Shuffling Cards

1 Some card shuffling methods

Let's assume here that we have n cards, where they are labelled $1, 2, \ldots, n$ (where we will mostly think of a standard deck with $n = 52$).

One of the most standard methods of shuffling cards is the *Riffle shuffle*, which, roughly speaking, repeats the following two-step procedure:

(i) cut the deck into two (possibly unequal) parts;

(ii) interleave cards (somehow at random, taking some random number of cards at a time from each half in the order in which they appear in that half – one could define this more precisely, but we will not do this) from the two parts to produce a new deck.

Another method of shuffling one could use is the *Random-To-Top shuffle*: At each step, take a randomly chosen card from the deck and insert it at the top of the deck (we are allowed to choose a top card as a random card and place it on top, and in this case, the order of cards just stays the same). Or we could use, in some sense, the opposite of this, which is the *Top-To-Random shuffle*, which at each step picks a card from the top and moves it to a random location (equally likely to the bottom, top or between any two consecutive cards). Both of these seem a bit silly and probably much less efficient than *Riffle shuffle*, as they only move one card at a time, but mathematically it is often interesting to first study some simple models, and one could expect that these are easier to analyse formally.

Which of the above methods would you use to shuffle cards? Could you guess how many shuffles one needs to shuffle the deck well for each of these methods?

Roughly speaking, we want to think about developing some theory which would allow us to understand, or even be able to prove, which method of shuffling works better and how efficient they are.

2 What is a well or even perfectly shuffled deck?

Before we come up with any sensible formal setup, let's consider the following question:

Example 2.1. Alice and Bob start with an ordered deck of 52 cards each. Alice is allowed to perform one random-to-top shuffle on her deck, and Bob is allowed to perform one riffle shuffle on his deck. Alice then asks Bob to guess the value of the top card of her deck, and

Bob asks Alice to guess the top card of his deck. Assuming they both guess "optimally", who is more likely to guess correctly? Which card should Alice guess? What about Bob?

In a riffle shuffle, Bob splits the deck into two, which means that at the top of one side is card 1 and at the top of the other is some card with label which is likely close to $\frac{n}{2}$ and equally likely smaller or larger than it. We have not defined a riffle shuffle precisely, but it makes sense that there is a $\frac{1}{2}$ chance that, once the cards from two halves are interleaved, the card with label 1 ends up at the top of the deck. So if Alice guesses 1, she has a $\frac{1}{2}$ chance to guess correctly, which is not too bad. On the other hand, at the top of Alice's deck is a uniformly randomly picked card. So whatever Bob guesses, there is a chance $\frac{1}{52}$ for him to be correct. Does this mean that a random-to-top shuffle is better than a riffle shuffle? Intuitively, that should really NOT be the case. But we have just shown that if you want to play a "guess the top card game", a riffle shuffle performs worse than a random-to-top shuffle, as one has a lot lower chance of guessing after the random-to-top shuffle. And a smaller chance of making a correct guess should suggest that the deck is shuffled better, at least intuitively. But this is just one game, so of course this should not mean that a riffle shuffle is worse overall; it is just worse at "making the top card random". And indeed, if instead of guessing the top card, Alice and Bob were guessing the bottom card, Alice would have a $\frac{1}{2}$ chance of guessing correctly if she guessed 52, but Bob would have a $1 - \frac{1}{52}$ chance of guessing correctly if he guessed 52. So for "guess the bottom card game" riffle shuffle is a lot better, as with random-to-top the bottom card is extremely likely to stay 52.

Similarly, one could consider a "guess the kth card game" for any k and it is natural to think that the deck is not very well shuffled if the probability of guessing the kth card is really large. On the other hand, it is also not possible that for all k this probability is incredibly small. For example, if my guess for the kth card is a card with label 1, for each k, the probability $p_k(1) = \mathbb{P}(k\text{th card has label } 1)$ has to satisfy that $\sum_{k=1}^{52} p_k(1) = 1$ as the kth card has to be somewhere and it can not be in more than one place in any given deck. So it would make sense that in a "perfectly shuffled deck" $p_k(1) = \frac{1}{52}$, as if the deck is perfectly shuffled, we should not expect to have 1 in any specific location more likely than in any other specific location. If we set, for $i \leq 52$, $p_k(i) = \mathbb{P}(k\text{th card has label } i)$ in a "perfectly shuffled deck" one would expect $p_k(i) = \frac{1}{52}$. So should one say that a deck is "perfectly shuffled" if for all k and i, $p_k(i) = \frac{1}{52}$?

Example 2.2. Consider a deck obtained by starting from an ordered deck of 52 cards, picking a number k uniformly from $1, 2, \ldots, 52$ (i.e. picking any of the numbers equally likely) and then moving first k cards to the end of the deck but keeping their relative order, so that one obtains $k+1, k+2 \ldots, 52, 1, 2, 3, \ldots k$. Is this deck "well shuffled"? What is the probability that the card 1 appears in this deck before the card 2?

The value of previously defined $p_k(i) = \frac{1}{52}$ for this deck. So the deck is as "well shuffled" as it could be for playing the "guess the kth card game". On the other hand, the probability that 1 appears before 2 is $1 - \frac{1}{52}$, as that only fails to happen if we have picked $k = 1$. In a "perfectly shuffled deck" there is no reason to expect to see 1 before 2 more likely than 2 before 1, and surely we would have that the probability of seeing 1

before 2 should be $\frac{1}{2}$. So $p_k(i) = \frac{1}{52}$ is definitely not enough to determine if the deck is "perfectly or well shuffled". In a "perfectly shuffled deck", we would expect to see card i before card j with probability $\frac{1}{2}$ for any i and j, and we would also expect to see i before j before k equally likely for any i, j, k, and as there are $3! = 6$ ways to order 3 numbers the probability in a "perfectly shuffled deck" to see cards i, j, k in exactly that order should be $\frac{1}{6}$. We can now think about this with 4 cards, or 5 cards or more. And in fact, surely in a "perfectly shuffled" deck we would just expect to see any order of the 52 cards equally likely. Of course, the only sensible definition for what it means for a deck to be "perfectly shuffled" is that it is a "uniform deck", i.e. a (random) deck of cards in which any order of cards is equally likely. There are 52! possible ways in which one can order 52 cards. So a "perfectly shuffled" or a "uniform deck" of cards is a deck of cards which are ordered at random, such that each possible order appears with probability $\frac{1}{52!}$. Or in other words, to get a "uniform or perfectly shuffled deck" one can just write all 52! possible orders and then pick any one of them with equal chance. The only problem with this is that 52! is an insanely large number (it has 68 digits), so of course nobody can write down all 52! orders and then pick one.

Surely we can get a perfectly shuffled deck of cards without having to think of insanely large numbers like 52!. For example, to get a "perfectly shuffled, i.e. uniform deck", one could do the following: Start with an ordered deck. Pick one card from it uniformly at random (any card with probability $\frac{1}{52}$) and place it on the side. The original deck now has 51 cards, and you pick another card from it uniformly at random (each card with probability $\frac{1}{51}$). Place the second picked card over the previously picked card. Now pick a card uniformly from the remaining 50 cards and place it on top of the two already picked cards. Continue doing this. After 52 steps, you will get one deck with 52 cards in which any order is equally likely. Indeed, what is the probability that we get a deck $a_1, a_2, \ldots, a_{52}$? (Here $a_1, a_2, \ldots, a_{52}$ is just some fixed ordering of $1, 2, \ldots, 52$.) This means we have picked a_{52} as the first card, which has a chance $\frac{1}{52}$; then we have picked a_{51} from the 51 remaining cards, so with the chance $\frac{1}{51}$ and so on, and the overall probability is $\frac{1}{52 \cdot 51 \cdot \ldots \cdot 2 \cdot 1} = \frac{1}{52!}$, i.e. we get any order equally likely. So we can get a truly perfectly shuffled deck by doing 52 "moves", which is a lot better than writing all 52! options, but is still a fairly long process. And surely nobody ever shuffles cards like this in practice.

So far we have understood that to shuffle cards "perfectly" we want to get the uniform deck. And we understand that the uniform deck means any of the 52! combinations equally likely. But if you are playing cards, you probably do not care to have an absolutely perfectly uniform deck. Surely, if a couple of permutations have a slightly higher or lower chance of appearing than the incredibly small chance $\frac{1}{52!}$, one would not really practically be able to spot the difference. So let's try to think about the following natural questions: How many times should we repeat the Riffle/Random-to-Top/Top-to-Random shuffle such that, starting from an ordered deck, we reach a (sufficiently) "well-shuffled" deck, i.e. to have a deck which is sufficiently close to the uniform deck? How close to the uniform deck is the deck we have after k shuffles of an ordered deck? Is any of these methods of shuffling more or less efficient than any other method we could think of? (Or in other words, does any get sufficiently close to the uniform deck after fewer or more shuffles than if we use some other method?)

To answer any of these questions formally, we first want to understand what the questions are actually asking precisely. So we now need to think about what a good definition of "well shuffled" is, before we can say anything about how quickly the deck is "well shuffled". We understand that "well shuffled" should mean close to the uniform deck in some way. And now we want to think of a sensible definition of a deck of cards being close to the uniform deck. (Here, whenever we say a deck of cards, we really talk about a random deck, i.e. it is not a specific deck, but rather a deck which has various probabilities of having cards in any specific order.) We can first be more specific about how close we want to be to the uniform deck and say that our target is to be ε-close to it, or ε-well shuffled, for some $\varepsilon > 0$ (in maths, ε is essentially always used to denote a small number).

Maybe a sensible definition could be to say that the deck is ε-well shuffled if, for any ordering of cards, the probability that the deck takes exactly that order is at least $\frac{1}{52!} - \varepsilon$ and at most $\frac{1}{52!} + \varepsilon$, i.e. ε away from $\frac{1}{52!}$. The problem with this definition is that if we take $\varepsilon = \frac{1}{52!}$, which is an insanely small number, a deck where half of the options have probability $\frac{2}{52!}$ of appearing and half have probability 0 would be according to this definition $\frac{1}{52!}$-well shuffled. So, for example, we could make all orderings where 1 is before 2 have probability $\frac{2}{52!}$ and the ones with 2 before 1 have probability 0 (this is valid as we know 1 is before 2 in exactly half of the orderings). This means that if we play a game of "guess if 1 is before 2" we should always guess "Yes, it is" and we will certainly be correct, but for the uniform deck the answer is yes only with a $\frac{1}{2}$ chance. So, even though the probability for the random deck to take any specific permutation is incredibly close to the one we want, somehow looking at many possible permutations at once, this tiny probability can add up to something significant, and we end up having a deck which is a lot worse than uniform one in some scenarios (in our example, it has an "error" of $\frac{1}{2}$ compared to the uniform deck in the "guess if 1 is before 2 game").

It seems that to call the deck ε-well shuffled, we would want the deck to have a small "error" (maybe $\leq \varepsilon$?) when compared to the uniform deck, for any possible "game we could play". So why not just make this precisely the definition of being ε-well shuffled? More precisely, we could say that a (random) deck of cards is ε-well shuffled if the probability of any question we could possibly ask about any deck (i.e. the probability of winning any possible game) differs by at most ε when we ask it for our (of course, random) deck and for the uniform deck. So we have just come up with a definition of some sort of distance on random orders of cards. This is known in probability as a total variation distance, and it can be defined in a more general setting for any two distributions, but we will just define it in the context of cards here.

Definition 2.1 (Total variation distance)**.** Let n be a positive integer and let X and Y be two random variables taking values in the set S_n, which is the set of $n!$ elements which are all possible permutations of the numbers $1, 2, 3 \ldots, n$. (In other words, X and Y are random decks of cards.) We define the total variation distance, or the TV distance, of the distributions of X and Y as follows:

$$\| \mathbb{P}(X \in \cdot) - \mathbb{P}(Y \in \cdot) \|_{\mathrm{TV}} = \max_{A \subseteq S_n} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|$$

Let us unpack what this definition means. Say $n = 52$ as usual, then S_n is this enormous set of all 52! orders of cards. A possible $A \subset S_n$ is, for example, a set of one element, which is just the increasing order of cards, i.e. $\{(1, 2, \dots, 52)\}$. Then, if the total variation distance of two decks is at most ε that would mean that for this set A the difference of probabilities that decks are in that set, i.e. as it is a singleton set, this would just be the difference of probabilities that decks are exactly $1, 2 \dots, 52$, has to be at most ε. But we could also consider a set $A' \subset S_n$ which is $\{(1, 2, \dots, 52), (52, 51, 50, \dots, 2, 1)\}$ i.e. set of two orderings where one is increasing and the other is decreasing. So for decks X and Y to have total variation distance less than ε we would need

$$|\mathbb{P}(X \in A') - \mathbb{P}(Y \in A')| = |\mathbb{P}(X = 1, \dots, 52 \text{ or } 52, \dots, 1) - \mathbb{P}(Y = 1, \dots, 52 \text{ or } 52, \dots, 1)| \leq \varepsilon.$$

We would need to have the same for subsets of S_n which have 3 possible orderings in them, or 4, or 50! or really any number which is at most 51!. So to call two decks ε-close, we want their total variation distance to be at most ε, which means that the difference in probabilities that the two decks take values from any specific subset is at most ε. So does that mean that if we played some of the previously mentioned games, with two decks X and Y which have total variation distance $\leq \varepsilon$, we would get probabilities which differ by at most ε? Let's consider a game of "guess the kth card" and assume we are guessing it is i. We would like X and Y to satisfy $|\mathbb{P}(k\text{th card in } X \text{ is } i) - \mathbb{P}(k\text{th card in } Y \text{ is } i)| \leq \varepsilon$. Hopefully, this would be implied by the total variation distance being at most ε. So we would like to know if kth card in X is i can be written as $X \in A$ for some $A \in S_n$? Of course, it can. Just take A to be the set of all 51! orderings which have card i in location k and the remaining 51 cards in any possible order on the remaining 51 locations! I can, of course, not write this set down, as it is too large, but if $k = i = 1$ it should look like this $A = \{(1, 2, \dots, 52), (1, 2, \dots, 50, 52, 51), \dots, (1, 52, 51, \dots, 2)\}$ i.e. its elements are all things that look like (1, any possible order of $2, \dots 52$).

What about playing the game "is 1 before 2"? Take A to be the set of all orderings where 1 is before 2, i.e. all $\frac{52!}{2}$ orderings which look like (anything 1 anything 2 anything). Then if total variation distance of decks X and Y is at most ε we have that

$$|\mathbb{P}(1 \text{ is before } 2 \text{ in } X) - \mathbb{P}(1 \text{ is before } 2 \text{ in } Y)| = |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| \leq \varepsilon.$$

Similarly, if the total variation distance is $\leq \varepsilon$ any game we could think of would have the difference in probabilities of specific scenarios for the two decks at most ε. Evaluating the probability that a specific scenario in a game holds for some deck is actually just evaluating the probability that the deck takes one of the values in a set A, where set A is picked to be made of all orderings where exactly that scenario holds.

We can *FINALLY* define what it means for the deck to be ε-well shuffled.

Definition 2.2 (ε-well shuffled)**.** Let n be a positive integer and let X be a random variable taking values in the set S_n, which is the set of $n!$ elements which are all possible permutations of the numbers $1, 2, 3 \dots, n$ and let Y be a uniform random variable on S_n. (In other words, X is a random deck of cards and Y is the uniform deck.) We say that X is ε-well shuffled if the total variation distance between (the distributions of) X and Y is at most ε.

3 How long should we shuffle a deck of cards?

Definition 3.1 (ε-well shuffling time)**.** For a given shuffling method, we say that the ε-well shuffling time is the smallest number of shuffles we need, t, such that, starting from any possible deck, the random deck we have after t shuffles is ε-well shuffled. More precisely, if we call X_0 the starting deck, and call X_1 the deck after 1 shuffle, X_2 after $2\ldots$ we say that the ε-well shuffling time is the smallest t for which X_t is ε-well shuffled, for any starting value of X_0.

We can actually notice that for any starting deck, we could just choose to relabel the cards such that they are in order, so in the above definition we could actually just start from the ordered deck. The above is actually known as the mixing time, and it has a lot broader context and can be considered in the case of more general random processes (known as Markov chains), not just card shufflings. It is a famous result that for a riffle shuffle, with a suitable formal definition, one needs 7 shuffles to make the deck $\frac{1}{4}$-well shuffled. This can be proved formally, and in fact, in the case of a deck of n cards it can also be shown that one needs an order of $\log n$ shuffles, which is a lot smaller than the total number of cards. Unfortunately, this proof is difficult, and we will of course not be able to prove this now. In fact, to understand the well-shuffling time for any deck fully formally, we would need to develop a few more results for which we do not have time now. But we will give an intuitive explanation or a "sketch of a proof" that the well-shuffling time of the random-to-top shuffle is of order $n \log n$ for a deck of n cards.

4 Shuffling time of random-to-top shuffle

Start from a deterministic deck $X_0 = 1, 2, \ldots n$. Perform a random-to-top shuffle to get a deck X_1. How does deck X_1 look? For $n = 52$ it is equally likely (probability $\frac{1}{52}$) any of $(1, 2, \ldots 52), (2, 1, 3, \ldots 52), (3, 1, 2, 4, \ldots, 52), \ldots, (51, 1, 2, \ldots, 50, 52), (52, 1, 2, \ldots, 50, 51)$. What we have already noticed is that the top card is uniform, and the rest is not at all. Now make a second random-to-top shuffle to get X_2. What could happen (with a $\frac{1}{52}$ chance) is that we have taken the top card, so we have the deck X_1 again, but if not, we have taken a new card and put it on top. Let us assume we have put a new card on top (which is actually fairly likely). In that case, what can we say about the top two cards? Well, the second card is the previously first card in X_1, and that was a uniform card, but now the first card is uniform from the set of all other cards. So the top two cards are just obtained by picking one uniform card and then another one uniformly at random from the rest. The top two cards are then just a uniform order of two uniform cards. Let's say we continue shuffling the deck until, at some point, we place on top a card which was not a first or a second card in X_2 (until then we have actually just been picking a uniform card from the top two cards of X_2 and reshuffling them, but throughout this process the top two cards actually stayed uniform cards in uniform order). At that point, our top three cards will actually be three uniform cards and have uniform order. We can continue like this. At some point in time, we will arrive at the situation where each card was touched

(i.e. moved from somewhere to the top) at least once. Intuitively, it is not too hard to convince yourself that at this point the deck is uniform. Indeed, at every point in time the set of previously touched cards is uniform and they are in uniform order and when we add a new card to that set, it is true that it is added on top, but the cards which are already there are uniform, their labels are not fixed, and we are now just adding another uniform card, so the set of touched cards is uniform but so is their order. However, it is more difficult to convince yourself that this can be used to bound the well-shuffling time. Indeed, we have found a time when the deck is well shuffled, as it is exactly uniform. But this time is random! It is not a fixed time; it is the first time when all cards have been touched, which depends on how we perform the shuffling. Let's call this random time, i.e. the first time we have picked all the cards at least once, τ. It is natural to expect that finding the expectation of τ would lead to an upper bound of the well-shuffling time (and one can show all of this very formally, with a few more spare hours and slightly more advanced probability knowledge). Our goal for now is to find $\mathbb{E}[\tau]$. We could try to do this from the definition, or we can try to study some random variables which are easier to understand and hope to relate them to τ. Notice that in the first step we always touch a new (first touched) card. Then in the second step, we either touch a new (second-touched) card with probability $1-\frac{1}{52}$, as it happens when we pick any card other than the top card, or we touch the first-touched card again. Let us consider τ_2 the time (here, time always means just the number of steps) it takes between touching the first card and touching the second card. $\tau_2 = 1$ with probability $1 - \frac{1}{52}$ and it is 2 with probability $\frac{1}{52}(1 - \frac{1}{52})$ (as we touch the first touched-card in the second step and then touch a new card in the third) and it is 3 with probability $\frac{1}{52^2}(1 - \frac{1}{52})$ and so on.... i.e. it is a geometric random variable with parameter $1 - \frac{1}{52} = \frac{51}{52}$. Now we can think of τ_3, which is the time it takes between touching the second-touched card and a new (third-touched) card. Similarly, this will be a geometric distribution with parameter $\frac{50}{52}$ (as we fail to get a new card if we take one of the two previously touched cards and succeed otherwise). Similarly, if τ_k is the time it takes between the $(k-1)$-th-touched card and the k-th-touched card, we can see that its distribution is geometric with parameter $\frac{52-k+1}{52}$. Finally,

$$\mathbb{E}[\tau] = \mathbb{E}[\tau_1 + \ldots + \tau_{52}] = 1 + \frac{52}{51} + \frac{52}{50} \ldots + \frac{52}{1}$$

where $\tau_1 = 1$ as we always get the first new card in the first step and we are using that the expectation of a geometric with parameter p is $\frac{1}{p}$ (which can be found using the "differentiation of a sum idea" and you should show it!). If we get back to the case of a general n cards, the same logic gives

$$\mathbb{E}[\tau] = \mathbb{E}[\tau_1 + \ldots + \tau_{52}] = \mathbb{E}[\tau_1] + \ldots + \mathbb{E}[\tau_{52}] = 1 + \frac{n}{n-1} + \frac{n}{n-2} \ldots + \frac{n}{1} \approx n \log n$$

where we are using that the sum of $\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} \ldots + \frac{1}{1} \approx \log n$ which one can estimate by thinking of how this bound compares to the integral of $\frac{1}{x}$ with suitable limits for lower and upper bound (try doing it formally!).

In fact, it is not hard to convince yourself that times $\tau_1, \tau_2, \ldots$ do not depend on each other. By this we mean that regardless of whatever the exact values of times $\tau_1, \tau_2, \ldots, \tau_k$ are, the time τ_{k+1} has exactly the same geometric distribution. In the case of independent

random variables, the variance also satisfies the linearity property (check this!). And so we leave it as an exercise to compute the variance of geometric (differentiation of sums again?) and get that

$$\mathrm{Var}(\tau) = \sum_{i=1}^{n} \mathrm{Var}(\tau_i) \leq \frac{n^2}{n^2} + \frac{n^2}{(n-1)^2} + \ldots + \frac{n^2}{1^2} \leq 2n^2$$

where we can now upper bound the sum of $\frac{1}{k^2}$ by 1+ intergal of $\frac{1}{x^2}$ from 1 to infinity.

We have now found that $\mathbb{E}[\tau] \approx n \log n$ and $\mathrm{Var}(\tau) \leq 2n^2$. But remember that τ is a random time, so knowing its expectation does not by itself tell us that it is usually close to $n \log n$. We now want to find some way of turning information about the mean and variance into statements about the probability that τ is unusually large.

The first such result is Markov's inequality. It says, roughly, that a non-negative random variable with a small mean cannot be extremely large with high probability.

Proposition 4.1 (Markov's inequality)**.** *Let τ be a non-negative random variable. Then, for every $a > 0$,*

$$\mathbb{P}(\tau \geq a) \leq \frac{\mathbb{E}[\tau]}{a}.$$

Proof. If we define $\mathbf{1}\{\tau \geq a\}$ to be a random variable which is 1 when $\tau > a$ and 0 otherwise (this is called indicator, as it "indicates" if something is true),

$$\tau \mathbf{1}_{\{\tau \geq a\}} \geq a \mathbf{1}_{\{\tau \geq a\}}.$$

Taking expectations gives

$$\mathbb{E}[\tau] \geq \mathbb{E}\big[a\mathbf{1}_{\{\tau \geq a\}}\big] = a\mathbb{E}[\mathbf{1}\{\tau \geq a\}] = a\mathbb{P}(\tau \geq a)\,,$$

as $\mathbf{1}\{\tau \geq a\}$ is 1 with probability $\mathbb{P}(\tau \geq a)$ and is 0 otherwise, and hence

$$\mathbb{P}(\tau \geq a) \leq \frac{\mathbb{E}[\tau]}{a}.$$

□

For example, taking $a = 2\mathbb{E}[\tau]$ gives

$$\mathbb{P}(\tau \geq 2\mathbb{E}[\tau]) \leq \frac{1}{2}.$$

Thus Markov tells us that τ cannot be much larger than its mean too often. However, this bound is quite crude. We have also gone to the trouble of calculating the variance of τ, and we can use this extra information to get a stronger statement.

Chebyshev's inequality formalises the idea that a random variable with small variance is typically close to its mean. In particular, it says that the probability of being a distance a or more from the mean is at most the variance divided by a^2.

Proposition 4.2 (Chebyshev's inequality)**.** *Let τ be a random variable with finite variance. Then, for every $a > 0$,*

$$\mathbb{P}(|\tau - \mathbb{E}[\tau]| \geq a) \leq \frac{\mathrm{Var}(\tau)}{a^2}.$$

Proof. Apply Markov's inequality to the non-negative random variable

$$(\tau - \mathbb{E}[\tau])^2.$$

We obtain

$$\mathbb{P}(|\tau - \mathbb{E}[\tau]| \geq a) = \mathbb{P}\big((\tau - \mathbb{E}[\tau])^2 \geq a^2\big) \leq \frac{\mathbb{E}[(\tau - \mathbb{E}[\tau])^2]}{a^2} = \frac{\mathrm{Var}(\tau)}{a^2}.$$

□

We can now apply Chebyshev's inequality to our random time. We know that

$$\mathbb{E}[\tau] \approx n \log n \qquad \text{and} \qquad \mathrm{Var}(\tau) \leq 2n^2.$$

Therefore, for any $a > 0$,

$$\mathbb{P}(|\tau - \mathbb{E}[\tau]| \geq a) \leq \frac{2n^2}{a^2}.$$

In particular, if we take $a = n\sqrt{2/\varepsilon}$, then

$$\mathbb{P}\left(|\tau - \mathbb{E}[\tau]| \geq n\sqrt{\frac{2}{\varepsilon}}\right) \leq \varepsilon.$$

So, for example, with very high probability τ is within a distance which is only a constant times n from its mean. Since the mean is of order $n \log n$, this means that τ is very strongly concentrated around $n \log n$ when n is large. In other words, take $a = cn$ for a sufficiently large constant c. Then

$$\mathbb{P}(|\tau - \mathbb{E}[\tau]| \geq cn) \leq \frac{2}{c^2}.$$

For instance, taking $c = 10$ gives

$$\mathbb{P}(|\tau - \mathbb{E}[\tau]| \geq 10n) \leq \frac{1}{50}.$$

Thus, with probability at least 49/50, the time at which every card has been touched is at most $\mathbb{E}[\tau] + 10n$.

Recall why this matters for our original shuffling problem. At time τ, every card has been touched at least once, and we argued that the deck is then exactly uniformly random. So, although τ is random, we have now shown that it is very unlikely to be much larger than its typical value $n \log n$. This gives strong evidence that after roughly $n \log n$ random-to-top shuffles, the deck should be well shuffled. In fact, with a little more probability theory, this intuition can be turned into a precise statement about how well shuffled the deck is at a fixed time.

5 Permutations

There is a useful connection between what we have been doing and the mathematics of *permutations*. If we label the cards by $1, \ldots, n$, then a particular ordering of the deck is simply a permutation

$$\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(n)),$$

where $\sigma(i)$ is the card in position i. For example,

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

means that card 3 is on top, followed by cards $1, 5, 2, 4$. The set of all permutations of n objects is denoted by S_n, so the possible states of our deck are precisely the elements of S_n, and, of course, we have intentionally used this notation already.

Permutations can also be written in *cycle notation*. For example,

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

means that 1 is sent to 3, 3 is sent to 5, and 5 is sent back to 1, while 2 and 4 are swapped. The notation $(1\,3\,5)$ is called a 3-cycle. If you have seen groups before, one can actually check that S_n is a group, and it is known as the *symmetric group*: two permutations can be combined ("multiplied") by performing one after the other.

This also gives a neat way to describe our shuffles. A shuffle is simply an operation which takes one permutation (one ordering of the deck) and produces another by just applying an additional permutation, corresponding to the shuffle, to the current permutation of the deck. You do not need this notation for anything in the notes, but it is worth knowing that the random shuffling processes we have studied can also be viewed as random processes on the symmetric group S_n. While we did not define anything here precisely, if you want to know more I encourage you to read about the symmetric group (online or from any text book on group theory).