Shuffling Cards for Blackjack, Bridge, and Other Card Games

Mark Conger and D. Viswanath *

March 23, 2006

Abstract

This paper is about the following question: How many riffle shuffles mix a deck of card for games such as blackjack and bridge? An object that comes up in answering this question is the descent polynomial associated with pairs of decks, where the decks are allowed to have repeated cards. We prove that the problem of computing the descent polynomial given a pair of decks is $\#P$-complete. We also prove that the coefficients of these polynomials can be approximated using the bell curve. However, as must be expected in view of the $\#P$-completeness result, approximations using the bell curve are not good enough to answer our question. Some of our answers to the main question are supported by theorems, and others are based on experiments supported by heuristic arguments. In the introduction, we carefully discuss the validity of our answers.

1 Introduction

1.1 Probability theory, card shuffling, and computational complexity theory

The representation of a sequence of coin tosses as the binary digits of a real number appears in a 1909 paper by Borel [5]. This representation was a step towards the formalization of the notion of probability using measure theory. Card shuffling was another example that was discussed around that time by Borel, Poincaré, and others. However, the analysis of card shuffling, unlike that of coin tosses, was not advanced very far at that time. Even as he noted that a large number of shuffles brings the distribution of a deck close to the uniform distribution, Rényi wrote in the 1960s that he would not deal with "what is meant by a large enough number of movements" [25]. Much is now known about the number of shuffles for mixing a deck of cards thanks to the development of the convergence theory of Markov chains by Aldous, Diaconis, and others [1] [8].

A notable result about riffle shuffles is due to Bayer and Diaconis [3]. The riffle shuffle, where a deck of cards is cut into two packets and cards are dropped from the two packets in some order, is the most common method of shuffling cards. For a certain model of the riffle shuffle, Bayer and Diaconis gave a complete analysis of riffle shuffles. This result assumes that all 52 cards are distinct and that each of the 52! possible permutations must be nearly equally likely for the deck to considered well mixed.

However, in card games such as blackjack, the distinction between suits (a standard deck of 52 cards has 4 suits with cards labeled $\mathtt{A}, \mathtt{2}, \ldots, \mathtt{10}, \mathtt{J}, \mathtt{Q}, \mathtt{K}$ in each suit) is ignored. In other card games

*Department of Mathematics, University of Michigan, 530 Church Street, Ann Arbor, MI 48109, U.S.A., mconger@umich.edu and divakar@umich.edu. This work was supported by a research fellowship from the Sloan Foundation. Computations on the Teragrid networks at NCSA and SDSC were supported by a DAC award.

such as bridge, all 52 cards are distinct, but there are only 4 players. In bridge, it is enough if each player receives a random *set* of 13 cards — the order in which those 13 cards are received by a player is inconsequential. Therefore, in some card games, not all cards in the standard deck are distinct, and in others, not all 52! permutations need to be nearly equally likely for the deck to be considered well mixed. The mixing times for such card games is the topic of this paper.

The popularity of card games is a reason to study card shuffling, but it is not the only reason. Examples based on card shuffling are an inextricable part of the convergence theory of finite state Markov chains. One application of this theory is to theoretical computer science, and in particular, to the problem of computing the permanent of a matrix all of whose entries are 0 or 1.

Let A be an $n \times n$ matrix whose ijth entry is a_{ij}. Then its permanent is defined as

$$\text{perm}(A) = \sum_{\pi} a_{1\pi(1)} a_{2\pi(2)} \dots a_{n\pi(n)}, \tag{1.1}$$

where π ranges over the permutations of $\{1, 2, \dots, n\}$. If the term under the summation in (1.1) were multiplied by ± 1, with the sign being $+$ for even π and $-$ for odd π, we would have a definition of the determinant of A. In spite of this resemblance to the determinant, computing the permanent is much harder than computing the determinant. As every student of linear algebra knows, the determinant can be computed using $O(n^3)$ arithmetic operations by carrying out Gaussian elimination or row reduction. But there is no known way of computing the permanent that is much better than using its definition (1.1) directly. The operation count for the direct method is more than $n!$. Even for $n = 50$, such an operation count is far out of the reach of today's computers.

The permanent belongs to a class of counting problems known as $\#P$. It is also $\#P$-complete which means that every problem in $\#P$ can be reduced to it in time that is polynomial in the size of the problem. There are many other $\#P$-complete problems, some of them of considerable importance to trade and industry, but it is believed that there is no polynomial time algorithm for any of these problems. The $P \neq NP$ conjecture is closely related to this belief. Proof or disproof of either conjecture would be a major advance in computational complexity theory.

If it is only desired to approximate the permanent, rather than compute it exactly, Markov chains are of help. Jerrum, Sinclair, and Vigoda [17] have devised a Markov chain, analyzed its convergence, and proved that the permanent can be approximated accurately with high probability in polynomial time.

This relationship between Markov chains and computational complexity theory is inverted in our work. To find out the number of riffle shuffles that mix a deck of cards for blackjack, bridge and other card games, we need to be able to find the transition probability between two given decks, with some cards repeated, under a given number of riffle shuffles. There are simple formulas to pass back and forth between transition probabilities between two decks and a polynomial that will be called the descent polynomial for those two decks. We prove that given two decks, the problem of computing their descent polynomial is $\#P$-complete.

A graph of the coefficients of the descent polynomial looks strikingly like a bell curve for most pairs of decks. We prove a theorem showing that coefficients of the descent polynomial are approximated by the normal law in very general circumstances. The coefficients of the descent polynomial that we need to figure out the mixing times for games such as blackjack and bridge lie in the tail of the normal fit. Unfortunately, but not unexpectedly, the bounds on the normal approximation are not sharp enough in this region. The proof of $\#P$-completeness in Section 4 suggests that these must be the coefficients that are hard to compute accurately, and indeed

	1	2	3	4	5	6	7	8	9	10
BayerDiaconis	**1**	**1**	**1**	**1**	**.924**	**.614**	**.334**	**.167**	**.085**	**.043**
Blackjack1	**1**	**1**	**1**	**.481**	**.215**	**.105**	**.052**	**.026**	**.013**	**.007**
Bridge2			.45?	.16	.08	.04	.02	.01	.00	.00

Table 1: Table of total variation distances after 1 to 10 riffle shuffles for three scenarios. The meaning and validity of these numbers is discussed in the text.

they are. It would be too simplistic to expect a probablistic technique useful for proving normal approximation to provide a way around a $\#P$-complete problem.

We find a way around using computations justified by heuristic arguments. The normal approximation result helps in two ways — it suggests a probablistic approach to finding the descent polynomial given a pair of decks and our computations of the descent polynomial are guided by the form of the bell curve. We state some of our results and discuss the validity of our computations in the second part of this introduction.

Apart from computational complexity theory and probability theory, the analysis of the mixing time for games such as blackjack and bridge is also connected to the theory of descents. The definition of descents of permutations is given in the next section and the central role of descents in the analysis of riffle shuffles will become clear. The systematic study of descents was begun by MacMahon [22]. More recent references on this topic are Foata and Schützenberger [12], Gessel and Reutenauer [13], and Knuth [20].

1.2 Theorems, experiments, and card games

Each of the three lines of Table 1 corresponds to a different scenario. In the first scenario, which was considered by Bayer and Diaconis [3], there are 52 distinct cards and we want all 52! permutations to be nearly equally likely. In the second scenario (blackjack), the distinction between the suits is ignored and the source deck is assumed to be four aces on top of four 2s, on top of four 3s, and so on. We want all $52!/4!^{13}$ possible permutations to be nearly equally likely. In the third scenario (bridge), all 52 cards are distinct but they are dealt to four players in cyclic order, and we only want the partition of the 52 cards to the four players to be random.

The total variation distances in Table 1 are a measure of how well mixed the deck is after a certain number of riffle shuffles. After 7 riffle shuffles the total variation distance is 0.334 for the Bayer-Diaconis scenario. The total variation distance is lesser after only 5 riffle shuffles for blackjack and after only 4 riffle shuffles for bridge.

As already mentioned, we had to resort to careful computations and heuristic arguments to determine some of the numbers in Table 1. How reliable are those numbers? How rigorous are the methods used to find them? We now answer these questions.

Each number in the first line of Table 1 is given by a formula derived by Bayer and Diaconis [3]. The formula is quite simple to implement and the rounding errors in finding the numbers in the first line can be bounded easily. We are happy to accept the numbers in that line as theorems.

The numbers in the second line of Table 1 are about blackjack. We prove that each blackjack

number has an error less than .001 with a probability greater than 96%, or an error less than .01 with a probability greater than 99.9996%.

The proofs of those error estimates are valid only with an assumption, however. The numbers are generated after permuting the blackjack deck randomly 10 million times, and the proofs of the error estimates assume those permutations to be independent of each other. The issue is whether the computer generated pseudorandom numbers match the idealizations found in probability theory.

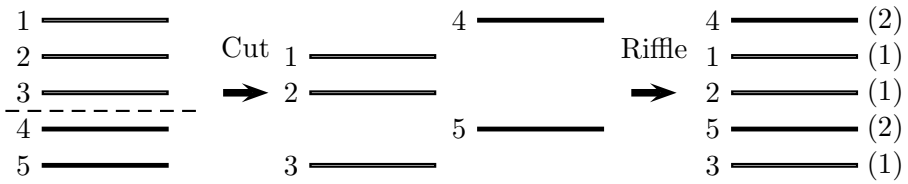
Pseudorandom numbers have been studied and used extensively [19]. A specific example is given by the coupling from the past construction by Propp and Wilson [24]. They prove that a sequence of random steps taken in a particular way is guaranteed to terminate at a state of the Ising model with a particular probability distribution. Implementations of that method, however, make do with pseudorandom numbers. The situation with our error estimates is similar.

Finally, what of the bridge numbers found in the third line of Table 1? We prove no error estimates for these numbers. While the blackjack problem fits into a subclass of pairs of decks for which we can find the descent polynomial quickly [7], the bridge problem does not. Considering that the problem of finding the descent polynomial given any pair of decks is $\#P$-complete and that "counting problems that can be solved exactly in polynomial time are few and far between" (see [16]), one begins to suspect that theorems that assert error bounds for the bridge numbers might be out of reach for a long time.

The bridge numbers in Table 1 were obtained using careful computations guided by heuristic arguments. These numbers have the same type of validity as the numbers produced in experimental physics and chemistry. The method for producing the bridge numbers has been checked for internal consistency in a variety of ways in Section 7, and the method is reported in enough detail to permit others to reproduce our results. The numbers are open to refutation, a quality of experimental results that has sometimes been emphasized [23].

2 A model of riffle shuffles

The figure below shows a riffle shuffle of a deck of 5 cards numbered 1 through 5.



The first step in a riffle shuffle is the *cut*. In the picture above, the deck is cut below the third card to get two packets with 3 and 2 cards, respectively. The second step is the *riffle*. The packets are riffled by repeatedly dropping cards from the bottom of one of the two packets until a new shuffled deck is obtained. In the picture above, the second card to be dropped and the last card to be dropped are from the second packet and all others are from the first, as indicated by the parenthesized numbers on the right.

In a riffle shuffle, the cut can be placed anywhere in the deck and the two packets can be riffled together in different ways. To model riffle shuffles, it is necessary to assign probabilities to the various ways of riffle shuffling a deck of n cards.

In the model we use, a random riffle shuffle of a deck of n cards is obtained by first generating a sequence of n independent random numbers each of which is either 1 or 2 with probability 1/2.

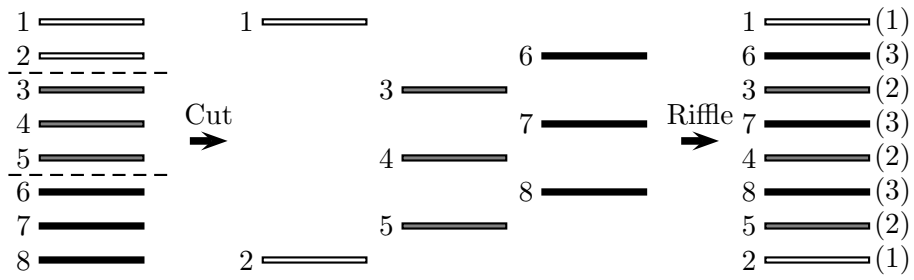


Figure 1: A 3-shuffle of 8 cards. The numbers in parentheses near the right edge of the figure give the packet the card comes from.

For the riffle shuffle depicted above, the corresponding sequence is given in parenthesis on the right. The cut must be placed such that the number of cards in the first packet equals the number of 1s in the sequence, and the number in the second packet equals the number of 2s. During the riffle, a card must be dropped from the first or second packet to get the ith card in the shuffled deck according as the ith number in the sequence is 1 or 2. The riffle shuffle depicted above results from the sequence $2, 1, 1, 2, 1$.

This model can be extended to assign probabilities to *a-shuffles*, in which the deck is cut into a packets, for any positive integer a. Figure 1 depicts a 3-shuffle of 8 cards. The extension is as follows. First, consider a sequence of n random numbers each of which is uniformly distributed over the set $\{1, 2, \ldots, a\}$ and independent of the $n-1$ other random numbers. When the deck of n cards is cut into a packets, the number of cards in the pth packet must be equal to the number of ps in the random sequence. If the number in the ith position of the random sequence is p, then the card that ends up in that position in the shuffled deck must be dropped from the pth packet. The random sequence that corresponds to the 3-shuffle depicted in Figure 1 is $1, 3, 2, 3, 2, 3, 2, 1$. From here onwards, the phrases *riffle shuffle* and *2-shuffle* are used interchangeably.

In this model of the a-shuffle, the pth packet will be empty if p does not appear at all in the random sequence. If every instance of p precedes every instance of $p+1$ in the random sequence, then all cards from the $p+1$st packet must be dropped before any card is dropped from the pth packet. An a-shuffle has $a-1$ cuts, and when some packets are empty, some of these cuts fall between the same positions.

According to Bayer and Diaconis [3], this model was described by Gilbert and Shannon in 1955 and independently by Reeds in 1971. It is also described by Epstein [11], who assumes $a = 2$ and calls it the amateur shuffle. It is sometimes called the GSR-model, after three of the people who proposed it. Many aspects of this model became clear only with the work of Bayer and Diaconis.

One of the descriptions of this model of the 2-shuffle given by Bayer and Diaconis [3] is as follows. A deck of n cards is cut after the kth card with probability $\frac{1}{2^n}\binom{n}{k}$; in other words, the cut is binomial. During the riffle, if the first packet has a cards and the second packet has b cards, the next card to be dropped is from the bottom of the first packet with probability $a/(a+b)$, and from the bottom of the second packet with probability $b/(a+b)$.

Every a-shuffle is a rearrangement of n cards and can therefore be thought of as a permutation π of $\{1, 2, \ldots, n\}$. If the ith card ends up in the jth position, then $\pi(i) = j$. The number of *descents* of a permutation turns out to be an essential concept and we will explain it in three different ways.

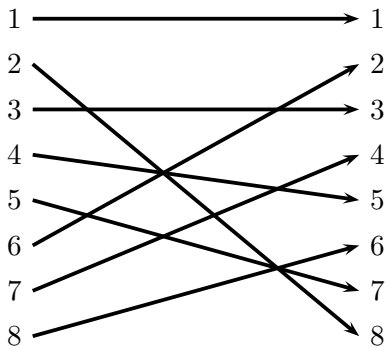


Figure 2: Depiction of the permutation $\pi(1), \pi(2), \ldots, \pi(8) = 1, 8, 3, 5, 7, 2, 4, 6$, which has 2 descents. Arrows that originate at i and $i+1$ on the left intersect if $i = 2$ or $i = 5$. If this permutation is to be realized as an a-shuffle, cuts must be placed in the two positions shown in Figure 1.

- Firstly, the number of descents of π is equal to the number of positions i, $1 \leq i \leq n-1$, with $\pi(i) > \pi(i+1)$. If we walk along the sequence $\pi(1), \pi(2), \ldots, \pi(n)$ from beginning to end, a descent must be recorded every time there is a decrease in value.

- For the second explanation, the permutation π must be represented as shown in Figure 2. The numbers 1 to n are listed twice with the listing on the left corresponding to positions of the unshuffled deck, or the *source deck*, and that on the right to positions of the shuffled deck, or the *target deck*. If $\pi(i) = j$, an arrow is drawn that originates at the i on the left and terminates at the j on the right. The number of descents is equal to the number of pairs of arrows that originate at consecutive positions i and $i+1$ and cross each other.

- For the third explanation, we realize the permutation π as an a-shuffle for some a. If π is depicted as shown in Figure 2, each packet is a block of contiguous cards on the left, and the arrows coming out of the same packet or block may not cross each other. A cut must be placed wherever two arrows that originate at consecutive positions cross each other. Therefore the minimum number of cuts necessary to realize π as an a-shuffle is equal to the number of descents in π, and a equals the number of cuts plus 1.

Bayer and Diaconis [3] proved that the probability that an a-shuffle results in a permutation π, with $\mathrm{des}(\pi) = d$, is given by

$$\frac{1}{a^n}\binom{a+n-d-1}{n}. \tag{2.1}$$

What we need to understand is the effect of repeated 2-shuffles of a deck. The above formula alone is not enough. The result stating that an a-shuffle followed by a b-shuffle is equivalent to an ab-shuffle, proved by Aldous [1], Bayer and Diaconis [3], and Reeds, is also needed. With that result, it follows immediately that k 2-shuffles are equivalent to a single 2^k-shuffle. The probability that k 2-shuffles result in a permutation π can then be found using (2.1). In addition, (2.1) tells us that the transition probability can take on only n different values corresponding to $d = 0, 1, \ldots, n-1$ for a fixed.

3 Decks with repeated cards and the descent polynomial

In this paper, the term *deck* refers to an ordered sequence of cards. Let D be a deck with cards labeled $1, 2, \ldots, h$. If the number of cards labeled 1 is n_1, the number labeled 2 is n_2, and so on, the total number of cards in D is $n = n_1 + n_2 + \cdots + n_h$. For example, the deck $D = 1, 2, 1, 2, 1, 1$ has $n_1 = 4$, $n_2 = 2$, and $n = 6$. The deck

$$D = \underbrace{1 \ldots 1}_{n_1} \underbrace{2 \ldots 2}_{n_2}, \ldots, \underbrace{h \ldots h}_{n_h},$$

which we will abbreviate as $1^{n_1}, 2^{n_2}, \ldots, h^{n_h}$, has n_1 cards labeled 1 above n_2 cards labeled 2 and so on. The deck $D = (1, 2)^{n_0}$ has $n = 2n_0$ cards with cards labeled 1 and 2 alternating. Given D, $D(i)$ denotes the label of the card in the ith position in D. For example, if $D = 1, 2, 1, 2, 1, 1$ then $D(1) = 1$ and $D(4) = 2$.

Normally, the cards of a deck will be listed from left to right, with the label of the topmost card appearing first in the list, and with the labels separated by commas, as in the previous paragraph. However, sometimes the commas will be omitted.

If a permutation π is applied to a deck $D = e_1, \ldots, e_n$, it sends the card e_i in position i to position $\pi(i)$. Therefore the resulting deck is $e_{\pi^{-1}(1)}, e_{\pi^{-1}(2)}, \ldots, e_{\pi^{-1}(n)}$, where π^{-1} is the inverse of the permutation π.

Let D_1 and D_2 be decks of n cards, n_c of which are labeled c for $1 \le c \le h$. We say that a permutation π of $\{1, 2, \ldots, n\}$ belongs to the set of permutations from D_1 to D_2 if $D_1(i) = D_2(\pi(i))$ for $1 \le i \le n$. That set will be denoted by $\Pi(D_1; D_2)$. It includes all the permutations which when applied to D_1 result in D_2 and only those. For example, if $D_1 = 1, 1, 2, 2$ to $D_2 = 1, 2, 2, 1$, $\Pi(D_1; D_2)$ has 4 members, given by $\pi(1), \pi(2), \pi(3), \pi(4)$ equal to $1, 4, 2, 3$ or $1, 4, 3, 2$ or $4, 1, 2, 3$ or $4, 1, 3, 2$. In general the cardinality of $\Pi(D_1; D_2)$ is $n_1! n_2! \ldots n_h!$.

The descent polynomial of $\Pi(D_1; D_2)$, the set of permutations from D_1 to D_2, is defined as $\sum_\pi x^{\mathrm{des}(\pi)}$, where π ranges over the set $\Pi(D_1; D_2)$ and $\mathrm{des}(\pi)$ is the number of descents of π. Let the descent polynomial be

$$c_0 + c_1 x + \cdots + c_{n-1} x^{n-1}. \tag{3.1}$$

The coefficient c_d equals the number of permutations in $\Pi(D_1; D_2)$ with d descents. The probability that an a-shuffle of D_1 results in the deck D_2 is therefore given by

$$p_a = \sum_{d=0}^{n-1} \frac{c_d}{a^n} \binom{a + n - d - 1}{n}, \tag{3.2}$$

a formula obtained by using (2.1) and summing over all the permutations in $\Pi(D_1; D_2)$. Setting $a = 1, 2, \ldots, n$ gives a triangular system of equations for $p_1, p_2, \ldots, p_n$ in terms of the coefficients $c_0, c_1, \ldots, c_{n-1}$. This triangular system can be inverted using a binomial identity (see [15, p. 269]) to get

$$c_d = p_{d+1}(d+1)^n - p_d d^n \binom{n+1}{1} + p_{d-1}(d-1)^n \binom{n+1}{2} - \cdots + (-1)^d p_1 1^n \binom{n+1}{d}, \tag{3.3}$$

for $1 \le d < n$. Using (3.2) and (3.3), it is easy to pass back and forth between the transition probabilities p_a and the descent polynomial (3.1).

Suppose we are given a source deck D_1 and asked to determine how many riffle shuffles mix that deck. We will examine a definition of mixing later on, but surely we need to be able to determine the transition probability from D_1 to any rearrangement of its cards under an a-shuffle. Since it is easy to pass back and forth between the transition probabilities and the descent polynomial, we need to be able to determine the descent polynomial of the permutations from D_1 to any rearrangement of it. This begs the question, given decks D_1 and D_2 is it possible to compute the descent polynomial (3.1) of permutations from D_1 to D_2 efficiently? To answer this question, we take a trip through computational complexity theory.

4 An excursion to computational complexity theory

We begin with a decision problem seemingly unrelated to our concerns:

THREE DIMENSIONAL MATCHING (3DM): Finite sets $X = \{x_1, \ldots, x_m\}$, $Y = \{y_1, \ldots, y_m\}$, and $Z = \{z_1, \ldots, z_m\}$ of equal cardinality are given. A subset T of $X \times Y \times Z$ is also given. Decide if there exists a subset M of T of cardinality m such that every element of X occurs exactly once as the first element of a triple in M, every element of Y occurs exactly once as the second element of a triple in M, and every element of Z occurs exactly once as the third element of a triple in M.

3DM belongs to the class of decision problems known as NP. Karp [18] showed a way to reduce every problem in NP to 3DM in time polynomial in problem size and thus proved 3DM to be NP-complete. Karp proved the NP-completeness of a number of other decision problems as well; his paper is a cornerstone of computational complexity theory. The P $\neq$ NP conjecture implies that there is no algorithm for 3DM whose running time is polynomial in m. Finding a polynomial time algorithm or a super-polynomial lower bound for a single NP-complete problem suffices to resolve this conjecture since all NP-complete problems can be reduced to each other in polynomial time. Kozen's book [21] has a good introduction to some of the basic topics of computational complexity such as NP-completeness and #P-completeness.

The following decision problem is related to card shuffling:

MIN CUTS: Given two decks, D_1 and D_2, and a positive integer d, decide if there exists a permutation $\pi \in \Pi(D_1; D_2)$ with d or fewer descents.

To determine the transition probability from D_1 to D_2 under an a-shuffle using (3.2), we need to know the coefficients c_i of the descent polynomial (3.1). MIN CUTS asks if one of the coefficients c_i, $0 \leq i \leq d$, is nonzero for a given d.

Another decision problem related to card shuffling is the following:

RIFFLE: Given p nonempty packets of cards $P_1, P_2, \ldots, P_p$ and a deck D, decide if it is possible to riffle the packets and get D.

Each packet of cards P_i is a set of cards with labels placed one above the other. We may also call it a deck, although we do not. All the cards in all the packets will appear somewhere in the new deck. RIFFLE can also be worded in terms of sequences and subsequences. RIFFLE has a solution if and only if each packet P_i corresponds to a subsequence of D, such that each packet equals the corresponding subsequence and every card of D appears in exactly one subsequence.

Both RIFFLE and MIN CUTS are NP-complete. To prove that, we give polynomial time

many-one reductions from 3DM to RIFFLE and from RIFFLE to MIN CUTS.

3DM reduces to RIFFLE. Suppose we are given an instance of 3DM. It can be reduced to RIFFLE as follows. The elements x_i, y_i, z_i of X, Y, Z become card labels and we assume X, Y, Z to be pairwise disjoint with no loss of generality. For every triple $(x_i, y_j, z_k) \in T$ create the packet x_i, y_j, z_k, L, where L is a new card label. For example, if $(x3, y1, z2) \in T$, we create the packet $x3, y1, z2, L$. If there are t triples in T, we create t packets totally in the instance of RIFFLE. The deck D for this instance of RIFFLE is given by

$$D = x_1, \ldots, x_m, y_1, \ldots, y_m, z_1, \ldots, z_m, L^m, x_1^{\alpha_1}, \ldots, x_m^{\alpha_m}, y_1^{\beta_1}, \ldots, y_m^{\beta_m}, z_1^{\gamma_1}, \ldots, z_m^{\gamma_m}, L^{m^*},$$

where α_i, β_i, and γ_i are one less than the number of occurrences of x_i, y_i, and z_i as elements of triples in T , respectively, but if the number of occurrences is zero those numbers must also be zero. Additionally, $m^* = \max(0, t - m)$, where t is the number of triples in T. We claim that the given instance of 3DM has a matching M if and only if the t packets in the created instance of RIFFLE can be riffled to get D.

Suppose the matching M is a solution of 3DM and $(x_i, y_j, z_k) \in M$. Then the card labeled L in the packet x_i, y_j, z_k, L must be dropped to get one of the m Ls that occur as the first block of Ls in D, and the cards x_i, y_j, z_k must be dropped from this packet to get the cards with those labels that precede the first block of Ls in D. If the triple (x_i, y_j, z_k) belongs to T but not to M, the card labeled L in the packet x_i, y_j, z_k, L must be dropped to get one of the m^* Ls that occurs at the very bottom of D and the cards x_i, y_j, z_k must be dropped from this packet to get cards with those labels that follow the first block of Ls in D. We can riffle the packets in this way to get D. Suppose the instance of RIFFLE that was created from the given instance of 3DM has a solution. Consider the m packets whose Ls are dropped to get the m Ls that occur as the first block of Ls in D. The triples that correspond to these packets must form a matching M of 3DM. The proof of validity of the reduction from 3DM to RIFFLE is now complete.

The reduction from 3DM to RIFFLE uses cards with $3m + 1$ different labels. What if cards are not allowed to have arbitrarily many labels? We modify the reduction to use only cards with the four labels [,], c, and L. The modification is to replace each x_i that occurs in a packet or in the deck D by the list of cards $[c^i]$; each y_i by $[c^{m+i}]$; and each z_i by $[c^{2m+i}]$. The opening and closing brackets [and] that occur in D and in each of the packets are properly matched with no nesting. Therefore, in any solution of RIFFLE, if a packet drops a [to get the very last or bottommost [that occurs in D, that dropped [must be the bottommost [in that packet and the same packet must have dropped the matching] to get the bottommost] in D. By induction, we conclude that matching parentheses [and] in a packet must be dropped to get matching parentheses in D. Therefore the list of cards that code for x_i or y_i or z_i in any packet must be dropped all at once to get a contiguous sequence of cards in D. The argument for the validity of the earlier reduction can now be applied to its modification. A little thought will convince the reader that cards labeled L in this reduction can be replaced by cards labeled c. We have proved that RIFFLE is NP-complete even if the cards are allowed to have only three different labels.

RIFFLE reduces to MIN CUTS. RIFFLE can be reduced to MIN CUTS as follows. Given an instance $(P_1, \ldots, P_p; D)$ of RIFFLE, consider the decks

$$D_1 = P_1 L P_2 L \ldots P_p$$
$$D_2 = D L^{p-1},$$

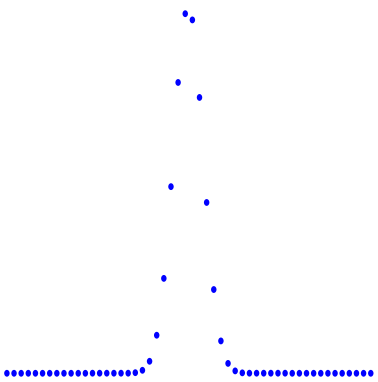


Figure 3: A graph of the coefficients of a descent polynomial looks like the bell curve.

where L is a new label that does not occur in the given instance of RIFFLE. The instance of MIN CUTS uses these decks D_1 and D_2 and chooses $d = p - 1$. We claim that RIFFLE has a solution if and only if there exists a permutation in $\Pi(D_1; D_2)$ with $p - 1$ descents. For a proof, it suffices to note that any permutation in $\Pi(D_1; D_2)$ must cut D_1 after each of the $p - 1$ Ls in D_1 as all the Ls in D_2 occur at the bottom. This reduction uses only a single new label. We have now proved the following theorem.

Theorem 4.1. *MIN CUTS is NP-complete. MIN CUTS remains NP-complete even if cards are allowed only four different labels.*

The problem of finding the descent polynomial (3.1) of $\Pi(D_1; D_2)$ is the counting version of MIN CUTS since the coefficient c_d equals the number of permutations in $\Pi(D_1; D_2)$ with d descents. Once the definition of #P-completeness in terms of Turing machines is understood, it will be clear that the proof of Theorem 4.1, with slight modifications, implies that finding the descent polynomial is #P-complete even if only four different labels are allowed. We conjecture that finding the descent polynomial is #P-complete even if D_1 and D_2 have cards with only two different labels.

If all cards in D_1 are distinct, there can be only one permutation π in $\Pi(D_1; D_2)$ and the descent polynomial can be easily found. There are other interesting cases where the descent polynomial can be found efficiently. If either D_1 or D_2 is a deck where all cards with the same label occur in a sequence of contiguous positions, there is an efficient algorithm for finding the descent polynomial. See [7]. There may be yet more cases.

There is an algorithm of time complexity $O(n^{2d})$ for determining the coefficient of x^d in the descent polynomial of permutations from a given deck D_1 to another given deck D_2. A detailed description of this algorithm will appear in the first author's Ph.D. thesis. As the running time of this algorithm is exponential in d, its use to determine the mixing times for games like blackjack and bridge is impractical on today's computers.

5 A brief introduction to Stein's method of auxiliary randomization

As the task of finding the descent polynomial is #P-complete and therefore unlikely to have an efficient algorithm, it is necessary to find a way to approximate the descent polynomial.

To approximate the descent polynomial, we turn to Stein's method of auxiliary randomization. Stein's method can be used to approximate some probability distributions by the normal distribution or the Poisson distribution or the binomial distribution.

Graphs of coefficients of most descent polynomials look like bell curves, as shown in Figure 3 and as proved in the next section. Although the normal approximations give us a good sense for what these polynomials look like, the approximations are not accurate for the first few coefficients of the descent polynomial. We will need those coefficients with good relative accuracy, and in Section 7, we obtain usable approximations for those coefficients based on heuristic arguments. The methods of Section 7 are guided by normal approximation results.

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables with $\mathrm{P}(X_i = 0) = \mathrm{P}(X_i = 1) = 1/2$. By elementary probability theory, the distribution of the sum $S = X_1 + X_2 + \cdots + X_n$ is binomial and can be approximated by the standard normal distribution after suitable normalization. We use this simple example to give a brief introduction to Stein's method [26] [27]. The introduction is broken down into three steps.

Characterization of the standard normal distribution. The standard normal distribution has the probability density function $(1/\sqrt{2\pi})\exp(-x^2/2)$. Another characterization would be in terms of its moments. There are ways to prove central limit theorems using these characterizations of the normal distribution, but neither is used by Stein's method. The characterization of the normal distribution used by Stein's method is given in the lemma below.

Lemma 5.1. *A random variable W has the standard normal distribution if and only if the expectations $\mathrm{E}(Wf(W))$ and $\mathrm{E}(f'(W))$ are equal for all bounded continuous functions f whose first and second derivatives are bounded and piecewise continuous.*

The proof in one direction is a simple use of integration by parts. The proof in the other direction is also elementary, but more involved. Stein [27, chapter 2] has given a proof of a much refined version of this lemma.

For the sum S, $\mu = \mathrm{E}S = n/2$ and $\sigma^2 = \mathrm{Var}(S) = n/4$. We want to show that the distribution of $T = (S - \mu)/\sqrt{\sigma}$ is close to normal. The expectations $\mathrm{E}(Tf(T))$ and $\mathrm{E}(f'(T))$, for functions f as in Lemma 5.1, will not be exactly equal, since the distribution of T is not exactly normal, but if it is possible to show that the expectations are close to one another then there might be a way to show that the distribution of T is close to normal. But how to show that the expectations $\mathrm{E}(Tf(T))$ and $\mathrm{E}(f'(T))$ are close? The answer is to proceed by throwing some extra randomness into the problem and then use the familiar Taylor series expansion.

Auxiliary randomization. There are many ways to add extra randomness with a view to showing the expectations $\mathrm{E}(Tf(T))$ and $\mathrm{E}(f'(T))$ to be close to each other. The way described here is related to size biasing.

To explain size biasing, we consider a random variable W with a continuous density function $p(x)$ and finite expectation. Assume that $W \geq 0$ or $p(x) = 0$ for $x < 0$. Then $\mathrm{E}W = \int_0^\infty xp(x)dx$. Therefore $xp(x)/\mathrm{E}W$ is also a probability density function. A variable with that density is said to

be W-size biased. More generally, W^* is said to have the W-size biased distribution if $\mathrm{E}(Wf(W)) = \mathrm{E}W\mathrm{E}(f(W^*))$ for all continuous f for which the expectation on the left hand side exists. The size biased distribution is defined only for non-negative random variables. If W is a 0-1 valued random variable which is 1 with a positive probability, its size biased distribution assigns probability 1 to the value 1.

The size biasing of sums such as S above is facilitated by the following lemma [2].

Lemma 5.2. *Let $W = X_1 + X_2 + \cdots + X_n$ where the X_i are all 0-1 valued random variables, but not necessarily independent or identically distributed. Let I be a random variable which is independent of the X_i and which satisfies $\mathrm{P}(I = i) = \mathrm{E}X_i/\mathrm{E}W$. Let $X_1^*, X_2^*, \ldots, X_n^*$ be a sequence of random variables such that $X_I^* = 1$ and*

$$\mathrm{P}\big((X_1^*, X_2^*, \ldots, X_n^*) \in A \big| I = i\big) = \mathrm{P}\big((X_1, X_2, \ldots, X_n) \in A \big| X_i = 1\big),$$

for all possible sets A. Then $W^* = X_1^* + X_2^* + \cdots + X_n^*$ has the W-size biased distribution

Proof. The sequence of equalities below proves the lemma. To justify the second equality below, note the assumption about the conditional distributions of the X_i^* and the X_i in the lemma. For the fourth equality below, note that $\mathrm{E}X_i = \mathrm{P}(X_i = 1)$ since X_i is 0-1 valued.

$$\begin{aligned}
\mathrm{E}f(W^*) &= \sum_{i=1}^n \mathrm{E}\big(f(W^*)\big|I = i\big)\mathrm{P}\big(I = i\big) \\
&= \sum_{i=1}^n \mathrm{E}\big(f(W)\big|X_i = 1\big)\mathrm{P}\big(I = i\big) \\
&= \sum_{i=1}^n \mathrm{E}\big(f(W)\big|X_i = 1\big)\mathrm{E}X_i/\mathrm{E}W \\
&= \frac{1}{\mathrm{E}W}\sum_{i=1}^n \mathrm{E}\big(f(W)\big|X_i = 1\big)\mathrm{P}\big(X_i = 1\big) \\
&= \frac{1}{\mathrm{E}W}\sum_{i=1}^n \mathrm{E}\big(X_i f(W)\big) = \frac{\mathrm{E}(Wf(W))}{\mathrm{E}W}.
\end{aligned}$$

□

Recall that the random variable S was defined to be a sum of independent and identically distributed random variables $X_1, \ldots, X_n$. Let I be independent of the X_i and uniformly distributed over the set $\{1, 2, \ldots, n\}$. Define $X_i^* = X_i$ if $i \neq I$ and $X_I^* = 1$. Now the hypotheses of Lemma 5.2 are easily verified and we may assert that the random variable $S^* = X_1^* + \cdots + X_n^*$ has the S-size biased distribution.

The random variable S^* was built up using the X_i and some extra randomness, namely the random variable I. Taylor series expansion must be used to see how having S^* around helps in proving the distribution of S to be normal.

Taylor series expansion. Let W be a non-negative random variable. Let W^* have the W-size biased distribution. Assume $\mu = \mathrm{E}W$ and $\sigma^2 = \mathrm{Var}(W) = \mathrm{E}(W - \mu)^2$. We want to compare the

distribution of $(W-\mu)/\sigma$ with the standard normal distribution. Consider the following calculation.

$$\begin{aligned}
\mathrm{E}\left(\left(\frac{W-\mu}{\sigma}\right)f\left(\frac{W-\mu}{\sigma}\right)\right) &= \mathrm{E}\left(\frac{W}{\sigma}f\left(\frac{W-\mu}{\sigma}\right)\right) - \frac{\mu}{\sigma}\mathrm{E}f\left(\frac{W-\mu}{\sigma}\right) \\
&= \frac{\mu}{\sigma}\left(\mathrm{E}f\left(\frac{W^*-\mu}{\sigma}\right) - \mathrm{E}f\left(\frac{W-\mu}{\sigma}\right)\right) \\
&= \frac{\mu}{\sigma}\mathrm{E}\left(\frac{(W^*-W)}{\sigma}f'\left(\frac{W-\mu}{\sigma}\right)\right) \\
&\quad + \frac{\mu}{\sigma}\mathrm{E}\left(\frac{(W^*-W)^2}{\sigma^2}f''\left(\frac{W-\mu}{\sigma}+\left(\frac{W^*-W}{\sigma}\right)U\right)(1-U)\right),
\end{aligned}$$

where U is uniformly distributed over $[0,1]$ and independent of all other random variables. In the last line above, the remainder term from Taylor's theorem, which is usually written as an integral, is written as an expectation by using U. A simple calculation shows that $\mathrm{E}(W^*-W) = (\mathrm{E}W^2/\mathrm{E}W) - EW = \sigma^2/\mu$. This calculation makes it plausible that the first term on the right hand side of the last equality above could be close to $\mathrm{E}f'((W-\mu)/\sigma)$. The second term will be small only if W^*-W is small. This is an important requirement. Not only must W^* have the W-size biased distribution, the joint distribution of W and W^* must be such that W^* is close to W.

Estimates carried out after Taylor series expansion imply the following theorem. It is due to Goldstein [14]. The size biased version of Stein's method was derived by Baldi, Rinott and Stein [2].

Theorem 5.3. *Let W be a non-negative random variable with $\mathrm{E}W = \mu$ and $\mathrm{Var}(W) = \sigma^2$. Let W^* be jointly defined with W such that its distribution is W-size biased. Let $|W^*-W| \le B$ and let $A = B/\sigma$. Let $B \le \sigma^{3/2}/\sqrt{6\mu}$. Then*

$$\left|\mathrm{P}\left(\frac{W-\mu}{\sigma} \le x\right) - \Phi(x)\right| \le 0.4A + \frac{\mu}{\sigma}(64A^2+4A^3) + \frac{23\mu}{\sigma^2}\sqrt{\mathrm{Var}\big(\mathrm{E}(W^*-W|W)\big)},$$

where Φ is the standard normal distribution.

Using Theorem 5.3, we can prove that the distribution of $(S-\mu)/\sigma$ is close to the normal distribution. We state again that $\mu = \mathrm{E}S = n/2$ and $\sigma^2 = \mathrm{Var}(S) = n/4$. From the construction of S^*, $|S^*-S| \le B$ for $B=1$. Given S, $S^*-S = 1$ if $I = i$ and $X_i = 0$. This happens with probability $(n-S)/n$. Otherwise, $S^*-S = 0$. Therefore, $\mathrm{E}\big(S^*-S\big|S\big) = 1-\frac{S}{n}$ and $\mathrm{Var}\big(\mathrm{E}\big(S^*-S\big|S\big)\big) = 1/(4n)$. By Theorem 5.3,

$$\left|\mathrm{P}\left(\frac{S-\mu}{\sigma} \le x\right) - \Phi(x)\right| \le \frac{C}{\sqrt{n}},$$

for some constant C.

6 Normal approximations for descent polynomials

A number of normal approximation theorems for descents and inversions of permutations of sets and multisets can be proved using the size biased version of Stein's method [6]. In this section,

we obtain a normal approximation for the coefficients of the descent polynomial (3.1) in a quite general setting. We use the technique developed in [6].

Let π be a uniformly distributed permutation in $\Pi(D_1; D_2)$, where D_2 is obtained by rearranging the cards of D_1 in some order. Let the decks have n_c cards with label c for $1 \le c \le h$ and let the total number of cards be n. The random variables $X_1, X_2, \ldots, X_{n-1}$ are defined as follows: $X_i = 1$ if $\pi(i) > \pi(i+1)$ but $X_i = 0$ otherwise. Let $W = \mathrm{des}(\pi) = X_1 + X_2 + \cdots + X_{n-1}$. Then

$$\mathrm{P}(W = d) = \frac{c_d}{n_1! n_2! \ldots n_h!}, \tag{6.1}$$

where c_d is the coefficient of the x^d term of the descent polynomial (3.1). If an approximation for the distribution of W is available, (6.1) can be used to approximate c_d.

We will construct W^* so that its distribution is W-size biased. Other constructions of this type can be found in [6]. Let I be a random variable independent of π with $\mathrm{P}(I = i) = \mathrm{E}X_i/\mathrm{E}W$ for $1 \le i \le n-1$. We assume $\mathrm{E}W > 0$ so that I is well defined. If c and e are two distinct card labels it is useful to define the following set:

$$S(c, e) = \big\{(x, y) \big| x > y, D_2(x) = c, D_2(y) = e\big\}, \tag{6.2}$$

where $D(i)$, as noted earlier, stands for the label of the ith card in the deck D. Note $\mathrm{E}X_i = \big|S(c,e)\big|/(n_c n_e)$ if $D_1(i) = c$, $D_1(i+1) = e$ and $c \ne e$, but $\mathrm{E}X_i = 1/2$ if $D_1(i) = D_1(i+1)$.

Given π and I, we define a permutation π^*, with $\pi^* \in \Pi(D_1; D_2)$. If $\pi(I) > \pi(I+1)$, then $\pi^* = \pi$. If $\pi(I) < \pi(I+1)$ and $D_1(I) = D_1(I+1)$, then $\pi^*(I) = \pi(I+1)$, $\pi^*(I+1) = \pi(I)$, and $\pi^*(i) = \pi(i)$ for $i \ne I, I+1$. The remaining case is $\pi(I) < \pi(I+1)$, $D_1(I) = c$, $D_1(I+1) = e$, and $c \ne e$. Let J be a random variable independent of π and I, and uniformly distributed over the set $S(c,e)$ of (6.2). Let $J = (x, y)$. Consider the list $\pi(1), \pi(2), \ldots, \pi(n-1)$. Exchange $\pi(I)$ and x and exchange $\pi(I+1)$ and y to get a new list. The permutation π^* from D_1 to D_2 is defined by setting $\pi^*(1), \pi^*(2), \ldots, \pi^*(n-1)$ equal to this new list.

There is another way to describe π^* in the last case above, which is $\pi(I) < \pi(I+1)$, $D_1(I) = c$, $D_1(I+1) = e$, and $c \ne e$. Define

$$S_\pi(c, e) = \big\{(k, l) \big| (\pi(k), \pi(l)) \in S(c, e)\big\}. \tag{6.3}$$

We can pick (k, l) uniformly from this set, exchange $\pi(I)$ with $\pi(k)$, and exchange $\pi(I+1)$ with $\pi(l)$ to get π^*. The set over which (k, l) is distributed depends upon π and I, but the distribution is always uniform.

Lemma 6.1. $\mathrm{P}\big(\pi^* \in A \big| I = i\big) = \mathrm{P}\big(\pi \in A \big| \pi(i) > \pi(i+1)\big)$ *for any possible set* A.

Proof. If $I = i$ and $\pi(i) > \pi(i+1)$, then $\pi^* = \pi$. Therefore it suffices to show that

$$\mathrm{P}\big(\pi^* \in A \big| I = i, \pi(i) < \pi(i+1)\big) = \mathrm{P}\big(\pi \in A \big| \pi(i) > \pi(i+1)\big).$$

The first case is when $D_1(i) = D_1(i+1)$. In this case, we have

$$\begin{aligned}
&\mathrm{P}\big(\pi^* \in A \big| I = i, \pi(i) < \pi(i+1)\big) \\
=&\mathrm{P}\big(\pi(1), \ldots, \pi(i-1), \pi(i+1), \pi(i), \pi(i+2), \ldots, \pi(n-1) \in A \big| I = i, \pi(i) < \pi(i+1)\big) \\
=&\mathrm{P}\big(\pi(1), \ldots, \pi(i-1), \pi(i+1), \pi(i), \pi(i+2), \ldots, \pi(n-1) \in A \big| \pi(i) < \pi(i+1)\big) \\
=&\mathrm{P}\big(\pi \in A \big| \pi(i) > \pi(i+1)\big)
\end{aligned}$$

The last equality above holds because if we pick a uniformly distributed permutation in $\Pi(D_1; D_2)$ and exchange $\pi(i)$ and $\pi(i+1)$, the new permutation is also a uniformly distributed permutation in $\Pi(D_1; D_2)$ provided $D_1(i) = D_1(i+1)$.

The second case is when $D_1(i) = c$, $D_1(i+1) = e$, and $c \neq e$. Below each of the summations is taken over $(x, y) \in S(c, d)$.

$$\begin{aligned}
&\mathrm{P}\big(\pi^* \in A \big| I = i, \pi(i) < \pi(i+1)\big) \\
=&\sum \mathrm{P}\big(\pi^* \in A \big| I = i, \pi(i) < \pi(i+1), J = (x,y)\big)\mathrm{P}\big(J = (x,y)\big) \\
=&\sum \mathrm{P}\big(\pi^* \in A \big| I = i, \pi(i) < \pi(i+1), J = (x,y)\big)\mathrm{P}\big(\pi(i) = x, \pi(i+1) = y \big| \pi(i) > \pi(i+1)\big) \\
=&\sum \mathrm{P}(\pi \in A \big| \pi(i) = x, \pi(i+1) = y)\mathrm{P}\big(\pi(i) = x, \pi(i+1) = y \big| \pi(i) > \pi(i+1)\big) \\
=&\mathrm{P}\big(\pi \in A \big| \pi(i) > \pi(i+1)\big).
\end{aligned}$$

The second equality above holds because J is uniformly distributed over the set $S(c, d)$ of (6.2) and because π is a uniformly distributed permutation in $\Pi(D_1; D_2)$. The third equality above holds because π is a uniformly distributed permutation in $\Pi(D_1; D_2)$ and because of the way π^* is generated using π, I, and J. □

For $i = 1, 2, \ldots, n-1$, define $X_i^* = 1$ if $\pi^*(i) > \pi^*(i+1)$, but $X_i^* = 0$ otherwise. Let $W^* = \mathrm{des}(\pi^*) = X_1^* + X_2^* + \cdots + X_{n-1}^*$.

Lemma 6.2. *W^* has the W-size biased distribution.*

Proof. Follows from Lemmas 5.2 and 6.1. □

If Theorem 5.3 is to be applied to approximate the distribution of W, it is necessary to find an upper bound for $\mathrm{Var}\big(\mathrm{E}(W^* - W|W)\big)$. By [4, p. 477], $\mathrm{Var}\big(\mathrm{E}(W^* - W|W)\big) \leq \mathrm{Var}\big(\mathrm{E}(W^* - W|\pi)\big)$, since W is a function of π. Let

$$Q = \mathrm{E}(W^* - W|\pi). \tag{6.4}$$

We will upper bound $\mathrm{Var}(Q)$.

The identity

$$\mathrm{Var}\Big(\sum_{i=1}^{n} Y_i\Big) = \left(\sum_{i=1}^{n} \mathrm{Var}(Y_i) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Covar}(Y_i, Y_j)\right) \tag{6.5}$$

holds because $\mathrm{Var}(Y) = \mathrm{E}Y^2 - \big(\mathrm{E}Y\big)^2$ and $\mathrm{Covar}(Y, Z) = \mathrm{E}YZ - \mathrm{E}Y\mathrm{E}Z$. Along with (6.5), the following Lemma 6.3 is useful for upper bounding $\mathrm{Var}(Q)$.

The argument to upper bound $\mathrm{Var}(Q)$ simplifies a great deal if it is assumed that cards with any label occur the same number of times in D_1 or D_2; in other words, $n_1 = n_2 = \cdots = n_h = n_0$ with $n_0 \geq 1$. We make this assumption from here onwards and put it to use immediately in the lemma below. With this assumption, the total number of cards is $n = hn_0$.

In the lemma below, as will become evident from its proof, the constants 7 and 10 can be replaced by other positive integers. We take the constants as 7 and 10 to facilitate later use of the lemma. The constants were chosen to simplify the exposition and are not the best possible.

Lemma 6.3. *Let π be a uniformly distributed permutation in $\Pi(D_1; D_2)$. Let $f(\pi)$ depend only upon the relative order of $\pi(i_1), \pi(i_2), \ldots, \pi(i_r)$. Let $g(\pi)$ depend only upon the relative order of $\pi(j_1), \pi(j_2), \ldots, \pi(j_s)$. Assume $|f| \leq 7$, $|g| \leq 7$, $r \leq 10$, and $s \leq 10$. There are three cases:*

1. $\{i_1, \ldots, i_r\} \cap \{j_1, \ldots, j_s\} \neq \phi$,

2. $\{i_1, \ldots, i_r\} \cap \{j_1, \ldots, j_s\} = \phi$, *but* $D_1(i_\rho) = D_1(j_\psi)$ *for some* $1 \leq \rho \leq r$ *and* $1 \leq \psi \leq s$,

3. $\{i_1, \ldots, i_r\} \cap \{j_1, \ldots, j_s\} = \phi$, *and* $D_1(i_\rho) \neq D_1(j_\psi)$ *for any* $1 \leq \rho \leq r$ *and* $1 \leq \psi \leq s$.

In these three cases, we have

1. $\mathrm{Covar}(f, g) \leq C$ *for some constant* C *that depends on neither* n_0 *nor* h,

2. $\mathrm{Covar}(f, g) \leq C/n_0$ *for some constant* C *that depends on neither* n_0 *nor* h,

3. $\mathrm{Covar}(f, g) = 0$,

respectively.

Proof. For the first case, it is enough to note that f and g are bounded in magnitude. For the third case, it is enough to note that f and g are independent of each other. The second case remains to be proved.

It is enough to consider f and g to be indicator functions that are 1 for a particular relative ordering and 0 for others. This is because any f or g that is bounded in magnitude by a constant can be written as a linear combination of a constant number of such indicator functions with coefficients that are bounded in magnitude by a constant. We show the proof assuming $f = 1$ if $\pi(i) < \pi(i+1)$ and $f = 0$ otherwise; and $g = 1$ if $\pi(j) < \pi(j+1)$ and $g = 0$ otherwise. For the second case to apply, either $i + 1 < j$ or $j + 1 < i$ must hold, and at least one of $D_1(i), D_1(i+1)$ must equal one of $D_1(j), D_1(j+1)$.

Let $P_1 = \mathrm{P}\big(\pi(i) < \pi(i+1)\big)$ and $P_2 = \mathrm{P}\big(\pi(j) < \pi(j+1)\big)$. We claim that

$$\mathrm{P}\big(\pi(j) < \pi(j+1) \big| \pi(i) = x, \pi(i+1) = y\big) = P_2 + \epsilon, \tag{6.6}$$

with $|\epsilon| < 4/n_0$. In (6.6), $1 \leq x, y \leq n$, $x \neq y$, $D_2(x) = D_1(i)$, and $D_2(y) = D_1(i+1)$. The claim is true because of the following argument. If $D_1(j) = D_1(j+1)$, the pair $(\pi(j), \pi(j+1))$ can take $n_0(n_0 - 1)$ different values. Otherwise, it can take n_0^2 values. For some of these values, $\pi(j) < \pi(j+1)$. Given $\pi(i) = x$ and $\pi(i+1) = y$, at least one of x or y is not allowed to appear in a possible value for $(\pi(j), \pi(j+1))$. Thus at most $2n_0$ possible values for this pair must be excluded.

The proof for this f and g can be completed by noting that $\mathrm{Covar}(f, g)$ equals

$$\mathrm{P}\big(\pi(i) < \pi(i+1), \pi(j) < \pi(j+1)\big) - \mathrm{P}\big(\pi(i) < \pi(i+1)\big)\mathrm{P}\big(\pi(j) < \pi(j+1)\big),$$

and by writing the first of the three probabilities in the line above in terms of the conditional probabilities on the left hand side of (6.6). The proof for more general f and g is similar. □

We now go back to Q defined by (6.4) and the construction of the size biased random variable W^*. Let us suppose that π is a given permutation in $\Pi(D_1; D_2)$. To write Q as a sum, we introduce the quantities $\chi_\pi(i, i+1)$ and $\chi_\pi(i, i+1, k, l)$, where $k \neq l$. The first of these is defined as the change in the number of descents when $\pi(i)$ and $\pi(i+1)$ are exchanged. The second is defined as the change in the number of descents when $\pi(i)$ is exchanged with $\pi(k)$ and $\pi(i+1)$ is exchanged with $\pi(l)$. With a view to applying Lemma 6.3 later on, we note that $\chi_\pi(i, i+1)$ depends only on the relative order of at most 4 numbers, namely $\pi(i-1), \pi(i), \pi(i+1), \pi(i+2)$. Similarly,

$\chi_\pi(i, i+1, k, l)$ depends only on the relative order of at most 10 numbers of the form $\pi(i)$. The magnitude of both of these quantities is always bounded by 7.

From here onwards, we denote $\mathrm{E}W$ and $\mathrm{Var}(W)$ by μ and σ^2, respectively.

Let A be the set of values of i for which $\pi(i) < \pi(i+1)$ and $D_1(i) = D_1(i+1)$. Let B be the set of values of i for which $\pi(i) < \pi(i+1)$, $D_1(i) \neq D_1(i+1)$, and $S(D_1(i), D_1(i+1)) \neq \phi$. We must have $A \cap B = \phi$. By the construction of W^* using π, I and J, if $i \in A$, we have

$$\mathrm{P}(I = i) = \frac{1}{2\mu} \quad \text{and} \quad \mathrm{E}\big(W^* - W \big| \pi, I = i\big) = \chi_\pi(i, i+1). \tag{6.7}$$

If $i \in B$, let $D_1(i) = c$ and $D_1(i+1) = e$. We have,

$$\mathrm{P}(I = i) = \frac{\big|S(c,e)\big|}{n_0^2} \frac{1}{\mu} \tag{6.8}$$

and

$$\begin{aligned} \mathrm{E}\big(W^* - W \big| \pi, I = i\big) &= \sum \mathrm{E}\big(W^* - W \big| \pi, I = i, J = (x,y)\big) \mathrm{P}(J = (x,y)) \\ &= \sum \frac{\chi_\pi(i, i+1, k, l)}{\big|S(c,e)\big|}, \end{aligned} \tag{6.9}$$

where the first summation is over all (x, y) in the set $S(c, e)$ of (6.2) and the second summation is over all (k, l) in the set $S_\pi(c, e)$ of (6.3). If $i \notin A$ and $i \notin B$, either $P(I = i) = 0$ or $\mathrm{E}\big(W^* - W \big| \pi, I = i\big) = 0$.

By (6.4), (6.7), (6.8) and (6.9), we have

$$Q = \sum_{i=1}^{n-1} \mathrm{E}\big(W^* - W \big| \pi, I = i\big) \mathrm{P}\big(I = i\big)$$

and

$$\mu Q = \sum_{i \in A} \frac{\chi_\pi(i, i+1)}{2} + \sum_{i \in B \text{ and } (k,l) \in S_\pi(D_1(i), D_1(i+1))} \frac{\chi_\pi(i, i+1, k, l)}{n_0^2}. \tag{6.10}$$

We will use (6.10) with (6.5) to upper bound $\mathrm{Var}(\mu Q)$.

If (6.5) and (6.10) are used to write $\mathrm{Var}(\mu Q)$ as a sum of variance and covariance terms, each term on the right hand side of (6.10) will contribute a single variance term. The total contribution of these variance terms will be bounded by Cn, for some constant C, because each χ_π is bounded in magnitude by 7.

Some of the covariance terms will be of the form

$$\mathrm{Covar}\left(\frac{\chi_\pi(i_1, i_1+1, k_1, l_1)}{n_0^2}, \frac{\chi_\pi(i_2, i_2+1, k_2, l_2)}{n_0^2} \right), \tag{6.11}$$

where $i_1 \in B$ and $i_2 \in B$. The value of $\chi_\pi(i_1, i_1+1, k_1, l_1)$ depends only upon the relative order of the $\pi(i)$, with i equal to i_1 or $i_1 + 1$ or k_1 or l_1, or differing from one of those 4 integers by at most 1. There can be at most 10 such values for i and we will denote this set of values by $\mathrm{nghd}\{i_1, i_1+1, k_1, l_1\}$. As for the magnitude of the covariance term (6.11), there are three cases.

1. Suppose $\text{nghd}\{i_1, i_1+1, k_1, l_1\} \cap \text{nghd}\{i_2, i_2+1, k_2, l_2\} \neq \phi$. This corresponds to the first case of Lemma 6.3. Therefore the magnitude of the covariance term (6.11) is bounded by C/n_0^4 in this case.

   To count the number of covariance terms of this type, note that i_1 can take at most $n-1$ different values. As we require $D_1(k_1) = D_1(i_1)$ and $D_1(l_1) = D_1(i_1+1)$, for given i_1, there are at most n_0^2 possible values of (k_1, l_1). For the covariance term (6.11) to fall under this type, at least one of i_2, k_2, l_2 must differ from one of i_1, k_1, l_1 by less than 2. Therefore, given i_1, k_1, l_1, there are at most a constant number of choices for one of i_2, k_2, l_2. Having chosen one of i_2, k_2, l_2, there are at most n_0^2 ways to choose the other two. For example, suppose k_2 has been chosen. This restricts i_2 to at most n_0 possibilities since we require $D_1(i_2) = D_1(k_2)$. Given i_2, there are at most n_0 possibilities for l_2 since we require $D_1(l_2) = D_1(i_2+1)$. Thus the number of covariance terms of this type is bounded by Cnn_0^4 for some constant C.

2. Suppose $\text{nghd}\{i_1, i_1+1, k_1, l_1\} \cap \text{nghd}\{i_2, i_2+1, k_2, l_2\} = \phi$, but $D_1(i_1^*) = D_1(i_2^*)$ for some $i_1^* \in \text{nghd}\{i_1, i_1+1, k_1, l_1\}$ and some $i_2^* \in \text{nghd}\{i_2, i_2+1, k_2, l_2\}$. This corresponds to the second case of Lemma 6.3. Therefore the magnitude of the covariance term (6.11) is bounded by C/n_0^5 for some constant C in this case.

   The number of such covariance terms is bounded by Cnn_0^5. The argument is the same as that in the previous case except for one difference. Given i_1, k_1, l_1, in this case, we require one of i_2, k_2, l_2 to differ by 2 or less from some position i^* such that identical cards occur at i^* and at one of the positions in $\text{nghd}\{i_1, i_1+1, k_1, l_1\}$ in the deck D_1. Therefore, there are at most Cn_0 possibilities for one of i_2, k_2, l_2 and not just a constant number of possibilities as in the previous case.

3. Suppose $\text{nghd}\{i_1, i_1+1, k_1, l_1\} \cap \text{nghd}\{i_2, i_2+1, k_2, l_2\} = \phi$, and $D_1(i_1^*) \neq D_1(i_2^*)$ for any $i_1^* \in \text{nghd}\{i_1, i_1+1, k_1, l_1\}$ and any $i_2^* \in \text{nghd}\{i_2, i_2+1, k_2, l_2\}$. This corresponds to the third case of Lemma 6.3. Therefore the covariance term (6.11) is 0 in this case.

Consequently, the total contribution of covariance terms of the form (6.11) to $\text{Var}(\mu Q)$ is bounded by Cn for some positive constant C. Apart from (6.11), covariance terms can also be of the form

$$\text{Covar}\left(\chi_\pi(i_1, i_1+1), \frac{\chi_\pi(i_2, i_2+1, k, l)}{n_0^2}\right),$$

with $i_1 \in A$ and $i_2 \in B$, or

$$\text{Covar}\big(\chi_\pi(i_1, i_1+1), \chi_\pi(i_2, i_2+1)\big),$$

with $i_1 \in A$ and $i_2 \in A$. The proof that the total contribution of such terms to $\text{Var}(\mu Q)$ is also bounded by Cn is similar to and simpler than the case that has already been dealt with. The bound on $\text{Var}(\mu Q)$ is stated as a lemma below.

Lemma 6.4. *$\text{Var}(\mu Q) < Cn$ for some positive constant C.*

Theorem 6.5. *Let D_1 be a deck of cards with h different labels with each label occurring n_0 times. Let D_2 be another deck with the same $n = hn_0$ cards in a different order. Let π be a uniformly distributed permutation in $\Pi(D_1; D_2)$. Let the random variable W be the number of descents of π.*

Let $\mu = \mathrm{E}W$ and $\sigma^2 = \mathrm{Var}(W)$. Assume $\sigma > 7^{2/3}(6\mu)^{1/3}$. Then

$$\left| \mathrm{P}\Big(\frac{W-\mu}{\sigma} \le x\Big) - \Phi(x) \right| \le C\Big(\frac{1}{\sigma} + \frac{\mu}{\sigma^3} + \frac{\mu}{\sigma^4} + \frac{\sqrt{n}}{\sigma^2}\Big),$$

for some positive constant C and $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-u^2/2)du$.

Proof. By Lemma 6.2, W^* has the W-size biased distribution. By construction of W^*, $|W^* - W| \le 7$. By [4, p. 477], $\mathrm{Var}\big(\mathrm{E}\big(W^* - W \big| W\big)\big) \le \mathrm{Var}(Q)$ with Q defined by (6.4). The proof can be completed using Theorem 5.3, if Lemma 6.4 is used to note that $\mathrm{Var}(Q) \le Cn/\mu^2$. □

Consider the atypical case where $D_1 = (1,2)^{n_0}$ and $D_2 = 1^{n_0}, 2^{n_0}$. Then every permutation in $\Pi(D_1; D_2)$ has exactly $n_0 - 1$ descents, with each descent corresponding to an occurrence of 2 immediately before a 1 in the deck D_1. In such a case $\sigma^2 = 0$, but typically σ^2 will be of the order of n. As $\mu < n$ always, in such cases, the normal approximation to W given by Theorem 6.5 will have $O\big(n^{-1/2}\big)$ error. The computation of μ and σ given D_1 and D_2 will be described presently.

Calculating the mean and the variance of W. The expectation $\mathrm{E}W$ can be computed by summing $\mathrm{E}X_i$ over $1 \le i \le n-1$. If $D_1(i) = a$ and $D_1(i+1) = b$, the expectation $\mathrm{E}X_i$ is 0.5 if $a = b$. If $a \ne b$,

$$\mathrm{E}X_i = \frac{R(a,b)}{N(a)N(b)},$$

where $N(a)$ and $N(b)$ are the number of cards with labels a and b in the deck D_2, and $R(a,b) = \sum_{1 \le l_b < l_a \le n} \chi(l_a, l_b)$ with $\chi(l_a, l_b) = 1$ if $D_2(l_a) = a$ and $D_2(l_b) = b$ and $\chi(l_a, l_b) = 0$ otherwise.

The variance $\mathrm{Var}(W)$ can be obtained from $\mathrm{E}(X_iX_j)$ for $1 \le i < j \le n-1$. The computation of these joint expectations involves many cases. First suppose that $j > i+1$. Denote $D_1(i)$, $D_1(i+1)$, $D_1(j)$, and $D_1(j+1)$ by a, b, c, and d, respectively. If $a = b$, then $\mathrm{E}(X_iX_j) = \mathrm{E}X_j/2$. Likewise if $c = d$, $\mathrm{E}(X_iX_j) = \mathrm{E}X_i/2$. There are seven cases when $a \ne b$ and $c \ne d$. One of these is when $a = d$ and a, b, c are distinct. In that case,

$$\mathrm{E}(X_iX_j) = \frac{R(a,b)R(c,a) - R(c,a,b)}{N(a)N(b)N(c)(N(a)-1)},$$

where $R(c,a,b) = \sum_{1 \le l_c < l_a < l_b \le n-1} \chi(l_c, l_a, l_b)$ with $\chi(l_c, l_a, l_c) = 1$ if $D_2(l_c) = c$, $D_2(l_a) = a$ and $D_2(l_b) = b$ but $\chi(l_c, l_a, l_c) = 0$ otherwise. The other cases are handled similarly. The $j = i+1$ case is also handled similarly.

7 Shuffling cards for blackjack and bridge

The rules for blackjack vary with the gambling house, but the distinction between the four suits is always ignored. Most of the time all the face cards are equivalent to cards with the number 10, but there are some obscure situations where 10s, jacks, queens, and kings must all be considered distinct. We take a blackjack source deck to be any permutation of the multiset $\{1^4, \ldots, 13^4\}$. We consider two of these — $1^4, \ldots, 13^4$ and $(1, \ldots, 13)^4$ — which are notable for their symmetry. These two will be called *Blackjack1* and *Blackjack2*, respectively. Let p_i be the transition probability from one of these source decks to the ith possible ordering of that deck after a certain number of riffle shuffles. Ideally, we would like all of these p_i to be equal.

The situation for bridge is different. All the 52 cards in the source deck are distinct but there are only four players. Each player must be dealt a random set of 13 cards but the order in which a player receives his cards is immaterial.

Suppose the cards are dealt to players $\mathcal{N}$, $\mathcal{E}$, $\mathcal{S}$, and $\mathcal{W}$ in cyclic order, as is the common practice. Let the deck which needs to be dealt to the four players be $1, 2, \ldots, 52$. Let the sets $\mathcal{P}_\mathcal{N}$, $\mathcal{P}_\mathcal{E}$, $\mathcal{P}_\mathcal{S}$, and $\mathcal{P}_\mathcal{W}$ be a partition of $\{1, 2, \ldots, 52\}$ with the cardinality of each set being 13. There are $52!/13!^4$ such partitions. Ideally, we would like the probability that $\mathcal{N}$, $\mathcal{E}$, $\mathcal{S}$, and $\mathcal{W}$ receive cards in $\mathcal{P}_\mathcal{N}$, $\mathcal{P}_\mathcal{E}$, $\mathcal{P}_\mathcal{S}$, and $\mathcal{P}_\mathcal{W}$, respectively, to be equal for all those partitions.

Suppose that the probability that $\mathcal{N}$, $\mathcal{E}$, $\mathcal{S}$, and $\mathcal{W}$ receive cards in $\mathcal{P}_\mathcal{N}$, $\mathcal{P}_\mathcal{E}$, $\mathcal{P}_\mathcal{S}$, and $\mathcal{P}_\mathcal{W}$, respectively, is equal to $p_\mathcal{P}$, when the deck $1, 2, \ldots, 52$ is a-shuffled and then dealt to $\mathcal{N}$, $\mathcal{E}$, $\mathcal{S}$, and $\mathcal{W}$ is cyclic order. Let the set of permutations of $\{1, 2, \ldots, 52\}$ that result in such a deal be $S_\mathcal{P}$. Now consider the deck D with 52 cards such that if i belongs to $\mathcal{P}_\mathcal{N}$, $\mathcal{P}_\mathcal{E}$, $\mathcal{P}_\mathcal{S}$, or $\mathcal{P}_\mathcal{W}$, the ith card of D is $\mathcal{N}$, $\mathcal{E}$, $\mathcal{S}$, or $\mathcal{W}$, respectively. Then the set of permutations $S_\mathcal{P}$ equals $\Pi(D; (\mathcal{NESW})^{13})$. Therefore, $p_\mathcal{P}$ is equal to the probability that an a-shuffle of D results in $(\mathcal{NESW})^{13}$. Ideally, we would like these transition probabilities to be equal for all $52!/13!^4$ possible decks D.

In both situations, we have N probabilities p_i, $1 \le i \le N$, which sum to 1, and ideally we would like all of them to be $1/N$. Unfortunately, that would require an infinite number of riffle shuffles. We have to determine how close the probabilities are to the uniform distribution after a certain number of riffle shuffles, and then decide how close is close enough.

There are many ways to measure the closeness between probability distributions, but the notion of closeness must be chosen with care. Some metrics such as the Euclidean distance or the L^2 norm are inappropriate as the following example shows. Suppose one probability distribution always picks the first out of N possibilities and another always picks the second. The Euclidean distance between these two distributions is $\sqrt{2}$. On the other hand, suppose a distribution picks one of the even numbered possibilities with equal likelihood and another distribution picks one of the odd possibilities with equal likelihood. The Euclidean distance between the third and fourth distributions is $2/\sqrt{N} + O(1/N)$, which incorrectly suggests that these two distributions are much closer to each other for large N.

7.1 A table of total variation distances

The notion of closeness we use here is the *total variation distance*. The total variation distance from the probability distribution defined by the p_i to the uniform distribution is given by

$$\sum_{i=1}^{N} \Big(\frac{1}{N} - p_i\Big)^+, \tag{7.1}$$

where $x^+ = x$ if $x \ge 0$ and $x^+ = 0$ if $x < 0$. This total variation distance always lies between 0 and 1. It also has a probabilistic meaning — if $P_1(A)$ is the probability of a certain subset A of the N possibilities under the distribution given by the p_i and if $P_2(A)$ is the probability of that set under the uniform distribution, then the total variation distance equals the maximum of $|P_1(A) - P_2(A)|$ over all possible A.

Table 2 gives the total variation distances from the uniform distribution for nine scenarios termed *BayerDiaconis*, *Blackjack1*, *Blackjack2*, *Bridge1*, *Bridge2*, *RedBlack1*, *RedBlack2*, *Alice-Bob1*, and *AliceBob2*. That table allows the number of riffle shuffles (i.e., 2-shuffles) to vary from 1 to 10. In five of the scenarios, the source deck is fixed and the target deck is allowed to vary.

	1	2	3	4	5	6	7	8	9	10
BayerDiaconis	**1**	**1**	**1**	**1**	**.924**	**.614**	**.334**	**.167**	**.085**	**.043**
Blackjack1	**1**	**1**	**1**	**.481**	**.215** /.23	**.105** /.11	**.052** /.05	**.026** /.03	**.013** /.01	**.007** /.00
Blackjack2					.60	.32	.16	.08	.04	.02
Bridge1	**1**	**1**	**1**	**.990**	**.748**/.75	**.423**/.42	**.218**/.21	**.110**/.11	**.055**/.05	**.027**/.03
Bridge2			.45?	.16	.08	.04	.02	.01	.00	.00
RedBlack1	**.580**	**.360**	**.208**	**.105**	**.052**/.05	**.026**/.03	**.013**/.01	**.007**/01	**.003**/.00	**.002**/.00
RedBlack2					.10	.03	.01	.01	.00	.00
AliceBob1	**1**	**1**	**.999**	**.725**	**.308** /.31	**.130**/.13	**.059**/.06	**.028**/.03	**.013**/.01	**.007**/.01
AliceBob2					.02	.01	.00	.00	.00	.00

Table 2: Table of total variation distances after 1 to 10 riffle shuffles for nine scenarios from BayerDiaconis to AliceBob2. The numbers in boldface have an error less that .01 with a probability greater than 99.9996%. The numbers that are not in boldface were determined using a less accurate and heuristic method. Entries of the table with a number in boldface and in ordinary type separated by a slash can be used to form an idea of the reliability of the less accurate method.

These are *BayerDiaconis*, where the source deck is fixed as $D_1 = 1, 2, \ldots, 52$; *Blackjack1* with $D_1 = 1^4, 2^4, \ldots, 13^4$; *Blackjack2* with $D_2 = (1, 2, \ldots, 13)^4$; *RedBlack1* with $D_1 = R^{26}B^{26}$; and *RedBlack2* with $D_1 = (RB)^{26}$. In the other four scenarios the target deck is fixed. These are *Bridge1*, where the target deck is fixed as $D_2 = \mathcal{N}^{13}\mathcal{E}^{13}\mathcal{S}^{13}\mathcal{W}^{13}$; *Bridge2* with $D_2 = (\mathcal{NESW})^{13}$; *AliceBob1* with $D_2 = A^{26}B^{26}$; and *AliceBob2* with $D_2 = (AB)^{26}$.

How were the total variation distances shown in Table 2 determined? For *BayerDiaconis* there is a simple and elegant formula for the total variation distance after a certain number of riffle shuffles that was derived by Bayer and Diaconis [3]. Some of the other numbers require more extensive computation. Two numbers for *Bridge2*, with the number of riffle shuffles being 3 or 4, were determined using more than 10000 hours of CPU time on the Teragrid computer network. From the point of view of card players, those two numbers are perhaps the most interesting results of this paper. First, we explain how the numbers in the rows of Table 2, other than the first *BayerDiaconis* row, were computed, and why they can be trusted.

7.2 Monte Carlo estimation of total variation distance

The determination of the numbers in Table 2 is impeded by two problems. The first of these is that the summation in (7.1) has so many terms that it is impractical to use (7.1) to find the total variation distances shown in Table 2.

This first problem is easy to overcome. Denote the sum in (7.1) by S. Let X_1 be a random variable that is equal to $\bigl(1/N - p_i\bigr)^+$, the ith term in (7.1), with probability $1/N$ for $1 \le i \le N$. Such a random variable can be easily generated *if* we can determine the transition probabilities p_i efficiently. We have $\mathrm{E}X_1 = S/N$ and $\mathrm{Var}(X_1) \le 1/N^2$. Let $X_1, \ldots, X_k$ be independent and identically distributed, and let $Y_k = (X_1 + \cdots + X_k)N/k$. Then

$$\mathrm{E}Y_k = S. \tag{7.2}$$

The theorem below tells us how good an estimate of S can be obtained from a single instance of the random variable Y_k.

Theorem 7.1. *For $\alpha > 0$ and $k > 2$,*

$$\mathrm{P}\Bigl(|Y_k - S| \ge \frac{\alpha}{\sqrt{k}}\Bigr) < \frac{4}{\alpha^4}.$$

Proof. Consider the following calculation.

$$\begin{aligned}
\mathrm{E}(|Y_k - S|^4) &= \mathrm{E}\left(\left|\Bigl(X_1 - \frac{S}{N}\Bigr) + \cdots + \Bigl(X_k - \frac{S}{N}\Bigr)\right|^4\right).\frac{N^4}{k^4} \\
&= \frac{N^4}{k^4}\sum_{i=1}^{k}\mathrm{E}\Bigl(X_i - \frac{S}{N}\Bigr)^4 + \frac{6N^4}{k^4}\sum_{1\le i<j\le k}\mathrm{E}\Bigl(X_i - \frac{S}{N}\Bigr)^2\Bigl(X_j - \frac{S}{N}\Bigr)^2 \\
&\le (1/k^3 + 3/k^2) < 4/k^2
\end{aligned}$$

The second equality above follows from $\mathrm{E}(X_i - S/N) = 0$ and from the independence of X_i and X_j for $i \ne j$. To deduce the first inequality in the last line above, note that X_i has the range $[0, 1/N]$ with $\mathrm{E}X_i = S/N$. The proof can be completed using Markov's inequality [4]. □

In some special cases that include *Blackjack1*, *Bridge1*, *RedBlack1*, and *AliceBob1*, we have derived efficient polynomial time algorithms for finding the descent polynomials [7]. Those algorithms and (3.2) can be used to determine the transition probabilities efficiently. The numbers given for these four cases in Table 2 in **boldface** were obtained using $k = 10^7$ random permutations of the source deck or the target deck, exact computation of the transition probabilities, and Monte Carlo summation (7.2). Theorem 7.1 with $k = 10^7$ and $\alpha = \sqrt{10}$ implies that the boldface numbers have errors less than .001 with probability greater than 96%. Theorem 7.1 with $k = 10^7$ and $\alpha = 10\sqrt{10}$ implies that the boldface numbers have errors less than .01 with probability greater than 99.9996%.

The more significant problem is that there may be no efficient way to determine the transition probabilities p_i. Indeed there is very probably no efficient method for determining these transition probabilities in general, as proved in Section 4.

7.3 Approximation of the descent polynomial

In the other four cases — *Blackjack2*, *Bridge2*, *RedBlack2*, and *AliceBob2* — we turn to the Monte Carlo method once again to approximate the descent polynomials. Suppose we are given decks D_1 and D_2 and it is required to approximate the descent polynomial of permutations in $\Pi(D_1; D_2)$. If the decks D_1 and D_2 have n_c cards with label c for $1 \le c \le h$, then the total number of permutations in $\Pi(D_1; D_2)$ is $n_1! \dots n_h!$. We generate l random permutations $\pi_1, \dots, \pi_l$ from this collection and form the polynomial

$$P = \sum_{i=1}^{l} x^{\mathrm{des}(\pi_i)} = \gamma_0 + \gamma_1 x + \cdots + \gamma_{n-1} x^{n-1}. \tag{7.3}$$

The coefficient γ_d counts the number of random permutations with d descents. The approximation to the descent polynomial is taken to be $\frac{n_1! \dots n_h!}{l} P$; in other words, the polynomial P given by (7.3) is normalized to get an approximation to the descent polynomial.

Once the polynomial P defined by (7.3) is formed, it carries within itself an estimate of the accuracy of the coefficients of the descent polynomial obtained by normalizing P, as we will explain.

Suppose that X is a Bernoulli random variable with $\mathrm{P}(X = 1) = p$ and that p is unknown. Suppose that $X_1, \dots X_l$ are independent with the same distribution as that of X, and that in one experiment m out of these l random variables equal 1. We can estimate $p \approx m/l$, but how accurate is this estimate? Let $Y = (X_1 + \ldots + X_l)/l$. Then $\mathrm{E}Y = p$ and $\mathrm{Var}(Y) = p(1-p)/l$. Therefore the fluctuations of Y about its mean are of the order $\sqrt{p(1-p)/l}$. If we use a single instance of Y to estimate its mean, which is p, then we expect an absolute error of about $\sqrt{p(1-p)/l}$ and a relative error of about $\sqrt{(1-p)/lp}$. If we substitute $p = m/l$, we find that the relative error will be about $\sqrt{(l-m)/lm}$. If p is very small, then $m << l$ and we may expect a relative error of about $1/\sqrt{m}$.

If we define $X(\pi) = 1$ if $\mathrm{des}(\pi) = d$ and $X(\pi) = 0$ otherwise, where π is a uniformly distributed permutation in $\Pi(D_1; D_2)$, then γ_d defined by (7.3) is $X_1 + \cdots + X_l$, where $X_i = X(\pi_i)$ are independent with the same distribution as that of $X(\pi)$. Then by the argument of the preceding paragraph, the relative error in the estimate of the x^d coefficient of the descent polynomial will be about $1/\sqrt{\gamma_d}$.

To illustrate this estimate in practice, we take the target deck to be $D_2 = (\mathcal{NESW})^{13}$, which is fixed for *Bridge2*, and the source deck D_1 to be

NSEENNWEWSSWESWNNNEESSSSSESWWNNSENWSEWSWWWEENEWNNNWE.

We computed P, which is defined by (7.3), with $l = 10^9$, and got the coefficients of x^{14}, x^{15}, and x^{16} to be 17, 397, and 4560, respectively. If the descent polynomial of permutations from D_1 to D_2 is approximated as $13!^4 P/10^9$, we expect the relative errors in the coefficients of x^{14}, x^{15}, and x^{16} to be about 25%, 5%, and 1.6%, respectively. When we computed P with $l = 10^{11}$, we got the coefficients of x^{14}, x^{15}, and x^{16} to be 2334, 38418, and 468359, respectively. These numbers are 100 times the counts for $l = 10^9$, if allowance is made for the relative errors when the counts are scaled and interpreted as coefficients of the descent polynomial. We conclude that descent polynomials can be approximated using (7.3) and that the accuracy of the approximation can be gauged by looking at the coefficients of P.

By (2.1), the probability that an a-shuffle leads to a permutation with d descents is equal to

$$\frac{1}{a^n}\binom{a+n-d-1}{n}\left\langle \begin{array}{c} n \\ d \end{array} \right\rangle,$$

where $\left\langle \begin{array}{c} n \\ d \end{array} \right\rangle$ is the Eulerian number that counts the number of permutations of $1, 2, \ldots, n$ with d descents. The Eulerian numbers can be calculated using simple recurrences [15].

The probability that an a-shuffle with $a = 32$ has 16 or more descents is more than 0.95. Coefficients of terms from x^{16} to x^{36} of descent polynomials that correspond to the scenarios from *Blackjack1* to *AliceBob2* in Table 2 can be approximated well using (7.3) with $l = 10^7$. The numbers reported in Table 2 for these scenarios, with the number of riffle shuffles varying from 5 to 10, that are not in boldface were computed by approximating descent polynomials in this manner. The transition probabilities were computed using these approximate descent polynomials and (3.2), and the total variation distances were computed using (7.2) with $k = 1000$. The total variation distances computed this way are reported with two digits after the decimal point. They compare well with more accurate estimates of the total variation distances, which are found in boldface in Table 2, for *Blackjack1*, *Bridge1*, *RedBlack1*, and *AliceBob1*.

We still need to explain the computation of the total variation distances for *Bridge2* when the number of riffle shuffles is 3 or 4. Four riffle shuffles are equivalent to an a-shuffle with $a = 16$, and it is necessary to accurately compute the coefficients of terms from x^{12} to x^{16} of the descent polynomials to find the total variation distance for *Bridge2* after four riffle shuffles. Obtaining an accurate estimate for the coefficient of x^{12} using (7.3) would require an l that is beyond the reach of today's computers. We used (7.3) with $l = 10^{10}$, and with this l the coefficient of x^{15} is approximated with a relative error that is less than 5% with a probability greater than 95%. We took the logs of the coefficients of the terms from x^{15} to x^{22} and computed degree 4 polynomials that were least squares fits to these numbers. These polynomials were nearly quadratic in accord with Theorem 6.5. We got the coefficients of the x^{12}, x^{13}, and x^{14} terms by extrapolation. We feel sure that the extrapolated coefficients had relative errors smaller than 10%.

For *Bridge2* and three riffle shuffles, we got the coefficients of the x^7 terms in the descent polynomials using the coefficients of terms from x^{15} to x^{26} and polynomial fits of degree 6. We are less sure that the estimated total variation distance for this case is accurate.

8 Three open problems

1. The first of the three open problems that we mention in this section is about a more general model of riffle shuffles. In the model of riffle shuffles we have used so far a random riffle shuffle of a deck of n cards is obtained by first generating a random sequence of n numbers. The numbers in

the sequence are independent of each other and each number is either 1 or 2 with probability 1/2. The model can be changed by requiring the first number in the sequence to be either 1 or 2 with probability 1/2. Every later number in the sequence equals the preceding number with probability $1-p$ and it is of the opposite kind with probability p. This model is described by Aldous [1] and Diaconis [9].

When $p = 1/2$, we get back the GSR-model which was described in Section 2. If $p > 1/2$, then the riffle shuffles are neat — if a card is dropped from one hand the next card is likelier to be dropped from the other hand. If $p < 1/2$, the shuffling is clumsy. The problem is to determine how the mixing times depend upon p.

Any sequence of 1s and 2s where every 1 occurs before every 2 corresponds to the identity permutation. All other sequences correspond to distinct permutations. The probability of one of these permutations under this model will be $0.5 * p^k(1-p)^{n-1-k}$ if there are k places where a 2 follows a 1 or a 1 follows a 2 in the corresponding sequence. The probability of the identity permutation is also easy to determine. Suppose we want to know the probability that a given permutation π can be obtained as the composition of m riffle shuffles none of which is the identity. This number will be a polynomial in p of the form

$$\sum_{k=0}^{m(n-1)} c_k p^k (1-p)^{m(n-1)-k}$$

with the coefficients c_k depending upon π and m. A good place to start might be by asking if the coefficients c_k can be determined in polynomial time.

The value of this polynomial when $p = 1/2$ can be deduced from the work of Bayer and Diaconis [3]. When $p = 1$, each riffle shuffle is either a perfect in-shuffle or a perfect out-shuffle with probability 1/2. Much information about this case can be found in the work of Diaconis, Graham, and Kantor [10].

2. The second problem is purely combinatorial and is related to the riffle shuffles of the deck $(1,2)^n$ [7]. Consider all sequences of length $2n$ with n 1s and n 2s. Define any two sequences $D_1 = \alpha\beta\gamma$ and $D_2 = \alpha\beta^*\gamma$ to be R-related if β has the same number of 1s as 2s and if β^* is obtained from β by changing 1s to 2s and 2s to 1s. This is an equivalence relation. We conjecture that the number of equivalence classes is $(n+3)2^{n-2}$. If the equivalence relation is modified by requiring α and γ to be sequences of the same length, the number of equivalence classes is the Catalan number $\frac{1}{n+2}\binom{2n+2}{n+1}$ [7].

3. The third problem too is purely combinatorial. Consider all permutations π of the numbers $1, \ldots, nh$ such that $\pi(i) \equiv i \mod h$ for $1 \le i \le nh$. The problem is to derive a recurrence for the number of these permutations that have exactly d descents. If $h = 1$, the familiar recurrences for the Eulerian numbers (see [15]) solve this problem. A solution of this problem will make it possible to compute the transition probability from the deck $(1, 2, \ldots, h)^n$ to itself under an a-shuffle.

9 Acknowledgements

The authors thank P. Diaconis, S. Fomin, J. Fulman, C. Mulcahy, and J. Stipins for helpful discussions. In addition, the authors thank Prof. Diaconis for his work, which got them interested in this problem, and for telling them about Stein's method.

References

[1] D. Aldous. Random walks on finite groups and rapidly mixing Markov chains. *Seminar on Probability XVII, Lecture Notes in Mathematics*, 986:243–297, 1983.

[2] P. Baldi, Y. Rinott, and C. Stein. A normal approximation for the number of local maxima of a random function on a graph. In *Probability, Statistics, and Mathematics: Papers in Honor of Samuel Karlin*, pages 59–81. Academic Press, 1989.

[3] D. Bayer and P. Diaconis. Trailing the dovetail shuffle to its lair. *The Annals of Applied Probability*, 2:294–313, 1992.

[4] P. Billingsley. *Probability and Measure*. Wiley, New York, 2nd edition, 1986.

[5] E. Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rend. Circ. Mat. Palermo*, 27:247–271, 1909.

[6] M. Conger and D. Viswanath. Normal approximations for descents and inversions of permutations of sets and multisets. `www.arxiv.org`.

[7] M. Conger and D. Viswanath. Riffle shuffles of decks with repeated cards. *Annals of Probability*. To appear.

[8] P. Diaconis. *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics, USA, 1988.

[9] P. Diaconis. Mathematical developments from the analysis of riffle shuffling. *Groups, combinatorics and geometry (Durham 2001)*, pages 73–97, 2003.

[10] P. Diaconis, R. Graham, and W. Kantor. The mathematics of perfect shuffles. *Advances in Applied Mathematics*, 4:175–193, 1983.

[11] E. Epstein. *The Theory of Gambling and Statistical Logic*. Academic Press, New York, revised edition, 1977.

[12] D. Foata and M.-P. Schützenberger. *Théorie Géométrique des Polynômes Eulériens*. Springer-Verlag, New York, 1970.

[13] I.M. Gessel and C. Reutenauer. Counting permutations with given cycle structure and descent set. *Journal of Combinatorial Theory, Series A*, 64:189–215, 1993.

[14] L. Goldstein. Berry Esseen bounds for combinatorial central limit theorems and pattern occurences, using zero and size biasing. *Journal of Applied Probability*, 2005.

[15] R.L. Graham, D.E. Knuth, and O. Patashnik. *Concrete Mathematics*. Addison-Wesley, Massachusetts, 2nd edition, 1994.

[16] M. Jerrum. *Counting, sampling and integrating: algorithms and complexity*. Available on the internet, 2005. Lecture Notes from a Nachdiplomvorlesung at ETH-Zürich.

[17] M. Jerrum, A. Sinclair, and E. Vigoda. A polynomial time approximation algorithm for the permanent of a matrix with non-negative entries. *J. Assoc. Comput. Mach.*, 51:671–697, 2004.

[18] R. Karp. Reducibility among combinatorial problems. In R.E. Miller and J.W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972.

[19] D.E. Knuth. *The Art of Computer Programming, Vol. 2.* Addison-Wesley, Massachusetts, 3rd edition, 1998.

[20] D.E. Knuth. *The Art of Computer Programming, Vol. 3.* Addison-Wesley, Massachusetts, 3rd edition, 1998.

[21] D. Kozen. *The Design and Analysis of Algorithms.* Springer-Verlag, 1992.

[22] P.A. MacMahon. *Combinatory Analysis, Vol. 1.* Cambridge University Press, Cambridge, 1915.

[23] K. Popper. *Conjectures and Refutations: the Growth of Scientific Knowledge.* Routledge, London, 5 edition, 2002.

[24] J.G. Propp and D.B. Wilson. Exact sampling with coupled markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, pages 223–252, 1996.

[25] A. Rényi. *A Diary on Information Theory.* Akadémiai Kiado, Budapest, 1984.

[26] C. Stein. A bound for the error in the normal approximation of the distribution of a sum of dependent random variables. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, 2:583–602, 1972.

[27] C. Stein. *Approximate Computation of Expectations.* Institute of Mathematical Statistics, Hayward, CA, 1986.