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 $\geq 1-\delta$,t 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
  $T \leq \sum_{i \neq i^*}^{n} c.\Delta_i^{-2} log(n.log(\Delta_i^{-2})$ for some c.

Thank You!