Stochastic Recursive Optimization: A Structured Multi-Armed Bandit Problem

Roland Szögi
Supervisor: Balázs Csanád Csáji

Multi-Armed Bandits

- The multi-armed bandit model consists of a set of arms $\mathcal{A}$ ($n = |\mathcal{A}|$).
- To every arm $a \in \mathcal{A}$ belongs a distribution $\nu(a)$.
- In each round an arm $a \in \mathcal{A}$ is chosen and a reward $R(a)$ is sampled from distribution $\nu(a)$.
- An arm is called optimal, if it has the highest expected reward among all of the arms.
- The goal is to identify the optimal arm.

Further assumptions

1. There are $n = 2^m + 1$, $m \geq 1$ arms numbered from 0 to 2^m: $a_0, a_1, \ldots, a_{2^m}$.
   (The expectation of arm a_i will be denoted by μ_i.)
2. There exists a $k \in \{0, 1, \ldots, 2^m\}$ such that

$$\mu_0 < \mu_1 < \ldots < \mu_{k-1} < \mu_k > \mu_{k+1} > \mu_{k+2} > \ldots > \mu_{2^m}.$$

3. There exists a $\Delta > 0$ such that

$$|\mu_{i+1} - \mu_i| \geq \Delta \quad \forall i \in \{0, 1, \ldots, 2^m - 1\},$$

   and it is known in advance.
4. The arms are 1-subgaussian.

Definition 2.1

A random variable X is σ-subgaussian if for all $\lambda \in \mathbb{R}$:

$$\mathbb{E}[\exp(\lambda X)] \leq \exp(\lambda^2 \sigma^2 / 2).$$

Algorithm

Input: $\delta > 0$
Output: an arm which is optimal with probability at least $1 - \delta$

Set $S_1 = \mathcal{A}$, $\delta_1 = \delta/2$, $\ell = 1$.
while $|S_\ell| > 3$ **do**
  Sample the three arms: a_j , $j \in \{i \cdot 2^{m-\ell-1}, i = 1, 2, 3\}$
  $n_\ell = \lceil \log(4/\delta_\ell)/(2^{2m-2\ell-5}\Delta^2) \rceil$ times each, and let $\hat{\mu}_j^\ell$ denote their empirical values
  $i_\ell^* = \arg\max_j \hat{\mu}_j^\ell$
  $S_{\ell+1} = \left\{a_i : i_\ell^* - 2^{m-\ell-1} \leq i \leq i_\ell^* + 2^{m-\ell-1}\right\}$
  Renumber the arms from 0 to $2^{m-\ell}$
  $\delta_{\ell+1} = \delta_\ell/2$, $\ell = \ell + 1$
end while
Sample each of the three remaining arms a_j, $j \in \{0, 1, 2\}$
$n_m = \lceil \log(4/\delta_m)/(2^{-3}\Delta^2) \rceil$ times, and let $\hat{\mu}_j^m$ denote their empirical values
$i_m^* = \arg\max_j \hat{\mu}_j^m$
return $a_{i_m^*}$

Theorem 1

Under the assumptions, the algorithm finds the optimal arm with probability at least $1 - \delta$ *and its sample complexity is*

$$\mathcal{O}\left(\log n + \frac{1}{\Delta^2} \log \frac{n}{\delta}\right).$$

General Case

In the general case when $2^{m-1}+1 < n \leq 2^m + 1$ we can do the following:

1. Update the indices:
$$i \leftarrow i + \left\lfloor \frac{2^m + 1 - n}{2} \right\rfloor .$$
2. Sample the arms a_j, $j \in \{i \cdot 2^{m-2},\ i = 1, 2, 3\}$, $n_1 = \lceil \log(8/\delta)/(2^{2m-7}\Delta^2) \rceil$ times each.
3. Let $\hat{\mu}_j^1$ denote their empirical values and let $i_1^* = \arg\max_j \hat{\mu}_j^1$.
4. Keep the $2^{m-1}+1$ arms closest to the arm $a_{i_1^*}$.
5. Renumber the arms starting from 0.
6. Continue with the algorithm.

Conclusion

In the special case I have investigated, the optimal arm could be found using the Median Elimination algorithm, however this new algorithm finds the optimal arm much faster, with a sample complexity of

$$\mathcal{O}\left(\log n + \frac{1}{\Delta^2} \log \frac{n}{\delta}\right)$$

instead of

$$\mathcal{O}\left(\frac{n}{\Delta^2} \log \frac{1}{\delta}\right).$$

Thank You For Your Attention!