EE675A - IITK, 2022-23-II

Lecture : Sample Quiz

Lecturer: Subrahmanya Swamy Peruru *Scribe: Harsha Kurma, Aravindasai Bura*

1 Exercise

Let X be a random variable which takes values in [0, 1]. Let $\mathbb{E}[X] = \mu$, and $\hat{\mu}_N$ denotes the sample average obtained by observing N i.i.d samples of X. Suggest a number N such that $\bar{\hat{\mu}}_N$ does not deviate from μ by more than 0.01 with a very high probability (let us say with 0.99 probability).

Hoeffding's Inequality: It states that the probability of deviation of estimated mean $\hat{\mu}(a)$ from true mean greater than ϵ is upper bound by $2e^{-2\epsilon^2 N}$ where N is number of samples

$$P[|\bar{\mu}(a) - \mu(a)| \geq \epsilon] <= 2e^{-2\epsilon^2 N}$$

Given $\hat{\mu}_N$ does not deviate from μ by more than 0.01 with a very high probability (let us say with 0.99 probability)

$\epsilon = 0.01$ does not deviate with probability 0.99 $\implies$ deviates with probability 0.01

$$P[|\bar{\mu}(a) - \mu(a)| \geq 0.01] \leq 0.01$$
$$0.01 \leq 2e^{-2\epsilon^2 N}$$
$$0.01 \leq 2e^{-2(0.01)^2 N}$$

taking log on both sides ans solving further

$$N \geq 24692$$

2 Exercise

Prove that ϵ-greedy algorithm, where ϵ is some fixed constant, incurs a regret that grows linearly in T .

In ϵ - Greedy Algorithm, in every round with probability ϵ an arm is played randomly,and with probability $1 - \epsilon$ the optimal arm is played.

Therefore, $\Delta(a)$ is bounded by,

$$\begin{aligned}
\Delta(a(t)) &= \mu(a^*(t)) - \mu(a(t)) \\
R(t) &= \sum_{t \in T} \Delta(a(t)) \\
E[R(t)] &= E[\sum_{t \in T} \Delta(a(t))] \\
E[R(t)] &= \sum_{t \in T} E[\Delta(a(t))] \\
&= \sum_{t \in T} (\sum_{a \in A} \Delta(a(t)) P(a(t) = a))
\end{aligned}$$

$\hat{\boldsymbol{a}}(t) = \arg\max_{a \in A} \overline{\mu_t}(a)$
(Note that $\hat{\boldsymbol{a}}$ is the best arm based on sample estimate, whereas a^* is the actual best arm that gives us the maximum reward.)

consider there are **k** arms

The probability to choose an arm at random is 1/k with ϵ, the probability a random arm is selected which includes the best arm, so the probability of choosing a random arm is ϵ/k

In Greedy case there is only one best arm and it is played with probability 1-ϵ

$$\begin{aligned}
&P(a(t) = \hat{\boldsymbol{a}}) = (1 - \epsilon) + \epsilon/k \\
&P(a(t) = a) = \epsilon/k \\
&P(a(t) = a) \geq \epsilon/k \\
&\quad E[R(t)] = \sum_{t \in T} (\sum_{a \in A} \Delta(a(t)) P(a(t) = a)) \\
&\quad E[R(t)] \geq \sum_{t \in T} (\sum_{a \in A} \Delta(a(t)) \epsilon/k) \\
&\quad E[R(t)] \geq (\epsilon T/k) \sum_{a \in A} \Delta(a(t))
\end{aligned}$$

ϵ is constant in this case, the expected regret grows linearly **T**

3 Exercise

In the UCB algorithm discussed in our class, we had

$$UCB_t(a) := \overline{\mu}_t(a) + \sqrt{\frac{2lnT}{n_t(a)}} \tag{2}$$

To execute this algorithm, we need to know the value of T , *i.e.,* we should know the total number of rounds we are going to play. Let us assume that we do not know the value of T in advance and consider the following variant of UCB.

$$UCB_t(a) := \overline{\mu}_t(a) + \sqrt{\frac{lnt}{n_t(a)}} \tag{3}$$

Prove that this UCB variant also has a similar regret bound. **HINT**: The proof proceeds in a similar fashion. Just make necessary changes to reflect this new formulation

step-1 :

We have $UCB_t(a) \geq UCB_t(a^*)$

calculate the probability that the true mean $\mu(a)$ lies outside the confidence interval that we have at time 't' i.e., $\forall$ actions $a \in \boldsymbol{A}$ and $0 < t < \boldsymbol{T}$

$$\mathbb{P}\left(\overline{\mu}_t(a) \notin \left(\overline{\mu}_t(a) - \sqrt{\frac{ln\mathbf{t}}{n_t(a)}}\right), \left(\overline{\mu}_t(a) + \sqrt{\frac{ln\mathbf{t}}{n_t(a)}}\right)\right) \tag{4}$$

Using Hoeffding's Inequality

$$P[|\bar{\mu}(a) - \mu(a)| \geq \epsilon] <= 2e^{-2\epsilon^2 N}$$

considering ϵas$\sqrt{\frac{lnt}{n_t(a)}}$ we get

$$2e^{-2\epsilon^2 N} \ = \frac{2}{t^2}$$

step-2:

In the proof discussed in class for UCB, we assumed that for every time $1 \leq t \leq \mathbf{T}$, and for every arm $a \in \boldsymbol{A}$, our confidence intervals are correct with high probability, specifically if we use union bound

$\mathbb{P}$(our confidence interval is wrong for atleast one arm (or) one time slot) $\leq \dfrac{2}{T^4}Tk$

for a given arm 'a' at a particular round 't' the probability of deviation of true mean out side the bound is $2/\mathbf{T}^4$

Union bound over all time slots 1<t<**T** and all the 'k' arms

Assuming k $\leq$ **T** which is a reasonable assumption since we should have enough rounds to play each arm at least once

P(confidence interval getting voilated atleast in one slot or for one arm)$\leq \mathcal{O}\left(\frac{1}{T^2}\right)$

For UCB= $\sqrt{\frac{logt}{n_t(a)}}$:

If we do a similar analysis for $\sqrt{\frac{logt}{n_t(a)}}$ version of UCB
Using Hoeffding's inequality

$$\mathbf{P}(\text{voilation for arm 'a' at round 't'}) \leq \frac{2}{T^2} \forall a \in \mathcal{A} \tag{6}$$

$$\begin{aligned} \mathbf{P}(\text{ voilation for atleast one arm at least on o=round}) \leq \sum_{t=1}^{T} \sum_{a \in A} \frac{2}{t^2} \\ \mathbf{P} = K \sum_{t=1}^{T} \frac{2}{t^2} > K \\ \text{as} T \rightarrow \infty \end{aligned} \tag{7}$$

Conclusion: This bound is greater than 1,which is useless,we cant hopw our confidence intervals are correct all the times

Step-3: For the UCB version done in class we have shown that

i)A bad arm cannot be played too many times.
ii)If any arm is played a lot of times, its not a very bad arm.

similarly , " If a suboptimal arm is played sufficient number of times then the probability of playing it further is very less

$$\mathbb{P}\left(\frac{a(t+1) = a}{n_t(a) \geq \frac{4lnt}{(\triangle(a))^2}})\right) \leq \frac{4}{t^2} \tag{8}$$

Note: using the following properties 1. If an arm has to be played at time 't',

$$UCB_t(a) \geq UCB_t(a^*) \tag{9}$$

$$\bar{\mu}_t(a) + \epsilon_t(a) \geq \bar{\mu}_t(a^*) + \epsilon_t(a^*) \tag{10}$$

$$\bar{\mu}_t(a^*) - \bar{\mu}_t(a) \leq \epsilon_t(a) - \epsilon_t(a^*) \tag{11}$$

Also we have ,

$$\Delta(a) = \mu(a^*) - \mu(a) \tag{12}$$
$$\Delta(a) \leq \bar{\mu}_t(a^*) + \epsilon_t(a^*) - (\bar{\mu}_t(a) - \epsilon_t(a)) \tag{13}$$
$$\Delta(a) \leq 2\epsilon_t(a)) \tag{14}$$
$$\Delta(a) \leq 2\sqrt{\frac{\log t}{n_t(a)}} \tag{15}$$
$$\boldsymbol{n}_t(a) \geq \frac{4\log t}{(\Delta(a))^2} \tag{16}$$

$$\mathbb{P}(\text{confidence interval voilation for arm a, at t}) \leq \frac{2}{t^2} \text{using this property for both a, a*}$$

P(confidence interval going wrong either for a or a*) $\leq \frac{4}{t^2}$
i.e.,

$$\left(|\mu(a) - \overline{\mu}_t(a))| \sqrt{\frac{lnt}{n_t(a)}} \right) \leq \frac{2}{t^2} \tag{17}$$
$$\tag{18}$$

using for both a and a^* we get,

$$\mathbf{P}(\text{confidence interval going wrong for either a or a*}) \leq \frac{4}{t^2} \tag{19}$$
$$\mathbf{P}(\text{confidence interval correct in both cases}) \leq 1 - \frac{4}{t^2} \tag{20}$$
$$\tag{21}$$

we can say that

$$\mathbb{P}\left(\frac{a(t+1) = a}{n_t(a) \geq \frac{4lnt}{(\triangle(a))^2}}) \right) \leq \frac{4}{t^2} \tag{22}$$

Step 4:
$\mathrm{E}[n_t(a)] \leq \frac{4lnT}{(\triangle(a))^2} + 8$

This is nothing but "A bad arm in not played many times".
Express
$\mathrm{E}[n_t(a)] = 1 + \mathbf{E}\left[\sum_{t=K+1}^{T} 1_{(a(t)=a)}\right]$

every arm is played once in first K rounds so 1
Divide the second term into two parts
$\mathrm{E}\left[\sum_{t=K+1}^{T} 1_{a(t)=a, n_t(a) \leq \frac{4lnt}{(\triangle(a))^2}}\right] + \mathbf{E}\left[\sum_{t=K+1}^{T} 1_{a(t)=a, n_t(a) \geq \frac{4lnt}{(\triangle(a))^2}}\right]$

$$\mathbf{E}\left[\sum_{t=K+1}^{T} 1_{a(t)=a, n_t(a) \leq \frac{4lnt}{(\triangle(a))^2}}\right] \text{the total contribution of this term is} \frac{4lnT}{((a))^2} \tag{23}$$

$\mathrm{E}\left[\sum_{t=K+1}^{T} 1_{a(t)=a, n_t(a) \geq \frac{4lnt}{(\triangle(a))^2}}\right]$
$\sum_{t=k+1}^{T} \mathbb{P}\left(\frac{a(t)}{n_t(a) \geq \frac{4lnt}{(\triangle(a))^2}}\right) \mathbb{P}\left(n_t(a) \geq \frac{4lnt}{(\triangle(a))^2}\right)$
$\sum_{t=k+1}^{T} \frac{4}{t^2} = 8$

step 5: calculate $\mathbf{E}\left[R(T;a)\right]$ using $\mathbf{E}[n_t(a)]$ from step 4 use $\mathbf{E}[R(T)] = \sum_a \mathbf{E}[R(T;a)]$

$$\sum_a \mathbf{E}[R(T;a)] = \sum_a \mathbf{E}[n_t(a)]\triangle(a) \tag{24}$$

$$\sum_a \mathbf{E}[R(T;a)] \leq \sum_a \frac{4\log t}{(\Delta(a))} + 8(\Delta(a)) \tag{25}$$