Evita Nestoridi - Coupon collecting and strong stationary times

Lecture 1 -

n coupons

Every day a collector buys a coupon.

All coupons appear with same probability. Let T be the first time we have collected them all. Want to know how big T must be.

Lemma: a) $P(T>t) \le n\left(1-\frac{1}{n}\right)^t$

b) $P(T > n\log n + cn) \le e^{-c}$ (doesn't take longer than $n\log n$ and a little bit)

c) $P(T < n\log n - cn) \ge e^{-c}$ [we are concentrated along $n\log n$]

Proof: a) $P\left(\text{coupon } i \text{ was purchased at time } t\right) = \frac{1}{n}$ All the purchases are independent

$\Rightarrow P(\text{not getting it at time } t) = 1 - \frac{1}{n}$

$$P(T>t) = \mathbb{P}\left(\bigcup_{i=1}^{n} \left\{\begin{array}{l}\text{coupon } i \text{ has}\\ \text{not been collected}\\ \text{at time } t\end{array}\right\}\right) \le \sum P\left(\text{coupon } i \text{ not collected by time } t\right) = \sum_{i=1}^{n}\left(1-\frac{1}{n}\right)^t = n\left(1-\frac{1}{n}\right)^t$$

b) $\left(1-\frac{1}{n}\right)^n \le \frac{1}{e}$

$$P(T > n\log n + cn) \le n\left(1-\frac{1}{n}\right)^{n[\log n + c]} \le n\left(\frac{1}{e}\right)^{\log n + c} \le e^{-c}$$

c) Let T_i = ~~first time that~~ i-th time we got a new coupon

$$T = T_n = \sum_{i=1}^{n}(T_i - T_{i-1}) \quad \text{as } T_0 = 0$$

Chernov bound

Markov's inequality: X random variable $P(X > a\overbrace{E(X)}^{\text{expected value}}) \le \frac{1}{a}$

famous trick - s to be chosen later

$$\mathbb{P}(T<t) \underset{s>0}{=} \mathbb{P}(-sT > -st) = \mathbb{P}(e^{-sT} > e^{-st}) = \mathbb{P}\left(e^{-sT} > \frac{e^{-st}}{E(e^{-sT})}E(e^{-sT})\right) \overset{\text{Markov}}{\le} e^{st}E(e^{-sT})$$

$$= e^{st}E\left(e^{-s\sum_{i=1}^{n}\underbrace{T_i - T_{i-1}}_{X_i}}\right) = e^{st}E\left(e^{-sX_1}e^{-sX_2}\cdots e^{-sX_n}\right) = e^{st}\prod_{i=1}^{n}E\left(e^{-sX_i}\right)$$

$X_i = T_i - T_{i-1}$ (how long it took for the i-th coupon to appear after the (i-1)st)

Geometric distribution $\sim$ Geometric$\left(\frac{n-i-1}{n}\right)$

$$= e^{st}\prod_{i=1}^{n}\frac{\frac{i-1}{n}}{e^s - 1 + \frac{i-1}{n}}$$ ← check in exercise session

If $t = n\log n - cn$, $s = \frac{1}{n}$ $\qquad e^s = e^{\frac{1}{n}} = 1 + \frac{1}{n} + \dots \ge 1 + \frac{1}{n}$

$$\Rightarrow \mathbb{P}(T<t) \le e^{\log n - c}\prod_{i=2}^{n+1}\frac{(i-1)/n}{i/n} = n e^{-c}\prod\frac{i-1}{i} = \frac{n}{n+1}e^{-c} < e^{-c}$$ $\square$

Will be using this result about coupon collecting over and over again.

This type of argument can be found on Feller - introduction to probability (also stack exchange is a good resource)

Card shuffling: Random-to-top: deck of n cards. Pick a card uniformly at random and move it to the top. How many times do we need to do this before the cards have been well shuffled.

People don't shuffle like this but computers often perform tasks in this way.

$$\begin{array}{ccc} 1 & 2 & 3 \\ 2 & 1 & 1 \\ 3 & 3 & 2 \\ \vdots & \vdots & 4 \\ & & \vdots \\ n & n & n \\ \rightsquigarrow \mathrm{id} & \rightsquigarrow (12) & \rightsquigarrow (123) \end{array} \qquad \in S_n \qquad |S_n| = n!$$

Transition matrix : $x, y \in S_n$ $P(x,y)$ = probability of $x \to y$ after one step

$$= \begin{cases} \frac{1}{n} & \text{if } y = x(1\,2 \cdots i) \quad \text{for } i = 1, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

We are interested in $P^t(x,y)$ = probability $x \to y$ after t steps. Now we can't write down a formula. It's useful to compute eigenvectors and eigenvalues

(P^t: t-th power)

$P^t_x(y)$ is a probability measure on S_n ($\sum P^t_x(y) = 1$).

Tomorrow we'll see this converges to the uniform measure as $t \to \infty$. We'll study the convergence

Used metric is the total variation distance Persi introduced

$$\| P^t_x - U \|_{TV} = \frac{1}{2} \sum_{y \in S_n} \left| P^t_x(y) - \frac{1}{n!} \right|$$

(U: uniform measure on S_n)

Separation distance : $$s_x(t) = \max_{y \in S_n} \left\{ 1 - \frac{P^t_x(y)}{U(y)} \right\}$$ exercise : this is positive (but not a measure)

$$s(t) = \max_x \{ s_x(t) \}$$

Lemma (Aldous – Diaconis) $\boxed{\| P^t_x - U \|_{TV} \leq s(t)}$ for every $x \in S_n$

~~Mixing~~ The mixing time is defined as

$$t_{mix}(\varepsilon) = \min \{ t > 0 : \| P^t_x - U \|_{TV} < \varepsilon \}$$

Each time we bring k distinct cards to the top, ~~then~~ all $k!$ orderings of the first k cards should appear with same probability. How long should we wait until we move all n cards? $n \log n$! (coupon collecting).

T = first time that all cards have been selected

T_i = first time that the i-th new card appeared

What we have said can be expressed as

$$\mathbb{P}^t_x(y \mid T \leq t) = \frac{1}{n!}$$

Want to link this to $P^t_x(y) = \mathbb{P}^t_x(y \mid T \leq t)\, P(T \leq t) + P^t_x(y \mid T > t)\, P(T > t)$ (Bayes formula)

$$\geq \mathbb{P}^t_x(y \mid T \leq t)\, P(T \leq t) = \frac{1}{n!} \mathbb{P}(T \leq t)$$

Plugging this into the separation distance we have $$s_x(t) \leq \max_{y \in S_n} \left\{ 1 - \frac{\frac{1}{n!} P(T \leq t)}{\frac{1}{n!}} \right\} = \mathbb{P}(T > t)$$

So we have proved

Theorem (Aldous - Diaconis) If $t_{n,c} = n \log n + cn$ then $s(t_{n,c}) \leq e^{-c}$

Summing up and generalizing :

Suppose we have a process $X_1, X_2, \dots, X_t, \dots$ (sequence of random variables)

If $P(X_{t+1} = x \mid X_1 = x_1, \dots, X_t = x_t) = P(X_{t+1} = x \mid X_t = x_t)$

we say this is a <u>Markov process</u> on Ω (before $\Omega = S_n$)

$P = (P(x,y))_{x,y \in \Omega}$ if $\exists$ measure $\pi : \Omega \to [0,1]$ such that $P^t_x \to \pi$ as $t \to \infty$.

$t_{mix}(\varepsilon) \overset{def}{=} \min\{t \geq 0 : \max_x d_x(t) < \varepsilon\}$

some distance

$T : \Omega \times \dots \times \Omega \times \dots \to \mathbb{N}$ random variable, is a <u>stopping time</u> for (X_t) if $\{T \leq t\}$ depends only on $X_1, X_2, \dots, X_t$.

<u>Example</u> : T = first time that card 1 is on top.

<u>Definition</u> : T is a <u>strong stationary time</u> for (X_t) if $\mathbb{P}_x(X_t = y \mid T \leq t) = \pi(y)$

<u>Lemma</u> (A-D) $s(t) \leq P(T > t)$.

The proof is exactly the same as in the special case discussed above.

Tomorrow : 1) Eigenvalues and eigenfunctions of transition matrix in mixing
2) Comparison