CS6700: Reinforcement Learning - Bandits Assignment

Atul Balaji - EE16B002

February 9, 2019

1 Median Elimination by dropping one-fourth of the arms

We are required to show the working of the median elimination algorithm, when we remove only the worst one-fourth of the arms instead of half.

1.1 Notation

1. **n** = no.of arms
2. $Q_l(j)$ = estimated value of j^{th} non-optimal arm
3. $Q(a_l^*)$ = estimated value of optimal arm
4. $q_*(a_l^*)$ = true value of optimal arm
5. $q_*(j)$ = true value of j^{th} non-optimal arm

1.2 Algorithm

The algorithm changes to the following:

Algorithm : Median Elimination(ϵ, δ)
1. Set S = A
2. $\epsilon_1 = \epsilon/8, \delta_1 = \delta/2, l = 1$
3. Sample every arm a $\in$ S for $\frac{2}{\epsilon_l^2}\ ln(\frac{7}{3\delta_l})$ times, and let $Q_l(a)$ denote its empirical value
4. Find the median of $Q_l(a)$. For the values below the median, find the median again and denote it by m_l
5. $S_{l+1} = S_l \setminus$ {a : $Q_l(a) < m_l$ }
6. If $|S_l| = 1$, then output S_l. Else $\epsilon_{l+1} = \frac{7}{8}\epsilon_l$; $\delta_{l+1} = \delta_l/2$; $l = l + 1$; Go to 3.

1.3 Justification for number of samples

Let us assume that the number of times we need to sample every arm in a step is $\frac{2}{\epsilon_l^2}\ ln(\frac{y}{\delta_l})$.
We consider the proof of Median Elimination being a (ϵ, δ) - PAC Algorithm.
We will now calculate the probability that the best arm at the l^{th} step - a_l^* is eliminated in that step.
This is covered by two events:

Event 1 (E_1) : We grossly **underestimate** the value of the **optimal** arm, i.e
$E_1 = Q(a_l^*) < q_*(a_l^*) - \epsilon_l/2$. $P[E_1]$ can be calculated using the Chernoff-Hoeffding bound, and let its value be $f\delta_l$.

Event 2 (E_2) : We **overestimate** the value of some **non-optimal** arms.
In case E_1 does not hold, we calculate the probability that an arm which is not an optimal arm is empirically better than the true best arm. We have:

$$\mathbf{P}[Q_l(j) \geq Q(a_l^*)|Q(a_l^*) \geq q_*(a_l^*) - \epsilon_l/2)] \leq \mathbf{P}[Q_l(j] \geq q_*(j) + \epsilon_l/2|Q(a_l^*) \geq q_*(a_l^*) - \epsilon_l/2] \leq f\delta_l \tag{1}$$

Let b be the number of non-optimal arms which are empirically better than the best arm. We have :

$$E[b|Q(a_l^*) \geq q_*(a_l^*) - \epsilon_l/2] \leq n\delta_l/3 \tag{2}$$

For the optimal arm to be eliminated, we need three-quarters of the arms to have value above it, i.e $b \geq 3n/4$.
Next, by **Markov inequality**, we have:

$$P[b \geq 3n/4|Q(a_l^*) \geq q_*(a_l^*) - \epsilon_l/2] \leq \frac{nf\delta_l}{3n/4} = 4f\delta_l/3 \tag{3}$$

Therefore, $P[E_2] \leq 4f\delta_l/3$.

However, for the algorithm to be (ϵ, δ) - PAC, $P[E_1] + P[E_2] \leq \delta_l$. This yields max. value of $\mathbf{f} = \mathbf{3/7}$.
Using the **Chernoff-Hoeffding** bound, we therefore obtain y = 7/3.
This means that in each round, we must sample every arm for $\frac{2}{\epsilon_l^2}\ln(\frac{7}{3\delta_l})$ times.

1.4 ϵ-update and Complexity

We must choose our ϵ such that two conditions are satisfied:

- Sum of the ϵ_l's must be $\leq \epsilon$.
- The series represented by the complexity summation must converge.

Let us consider the **update rule** as: $\epsilon_1 = \epsilon/k$ and $\epsilon_{l+1} = \epsilon_l(k-1)/k$. This ensures that the infinite summation of the ϵ_l's is ϵ, and therefore, any finite sum will be $\leq \epsilon$.
We must choose some suitable value of k such that the other condition is satisfied.

The number of samples in the l^{th} round is $\frac{2n_l ln(7/3\delta_l)}{\epsilon_l^2}$. Then, we have:
1. $\delta_1 = \delta/2$; $\delta_l = \delta_{l-1}/2 = \delta/2^l$
2. $n_1 = n$; $n_l = 3n_{l-1}/4 = n(\frac{3}{4})^{l-1}$
3. $\epsilon_1 = \epsilon/k$ and $\epsilon_l = \epsilon_{l-1}(k-1)/k = (\frac{k-1}{k})^{l-1}\epsilon/k$
Therefore, we have:

$$\sum_{l=1}^{log_{4/3}(n)} \frac{2n_l ln(3/\delta_l)}{\epsilon_l^2} = 32 \sum_{l=1}^{log_{4/3}(n)} n(\frac{3}{4}(\frac{k}{k-1})^2)^{l-1}(\frac{\ln(1/\delta)}{\epsilon^2} + \frac{\ln(7/3)}{\epsilon^2} + \frac{l\ln(2)}{\epsilon^2})$$
$$\leq 32\frac{nln(1/\delta)}{\epsilon^2}\sum_{l=1}^{\infty}(\frac{3}{4}(\frac{k}{k-1})^2)^{l-1}(lC' + C)$$

This sum will converge only if $(\frac{3}{4}(\frac{k}{k-1})^2)^{l-1} \leq 1$. The smallest value of k that satisfies this is k = 8.
Therefore, we choose $\epsilon_1 = \epsilon/8$ and $\epsilon_{l+1} = 7\epsilon_l/8$.
Now, the summation becomes:

$$32\frac{nln(1/\delta)}{\epsilon^2}\sum_{l=1}^{\infty}(\frac{48}{49})^{l-1}(lC' + C) \leq O(\frac{n\ln(1/\delta)}{\epsilon^2}),$$

which is the sample complexity of the algorithm.

2 Conclusion

A modified version of Median Elimination Algorithm has been presented for the case where the bottom one-fourth of the arms are dropped in every round.
The algorithm has a sample complexity of $O(\frac{n\ln(1/\delta)}{\epsilon^2})$, same as that of the original algorithm.