STATS 710 – Sequential Decision Making with mHealth Applications

Note: Bandits with heavy tailed distributions

Instructors: Susan Murphy and Ambuj Tewari *Scribe: Anurag Beniwal*

1 Introduction

In the bandit algorithms discussed so far in the class, the reward distribution was assumed to be sub-gaussian. In this note, we discuss the bandit problems for heavy tailed distributions and robust estimators of mean which can enable similar regret bounds as that of sub-gaussian setting in heavy tailed distributions.We will see that moment of order 2 and moment generating function not necessarily being finite (a weaker condition than sub-gaussian) is sufficient to prove regret bounds of same order as sub-gaussian with these robust estimators. With the existing estimator also it is possible to achieve a logarithmic bound in n but the dependence on Δ_i deteriorates with the tail of distribution becoming heavier .

If the 2^{nd} moment is not finite but the moment of order $(1+\epsilon)$ exists with $\epsilon \in (1,2)$, we can achieve logarithmic guarantees in n but the dependence on Δ_i worsens with ϵ_i getting smaller.

We will discuss the performance and regret bounds of the robust estimators in the context of UCB algorithm. In UCB, we choose an arm for which the sum of mean and confidence interval is maximum. When reward distribution are sub-Gaussian with a common variance factor v, then this confidence interval is easy to obtain as we can use Chernoff bounds to get bound on the sample mean. For the rewards $X_{i,t}$ where i is the i_{th} arm and t is the time point such that $\mathbf{E}X_{i,t} = \mu_i$, using Chernoff bounds, for any $\delta \in (0,1)$, the empirical mean satisfies with probability at least $1-\delta$,

$$\frac{1}{s}\sum_{r=1}^{s} X_{i,r} \leq \mu_i + \sqrt{\frac{2v\log\frac{1}{\delta}}{s}}$$

However, we cannot have the sample mean to behave in a similar manner for heavy tailed distributions. We would want to have similar guarantees for the robust estimators for heavy tailed distributions. We need the mean estimator with following property. This assumption would serve as a pivot in proving regret bounds for all robust estimators

Assumption 1: Let $\epsilon \in (0,1]$ be a positive parameter and let c, v be positive constants.Let $X_1 ...., X_n$ be i.i.d. random variables with finite mean μ Suppose that for all $\delta \in (0,1)$ there exists an estimator $\hat{\mu} = \hat{\mu}(n,\delta)$ such that ,with probability at least $1-\delta$

$$\hat{\mu} \leq \mu + v^{\frac{1}{(1+\epsilon)}} (\frac{c\log\frac{1}{\delta}}{n})^{\frac{\epsilon}{(1+\epsilon)}}$$

and also with probability at least $1-\delta$

$$\mu \leq \hat{\mu} + v^{\frac{1}{(1+\epsilon)}} (\frac{c\log\frac{1}{\delta}}{n})^{\frac{\epsilon}{(1+\epsilon)}}$$

The condition for sub-gaussian rewards can be derived from this more general condition by plugging $\epsilon = 1$ and $c = 2$. With robust estimators, this condition can be satisfied by a more general class of distributions.

Therefore we can write a generalize version of UCB with robust estimator $\hat{\mu}_{i,s,t}$ as

Robust UCB:
Parameter: $\epsilon \in (0,1]$, mean estimator $\hat{\mu}(t,\delta)$.
For arm i, define $\hat{\mu}_{i,s,t}$ as the estimate $\hat{\mu}(s,t^{-2})$ based on the first s observed values $X_{i,1},....X_{i,s}$ of the rewards of arm i.
Define the index

$$B_{i,s,t} = \hat{\mu}_{i,s,t} + v^{\frac{1}{1+\epsilon}}(\frac{c\log t^2}{s})^{\frac{\epsilon}{(1+\epsilon)}}$$

for $s,t \geq 1$ and $B_{i,0,t} = +\infty$

For the Robust-UCB algorithm the regret bound is given by the following result.

Theorem 1. *Let $\epsilon \in (0,1]$ and let $\hat{\mu}(s,\delta)$ be a mean estimator . Suppose that the distributions $\nu_1,....\nu_K$ are such that the mean estimator satisfies Assumption 1 for all i = 1,......,K. Then the regret of the robust UCB policy satisfies*

$$\mathcal{R}_n \leq \sum_{i:\Delta_i>0} (2c(\frac{v}{\Delta_i})^{\frac{1}{\epsilon}} \log n + 5\Delta_i)$$

Also if n is such that $\log n \geq \max_i(\frac{5\Delta_i^{\frac{(1+\epsilon)}{\epsilon}}}{2cv^{\frac{1}{\epsilon}}})$ *then*

$$\mathcal{R}_n \leq n^{\frac{1}{1+\epsilon}}(4Kc\log n)^{\frac{\epsilon}{1+\epsilon}} v^{\frac{1}{(1+\epsilon)}}$$

1.1 Truncated empirical mean

Truncated mean is a robust estimator. It involves the calculation of the mean after discarding given parts of a probability distribution or sample at the high and low end, and typically discarding an equal amount of both.

Let $\delta \in (0,1)$, $\epsilon \in (0,1]$, and $u > 0$. The truncated empirical mean $\hat{\mu}_T$ is defined as:
$\hat{\mu}_T = \frac{1}{n}\sum_{t=1}^{n} X_t \mathbb{1}[\mid X \mid \leq (\frac{ut}{\log(\delta^{-1})})^{\frac{1}{1+\epsilon}}]$

Concentration Inequality:
If the $(1+\epsilon)^{th}$ moment is bounded, i.e.,$\mathbf{E} \mid X \mid^{1+\epsilon} \leq u$, then with probability at least $1-\delta$

$\hat{\mu}_T \leq \mu + 4u^{\frac{1}{1+\epsilon}}(\frac{\log(\delta^{-1})}{n})^{\frac{\epsilon}{1+\epsilon}}$

Regret Bound
Let $\epsilon \in (0,1]$ and $u > 0$. Assume that the reward distributions $\nu_1,.....\nu_K$ satisfy
$\mathbf{E}_{X\sim\nu_i} \mid X \mid^{1+\epsilon} \leq u \; \forall i \in \{1,.....K\}$
then the regret of Robust-UCB policy used with the truncated mean estimator defined above satisfies

$\mathcal{R}_n \leq \sum_{i:\Delta_i>0} (8(\frac{4u}{\Delta_i})^{\frac{1}{\epsilon}} \log n + 5\Delta_i)$

Merits & Demerits:
Merits: Computationally efficient, requires constant time and space for each update
Demerits: The bound depends on raw $(1+\epsilon)$-moments which makes it not translational invariant. The arms selected by the strategy might change if all reward distributions are shifted by same constant amount. It requires the bound u on these moments.

1.2 Median of Means

The data is divided into various disjoint blocks or epochs. Within each epoch mean of epoch is calculated. The median of means of all epochs is used as a robust estimator .

Concentration Inequality:
Let $\delta \in (0,1)$ and $\epsilon \in (0,1]$. Let $X_1, ...X_n$ be i.i.d. random variables with mean $\mathbf{E}X = \mu$ and centered $(1+\epsilon)$-th moment $\mathbf{E} \mid X - \mu \mid^{1+\epsilon} = u$. Let $k = \lfloor 8 \log(\frac{e^{\frac{1}{8}}}{\delta}) \wedge \frac{n}{2} \rceil$, $N = \lfloor \frac{n}{k} \rceil$ and

$$\hat{\mu}_1 = \frac{1}{N} \sum_{t=1}^{N} X_t,\ \hat{\mu}_2 = \frac{1}{N} \sum_{t=N+1}^{2N} X_t,\ \ldots,\ \hat{\mu}_k = \frac{1}{N} \sum_{t=(k-1)N+1}^{kN} X_t.$$

Then, $\hat{\mu}_M \leq \mu + (12\upsilon)^{\frac{1}{1+\epsilon}} (\frac{16 \log(e^{\frac{1}{8}} \delta^{-1})}{n})^{\frac{\epsilon}{1+\epsilon}}$

Regret Bound
Let $\epsilon \in (0,1]$ and $\upsilon > 0$. Assume that the reward distributions $\nu_1, ...\nu_K$ satisfy
$\mathbf{E}_{X \sim \nu_i} \mid X - \mu_i \mid^{1+\epsilon} \leq \upsilon, \forall i \{1, ...K\}$
Then the regret of the Robust-UCB policy used with the median of means mean estimator satisfies

$\mathcal{R}_n \leq \sum_{i:\Delta_i > 0} 32 (\frac{12\upsilon}{\Delta_i})^{\frac{1}{\epsilon}} \log n + 5\Delta_i)$

Merits & Demerits:
Merits:
The regret bound is dependent on central moments unlike truncated mean . This makes it a translation invariant estimator. It doesn't require knowledge of bounds of the $(1+\epsilon)$ moments.
Demerits:
It is not as computationally efficient as truncated mean estimator as it requires $O(\log \delta^{-1})$ space and $O(\log \log \delta^{-1})$ time per update

1.3 Catoni's M-estimator

This estimator has similar performance guarantees as Median of Means but has better, near optimal numerical constants. The estimator can be defined as follows. Let $\psi : \mathbf{R} \to \mathbf{R}$ be a continuously strictly increasing function satisfying

$$-\log(1 - x + \frac{x^2}{2}) \leq \psi(x) \leq \log(1 - x + \frac{x^2}{2}).$$

Let $\delta \in (0,1)$ be such that $n > 2\log(\frac{1}{\delta})$ and introduce

$$\alpha_\delta = \sqrt{\frac{2\log(\frac{1}{\delta})}{n(v + \frac{2v\log(1/\delta)}{n-2\log(1/\delta)})}}.$$

If $X_1, ...X_n$ be i.i.d random variables , then Cantoni's estimator is defined as the unique value $\hat{\mu}_C = \hat{\mu}_C(n,\delta)$ such that

$$\sum_{t=1}^{n} \psi(\alpha_\delta(X_i - \hat{\mu}_C)) = 0.$$

Concentration Inequality:

If $n \geq 4\log(1/\delta)$ and X_i have a mean μ and variance at most v, then with probability at least $1-\delta$

$$\hat{\mu}_C = \mu + 2\sqrt{\frac{v\log\frac{1}{\delta}}{n}}.$$

Regret Bounds:
Let $v > 0$ and assume that the reward distributions $\nu_1, ..., \nu_K$ satisfy

$$\mathbf{E}_{X\sim\nu_i} \mid X - \mu_i \mid^{1+\epsilon} \leq v, \forall i\{1, ...K\}.$$

Then,

$$\mathcal{R}_n \leq \sum_{i:\Delta_i>0} (\frac{8vlogn}{\Delta_i} + 8\Delta_i logn + 5\Delta_i).$$

Merits & Demerits:

Merits: Better Numerical constants in terms of optimality and same performance as Median of Means.

Demerits: Has validity on a restricted range which leads to an extra term $\sum_i \Delta_i \log n$ term.

This note is a summary of [BCBL13]

References

[BCBL13] Sébastien Bubeck, Nicolo Cesa-Bianchi, and Gábor Lugosi. Bandits with heavy tail. *IEEE Transactions on Information Theory*, 59(11):7711–7717, 2013.