Then*

$$L_\alpha\left(x^{(1)} \otimes \cdots \otimes x^{(k)}\right) = \sum_{\substack{u_1,\dots,u_k\in S_n\\ u_k u_{k-1}\cdots u_1 = w}} L_{\mathrm{co}(u_1)}(x^{(1)}) \cdots L_{\mathrm{co}(u_k)}(x^{(k)}).$$

By setting α to certain compositions, we can derive interesting identities. For instance, if $\alpha = (n)$, so that $w = \mathrm{id}$, then $L_\alpha = h_n$, and Proposition 11 gives us

$$h_n(x \otimes y) = \sum_{u\in S_n} L_{\mathrm{co}(u)}(x)L_{\mathrm{co}(u^{-1})}(y).$$

The symmetric function $h_n(x \otimes y)$ is special. Indeed, using $\sum_{n\ge 0} h_n(z) = \prod_i (1 - z_i)^{-1}$, we find that

$$\sum_{n\ge0} h_n(x \otimes y) = \prod_{i,j} (1 - x_i y_j)^{-1} = \sum_\lambda s_\lambda(x)s_\lambda(y),$$

where the last equality is the Cauchy identity [15, Thm. 7.12.1]. By taking the n-th degree homogeneous component, we obtain the following result, which can also be found in [15, Thm. 7.23.2] (although proved there without using riffle shuffles).

Proposition 13. *Let $n \in \mathbb{N}$, then*

$$\sum_{\lambda \vdash n} s_\lambda(x) s_\lambda(y) = \sum_{u \in S_n} L_{\mathrm{co}(u)}(x) L_{\mathrm{co}(u^{-1})}(y).$$

Setting $\alpha = (n)$ in Proposition 12 gives us the following multivariable generalization of 13.

Proposition 14. *Let $x^{(1)}, \ldots, x^{(k)}$ be disjoint sets of variables. For $\lambda \vdash n$, let χ^λ be the irreducible character of S_n corresponding to λ ([15, Sec. 7.18]), and let 1_{S_n} be the trivial character. Then*

$$\sum_{\lambda^1, \ldots, \lambda^k \vdash n} \left\langle 1_{S_n}, \chi^{\lambda^1} \cdots \chi^{\lambda^k} \right\rangle s_{\lambda^1}(x^{(1)}) \cdots s_{\lambda^k}(x^{(k)}) = \sum_{\substack{u_1, \ldots, u_k \in S_n \\ u_k u_{k-1} \cdots u_1 = \mathrm{id}}} L_{\mathrm{co}(u_1)}(x^{(1)}) \cdots L_{\mathrm{co}(u_k)}(x^{(k)}).$$

Proof. As before, we have

$$\sum_{\substack{u_1, \ldots, u_k \in S_n \\ u_k u_{k-1} \cdots u_1 = \mathrm{id}}} L_{\mathrm{co}(u_1)}(x^{(1)}) \cdots L_{\mathrm{co}(u_k)}(x^{(k)}) = h_n \left(x^{(1)} \otimes \cdots \otimes x^{(k)} \right)$$

and

$$\begin{aligned} \sum_{n \geq 0} h_n \left(x^{(1)} \otimes \cdots \otimes x^{(k)} \right) &= \prod_{i_1, \ldots, i_k} \left(1 - x^{(1)}_{i_1} \cdots x^{(k)}_{i_k} \right)^{-1} \\ &= \sum_{n \geq 0} \sum_{\lambda^1, \ldots, \lambda^k \vdash n} \left\langle 1_{S_n}, \chi^{\lambda^1} \cdots \chi^{\lambda^k} \right\rangle s_{\lambda^1}(x^{(1)}) \cdots s_{\lambda^k}(x^{(k)}) \end{aligned}$$

where the the last step follows from [15, Exer. 7.78]. Taking the n-th homogeneous component of both sides yields the result. □

In particular, the RHS of the equation in Proposition 14 is actually a symmetric function in each $x^{(i)}$! This is not at all obvious from just looking at the expression.

On the other hand, suppose that we set $\alpha = (1^n)$ in Proposition 11. Then $w \in S_n$ needs to be the permutation $w(k) = w^{-1}(k) = n + 1 - k$. So that $uw^{-1} = u^r$, where u^r denotes the permutation obtained from u by reversing its word representation. On the other hand, $L_\alpha = e_n$. So

$$e_n(x \otimes y) = \sum_{u \in S_n} L_{\mathrm{co}(u)}(x) L_{\mathrm{co}((u^r)^{-1})}(y). \tag{1}$$

For the LHS, we observe that

$$\sum_{n \geq 0} e_n(x \otimes y) = \prod_{i,j} (1 + x_i y_j) = \sum_\lambda s_\lambda(x) s_{\lambda'}(y)$$

where the last step is the dual Cauchy identity [15, Thm. 7.14.3]. Taking the n-th degree homomgeneous component gives

$$e_n(x \otimes y) = \sum_{\lambda \vdash n} s_\lambda(x) s_{\lambda'}(y)$$

For the RHS of (1), let us re-index the summation by switching the roles of u and u^r. Then, we get the following result, which is equivalent to the second half of [15, Thm. 7.23.2]

Proposition 15. *Let $n \in \mathbb{N}$, then*

$$\sum_{\lambda \vdash n} s_\lambda(x) s_{\lambda'}(y) = \sum_{u \in S_n} L_{\mathrm{co}(u^r)}(x) L_{\mathrm{co}(u^{-1})}(y).$$

6. Eigenvalues of the shuffling Markov Chain

Biased shuffles clearly induce a Markov chain on orderings of a deck of cards. That is, we have a Markov chain $\mathcal{M}_n$ on S_n, where the probability of transition from u to uw is $QS(w) = L_{\mathrm{co}(w^{-1})}(x)$. The following result gives the eigenvalues of $\mathcal{M}_n$.

Theorem 16. [16, Thm. 2.2] *The eigenvalues of $\mathcal{M}_n$ are the power sum symmetric functions $p_\mu(x)$ for $\mu \vdash n$. The eigenvalue $p_\mu(x)$ occurs with multiplicity $n!/z_\mu$, the number of elements of S_n of cycle type μ.*

In this section, we show how to deduce Theorem 16 from a more general result by Bidigare, Hanlon, and Rockmore [5], where a more general model of shuffling, called *pop shuffles*, is studied. (In fact, the authors study the problem in an even more general context of hyperplane arrangements, but we restrict ourselves here to pop shuffles). A slightly different explanation is given in [16, Sec. 2].

An *ordered partition* of a set S is a vector $(T_1, \dots, T_k)$ of nonempty subsets $T_i \subset S$ such that the sets $T_1, \dots, T_k$ partition S, that is, $T_1 \cup \cdots \cup T_k = S$ and $T_i \cap T_j = \emptyset$ whenever $i \neq j$. Suppose that $\mathcal{T} = (T_1, \dots, T_m)$ is an ordered partition. If we have a deck of cards with face values $\{1, 2, \dots, n\}$ (not necessarily sorted), define the pop shuffle indexed by $\mathcal{T}$ to be the following process: for each $1 \leq i \leq k$, take all the cards with face values in T_i, and put them into the i-th pile without changing their order, and afterwards, stack all the piles together in order, with the 1st pile on top and the k-th pile on bottom.

Example 17. If $\mathcal{T} = (\{3,4,8\}, \{2\}, \{1,6\}, \{5,7,9\})$, then the deck 624198735 is first divided into piles $(483, 2, 61, 975)$, and then assembled as 483261975.

Let $(w_\mathcal{T})$ be a probability distribution on the ordered partitions of $[n]$. Then pop shuffles induce a Markov chain on S_n. Before we state the result about the eigenvalues of this Markov chain, we need some more notation.

Given an ordered partition $\mathcal{T}$, let $\mathcal{T}^*$ denote the underlying set partition obtained by removing the order of the blocks. For instance, using the $\mathcal{T}$ from Example 17, we have $\mathcal{T}^* = \{\{1,6\}, \{2\}, \{3,4,8\}, \{5,7,9\}\}$.

The set partitions of $[n]$ form a partially ordered set under the refinement ordering. We write $\mathcal{A} \leq \mathcal{B}$ if each block of $\mathcal{A}$ is contained in a block of $\mathcal{B}$.

The following theorem, which we state without proof, completely characterizes the eigenvalues of the Markov chain induced by pop shuffles.

Theorem 18. [5, Thm. 2.1] *The distinct eigenvalues of the Markov chain on S_n induced by pop shuffles selected using the probability distribution $(w_\mathcal{T})$ can be indexed by set partitions of $[n]$. If $\lambda_\mathcal{A}$ denotes the eigenvalue associated to the set partition $\mathcal{A}$, then*

$$\lambda_\mathcal{A} = \sum_{\mathcal{T}:\mathcal{A}\leq\mathcal{T}^*} w_\mathcal{T}.$$

The multiplicity of $\lambda_\mathcal{A}$ is the number of permutations with cycle partition equal to $\mathcal{A}$.

Now we apply Theorem 18 to prove Theorem 16.

Consider the following process for producing a random ordered partition $\mathcal{T}$ of $[n]$. Set up a collection of blocks indexed by a totally ordered set I, and, as usual, let $(x_i)_{i\in I}$ be a probability distribution on I. For each $j \in [n]$ independently, assign it to one of the block, such that block i is chosen with probability x_i. Finally, take all the nonempty blocks, and form the ordered partition $\mathcal{T}$ of $[n]$ where the blocks are ordered in accordance to I.

If $\mathcal{T} = (T_1, \dots, T_k)$, and $\alpha = (\alpha_1, \dots, \alpha_k) \in \mathrm{Comp}(n)$ is the type of $\mathcal{T}$, i.e., $\alpha_j = \#T_i$ for each i, then the probability that $\mathcal{T}$ is chosen under previous process is

$$w_\mathcal{T} = \sum_{i_1<\cdots<i_k} x_{i_1}^{\alpha_1}\cdots x_{i_k}^{\alpha_k}.$$

Indeed, we need all the elements in T_j to be in the same pile, i_j, and such that $i_1 < \cdots < i_k$.

The above process for choosing a random $\mathcal{T}$, and then followed by the corresponding pop shuffle, is exactly the description for the inverse riffle shuffle. Therefore, the Markov chain on S_n induced by pop shuffles with the above weights is just the reverse of the Markov chain on S_n induced by the QS-distribution (i.e., biased shuffles). So the two transition matrices are transposes of each other, and hence they have the same multiset of eigenvalues. Let us now compute these eigenvalues.

For a set partition $\mathcal{A}$ of $[n]$, define its type μ to be the partition of n whose parts correspond to the sizes of the blocks in $\mathcal{A}$. For notational convenience in our next lemma, assume that $\mathcal{A} = \{A_1, \dots, A_k\}$ where $\#A_j = \mu_j$.

Lemma 19. *For any set partition $\mathcal{A}$ of type μ, we have*

$$p_\mu = \sum_{\mathcal{T}:\mathcal{A}\leq\mathcal{T}^*} w_\mathcal{T}.$$

Proof. (Sketch) Let $\mu = (\mu_1, \dots, \mu_k)$. Consider the monomials in the product

$$p_\mu = \left(\sum_{i\in I} x_i^{\mu_1}\right)\left(\sum_{i\in I} x_i^{\mu_2}\right)\cdots\left(\sum_{i\in I} x_i^{\mu_k}\right)$$

formed by picking one term from each bracket. If the resulting product is equal to $x_{i_1}^{a_1}\cdots x_{i_m}^{a_m}$, where $i_1 < i_2 < \cdots < i_m$ and $a_1, \dots, a_m > 0$, then associate it to the ordered partition $\mathcal{T} = (T_1, \dots, T_m)$, such that T_j is the union of all A_l where $x_{i_j}^{\mu_l}$ is selected from the l-th bracket. Then $|T_j| = a_j$, $\mathcal{A} \leq \mathcal{T}^*$, and the monomial $x_{i_1}^{a_1}\cdots x_{i_m}^{a_m}$ appears with coefficient 1 in $w_{\mathcal{T}}$. It is easy to see that this description gives a bijection between the monomials in the two sides of equation, thereby showing the result. □

It follows Proposition 18 and Lemma 19 that the eigenvalues of $\mathcal{M}_n$ are equal to p_μ, for $\mu \vdash n$. As for the multiplicity of the eigenvalue p_μ, from Theorem 18 where we sum over all $\mathcal{A}$ with type μ, we see that the multiplicity is equal to the number of permutations with cycle type μ. This completes the proof of Theorem 16.

References

1. D. Aldous and P. Diaconis, *Shuffling cards and stopping times*, Amer. Math. Monthly **93** (1986), no. 5, 333–348.
2. S. Assaf, P. Diaconis, and K. Soundararajan, *A rule of thumb for riffle shuffling*, preprint, avaiable `http://stat.stanford.edu/~cgates/PERSI/papers/redblack.pdf`, 2008.
3. ______, *Riffle shuffles of a deck with repeated cards*, submitted, avaiable `http://www-stat.stanford.edu/~cgates/PERSI/papers/repeatcards.pdf`, 2009.
4. D. Bayer and P. Diaconis, *Trailing the dovetail shuffle to its lair*, Ann. Appl. Probab. **2** (1992), no. 2, 294–313.
5. 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** (1999), no. 1, 135–174.
6. P. Diaconis, *The cutoff phenomenon in finite Markov chains*, Proc. Nat. Acad. Sci. U.S.A. **93** (1996), no. 4, 1659–1664.
7. P. Diaconis, J. Fill, and J. Pitman, *Analysis of top to random shuffles*, Combin. Probab. Comput. **1** (1992), no. 2, 135–155.
8. P. Diaconis and J. Fulman, *Carries, shuffling, symmetric function*, Adv. in Appl. Math. (to appear), math.CO/0902.0179.
9. P. Diaconis, M. McGrath, and J. Pitman, *Riffle shuffles, cycles, and descents*, Combinatorica **15** (1995), no. 1, 11–29.
10. J. Fulman, *The combinatorics of biased riffle shuffles*, Combinatorica **18** (1998), no. 2, 173–184.
11. ______, *Applications of symmetric functions to cycle and increasing subsequence structure after shuffles*, J. Algebraic Combin. **16** (2002), no. 2, 165–194.
12. E. Gilbert, *Theory of shuffling*, Technical memorandum, Bell Laboratories, 1955.
13. J. Reeds, *Theory of riffle shuffling*, Unpublished manuscript, 1955.
14. *Sage mathematical software*, Version 3.4, `http://www.sagemath.org`.
15. R. Stanley, *Enumerative Combinatorics, vol. 2*, Cambridge University Press, Cambridge, 1999.
16. ______, *Generalized riffle shuffles and quasisymmetric functions*, Ann. Comb. **5** (2001), 479–491.

Appendix A. SAGE computation of total variation distance

This appendix explains how the values in Table 3 were computed using SAGE.

The computation was based on the formula

$$\|P - P_U\|_{TV} = \frac{1}{2}\sum_{\lambda\vdash n}|P(\lambda) - P_U(\lambda)| = \frac{1}{2}\sum_{\lambda\vdash n}\left|f^\lambda s_\lambda(x) - \frac{(f^\lambda)^2}{n!}\right|.$$

We are interested in the case when $x = (\frac{1}{a}, \ldots, \frac{1}{a})$. Using [15, Cor. 7.21.4], we have

$$s_\lambda(x) = \frac{1}{a^n}s(1^a) = \frac{1}{a^n}\prod_{u\in\lambda}\frac{a + c(u)}{h(u)},$$

where $c(u)$ and $h(u)$ are respectively the content and hook-length of square u in the Young diagram λ.

The following function computes $s_\lambda(1^a)$.

```
def s_ones(lamb, a):
    X = 1
    for i in range(len(lamb)):
        for j in range(lamb[i]):
            X *= (a + i - j)/Partition(lamb).hook(i,j)
    return X
```

And the following function computes $\|P_a - P_U\|_{TV}$ for a deck of n cards.

```
def TV(n,a):
    sum = 0.
    n_fact = factorial(n)
    Par = Partitions(n)

    for i in range(Par.count()):
        lamb = Par[i]
        f_lamb = StandardTableaux(lamb).count()
        s = s_ones(lamb, a) / a^n
        diff = f_lamb * s - f_lamb^2/n_fact
        sum += abs(diff)
    return sum/2
```

The running time of this algorithm grows faster than the number of partitions of n. It would be nice to have a more efficient algorithm (exact or Monte-Carlo) that works for larger decks.

E-mail address: yufeiz@mit.edu