Theorems and Proofs

Tim Radtke

4/23/2017

0.1 Theorem 1

In Theorem 1 of Locatelli et al. (2016), a lower bound for the thresholding bandit problem is shown. The basic idea is as follows. The problem formulation considered in the paper describes bandit problems for Sub-Gaussian distributions. To show a lower bound, we show that every algorithm is expected to make a certain error even if we restrict ourselves to a heavily specified scenario. Even then, any algorithm will have a certain level of risk with which it cannot distinguish between the actual setting and a slightly changed setting (under which the former answer leads to regret).

To be more specific, consider a case in which we set $\tau = 0$ and $\epsilon = 0$ w.l.o.g. If we then choose all arms to have normally distributed feedback with means μ_i, we see that for each arm i the distance from the threshold is given by $\Delta_i = |\mu_i - \tau| + \epsilon = \mu_i$. Thus, assuming a variance of 1 for every arm, we have for every arm the distribution $\nu_i = \mathcal{N}(\mu_i, 1) = \mathcal{N}(\Delta_i, 1)$. Furthermore, define $\nu_i' = \mathcal{N}(-\Delta_i, 1)$.

The idea now is that every arm $i \in \{1, \ldots, K\}$ generates a sample for every time point $t \in \{1, \ldots, T\}$ from its distribution ν_i before the algorithm starts to choose arms. Thus, we have a $T \times K$ table of realized observations from the product distribution given by $\mathcal{B}^0 = \nu_1^0 \otimes \ldots \otimes \nu_K^0 = \nu_1^0 \otimes \ldots \otimes \nu_K$. This is a product distribution where the K different means all lie above the threshold. We would hope that an algorithm classifies the means accordingly. However, we will show in what follows that any algorithm makes a certain error by not being able to distinguish $\mathcal{B}^0$ from a product distribution $\mathcal{B}^i$ which differs only slightly and has the same overall complexity (as defined by $H = \sum_{i=1}^{K} (\Delta_i)^{-2}$).

Before the start of the algorithm, we introduce a slight variation in the setup. Similar to how one can switch arms in the pure exploration setup for multi-armed bandits (compare Audibert et al., Garivier & Kaufmann) to "confuse" the algorithm without introducing additional complexity to the specified problem. However, since we do not compare arms with each other directly in the thresholding bandit problem (switching arms would still keep all means above the threshold), we instead need to flip an arm with respect to the threshold. Since the threshold is $\tau = 0$, we have for the flipped arm the new distribution ν_i' as defined above. Indeed, we can define the new product distribution $B^i = \nu_1^i \otimes \cdots \otimes \nu_K^i$ where for $k \leq K, \nu_k' := \nu_i \mathbb{1}_{k \neq i} + \nu_i' \mathbb{1}_{k=i}$. This flipping of the arm corresponds to multiplying all random variables $X_{i,t}$ of arm i by -1.

As in Locatelli et al. (2016), we let for $i \leq K$, $\mathbb{P}_{\mathcal{B}^i}$ define the probability distribution describing all samples that a bandit strategy can potentially collect up to horizon T, i.e. according to the samples $(X_{k,s})_{k \leq K, s \leq T} \sim (\mathcal{B}^i)^{\otimes T}$. $(T_k)_{k \leq K}$ denotes the number of samples collected on

1

arm k until time T.

0.1.0.1 Reminder about the goal of the proof

At this point, we have defined a finite set of problems of Gaussian distributions with fixed variance 1 for a given set of set of gaps between the means of the distributions and the threshold, $(d_k)_k$. We will now proof a lower bound on the worst case regret of any algorithm; that is, the maximum error probability that even the best algorithm makes on this problem.

0.1.0.2 Statistical Decision

Given our formulation of the regret, the expected regret $\mathbb{E}[\mathcal{L}(T)]$ corresponds directly to the probability of making a wrong classification of any of the K means. In general, let T be a binary decision function, and define $\Omega = \{T = 0\}$ and $\Omega^C = \{T = 1\}$. Assume that the decision T distinguishes between two states of the world, H_0 and H_1. Then our overall error probability is given by $P_{H_0}(T = 1) + P_{H_1}(T = 0) = P_{H_0}(\Omega^C) + P_{H_1}(\Omega)$.

In our case, H_0 may correspond to the state in which we have $\mathcal{B}^0$, and H_1 corresponds to the case where any arm has been flipped, $\mathcal{B}^i$. Furthermore, we consider the events in which the algorithm, making the decision, classifies arm i as being above the threshold: $\mathcal{A}_i = \{i \in \hat{S}_\tau$. The event on which all arms are being classified as above the threshold is consequently given by $\mathcal{A} = \cap \mathcal{A}_i$. However, Theorem 1 does not bound the overall regret, but the maximum error probability. We can write this now as

$$\max_{i \in \{0,\dots,K\}} \mathbb{E}_{\mathcal{B}^i}(\mathcal{L}(T)) \geq \max \left(\max_{i \in \{1,\dots,K\}} \mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i), \mathbb{P}_{\mathcal{B}^0}(\mathcal{A}^C) \right) = \max \left(\max_{i \in \{1,\dots,K\}} \mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i), 1 - \mathbb{P}_{\mathcal{B}^0}(\mathcal{A}) \right)$$

$$(1)$$

In order to give a bound on above equation, we start by considering $\mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i)$. This probability can be measured as the expected change in the likelihood of observed data times the probability under the original state. The measure of event $\mathcal{A}_i$ under state $\mathcal{B}^i$ is given as the probability of $\mathcal{A}_i$ under the (original) state $\mathcal{B}^0$ multiplied by a factor. This factor is the change in the data likelihoods due to moving from $\mathcal{B}^0$ to $\mathcal{B}^i$, and is expressed as the expected likelihood ratio under $\mathcal{B}^0$. By definition though, this is equal to the empirical Kullback-Leibler divergence of the two measures.

First, let us define the change of measure in terms of the empirical log-likelihood for our case. Then, we will show how to replace the log-likelihood by the empirical Kullback-Leibler divergence.

Let every ν_i, ν_i' be dominated by the Lebesgue measure λ. Then their density functions are given by $f_i = \frac{d\nu_i}{d\lambda}$ and $f_i' = \frac{d\nu_i'}{d\lambda}$ respectively. For $T \geq t \geq 0$, define the empirical Kullback-Leibler divergence as

$$\hat{\text{KL}}_{k,t} = \frac{1}{t} \sum_{s=1}^{t} \log\left(\frac{\mathrm{d}\nu'_k}{\mathrm{d}\nu_k}(X_{k,s})\right)$$

$$= \frac{1}{t} \sum_{s=1}^{t} \log\left(\frac{f'_k(X_{k,s})}{f_k(X_{k,s})}\right)$$

$$= \frac{1}{t} \log\left(\prod_{s=1}^{t} \frac{f'_k(X_{k,s})}{f_k(X_{k,s})}\right)$$

$$\stackrel{\text{iid}}{=} \frac{1}{t} \log\left(\frac{f'_k(X_{k,1}, \ldots, X_{k,t})}{f_k(X_{k,1}, \ldots, X_{k,t})}\right)$$

$$= \frac{1}{t} \log\left(\text{LR}((X_{k,1}, \ldots, X_{k,t}), \nu'_k, \nu_k)\right)$$

Since we consider Gaussian distributions with variance 1, we have density functions $f_i(x) \propto \exp\left(-\frac{1}{2}(x - \Delta_i)^2\right)$ and $f'_i(x) \propto \exp\left(-\frac{1}{2}(x + \Delta_i)^2\right)$. Plugging this into the above definition of the empirical Kullback-Leibler divergence, we easily see

$$\hat{\text{KL}}_{k,t} = \frac{1}{t} \sum_{s=1}^{t} \log\left(\frac{\exp\left(-\frac{1}{2}(X_{k,s} + \Delta_k)^2\right)}{\exp\left(-\frac{1}{2}(X_{k,s} - \Delta_k)^2\right)}\right)$$

$$= \frac{1}{t} \sum_{s=1}^{t} \log\left(\exp\left(-\frac{1}{2}(X_{k,s} + \Delta_k)^2 + \frac{1}{2}(X_{k,s} - \Delta_k)^2\right)\right)$$

$$= -\frac{1}{t} \sum_{s=1}^{t} 2X_{k,s}\Delta_k$$

0.1.0.3   Change of Measure

We are now ready to state the change of measure. As alluded to above, we consider how the probability changes for the event $\mathcal{A}_i$ when we move from $\mathcal{B}^0$ to $\mathcal{B}^i$. When flipping arm i, we only changed the samples of this arm, that is, the first T_i observations we got from arm i the algorithm took. The change of measure is then defined as:

$$\mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i) = \mathbb{E}_{\mathcal{B}^0}\left[\mathbb{1}_{\mathcal{A}_i} \text{LR}((X_{i,1}, \ldots, X_{i,T_i}), \nu'_i, \nu_i)\right]$$

$$= \mathbb{E}_{\mathcal{B}^0}\left[\mathbb{1}_{\mathcal{A}_i} \exp(-T_i \hat{\text{KL}}_{i,T_i})\right]$$

Obviously, $\hat{\text{KL}}_{i,T_i}$ depends on the realized samples $X_{i,1}, \ldots, X_{i,T_i}$. Consequently, we need to find a bound on $\hat{\text{KL}}_{i,T_i}$ itself to incorporate the uncertainty involved in this metric.

0.1.0.4 Kullback-Leibler Divergence

For our problem, the Kullback-Leibler divergence of $/nu'_k$ from ν_k is given by $\text{KL}_k := KL(\nu'_k, \nu_k) = 2\Delta_k^2$.

0.1.0.5 Concentration of the Empirical Kullback-Leibler Divergence

It is easy to see that

$$|\hat{\text{KL}}_{k,t} - \text{KL}_{k,t}| = |-\frac{2}{t}\Delta_k \sum_{s=1}^{t}(X_{k,s} - \Delta_k)|$$

itself is a normally distributed random variable with mean zero and a variance decreasing in t. Thus, one can show that $\mathbb{P}_{\mathcal{B}^i}(\xi) \geq 3/4$ for the event

$$\xi = \{\forall k \leq K, \forall t \leq T, |\hat{\text{KL}}_{k,t} - \text{KL}_{k,t}| \leq 4\Delta_k\sqrt{\frac{\log(4(\log(T)+1)K)}{t}}\}. \tag{2}$$

Thus, in order to bound the empirical Kullback-Leibler divergence in our change of measure, we can consider the intersection of $\mathcal{A}_i$ and ξ as follows and plug in the bound on event ξ:

$$\begin{aligned}
\mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i) &= \mathbb{E}_{\mathcal{B}^0}\left[\mathbb{1}_{\mathcal{A}_i}\exp(-T_i\hat{\text{KL}}_{i,T_i})\right]\\
&\geq \mathbb{E}_{\mathcal{B}^0}\left[\mathbb{1}_{\mathcal{A}_i\cap\xi}\exp(-T_i\hat{\text{KL}}_{i,T_i})\right]\\
&\geq \mathbb{E}_{\mathcal{B}^0}\left[\mathbb{1}_{\mathcal{A}_i\cap\xi}\exp(-2\Delta_i^2 T_i - 4\Delta_i\sqrt{T_i}\sqrt{\log(4(\log(T)+1)K)})\right]
\end{aligned}$$

Two things are left to do: With T_i, we still have one random variable left which we cannot leave in our bound. Second, we are actually not interested in $\mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i)$ but in $\max_{i\in\{1,\dots,K\}}\mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i)$.

0.1.0.6 Bounding T_i for $\max_{i\in\{1,\dots,K\}}\mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i)$

Part of our bound is $\max_{i\in\{1,\dots,K\}}\mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i)$, so let's continue with that. In general it holds that $\max_{1\leq i\leq K}(X_1,\dots,X_K) \geq \bar{X}$. Thus we can write:

$$\begin{aligned}
\max_{i\in\{1,\dots,K\}}\mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i) &\geq \frac{1}{K}\sum_{i=1}^{K}K\mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i)\\
&\geq \frac{1}{K}\sum_{i=1}^{K}K\mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i\cap\xi)\\
&\geq \frac{1}{K}\sum_{i=1}^{K}K\mathbb{E}_{\mathcal{B}^0}\left[\mathbb{1}_{\mathcal{A}_i\cap\xi}\exp(-2\Delta_i^2 T_i - 4\Delta_i\sqrt{T_i}\sqrt{\log(4(\log(T)+1)K)})\right]
\end{aligned}$$

where the second line is again the intersection of $\mathcal{A}_i$ and ξ as before in order to apply the change of measure again in the third line.

What is left to show before we have our final bound is a bound on T_i. To show that any algorithm has a certain bound on its error, it only makes sense that every arm has to pulled at least a certain amount of times. Consequently, no arm should be pulled exclusively.

First, let $a = \Delta_i\sqrt{T_i}$ and $b = 4\sqrt{\log((4\log(T)+1)K)}$. Then use $ab \le a^2 + b^2$ to write $ab \le \Delta_i^2 T_i + 16\log((4\log(T)+1)K)$. And so:

$$\max_{i\in\{1,\dots,K\}} \mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i) \ge \frac{1}{K}\sum_{i=1}^{K} K\mathbb{E}_{\mathcal{B}^0}\left[\mathbb{1}_{\mathcal{A}_i\cap\xi}\exp(-3\Delta_i^2 T_i - 16\log(4(\log(T)+1)K))\right]$$

Furthermore, we know that the sum of the individual arm pulls has to be equal to the overall number of pulls (equal to the budget in the fixed budget sense), $\sum_i T_i = T$, and furthermore we know that $T_i > 0$. The latter holds since any reasonable algorithm has to check each arm at least once in our setup. This is because we need to check each arm's performance in comparison to the threshold. Thus it is easy to show that $\exists i : T_i \le \frac{T}{H\Delta_i^2} = \frac{T}{(\sum_{i=1}^{K}\Delta_i^{-2})\Delta_i^2}$. To see this, assume the opposite: $\forall i : T_i > T\frac{\sum_{i=1}^{K}\Delta_i^2}{\Delta_i^2}$. This implies however $\sum_{i=1}^{K} T_i > \sum_{i=1}^{K} T\frac{\sum_{i=1}^{K}\Delta_i^2}{\Delta_i^2} = T\cdot 1 = T$ which is a contradiction.

So we can write $(\Delta_i\sqrt{T_i})^2 \le \frac{T}{H}$. We use $\mathcal{A} = \cap_{i=1}^{K}\mathcal{A}_i$ to write:

$$\max_{i\in\{1,\dots,K\}} \mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i) \ge \frac{1}{K}\sum_{i=1}^{K}\mathbb{E}_{\mathcal{B}^0}\left[\mathbb{1}_{\mathcal{A}_i\cap\xi}\exp(-3\Delta_i^2 T_i - 16\log(4(\log(T)+1)K))\right]$$
$$= \mathbb{E}_{\mathcal{B}^0}\left[\mathbb{1}_{\mathcal{A}\cap\xi}\frac{1}{K}\sum_{i=1}^{K}K\exp(-3\Delta_i^2 T_i - 16\log(4(\log(T)+1)K))\right]$$

We write $\frac{1}{K}\sum_{i=1}^{K}\exp(-3T_i\Delta_i^2) \ge \frac{1}{K}\sum_{i=1}^{K}\exp(-3T/H) = \exp(-3T/H)$. Thus:

$$\max_{i\in\{1,\dots,K\}} \mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i) = \mathbb{E}_{\mathcal{B}^0}\left[\mathbb{1}_{\mathcal{A}\cap\xi}\frac{1}{K}\sum_{i=1}^{K}\exp(-3\Delta_i^2 T_i)\exp(-16\log(4(\log(T)+1)K))\right]$$
$$\ge \mathbb{E}_{\mathcal{B}^0}\left[\mathbb{1}_{\mathcal{A}\cap\xi}\frac{1}{K}\sum_{i=1}^{K}\exp(-3\frac{T}{H})\exp(-16\log(4(\log(T)+1)K))\right]$$
$$= \mathbb{E}_{\mathcal{B}^0}\left[\mathbb{1}_{\mathcal{A}\cap\xi}\exp(-3\frac{T}{H})\exp(-16\log(4(\log(T)+1)K))\right]$$
$$= \mathbb{P}_{\mathcal{B}^0}(\mathcal{A}\cap\xi)\exp(-3\frac{T}{H} - 16\log(4(\log(T)+1)K))$$

0.1.0.7 Conclusion

Consider again the risk which we need to bound from equation (1). We bound the maximum expected loss over all $\mathcal{B}_i$. By again using the fact that in general $\max_{1 \leq i \leq K}(X_1, \ldots, X_K) \geq \bar{X}$, we plug in our previous calculations:

$$\max_{i \in \{0,\ldots,K\}} \mathbb{E}_{\mathcal{B}^i}(\mathcal{L}(T)) \geq \max \left(\max_{i \in \{1,\ldots,K\}} \mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i), 1 - \mathbb{P}_{\mathcal{B}^0}(\mathcal{A}) \right) \tag{3}$$

$$\geq \frac{1}{2}\mathbb{P}_{\mathcal{B}^0}(\mathcal{A} \cap \xi) \exp(-3\frac{T}{H} - 16\log(4(\log(T)+1)K)) + \frac{1}{2}(1 - \mathbb{P}_{\mathcal{B}^0}(\mathcal{A})) \tag{4}$$

From (2) we know that for any $i \in \{0,\ldots,K\}$, $\mathbb{P}_{\mathcal{B}^i}(\xi) \geq 3/4$. Following the argument in Locatelli et al. (2016), we now consider two cases: $\mathbb{P}_{\mathcal{B}^0}(\mathcal{A}) \geq 1/2$ and $\mathbb{P}_{\mathcal{B}^0}(\mathcal{A}) \leq 1/2$. In the former case, we can plug in the two probabilities into (4). If $\mathbb{P}_{\mathcal{B}^0}(\mathcal{A}) \geq 1/2$ and $\mathbb{P}_{\mathcal{B}^0}(\xi) \geq 3/4$, then $\mathbb{P}_{\mathcal{B}^0}(\mathcal{A} \cap \xi) \geq 1/4$. Thus:

$$\max_{i \in \{0,\ldots,K\}} \mathbb{E}_{\mathcal{B}^i}(\mathcal{L}(T)) \geq \max \left(\max_{i \in \{1,\ldots,K\}} \mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i), 1 - \mathbb{P}_{\mathcal{B}^0}(\mathcal{A}) \right)$$

$$\geq \frac{1}{2}\mathbb{P}_{\mathcal{B}^0}(\mathcal{A} \cap \xi) \exp(-3\frac{T}{H} - 16\log(4(\log(T)+1)K)) + \frac{1}{2}(1 - \mathbb{P}_{\mathcal{B}^0}(\mathcal{A}))$$

$$\geq \frac{1}{8} \exp(-3\frac{T}{H} - 16\log(4(\log(T)+1)K))$$

where we drop the second summand of the second line so that the inequality holds.

On the other hand, when $\mathbb{P}_{\mathcal{B}^0}(\mathcal{A}) \leq 1/2$, then we know that $\mathbb{P}_{\mathcal{B}^0}(\mathcal{A} \cap \xi) \leq 1/2$ for any $\mathbb{P}_{\mathcal{B}^0}(\xi)$. At the same time $1 - \mathbb{P}_{\mathcal{B}^0}(\mathcal{A}) = \mathbb{P}_{\mathcal{B}^0}(\mathcal{A}^{\mathcal{C}}) \geq 1/2$. Consequently, $1/4 \leq \frac{1}{2}(1 - \mathbb{P}_{\mathcal{B}^0}(\mathcal{A})) \leq 1/2$, while $\frac{1}{2}\mathbb{P}_{\mathcal{B}^0}(\mathcal{A} \cap \xi) \exp(-3\frac{T}{H} - 16\log(4(\log(T)+1)K)) \geq 0$. And so the stronger lower bound in that case is, as noticed by Locatelli et al. (2016), given simply by

$$\max_{i \in \{0,\ldots,K\}} \mathbb{E}_{\mathcal{B}^i}(\mathcal{L}(T)) \geq \max \left(\max_{i \in \{1,\ldots,K\}} \mathbb{P}_{\mathcal{B}^i}(\mathcal{A}_i), 1 - \mathbb{P}_{\mathcal{B}^0}(\mathcal{A}) \right)$$

$$\geq 1 - \mathbb{P}_{\mathcal{B}^0}(\mathcal{A}) = \mathbb{P}_{\mathcal{B}^0}(\mathcal{A}^{\mathcal{C}})$$

$$\geq \frac{1}{2}$$