On Shuffles and their Mixing Time

Ayush Goyal

Euler Circle Independent Research Programme

July 9, 2026

Why Study Card Shuffling?

How many shuffles are required before a deck is random?

Why Study Card Shuffling?

How many shuffles are required before a deck is random?

What kinds of shuffles exist?

Why Study Card Shuffling?

How many shuffles are required before a deck is random?

What kinds of shuffles exist?

How do we mathematically define "random"?

Why Study Card Shuffling?

How many shuffles are required before a deck is random?

What kinds of shuffles exist?

How do we mathematically define "random"?

When is a deck *truly* random?

Why Study Card Shuffling?

How many shuffles are required before a deck is random?

What kinds of shuffles exist?

How do we mathematically define "random"?

When is a deck *truly* random?

Goal of the talk

Understand the different types of shuffles and mathematical approaches to them. To look at the ideas behind the proofs of such arguments

A Very Brief Introduction to Probability

Probability provides a mathematical framework for studying random events.
If A is an event,

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

Conditional probability:

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

Law of Total Probability:

$$P(A) = \sum_i P(B_i) P(A \mid B_i),$$

where $\{B_i\}$ forms a partition of the sample space.

A Very Brief Introduction to Probability

Probability provides a mathematical framework for studying random events.
If A is an event,

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

Conditional probability:

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

Law of Total Probability:

$$P(A) = \sum_i P(B_i) P(A \mid B_i),$$

where $\{B_i\}$ forms a partition of the sample space.

Point of looking at probability for a probem like shuffling

Every shuffle is a random experiment (unlike the deterministic one we saw in the guest lecture!), so probability forms the language in which we describe card shuffling.

Markov Chains

A Markov chain is a random process satisfying the **Markov Property**:

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

In words,

The future depends only on the present, not the past.

Markov Chains

A Markov chain is a random process satisfying the **Markov Property**:

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

In words,

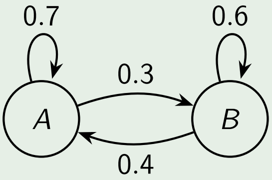
The future depends only on the present, not the past.

Example

Transition Matrix

$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}$$

Each row sums to 1.



Every card shuffle defines a Markov chain whose states are the possible orderings of the deck.

Markov Chains

A Markov chain has the memory of a goldfish.

The next state depends only on the current state, not on how it got there.

Stationary Distributions

A probability distribution

$$\pi = (\pi_1, \pi_2, \ldots, \pi_n)$$

is called a **stationary distribution** if

$$\boxed{\pi P = \pi.}$$

Stationary Distributions

A probability distribution

$$\pi = (\pi_1, \pi_2, \ldots, \pi_n)$$

is called a **stationary distribution** if

$$\boxed{\pi P = \pi.}$$

In other words,

If the chain starts in the distribution π, it remains in the same distribution after one transition.

Interpretation

A stationary distribution represents the long-term equilibrium of the Markov chain.

Irreducibility and Aperiodicity

Irreducible Every state can eventually be reached from every other state.

Irreducibility and Aperiodicity

Irreducible Every state can eventually be reached from every other state.

Aperiodic The chain is not trapped in deterministic cycles.

Irreducibility and Aperiodicity

Irreducible Every state can eventually be reached from every other state.

Aperiodic The chain is not trapped in deterministic cycles.

Theorem (Fundamental Theorem of Finite Markov Chains)

Every finite, irreducible and aperiodic Markov chain possesses a unique stationary distribution. Furthermore,

$$P^n(x, \cdot) \longrightarrow \pi$$

for every initial state.

The Symmetric Group S_n

Each ordering of a deck is a permutation.

The set of all permutations of

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

is called the **symmetric group**

$$S_n$$

The Symmetric Group S_n

Each ordering of a deck is a permutation.

The set of all permutations of

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

is called the **symmetric group**

$$S_n$$

Example

For a deck of 3 cards, the symmetric group S_3 is

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

The Symmetric Group S_n

Each ordering of a deck is a permutation.

The set of all permutations of

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

is called the **symmetric group**

$$S_n$$

Example

For a deck of 3 cards, the symmetric group S_3 is

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

A deck of n cards therefore has following possible configurations

$$\boxed{n!}$$

Card Shuffling as a Markov Chain

Each deck configuration is a state.

One shuffle corresponds to one transition between states.

Card Shuffling as a Markov Chain

Each deck configuration is a state.

One shuffle corresponds to one transition between states.

The transition matrix is

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

where

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

Idea

Repeated shuffling is simply a random walk on the symmetric group.

Properties of the Transition Matrix

Proposition

Proposition 1

Every row of the transition matrix sums to one.

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

Proof

Each shuffle must end in exactly one deck configuration.

Properties of the Transition Matrix

Proposition

Proposition 2

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

Proof

Induction on k using the Law of Total Probability.

The Top-to-Random Shuffle

A **top-to-random shuffle** is performed as follows:

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

The Top-to-Random Shuffle

A **top-to-random shuffle** is performed as follows:

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

Example

Initial configuration: Remove top card

$$12345 : 2345$$

Possible outcomes:

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

The Top-to-Random Shuffle

A **top-to-random shuffle** is performed as follows:

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

Example

Initial configuration: Remove top card

$$12345 : 2345$$

Possible outcomes:

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

Each occurs with probability

$$\boxed{\frac{1}{n}}.$$

Transition Structure

The state space is

$$S_n.$$

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

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

Transition Structure

The state space is

$$S_n.$$

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

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

For the top-to-random shuffle,

$$P(\sigma, \tau) = \begin{cases} \frac{1}{n}, & \text{if } \tau \text{ is reachable in one shuffle,} \\ 0, & \text{otherwise.} \end{cases}$$

Observation

Every row contains exactly n non-zero entries.

Why is the Chain Irreducible and Aperiodic?

Irreducibility

Any permutation can be transformed into any other by repeatedly moving the top card to appropriate positions.

Why is the Chain Irreducible and Aperiodic?

Irreducibility

Any permutation can be transformed into any other by repeatedly moving the top card to appropriate positions.

Aperiodicity

There is probability

$$\frac{1}{n}$$

of reinserting the top card back on top. Hence, for every state

$$P(\sigma, \sigma) > 0$$

Why is the Chain Irreducible and Aperiodic?

Irreducibility

Any permutation can be transformed into any other by repeatedly moving the top card to appropriate positions.

Aperiodicity

There is probability

$$\frac{1}{n}$$

of reinserting the top card back on top. Hence, for every state

$$P(\sigma, \sigma) > 0$$

Therefore,

$$\boxed{\text{The chain is irreducible and aperiodic.}}$$

The Uniform Stationary Distribution

Consider the uniform distribution

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

Then

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

The Uniform Stationary Distribution

Consider the uniform distribution

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

Then

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

Each permutation has exactly n predecessors.

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

Therefore,

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

$$\boxed{\pi P = \pi}$$

Stopping Times

A **stopping time** is a random variable

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

such that, for every $t \geq 0$,

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

can be determined solely from

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

In words: Whether the process stops at time t can be decided without looking into the future.

Example

The first time a random walk reaches a given state.

Strong Stationary Times

A stopping time τ is called a **Strong Stationary Time (SST)** if

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

for every state x.

Strong Stationary Times

A stopping time τ is called a **Strong Stationary Time (SST)** if

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

for every state x.

Equivalently,

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

Interpretation

When the stopping time occurs, the chain is **exactly** at stationarity, independent of when it stopped.

Total Variation Distance

For two probability distributions μ and ν,

$$\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2} \sum_{x} |\mu(x) - \nu(x)|.$$

It measures how far a distribution is from equilibrium.

Total Variation Distance

For two probability distributions μ and ν,

$$\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2}\sum_{x} |\mu(x) - \nu(x)|.$$

It measures how far a distribution is from equilibrium.

Theorem

If τ is a strong stationary time,then

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

A Strong Stationary Time

Define

$$\tau_{\text{top}}$$

to be one shuffle after the original bottom card first reaches the top

A Strong Stationary Time

Define

$$\tau_{\text{top}}$$

to be one shuffle after the original bottom card first reaches the top

Theorem

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

Main Idea

By the time the original bottom card reaches the top, the remaining cards are already uniformly ordered.

Coupon Collector Interpretation

Suppose there are k cards beneath the original bottom card.

Coupon Collector Interpretation

Suppose there are k cards beneath the original bottom card.

Probability that it rises one position:

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

Coupon Collector Interpretation

Suppose there are k cards beneath the original bottom card.

Probability that it rises one position:

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

Expected waiting time:

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

Coupon Collector Interpretation

Suppose there are k cards beneath the original bottom card.

Probability that it rises one position:

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

Expected waiting time:

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

This is exactly the same sequence of waiting times as the classical **Coupon Collector Problem**

Coupon Collector Interpretation

Suppose there are k cards beneath the original bottom card.

Probability that it rises one position:

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

Expected waiting time:

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

This is exactly the same sequence of waiting times as the classical **Coupon Collector Problem**

Problem (Coupon Collector Interpretation)

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

Expected Stopping Time

Proposition

The expect value of the strong stationary time is

$$E[\tau_{\mathrm{top}}] = nH_n$$

where H_n is the n^{th} harmonic number

Expected Stopping Time

Proposition

The expect value of the strong stationary time is

$$E[\tau_{\mathrm{top}}] = nH_n$$

where H_n is the n^{th} harmonic number

Theorem

The mixing time of the top-to-random shuffle satisfies

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

Why Riffle Shuffles?

So far, we have analysed the **top-to-random shuffle**.

Top-to-Random

Remove the top card

$\downarrow$

Insert uniformly

$\downarrow$

Repeat

Why Riffle Shuffles?

So far, we have analysed the **top-to-random shuffle**.

Top-to-Random	**Real Life**
Remove the top card	Split the deck
$\downarrow$	$\downarrow$
Insert uniformly	Interleave
$\downarrow$	$\downarrow$
Repeat	Repeat

Question

Can mathematics explain the shuffle that people actually use?

The Gilbert–Shannon–Reeds (GSR) Model

An ideal riffle shuffle consists of two steps.

Step 1: Random Cut

12345678 to 1234 and 5678

The Gilbert–Shannon–Reeds (GSR) Model

An ideal riffle shuffle consists of two steps.

Step 1: Random Cut

12345678 to 1234 and 5678

Step 2: Uniform Interleaving

15263748

The Gilbert–Shannon–Reeds (GSR) Model

An ideal riffle shuffle consists of two steps.

Step 1: Random Cut

$$12345678 \text{ to } 1234 \text{ and } 5678$$

Step 2: Uniform Interleaving

$$15263748$$

Properties:

- Cut size follows a Binomial distribution.
- Relative order within each packet is preserved.
- Every cut–interleaving pair occurs with probability

$$\frac{1}{2^n}.$$

Rising Sequences

Definition

A *rising sequence* is a maximal increasing consecutive subsequence of a permutation.

Example

$$15263748$$

splits into

$$15 \mid 26 \mid 37 \mid 48.$$

Hence this permutation has 4 rising sequences

Why are they important?

The probability of obtaining a permutation after a riffle shuffle depends *only* on its number of rising sequences.

Repeated Riffle Shuffles

One riffle shuffle is called a **2-shuffle**.
Repeating the shuffle produces increasingly finer interleavings.

Repeated Riffle Shuffles

One riffle shuffle is called a **2-shuffle**.
Repeating the shuffle produces increasingly finer interleavings.

Theorem (Multiplication Theorem)

Performing an a-shuffle followed by a b-shuffle is equivalent to performing a single

$$ab\text{-shuffle.}$$

Repeated Riffle Shuffles

One riffle shuffle is called a **2-shuffle**.
Repeating the shuffle produces increasingly finer interleavings.

Theorem (Multiplication Theorem)

Performing an a-shuffle followed by a b-shuffle is equivalent to performing a single

$$ab\text{-shuffle}.$$

Hence,

$$\boxed{k \text{ riffle shuffles} = 2^k\text{-shuffle}}$$

This remarkable fact allows repeated shuffles to be analysed as a single random process.

The Bayer–Diaconis Theorem

The celebrated theorem of Bayer and Diaconis completely characterizes the mixing behaviour of riffle shuffles.

The Bayer–Diaconis Theorem

The celebrated theorem of Bayer and Diaconis completely characterizes the mixing behaviour of riffle shuffles.

Theorem (Bayer–Diaconis, 1992)

For a deck of n cards,

$$\boxed{\textit{approximately } \frac{3}{2}\log_2 n \textit{ riffle shuffles}}$$

are necessary and sufficient for the deck to become close to uniformly random.

The Bayer–Diaconis Theorem

The celebrated theorem of Bayer and Diaconis completely characterizes the mixing behaviour of riffle shuffles.

Theorem (Bayer–Diaconis, 1992)

For a deck of n cards,

$$\boxed{\textit{approximately } \frac{3}{2}\log_2 n \textit{ riffle shuffles}}$$

are necessary and sufficient for the deck to become close to uniformly random.

For a standard deck,

$$n = 52,$$

giving the famous

$$\boxed{\textbf{7 riffle shuffles.}}$$

Why Seven?

1 Shuffle	Still highly structured

Why Seven?

1 Shuffle	Still highly structured
3 Shuffles	Significant order remains

Why Seven?

1 Shuffle	Still highly structured
3 Shuffles	Significant order remains
5 Shuffles	Nearly random

Why Seven?

1 Shuffle	Still highly structured
3 Shuffles	Significant order remains
5 Shuffles	Nearly random
7 Shuffles	**Indistinguishable from random**

Why Seven?

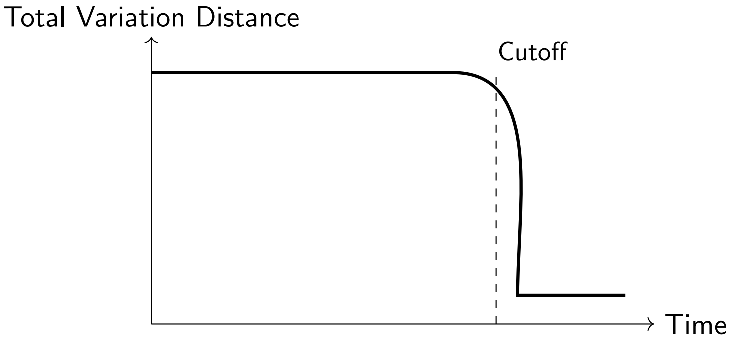
1 Shuffle	Still highly structured
3 Shuffles	Significant order remains
5 Shuffles	Nearly random
7 Shuffles	**Indistinguishable from random**

Interpretation

The transition from "ordered" to "random" occurs very abruptly.

The Cutoff Phenomenon

Many Markov chains do *not* converge gradually. Instead, they remain far from stationarity before becoming random over a very short interval.



Key Takeaways

- Probability provides a mathematical language for randomness.
- Card shuffling can be modelled as a finite Markov chain on the symmetric group S_n.
- Strong stationary times give elegant bounds on mixing time.
- The top-to-random shuffle mixes in the following amount of steps

$$O(n \log n)$$

- The Gilbert–Shannon–Reeds model explains realistic riffle shuffles.
- The celebrated Bayer–Diaconis theorem shows that

$$\boxed{\textbf{approximately seven riffle shuffles}}$$

randomize a standard deck.

Thank You!

Questions?

Ayush Goyal
Euler Circle Independent Research Programme