Snapshots of modern mathematics
from Oberwolfach

№ xx/xxxx

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

Isabelle Baraquin • Nadia Lafrenière
Katharina Schuh

We study random processes, including shuffling a deck of cards, and what happens after a long time. For example, when doing the riffle shuffle on a deck of cards, we see a sudden transition from an unmixed to a well-mixed deck. The so-called *cutoff phenomenon* is also detected for Markov chains arising in a variety of contexts.

1 A game with cards

Whenever you want to play cards, a natural question arises: how long one has to shuffle the deck until the cards are well mixed? One clearly does not want to have a tidy deck of cards, otherwise it is easy to guess which card is the next one. But you also do not want to spend too much time shuffling the deck. So, the question becomes: what is the minimal number of times one has to repeat the shuffle to get something random?

To be able to answer this question, we need to be more precise: specify the size of the deck, choose a shuffling technique and explain what it means to be close to random. For the moment, let us say that the deck is close to random if all the possible orders are almost equally likely.

Mathematicians look at our question with the help of probability theory. For instance, the familiar overhand shuffle, letting small bunches of cards dropping from one hand to the other, is believed to take thousands of shuffles to be effective on a standard deck with 52 cards!

Now consider a 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 (see Figure 1). This technique, called the *riffle shuffle*, is known to be very efficient, meaning we would have to shuffle less than ten times [3], and is used by casual card players and casino dealers along.

Figure 1: The step of interlacing the two smaller decks into one deck of the original size.

When we try to quantify the number of times we must repeat the shuffle to get something really random, it is interesting to look at some features. Since the quantity of information provided by the deck is hard to analyze, we look at one piece of information, and figure out if the behaviour is really random. A good example of this for the riffle shuffle is the number of rising sequences. These are sequences of cards that were adjacent and in a certain order, that are still in that order once the deck is mixed (see Figure 2).

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 2. 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 2 to get n). That shows that it is *impossible* to get a well-mixed deck of cards in less than $\log_2(n)$ times, but showing that a given number of times is sufficient is also very hard. We describe a way in the next paragraph.

---

[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.

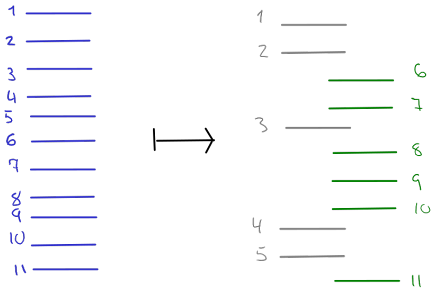

Figure 2: On this 11-card deck, the rising sequences are pictured in two different columns, each column corresponding to one hand in the riffle shuffle. Each time we shuffle, the number of rising sequences is at most multiplied by 2.

Mathematicians have a way to quantify how close a deck is from being perfectly mixed. Since this depends on the number of times we have repeated the shuffling procedure, we express the randomness as a function of the number of iterations. To compute this function, we compare the probability of each ordering of the deck happening following our shuffling technique with the probability of that ordering if the perfect mixing had happened.

We obtain 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: $\frac{1}{4}$, for example.

We chose $\frac{1}{4}$, but would we get a radically different answer had we chosen $\frac{1}{3}$? For many shuffling techniques, that has absolutely no importance, as long as we choose a number below $\frac{1}{2}$! This is due to the occurrence of the *cutoff phenomenon*, which is a sharp decrease in the total variation distance, as exhibited on Figure 3.

On the graph in Figure 3, the number of repetitions of the shuffling procedure needed to obtain a total variation distance of $\frac{1}{2}$, $\frac{1}{3}$ or $\frac{1}{4}$ are all very close. 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.

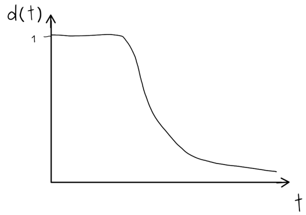

Figure 3: A typical profile of the total variation distance, for a chain that exhibits cutoff

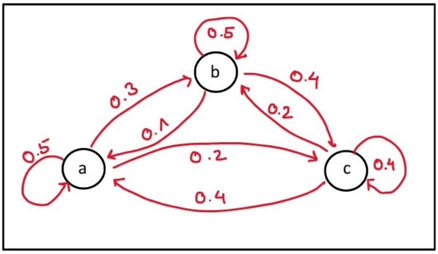

Figure 4: Toy example: Transition probabilities

2 Markov chains and stationarity

Next, we express the card shuffling in a mathematical concept and we introduce the notion of a Markov chain.

As the card shuffling is already a quite involved model, we first introduce the definition of a Markov chain via an explicit toy model. As illustrated in Figure 4, we consider a state space S with three states a, b, c and a process given in the following way: If the process is currently at state a, then it stays in state 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. The transition probabilities starting in b and c are given analogously as drawn in Figure 4. 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 resulting process is memoryless, i.e., the next state depends only on the current state but not on the previous ones. We call a process satisfying this property a *Markov chain.* Then, if the chain starts in the state a, the probability that the chain X_0, X_1, X_2, X_3 attains for instance the values (a, b, a, c), in that order, is given by

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

The concept of Markov chains generalizes to card shuffling and to other models. For card shuffling, the state space S is the set of all orderings of 52 cards, which are $|S| = 52 \cdot 51 \cdots 2 \cdot 1$ states.[2] As we saw in the case of the toy example, 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. Hence, the card shuffling forms a Markov chain, and its transition probabilities are determined by the shuffling mechanism.

As with card shuffling, one is often interested in how a Markov chain evolves and whether one can obtain a similar phenomenon as the abrupt transition from 'not shuffled' to 'well shuffled'. In the probabilistic language a 'well shuffled card deck' translates to a Markov chain staying in *stationarity.* For the toy example, illustrated in Figure 4, the stationarity is the uniform distribution, i.e., all states occur with the same probability $\mu(a) = \mu(b) = \mu(c) = 1/3$ 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.

As one would intuitively guess, the stationarity for the riffle shuffle is also the uniform distribution, i.e., $\mu(x) = \frac{1}{|S|} = \frac{1}{52!}$ 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 [6]. 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 stationarity.

3 Two examples

As it was already said, 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 and converges very quickly to 0[3]. 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.

Let us present two Markov chains taking values in $\{0, 1, 2, ..., n\}$, one of which has a cutoff unlike the other.

---

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

[3] To be more precise, it converges exponentially fast to 0.

3.1 The classical model

The model was introduced by Paul and Tatiana Ehrenfest 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 $\frac{\#\{\text{balls in urn } 2\}}{\#\{\text{balls}\}} = \frac{n-i}{n}$.

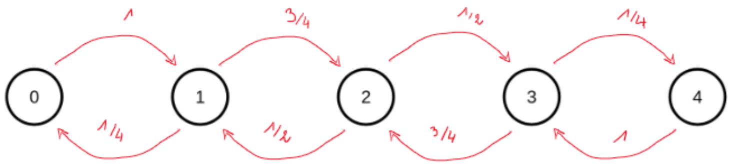

Figure 5: Graph representing the Markov chain associated to the Ehrenfest urns, for $n = 4$ balls

Mathematicians were able to show that this Markov chain admits a stationary distribution, which is the binomial distribution of parameters n and $\frac{1}{2}$, already seen in Footnote 1. However, in order to study the convergence of the Markov chain, we need to rule out a small periodicity problem, and slightly modify the transition probabilities.

3.2 The modified model

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

$$\pi(i, i-1) = \frac{i}{n+1} \quad \pi(i, i) = \frac{1}{n+1} \quad \pi(i, i+1) = \frac{n-i}{n+1}. \tag{1}$$

Mathematicians were able to show that this Markov chain has the same stationary distribution as the classical Ehrenfest urn, and converges to its stationarity.

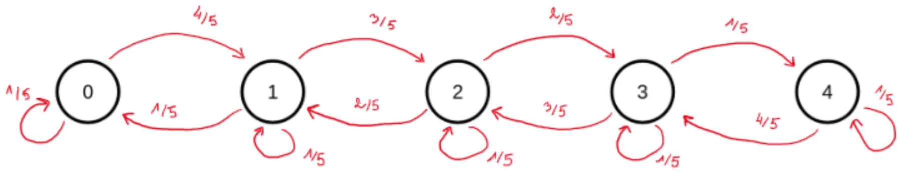

Figure 6: Graph representing the Markov chain associated to the modified Ehrenfest urn, for $n = 4$

Moreover, if we assume that all the balls are in urn 2 at the beginning, they proved that it presents a cutoff[4] [4]. It means that they studied the Markov chain $(X_m)_m$ where $X_0 = 0$ and the transition probabilities are given by (1), and observed a sudden convergence to the stationarity, as in Figure 3.

Surprisingly, using the same techniques, they also proved that if the Markov chain starts at $\frac{n}{2}$, then it does not have a cutoff. They were then 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 $(Y_m)_m$, given by $Y_0 = \frac{n}{2}$ and the transition probabilities in (1), converges smoothly to the stationary distribution, as we can see in Figure 7.

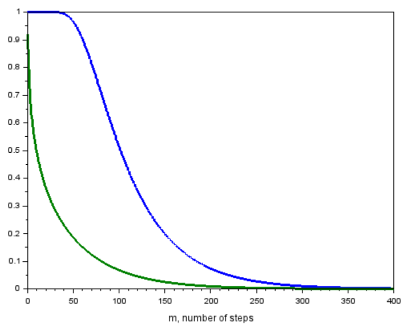

Figure 7: Distance between the distribution of X_m (in blue), or Y_m (in green), and the stationary distribution, for $n = 100$. The random walk X_m exhibits the cutoff phenomenon, while Y_m does not.

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

4 Criterion for cutoff

Since the idea of the cutoff phenomenon was introduced in the 80's by Aldous and Diaconis [1, 2, 4], mathematicians have been interested in understanding for which Markov chain 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. Furthermore, using analytical and probabilistic techniques [5, 7], several attempts have been made at giving a criterion for cutoff for classes of Markov chains. However, a general criterion for all Markov chains is still missing, and the problem remains widely open for many models.

Image credits

Figure 1 Photo of card shuffling by LDC, released to public domain, on Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Card_shuffling.jpg

References

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

[2] David Aldous and Persi Diaconis, *Shuffling cards and stopping times*, Amer. Math. Monthly **93** (1986), no. 5, 333–348, ISSN 0002-9890.

[3] Dave Bayer and Persi Diaconis, *Trailing the dovetail shuffle to its lair*, The Annals of Applied Probability **2** (1992), no. 2, 294–313, ISSN 1050-5164.

[4] Persi Diaconis, *The cutoff phenomenon in finite Markov chains*, Proc. Nat. Acad. Sci. U.S.A. **93** (1996), no. 4, 1659–1664, ISSN 0027-8424.

[5] Cyril Labbé and Hubert Lacoin, *Mixing time and cutoff for the weakly asymmetric simple exclusion process*, Ann. Appl. Probab. **30** (2020), no. 4, 1847–1883, ISSN 1050-5164.

[6] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, Providence, RI, 2009, ISBN 978-0-8218-4739-8, With a chapter by James G. Propp and David B. Wilson.

[7] Justin Salez, *Cutoff for non-negatively curved Markov chains*, arXiv:2102.05597 (2021).

Isabelle Baraquin *is an Attaché Temporaire d'Enseignement et de Recherche of mathematics at École Nationale Supérieure de Mécanique et des Microtechniques and Laboratoire de mathématiques de Besançon.*

Nadia Lafrenière *is a research instructor at Dartmouth College.*

Katharina Schuh *is a PhD student at the Institute for Applied Mathematics at the University of Bonn.*

Mathematical subjects
will be filled out by the editors

Connections to other fields
will be filled out by the editors

License
will be filled out by the editors

DOI
will be filled out by the editors

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 Editor
will be filled out by the editors
junior-editors@mfo.de

Senior Editor
Sophia Jahns
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