# Lecture 33: Riffle shuffling

Last time: for top-in-at-random shuffling, we showed

$$\text{Prob}[T > k] \leq e^{-c}$$

for $k = \lceil n \log n + cn \rceil$

$n$ = number of cards

$T$ = stopping time for strong uniform stopping rule.

**Lemma:** Let $Q : S_n \to \mathbb{R}$ be a prob. distribution that defines a shuffling process $Q^{*k}$ with a strong uniform stopping rule whose stopping time is $T$. Then

$$\| Q^{*k} - U \| \leq \text{Prob}[T > k] \quad \forall k \geq 0$$

**proof:** $X$ random var. with values in $S_n$, prob dist. $Q$. For $S \subseteq S_n$,

$$U(S) = \text{Prob}[X \in S] = \frac{|S|}{n!}$$

← cardinality

$$Q^{*h}(S) = \text{Prob}[X_h \in S]$$

$$= \sum_{j \le h} \text{Prob}[X_h \in S \wedge T = j]$$

$$+ \text{Prob}[X_h \in S \wedge T > h]$$

Recall: $\text{Prob}[X_j \in S \mid T = j] = U(S)$

$$\text{Prob}[X_j \in S \wedge T = j] = U(S)\,\text{Prob}[T = j]$$

↘ can replace by $h$, since $h \ge j$

$$Q^{*h}(S) = \sum_{j \le h} U(S)\,\text{Prob}[T = j]$$

$$+ \text{Prob}[X_h \in S \mid T > h]\,\text{Prob}[T > h]$$

$$= U(S)\,(1 - \text{Prob}[T > h]) + \cdots$$

$$= U(S) + (\text{Prob}[X_h \in S \mid T > h] - U(S))\,\text{Prob}[T > h]$$

$|\cdot| \le 1$ (difference of probs.)

Hence,

$$|Q^{*h}(S) - U(S)| \le \text{Prob}[T > h].$$

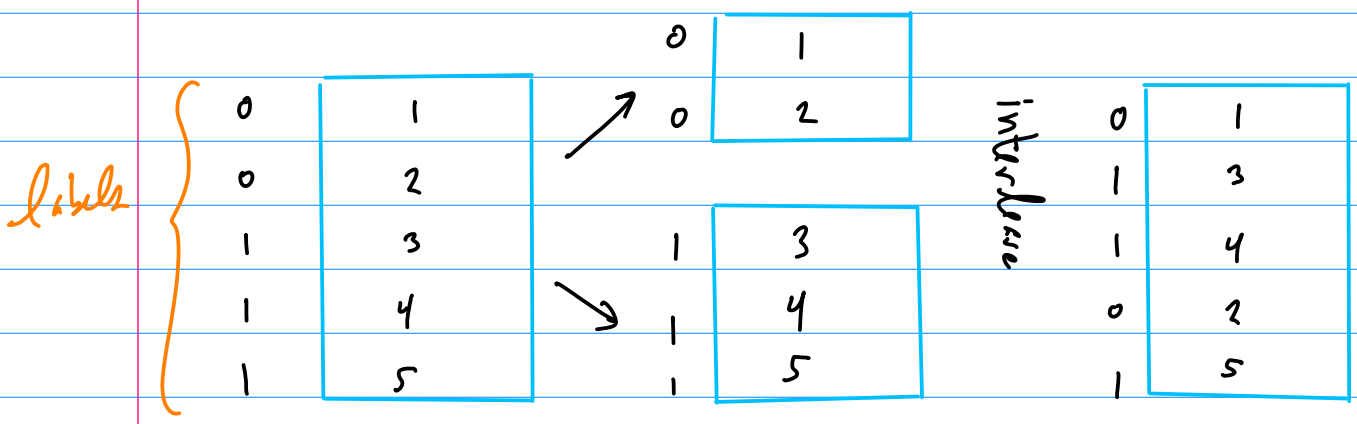
Conclude: $d(k) = \| Top^{*k} - U \| \leq e^{-c}$.

Riffle shuffle: • split deck into two parts
• interleave

Model: riffle shuffle consists of permutations $\pi \in S_n$ such that

$$(\pi(1), \pi(2), \ldots, \pi(n))$$

consists of two interlaced increasing sequences.
(except if it is the identity)




$t$ cards in right hand $(0 \leq t \leq n)$
$n-t$ " " left hand

$\binom{n}{t}$ ways to interleave

All distinct, except for identity.

total # of ways. $\sum_{t=0}^{n} \binom{n}{t} = 2^n$

But $n+1$ of these are the identity $\rightarrow$ keep only one

$\Rightarrow 2^n - n$ distinct ways. (one is the identity)

What probability distribution to assign?

Gilbert & Shannon (1955) [simple]

$Rif : S_n \rightarrow \mathbb{R}$;

$$Rif(\pi) = \begin{cases} \frac{1}{2^n}(n+1), & \pi = id \\ \frac{1}{2^n}, & \pi \text{ consists of two increasing sequences} \\ 0, & \text{otherwise} \end{cases}$$

Check: $\sum_{\pi} Rif(\pi) = \frac{1}{2^n}(n+1) + \frac{1}{2^n}(2^n - n - 1)$

$= 1$

Inverse shuffle: assign label 0 or 1 to each card randomly with prob. 1/2.
Move 0 cards to the top.

Inverse shuffling yields the same probability distribution as Rif.

For instance, inverse shuffling gives identity whenever all the 0 cards are already on top:

$\frac{n+1}{2^n}$ ← ways to have 0's on top (including none)
← total # of configs

which is the same as $Rif(id)$.

Show: $\|Rif^{*k} - U\| \leq 1 - \prod_{i=1}^{n-1} \left(1 - \frac{i}{2^k}\right)$

proof: Analyze inverse shuffles instead:

$\overline{Rif}(\pi) := Rif(\pi^{-1})$

Since every perm has unique inverse, and $U(\pi) = U(\pi^{-1})$:

$\|Rif^{*k} - U\| = \|\overline{Rif}^{*k} - U\|$

In every inverse riffle shuffle, each card gets a bit: remember these digits.

Stopping rule:

STOP as soon as all cards have distinct strings.

The cards are sorted according to the binary numbers

$$b_k b_{k-1} \cdots b_2 b_1 \qquad b_i = i^{\text{th}} \text{ shuffle}$$

But since these bits are independent and random, the order of the deck must then be random!
$\rightarrow$ stopping rule is strong uniform.

But how long does this take?

$K = 2^k$ configurations (boxes)

Put 2 cards in the same box if they have the same label $b_k b_{k-1} \cdots b_2 b_1$

What is the probability that some box gets more than one card?

Birthday paradox!

$$\text{Prob}[T > k] = 1 - \prod_{i=1}^{n-1} \left(1 - \frac{i}{2^k}\right)$$

and this bounds the variation distance. ///

How many shuffles do we need? ($n$ large)

Let $k = 2\log_2(cn)$, $c \geq 1$

$$1 - \frac{i}{2^k} = 1 - \frac{i}{(cn)^2}$$

$$\sum_{i=1}^{n-1} \log\left(1 - \frac{i}{(cn)^2}\right) \approx \sum_{i=1}^{n-1} \frac{-i}{(cn)^2}$$

$$= \frac{-1}{(cn)^2} \frac{1}{2} n(n-1) \approx -\frac{1}{2c^2}$$

Hence,

$$\text{Prob}[T > k] \approx 1 - e^{-1/2c^2} \approx \frac{1}{2c^2}$$

We get from this

$$d(12) \leq 0.28$$

This is actually a pretty bad bound.
In practice,

$$d(6) \leq 0.614$$

$$d(7) \leq 0.334 \quad \leftarrow \text{magic \# of shuffles}$$

$$d(8) \leq 0.167$$