2 Lower Bounds
2.1 Background on KL-divergence
2.2 A simple example: flipping one coin
2.3 Flipping several coins: "best-arm identification"
2.4 Proof of Lemma 2.5 for K ≥ 24 arms
2.5 Instance-dependent lower bounds (without proofs)
2.6 Bibliographic remarks and further directions
2.7 Exercises and Hints

2 Lower Bounds

Notation	Name[19]	Description	Formal Definition	Limit Definition[20][21][22][19][14]
$f(n) = O(g(n))$	Big O; Big Oh; Big Omicron	$\lvert f\rvert$ is bounded above by g (up to constant factor) asymptotically	$\exists k > 0\ \exists n_0\ \forall n > n_0\ \lvert f(n)\rvert \leq k \cdot g(n)$	$\limsup_{n\to\infty} \frac{\lvert f(n)\rvert}{g(n)} < \infty$
$f(n) = \Theta(g(n))$	Big Theta	f is bounded both above and below by g asymptotically	$\exists k_1 > 0\ \exists k_2 > 0\ \exists n_0\ \forall n > n_0$ $k_1 \cdot g(n) \leq f(n) \leq k_2 \cdot g(n)$	$f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$ (Knuth version)
$f(n) = \Omega(g(n))$	Big Omega in complexity theory (Knuth)	f is bounded below by g asymptotically	$\exists k > 0\ \exists n_0\ \forall n > n_0\ f(n) \geq k \cdot g(n)$	$\liminf_{n\to\infty} \frac{f(n)}{g(n)} > 0$

2.1 Background on KL-divergence

KL-divergence: Throughout, consider a finite sample space Ω, and let p, q be two probability distributions on Ω. Then, the Kullback-Leibler divergence or KL-divergence is defined as $\mathrm{KL}(p,q) = \sum_{x\in\Omega} p(x) \ln \frac{p(x)}{q(x)} = \mathbb{E}_p \left[\ln \frac{p(x)}{q(x)}\right]$

let $\mathrm{RC}_\epsilon\ \epsilon \geq 0$, denote a biased random coin with bias $\frac{\epsilon}{2}, i.e.$, a distribution over {0,1} with expectation $(1+\epsilon)/2$

Theorem 2.2. KL-divergence satisfies the following properties:

1. Gibbs' Inequality: $\mathrm{KL}(p,q) \geq 0$ for any two distributions p, q with equality if and only if $p = q$
2. Chain rule for product distributions: Let the sample space be a product $\Omega = \Omega_1 \times \Omega_1 \times \cdots \times \Omega_n$. Let p and q be two distributions on Ω such that $p = p_1 \times p_2 \times \cdots \times p_n$ and $q = q_1 \times q_2 \times \cdots \times q_n$, where p_j, q_j are distributions on Ω_j, for each $j \in [n]$. $\mathrm{KL}(p,q) = \sum_{j=1}^n \mathrm{KL}\left(p_j, q_j\right)$
3. Pinsker's inequality: for any event $A \subset \Omega$ we have $2(p(A) - q(A))^2 \leq \mathrm{KL}(p,q)$

4. Random coins: $\mathrm{KL}(\mathrm{RC}_\epsilon, \mathrm{RC}_0) \leq 2\epsilon^2$, and $\mathrm{KL}\left(\mathrm{RC}_0, \mathrm{RC}_\epsilon\right) \leq \epsilon^2$ for all $\epsilon \in \left(0, \frac{1}{2}\right)$

Lemma 2.3. Consider sample space $\Omega = \{0,1\}^n$ and two distributions on $\Omega, p = \mathrm{RC}_\epsilon^n$ and $q = \mathrm{RC}_0^n$, for some $\epsilon > 0$. Then $|p(A) - q(A)| \leq \epsilon\sqrt{n}$ for any event $A \subset \Omega$

2.2 A simple example: flipping one coin

Consider a biased random coin: a distribution on {0,1} with an unknown mean $\mu \in [0,1]$. Assume that $\mu \in \{\mu_1, \mu_2\}$ for two known values $\mu_1 > \mu_2$. The coin is flipped T times. The goal is to identify if $\mu = \mu_1$ or $\mu = \mu_2$ with low probability of error.

具体来说，Define $\Omega := \{0,1\}^T$ to be the sample space for the outcomes of T coin tosses.

想找一个Rule $\text{Rule} : \Omega \rightarrow \{\text{ High, Low }\}$满足这两个条件：

$$\Pr[\text{ Rule (observations) } = \text{High} \mid \mu = \mu_1] \geq 0.99$$
$$\Pr[\text{ Rule (observations) } = \text{Low} \mid \mu = \mu_2] \geq 0.99$$

从1.7能得到，$T \sim (\mu_1 - \mu_2)^{-2}$这么大足够分辨到底是$\mu_1$还是$\mu_2$了，这里其实还能证明这个数量的观测也是necessary的

Lemma 2.4. Let $\mu_1 = \frac{1+\epsilon}{2}$ and $\mu_2 = \frac{1}{2}$. Fix a decision rule which satisfies (2.2) and (2.3). Then $T > \frac{1}{4\epsilon^2}$

2.3 Flipping several coins: "best-arm identification"

Let us extend the previous example to multiple coins. We consider a bandit problem with K arms, where each arm is a biased random coin with unknown mean. More formally, the reward of each arm is drawn independently from a fixed but unknown Bernoulli distribution. After T rounds, the algorithm outputs an arm y_T : a prediction for which arm is optimal (has the highest mean reward). We call this version "best-arm identification". We are only be concerned with the quality of prediction, rather than regret.

As a matter of notation, the set of arms is $[K]$, $\mu(a)$ is the mean reward of arm a, and a problem instance is specified as a tuple $\mathcal{I} = (\mu(a) : a \in [K])$

For concreteness, let us say that a good algorithm for "best-arm identification" should satisfy $\Pr[\text{ prediction } y_T \text{ is correct } |\mathcal{I}] \geq 0.99$ (2.5) for each problem instance $\mathcal{I}$.

We will use the family (2.1) of problem instances, with parameter $\epsilon > 0,$ to argue that one needs $T \geq \Omega\left(\frac{K}{\epsilon^2}\right)$ for any algorithm to "work", i.e., satisfy property (2.5), on all instances in this family.

(2.1): We consider $0-1$ rewards and the following family of problem instances, with parameter $\epsilon > 0$ to be adjusted in the analysis:
$\mathcal{I}_j = \begin{cases} \mu_i = (1+\epsilon)/2 & \text{for arm } i = j \\ \mu_i = 1/2 & \text{for each arm } i \neq j \end{cases}$ for each $j = 1, 2, \ldots, K$. (Recall that K is the number of arms.)

Lemma 2.5. Consider a "best-arm identification" problem with $T \leq \frac{cK}{\epsilon^2}$ for a small enough absolute constant $c > 0$. Fix any deterministic algorithm for this problem. Then there exists at least $\lceil K/3 \rceil$ arms a such that, for problem instances $\mathcal{I}_a$ defined in (2.1), we have $\Pr[y_T = a|\mathcal{I}_a] < \frac{3}{4}$

Corollary 2.6. Assume T is as in Lemma 2.5. Fix any algorithm for "best-arm identification". Choose an arm a uniformly at random, and run the algorithm on instance $\mathcal{I}_a$. Then $\Pr[y_T \neq a] \geq \frac{1}{12}$, where the probability is over the choice of arm a and the randomness in rewards and the algorithm.

Theorem 2.7. Fix time horizon T and the number of arms K. Fix a bandit algorithm. Choose an arm a uniformly at random, and run the algorithm on problem instance $\mathcal{I}_a$. Then $\mathbb{E}[R(T)] \geq \Omega(\sqrt{KT})$ where the expectation is over the choice of arm a and the randomness in rewards and the algorithm.

2.4 Proof of Lemma 2.5 for K ≥ 24 arms

Notation summary:

$\mathcal{I}_0 = \{\mu_i = \frac{1}{2} \text{ for all arms } i\}$

T_a be the total number of times arm a is played.

y_T: After T rounds, the algorithm outputs an arm y_T : a prediction for which arm is optimal (has the highest mean reward). We call this version "best-arm identification".

Let $P_j^{a,t}$ be the distribution of $r_t(a)$ under instance $\mathcal{I}_j$

Sample space: 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 $\Omega = \prod_{a \in [K]} \Omega_a^T$.

Consider a "reduced" sample space in which arm j is played only $m = \frac{24T}{K}$ times: $\Omega^* = \Omega_j^m \times \prod_{\text{arms } a \neq j} \Omega_a^T$

Each problem instance $\mathcal{I}_j$ defines distribution P_j on $\Omega : P_j(A) = \Pr[A|\mathcal{I}_j]$for each $A \subset \Omega$

For each problem instance $\mathcal{I}_\ell$, we define distribution P_ℓ^* on Ω^* as follows $P_\ell^*(A) = \Pr[A|\mathcal{I}_\ell]$ for each $A \subset \Omega^*$. In other words, distribution P_ℓ^* is a restriction of P_ℓ to the reduced sample space Ω^*.

2.5 Instance-dependent lower bounds (without proofs)

Theorem 2.8. No algorithm can achieve regret $\mathbb{E}[R(t)] = o(c_\mathcal{I} \log t)$ for all problem instances $\mathcal{I}$, where the "constant" $c_\mathcal{I}$ can depend on the problem instance but not on the time t

---

Theorem 2.9. Fix K, the number of arms. Consider an algorithm such that

$\mathbb{E}[R(t)] \leq O(C_{\mathcal{I},\alpha} t^\alpha)$ for each $\alpha > 0$ and problem instance $\mathcal{I}$. (2.15)

Here the "constant" $C_{\mathcal{I},\alpha}$ can depend on the problem instance $\mathcal{I}$ and the α, but not on time t.

Fix an arbitrary problem instance $\mathcal{I}$. For this problem instance:

There is time t_0 such that for any $t \geq t_0 \quad \mathbb{E}[R(t)] \geq C_\mathcal{I} \ln(t)$, (2.16)

for some constant $C_\mathcal{I}$ that depends on the problem instance, but not on time t.

Theorem 2.10. For each problem instance $\mathcal{I}$ and any algorithm that satisfies (2.15)

1. the bound (2.16) holds with

$$C_{\mathcal{I}} = \sum_{a:\Delta(a)>0} \frac{\mu^* (1 - \mu^*)}{\Delta(a)}$$

2. for each $\epsilon > 0$, the bound (2.16) holds with

$$C_{\mathcal{I}} = \sum_{a:\Delta(a)>0} \frac{\Delta(a)}{\mathrm{KL}(\mu(a), \mu^*)} - \epsilon$$

2.6 Bibliographic remarks and further directions

2.7 Exercises and Hints