CMSC 858G: Bandits, Experts and Games 09/19/16

# Lecture 3: Lower Bounds for Bandit Algorithms

Instructor: Alex Slivkins Scribed by: Soham De & Karthik A Sankararaman

## 1 Lower Bounds

In this lecture (and the first half of the next one), we prove a $\Omega(\sqrt{KT})$ lower bound for regret of bandit algorithms. This gives us a sense of what are the best possible upper bounds on regret that we can hope to prove.

On a high level, there are two ways of proving a lower bound on regret:

(1) Give a family $\mathcal{F}$ of problem instances, which is the same for all algorithms, such that any algorithm 'fails' (has high regret) on some instance in $\mathcal{F}$.

(2) Give a distribution over problem instances, and show that, in expectation over this distribution, any algorithm with fail.

Note that (2) implies (1) since: if regret is high in expectation over problem instances, then there exists at least one problem instance with high regret. Also, (1) implies (2) if $|\mathcal{F}|$ is a constant. This can be seen as follows: suppose we know that for any algorithm we have high regret (say $H$) with one problem instance in $\mathcal{F}$ and low regret with all other instances in $\mathcal{F}$, then, taking a uniform distribution over $\mathcal{F}$, we can say that any algorithm has expected regret at least $H/|\mathcal{F}|$. (So this argument breaks if $|\mathcal{F}|$ is large.) If we prove a stronger version of (1) that says that for any algorithm, regret is high for a *constant fraction* of the problem instances in $\mathcal{F}$, then, considering a uniform distribution over $\mathcal{F}$, this implies (2) regardless of whether $|\mathcal{F}|$ is large or not.

In this lecture, for proving lower bounds, we consider 0-1 rewards and the following family of problem instances (with fixed $\epsilon$ to be adjusted in the analysis):

$$\mathcal{I}_j = \begin{cases} \mu_i = 1/2 & \text{for each arm } i \neq j, \\ \mu_i = (1+\epsilon)/2 & \text{for arm } i = j \end{cases} \qquad \text{for each } j = 1, 2, \ \ldots \ , K. \tag{1}$$

(Recall that $K$ is the number of arms.)

In the previous lecture, we saw that sampling each arm $\tilde{O}(1/\epsilon^2)$ times is sufficient for the upper bounds on regret that we derived. In this lecture, we prove that sampling each arm $\Omega(1/\epsilon^2)$ times is necessary to determine whether an arm is bad or not.

The proof methods will require *KL divergence*, an important tool from Information Theory. In the next section, we briefly study the KL divergence and some of its properties.

## 2 KL-divergence

Consider a finite sample space $\Omega$, and let $p, q$ be two probability distributions defined on $\Omega$. Then, the Kullback-Leibler divergence or *KL-divergence* is defined as:

$$KL(p,q) = \sum_{x \in \Omega} p(x) \ln \frac{p(x)}{q(x)} = \mathbb{E}_p \left[ \ln \frac{p(x)}{q(x)} \right].$$

The KL divergence is similar to a notion of distance with the properties that it is non-negative, 0 iff $p = q$, and small if the distributions $p$ and $q$ are close. However, it is not strictly a distance function since it is not symmetric and does not satisfy the triangle-inequality.

The intuition for the formula is as follows: we are interested in how certain we are that data, with underlying distribution $q$, can be generated from distribution $p$. The KL divergence effectively answers this question by measuring the average log likelihood of observing data with distribution $p$ when the underlying distribution of the data is actually given by $q$.

*Remark* 2.1. The definition of KL-divergence, as well as the properties discussed below, extend to infinite sample spaces. However, KL-divergence for finite sample spaces suffices for this class, and is much easier to work with.

### Properties of KL-divergence

We present several basic properties of KL-divergence that will be needed later. The proofs of these properties are fairly simple, we include them here for the sake of completeness.

1. *Gibbs' Inequality*: $KL(p, q) \geq 0, \forall p, q$. Further, $KL(p, q) = 0$ iff $p = q$.

   *Proof.* Let us define: $f(y) = y \ln(y)$. $f$ is a convex function under the domain $y > 0$. Now, from the definition of the KL divergence we get:

   $$\begin{aligned} KL(p, q) &= \sum_{x \in \Omega} p(x) \ln \frac{p(x)}{q(x)} \\ &= \sum_{x \in \Omega} q(x) f\left(\frac{p(x)}{q(x)}\right) \\ &\geq f\left(\sum_{x \in \Omega} q(x) \frac{p(x)}{q(x)}\right) && \text{[follows from Jensen's inequality]} \\ &= f\left(\sum_{x \in \Omega} p(x)\right) = f(1) = 0, \end{aligned}$$

   where Jensen's inequality states that $\varphi(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 \varphi(x_1) + \lambda_2 \varphi(x_2)$, if $\varphi$ is a convex function and $\lambda_1 + \lambda_2 = 1$ with $\lambda_1, \lambda_2 > 0$. Jensen's inequality further has the property that $\varphi(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 \varphi(x_1) + \lambda_2 \varphi(x_2)$ iff $x_1 = x_2$ or if $\varphi$ is a linear function. In this case, since $f$ is not a linear function, equality holds (i.e., $KL(p, q) = 0$) iff $p(x) = q(x), \forall x$. □

2. Let the sample space $\Omega$ be composed as $\Omega = \Omega_1 \times \Omega_1 \times \cdots \times \Omega_n$. Further, let $p$ and $q$ be two distributions defined on $\Omega$ as $p = p_1 \times p_2 \times \cdots \times p_n$ and $q = q_1 \times q_2 \times \cdots \times q_n$, such that $\forall j = 1, \ldots, n$, $p_j$ and $q_j$ are distributions defined on $\Omega_j$. Then we have the property: $KL(p, q) = \sum_{j=1}^{n} KL(p_j, q_j)$.

   *Proof.* Let $x = (x_1, x_2, \ldots, x_n) \in \Omega$ st $x_i \in \Omega_i, \forall i = 1, \ldots, n$. Let $h_i(x_i) = \ln \frac{p_i(x_i)}{q_i(x_i)}$. Then:

   $$KL(p, q) = \sum_{x \in \Omega} p(x) \ln \frac{p(x)}{q(x)}$$

$$= \sum_{i=1}^{n} \sum_{x \in \Omega} p(x) h_i(x_i) \qquad \left[\text{since } \ln \frac{p(x)}{q(x)} = \sum_{i=1}^{n} h_i(x_i)\right]$$

$$= \sum_{i=1}^{n} \sum_{x_i^\star \in \Omega_i} h_i(x_i^\star) \sum_{\substack{x \in \Omega, \\ x_i = x_i^\star}} p(x)$$

$$= \sum_{i=1}^{n} \sum_{x_i \in \Omega_i} p_i(x_i) h_i(x_i) \qquad \left[\text{since } \sum_{x \in \Omega,\ x_i = x_i^\star} p(x) = p_i(x_i^\star)\right]$$

$$= \sum_{i=1}^{n} KL(p_i, q_i). \qquad \square$$

3. *Weaker form of Pinsker's inequality*: $\forall A \subset \Omega : \ 2\left(p(A) - q(A)\right)^2 \leq KL(p, q)$.

*Proof.* To prove this property, we first claim the following:

**Claim 2.2.** *For each event $A \subset \Omega$,*

$$\sum_{x \in A} p(x) \ln \frac{p(x)}{q(x)} \geq p(A) \ln \frac{p(A)}{q(A)}.$$

*Proof.* Let us define the following:

$$p_A(x) = \frac{p(x)}{p(A)} \quad \text{and} \quad q_A(x) = \frac{q(x)}{q(A)} \quad \forall x \in A.$$

Then the claim can be proved as follows:

$$\sum_{x \in A} p(x) \ln \frac{p(x)}{q(x)} = p(A) \sum_{x \in A} p_A(x) \ln \frac{p(A) p_A(x)}{q(A) q_A(x)}$$

$$= p(A) \left( \sum_{x \in A} p_A(x) \ln \frac{p_A(x)}{q_A(x)} \right) + p(A) \ln \frac{p(A)}{q(A)} \sum_{x \in A} p_A(x)$$

$$\geq p(A) \ln \frac{p(A)}{q(A)}. \qquad \left[\text{since } \sum_{x \in A} p_A(x) \ln \frac{p_A(x)}{q_A(x)} = KL(p_A, q_A) \geq 0\right]$$

$\square$

Fix $A \subset \Omega$. Using Claim 2.2 we have the following:

$$\sum_{x \in A} p(x) \ln \frac{p(x)}{q(x)} \geq p(A) \ln \frac{p(A)}{q(A)},$$

$$\sum_{x \notin A} p(x) \ln \frac{p(x)}{q(x)} \geq p(\bar{A}) \ln \frac{p(\bar{A})}{q(\bar{A})},$$

where $\bar{A}$ denotes the complement of $A$. Now, let $a = p(A)$ and $b = q(A)$. Further, assume $a < b$. Then, we have:

$$\begin{aligned} KL(p,q) &= a \ln \frac{a}{b} + (1-a) \ln \frac{1-a}{1-b} \\ &= \int_a^b \left( -\frac{a}{x} + \frac{1-a}{1-x} \right) dx \\ &= \int_a^b \frac{x-a}{x(1-x)} dx \\ &\geq \int_a^b 4(x-a)dx = 2(b-a)^2. \qquad \text{[since } x(1-x) \leq 1/4\text{]} \end{aligned}$$

This proves the property. □

4. Let $p_\epsilon$ denote a distribution on $\{0,1\}$ such that $p_\epsilon(1) = (1+\epsilon)/2$. Thus, $p_\epsilon(0) = (1-\epsilon)/2$. Further, let $p_0$ denote the distribution on $\{0,1\}$ where $p_0(0) = p_0(1) = 1/2$. Then we have the property: $KL(p_\epsilon, p_0) \leq 2\epsilon^2$.

   *Proof.*

$$\begin{aligned} KL(p_\epsilon, p_0) &= \frac{1+\epsilon}{2} \ln(1+\epsilon) + \frac{1-\epsilon}{2} \ln(1-\epsilon) \\ &= \frac{1}{2} \left( \ln(1+\epsilon) + \ln(1-\epsilon) \right) + \frac{\epsilon}{2} \left( \ln(1+\epsilon) - \ln(1-\epsilon) \right) \\ &= \frac{1}{2} \ln(1-\epsilon^2) + \frac{\epsilon}{2} \ln \frac{1+\epsilon}{1-\epsilon}. \end{aligned}$$

   Now, $\ln(1-\epsilon^2) < 0$ and we can write $\ln \frac{1+\epsilon}{1-\epsilon} = \ln \left( 1 + \frac{2\epsilon}{1-\epsilon} \right) \leq \frac{2\epsilon}{1-\epsilon}$. Thus, we get:

$$KL(p_\epsilon, p_0) < \frac{\epsilon}{2} \cdot \frac{2\epsilon}{1-\epsilon} = \frac{\epsilon^2}{1-\epsilon} \leq 2\epsilon^2.$$

□

**How are these properties going to be used?**

We start with the same setting as in Property 2. From Property 3, we have:

$$\begin{aligned} 2(p(A) - q(A))^2 &\leq KL(p,q) \\ &= \sum_{j=1}^{n} KL(p_j, q_j). \qquad \text{(follows from Property 2)} \end{aligned}$$

For example, we can define $p_j$ and $q_j$ to be distributions of a biased coin with small $\epsilon$ ($p_j(1) = (1+\epsilon)/2,\ p_j(1) = (1-\epsilon)/2$) vs an unbiased coin ($q_j(0) = q_j(1) = 1/2$). Then, we can use Property 4 to bound the above as:

$$2(p(A) - q(A))^2 \leq \sum_{j=1}^{n} KL(p_j, q_j) \leq \sum_{j=1}^{n} \delta = n\delta,$$

where $\delta = 2\epsilon^2$ Thus, we arrive at the following bound:

$$|p(A) - q(A)| \leq \sqrt{n\delta/2}.$$

## 3 A simple example: flipping one coin

We start with a simple example, which illustrates our proof technique *and* is interesting as a standalone result. We have a single coin, whose outcome is a 0 or 1. The coin's mean is unknown. We assume that the true mean $\mu \in [0,1]$ is either $\mu_1$ or $\mu_2$ for two known values $\mu_1 > \mu_2$. The coin is flipped $T$ times. The goal is to identify if $\mu = \mu_1$ or $\mu = \mu_2$.

Define $\Omega := \{0,1\}^T$ to be the sample space of the outcomes of the $T$ coin tosses. We need a decision rule $Rule : \Omega \to \{High, Low\}$ with the following two properties:

$$\Pr[Rule(observations) = High | \mu = \mu_1] \geq 0.99 \tag{2}$$

$$\Pr[Rule(observations) = Low | \mu = \mu_2] \geq 0.99 \tag{3}$$

The question is how large should $T$ be for for such a *Rule* to exist? We know that if $\delta = |\mu_1 - \mu_2|$, then $T \geq \Omega(\frac{1}{\delta^2})$ is sufficient. We will prove that it is also necessary. We will focus on the special case when both $\mu_1$ and $\mu_2$ are close to $\frac{1}{2}$.

**Claim 3.1.** *Let $\mu_1 = \frac{1+\epsilon}{2}$ and $\mu_2 = \frac{1}{2}$. For any rule to "work" (i.e., satisfy equations (2) and (3)) we need*

$$T \geq \Omega(\frac{1}{\epsilon^2}) \tag{4}$$

*Proof.* Define for any event $A \subseteq \Omega$, the following quantities

$$P_1(A) = \Pr[A | \mu = \mu_1]$$

$$P_2(A) = \Pr[A | \mu = \mu_2]$$

To prove this claim, we will consider the following equation. For an event $A \subseteq \Omega$, such that $Rule(A) = High$ (i.e. A $= \{\omega \subseteq \Omega : Rule(\omega) = High\}$),

$$P_1(A) - P_2(A) \geq 0.98 \tag{5}$$

We prove the claim by showing that if (4) is false, then (5) is false, too. Specifically, we will assume that $T < \frac{1}{4\epsilon^2}$.(In fact, the argument below holds for an arbitrary event $A \subset \Omega$.)

Define, for all $i \in \{1,2\}$, $P_{i,t}$ to be the distribution of the $t^{th}$ coin toss for $P_i$. Then, $P_i = P_{i,1} \times P_{i,2} \times \ldots \times P_{i,T}$. From KL divergence property, we have

$$\begin{aligned}
2(P_1(A) - P_2(A))^2 &\leq KL(P_1, P_2) && \text{From KL divergence property 3} \\
&= \sum_{t=1}^{T} KL(P_{1,t}, P_{2,t}) && \text{From KL divergence property 2} \\
&\leq 2T\epsilon^2 && \text{From KL divergence property 4}
\end{aligned}$$

Hence, we have $|P_1(A) - P_2(A)| \leq \sqrt{T} \cdot \epsilon < \frac{1}{2}$. □

# 4 Flipping several coins: "bandits with prediction"

Let us extend the previous example to flipping multiple coins. More formally, we consider a bandit problem with $K$ arms (where each arm corresponds to a coin). Each arm gives a 0-1 reward, drawn independently from a fixed but unknown distribution. After $T$ rounds, the algorithm outputs a guess $y_T \in \mathcal{A}$ for which arm is the best arm (where $\mathcal{A}$ is the set of all arms).[1] We call this version "bandits with predictions". In this section, we will only be concerned with a quality of prediction, rather than accumulated rewards and regret.

For each arm $a \in \mathcal{A}$, the mean reward is denoted as $\mu(a)$. (We will also write it as $\mu_a$ whenever convenient.) A particular problem instance is specified as a tuple $\mathcal{I} = (\mu(a) : \forall a \in \mathcal{A})$.

A good algorithm for the "bandits with prediction" problem described above should satisfy

$$\Pr[y_T \text{ is correct } |\mathcal{I}] \geq 0.99 \tag{6}$$

for each problem instance $\mathcal{I}$. We will use the family (1) of problem instances to argue that one needs $T \geq \Omega\left(\frac{K}{\epsilon^2}\right)$ for any algorithm to "work", *i.e.,* satisfy (6), on all instances in this family.

**Lemma 4.1.** *Suppose an algorithm for "bandits with predictions" satisfies* (6) *for all problem instances* $\mathcal{I}_1, \ldots, \mathcal{I}_K$. *Then* $T \geq \Omega\left(\frac{K}{\epsilon^2}\right)$.

This result is of independent interest (regardless of the lower bound on regret). In fact, we will prove a stronger lemma which will (also) be the crux in the proof of the regret bound.

**Lemma 4.2.** *Suppose* $T \leq \frac{cK}{\epsilon^2}$, *for a small enough absolute constant* $c$. *Fix any deterministic algorithm for "bandits with prediction". Then there exists at least* $\lceil K/3 \rceil$ *arms* $j$ *such that*

$$\Pr[y_T = j \mid \mathcal{I}_j] < \tfrac{3}{4}$$

*Remark* 4.3. The proof for $K = 2$ arms is particularly simple, so we will do it first. We will then extend this proof to arbitrary $K$ with more subtleties. While the lemma holds for an arbitrary $K$, we will present a simplified proof which requires $K \geq 24$.

We will use a standard shorthand $[T] := \{1, 2, \ldots, T\}$.

Let us set up the sample space to be used in the proof. Let $(r_t(a) : a \in \mathcal{A}, t \in [T])$ be mutually independent 0-1 random variables such that $r_t(a)$ has expectation $\mu(a)$. We refer to this tuple as the *rewards table*, where we interpret $r_t(a)$ as the reward received by the algorithm for the $t$-th time it chooses arm $a$. The sample space is $\Omega = \{0, 1\}^{K \times T}$, where each outcome $\omega \in \Omega$ corresponds to a particular realization of the rewards table. Each problem instance $\mathcal{I}_j$ defines distribution $P_j$ on $\Omega$:

$$P_j(A) = \Pr[A|\mathcal{I}_j] \quad \text{for each } A \subset \Omega.$$

Also, let $P_j^{a,t}$ be the distribution of $r_t(a)$ under instance $\mathcal{I}_j$, so that $P_j = \prod_{a \in \mathcal{A},\, t \in [T]} P_j^{a,t}$.

*Proof ($K = 2$ arms).* Define $A = \{\omega \subseteq \Omega : y_T = 1\}$. In other words, $A$ is the set of all events such that the "correct" arm is arm 1. (But the argument below holds for any any event $A \subset \Omega$.)

Similar to the previous section, we use the properties of KL divergence as follows:

$$2(P_1(A) - P_2(A))^2 \leq KL(P_1, P_2)$$

[1] Recall that the "best arm" is the arm with the highest mean reward.

$$= \sum_{a=1}^{K=2} \sum_{t=1}^{T} KL(P_1^{a,t}, P_2^{a,t})$$
$$\leq 2T \cdot 2\epsilon^2 \tag{7}$$

The last inequality is because KL divergence of $P_1^{a,t}$ and $P_2^{a,t}$ is non-zero if and only if $P_1^{a,t} \neq P_2^{a,t}$. And when they are non-equal, their KL divergence is $2\epsilon^2$. Hence,

$$|P_1(A) - P_2(A)| \leq 2\epsilon\sqrt{T} < \tfrac{1}{2}.$$

The last inequality holds whenever $T \leq (\frac{1}{4\epsilon})^2$.

To complete the proof, observe that if $\Pr[y_T = j \mid \mathcal{I}_j] \geq \frac{3}{4}$ for both problem instances, then $P_1(A) \geq \frac{3}{4}$ and $P_2(A) < \frac{1}{4}$, so their difference is at least $\frac{1}{2}$, contradiction. □

*Proof ($K \geq 24$).* Compared to the 2-arms case, time horizon $T$ can be larger by a factor of $O(K)$. The crucial improvement is a more delicate version of the KL-divergence argument in (7) which results in the right-hand side of the form $O(T\epsilon^2/K)$.

For the sake of the analysis, we will consider an additional problem instance

$$\mathcal{I}_0 = \{\ \mu_i = \tfrac{1}{2}\ \ :\ \text{for all arms a}\ \}$$

which we call the "base instance". Let $\mathbb{E}_0[\cdot]$ be the expectation given this problem instance. Also, let $T_a$ be the total number of times arm $a$ is played.

We consider the algorithm's performance on problem instance $\mathcal{I}_0$, and focus on arms $j$ that are "neglected" by the algorithm, in the sense that the algorithm does not choose arm $j$ very often *and* is not likely to pick $j$ for the guess $y_T$. Formally, we observe that

$$\exists \geq \frac{2K}{3} \text{ arms } j \text{ such that } \mathbb{E}_0(T_j) \leq \frac{3T}{K} \tag{8}$$
$$\exists \geq \frac{2K}{3} \text{ arms } j \text{ such that } P_0(y_T = j) \leq \frac{3}{K}. \tag{9}$$

(To prove (8), assume for contradiction that we have more than $\frac{K}{3}$ arms with $\mathbb{E}_0(T_j) > \frac{3T}{K}$. Then the expected total number of times these arms are played is strictly greater than $T$, which is a contradiction. (9) is proved similarly.) By Markov inequality,

$$\mathbb{E}_0(T_j) \leq \tfrac{3T}{K} \text{ implies that } \Pr[T_j \leq \tfrac{24T}{K}] \geq \tfrac{7}{8}.$$

Since the sets of arms in (8) and (9) must overlap on least $\frac{K}{3}$ arms, we conclude:

$$\exists \geq \frac{K}{3} \text{ arms j such that } \Pr[T_j \leq m] \geq \tfrac{7}{8} \text{ and } P_0(y_T = j) \leq \tfrac{3}{K}, \tag{10}$$

where $m = \frac{24T}{K}$.

We will now refine our definition of the sample space to get the required claim. For each arm $a$, define the $t$-round sample space $\Omega_a^t = \{0,1\}^t$, where each outcome corresponds to a particular realization of the tuple $(r_s(a) :\ s \in [t])$. (Recall that we interpret $r_t(a)$ as the reward received by the algorithm for the $t$-th time it chooses arm $a$.) Then the "full" sample space we considered before can be expressed as $\Omega = \prod_{a \in \mathcal{A}} \Omega_a^T$.

Fix an arm $j$ satisfying the two properties in (10). We will consider a "reduced" sample space in which arm $j$ is played only $m = \frac{24T}{K}$ times:

$$\Omega^* = \Omega_j^m \times \prod_{\text{arms } a \neq j} \Omega_a^T. \tag{11}$$

Each problem instance $\mathcal{I}_\ell$ defines a distribution $P_\ell^*$ on $\Omega^*$:

$$P_\ell^*(A) = \Pr[A|\mathcal{I}_\ell] \quad \text{for each } A \subset \Omega^*.$$

In other words, distribution $P_\ell^*$ is a restriction of $P_\ell$ to the reduced sample space $\Omega^*$.

We apply the KL-divergence argument to distributions $P_0^*$ and $P_j^*$. For each event $A \subset \Omega^*$:

$$\begin{aligned} 2(P_0^*(A) - P_j^*(A))^2 &\leq KL(P_0^*, P_j^*) \\ &= \sum_{arm\ a \neq j} \sum_{t=1}^{T} KL(P_0^{a,t}, P_j^{a,t}) + \sum_{t=1}^{m} KL(P_0^{j,t}, P_j^{j,t}) \\ &\leq 0 + m * 2\epsilon^2. \end{aligned}$$

Note that each arm $a \neq j$ has identical distributions under instances $\mathcal{I}_0$ and $\mathcal{I}_j$ (namely, its mean reward is $\frac{1}{2}$). So distributions $P_0^{a,t}$ and $P_j^{a,t}$ are the same, and therefore their KL-divergence is 0. Whereas for arm $j$ we only need to sum up over $m$ samples.

Therefore, assuming $T \leq \frac{cK}{\epsilon^2}$ with small enough constant $c$, we can conclude that

$$|P_0^*(A) - P_j^*(A)| \leq \epsilon\sqrt{m} < \tfrac{1}{8} \quad \text{for all events } A \subset \Omega^*. \tag{12}$$

To apply (12), we need to make sure that the event $A$ is in fact contained in $\Omega^*$, *i.e.,* whether $A$ holds is completely determined by the first $m$ samples of arm $j$ (and arbitrarily many samples of other arms). In particular, we cannot take $A = \{y_t = j\}$, which would be the most natural extension of the proof technique from the 2-arms case. Instead, we apply (12) twice: to events

$$A = \{y_T = j \text{ and } T_j \leq m\} \text{ and } A' = \{T_j > m\}. \tag{13}$$

Indeed, note that whether the algorithm samples arm $j$ more than $m$ times is completely determined by the first $m$ coin tosses!

We are ready for the final computation:

$$\begin{aligned} P_j(A) &\leq \tfrac{1}{8} + P_0(A) && \text{by (12)} \\ &\leq \tfrac{1}{8} + P_0(y_T = j) \\ &\leq \tfrac{1}{4} && \text{by our choice of arm } j. \\ P_j(A') &\leq \tfrac{1}{8} + P_0(A') && \text{by (12)} \\ &\leq \tfrac{1}{4} && \text{by our choice of arm } j. \\ P_j(Y_T = j) &\leq P_j^*(Y_T = j \text{ and } T_j \leq m) + P_j^*(T_j > m) \\ &= P_j(A) + P_j(A') \leq \tfrac{1}{4}. \end{aligned}$$

Recall that this holds for any arm $j$ satisfying the properties in (10). Since there are at least $K/3$ such arms, the lemma follows. □

**Next lecture: Lemma 4.2 is used to derive the $\Omega(\sqrt{KT})$ lower bound on regret.**

## 5 Bibliographic notes

The $\Omega(\sqrt{KT})$ lower bound on regret is from Auer et al. (2002). KL-divergence and its properties is "textbook material" from Information Theory, *e.g.,* see Cover and Thomas (1991). The present exposition — the outline and much of the technical details — is based on Robert Kleinberg's lecture notes from (Kleinberg, 2007).

We present a substantially simpler proof compared to (Auer et al., 2002; Kleinberg, 2007) in that we avoid the general "chain rule" for KL-divergence. Instead, we only use the special case of independent distributions (Property 2 in Section 2), which is much easier to state and to apply. The proof of Lemma 4.2 (for general $K$), which in prior work relies on the general "chain rule", is modified accordingly. In particular, we define the "reduced" sample space $\Omega^*$ with only a small number of samples from the "bad" arm $j$, and apply the KL-divergence argument to carefully defined events in (13), rather than a seemingly more natural event $A = \{y_T = j\}$.

## References


Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM J. Comput.*, 32(1):48–77, 2002. Preliminary version in *36th IEEE FOCS*, 1995.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* John Wiley & Sons, New York, 1991.

Robert Kleinberg. Lecture notes: *CS683: Learning, Games, and Electronic Markets* (week 9), 2007. Available at `http://www.cs.cornell.edu/courses/cs683/2007sp/lecnotes/week9.pdf`.