Supervised Research Exposition

Puranjay Datta ,19D070048

14-03-2022

Problem Statement

Multiarm Bandits

We have K groups of arms where each group has N arms. In each round, we select a group and are randomly assigned one of the N arms in that group. The identity and the reward corresponding to this arm are revealed to us by the end of the round. The goal is to identify the group with the highest mean reward (average across the N arms) subject to the constraint that the minimum mean reward in the group exceeds a given threshold.

General Successive Elimination

- Player maintains an active set of arms $'S'$.
- At every round player first samples from the reward distribution of every arm in the active set.
- Player then removes all arms in the active set with estimated rewards that are outside the anytime confidence interval around the highest estimated reward in active set.
- When active set has 1 arm,the player identifies this arm with high probability as the best arm.

Algorithm

Successive Elimination($\{1, 2, 3..., n\}, \delta$)
$S \leftarrow \{1, 2, 3..., n\}$
while $1 \leq t \leq \infty$ **do**
    Pull arms in S
    $S \leftarrow S - \{i \in S; \exists j \in S : \hat{\mu}_{j,t} - U(t, \frac{\delta}{n}) \geq \hat{\mu}_{i,t} + U(t, \frac{\delta}{n})\}$
    Stop when $|S| = 1$
end while
return S
end procedure

Notation

- S:Active set of arms
- Estimated mean reward for arm i after t pulls: $\hat{\mu}_{i,t} = \frac{1}{t}\sum_{j=1}^{t} X_{i,j}$
- $U(t,\delta)$ =Confidence bound $\mathbb{P}(\{\cup_{t=1}^{\infty}|\hat{\mu}_{i,t} - \mu_i| > U(t,\delta)\}) \leq \delta$
  With high probability these bounds hold for all time rather than independently holding eith high probability at each time step individually.

Theorem 1

Show that $w.p \geq 1 - \delta$ Successive Elimination Identifies the best arm in $O(\sum_{i \neq i^*}^{n} \Delta_i^{-2} log(n.log(\Delta_i^{-2}))$ Samples.

- Proof:To prove this we show $w.p \geq 1 - \delta$
  - Arm with highest expected reward μ^* will always remain in active set S.
  - All non optimal arms i with reward $\mu_i \leq \mu^*$ will be dropped from S after $O(\sum_{i \neq i^*}^{n} \Delta_i^{-2} log(n.log(\Delta_i^{-2}))$ pulls.

Lemma 1

Let Event $\mathcal{E}$ be the case that for any arm at anytime t,the estimated reward $\hat{\mu}_{i,t}$ is outside the confidence bound around true mean μ_i
$\mathcal{E} = \cup_{i=1}^{n} \cup_{t=1}^{\infty} \{|\hat{\mu}_{i,t} - \mu_i| > U(t, \frac{\delta}{n})\}$
The Event will happen with $\mathbb{P}(\mathcal{E}) \leq \delta$

- Proof by Union Bound
- $\mathbb{P}(\mathcal{E}) \leq \sum_{i=1}^{n} \cup_{t=1}^{\infty} \{|\hat{\mu}_{i,t} - \mu_i| > U(t, \frac{\delta}{n})\} \leq \sum_{i=1}^{n} \frac{\delta}{n} \leq n.\frac{\delta}{n} \leq \delta$

Theorem 1:Part 1

With probability $\geq 1-\delta$, the best arm remians in the active set S until termination.

- Proof:Arm i will only be dropped from set S if $\exists j \; s.t$ $\hat{\mu}_{j,t} - U(t, \frac{\delta}{n}) \geq \hat{\mu}_{i,t} + U(t, \frac{\delta}{n})$
- Additionally when $\mathcal{E}^c$ holds we know that estimated rewards are always within a confidence bound around the true mean and so $\mu_j + U(t, \frac{\delta}{n}) \geq \hat{\mu}_{j,t}$ and $\mu_j - U(t, \frac{\delta}{n}) \leq \hat{\mu}_{i,t}$
- Plugging in the above equation $\implies \mu_j \geq \mu_i$
- Using Lemma 1 we have $\mathbb{P}(\mathcal{E}^c) \geq 1-\delta$

Theorem 1:Part 2

All non optimal arms i with reward $\mu_i \leq \mu^*$ will be dropped from S after $O(\sum_{i \neq i^*}^{n} \Delta_i^{-2} log(n.log(\Delta_i^{-2}))$ pulls.

- By the rules of Successive Elimination described above ,arm i will be removed from the set S if $\hat{\mu}_t^* - U(t, \frac{\delta}{b}) \geq \hat{\mu}_{i,t} + U(t, \frac{\delta}{n})$ where $\hat{\mu_t^*}$ is the estimated reward of the arm with highest expected reward.
- $\therefore$ if $\mathcal{E}^c$ holds estimated rewards are within the confidence bound around true mean.
- $\implies \hat{\mu}_t^* \geq \mu^* - U(t, \frac{\delta}{n})$ and $\hat{\mu}_{i,t} \leq \mu_{i,t} + U(t, \frac{\delta}{n})$
- $\implies \mu^* - 2U(t, \frac{\delta}{n}) \geq \mu_i + 2U(t, \frac{\delta}{n})$
- $\implies \Delta_i \geq 4U(t, \frac{\delta}{n})$
- By Solving minimum value of T we get $T \leq \sum_{i \neq i^*}^{n} c.\Delta_i^{-2} log(\frac{n.log(\Delta_i^{-2}}{\delta})$ for some c.

Median Elimination Algorithm

Median Elimination$(\{1, 2, 3..., n\}, \epsilon, \delta)$
$S \leftarrow \{1, 2, 3..., n\}, \epsilon_1 = \frac{\epsilon}{4}, \delta_1 = \frac{\delta}{2}, l = 1$
while $1 \leq l \leq \infty$ **do**
    Sample every arm a in S for $\frac{1}{(\frac{\epsilon_l}{2})^2} \log(\frac{3}{\delta_l})$
    Let p_a denote its received reward
    Find the median(m_l) of received rewards p_a of all arms a in S
    $\epsilon_{l+1} = \frac{3}{4}\epsilon_l, \delta_{l+1} = \frac{\delta_l}{4}, l = l + 1$
    Stop when $|S| = 1$
end while
return S
end procedure

Theorem 1

Theorem 1: Median Elimination(ϵ, δ) is (ϵ, δ) PAC Algorithm with Average Sample Complexity $O(\frac{n}{\epsilon^2}\log(\frac{1}{\delta}))$.
Lemma1: For every phase l in Meadian Elimination Algorithm where $\mathbb{P}(max_{j\in S_l}\ p_j \leq max_{i\in S_{l+1}}\ p_i + \epsilon_l) \geq 1 - \delta_l$

- Proof Lemma1:$w.l.o.g$ consider $l = 1$
- E_1 :$\hat{p}_1 \leq p_1 - \frac{\epsilon}{2}$ where arm 1 is the best arm at level l
- $\mathbb{P}(E_1) \leq \frac{\delta_1}{3}$
- In the case E_1 does not hold ,calculate the probability that an arm j which is not ϵ_1 optimal arm is empirically better than best arm.
- $\mathbb{P}(\hat{p}_j \geq \hat{p}_1 | \hat{p}_1 \geq p_1 - \frac{\epsilon_1}{2}) \leq \mathbb{P}(\hat{p}_j \geq p_j + \frac{\epsilon_1}{2} | \hat{p}_1 \geq p_1 - \frac{\epsilon_1}{2}) \leq \frac{\delta}{3}$

Lemma1 Proof Contd

- Let the $\#bad$ be the number of arms which are not ϵ_1 optimal but are empirically better than the best arm.We have
- $E[\#bad|\hat{p}_1 \geq p_1 - \frac{\epsilon}{2}] \leq \frac{n\delta_1}{3}$
- $\mathbb{P}(\#bad \geq \frac{n}{2}|\hat{p}_1 \geq p_1 - \frac{\epsilon}{2}] \leq \frac{\frac{n\delta_1}{3}}{\frac{n}{2}} = \frac{2\delta_1}{3}$
- Using Union Bound gives $\mathbb{P}(Failure) \leq \delta_1$

Lemma 2

Lemma 2: Sample Complexity is $O(\frac{n}{\epsilon^2}\log(\frac{1}{\delta}))$.

- $\sum_{l=1}^{log_2(n)} n_l \frac{log(\frac{3}{\delta_l})}{(\frac{\epsilon_l}{2})^2} = 4\sum_{l=1}^{log_2(n)} \frac{\frac{n}{2^{l-1}} log(\frac{2^l.3}{\delta})}{((\frac{3}{4})^{l-1}\frac{\epsilon}{4})^2}$
- $\leq \frac{64n\log(\frac{1}{\delta})}{\epsilon^2}\sum_{l=1}^{log_2(n)} \frac{8}{9}^{l-1}(lC' + C) = O(\frac{n}{\epsilon^2}\log(\frac{1}{\delta}))$

Theorem 1 proof contd

- Proved Sample Complexity is $O(\frac{n}{\epsilon^2}\log(\frac{1}{\delta}))$.
- Algorithm fails with probability δ_l in every round ,therefore $\sum_{l=1}^{log_2(n)} \delta_l \leq \delta$
- In each round we reduce the optimal reward of surviving arms by atmost ϵ_l.Therefore total error is bounded by $\sum_{l=1}^{log_2(n)} \epsilon_l \leq \epsilon$

Group Successive elimination Algorithm

- $\mathcal{E} = \cup_{j=1}^{k} \cup_{i=1}^{n} \cup_{t=1}^{\infty} \{|\hat{\mu}_{j,i,t} - \mu_{j,i}| > U(t, \frac{\delta}{nk})\}$ where n:arms/group, k:number of groups
- $\mathbb{P}(\mathcal{E}) \leq \delta$ Using Union bound argument
- A group will be dropped from the set if the following two events take place
- Event1:Sum of UCB's of all the arms in the group is less than sum of LCB's of all the arms in some other group.
- Event2:UCB of any arm in the group goes below the minimum threshold specified.
- We need to come up with another event in order to take care of a case where some Group i is eliminated by Group j but group j has some min_violating arm but its UCB exceeds group i.Hence we come up with following
  Event3:Wait until the LCB of all the arms in the group is above the specified threshold.

Event 1

- $\sum_{i=1}^{n}\{|\hat{\mu}_{j,t} - \mu_j| > U(t, \frac{\delta}{nk})\} \geq \sum_{i=1}^{n}\{|\hat{\mu}_{i,t} - \mu_i| > U(t, \frac{\delta}{nk})\}$
- $\mathcal{E}^c$ holds $\implies \hat{\mu}_{j,l,t} < \mu_{j,l} + U(t, \frac{\delta}{nk}) \ \ \forall l \in [1, n]$
- $\mathcal{E}^c$ holds $\implies \hat{\mu}_{i,l,t} > \mu_{i,l} - U(t, \frac{\delta}{nk}) \ \ \forall l \in [1, n]$
- $\implies \sum_{l=1}^{n} \mu_{j,l} > \sum_{l=1}^{n} \mu_{i,l}$
- Define $\Delta_{G,i} = \frac{1}{2n}\sum_{l=1}^{n} \mu_{*,l} - \mu_{i,l}$
- Sine we have $\sum_{i=1}^{n}\{\hat{\mu}_{*,l,t} - U(t, \frac{\delta}{nk})\} \geq \sum_{i=1}^{n}\{\hat{\mu}_{i,l,t} + U(t, \frac{\delta}{nk})\}$ and $\mathcal{E}^c$ holds we have
  $\hat{\mu}_{*,l,t} \geq \mu_{*,l} - U(t, \frac{\delta}{nk})$ and $\hat{\mu}_{i,l,t} \leq \mu_{i,l} + U(t, \frac{\delta}{nk})$
- Plugging into the above equation we get $\Delta_{G,i} \geq 4U(t, \frac{\delta}{n})$
- $T_{i,1} \leq c.\Delta_{G,i}^{-2} log(\frac{nk.log(\Delta_{G,i}^{-2})}{\delta})$

Event 2 & 3

- $\hat{\mu}_{i,I,t} + U(t, \frac{\delta}{nk}) \leq \mu_{th} \ \ \forall i \in [1, k] \ \ , I \in [1, n]$
- If $\mathcal{E}^c$ holds $\implies \hat{\mu}_{i,I,t} \geq \mu_{i,I,t} - U(t, \frac{\delta}{nk}) \implies \mu_{i,I,t} \leq \mu_{th}$
- If $\mathcal{E}^c$ holds
  $\implies \hat{\mu}_{i,I,t} \leq \mu_{i,I,t} + U(t, \frac{\delta}{nk}) \implies 2U(t, \frac{\delta}{nk}) \leq \mu_{th} - \mu_{i,I} = \Delta_{th,i,I}$
- $T_{i,2} = min_I c.\Delta_{th,i,I}^{-2} log(\frac{nk.log(\Delta_{th,i,I}^{-2})}{\delta})$
- $T_{i,3} = max_I c.\Delta_{th,i,I}^{-2} log(\frac{nk.log(\Delta_{th,i,I}^{-2})}{\delta})$
- $T_i = min(T_{i,2}, max(T_{i,2}, T_{i,3}))$

Thank You!