Lecture 31: Shuffling cards

[Aigner & Ziegler]

Many similarities with mixing. But exactly what is the connection?

Goal: define shuffling method. How long before deck is shuffled "enough"?

Two warmup problems, which at first appear unrelated.

Birthday paradox: n people

prob. that they all have different birthdays?

(365 days a year, no seasonal effects)

two people: $p(2) = 1 - \frac{1}{365}$

three people: $p(3) = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right)$

$\vdots$

$$p(n) = \prod_{i=1}^{n-1} \left(1 - \frac{i}{365}\right)$$

We have $p(n) < \frac{1}{2}$ for $n = 23$.

n balls placed independently in K boxes.
prob that no box has > 1 ball is

$$p(n, K) = \prod_{i=1}^{n-1} \left(1 - \frac{i}{K}\right)$$

Coupon collector:

Take balls from a bowl with n distinguishable balls, put back each time

How many draws expected until you've drawn each ball at least once?

If you've drawn k distinct balls, then prob. of not getting a new one in next draw is k/n.

So prob. to need exactly s drawings for the next ball is:

$$\left(\frac{k}{n}\right)^{s-1}\left(1 - \frac{k}{n}\right)$$

Expected number of drawings for next ball is

$$\sum_{s \geq 1} \left(\frac{k}{n}\right)^{s-1} \left(1 - \frac{k}{n}\right) s$$

We can evaluate this: $x = k/n$

$$\sum_{s \geq 1} x^{s-1} (1-x) s = \sum_{s \geq 1} x^{s-1} s - \sum_{s \geq 1} x^s s$$

$$= \sum_{s \geq 0} x^s (s+1) - \sum_{s \geq 0} x^s s$$

$$= \sum_{x \geq 0} x^s = \frac{1}{1-x}$$

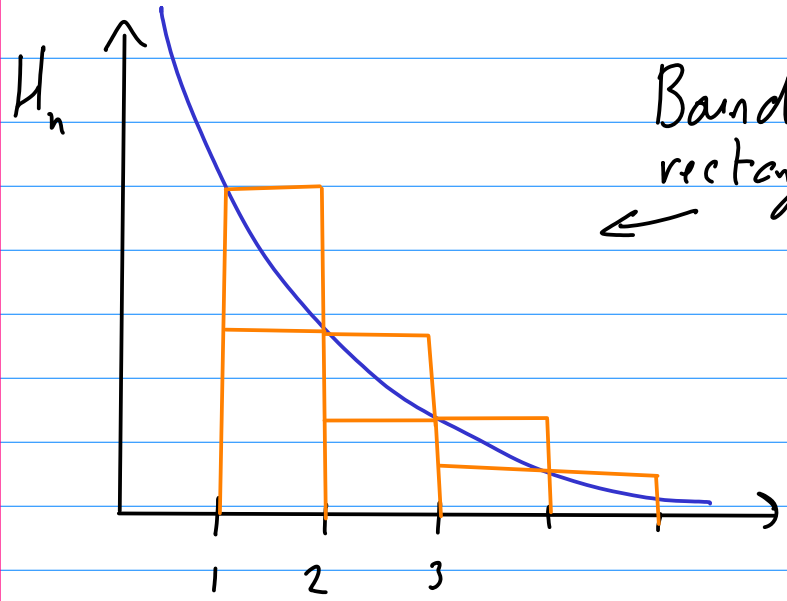
Hence, $$\sum_{s \geq 1} \left(\frac{k}{n}\right)^{s-1} \left(1 - \frac{k}{n}\right) s = \frac{1}{1 - k/n}$$

The expected number of draws until we have drawn each of the n different balls is then

$$\sum_{k=0}^{n-1} \frac{1}{1 - \frac{k}{n}} = \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{2} + \frac{n}{1}$$

$$= n H_n$$

$$H_n = \sum_{k=1}^{n} \frac{1}{k}$$ Harmonic series



Bound this using rectangles

$$H_n - 1 = \sum_{k=2}^{n} \frac{1}{k} < \int_1^n \frac{dt}{t} = \log n$$

$$H_n - \frac{1}{n} = \sum_{k=1}^{n-1} \frac{1}{k} > \int_1^n \frac{dt}{t} = \log n$$

$$\log n + \frac{1}{n} < H_n < \log n + 1$$

So $H_n \approx \log n$, $n \gg 1$.

So the number of draws to get all the balls is

$$\approx n \log n$$

V_n = number of drawings needed to get all n balls

$$E[V_n] \approx n \log n$$

$A_{i,m}$ = {ball i not drawn in first m drawings}

$$\text{Prob}[V_n > m] = \text{Prob}\left[\bigcup_{i=1}^{n} A_{i,m}\right]$$

$$\leq \sum_{i=1}^{n} \text{Prob}\, A_{i,m}$$ ← these are all equal

$$= n\left(1 - \frac{1}{n}\right)^m$$ ← we replace, so always n balls

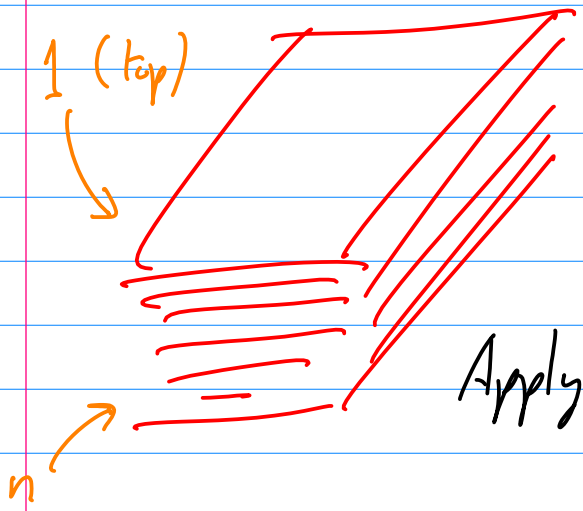
$$< n e^{-m/n}$$

[since $\left(1 - \frac{1}{n}\right)^n < \frac{1}{e}$ for all $n \geq 1$]

Let $m = \lceil n \log n + cn \rceil$. Then:

$$\text{Prob}[V_n > m] \leq e^{-c}$$ probability that it takes "long"

Deck of cards: number 1 to n in the order they first appear in the deck. (all different)



S_n = symmetric (permutation) group.

What does it mean to shuffle?

Apply random permutations to deck.

Say we pick $\pi \in S_n$ with equal prob. $\frac{1}{n!}$.

Order after one shuffle is $(\pi(1), \pi(2), \ldots, \pi(n))$.

Perfectly random!

But that's not how it's really done.

Only "certain" permutations occur.

example: top-in-at-random shuffle
(not very effective, but easy to analyze)

Take the top card, insert at random in one of n places, with prob $1/n$.

So one of

$$\tau_i = (2, 3, \cdots, i, 1, i+1, \cdots, n) \quad 1 \le i \le n$$

is applied, each with prob $1/n$.

Expect this to take a long time, but how do we measure this?

"variation distance":

We look at the probability distribution of the $n!$ different orderings of our deck.

Starting distribution I:

$$I(id) = 1$$

$$I(\pi) = 0, \text{ otherwise}$$

$$(\pi \in S_n)$$

Since the cards are unmixed

At the other extreme, we have the Uniform distribution:

$$U(\pi) = \frac{1}{n!}, \quad \pi \in S_n$$

Our goal is to see, after a certain number of shuffles, how "close" we are to the uniform distribution.

The variation distance between two probability distributions Q_1 and Q_2 is

$$\|Q_1 - Q_2\| = \frac{1}{2} \sum_{\pi \in S_n} |Q_1(\pi) - Q_2(\pi)|$$