On shuffles and their mixing times

Ayush Goyal

June 2026

Abstract

Card shuffling provides one of the classical examples of a random process whose analysis combines probability, combinatorics, and group theory. In this paper, we study card shuffling through the theory of finite Markov chains, viewing each permutation of a deck as a state in the symmetric group S_n. After developing the necessary probabilistic background, we analyze the top to random shuffle, proving its fundamental structural properties and establishing the uniform distribution as its stationary distribution. We then introduce the notion of strong stationary times and the coupon collector problem to study convergence to equilibrium. Finally, we discuss the Gilbert Shannon Reeds model of riffle shuffling and the celebrated 7 shuffles theorem, which explains why approximately seven riffle shuffles suffice to randomize a standard deck.

1 Introduction

How many shuffles are required before a deck is random? has motivated decades of research connecting probability theory and stochastic processes. Card shuffling models have found applications in games of chance, randomized algorithms, cryptography, and the study of random walks in finite groups.

A deck of n cards has $n!$ possible orderings, naturally identified with the elements of the symmetric group S_n. A shuffle therefore induces a random walk on S_n, where each shuffle specifies transition probabilities between permutations.

Among the many models of card shuffling, the *top-to-random shuffle* occupies a particularly important position. The principal mathematical result established in this paper is the following.

Theorem 1 (Informal)**.** *The top-to-random shuffle defines an irreducible and aperiodic Markov chain on the symmetric group S_n whose unique stationary distribution is the uniform distribution on all permutations. Furthermore, the chain mixes on the order of $n \log n$ shuffles*

The theorem above summarizes the main mathematical development of this paper. A precise formulation and proof are given in Sections 4 and 5.

Having developed the theory through the top-to-random shuffle, we then turn to the more realistic Gilbert Shannon Reeds model of riffle shuffling. Although its analysis is considerably more sophisticated, it culminates in one nice result in the mathematics of random processes: the Bayer Diaconis theorem, which shows that approximately seven riffle shuffles suffice to randomize a standard deck.

2 Mathematical Background

We briefly review elementary probability, finite Markov chains, and stationary distributions. Unless stated otherwise, all Markov chains considered in this paper have finite state spaces.

2.1 Basic Probability

A *sample space*, denoted by Ω, is the set of all possible outcomes of a random experiment. An *event* is any subset of Ω, and a *probability measure* assigns to each event $A \subseteq \Omega$ a number $P(A)$ satisfying

$$0 \leq P(A) \leq 1, \qquad P(\Omega) = 1$$

Definition 1. *Let A and B be events with $P(B) > 0$. The* conditional probability *of A given B is*

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Conditional probability allows probabilities to be updated when new information is known. An important consequence is the *Law of Total Probability.*

Theorem 2 (Law of Total Probability)**.** *Suppose $B_1, \ldots, B_n$ form a partition of the sample space and satisfy $P(B_i) > 0$ for each i. Then for every event A,*

$$P(A) = \sum_{i=1}^{n} P(B_i)\, P(A \mid B_i)$$

This theorem will be used repeatedly throughout the paper when analyzing transition probabilities of card-shuffling processes.

Definition 2. *A* random variable *is a function*

$$X : \Omega \to \mathbb{R},$$

which assigns a real number to each outcome of the experiment.

Definition 3. *If X is a discrete random variable taking values $x_1, x_2, \ldots$, then its* expected value *is*

$$\mathbb{E}[X] = \sum_{i} x_i P(X = x_i)$$

Expectation represents the average value obtained after repeating an experiment many times. Throughout this paper, expectation will arise naturally when discussing the average number of shuffles required to randomize a deck and the expected stopping times of certain random processes.

2.2 Finite Markov Chains

Many random processes evolve over time, with each step depending on the current state of the system. Such processes are modeled by Markov chains.

Definition 4. *A* Markov chain *is a sequence of random variables*

$$X_0, X_1, X_2, \ldots$$

taking values in a finite state space S such that

$$P(X_{n+1} = j \mid X_n = i, X_{n-1}, \ldots, X_0) = P(X_{n+1} = j \mid X_n = i)$$

for every n and all states $i, j \in S$.

This condition is known as the *Markov property*: the future evolution of the process depends only on its current state and not on the path by which that state was reached.

The transition probabilities of a Markov chain are collected into a matrix.

Definition 5. *The* transition matrix *of a finite Markov chain is the matrix*

$$P = (p_{ij}),$$

where

$$p_{ij} = P(X_{n+1} = j \mid X_n = i)$$

Each row of P sums to 1.

The entry $(P^k)_{ij}$ gives the probability of moving from state i to state j in exactly k steps.

One of the simplest examples is the simple random walk on the integers. At each step the walker moves one unit to the left or one unit to the right with equal probability. Although its definition is elementary, this process exhibits many of the phenomena that appear in more sophisticated stochastic models.

Card shuffling can also be viewed as a Markov chain. In this setting, each possible ordering of the deck represents a state, and each shuffle determines the transition probabilities between states. This viewpoint forms the mathematical foundation for the entire paper

2.3 Stationary Distributions

A central question in the study of Markov chains is their long-term behaviour.

Definition 6. *A probability distribution*

$$\pi = (\pi_i)_{i \in S}$$

is called a stationary distribution *if*

$$\pi P = \pi$$

In other words, if the initial state of the chain is distributed according to π, then the distribution remains unchanged after one transition.

Not every Markov chain possesses a unique stationary distribution. Two structural properties play an important role.

Definition 7. *A Markov chain is said to be* irreducible *if every state can be reached from every other state with positive probability after finitely many steps.*

Definition 8. *A state has period d if every return to that state occurs at times that are multiples of d, where d is the greatest common divisor of all such return times. A Markov chain is called* aperiodic *if every state has period* 1*.*

These conditions ensure that the chain does not decompose into disconnected components or become trapped in deterministic cycles.

Theorem 3. *Every finite, irreducible Markov chain possesses a unique stationary distribution. Moreover, if the chain is also aperiodic, then the distribution of the chain converges to this stationary distribution regardless of the initial state.*

This theorem provides the mathematical justification for speaking of a shuffled deck becoming 'random.' Once a card-shuffling process is modeled as a finite irreducible and aperiodic Markov chain, repeated shuffling causes the distribution of the deck to approach its unique stationary distribution.

3 Card Shuffling as a Markov Chain

Now with the basic concepts of probability and finite Markov chains, we now describe how a card-shuffling process may be modeled mathematically. The key observation is that every possible ordering of a deck can be regarded as a state of a Markov chain, while each shuffle determines probabilistic transitions between these states. Consequently, repeated shuffling corresponds to a random walk on a finite state space.

3.1 The Symmetric Group

Consider a deck consisting of n distinct cards labelled

$$1, 2, \ldots, n$$

Every ordering of these cards is a permutation of the set

$$\{1, 2, \ldots, n\}$$

The collection of all such permutations forms the *symmetric group*, denoted by

$$S_n$$

Since every permutation corresponds to exactly one arrangement of the deck, the state space of any card-shuffling process is naturally identified with S_n. As there are

$$|S_n| = n!$$

elements of the symmetric group, a deck of n cards has precisely $n!$ possible configurations.

Example 1. *For $n = 3$,*

$$S_3 = \{123, 132, 213, 231, 312, 321\},$$

where, for instance, the permutation 231 represents the deck whose top card is 2, followed by 3, then 1.

3.2 Card Shuffling as a Markov Chain

A shuffle specifies a random rule for transforming one permutation into another. Suppose that the current ordering of the deck is represented by the permutation $\sigma \in S_n$. After performing one shuffle, the deck moves to another permutation $\tau \in S_n$ according to some probability

$$P(\sigma, \tau)$$

These probabilities define a Markov chain on the finite state space S_n.

The Markov property follows naturally from the definition of a shuffle. Once the current ordering of the deck is known, the distribution of the next ordering depends only on the shuffle being performed and not on the sequence of previous shuffles that produced the current arrangement. Hence,

$$P(X_{k+1} = \tau \mid X_k = \sigma, X_{k-1}, \ldots, X_0) = P(X_{k+1} = \tau \mid X_k = \sigma),$$

where X_k denotes the configuration of the deck after k shuffles.

Thus, repeated shuffling defines a finite Markov chain whose state space is the symmetric group.

3.3 Transition Matrices

As with every finite Markov chain, the shuffle is completely described by its transition matrix.

Definition 9. *The transition matrix of a card-shuffling process is the matrix*

$$P = (P(\sigma, \tau))_{\sigma, \tau \in S_n},$$

where

$$P(\sigma, \tau) = P(X_{k+1} = \tau \mid X_k = \sigma)$$

is the probability that one shuffle transforms the deck from the ordering σ to the ordering τ.

Since the deck must move to exactly one new configuration after each shuffle,

$$\sum_{\tau \in S_n} P(\sigma, \tau) = 1$$

for every state σ.

Proposition 1. *Every row of the transition matrix sums to one.*

Proof. Fix a state $\sigma \in S_n$. After one shuffle, the deck must end in one of the $n!$ possible configurations. Since these events are mutually exclusive and exhaustive,

$$\sum_{\tau \in S_n} P(\sigma, \tau) = 1$$

Therefore every row of the transition matrix represents a probability distribution over the state space. □

Proposition 2. *For every positive integer k,*

$$(P^k)(\sigma, \tau) = P(X_k = \tau \mid X_0 = \sigma),$$

that is, the (σ, τ) entry of P^k gives the probability of moving from σ to τ in exactly k shuffles.

Proof. The proof starts by induction on k. The case $k = 1$ follows directly from the definition of the transition matrix. Assuming the statement holds for k, we compute

$$(P^{k+1})(\sigma, \tau) = \sum_{\gamma \in S_n} (P^k)(\sigma, \gamma) P(\gamma, \tau)$$

By the induction hypothesis, this is precisely the probability of reaching τ after $k + 1$ shuffles by conditioning on the intermediate state γ, which is exactly the Law of Total Probability. □

The transition matrix therefore contains all probabilistic information about the shuffling process.

4 The Top-to-Random Shuffle

Throughout this section we assume that the deck consists of n distinct cards.

4.1 Definition

The top-to-random shuffle is performed according to the following rule.

1. Remove the top card of the deck.
2. Choose one of the n possible positions uniformly at random.
3. Insert the removed card into the chosen position.

Every application of this procedure transforms one permutation of the deck into another. Consequently, the shuffle defines a Markov chain on the state space S_n.

Example 2. *Consider the deck*

$$12345.$$

Removing the top card leaves the deck

$$2345.$$

Reinserting the card uniformly at random produces one of the five configurations

$$12345,\ 21345,\ 23145,\ 23415,\ 23451,$$

each with probability $\frac{1}{n}$.

4.2 Transition Structure

For two permutations $\sigma, \tau \in S_n$, define

$$P(\sigma,\tau) = P(X_{k+1} = \tau \mid X_k = \sigma)$$

Since the removed card may be inserted into any of the n positions with equal probability,

$$P(\sigma,\tau) = \begin{cases} \frac{1}{n}, & \text{if } \tau \text{ can be obtained from } \sigma \text{ by one top-to-random shuffle,} \\ 0, & \text{otherwise.} \end{cases}$$

Thus every row of the transition matrix contains exactly n nonzero entries, each equal to $\frac{1}{n}$.

4.3 Irreducibility and Aperiodicity

We now establish two fundamental structural properties of the chain.

Theorem 4. *The top-to-random shuffle is irreducible.*

Proof. It suffices to show that every permutation can be transformed into any other permutation through finitely many top-to-random shuffles.

Observe that any card may be moved to the top by repeatedly moving the cards above it downward through successive top-to-random operations. Once a desired card is on top, it can be inserted into any position of the deck in a single shuffle.

Repeating this procedure finitely many times allows any prescribed permutation of the deck to be constructed from any initial permutation. Hence every permutation is reachable from every other permutation after finitely many shuffles, and therefore the chain is irreducible. □

Theorem 5. *The top-to-random shuffle is aperiodic.*

Proof. At every step there is probability $\frac{1}{n}$ of reinserting the top card back into the first position. In this case the deck remains unchanged.

Therefore every state has a positive probability of returning to itself in one step. Since a return in one step is possible, every state's period is equal to one, proving that the chain is aperiodic. □

4.4 The Uniform Stationary Distribution

The next theorem takes a look at the long-term equilibrium of the chain.

Theorem 6. *The uniform distribution on S_n,*

$$\pi(\sigma) = \frac{1}{n!}, \qquad \sigma \in S_n,$$

is a stationary distribution for the top-to-random shuffle.

Proof. Since every permutation has probability $\frac{1}{n!}$ under the uniform distribution,

$$(\pi P)(\tau) = \sum_{\sigma \in S_n} \pi(\sigma) P(\sigma,\tau) = \frac{1}{n!} \sum_{\sigma \in S_n} P(\sigma,\tau)$$

To evaluate the sum, observe that every permutation τ has exactly n predecessors under a single top-to-random shuffle. Each possible position of the top card before shuffling determines a unique predecessor. Each predecessor contributes probability $\frac{1}{n}$, so

$$\sum_{\sigma \in S_n} P(\sigma,\tau) = n \cdot \frac{1}{n} = 1$$

Hence

$$(\pi P)(\tau) = \frac{1}{n!} = \pi(\tau),$$

showing that the uniform distribution is stationary. □

Since the chain is finite, irreducible, and aperiodic, the Fundamental Theorem of Markov Chains implies that the uniform distribution is the unique stationary distribution and that the distribution of the deck converges to it regardless of the initial ordering.

The existence of a stationary distribution answers only part of the shuffling problem. In practice, one is interested in determining how many shuffles are required before the distribution of the deck is close to uniform.

For the top-to-random shuffle, convergence is remarkably slow compared to other common shuffling methods.To resolve this, we need better techniques

5 Strong Stationary Times

In the previous section, we showed that the top-to-random shuffle defines an irreducible and aperiodic Markov chain whose unique stationary distribution is the uniform distribution on the symmetric group S_n. While this guarantees that repeated shuffling eventually randomizes the deck, it does not quantify the rate at which convergence occurs. That is what we aim to achieve here

Unlike many approaches to convergence, strong stationary times do not compare probability distributions directly. Instead, they identify a random time at which the chain is guaranteed to have reached its stationary distribution exactly. Estimating the probability that this stopping time has not yet occurred then yields bounds on the distance from stationarity.

5.1 Stopping Times and Strong Stationary Times

We begin by recalling the notion of a stopping time.

Definition 10. *Let $(X_t)_{t\geq 0}$ be a Markov chain on a finite state space. A random variable*

$$\tau : \Omega \to \{0, 1, 2, \ldots\}$$

is called a stopping time *if, for every $t \geq 0$, the event*

$$\{\tau = t\}$$

is completely determined by the history

$$X_0, X_1, \ldots, X_t$$

In other words, whether the process stops at time t can be decided without any knowledge of future states.

Stopping times arise naturally. Examples include the first time a random walk reaches a specified state or the first time a particular event occurs. In the context of card shuffling, stopping times often describe the first time that a card reaches a certain position or the first time the deck satisfies some desired property.

The stopping times relevant to mixing theory possess a considerably stronger property.

Definition 11. *Let $(X_t)_{t\geq 0}$ be a Markov chain with stationary distribution π. A stopping time τ is called a* strong stationary time *if*

$$\Pr(X_\tau = x \mid \tau = t) = \pi(x)$$

for every state x and every t such that $\Pr(\tau = t) > 0$.

Equivalently,

$$\Pr(X_\tau = x, \tau = t) = \pi(x) \Pr(\tau = t),$$

which states that the position of the chain at the stopping time is independent of the value of the stopping time itself.

This condition is substantially stronger than merely requiring that the chain be stationary at time τ. Indeed, a strong stationary time guarantees that, regardless of whether the process stops early or late, the distribution of the state is already exactly the stationary distribution.

The importance of strong stationary times stems from the following fundamental result.

Theorem 7. *Let τ be a strong stationary time for a Markov chain with stationary distribution π. Then, for every time $t \geq 0$,*

$$\left\|P^t(x,\cdot) - \pi\right\|_{\mathrm{TV}} \leq \Pr_x(\tau > t),$$

where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation distance.

Proof. Let A be any subset of the state space. We decompose the probability of the event $\{X_t \in A\}$ according to whether the stopping time has already occurred:

$$\Pr(X_t \in A) = \Pr(X_t \in A, \tau \leq t) + \Pr(X_t \in A, \tau > t)$$

Since τ is a strong stationary time, conditioning on the event $\{\tau \leq t\}$ implies that the distribution of X_t is already stationary. Consequently,

$$\Pr(X_t \in A, \tau \leq t) = \pi(A)\Pr(\tau \leq t)$$

Therefore,

$$\Pr(X_t \in A) - \pi(A) = \Pr(X_t \in A, \tau > t) - \pi(A)\Pr(\tau > t)$$

Taking absolute values and using the fact that probabilities are bounded above by one gives

$$|\Pr(X_t \in A) - \pi(A)| \leq \Pr(\tau > t)$$

Finally, taking the supremum over all subsets A yields

$$\left\|P^t(x,\cdot) - \pi\right\|_{\mathrm{TV}} \leq \Pr_x(\tau > t),$$

as required. □

The theorem reveals the usefulness of strong stationary times. Instead of estimating the total variation distance directly, which is often difficult, it suffices to bound the probability that the stopping time has not yet occurred. Consequently, the analysis of mixing times is reduced to the study of an appropriate probabilistic stopping rule.

5.2 A Strong Stationary Time for the Top-to-Random Shuffle

Having introduced the notion of a strong stationary time, we now construct one for the top-to-random shuffle. The construction, due to Aldous and Diaconis, is remarkably simple and elegant and forms the foundation of the mixing-time analysis that follows.

Recall that the top-to-random shuffle repeatedly removes the top card of the deck and reinserts it uniformly among the n possible positions.

Let τ_{top} denote one shuffle after the first time that the original bottom card reaches the top of the deck.

The importance of this stopping time is not immediately obvious. The key observation is that as the original bottom card gradually rises through the deck, the relative ordering of the cards beneath it becomes increasingly randomized. By the time the original bottom card reaches the top, every card has been randomly inserted beneath it, completely destroying the memory of the initial ordering.

The following proposition makes this idea precise.

Proposition 3. *Suppose that after t shuffles there are k cards beneath the original bottom card. Then every one of the $k!$ possible orderings of these cards is equally likely.*

Proof. We proceed by induction on the number of shuffles.

Initially, at time $t = 0$, there are no cards beneath the original bottom card. Since there is only one possible ordering of the empty set, the claim holds trivially.

Assume that after t shuffles the proposition holds whenever there are k cards beneath the original bottom card.

Consider the next shuffle.

There are two possible outcomes.

1. The top card is inserted above the original bottom card.

   In this case, the collection of cards beneath the original bottom card is unchanged. Since no new card has been added, the inductive hypothesis implies that these cards remain uniformly ordered.

2. The top card is inserted beneath the original bottom card.

   The newly inserted card may occupy any of the $k + 1$ positions among the cards beneath the original bottom card. Since the insertion position is chosen uniformly at random, each of these positions occurs with probability
   $$\frac{1}{k+1}$$
   By the inductive hypothesis, the original k cards are already uniformly distributed among their $k!$ possible orderings.

   Therefore every one of the
   $$(k + 1) \cdot k! = (k + 1)!$$
   possible orderings of the $k + 1$ cards occurs with equal probability.

Thus the proposition holds after the $(t + 1)$st shuffle. By induction, it is true for every t. □

Theorem 8. *The stopping time τ_{top} is a strong stationary time for the top-to-random shuffle.*

Proof. Immediately before the stopping time occurs, the original bottom card occupies the top position of the deck.

By Proposition 5.1, every possible ordering of the remaining $n - 1$ cards beneath it is equally likely.

The stopping time is defined to occur after one additional shuffle. During this final shuffle, the original bottom card is removed from the top and inserted uniformly into one of the n possible positions.

For every fixed ordering of the remaining $n - 1$ cards, there are exactly n possible insertion positions for the original bottom card, producing every permutation of the deck exactly once.

Since

- each ordering of the remaining $n - 1$ cards is equally likely, and
- each insertion position is chosen with probability $1/n$,

every permutation of the deck occurs with probability
$$\frac{1}{(n-1)!} \cdot \frac{1}{n} = \frac{1}{n!}$$
Hence the distribution of the deck at time τ_{top} is uniform over the symmetric group S_n.

The event
$$\{\tau_{\text{top}} = t\}$$
depends only on the trajectory of the original bottom card up to time t, that is, on the first time at which it reaches the top of the deck. It does not depend on the ordering of the remaining $n - 1$ cards beneath it.

Conditioned on the event $\{\tau_{\text{top}} = t\}$, Proposition 5.1 shows that the remaining $n - 1$ cards are uniformly distributed over all $(n - 1)!$ possible orderings. During the final shuffle, the original bottom card is inserted independently and uniformly into one of the n possible positions.

Consequently, every permutation of the deck is obtained with probability

$$\frac{1}{(n-1)!} \cdot \frac{1}{n} = \frac{1}{n!},$$

independently of the value of t. Hence, for every permutation $\sigma \in S_n$,

$$P(X_{\tau_{\text{top}}} = \sigma, \tau_{\text{top}} = t) = P(\tau_{\text{top}} = t)\frac{1}{n!} = P(\tau_{\text{top}} = t)\pi(\sigma)$$

Therefore,

$$P(X_{\tau_{\text{top}}} = \sigma \mid \tau_{\text{top}} = t) = \pi(\sigma),$$

which is precisely the defining property of a strong stationary time. Hence τ_{top} is a strong stationary time. □

The theorem provides an exact randomization criterion that once the stopping time has occurred, the deck is not merely close to uniformly distributed, it is exactly uniform. Consequently, estimating the mixing time reduces to studying the distribution of the random variable τ_{top} itself.

5.3 The Coupon Collector Interpretation

Theorem 5.2 reduces the problem of determining the mixing time of the top-to-random shuffle to understanding the random variable τ_{top}. We now show that the distribution of this stopping time is identical to that of the classical coupon collector problem. This correspondence allows us to apply well-known probabilistic results to obtain strong quantitative bounds on the convergence of the shuffle.

5.3.1 The Motion of the Original Bottom Card

Recall that τ_{top} is defined as one shuffle after the original bottom card first reaches the top of the deck.

Suppose that at some stage of the shuffling process there are exactly k cards beneath the original bottom card, where

$$0 \leq k \leq n-1$$

The only way for the original bottom card to move upward by one position is for the current top card to be inserted somewhere beneath it.

Since the top card is inserted uniformly among the n possible insertion positions, exactly $k+1$ of these positions lie beneath the original bottom card. Consequently,

$$P(\text{bottom card rises by one position}) = \frac{k+1}{n}$$

Thus, while there are k cards beneath it, the original bottom card waits until the first successful insertion beneath it. The waiting time therefore follows a geometric distribution.

Proposition 4. *If there are k cards beneath the original bottom card, then the expected number of shuffles required for the card to move upward by one position is*

$$\frac{n}{k+1}$$

Proof. Let T_k denote the waiting time while there are k cards beneath the original bottom card.

Since each shuffle independently moves the card upward with probability

$$p = \frac{k+1}{n},$$

the random variable T_k has a geometric distribution with parameter p.

For a geometric random variable,

$$\mathbb{E}[T_k] = \frac{1}{p},$$

and therefore

$$\mathbb{E}[T_k] = \frac{n}{k+1}$$

□

The original bottom card begins with $n-1$ cards beneath it and must successively move upward through each position until no cards remain beneath it. Consequently,

$$\tau_{\text{top}} = T_{n-1} + T_{n-2} + \cdots + T_0$$

Using linearity of expectation immediately yields the following theorem.

Theorem 9. *The expected value of the strong stationary time is*

$$\mathbb{E}[\tau_{\text{top}}] = nH_n,$$

where

$$H_n = \sum_{i=1}^{n} \frac{1}{i}$$

is the nth harmonic number.

Proof. Applying linearity of expectation,

$$\mathbb{E}[\tau_{\text{top}}] = \sum_{k=0}^{n-1} \mathbb{E}[T_k]$$

Substituting the previous proposition,

$$= \sum_{k=0}^{n-1} \frac{n}{k+1} = n \sum_{i=1}^{n} \frac{1}{i} = nH_n$$

□

The harmonic numbers satisfy the well-known asymptotic expansion

$$H_n = \log n + \gamma + O\left(\frac{1}{n}\right),$$

where $\gamma \approx 0.57721$ denotes the Euler–Mascheroni constant.
Hence,

$$\boxed{\mathbb{E}[\tau_{\text{top}}] = n \log n + O(n)}$$

This establishes that the expected stopping time grows on the order of $n \log n$.

5.3.2 Connection with the Coupon Collector Problem

The preceding calculation is identical to that arising in the classical coupon collector problem.

Suppose that there are n distinct types of coupons, and at each trial one coupon is selected independently and uniformly at random. The objective is to determine the number of trials required until every type of coupon has been collected.

Initially, every coupon is new, so new coupons are obtained frequently. As the collection nears completion, however, the remaining unseen coupon types become increasingly rare. If k distinct coupon types have already been collected, then the probability that the next coupon is new is

$$\frac{n-k}{n},$$

and the expected waiting time to obtain a new coupon is

$$\frac{n}{n-k}$$

Summing these waiting times gives exactly

$$nH_n$$

Thus both processes are governed by the same sequence of geometric waiting times.

Levin Peres show that the stopping time τ_{top} has precisely the same distribution as the coupon collector completion time. Consequently, the probabilistic estimates developed for the coupon collector problem apply directly to the top-to-random shuffle.

The coupon collector interpretation provides considerably more than an expectation. Standard concentration results imply that, with high probability,

$$\tau_{\text{top}} = n \log n + O(n),$$

which strongly suggests that approximately $n \log n$ shuffles are sufficient for the top-to-random shuffle to become well mixed.

5.4 Mixing Time Bounds

We are now in a position to combine the results of the previous sections. Recall that Theorem 5.1 states that if τ is a strong stationary time for a Markov chain with stationary distribution π, then

$$\left\|P^t(x,\cdot) - \pi\right\|_{\text{TV}} \leq \Pr(\tau > t),$$

for every initial state x.

For the top-to-random shuffle, we have shown that the stopping time τ_{top} is a strong stationary time and that

$$\mathbb{E}[\tau_{\text{top}}] = nH_n = n \log n + O(n)$$

Thus, to bound the total variation distance, it suffices to estimate the probability that the stopping time has not yet occurred.

The coupon collector problem (Levin, Peres) provides precisely such an estimate. A classical result states that for every real number c,

$$\Pr(\tau_{\text{top}} > n \log n + cn) \leq e^{-c}$$

Combining this inequality with the strong stationary time bound yields

$$\left\|P^{n \log n + cn}(x,\cdot) - \pi\right\|_{\text{TV}} \leq e^{-c}$$

Consequently, after approximately

$$n \log n$$

top-to-random shuffles, the distribution of the deck is exponentially close to the uniform distribution. This establishes the principal asymptotic result of the chapter.

Theorem 10. *The mixing time of the top-to-random shuffle satisfies*

$$t_{\text{mix}} = O(n \log n)$$

Proof. From the preceding inequality,

$$\left\| P^{n \log n + cn}(x, \cdot) - \pi \right\|_{\text{TV}} \leq e^{-c}$$

Given any fixed $\varepsilon > 0$, choose

$$c = \log\left(\frac{1}{\varepsilon}\right)$$

Then

$$e^{-c} = \varepsilon,$$

so that

$$\left\| P^{n \log n + n \log(1/\varepsilon)}(x, \cdot) - \pi \right\|_{\text{TV}} \leq \varepsilon$$

Since ε is fixed, the additional term

$$n \log(1/\varepsilon)$$

is linear in n. Therefore the dominant term is

$$n \log n,$$

which proves that

$$t_{\text{mix}} = O(n \log n)$$

□

6 Riffle Shuffles

The top-to-random shuffle provides an introduction to the theory of finite Markov chains and mixing times. Its simple description allows many of its properties to be established rigorously, including irreducibility, the existence of a unique stationary distribution, and an $O(n \log n)$ upper bound on its mixing time. Nevertheless, the top-to-random shuffle is not representative of the way people ordinarily shuffle playing cards.

In practice, cards are almost always shuffled by splitting the deck into two packets and interleaving them. This procedure, known as a *riffle shuffle*, preserves much of the relative ordering within each packet while creating randomness through their interleaving. Although its physical description is simple, its mathematical analysis is considerably richer than that of the top-to-random shuffle. Remarkably, it is this model that leads to the celebrated theorem of Bayer and Diaconis explaining why approximately seven riffle shuffles suffice to randomize a standard deck of fifty-two cards.

In this section we introduce the Gilbert–Shannon–Reeds (GSR) model of riffle shuffling, explain how it is represented mathematically, and develop the combinatorial tools required for its analysis.

6.1 The Gilbert–Shannon–Reeds Model

The Gilbert–Shannon–Reeds model, commonly abbreviated as the *GSR model*, is the standard mathematical model for an ideal riffle shuffle. It was independently proposed by Gilbert, Shannon, and Reeds as a probabilistic abstraction of the way cards are shuffled in practice.

The shuffle consists of two independent stages.

1. The deck is cut into two packets.
2. The two packets are interleaved while preserving the relative order of the cards within each packet.

Unlike an ordinary cut, the position at which the deck is divided is itself random.

Definition 12 (Gilbert–Shannon–Reeds Shuffle)**.** *Suppose a deck contains n cards.*

First choose an integer

$$k \in \{0, 1, \dots, n\}$$

according to the binomial distribution

$$P(\textit{cut after } k \textit{ cards}) = \frac{\binom{n}{k}}{2^n}$$

The top k cards form the first packet, while the remaining $n - k$ cards form the second packet.

The two packets are then interleaved uniformly among all possible interleavings that preserve the relative order of the cards within each packet.

The choice of a binomial distribution needs explanation. Imagine independently assigning each card to either the left or right hand with probability $1/2$. The number of cards assigned to the left hand is then distributed as

$$\mathrm{Bin}(n, \tfrac{1}{2}),$$

which immediately gives

$$P(k) = \frac{\binom{n}{k}}{2^n}$$

Thus the GSR model captures the natural variability observed when a person cuts a deck before shuffling.

Once the cut has been made, the two packets are interleaved. Importantly, the cards within each packet never change their relative order.

For example, suppose the deck initially consists of

$$12345678,$$

and is cut after the fourth card, producing the packets

$$1234 \qquad \text{and} \qquad 5678$$

One possible riffle shuffle is

$$15263748.$$

Observe that the cards

$$1 < 2 < 3 < 4$$

still appear in this order, although they are separated by cards from the second packet. Likewise,

$$5 < 6 < 7 < 8$$

remain in their original order.

This preservation of relative order is the defining feature of a riffle shuffle and distinguishes it from simpler models such as the top-to-random shuffle.

For a fixed cut after k cards, the number of possible interleavings is

$$\binom{n}{k},$$

since one need only choose which k of the n positions are occupied by cards from the first packet; the remaining positions are automatically filled by the second packet.

Each of these interleavings is assumed to be equally likely. Therefore the probability of obtaining any particular interleaving is

$$\frac{1}{\binom{n}{k}}$$

Combining the probability of the cut with the probability of the subsequent interleaving gives

$$\frac{\binom{n}{k}}{2^n} \cdot \frac{1}{\binom{n}{k}} = \frac{1}{2^n}$$

Thus every possible pair consisting of a cut and a valid interleaving occurs with exactly the same probability.

This observation is surprisingly powerful. Although different permutations may arise from different numbers of cut–interleaving pairs, every individual pair is equally likely. Consequently, the probability of obtaining a particular permutation is determined entirely by the number of distinct ways in which it can be produced.

6.2 Rising Sequences

The probability distribution induced by a Gilbert–Shannon–Reeds shuffle possesses a remarkable combinatorial structure. Although the state space consists of all $n!$ permutations of the deck, the probability assigned to a permutation does not depend on the permutation itself, but rather on a single statistic known as the number of *rising sequences*. This observation, first used by Bayer and Diaconis, forms the foundation of the modern analysis of riffle shuffles.

Definition 13. *Let $\pi = (\pi_1, \pi_2, \ldots, \pi_n) \in S_n$. A* rising sequence *is a maximal consecutive increasing subsequence of π. Equivalently, a new rising sequence begins whenever*

$$\pi_i > \pi_{i+1}$$

The total number of rising sequences of π is denoted by $r(\pi)$.

A completely ordered deck has a single rising sequence, while increasingly disordered decks tend to possess a larger number of rising sequences.

Example 3. *Consider the permutation*

$$\pi = (1, 5, 2, 6, 3, 7, 4, 8)$$

Since descents occur after 5, 6, and 7, the permutation decomposes as

$$(1, 5) \mid (2, 6) \mid (3, 7) \mid (4, 8),$$

and hence

$$r(\pi) = 4$$

Similarly,

$$(1, 2, 3, 4, 5)$$

has one rising sequence, while

$$(3, 1, 4, 2, 5) = (3) \mid (1, 4) \mid (2, 5)$$

contains three rising sequences.

Proposition 5. *A permutation $\pi \in S_n$ can be produced by a single Gilbert–Shannon–Reeds 2-shuffle if and only if*

$$r(\pi) \leq 2$$

Proof. Suppose first that π is obtained from a single 2-shuffle.

The deck is divided into two packets,

$$A = (1, \dots, k), \qquad B = (k+1, \dots, n),$$

for some $0 \leq k \leq n$, after which the packets are interleaved while preserving the order of the cards within each packet.

Consequently, every card originating from packet A appears in increasing order, and the same is true for packet B. Any descent in the final permutation can therefore occur only when the interleaving switches from a card belonging to packet B to the next card belonging to packet A.

Since there is only one boundary separating the two packets, there can be at most one such switch that creates a new rising sequence. Hence the resulting permutation contains at most two rising sequences.

Conversely, suppose that π contains at most two rising sequences.

If $r(\pi) = 1$, then π is the identity permutation and is obtained by performing no effective shuffle.

Assume therefore that $r(\pi) = 2$. Write

$$\pi = R_1 \mid R_2,$$

where R_1 and R_2 are the two maximal increasing blocks.

Because each block is increasing, we may regard R_1 and R_2 as the two packets created by the initial cut. Interleaving these packets while preserving their internal order reproduces precisely the permutation π.

Thus every permutation with at most two rising sequences arises from a single 2-shuffle.

Therefore,

$$\pi \text{ is obtainable by one 2-shuffle} \iff r(\pi) \leq 2$$

$\square$

The preceding proposition reveals an unexpected connection between a probabilistic process and a purely combinatorial statistic. Rather than studying individual permutations, one may classify them according to the number of rising sequences they contain. This observation extends far beyond a single shuffle. After repeated riffle shuffles, the appropriate statistic is no longer two rising sequences, but rather the number of packets into which the deck may be divided while preserving the internal order of each packet. This naturally leads to the notion of an a-shuffle.

6.3 Composition of Riffle Shuffles

The Gilbert–Shannon–Reeds model describes a single riffle shuffle by cutting the deck into two packets and interleaving them uniformly at random. To analyse repeated shuffles, it is convenient to introduce a more general class of shuffles.

Definition 14. *Let $a \geq 2$ be an integer. An a-*shuffle *is performed as follows.*

1. *Each card is independently assigned one of the labels*

$$1, 2, \dots, a,$$

   each with probability $1/a$.

2. *The cards are collected into a packets according to their labels, preserving the relative order of the cards within each packet.*

3. *The packets are stacked in increasing order of their labels.*

When $a = 2$, this construction coincides with the Gilbert–Shannon–Reeds riffle shuffle.

The labelling description is entirely equivalent to the cut-and-interleave construction introduced in the previous subsection, but it possesses a major advantage: repeated shuffles become remarkably easy to analyse.

The key observation is that the labels assigned during successive shuffles may themselves be combined into a single label.

Theorem 11 (Multiplication Theorem)**.** *Performing an a-shuffle followed by a b-shuffle is equivalent in distribution to performing a single ab-shuffle.*

Proof. Suppose the first shuffle assigns each card a label

$$i \in \{1, \dots, a\},$$

while the second shuffle independently assigns each card a label

$$j \in \{1, \dots, b\}$$

Each card therefore receives an ordered pair

$$(i, j)$$

Since the two labellings are independent, every ordered pair occurs with probability

$$\frac{1}{a} \cdot \frac{1}{b} = \frac{1}{ab}$$

Now associate to the pair (i, j) the single label

$$k = (i - 1)b + j,$$

which takes values in

$$\{1, 2, \dots, ab\}$$

Because every pair (i, j) is equally likely, every value of k occurs with probability

$$\frac{1}{ab}$$

Moreover, ordering first by the first coordinate and then by the second coordinate is precisely the same as ordering by the integer k. Consequently, the final arrangement of the cards obtained after performing an a-shuffle followed by a b-shuffle is identical to that obtained by assigning each card one of the ab labels uniformly at random and collecting the packets according to these labels.

Hence the two procedures produce the same probability distribution on S_n. □

An immediate consequence of the multiplication theorem is that repeated riffle shuffles admit an especially simple description.

Corollary 1. *The composition of k independent Gilbert–Shannon–Reeds riffle shuffles is equivalent to a single*

$$2^k\text{-}\textit{shuffle}$$

Proof. Apply the multiplication theorem repeatedly. Two successive 2-shuffles form a 4-shuffle, three successive 2-shuffles form an 8-shuffle, and, by induction, k successive 2-shuffles form a

$$2^k\text{-shuffle}$$

□

The multiplication theorem is one of the principal reasons why the Gilbert–Shannon–Reeds model is mathematically tractable. Rather than analysing a complicated sequence of repeated shuffles, one may instead study a single 2^k-shuffle.

6.4 The Bayer–Diaconis Theorem

The theory developed in the preceding sections culminates in one of the most celebrated results in the study of random walks on finite groups. In a landmark paper published in 1992, Bayer and Diaconis completely determined the asymptotic mixing time of the Gilbert–Shannon–Reeds riffle shuffle.

Recall that if P^k denotes the distribution of the deck after k riffle shuffles and π denotes the uniform distribution on S_n, then the total variation distance

$$\|P^k - \pi\|_{\mathrm{TV}}$$

measures how far the deck remains from perfect randomness.

The breakthrough discovery of Bayer and Diaconis is that this distance undergoes an abrupt transition after approximately

$$\frac{3}{2}\log_2 n$$

riffle shuffles.

Theorem 12 (Bayer–Diaconis)**.** *Let*

$$k = \frac{3}{2}\log_2 n + c,$$

where c is a fixed constant independent of n. Then, as $n \to \infty$,

$$\|P^k - \pi\|_{\mathrm{TV}} \longrightarrow \begin{cases} 1, & c \to -\infty, \\ 0, & c \to +\infty \end{cases}$$

In particular, the mixing time of the Gilbert–Shannon–Reeds riffle shuffle satisfies

$$t_{\mathrm{mix}} = \frac{3}{2}\log_2 n + O(1)$$

The theorem demonstrates the existence of a *cutoff phenomenon.* Before approximately

$$\frac{3}{2}\log_2 n$$

shuffles have been performed, the deck remains far from uniformly distributed. However, after only a few additional shuffles, the total variation distance decreases rapidly to zero. Rather than converging gradually, the chain exhibits a sharp transition from ordered to random.

Although the complete proof lies beyond the scope of this paper, its principal ideas may be summarized as follows.

1. Repeated riffle shuffles are reduced to a single 2^k-shuffle using the Multiplication Theorem proved in the previous subsection.
2. The probability assigned to a permutation after an a-shuffle depends only upon the number of rising sequences contained in that permutation.
3. Explicit formulas are obtained for the probability of every permutation in terms of its rising sequences.
4. These probabilities are compared with the uniform distribution to estimate the total variation distance.
5. Finally, asymptotic estimates for the distribution of rising sequences establish that the transition from non-randomness to randomness occurs within a bounded window centred at

$$\frac{3}{2}\log_2 n$$

The second step constitutes the central combinatorial insight of the argument. Instead of analysing all $n!$ permutations individually, Bayer and Diaconis showed that it is sufficient to classify permutations according to their number of rising sequences.

6.5 The Seven-Shuffle Rule

The theorem of Bayer and Diaconis gives an asymptotic description of the mixing behaviour of riffle shuffles. A natural question is whether this asymptotic result accurately predicts the behaviour of an ordinary deck of fifty-two playing cards.

Substituting $n = 52$ into the expression

$$\frac{3}{2}\log_2 n,$$

gives

$$\frac{3}{2}\log_2(52) \approx 8.55$$

At first sight, one might therefore expect that approximately nine riffle shuffles are required before the deck becomes well mixed. Surprisingly, a more careful analysis of the exact transition probabilities reveals that the convergence to stationarity is already well advanced after only seven shuffles.

Bayer and Diaconis computed the total variation distance between the distribution of the deck after successive riffle shuffles and the uniform distribution on S_{52}. Their results are summarized in Table 1.

Number of riffle shuffles	Total variation distance
0	1.000
1	1.000
2	1.000
3	1.000
4	0.924
5	0.614
6	0.334
7	0.167
8	0.085
9	0.043
10	0.022
11	0.011

Table 1: Approximate total variation distance for repeated Gilbert–Shannon–Reeds riffle shuffles of a standard deck of 52 cards.

The table clearly illustrates the cutoff phenomenon described in the previous subsection. For the first several shuffles, the deck remains far from uniformly distributed. However, between approximately six and eight shuffles the total variation distance decreases dramatically, indicating a rapid transition from an ordered deck to one that is effectively random.

The value seven therefore occupies a distinguished position. Although it is not an exact mathematical threshold, it represents the point beyond which further shuffling produces only comparatively small improvements in randomness. Consequently, the statement that "seven riffle shuffles suffice" has become the accepted practical interpretation of the Bayer–Diaconis theorem.

It is important to emphasize that this conclusion depends upon the Gilbert–Shannon–Reeds model. Human shuffles are not perfectly random, and skilled card manipulators may intentionally introduce biases that invalidate the assumptions of the model. Nevertheless, experimental studies have demonstrated that ordinary riffle shuffles performed by non-experts are remarkably well approximated by the GSR model, making the seven-shuffle rule both mathematically justified and practically meaningful.

The analysis of card shuffling has given an elegant illustration of how 7 seven shuffles are said to be enough for the usual deck of 52 cards.

References

[1] Peter G. Doyle and J. Laurie Snell, *Random Walks and Electric Networks*, Mathematical Association of America, 1984. Available at: `https://math.dartmouth.edu/~doyle/docs/walks/walks.pdf`

[2] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer, *Markov Chains and Mixing Times*, 2nd ed., American Mathematical Society, Providence, RI, 2017.

[3] Dave Bayer and Persi Diaconis, "Trailing the Dovetail Shuffle to Its Lair," *The Annals of Applied Probability*, Vol. 2, No. 2, 1992, pp. 294–313.

[4] David Austin, "How Many Times Should You Shuffle a Deck of Cards?" *Feature Column*, American Mathematical Society. Available at: `https://www.ams.org/publicoutreach/feature-column/fcarc-shuffle`