MCAA lecture 8

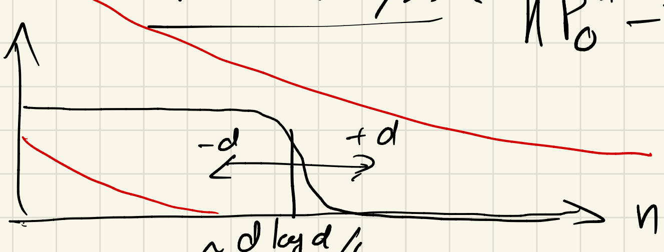
Recap: cut off phenomenon

1. RW on $S = \{0,1\}^d$ $\quad$ S very large
(prob $\frac{1}{d+1}$ to stay or to flip one bit)

Bounds: $\frac{1}{2}\exp\left(-\frac{2n}{d+1}\right) \le \| P_0^n - \pi \|_{TV} \le \frac{1}{2\sqrt{\pi_0}} \cdot \exp\left(-\frac{2n}{d+1}\right)$

pb

Finer analysis: $\| P_0^n - \pi \|_{TV} \sim \begin{cases} 0 & \text{if } n \gg \frac{d+1}{4}\log d \quad \text{(UB)} \\ 1 & \text{if } n \ll \frac{d+1}{4}\log d \quad \text{(LB)} \end{cases}$



Pf idea:

UB: $\| P_0^n - \pi \|_{TV} \leq \frac{1}{2} \Big(\sum_{z=0} \lambda_z^{2n} \Big)^{1/2} \cdots$ all the ev.

~ 0 si $n \gg \frac{d \log d}{4}$

LB: $\| P_0^n - \pi \|_{TV} = \max_{A \subset S} | P_0^n(A) - \pi(A) |$

$\geq | P_0^n(A) - \pi(A) | \quad \forall A \subset S$

Pick $A_\beta = \{ x \in S : \; \big| |x| - \frac{d}{2} \big| \leq \frac{\beta}{2} \sqrt{d} \}$

$\Longrightarrow \pi(A_\beta) \simeq 1$, $\quad P_0^n(A_\beta) \simeq 0$ if $n \ll \frac{d \log d}{4}$

Card shuffling

$S = \{ \text{permutations of } \{1..N\} \}$ $\quad |S| = N!$
large!

Shuffling method = Markov chain on S

Question: For a given method, how long does it take to decently shuffle the deck?

(i.e. to have a distribution ε-close in TV-dist to the uniform distribution on S)

Method 0 : cut repeatedly the deck

not an ergodic chain (51! equivalence classes)

Method 1 : "random to top"

- choose a number unif at random in $\{1 .. N\}$
- look for the card with this number & put it on top

This is an ergodic chain !

Claim : $\Theta(N \log N)$ shuffles are needed with this method.

"Pf": UB

Coupling

X_n = chain starting from the identity state Id

Y_n = chain starting from $\pi \sim$ uniform

choose a number $\in \{1..N\}$ unif. at random & look after the card with this number in each deck, and put it on top in each deck.

Former lemma:

$$\| \underbrace{P^n_{Id}}_{\text{dist of } X_n} - \underbrace{\pi}_{\text{dist of } Y_n} \|_{TV} \leq \mathbb{P}(X_n \neq Y_n)$$

Observation:

After each card number has been picked at least once, the two decks are the same

So if $\tau = \inf\{n \geq 1 : \text{each number has been picked once}\}$

Then $\mathbb{P}(X_n \neq Y_n) \leq \mathbb{P}(\tau > n)$

<u>Coupon collector pb:</u>

$$\mathbb{E}(\tau) = \frac{N}{N} + \frac{N}{N-1} + \frac{N}{N-2} + \dots + \frac{N}{3} + \frac{N}{2} + \frac{N}{1}$$

$$= N \cdot \sum_{k=1}^{N} \frac{1}{k} \sim N \cdot \log N$$

$$\mathrm{Var}(\tau) = \Theta(N^2) \qquad \tau \simeq N \log N \pm N$$

So $\mathbb{P}(\tau > n) \sim 0$ for $n \gg N \log N$

"Pf" (LB):

$$\| P_{Id}^{n} - \pi \|_{TV} = \max_{A \subset S} | P_{Id}^{n}(A) - \pi(A) |$$

$$\geq | P_{Id}^{n}(A) - \pi(A) | \quad \forall A \subset S$$

Choose $A_k = \{$ k bottom cards of the deck are ordered $\}$ k fixed (k=10)

$\pi(A_k) = \frac{1}{k!} \sim$ small

To check: $P_{Id}^{n}(A) \sim 1$ if $n \ll N \log N$

Observation: while k different cards have never been picked, at least k bottom cards of the deck will be ordered (because we started from the Id permutation)

The average to pick $N-k$ different cards is

$$\mathbb{E}(T_k) = \frac{N}{N} + \frac{N}{N-1} + \frac{N}{N-2} + \dots + \frac{N}{k} = N \sum_{\ell=k}^{N} \frac{1}{\ell} \simeq N(\log N - \log k)$$

$$= N \log\left(\frac{N}{k}\right)$$

$$\mathrm{Var}(T_k) = \Theta(N^2)$$

Conclusion:

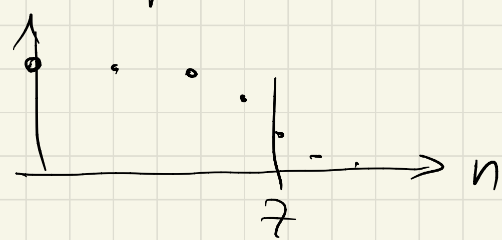
If $n \ll N \log N$, $P^n_{Id}(A_k) \sim 1$

So $\| P^n_{Id} - \pi \|_{TV} \sim 1$

<u>Method 2</u>: riffle shuffle

$\Rightarrow \Theta(\log N)$ suffice!

réf: P. Diaconis:



$(N=52)$

Sampling

$\pi_i \geq 0, \sum_{i=1} \pi_i = 1$

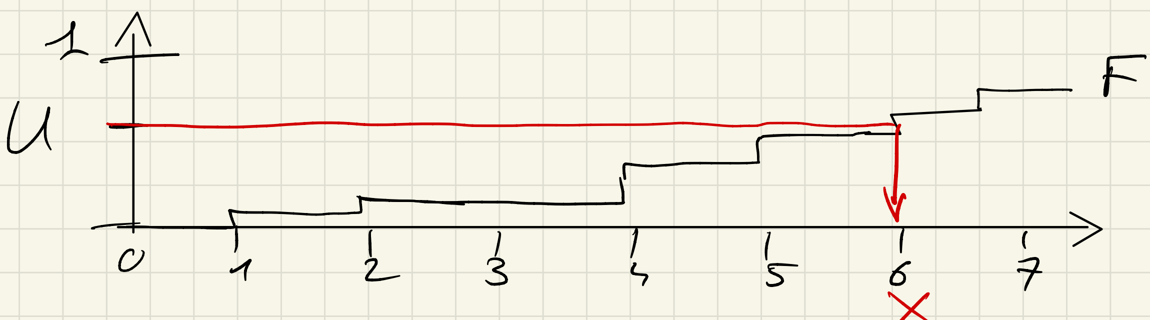
Given a distribution $(\pi_i, i \in S)$ on a state space S, how can we sample from it?

"Easy solution": let X be a r.v. with values in S such that $\mathbb{P}(X=i) = \pi_i$ $i \in S$ and assume $S = \mathbb{N}$

Generate $U \sim \mathcal{U}[0,1]$ and declare

$$X = \begin{cases} 0 & \text{if } \ 0 \le U \le \pi_0 \\ 1 & \text{if } \ \pi_0 < U \le \pi_0 + \pi_1 \\ 2 & \text{if } \ \pi_0 + \pi_1 < U \le \pi_0 + \pi_1 + \pi_2 \\ \vdots & \end{cases}$$

$$= F^{-1}(U) \qquad \text{where } F = \text{cdf of } X$$



Why to sample? 1. **Optimization of a complex fn**

$f : \{0,1\}^d \longrightarrow \mathbb{R}$

Aim: to maximize f, ie

to find $x_0 \in \{0,1\}^d$ st $f(x_0) = \max$.

Define $\pi(x) = \begin{cases} 0 & \text{if } f(x) \neq \max \\ c = \frac{1}{Z} & \text{if } f(x) = \max \end{cases}$

where Z = number of maxima of f

First idea: sample from π to get a maximum x

Second idea: instead of sampling from π, sample from π_β defined as follow:

$$\pi_\beta(x) = \frac{e^{\beta f(x)}}{Z_\beta} \qquad x \in \{0,1\}^d$$

where $\underbrace{Z_\beta = \sum_{x\in\{0,1\}^d} e^{\beta f(x)}}_{}$ normalization cst

"Partition function"

2. Compute averages (Monte Carlo method)

- X = r.v. with values in S and distribution π

  ie $\mathbb{P}(X=i) = \pi_i \quad i \in S$
- $f : S \to \mathbb{R}$
- Aim: Compute $\mathbb{E}(f(X)) = \sum_{i \in S} f(i)\pi_i$

MC method: draw independent samples $X_1 \dots X_M$

& compute $\frac{1}{M}\sum_{j=1}^{M} f(X_j) \sim \mathbb{E}(f(X))$

$$\mathrm{Var}\left(\frac{1}{M}\sum_{j=1}^{M} f(x_j)\right) = \frac{1}{M^2}\sum_{j=1}^{M} \underbrace{\mathrm{Var}(f(x_j))}_{= \mathrm{Var}(f(x)}$$

$$= \frac{\mathrm{Var}(f(x))}{M}$$

$$\text{std dev} \sim \Theta\left(\frac{1}{\sqrt{M}}\right)$$