Theorems and Proofs

Tim Radtke

4/23/2017

Theorem 2

Let $K \geq 0$, $T \geq 2K$, and consider a problem $\mathbb{B}$. Assume that all arms ν_k of the problem are R-sub-Gaussian with means μ_k. Let $\tau \in \mathbb{R}, \epsilon \geq 0$.

Algorithm APT's expected loss is upper bounded on this problem as

$$\mathbb{E}(\mathcal{L}(T)) \leq \exp(-\frac{1}{64R^2}\frac{T}{H} + 2\log(T) + 1)K))$$

where we remind that $H = \sum_i (|\mu_i - \tau| + \epsilon)^{-2}$ and where $\mathbb{E}$ is the expectation according to the samples of the problem.

Proof

To prove above theorem, we first define a favorable event ξ on which we hope to make a loss with only a small probability. Indeed, it is the event for which we want to upper bound the expected loss. As we want to find an upper bound for the expected loss of the APT algorithm, the event corresponds to the decision made by it. The APT algorithm's heuristic leads to a decision with low expected regret, if it finds empirical mean estimates $\frac{1}{s}\sum_{t=1}^{s} X_{i,t}$ of the arms that are close to the true mean mu_i of the arms. Then, we would expect arms which are over $\tau + \epsilon$ to be classified correctly, and arms that are under $\tau - \epsilon$ to be rejected.

Indeed, we have $\mathbb{E}(\mathcal{L}(T)) = P(S_{\tau+\epsilon} \cap \hat{S}_\tau^C \neq \emptyset \vee S_{\tau-\epsilon}^C \cap \hat{S}_\tau \neq \emptyset)$. But this just corresponds to $P(\{\exists i : ((\mu_i < \tau - \epsilon) \wedge (\hat{\mu}_i \geq \tau)) \vee ((\mu_i > \tau + \epsilon) \wedge (\hat{\mu}_i < \tau)))$. Each of the two $(...) \vee (...)$ conditions corresponds to $|\hat{\mu}_i - \mu_i| \geq \epsilon$. It follows that the expected loss can be rewritten as: $\mathbb{E}(\mathcal{L}(T)) = P(\exists i, \exists s : |\hat{\mu}_{i,s} - \mu_i| \geq \epsilon) = P(\xi^C)$. But this is exactly the probability of classifying at least one arm incorrectly which results in a loss of size 1.

Thus we define as the favorable event:

$$\xi = \{\forall i, \forall s : |\hat{\mu}_{i,s} - \mu_i| \leq \epsilon\} \tag{1}$$

Furthermore, if we write $\xi_{i,s} = \{|\hat{\mu}_{i,s} - \mu_i| \leq \epsilon\}$, then the following holds by union bound.

$$P(\xi) = P(\cap_i \cap_s \xi_{i,s}) = 1 - P(\cup_i \cup_s \xi_{i,s}^C) \geq 1 - \sum_i \sum_s P(|\hat{\mu}_{i,s} - \mu_i| \geq \epsilon) \tag{2}$$

Furthermore, for any $i \in \{1, \ldots, K\}$, for any $s \in \{1, \ldots, T(i)\}$, with X_{it} being the random sample from the R-sub-Gaussian distribution v_i at time t:

$$P(|\hat{\mu}_{i,s} - \mu_i| \geq \epsilon) = P(|(\frac{1}{s}\sum_{t=1}^{s} X_{it} - \mu_i)| \geq \epsilon) = P(|\sum_{t=1}^{s}(X_{it} - \mu_i)| \geq s\epsilon) \tag{3}$$

where the latter can be bounded by a Hoeffding inequality:

$$P(|\sum_{t=1}^{s}(X_{it} - \mu_i)| \geq s\epsilon) \leq 2\exp(-\frac{\epsilon^2 s}{2R^2}) \tag{4}$$

The Hoeffding inequality as used here is as follows:

...

Plugging (4) back into (2), we get:

$$P(\xi) \geq 1 - \sum_i \sum_s P(|\hat{\mu}_{i,s} - \mu_i| \geq \epsilon) \geq 1 - 2 \sum_i \sum_s \exp(-\frac{\epsilon^2 s}{2R^2}) \tag{5}$$

Now, in order to match the performance of the lower bound on the expected loss derived in theorem 1, we will choose $\epsilon = \sqrt{\frac{T\delta^2}{Hs}}$. Thus:

$$P(\xi) \geq 1 - 2 \sum_i \sum_s \exp(-\frac{\epsilon^2 s}{2R^2}) = 1 - 2 \sum_i \sum_s \exp(-\frac{T\delta^2}{2R^2 H}) \geq 1 - TK \exp(-\frac{T\delta^2}{2R^2 H}) \tag{6}$$

Thus, we don't exactly get the bound from Locatelli et al. (2016) because we used the Hoeffding inequality instead of the martingale bound; but the the method should be clear.

What is left now is to show that the favorable event still holds under the proposed ϵ, that is, we need to show that under $P(\xi) \geq 1 - TK \exp(-\frac{T\delta^2}{2R^2 H})$ we still reject arms that are under $\tau - \epsilon$ and accept those above $\tau + \epsilon$.

Step 2: Lower Bound on Number of Pulls of Difficult Arm

In order to see that arms are correctly classified with large probability, we need to show that they are are pulled sufficiently often, that is, we have a degree of exploration for each arm i that allows us to identify its characteristics. To see this, rewrite the favorable event of a single arm as follows with $s = T(i)$:

$$\xi_{i,s} = \{|\hat{\mu}_{i,s} - \mu_i| \leq \epsilon\} = \{|\hat{\mu}_{i,s} - \mu_i| \leq \sqrt{\frac{T\delta^2}{HT(i)}}\} \tag{7}$$

To make sure that this deviation is small enough, we need to show that at time T, all $T(i)$ have a sufficient lower bound and thus all arms i have been explored sufficiently to classify correctly with probability larger $TK \exp(-\frac{T\delta^2}{2R^2 H})$.

To start off, notice that due to the initialization of the APT algorithm, every $T_i(T) \geq 1$. Furthermore, we have assumed that $T > 2K$. Now, at time T, consider an arm k that has been pulled $T_k(T) - 1 \geq \frac{(T-K)}{H\Delta_k^2}$ times. That is, it has been pulled at least proportionally to it's contribution to the overall hardness of the problem H. Recall $H = \sum_{i=1}^{K}(|\mu_i - \tau| + \epsilon)^{-2}$ and $\delta_i^2 = (|\mu_i - \tau| + \epsilon)^2$. To see that such arm k exists, assume it does not. Then

$$T_i(T) - 1 < \frac{T - K}{H\Delta_i^2} \forall i$$

$$T - K = \sum_{i=1}^{K} T_i(T) - 1 < \sum_{i=1}^{K} \frac{T - K}{H\Delta_i^2} = \frac{T - K}{H\Delta_i^2} \sum_{i=1}^{K} \frac{1}{\Delta_i^2} = T - K$$

which is a contradiction.

Consequently, using the assumption $T > 2K$ from above, we have

$$T_k(T) \geq T_k(T) - 1 \geq \frac{T - K}{H\Delta_k^2} = \frac{T}{2H\Delta_k^2}$$

This states a lower bound on the amount of pulls $T_k(t)$ for an arm k that contributes more than average to the overall hardness of the problem.

Step 3: Lower Bound on Number of Pulls of Simple Arm

Next, we would like to derive bounds on the number of pulls of the remaining arms i at the time t, the last pull of arm k. To do so, consider that we are on the favorable event ξ. For every arm i, it holds:

$$|\hat{\mu}_i(t) - \mu_i| \leq \sqrt{\frac{T\delta^2}{HT_i(t)}} \tag{8}$$

As described in Locatelli et al., using the reverse triangle inequality, we can write

$$\begin{aligned}
|\hat{\mu}_i(t) - \mu_i| &= |(\hat{\mu}_i(t) - \tau) - (\mu_i - \tau)| \\
&\geq ||\hat{\mu}_i(t) - \tau| - |\mu_i - \tau|| \\
&= |(|\hat{\mu}_i(t) - \tau| + \epsilon) - (|\mu_i - \tau| - \epsilon)| \\
&= |\hat{\Delta}_i(t) - \Delta_i|
\end{aligned}$$

Consequently, we can again employ the hardness of the task to start deriving bounds on the number of pulls of the arms:

$$|\hat{\Delta}_i(t) - \Delta_i| \leq |\hat{\mu}_i(t) - \mu_i| \sqrt{\frac{T\delta^2}{HT_i(t)}}$$

From this, we get

$$\Delta_i - \sqrt{\frac{T\delta^2}{HT_i(t)}} \leq \hat{\Delta}_i(t) \leq \Delta_i + \sqrt{\frac{T\delta^2}{HT_i(t)}}$$

and in particular for the arm k discussed above

$$\Delta_k - \sqrt{\frac{T\delta^2}{HT_k(t)}} \leq \hat{\Delta}_k(t) \leq \Delta_k + \sqrt{\frac{T\delta^2}{HT_k(t)}} \tag{9}$$

To bound $T_i(t)$, it is helpful to compare Δ_i and Δ_k at time t. From the definition of the APT algorithm, we know that since arm k has been pulled at time t ("Pull arm $I_{t+1} = \arg\min_i \mathcal{B}_i(t+1)$"):

$$\begin{aligned}
\mathcal{B}_k(t) &\leq \mathcal{B}_i(t) \\
\sqrt{T_k(t)}\hat{\Delta}_k(t) &\leq \sqrt{T_i(t)}\hat{\Delta}_i(t)
\end{aligned}$$

Thus, every arm i has a lower bound on it's number of pulls given by the lower bound for arm k proportional with it's inverse relative estimated hardness $\hat{\Delta}_k(t)/\hat{\Delta}_i(t)$. We can easily derive a lower bound for the left hand side by plugging in (9):

3

$$\sqrt{T_k(t)}(\Delta_k - \sqrt{\frac{T\delta^2}{HT_k(t)}}) \leq \mathcal{B}_k(t) \tag{10}$$

$$\sqrt{\frac{T}{2H\Delta_k^2}}(\Delta_k - \sqrt{2}\Delta_k\delta) = \sqrt{\frac{T}{H}}(\frac{1}{\sqrt{2}} - \delta) \tag{11}$$

$$\leq \mathcal{B}_k(t) \tag{12}$$

since $T_k(t) \geq T/(2H\Delta_k^2)$ and $\sqrt{T\delta^2/(HT_k(t))} \geq \sqrt{T\delta^2/(H\frac{T}{2H\Delta_k^2})} = \sqrt{2}\Delta_k\delta$.

Next, upper bound $\mathcal{B}_i(t)$. In contrast to $T_k(t)$, there is no lower bound for $T_i(t)$ available yet that we could plug in. Thus we derive:

$$\mathcal{B}_i(t) = \sqrt{T_i(t)}\hat{\Delta}_i(t)$$
$$\leq \sqrt{T_i(t)}(\Delta_i + \sqrt{\frac{T\delta^2}{HT_i(t)}})$$
$$= \sqrt{T_i(t)}\Delta_i + \delta\sqrt{\frac{T}{H}}$$

Combining the two bounds, we get a lower bound on the pulls for every other arm i:

$$\Delta_i\sqrt{T_i(t)} + \delta\sqrt{\frac{T}{H}} \geq (\frac{1}{\sqrt{2}} - \delta)\sqrt{\frac{T}{H}}$$
$$\sqrt{T_i(t)} \geq (\frac{1}{\sqrt{2}} - 2\delta)\sqrt{\frac{T}{H}}\frac{1}{\Delta_i}$$
$$\text{(choose } 2\delta > 1/\sqrt{2} \text{ such that RHS is greater 0)}$$

Conclusion

As stated before, we require the lower bound on $T_i(t)$ in order to show that $|\hat{\mu}_i(t) - \mu_i| \leq \sqrt{frac{T\delta^2}{HT_i(t)}}$ holds for all i. So if we now simply plug in the derived lower bound, we get:

$$\mu_i - \Delta_i(\frac{\sqrt{2}\delta}{1 - 2\sqrt{2}\delta}) \leq \hat{\mu}_i(t) \leq \mu_i + \Delta_i(\frac{\sqrt{2}\delta}{1 - 2\sqrt{2}\delta})$$

and with $\lambda := (\frac{\sqrt{2}\delta}{1-2\sqrt{2}\delta}) > 0$:

$$\mu_i - \Delta_i\lambda \leq \hat{\mu}_i(t) \leq \mu_i + \Delta_i\lambda$$

Under this formulation on ξ, do we accept arms with $\mu_i \geq \tau + \epsilon$? For these arms holds that $\Delta_i = \mu_i - \tau + \epsilon$. Consequently:

$$\mu_i - (\mu_i - \tau + \epsilon)\lambda \leq \hat{\mu}_i(t)$$
$$\hat{\mu}_i(t) \geq \tau\lambda + \mu(1 - \lambda) - \epsilon\lambda$$

The arms are accepted if

$$\hat{\mu}_i(t) - \epsilon \geq 0$$

Or equivalently as long as

$$\tau\lambda + \mu(1 - \lambda) - \epsilon\lambda - \tau \geq 0$$
$$\lambda \leq \frac{1}{2}$$