Information and Coding Theory **Autumn 2014**

Lecture 6: October 16, 2014

Lecturer: Madhur Tulsiani Scribe: Takeshi Onishi

In this lecture, we will use KL-divergence to prove a lower bound for the multi-armed bandit problem. Recall that in the multi-armed bandit problem pronlem, we have N possible actions. At time t, each action i has a loss $l_t(i) \in \{0, 1\}$. However, we only get to see the loss of the action of the action we decide to take at time t (like the traffic on the road that we choose to take). If $a_1, \ldots, a_T \in [N]$ describe our actions for times $1, \ldots, T$, out goal is to minimize the loss of our actions as compared to the best *single* action. We define the regret as

$$\mathcal{R}_T \;=\; \sum_{t=1}^{T} l_t(a_t) - \min_{i \in [N]} \sum_{t=1}^{T} l_t(i)\,.$$

We will use the lower bound on distinguishing coin tosses to give a distribution over losses such that the regret for any action sequence is high in expectation. In particular, we will consider losses when we hide a biased coin in the i^{th} position such that $l_t(i) = 1$ with probability $1/2 - \varepsilon$ and 0 with probabiliy $1/2 + \varepsilon$. All other losses will be 0/1 with probability 1/2 each. We will also chooses $i \in [N]$ at random. We will use this to show that

$$\mathcal{R}_T \;=\; \Omega\left(\sqrt{NT}\right)\,.$$

This bound is almost tight since there exists a randomized algorithm which achieves an expected regret $O\left(\sqrt{N \log N \cdot T}\right)$.

First we develop a generalization of the lower bound in the previous class for disinguishing coins, to the setting where we have N instead of 2 coins. We follow the exposition by Kleinberg [K07] for this lecture.

1 Tossing N coins

We define a generalization of the problem of distinguishing two coins discussed in the previous lecture. Consider a scenario where we have N coins, $N - 1$ of which are fair coins which are heads and tails with probability 1/2 each, and one coin is biased and comes up heads with probability $1/2 - \varepsilon$. Moreover, the biased coin is hidden in one of the N positions at random. We will prove a lower bound on the number of coin tosses any algorithm needs to observe to guess the position of the biased with significant probability. Note that for $N = 2$, this is equivalent to the problem discussed in the previous lecture.

We consider a guessing algorithm, which at time t outputs a pair $(a_t, b_t) \in [N]^2$. Here, $b_t \in N$ is the algorithm's current guess for the position of the biased coin. Also, the algorithm gets to see the output of the toss for the coin in position a_t at time t. We can think of this as a model where we toss all N coins at time t, but algorithm can ask to see the output of only one coin (which it asks for by specifying a_t).

Let $P_1, \ldots, P_N$ denote the distributions for the view of the algorithm from time 1 to T, when the biased coin is hidden in the i^{th} position. Note that what the algorithm sees from time 1 to T is a seuqnce of T 0/1 values (say 0 for "tails" and 1 for "heads"). However, the distribution can be fairly complicated. In particular, which coin the algorithm asks to see at time t can depend on what it saw at times $1, \ldots, t-1$, which can in turn depend on the position of the biased coin. Nevertheless, the distribution is determined completely by the description of the algorithm and the position of the biased coin.

We will prove the following lemma which shows that for every algorithm A, there are at least $N/3$ places to hide the biased coin, such that algorithm fails to find it with probability at least $1/2$, after $T \leq N/(100\varepsilon^2)$ steps.

Lemma 1.1 *Let A be any guessing algorithm operating as specified above and let $T \leq \frac{N}{60 \cdot \varepsilon^2}$ for $\varepsilon \leq 1/4$ and $N \geq 14$. Then, there exists $J \subseteq [N]$ with $|J| \geq N/3$ such that*

$$\forall j \in J, \;\; \mathop{\mathbb{P}}_{D_j} [b_{T+1} = j] \;\; \leq \;\; \frac{1}{2}$$

Note that it is necessary, J depends on the algorithm. In particular, if the algorithm guesses the position of the biased coin to be 1 irrespective of what it sees, then the set J certainly cannot contain 1. In the proof below, we shall find J simply by eliminating all positions which the algorithm guesses with atypically high probability or which the algorithm queries too many times.

Proof: Define N_i to be the number of times the algorithm asks to see the output of the i^{th} coin

$$N_i \;\; := \;\; |\{t \in [T] \;\; | \;\; a_t = i\}| \; .$$

Also, let D_0 be the hypothetical distribution for the view of the algorithm when all the N coins are fair. This is simply the distribution for a sequence of T independent and uniform 0/1 values. We shall define the set J by considering the behavior of the algorithm if tosses it saw were according to the distribution D_0. We define

$$J_1 := \left\{ i \mid \mathop{\mathbb{E}}_{D_0} [N_i] \leq \frac{3T}{N} \right\}, \quad J_2 := \left\{ i \mid \mathop{\mathbb{P}}_{D_0} [b_{T+1} = i] \leq \frac{3}{N} \right\} \quad \text{and} \quad J = J_1 \cap J_2 \, .$$

Since $\sum_i \mathbb{E}_{D_0} [N_i] = T$ and $\sum_i \mathbb{P}_{D_0} [b_{T+1} = i] = 1$, an averaging argument gives $|J_1| \geq 2N/3$ and $|J_2| \geq 2N/3$, and hence $|J| \geq N/3$.

Consider any $j \in J$ and a function f on the view of the algorithm for times $1, \ldots, T$, which is 1 if $b_{T+1} = j$ and 0 otherwise. Then, we have that

$$\left| \mathop{\mathbb{P}}_{D_j} [b_{T+1} = j] - \mathop{\mathbb{P}}_{D_0} [b_{T+1} = j] \right| \;\; = \;\; \left| \mathop{\mathbb{E}}_{D_j} [f] - \mathop{\mathbb{E}}_{D_0} [f] \right| \;\; \leq \;\; \frac{1}{2} \cdot \| D_0 - D_j \|_1 \; .$$

Combining this with Pinsker's inequality, we have

$$\mathop{\mathbb{P}}_{D_j} [b_{T+1} = j] \;\; \leq \;\; \frac{3}{N} + \frac{1}{2} \cdot \sqrt{2 \ln 2 \cdot KL \left(D_0 || D_j \right)} \, ,$$

where we use the notation $KL(D_0||D_j)$ instead of $D(D_0||D_j)$ to denote KL-divergence to avoid confusion.

Let $x_1, \ldots, x_T$ be the outputs of the coin tosses seen by the algorithm A. Using the chain rule, we can write the KL-divergence as

$$KL(D_0||D_j) \;=\; \sum_{t=1}^{T} \sum_{x_1,\ldots,x_{t-1}} \mathop{\mathbb{P}}_{D_0}[x_1, ..., x_t] \cdot KL\left(D_0(x_t)||D_j(x_t) \;\mid\; x_1, ..., x_{t-1}\right)$$

where $D_0(x_t)$ and $D_j(x_t)$ denote the distribution of the t^{th} coin toss seen by the algorithm (given the outputs of the first $t-1$ tosses). Note that this is a single coin toss which is according to a biased coin in the distribution $D_|j$ if $a_t = j$ and is according to a fair coin if $a_t \neq j$. On the other hand, this toss is always according to a fair coin in the distribution D_0. Let $B = KL(P||Q)$ where P is the distribution which 0/1 with probability 1/2 each and Q is 1 with probability $1/2 - \varepsilon$ and 0 with probability $1/2 + \varepsilon$. Then we can write

$$KL(D_0||D_j) \;=\; \sum_{t=1}^{T} \sum_{x_1,\ldots,x_{t-1}} \mathop{\mathbb{P}}_{D_0}[x_1, ..., x_t] \cdot \mathbb{1}_{\{a_t=j\}} \cdot B \;=\; \mathop{\mathbb{E}}_{D_0}[N_j] \cdot B \;\leq\; \frac{3T}{N} \cdot B\,.$$

Using a calculation similar to the one in the previous lecture, we get that $B \leq \frac{5\varepsilon^2}{2\ln 2}$ for $\varepsilon \leq 1/4$. Combining this with the above bounds gives

$$\mathop{\mathbb{P}}_{D_j}[b_{T+1} = j] \;\leq\; \frac{3}{N} + \frac{1}{2} \cdot \sqrt{2\ln 2 \cdot \frac{3T}{N} \cdot \frac{5\varepsilon^2}{2\ln 2}} \;\leq\; \frac{3}{N} + \frac{1}{2} \cdot \sqrt{\frac{15T}{\varepsilon^2 N}} \;\leq\; \frac{3}{N} + \frac{1}{4}\,,$$

since $T \leq \frac{N}{60\varepsilon^2}$. For $N \geq 12$, the above bound is at most 1/2 which proves the lemma. ∎

2 The regret bound

Using the generalization of coin tossing in the previous section, it is now easy to derive the lower bound for multi-armed bandits. For any fixed algorithm A, we will give a *distribution* over losses such that the expected regret of the algorithm is $\Omega\left(\sqrt{NT}\right)$.

We define the distribution by chosing a random $i^* \in [N]$ and defining the losses $l_t(i)$ as

$$l_t(i) = \begin{cases} 1 & \text{w.p. } 1/2 \\ 0 & \text{w.p. } 1/2 \end{cases} \quad \text{if } i \neq i^* \qquad \text{and} \qquad l_t(i) = \begin{cases} 1 & \text{w.p. } 1/2 - \varepsilon \\ 0 & \text{w.p. } 1/2 + \varepsilon \end{cases} \quad \text{if } i = i^*$$

We choose $\varepsilon = \sqrt{60T/N}$ and think of an algorithm which chooses action $a_t \in [N]$ at time t as a coin-guessing algorithm which outputs the pair (a_t, a_t) at time t. Since $T \leq \frac{N}{60\varepsilon^2}$, $\forall t \in \{0, \ldots, T-1\}$, there exists a set $J_t \subseteq [N]$ with $|J_t| \geq N/3$ such that

$$\forall j \in J_t, \;\; \mathop{\mathbb{P}}_{D_j}[a_{t+1} = j] \;\leq\; \frac{1}{2}\,.$$

Hence, we have that for all $t \in \{0, \ldots, T-1\}$

$$\mathbb{E}[l_t(a_t)] \;\geq\; \frac{1}{3} \cdot \left(\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \left(\frac{1}{2} - \varepsilon\right)\right) + \frac{2}{3} \cdot \left(\frac{1}{2} - \varepsilon\right) \;\geq\; \frac{1}{2} - \frac{5\varepsilon}{6}\,.$$

Here, the expectation is also over the choice of i^*. On the other hand,

$$\mathbb{E}\left[\min_{i\in[N]}\sum_{t=1}^{T} l_t(i)\right] \;\leq\; \mathbb{E}\left[\sum_{t=1}^{T} l_t(i^*)\right] \;\leq\; \left(\frac{1}{2}-\varepsilon\right)\cdot T\,.$$

Thus, we have that

$$\mathbb{E}\left[\mathcal{R}_T\right] \;\geq\; \left(\frac{1}{2}-\frac{5\varepsilon}{6}\right)\cdot T - \left(\frac{1}{2}-\varepsilon\right)\cdot T \;\geq\; \frac{\varepsilon T}{6} \;\geq\; \frac{1}{6}\cdot\sqrt{\frac{NT}{60}}\,.$$

References

[K07] R. Kleinberg, "Multi-Armed Bandit Problems", *Lecture notes on "Learning, Games, and Electronic Markets"*.