Turkish Journal of Mathematics

http://journals.tubitak.gov.tr/math/

Research Article

Turk J Math
(2018) 42: 502 – 514
© TÜBİTAK
doi:10.3906/mat-1610-59

Descent-inversion statistics in riffle shuffles

Ümit IŞLAK*
Department of Mathematics, Boğaziçi University, İstanbul, Turkey

Received: 15.10.2016 • **Accepted/Published Online:** 21.05.2017 • **Final Version:** 24.03.2018

Abstract: The purpose of this paper is to answer a question of Fulman on the asymptotic normality of the number of inversions in riffle shuffles. We will also study asymptotics for the number of descents and the length of the longest alternating subsequences in the same shuffling scheme.

Key words: Random Permutations, random Words, riffle shuffles, inversions, descents, longest alternating subsequences

1. Introduction

For a sequence $\mathbf{x} = (x_1, \dots, x_n)$ of real numbers, the numbers of descents and inversions are defined by $des(\mathbf{x}) = \sum_{i=1}^{n-1} \mathbf{1}(x_i > x_{i+1})$ and $inv(\mathbf{x}) = \sum_{1 \le i < j \le n} \mathbf{1}(x_i > x_j)$, respectively. For a permutation π in the symmetric group S_n, we write $des(\pi)$ for the number of descents in the sequence $(\pi(1), \pi(2), \dots, \pi(n))$. Similar notation will be used for the number of inversions and other permutation statistics. In statistics literature, the number of inversions (or its slight variations) is commonly known as Kendall's tau. Both the number of descents and inversions and various other related statistics are standard tools in nonparametric statistics to test the distribution of random sequences.

In this paper, we will analyze $des(\rho)$ and $inv(\rho)$ when ρ is a random permutation with riffle shuffle distribution (which is defined in the next section precisely). In particular, we will answer a question of Fulman in [10] by proving the asymptotic normality of the number of inversions in riffle shuffles.

All statistics previously mentioned are well understood when the underlying permutation is picked uniformly at random. See [10] for a proof that both the number of descents and the number of inversions are asymptotically normal in a uniform setting and [16] for an extension of the results of [10] to generalized descents. Also, let us note that the number of inversions is also studied in random words [4, 14] and in unfair permutations [2].

The proofs that follow will be mainly based on tools from Stein's method, which is a technique used for obtaining convergence rates in distributional approximations. The method was first introduced for normal approximation in 1972 [20], but since then it has been applied to several other distributions. Stein's method in general makes use of characterizing differential equations of distributions and various coupling constructions to get error bounds with respect to certain probability metrics. See [18] for an introductory survey on Stein's method and [5] for an involved analysis of normal approximation.

*Correspondence: umit.islak1@boun.edu.tr
2010 *AMS Mathematics Subject Classification:* 05A05, 60C05, 60F05

Some notation is used below. For $a \in \mathbb{N}$, $[a]$ is defined to be the set $\{1, \ldots, a\}$. $\mathcal{G}$ stands for a standard normal random variable. For given random variables X and Y, $X \leq_s Y$ is used for the stochastic dominance of X by Y. We write $\rightarrow_d$ for convergence in distribution. C will be used for a constant that does not depend on the underlying parameter n, which may vary in different lines. Finally, the standard abbreviation i.i.d. is used for independent and identically distributed.

The organization of the manuscript is as follows. Section 2 provides background on riffle shuffles and makes the connection to random words via inverse shuffles. Inverse shuffles for a variation of top m to random shuffles is also introduced and discussed in the same section. Section 3 treats the asymptotic distribution of the number of inversions in riffle shuffles. Finally, in Section 4 we provide asymptotic results for the number of descents and the length of longest alternating subsequences in uniformly random permutations and riffle shuffles.

2. Riffle shuffles and connection to random words

The method most often used to shuffle a deck of cards is the following: first, cut the deck into two piles and then riffle the piles together, that is, drop the cards from the bottom of each pile to form a new pile. The first mathematical models for riffle shuffles were introduced in [11] and (Reeds J, *Theory of riffle shuffling*, unpublished manuscript, 1981), and they were later further insvestigated in [1, 3, 10]. Now, following [10], we give two equivalent rigorous descriptions of this shuffling scheme.

Description 1 *Cut the n card deck into a piles by picking pile sizes according to the mult$(a; \boldsymbol{p})$ distribution, where $\boldsymbol{p} = (p_1, \ldots, p_a)$. That is, choose $b_1, \ldots, b_a$ with probability*

$$\binom{n}{b_1, \ldots, b_a} \Pi_{i=1}^{a} p_i^{b_i}.$$

Then choose uniformly one of the $\binom{n}{b_1,\ldots,b_a}$ ways of interleaving the packets, leaving the cards in each pile in their original order.

Definition 2.1 *The probability distribution on S_n resulting from Description 1 will be called the* riffle shuffle distribution *and will be denoted by $P_{n,a,\boldsymbol{p}}$. When $\boldsymbol{p} = (1/a, 1/a, \ldots, 1/a)$, the shuffle is said to be* unbiased *and the resulting probability measure is denoted by $P_{n,a}$. Otherwise, the shuffle is said to be* biased.

Note that the usual way of shuffling n cards with two hands corresponds to $P_{n,2,\mathbf{p}}$. Before moving on to Description 2, let us give an example via unbiased 2-shuffles. The permutation

$$\rho_{n,2} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 2 & 5 & 3 & 6 & 7 & 4 \end{pmatrix}$$

is a possible outcome of the $P_{n,2}$ distribution. Here the first four cards form the first pile, the last three form the second one, and these two piles are riffled together. The following alternative description will be important in the sequel.

Description 2 *(Inverse a-shuffles) The inverse of a biased a-shuffle has the following description. Assign independent random digits from $\{1, \ldots, a\}$ to each card with distribution $\boldsymbol{p} = (p_1, \ldots, p_a)$. Then sort according to digit, preserving relative order for cards with the same digit.*

In other words, if σ is generated according to Description 2, then $\sigma^{-1} \sim P_{n,a,\mathbf{p}}$. A proof of the equivalence of these two descriptions (with two other formulations) for unbiased shuffles can be found in [3]. Extension to the biased case is straightforward. Now let us give an example of generating a random permutation with distribution $P_{n,2}$ using inverse shuffles.

Consider a deck of 7 cards. We wish to shuffle this deck with the unbiased 2-shuffle distribution using inverse shuffles. Let $\mathbf{X} = (X_1, \ldots, X_n) = (1,1,2,1,2,2,1)$ be a sample from the uniform distribution over $\{1,2\}^7$. Then sorting according to digits preserving relative order for cards with the same digit gives the new configuration of cards as $(1,2,4,7,3,5,6)$. In the usual permutation notation, the resulting permutation after the inverse shuffle is

$$\sigma = \left(\begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 2 & 4 & 7 & 3 & 5 & 6 \end{array} \right),$$

and the resulting sample from $P_{n,2}$ is

$$\rho_{n,2} := \sigma^{-1} = \left(\begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 2 & 5 & 3 & 6 & 7 & 4 \end{array} \right).$$

Letting $\rho_{n,a,\mathbf{p}}$ be a random permutation with distribution $P_{n,a,\mathbf{p}}$ that is generated using inverse shuffles with the random word $\mathbf{X} = (X_1, \ldots, X_n)$, observe that

$$\rho_{n,a,\mathbf{p}}(i) = \#\{j : X_j < X_i\} + \#\{j \leq i : X_j = X_i\}.$$

Lemma 2.2 *Let $\mathbf{X} = (X_1, \ldots, X_n)$ where X_is are independent with distribution $\boldsymbol{p} = (p_1, \ldots, p_a)$. Also let $\rho_{n,a,\boldsymbol{p}}$ be the corresponding random permutation having distribution $P_{n,a,\boldsymbol{p}}$. Then for $i < k$, $\rho_{n,a,\boldsymbol{p}}(i) > \rho_{n,a,\boldsymbol{p}}(k)$ if and only if $X_i > X_k$.*

Proof Note that for $i, k \in [n]$, we have $\rho_{n,a,\mathbf{p}}(i) > \rho_{n,a,\mathbf{p}}(k)$ if and only if

$$\#\{j : X_j < X_i\} + \#\{j : j \leq i, X_j = X_i\} > \#\{j : X_j < X_k\} + \#\{j : j \leq k, X_j = X_k\}. \tag{2.1}$$

Since $i < k$, it is now immediate that the relation in (2.1) can occur only when $X_i > X_k$. □

This has the following corollary:

Corollary 2.3 *Consider the setting in Lemma 2.2 and let $S \subset \{(i,j) \in [n] \times [n] : i < j\}$. Then*

$$\sum_{(i,j) \in S} \mathbf{1}(\rho_{n,a,\boldsymbol{p}}(i) > \rho_{n,a,\boldsymbol{p}}(j)) = \sum_{(i,j) \in S} \mathbf{1}(X_i > X_j). \tag{2.2}$$

In the following sections, we will make use of (2.2) to study various statistics of riffle shuffles. Before that, let us close this section by describing how the random word approach above can be employed to study a variation of top m to random shuffles, which was first introduced in Section 5 in [9]. Consider a deck of n cards and let $0 \leq m \leq n$ be fixed. Now cut off the top m cards and insert them randomly among the remaining $n - m$ cards, keeping both packets in the same relative order. We will call this shuffling method *ordered top m to random shuffles.*

An ordered top m to random shuffle is actually equivalent to a 2-shuffle in which exactly m cards are cut off (whereas for the 2-shuffles case, m is a binomial random variable). It is not hard to see that the following result gives an inverse description of ordered top m to random shuffles.

Proposition 2.4 *The inverse of an ordered top m to random shuffle has the following description. Assign card $i \in [n]$ a random bit X_i where the random vector $\mathbf{X} = (X_1, \ldots, X_n)$ is uniformly distributed over $\{1,2\}^n$ with the restriction that $\sum_{i=1}^n X_i = n - m$. Then sort according to digit, preserving relative order for cards with the same digit.*

Letting τ be a random permutation in S_n with ordered top m to random shuffle distribution, Proposition 2.4 allows us to rewrite $des(\tau)$ or $inv(\tau)$ in a useful way exactly as we did in Corollary 2.3. Namely, we have

$$des(\tau) =_d \sum_{i=1}^{n-1} \mathbf{1}(X_i > X_{i+1}) \text{ and } inv(\tau) =_d \sum_{i<j} \mathbf{1}(X_i > X_j),$$

where this time $\mathbf{X} = (X_1, \ldots, X_n)$ is uniformly distributed over $\{1,2\}^n$ with the restriction that $\sum_{i=1}^n X_i = n - m$. Hence, the problem is transformed into a problem of uniform permutations of a fixed multiset, which is well studied in the literature. See, for example, [8]. We will revisit this in Section 4.

3. Central limit theorem for the number of inversions

In this section we will discuss the asymptotic normality of the number of inversions in biased riffle shuffles and will provide a convergence rate of order $1/\sqrt{n}$ in the Kolmogorov distance. This will in particular show that the conjecture of Fulman in [10], the asymptotic normality of number of inversions in an unbiased setting, is true as well. Before getting into the main result, let us note that the asymptotic normality of the number of inversions in random words was recently proven by Bliem and Kousidis [4] without convergence rates in a more general framework. In [14], Janson gave equivalent descriptions of the random words problem and analyzed the asymptotic behavior again without converge rates.

Recalling that the Kolmogorov distance between two probability measures μ and ν on $\mathbb{R}$ is defined by

$$d_K(\mu, \nu) = \sup_{z \in \mathbb{R}} |\mu((-\infty, z]) - \nu((-\infty, z])|,$$

our main result on the number of inversions in riffle shuffles is as follows.

Theorem 3.1 *Let $\rho_{n,a,\mathbf{p}}$ be a random permutation with distribution $P_{n,a,\mathbf{p}}$ where $\mathbf{p}$ is nondegenerate. Then*

$$d_K\left(\frac{inv(\rho_{n,a,\mathbf{p}}) - e_n}{s_n}, \mathcal{G}\right) \leq \frac{C}{\sqrt{n}},$$

$$e_n = \mathbb{E}[inv(\rho_{n,a,\mathbf{p}})] = \binom{n}{2} \sum_{k=2}^{a} \left(\sum_{i=1}^{k-1} p_i\right) p_k,$$

and

$$\begin{aligned} s_n^2 \quad &\sim \quad \binom{n}{2} \sum_{i=1}^{a-1} \sum_{j=i+1} p_i p_j . + \binom{n}{2}\binom{n-2}{2} \left(\sum_{i=1}^{a-1} \sum_{j=i+1}^{a} p_i p_j\right)^2 \\ &\quad +10\binom{n}{3} \sum_{j=2}^{a-1} \left(\left(\sum_{i=1}^{j-1} p_i\right)\left(\sum_{k=j+1}^{a} p_k\right)\right) p_j - \left(\binom{n}{2} \sum_{k=2}^{a} \left(\sum_{i=1}^{k-1} p_i\right) p_k\right)^2 \\ &= \quad Var(inv(\rho_{n,a,\mathbf{p}})). \end{aligned}$$

Our strategy in the proof will be using Corollary 2.3 to transform the problem into random words language and then using Chen and Shao's results on asymptotics of U-statistics [7]. Before moving on to the proof of main result, we provide some necessary background on U-statistics. First, for a real valued symmetric function $h : \mathbb{R}^m \to \mathbb{R}$ and for a random sample $X_1, \dots, X_n$ with $n \geq m$, a *U-statistic with kernel* h is defined by

$$U_n = U_n(h) = \frac{1}{\binom{n}{m}} \sum_{C_{m,n}} h(X_{i_1}, \dots, X_{i_m})$$

where the summation is over the set $C_{m,n}$ of all $\binom{n}{m}$ combinations of m integers, $i_1 < i_2 < \dots < i_m$ chosen from $\{1, \dots, n\}$. The next result of Chen and Shao on U-statistics will turn out to be useful in our proof. Here, $h_1(X_1) := \mathbb{E}[h(X_1, \dots, X_m)|X_1]$. Also recall that a kernel $h : \mathbb{R}^n \to \mathbb{R}$ is said to be *symmetric* if $h(x_1, x_2, \dots, x_n) = h(x_{\tau(1)}, x_{\tau(2)}, \dots, x_{\tau(n)})$, for any $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, and for any $\tau \in S_n$.

Theorem 3.2 *[7] Let $X_1, \dots, X_n$ be i.i.d. random variables, U_n be a U-statistic with symmetric kernel h, $\mathbb{E}[h(X_1, \dots, X_m)] = 0, \sigma^2 = Var(h(X_1, \dots, X_m)) < \infty$, and $\sigma_1^2 = Var(h_1(X_1)) > 0$. If in addition $\mathbb{E}|h_1(X_1)|^3 < \infty$, then*

$$d_K\left(\frac{\sqrt{n}}{m\sigma_1}U_n, \mathcal{G}\right) \leq \frac{6.1\mathbb{E}|h_1(X_1)|^3}{\sqrt{n}\sigma_1^3} + \frac{(1+\sqrt{2})(m-1)\sigma}{(m(n-m+1))^{1/2}\sigma_1}. \tag{3.1}$$

Before moving on to the proof, let us finally note that σ_1 satisfies the following asymptotic relation

$$\frac{m\sigma_1}{\sqrt{n}} \sim \sqrt{Var(U_n)} \tag{3.2}$$

(see Theorem 1 of Ferguson T, U-Statistics, lecture notes, http://www.math.ucla.edu/ tom/Stat200C/Ustat.pdf).

Proof Let $\mathbf{p}$ be nondegenerate and $\rho_{n,a,\mathbf{p}}$ have distribution $P_{n,a,\mathbf{p}}$. Leaving the details for the computations regarding e_n and s_n to the end of the proof, we first prove that $inv(\rho_{n,a,\mathbf{p}})$ can indeed be represented as a U-statistic.

Let $U_1, \dots, U_n$ be independent random variables uniformly distributed over $(0,1)$. Let $\sigma \in S_n$ be a random permutation so that $U_{\sigma(1)} < U_{\sigma(2)} < \dots < U_{\sigma(n)}$. Noting that σ has uniform distribution over S_n, we have

$$(X_1, \dots, X_n) =_d (X_{\sigma(1)}, \dots, X_{\sigma(n)}). \tag{3.3}$$

Therefore,

$$\begin{aligned} inv(\rho_{n,a,\mathbf{p}}) = \sum_{i<j} \mathbf{1}(\rho_{n,a,\mathbf{p}}(i) > \rho_{n,a,\mathbf{p}}(j)) =_d \sum_{i<j} \mathbf{1}(X_i > X_j) \quad &=_d \quad \sum_{i<j} \mathbf{1}(X_{\sigma(i)} > X_{\sigma(j)}) \\ &= \quad \sum_{i,j=1}^{n} \mathbf{1}(X_{\sigma(i)} > X_{\sigma(j)}, i<j), \end{aligned}$$

where the third equality follows from (3.3). Observing $i < j$ if and only if $U_{\sigma(i)} < U_{\sigma(j)}$, we obtain

$$inv(\rho_{n,a,\mathbf{p}}) =_d \sum_{i,j=1}^{n} \mathbf{1}(X_{\sigma(i)} > X_{\sigma(j)}, U_{\sigma(i)} < U_{\sigma(j)}) = \sum_{i,j=1}^{n} \mathbf{1}(X_i > X_j)\mathbf{1}(U_i < U_j). \tag{3.4}$$

Let $Z_i = (X_i, U_i)$, $i = 1, \ldots, n$, and observe that Z_is are i.i.d. random variables. Define the functions f and g by

$$f((x_i, u_i), (x_j, u_j)) = \binom{n}{2} \mathbf{1}(x_i > x_j) \mathbf{1}(u_i < u_j)$$

and

$$g((x_i, u_i), (x_j, u_j)) = f((x_i, u_i), (x_j, u_j)) + f((x_j, u_j), (x_i, u_i)).$$

Then g is clearly a real valued symmetric function and

$$\sum_{k,l=1}^{n} \mathbf{1}(X_k > X_l) \mathbf{1}(U_k < U_l) = \frac{1}{\binom{n}{2}} \sum_{k<l} g(Z_k, Z_l). \tag{3.5}$$

Thus, by (3.4) and (3.5) we conclude that $inv(\rho_{n,a,\mathbf{p}})$ is a U-statistic with

$$inv(\rho_{n,a,\mathbf{p}}) =_d \binom{n}{2}^{-1} \sum_{i<j} \binom{n}{2} (\mathbf{1}(X_i > X_j) \mathbf{1}(U_i < U_j) + \mathbf{1}(X_i < X_j) \mathbf{1}(U_i > U_j)).$$

This in particular implies that $U_n := inv(\rho_{n,a,\mathbf{p}}) - \mathbb{E}[inv(\rho_{n,a,\mathbf{p}})]$ is also a U-statistic.

We now focus on the moments of $inv(\rho_{n,a,\mathbf{p}})$. First, we have

$$\mathbb{E}[inv(\rho_{n,a,\mathbf{p}})] = \mathbb{E}[inv(X)] = \mathbb{E}\left[\sum_{i<j} \mathbf{1}(X_i > X_j)\right] = \sum_{i<j} \mathbb{P}(X_i > X_j) = \binom{n}{2} \mathbb{P}(X_1 > X_2),$$

where the last equality follows since X_is are i.i.d.. Also,

$$\mathbb{P}(X_1 > X_2) = \sum_{k=2}^{a} \mathbb{P}(X_1 > X_2, X_1 = k) = \sum_{k=2}^{a} \left(\sum_{i=1}^{k-1} p_i\right) p_k,$$

from which the conclusion that

$$\mathbb{E}[inv(\rho_{n,a,\mathbf{p}})] = \binom{n}{2} \sum_{k=2}^{a} \left(\sum_{i=1}^{k-1} p_i\right) p_k \tag{3.6}$$

follows. For the computation of $Var(inv(\rho_{n,a,\mathbf{p}}))$, let $\mathcal{S} = \{(\alpha, \beta, \gamma, \delta) : 1 \leq \alpha < \beta \leq n, 1 \leq \gamma < \delta \leq n\}$. Further, define

$$\mathcal{S}_0 = \{(\alpha, \beta, \gamma, \delta) \in \mathcal{S} : \alpha = \gamma, \beta = \delta\},$$

$$\mathcal{S}_1 = \{(\alpha, \beta, \gamma, \delta) \in \mathcal{S} : \{\alpha, \beta\} \cap \{\gamma, \delta\} = \emptyset\}, \qquad \mathcal{S}_2 = \{(\alpha, \beta, \gamma, \delta) \in \mathcal{S} : \alpha = \delta\},$$

and

$$\mathcal{S}_3 = \{(\alpha, \beta, \gamma, \delta) \in \mathcal{S} : \alpha = \gamma, \delta < \beta\}, \qquad \mathcal{S}_4 = \{(\alpha, \beta, \gamma, \delta) \in \mathcal{S} : \beta = \delta, \alpha < \gamma\}.$$

Then a simple manipulation shows that

$$\mathbb{E}[(inv(X))^2] = \sum_{\mathcal{S}_0} P_{\alpha,\beta,\gamma,\delta} + \sum_{\mathcal{S}_1} P_{\alpha,\beta,\gamma,\delta} + 2\sum_{\mathcal{S}_2} P_{\alpha,\beta,\gamma,\delta} + 2\sum_{\mathcal{S}_3} P_{\alpha,\beta,\gamma,\delta} + 2\sum_{\mathcal{S}_4} P_{\alpha,\beta,\gamma,\delta},$$

where $P_{\alpha,\beta,\gamma,\delta} = \mathbb{P}(X_\alpha > X_\beta, X_\gamma > X_\delta)$. Here and below, we write $\sum_{\mathcal{S}.}$ instead of $\sum_{(\alpha,\beta,\gamma,\delta)\in\mathcal{S}.}$ for convention. The factors 2 in front of the sum over $\mathcal{S}_3$ and the sum over $\mathcal{S}_4$ are due to obvious symmetries. To see why we have the factor 2 in front of the sum over $\mathcal{S}_2$, observe that the summation of $P_{\alpha,\beta,\gamma,\delta}$ over the set $\{(\alpha,\beta,\gamma,\delta) \in \mathcal{S} : \beta = \gamma\}$ is the same as the summation over $\mathcal{S}_2$, and that it is not included any other set defined above.

Now we compute these five sums separately.

Sum over $\mathcal{S}_0$***.*** Clearly,

$$\sum_{\mathcal{S}_0} P_{\alpha,\beta,\gamma,\delta} = \mathbb{E}[inv(X)] = \binom{n}{2}\sum_{i=1}^{a-1}\sum_{j=i+1}^{a} p_i p_j. \tag{3.7}$$

Sum over $\mathcal{S}_1$***.*** We have

$$\sum_{\mathcal{S}_1} P_{\alpha,\beta,\gamma,\delta} = \sum_{\mathcal{S}_1} (\mathbb{P}(X_1 > X_2))^2 = \binom{n}{2}\binom{n-2}{2}\left(\sum_{i=1}^{a-1}\sum_{j=i+1}^{a} p_i p_j\right)^2. \tag{3.8}$$

Sum over $\mathcal{S}_2$***.***

$$\begin{aligned}\sum_{\mathcal{S}_2} P_{\alpha,\beta,\gamma,\delta} = \binom{n}{3}\mathbb{P}(X_1 > X_2 > X_3) &= \binom{n}{3}\sum_{j=2}^{a-1}\mathbb{P}(X_1 > j > X_3)\mathbb{P}(X_2 = j)\\ &= \binom{n}{3}\sum_{j=2}^{a-1}\left(\left(\sum_{i=1}^{j-1} p_i\right)\left(\sum_{k=j+1}^{a} p_k\right)\right)p_j. \end{aligned}\tag{3.9}$$

Sum over $\mathcal{S}_3$***.***

$$\begin{aligned}\sum_{\mathcal{S}_3} P_{\alpha,\beta,\gamma,\delta} &= \binom{n}{3}\mathbb{P}(X_1 > X_2, X_1 > X_3)\\ &= 2\binom{n}{3}\mathbb{P}(X_1 > X_2 > X_3)\\ &= 2\binom{n}{3}\sum_{j=2}^{a-1}\left(\left(\sum_{i=1}^{j-1} p_i\right)\left(\sum_{k=j+1}^{a} p_k\right)\right)p_j. \end{aligned}\tag{3.10}$$

Sum over $\mathcal{S}_4$***.*** Using the same approach we used for computation over $\mathcal{S}_3$,

$$\sum_{\mathcal{S}_4} P_{\alpha,\beta,\gamma,\delta} = 2\binom{n}{3}\sum_{j=2}^{a-1}\left(\left(\sum_{i=1}^{j-1} p_i\right)\left(\sum_{k=j+1}^{a} p_k\right)\right)p_j. \tag{3.11}$$

Combining (3.7), (3.8), (3.9), (3.10), and (3.11), we arrive at

$$\mathbb{E}[(inv(X))^2] = \binom{n}{2}\sum_{i=1}^{a-1}\sum_{j=i+1}^{a} p_i p_j. + \binom{n}{2}\binom{n-2}{2}\left(\sum_{i=1}^{a-1}\sum_{j=i+1}^{a} p_i p_j\right)^2$$
$$+10\binom{n}{3}\sum_{j=2}^{a-1}\left(\left(\sum_{i=1}^{j-1} p_i\right)\left(\sum_{k=j+1}^{a} p_k\right)\right)p_j. \quad (3.12)$$

Finally, the variance is found to be $Var(inv(X) = \mathbb{E}[(inv(X))^2] - (\mathbb{E}[inv(X])^2$, where the moments are as given in (3.6) and (3.12).

Once we have the first two moments, the asserted central limit theorem now follows by using (3.2) and estimating the right-hand side of (3.1) in an elementary way. □

Remark 3.3 *The discussion from Section 2 and a simple coupling argument gives the following stochastic dominance result for the number of inversions:*

$$inv(\rho_{n,2}) \leq_s inv(\rho_{n,a}) \leq_s inv(\pi_n),$$

where $a \geq 2$, π_n is a uniformly random permutation in S_n and $\leq_s$ denotes the standard stochastic ordering. Since the means and variances of these three statistics are of the same order, it would not be surprising to obtain the asymptotic normality of $inv(\rho_{n,a})$ by the corresponding results for $inv(\rho_{n,2})$ and $inv(\pi_n)$. We are planning to pursue this idea in a future work.

4. Two m-dependent statistics

The purpose of this section is to study the number of descents and the length of longest alternating subsequences after a riffle shuffle. Both of these two statistics will turn out to be much easier to handle than the number of inversions due to the underlying local dependence. Also, in this section we focus on unbiased shuffles for notational convenience, but the results can be extended to a biased setting in a straightforward way.

Before moving on to the main discussion, we discuss some preliminaries. First recall that, if we define the distance between two subsets of A and B of $\mathbb{N}$ by

$$\rho(A,B) := \inf\{|i-j| : i \in A, j \in B\},$$

the sequence of random variables $Y_1, Y_2, \ldots$ is said to be m-*dependent* if $\{Y_i, i \in A\}$ and $\{Y_j, j \in B\}$ are independent whenever $\rho(A,B) > m$ with $A, B \subset \mathbb{N}$. The following result from [6] about m-dependent random variables will be useful.

Theorem 4.1 *[6] Let $\{Y_i\}_{i\geq 1}$ be a sequence of zero mean m-dependent random variables with $W = \sum_{i=1}^n Y_i$ and $\mathbb{E}[W^2] = 1$. If $\mathbb{E}|Y_i|^p < \infty$, $i = 1, \ldots, n$, for some $p \in (2,3]$, then we have*

$$d_K(W, \mathcal{G}) \leq 75(10m+1)^{p-1}\sum_{i=1}^n \mathbb{E}|Y_i|^p.$$

4.1. Number of descents

Let $\rho_{n,a}$ have distribution $P_{n,a}$. We know from Corollary 2.3 that

$$des(\rho_{n,a}) = \sum_{i=1}^{n-1} \mathbf{1}(\rho_{n,a}(i) > \rho_{n,a}(i+1)) =_d \sum_{i=1}^{n-1} \mathbf{1}(X_i > X_{i+1}),$$

where X_is are independent and uniform over $[a]$. It is then easy to see that $\mathbb{E}[des(\rho_{n,a})] = (n-1)\frac{a-1}{2a}$. Also, setting $V = \sum_{i=1}^{n-1} V_i$ with $V_i = \mathbf{1}(X_i > X_{i+1}) - \frac{a-1}{2a}$, we have $Var(V_i) = \frac{a^2-1}{4a^2}$ and $Cov(V_i, V_{i+1}) = -\left(\frac{a^2-1}{12a^2}\right)$ for $i = 1, \dots, n-1$. Then

$$\begin{aligned} Var(V) = \sum_{i=1}^{n-1} Var(V_i) + 2\sum_{i<j} Cov(V_i, V_j) &= (n-1)\frac{a^2-1}{4a^2} - 2(n-1)\left(\frac{a^2-1}{12a^2}\right) \\ &= \frac{(a^2-1)(n-1)}{12a^2}. \end{aligned}$$

Letting $W = \frac{des(\rho_{n,a}) - (n-1)\frac{a-1}{2a}}{\sqrt{\frac{(a^2-1)(n-1)}{12a^2}}}$ and $Y_i = \frac{\mathbf{1}(X_i > X_{i+1}) - \frac{a-1}{2a}}{\sqrt{\frac{(a^2-1)(n-1)}{12a^2}}}$, we have $W = \sum_{i=1}^n Y_i$. Noting that Y_is are 1-dependent, and using Theorem 4.1 with $p = 3$, we arrive at:

Theorem 4.2 *Let $\rho_{n,a}$ be distributed according to $P_{n,a}$. Then*

$$d_K\left(\frac{des(\rho_{n,a}) - \frac{(a-1)(n-1)}{2a}}{\sqrt{\frac{(a^2-1)(n-1)}{12a^2}}}, \mathcal{G}\right) \leq \frac{C}{\sqrt{n}}.$$

Remark 4.3 *The number of d-descents in $\rho_{n,a}$ is defined by*

$$des_d(\rho_{n,a}) = \#\{(i,j) \in \{1, 2, \dots, n\}^2 : 1 \leq j - i \leq d, \rho_{n,a}(i) > \rho_{n,a}(j)\}.$$

The number of ordinary descents then corresponds to the case $d = 1$. When d is fixed, $des_d(\rho_{n,a})$ is still a sum of m-dependent random variables, and so one can employ Theorem 4.1 to show that a similar central limit theorem holds for $des_d(\rho_{n,a})$.

Remark 4.4 *Note that $\lim_{a\to\infty} \mathbb{E}[des(\rho_{n,a})] = \frac{n-1}{2} = \mathbb{E}[des(\pi_n)]$, where π_n is a uniformly random permutation in S_n. This is no surprise, and indeed with a little more effort, one can also conclude that*

$$des(\rho_{n,a}) \longrightarrow_d des(\pi_n),$$

as $a \to \infty$. The same also holds for other statistics discussed in this manuscript.

Remark 4.5 *We conclude this section with a discussion of the asymptotic normality of the number of descents after ordered top m to random shuffles, which were defined at the end of Section 2. We start by recalling a special case of a result of Congar and Viswanath [8] on multisets. Let $\beta \in [1/2, 1)$. Then there exists a* *constant* $C > 0$ *depending only on* β *so that whenever* τ *is a uniform permutation of the multiset* $\{0^{n_0}, 1^{n_1}\}$ *with* $n_0, n_1 \in \mathbb{N}$, $n_0 + n_1 = n$, $\max\{n_0, n_1\} \leq \beta n$,

$$d_K\left(\frac{des(\tau) - \mu}{\sigma}, \mathcal{G}\right) \leq \frac{C}{\sqrt{n}}$$

is satisfied where $\mu = \mathbb{E}[des(\tau)]$ *and* $\sigma^2 = Var(des(\tau))$ *(for details, see [8]). It is easily seen from this result and Theorem 2.4 that one can analyze the asymptotic behavior of the number of inversions in ordered top* m *to random shuffles under the assumption that* $\max\{m, n-m\} \leq \beta n$. *Note that this also suggests a natural generalization of riffle shuffles. To see this, consider the case where the number of cards in the hands is* (n_0, n_1), *where* (n_0, n_1) *is uniform over the set* $\{(n_0, n_1) \in [n] \times [n] : n_0 + n_1 = n, \min\{n_0, n_1\} \geq \alpha n\}$ *for some* $1 > \alpha \geq 0$. *When* $\alpha = 0$, *we get* $P_{n,2}$. *Using* $\alpha > 0$, *we get a different model, which can be meaningful since when one shuffles a deck it is highly likely that there will be at least a few cards in each hand.*

4.2. Another related statistic: longest alternating subsequences

In this section we will study the asymptotic behavior of lengths of longest alternating subsequences in uniform permutations and riffle shuffles. Letting $\mathbf{x} := (x_i)_{i=1}^n$ be a sequence of real numbers, a subsequence x_{i_k}, where $1 \leq i_1 < \ldots < i_k \leq n$, is called an *alternating subsequence* if $x_{i_1} > x_{i_2} < x_{i_3} > \ldots x_{i_k}$. The *length of the longest alternating subsequence of* $\mathbf{x}$ is defined by $LA_n(\mathbf{x}) := \max\{k : \mathbf{x} \text{ has an alternating subsequence of length } k\}$. For example, letting $\mathbf{x} = (3,1,7,4,2,6,5)$, one can easily see that $(3,1,7,2,6,5)$ is an alternating subsequence and that $LA_7(\mathbf{x}) = 6$. See [19] for a survey on the longest alternating subsequence problem. The following proposition, whose proof can be found in [13, 17], is quite useful for understanding $LA_n(\mathbf{x})$.

Proposition 4.6 *[17] Let* $\mathbf{x} := (x_i)_{i=1}^n$ *be a sequence of distinct real numbers. Then*

$$LA_n(\mathbf{x}) = 1 + \mathbf{1}(x_1 > x_2) + \sum_{k=2}^{n-1} \mathbf{1}(x_{k-1} > x_k < x_{k+1}) + \sum_{k=2}^{n-1} \mathbf{1}(x_{k-1} < x_k > x_{k+1}).$$

Example 4.7 *Let* $\mathbf{x} = (3,1,7,4,2,6,5)$. *Then the local maximums are* $\{x_3, x_6\} = \{7,6\}$ *and the local minimums are* $\{x_2, x_5\} = \{1,2\}$. *Noting that* $x_1 > x_2$ *and using Proposition 4.6, we get* $LA_n(\mathbf{x}) = 1+1+2+2 = 6$. *Indeed, the subsequence* $(3,1,7,2,6,5)$ *has length 6 and* $\mathbf{x}$ *does not have a longer alternating subsequence.*

For the longest alternating subsequence problem in a uniformly random permutation π_n in S_n, [13], [17], and [19] find the expectation and variance as

$$\mathbb{E}[LA_n(\pi_n)] = \frac{2n}{3} + \frac{1}{6} \quad \text{and} \quad Var(LA_n(\pi_n)) = \frac{8n}{45} - \frac{13}{180}.$$

Letting $X_1, \ldots, X_n$ be independent uniform random variables over $(0,1)$,

$$E_k = \{X_{k-1} > X_k < X_{k+1}\} \cup \{X_{k-1} < X_k > X_{k+1}\} \quad \text{for} \quad k = 2, \ldots, n-1, \tag{4.1}$$

we have

$$\begin{aligned} LA_n(\pi_n) &= 1 + \mathbf{1}(\pi_n(1) > \pi_n(2)) + \sum_{k=2}^{n-1} \mathbf{1}(\pi_n(k-1) > \pi_n(k) < \pi_n(k+1)) \\ &+ \sum_{k=2}^{n-1} \mathbf{1}(\pi_n(k-1) < \pi_n(k) > \pi_n(k+1)) =_d 1 + \mathbf{1}(X_1 > X_2) + \sum_{k=2}^{n-1} \mathbf{1}(E_k), \end{aligned}$$

where in the second equality, we used the discussion from the Introduction. Clearly, $LA_n(\pi_n)$ is a sum of 1-dependent random variables and so Theorem 4.1 can be used to improve results of [13] and [17] by obtaining convergence rates in the central limit theorem.

Theorem 4.8 *Let π_n be a uniformly random permutation in S_n. Then, for every $n \geq 1$,*

$$d_K\left(\frac{LA_n(\pi) - \left(\frac{2n}{3} + \frac{1}{6}\right)}{\sqrt{\frac{8n}{45} - \frac{13}{180}}}, \mathcal{G}\right) \leq \frac{C}{\sqrt{n}}$$

where C is a constant independent of n.

Next we work on alternating subsequences in riffle shuffles. Note that, with a close connection to the number of extremum points, longest alternating subsequences can be quite useful in nonparametric tests. Indeed, our motivation here comes from practical discussions of this issue (Nass C, *Running the Cheaters Out of Town: Counting Out Corrupt Coins, Dubious Dice, Shifty Shuffling, and Lying Lotteries*, unpublished manuscript) on cheating in card games.

We start by recalling the development of longest alternating subsequences in random words given in [13]. This time we need to be careful about defining maxima and minima properly as we may have repeated values in the sequence. We say that a sequence $\mathbf{x} = (x_1, \ldots, x_n) \in [a]^n$ has a local minimum at k if (i) $x_k < x_{k+1}$ or $k = n$, and if (ii) for some $j < k$, $x_j > x_{j+1} = \ldots = x_{k-1} = x_k$. Similarly, $\mathbf{x}$ has a local maximum at k if (i) $x_k > x_{k+1}$ or $k = n$, and if (ii) for some $j < k$, $x_j < x_{j+1} = \ldots = x_{k-1} = x_k$, or for all $j < k$, $x_j = x_k$. With these definitions, a useful representation of $LA_n(\mathbf{x})$ was found by Houdre and Restrepo [13] as

$$LA_n(\mathbf{x}) = \#\text{ of local maxima of } \mathbf{x} + \#\text{ of local minima of } \mathbf{x}.$$

Letting $\mathbf{X} = (X_1, \ldots, X_n)$ be a random word where X_is are independent and uniform over $[a]$, they also showed that

$$\frac{LA_n(\mathbf{X}) - n(2/3 - 1/3a)}{\sqrt{n}\gamma} \longrightarrow_d \mathcal{G}, \tag{4.2}$$

as $n \to \infty$ where

$$\gamma^2 = \frac{8}{45}\left(\frac{(1+1/a)(1-3/4a)(1-1/2a)}{1-2/(a+1)}\right). \tag{4.3}$$

(Note there is a typo in [13] for the expression of γ^2. This can be checked from [15] by taking limits in the corresponding variance formula.) Now Lemma 2.2, the discussion just before it, and (4.2) immediately give:

Theorem 4.9 *Let $\rho_{n,a}$ be a random permutation with distribution $P_{n,a}$. Then*

$$\frac{LA_n(\rho_{n,a}) - n(2/3 - 1/3a)}{\sqrt{n}\gamma} \longrightarrow_d \mathcal{G},$$

as $n \to \infty$ where γ is as defined in (4.3).

This result can be generalized to biased shuffles as in previous problems in a straightforward way. Asymptotic mean and variance of this case are described in detail in [13]. Also note that, due to the lack of local dependence, obtaining convergence rates is not as easy as the case of uniform random permutations for a shuffles and it will be studied in a subsequent work. However, when one focuses on $\rho_{n,2}$, one still has local dependence thanks to the following connection to the number of descents in $\rho_{n,2}$.

Proposition 4.10 *Let $\rho_{n,2,\boldsymbol{p}}$ be a random permutation with distribution $P_{n,2,\boldsymbol{p}}$ generated by inverse shuffling with the random vector $\mathbf{X} = (X_1, \ldots, X_n)$ where X_is are independent with distribution $\boldsymbol{p} = (p_1, p_2)$ with $0 < p_1 < 1$. Then for $k = 2, \ldots, n-1$,*

i. $\rho_{n,2,\boldsymbol{p}}$ has a local maximum at k if and only if $\rho_{n,2,\boldsymbol{p}}$ has a descent at k.

ii. $\rho_{n,2,\boldsymbol{p}}$ has a local minimum at k if and only if $\rho_{n,2,\boldsymbol{p}}$ has a descent at $k-1$.

Proof

i. ($\Rightarrow$) Obvious. ($\Leftarrow$) Assume $\pi(k) > \pi(k+1)$. We should show $\pi(k-1) < \pi(k)$. Since $\pi(k) > \pi(k+1)$, we see that the k^{th} card comes from the second pile and the $k+1^{st}$ from the first pile. Now whether card $k-1$ comes from the first pile or the second pile, we have $\pi(k-1) < \pi(k)$ since the relative orders of the piles are preserved.

ii. Proof is similar to the maximum case and we skip it.

□

Via Proposition 4.10, we now have

$$\begin{aligned}
LA_n(\rho_{n,2,\mathbf{p}}) &= 1 + \mathbf{1}(\rho_{n,2,\mathbf{p}}(1) > \rho_{n,2,\mathbf{p}}(2)) + \sum_{k=2}^{n-1} \mathbf{1}(\rho_{n,2,\mathbf{p}}(k-1) > \rho_{n,2,\mathbf{p}}(k) < \rho_{n,2,\mathbf{p}}(k+1)) \\
&+ \sum_{k=2}^{n-1} \mathbf{1}(\rho_{n,2,\mathbf{p}}(k-1) < \rho_{n,2,\mathbf{p}}(k) > \rho_{n,2,\mathbf{p}}(k+1)) \\
&=_d 1 + \mathbf{1}(X_1 > X_2) + \sum_{i=2}^{n-1} \mathbf{1}(X_i > X_{i+1}) + \sum_{i=1}^{n-2} \mathbf{1}(X_i > X_{i+1}) \\
&= 2\left(\sum_{i=1}^{n-1} \mathbf{1}(X_i > X_{i+1})\right) + \mathbf{1}(X_{n-1} < X_n). \qquad (4.4)
\end{aligned}$$

By the representation in (4.4), it is clear that we still have local dependence for $LA_n(\rho_{n,2,\mathbf{p}})$ and thus we can still use Theorem 4.1 with $p = 3$ to obtain a convergence rate of order $1/\sqrt{n}$ for $LA_n(\rho_{n,2,\mathbf{p}})$.

5. Concluding Remarks

In this note, after relating riffle shuffle statistics to random word statistics, we were able to obtain asymptotic normality results with convergence rates for the number of descents and inversions after an arbitrary number of a-shuffles. We also discussed how similar ideas can be used for a variant of top m to random shuffles and provided small contributions to Houdre and Restrepo's work on longest alternating subsequences [13].

In subsequent work, we will provide convergence rates for the length of longest alternating subsequences in a-shuffles for $a \geq 2$. We also hope to find a general framework for establishing the asymptotic normality of a large class of a-shuffle statistics. One possible direction for this can be using the stochastic dominance idea introduced in Remark 3.3 as in many cases it can be easier to prove the results for 2-shuffles and uniformly random permutations.

References

[1] Aldous D, Diaconis P. Shuffling cards and stopping times. Am Math Mon 1986; 93: 333-348.

[2] Arslan İ, Işlak Ü, Pehlivan C. On unfair permutations. arXiv: 1611.07275, 2016.

[3] Bayer D, Diaconis P. Trailing the dovetail shuffle. Ann Appl Probab 1992; 2: 294-313.

[4] Bliem T, Kousidis S. The number of flags in finite vector spaces: asymptotic normality and Mahonian statistics. J Algebraic Combin 2013; 37: 361-380.

[5] Chen LHY, Goldstein L, Shao, QM. Normal Approximation by Stein's Method. Berlin, Germany: Springer, 2011.

[6] Chen LHY, Shao QM. Normal approximation under local dependence. Ann Prob 2004; 32: 1985-2028.

[7] Chen LHY, Shao QM. Normal approximation for nonlinear statistics using a concentration inequality approach. Bernoulli 2007; 13: 581-599.

[8] Conger M, Viswanath D. Normal approximations for descents and inversions of permutations of multisets, J Theoret Prob 2007; 20: 309-325.

[9] Diaconis P, Fill JA, Pitman J. Analysis of top to random shuffles. Combin Probab Comput 1992; 1: 135-155.

[10] Fulman J. The combinatorics of biased riffle shuffles. Combinatorica 1998; 18: 173-184.

[11] Gilbert E. Theory of shuffling. Technical Memorandum. Murray Hill, NJ, USA: Bell Laboratories, 1955.

[12] Hoeffding W. A class of statistics with asymptotically normal distribution. Ann Math Statistics 1948; 19: 293-325.

[13] Houdre C, Restrepo R. A probabilistic approach to the asymptotics of the longest alternating subsequence. Electron J Combin 2010; 17: 1-19.

[14] Janson S. Generalized Galois numbers, inversions, lattice paths, Ferrers diagrams and limit theorems. Electron J Combin 2012; 19: P34.

[15] Mansour T. Longest alternating subsequences of k-ary words. Discrete Appl Math 2008; 156: 119-124.

[16] Pike J. Convergence rates for generalized descents. Electron J Combin 2011; 18: P236.

[17] Romik D. Local extrema in random permutations and the structure of longest alternating subsequences. In: 23rd International Conference on Formal Power Series and Algebraic Combinatorics, 2011, pp. 825-834.

[18] Ross NF. Fundamentals of Stein's method. Probab Surv 2011; 8: 210-293.

[19] Stanley R. Longest alternating subsequences of permutations. Michigan Math J 2008; 57: 675-687.

[20] Stein C. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley, CA, USA: University of California Press, 1972, pp. 583-602.