Snapshots of modern mathematics
from Oberwolfach

№ 5/2023

Cutoff phenomenon: Surprising behaviour in card shuffling and other Markov chains

Isabelle Baraquin • Nadia Lafrenière
Katharina Schuh

This snapshot compares two techniques of shuffling a deck of cards, asking how long it will take to shuffle the cards until a "well-mixed deck" is obtained. Surprisingly, the number of shuffles can be very different for very similar looking shuffling techniques.

1 Shuffling a deck of cards

When playing a card game, it is important to shuffle the card deck well in the beginning. But how long do you have to shuffle the deck until the cards are well-mixed? If you shuffle too few times, it will be easy to guess which card is the next one. But you also do not want to spend too much time shuffling the deck. Formulated more precisely, the question is:

> *What is the minimal number of times one has to repeat the shuffle to obtain a nearly random card order?*

To be able to answer this question, we need to specify the size of the deck, choose a shuffling technique and explain what it means to be close to random. Let us take a standard deck with 52 cards and shuffle them with the common overhand shuffling technique, which is the technique of letting small bunches of cards dropping from one hand to the other. A deck is said to be *close to random* if all the possible orders are almost equally likely.

Probability theory, a branch of mathematics which measures the likelihood of repeated random events like our overhand shuffle, tells us that not all shuffling techniques are efficient enough to be of practical use: For instance, shuffling with the overhand shuffle technique would take thousands of shuffles to be effective on a standard deck with 52 cards!

Now consider another common technique for card shuffling: first break the deck into two smaller decks, of approximately equal size, by taking the top block in one hand and the bottom block in the other one. Don't be too strict on splitting exactly in the middle, as we wish to create some randomness here.[1] Take one block in each hand, and interlace them neatly so they become only one deck. This technique, called the *riffle shuffle*, is known to be very efficient, meaning we would have to shuffle less than ten times [2], and is used by casual card players and casino dealers alike.

To try to quantify the number of times we must repeat the shuffle to get a truly random sequence of cards in our deck, we look at certain pieces of information that may indicate randomness. In the case of the riffle shuffle, the number of *rising sequences* is quite informative. Rising sequences are sequences of cards that were adjacent and in a certain order, and which are still in that order once the deck is mixed (see Figure 1).

Typically, a random deck of n cards has around $\frac{n}{2}$ rising sequences, and can have up to n such sequences. However, shuffling the deck one time creates only two rising sequences. At each step, the number of rising sequences is at most multiplied by two. Hence, if we want all arrangements of the deck to be possible, we need, at the very least, to shuffle enough times to get (hypothetically) n rising sequences. This is achieved in $\log_2(n)$ times (which is the number of times we must multiply by two to get n). That shows that we need at least $\log_2(n)$ times to get a well-mixed deck of cards. In the next paragraph, we will find out how many times we need to shuffle the card deck such that it will very likely be well-mixed.

The right number of shuffles

Determining the number of shuffles which are sufficient to get a well-mixed deck is closely related to determining how close a deck is from being perfectly mixed. As the distribution of the cards in the deck becomes more and more random each time we iterate the shuffling procedure, we can express the randomness as a function of the number of iterations. To compute this function, we compare

[1] Mathematically speaking, we cut the deck according to a binomial distribution, which means that the probability of the top deck having k cards is the same as the probability of getting k times "tail" after tossing a fair coin as many times as there are cards.

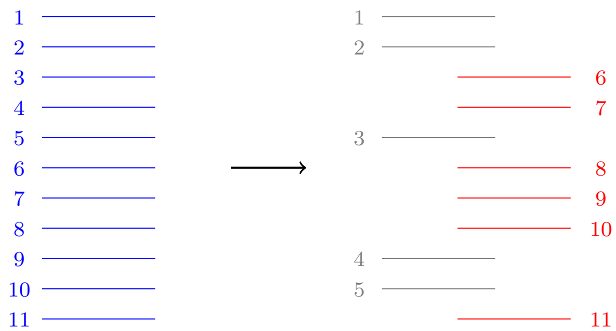

Figure 1: In this 11-card deck, the rising sequences are pictured in two different columns, each column corresponding to one hand in the riffle shuffle. After one riffle shuffle, the order of the cards in the deck is $(1, 2, 6, 7, 3, 8, 9, 10, 4, 5, 11)$, and the rising sequences are $(1, 2, 3, 4, 5)$, visualized in black, and $(6, 7, 8, 9, 10, 11)$, visualized in red. Each time we shuffle, the number of rising sequences is at most multiplied by 2.

the probability that each ordering of the deck occurs after our shuffling to the probability that it occurs after perfect mixing. Formally, for a given ordering σ, we denote by $\tau^t(\sigma)$ the probability that the deck is in that order after t shuffles. Moreover, we let $\pi(\sigma)$ be the probability that the deck follows the order σ if it was perfectly mixed (so this probability is the same for any ordering). Then, the randomness of the shuffle is computed using the following sum over all possible orderings:

$$d(t) = \frac{1}{2} \sum_{\sigma \text{ an ordering}} |\tau^t(\sigma) - \pi(\sigma)|.$$

This formula represents the *total variation distance*: this is a number between 0 and 1, where 0 means that a deck is perfectly well shuffled, and 1 is very close to what we get before we start shuffling. As we expect, this function is decreasing at each step, meaning that the deck is getting better shuffled. The deck is shuffled enough when the total variation distance falls below some fixed (but somewhat arbitrary) number, for example $1/4$.

We chose $1/4$, but would we get a radically different answer had we chosen $1/3$? For many shuffling techniques, the number of repetitions of the shuffling procedure needed to obtain a total variation distance of $1/2$, $1/3$, or $1/4$ are all very close. So, the exact fraction we choose is not important, as long as we choose a number below $1/2$. This is due to the occurrence of the *cutoff phenomenon*, which is a sharp decrease in the total variation distance, as exhibited on Figure 2.

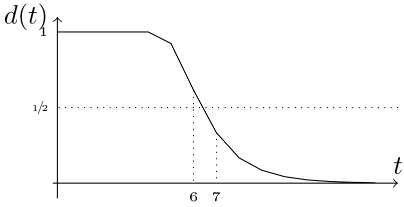

Figure 2: The profile of the total variation distance for the riffle shuffle of a deck of 52 cards exhibits cutoff. After seven shuffles, the total variation distance drops below $1/2$, suggesting that seven iterations of the riffle shuffle should suffice to get a well-shuffled deck of cards.

We can therefore focus our attention to the moment at which this sharp drop occurs: the short time frame in which the distance goes from 1 to 0 is where the deck gets well-mixed, regardless of the arbitrary threshold we chose.

To investigate the cutoff phenomenon in more detail, we express the card shuffling as a mathematical concept, the *Markov chain*, named after the Russian mathematician Andrey Markov (1856–1922).

2 Markov chains and stationarity

A *Markov chain* is a stochastic model describing a sequence of possible events happening at fixed times where the update of an event depends only on the current event but not on the history of previous events. So, the subsequent shuffling of a card deck is a Markov chain, as it is sufficient to know the current ordering to predict the evolution of the card shuffling from now on. The orderings that happened before our last shuffle do not influence the next orderings.

To illustrate the inner workings of a Markov chain, and to formulate our intuitive idea of a well-shuffled card deck in precise mathematical terms, we first give a name to the set of all orderings of 52 cards: We call this set the *state space* S and each element in this set a *state*. The states, so the elements of the set of all card orderings, each represent a particular order of our 52 cards. In total, there are $|S| = 52 \cdot 51 \cdots 2 \cdot 1$ possible states of a card deck.[2]

To formalize the process of a shuffle, we first give an example of a simpler Markov chain with only three states, which we illustrate graphically in Figure 3. Consider a state space S with three states a, b, c and the following set of rules: If at the present time our system is at state a, then it stays in state

[2] The product $52 \cdot 51 \cdots 2 \cdot 1$ is called "52 factorial", and can be abbreviated by 52!.

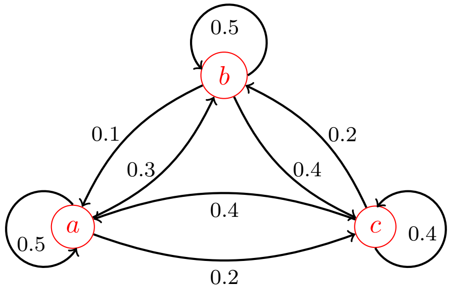

Figure 3: Toy example: Transition probabilities in a system with 3 states.

a with probability $\pi(a,a) = 0.5$ and it moves to state b or c with probability $\pi(a,b) = 0.3$ and $\pi(a,c) = 0.2$, respectively. We call $\pi(i,j)$ the *transition probability* from state i to state j.

The transition probabilities starting in b and c are given analogously as drawn in Figure 3. This determines the transition matrix

$$\pi = \begin{pmatrix} \pi(a,a) & \pi(a,b) & \pi(a,c) \\ \pi(b,a) & \pi(b,b) & \pi(b,c) \\ \pi(c,a) & \pi(c,b) & \pi(c,c) \end{pmatrix} = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.1 & 0.5 & 0.4 \\ 0.4 & 0.2 & 0.4 \end{pmatrix}$$

satisfying that all entries are between 0 and 1 and that the sum of each row is 1. We note that the next state depends only on the current state but not on the previous ones. This property is called *memorylessness* and is the characterizing property of a Markov chain.

We can then calculate the probability that a chain X_0, X_1, X_2, X_3, which starts in the state a at time zero (in formula, $X_0 = a$), attains the value b at time one ($X_1 = b$), followed by $X_2 = a$ and $X_3 = c$, as

$$\mathbb{P}[X_0 = a, X_1 = b, X_2 = a, X_3 = c] = \pi(a,b)\pi(b,a)\pi(a,c).$$

Coming back to our more involved Markov chain, describing the shuffling process of a card deck, we may determine the transition probabilities analogously, by the shuffling mechanism, but the transition matrix is now $52! \times 52!$ entries large, since there are $52!$ combination of cards. We won't write it out here.

Next, let us make precise what we mean by a "well-shuffled card deck": in the probabilistic language, this translates to a Markov chain staying in its *stationary distribution*, or being at least close to it in an appropriate sense. Therefore, to say that a deck cards is "sufficiently randomized" after m repeated shuffles corresponds to measuring the *distance between the distributions* of the cards after m steps of the Markov chain and their stationary distribution, see Section 1.

In the example of the Markov chain with three states, recall Figure 3, the stationary distribution is the uniform distribution, that is, all states occur with the same probability after a long time. Put in formulae, this means that the probability of reaching each state will be $\mu(a) = \mu(b) = \mu(c) = 1/3$ after a long time. We verify this with the equation $\mu(a) = \mu(a)\pi(a,a) + \mu(b)\pi(b,a) + \mu(c)\pi(c,a)$, and analogously for the states b and c.

Consequently, the stationary distribution for the riffle shuffle is one over the probability for each distribution of cards to happen, and, as we said above, we have 52! possibilities for that. We write this in formulae as $\mu(x) = \frac{1}{|S|} = \frac{1}{52!}$ for all states $x \in S$. It is known that, under mild conditions on the transition matrix, a Markov chain on a finite state space has a unique stationary distribution, see [5] for details. Evaluating the mixing quality of the m-th step of the shuffle, or of another Markov chain, corresponds to measuring the distance between the distribution after m steps of the Markov chain and its stationary distribution.

3 Two examples of cutoffs

To summarize our discussion above, a cutoff occurs when the distance to stationarity (also called the *total variation distance*) stays close to 1 for a number of steps and then it suddenly drops and converges very quickly to 0. But this is a fragile process: changing the rules, namely the initial state or the transition matrix of the Markov chain, can break the cutoff. In the following, we look at two Markov chains taking values in $\{0, 1, 2, ..., n\}$, with different cutoff phenomena, namely the classical Ehrenfest model, visualized in Figure 4, and the modified Ehrenfest model, visualized in Figure 5.

3.1 The classical model

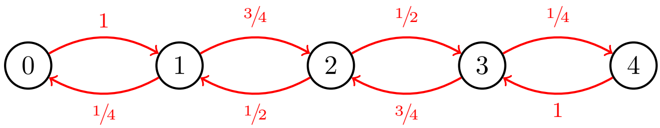

Figure 4: Graph representing the Markov chain associated to the Ehrenfest urns, for $n = 4$ balls.

The model was introduced by Paul Ehrenfest (1880–1933) and Tatyana Ehrenfest-Afanaseva (1876–1964) to study diffusion of gases. Consider two urns and n balls. Assume that we know where the balls are at the beginning. Then, at each step, a ball is chosen at random, among all the balls, and moved to the other urn. For instance, if the chosen ball is in urn 1, then it is moved to urn 2. The state of the associated Markov chain corresponds to the number of balls in urn 1. It goes from i either to $i-1$ if the chosen ball is in urn 1 (this happens with probability $\frac{\#\{\text{balls in urn 1}\}}{\#\{\text{balls}\}} = \frac{i}{n}$), or to $i+1$ if the chosen ball is in urn 2 – which happens with probability $\frac{\#\{\text{balls in urn 2}\}}{\#\{\text{balls}\}} = \frac{n-i}{n}$.

3.2 The modified model

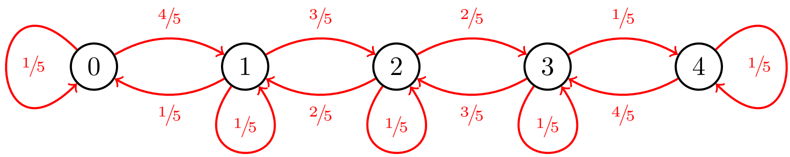

Figure 5: Graph representing the Markov chain associated to the modified Ehrenfest urn, for $n = 4$.

Let us assume that the experimenter is lazy, and sometimes leaves the ball in the same urn, with probability $\frac{1}{n+1}$. To be more precise, it means that the transition probabilities are now:

$$\pi(i, i-1) = \frac{i}{n+1} \quad \pi(i,i) = \frac{1}{n+1} \quad \pi(i, i+1) = \frac{n-i}{n+1}. \qquad (\pi_{\text{mod}})$$

In 1983, David Aldous proved in [1] that this Markov chain has the same stationary distribution as the classical Ehrenfest urn, and converges to the stationary distribution. Moreover, if we assume that all the balls are in urn 2 at the beginning, he proved that it presents a cutoff[3]. It means that he studied the Markov chain $(X_t)_t$ where $X_0 = 0$ and the transition probabilities are given by Formula (π_{mod}), and observed a sudden convergence to the stationarity, as in Figure 2.

Surprisingly, using the same techniques, Persi Diaconis also proved that if the Markov chain starts at $\frac{n}{2}$, then it does not have a cutoff, as explained in [3]. He was considering an even number n of balls and assumed that there are $\frac{n}{2}$ balls in each urn at the beginning. In that setting, the Markov chain $(Y_t)_t$, given

[3] The cutoff occurs at time $\frac{1}{4} n \log(n)$.

by $Y_0 = \frac{n}{2}$ and the transition probabilities in Formula (π_{mod}) above, converges smoothly to the stationary distribution, as we can see in Figure 6.

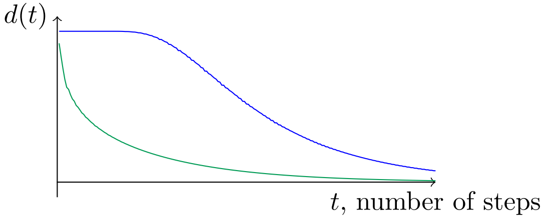

Figure 6: Distance between the distribution of X_t (in blue), or Y_t (in green), and the stationary distribution, for $n = 100$. The random walk X_t exhibits the cutoff phenomenon, while Y_t does not.

4 Outlook: a cutoff criterion

Since Aldous and Diaconis introduced the idea of the cutoff phenomenon, mathematicians have been interested in understanding the Markov chains for which a cutoff phenomenon occurs. Using a distinct approach for each Markov chain, this question is answered for a variety of models, like the riffle shuffle or the Ehrenfest urns.

However, as we have seen for the Ehrenfest model, a slight modification of the setting may result in a drastic change of behaviour. Therefore, understanding the general mechanisms for a cutoff to occur is an important question which would allow also to predict the behaviour for novel models without analysing each model individually from scratch.

Using analytical and probabilistic techniques, several attempts have been made at giving a criterion for cutoff for classes of Markov chains. However, these approaches still do not give a complete picture. For example, in [6], cutoff is established for Markov chains satisfying certain geometric properties, including random walks on certain “nice” graphs. Similarly in [4], techniques are evolved to provide a condition for the cutoff of the class of “weakly asymptotically simple exclusion processes”, which are particle processes on a line that jump to the left and the right with slightly different probability. These processes display a cutoff phenomenon if the segments of the line gets finer and finer. In that case, upper and lower bounds for the time at which the model comes close to equilibrium are analysed separately. However, the results cover only some classes of Markov chains. In many cases, one can only conjecture the existence of a cutoff.

Therefore, a general criterion for cutoff applicable to all Markov chains is still missing, and the problem remains widely open for the future.

5 Acknowledgements

We would like to thank Charles Bordenave, Persi Diaconis, Hubert Lacoin, and Justin Salez for organizing a seminar on the cutoff phenomenon for finite Markov chains, and for the opportunity to write this snapshot. We also thank the snapshot editors for their helpful comments on the exposition.

References

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

[2] D. Bayer and P. Diaconis, *Trailing the dovetail shuffle to its lair*, The Annals of Applied Probability **2** (1992), no. 2, 294–313.

[3] P. Diaconis, *The cutoff phenomenon in finite Markov chains*, Proceedings of the National Academy of Sciences of the United States of America **93** (1996), no. 4, 1659–1664.

[4] C. Labbé and H. Lacoin, *Mixing time and cutoff for the weakly asymmetric simple exclusion process*, The Annals of Applied Probability **30** (2020), no. 4, 1847–1883.

[5] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, With a chapter by James G. Propp and David B. Wilson.

[6] Justin Salez, *Cutoff for non-negatively curved Markov chains*, Journal of the European Mathematical Society (2023).

Isabelle Baraquin *'s research activity is hosted at the Laboratoire de mathématiques de Besançon.*

Nadia Lafrenière *is an Assistant Professor at Concordia University.*

Katharina Schuh *is a Postdoc at the Institute of Analysis and Scientific Computing at TU Wien.*

Mathematical subjects
Probability Theory and Statistics

Connections to other fields
Physics, Humanities and Social Sciences

License
Creative Commons BY-SA 4.0

DOI
10.14760/SNAP-2023-005-EN

Snapshots of modern mathematics from Oberwolfach provide exciting insights into current mathematical research. They are written by participants in the scientific program of the Mathematisches Forschungsinstitut Oberwolfach (MFO). The snapshot project is designed to promote the understanding and appreciation of modern mathematics and mathematical research in the interested public worldwide. All snapshots are published in cooperation with the IMAGINARY platform and can be found on www.imaginary.org/snapshots and on www.mfo.de/snapshots.

ISSN 2626-1995

Junior Editors
Carina Geldhauser
Anup Anand Singh
junior-editors@mfo.de

Senior Editor
Anja Randecker
senior-editor@mfo.de

Mathematisches Forschungsinstitut
Oberwolfach gGmbH
Schwarzwaldstr. 9–11
77709 Oberwolfach
Germany

Director
Gerhard Huisken

Mathematisches
Forschungsinstitut
Oberwolfach

Member of
Leibniz
Leibniz
Association

IMAGINARY
open mathematics