Alexander Rakhlin

LECTURE 18

1. THE MAGIC OF RANDOM PLAYOUT

Let us now describe a very useful technique for coming up with efficient randomized strategies. The technique is an analogue of "hallucinating future coin flips" in the Cover result. We call it a magical technique because it allows us to seemingly pass an expectation through min max, something that generally does not hold. In the following lecture, we will show that the famous "Follow the Perturbed Leader" technique is a particular consequence of the general approach.

Suppose we have an abstract optimization problem

$$\min_q \max_b \{\mathbb{E}_{a\sim q}\boldsymbol{\ell}(a,b) + \mathbb{E}_{z\sim p}\Phi(z,b)\} \tag{1}$$

where q is a mixed strategy for our choice a. The optimization problem might be computationally difficult due to the presence of the expected value $\mathbb{E}_{z\sim p}$. However, suppose we are able to simulate $z \sim p$. *The random playout technique says that it's "enough" to be able to solve a version of* (1) *without the expected value.* Specifically,

- ▷ Draw $z \sim p$
- ▷ Output $\tilde{q}(z)$, a solution to

$$\tilde{q}(z) = \operatorname*{argmin}_{\tilde{q}} \max_b \{\mathbb{E}_{a\sim q}\boldsymbol{\ell}(a,b) + \Phi(z,b)\} \tag{2}$$

This process defines a mixed strategy which we denote by $\widehat{q}$. Drawing $a \sim \widehat{q}$ requires double randomization: first draw z, then solve for $\tilde{q}(z)$, then draw $a \sim \tilde{q}(z)$. How good is this strategy $\widehat{q}$?

Lemma 1. *Let*

$$V(q) = \max_b \{\mathbb{E}_{a\sim q}\boldsymbol{\ell}(a,b) + \mathbb{E}_{z\sim p}\Phi(z,b)\}$$

be the value in (1) *attained by a strategy* q. *The strategy* $\widehat{q}$ *defined earlier attains*

$$V(\widehat{q}) \le \mathbb{E}_{z\in p} \min_q \max_b \{\mathbb{E}_{a\sim q}\boldsymbol{\ell}(a,b) + \Phi(z,b)\}. \tag{3}$$

This simple result should be a standard tool of anyone looking to develop an efficient randomized strategies. Of course, the question is how good the upper bound (3) is. It turns out, in the questions we consider, it's pretty much as good as it can get. Note that the difference is only in the location of $\mathbb{E}_{z\sim p}$: in (1) it is inside, while in the upper bound (3) it is outside. However, this is not Jensen's inequality: while $\mathbb{E}$ can pass through the max, it cannot pass through the min. The proof, however, is simple:

Proof. By taking $\widehat{q}$ as a particular choice in (1),

$$\min_q \max_b \{\mathbb{E}_{a\sim q}\boldsymbol{\ell}(a,b) + \mathbb{E}_{z\sim p}\Phi(z)\} \leq \max_b \left\{\mathbb{E}_{z\sim p}\mathbb{E}_{a\sim\tilde{q}(z)}\boldsymbol{\ell}(a,b) + \mathbb{E}_{z\sim p}\Phi(z)\right\} \tag{4}$$

$$\leq \mathbb{E}_{z\sim p} \max_b \left\{\mathbb{E}_{a\sim\tilde{q}(z)}\boldsymbol{\ell}(a,b) + \Phi(z)\right\} \tag{5}$$

$$= \mathbb{E}_{z\sim p} \min_{\tilde{q}} \max_b \{\mathbb{E}_{a\sim\tilde{q}}\boldsymbol{\ell}(a,b) + \Phi(z)\} \tag{6}$$

The first inequality is due to the particular (potentially suboptimal) choice of $\widehat{q}$. The second is by linearity and Jensen's inequality (using convexity of max). The third equality is by the definition of $\tilde{q}(z)$. □

The bottom line is that problems of the form (1) can be solved by drawing a random variable and solving (2), thus avoiding the costly integration over the random variable.

2. ONLINE LINEAR OPTIMIZATION

We are now back to the setting of online learning. Recall the experts problem: on round t, we decide on a distribution $\widehat{y}_t \in \Delta(N)$ on the N experts and then observe the costs incurred by the experts. We denote this vector of costs by $z_t \in [-1,1]^N$. The cumulative cost incurred by the learner is

$$\sum_{t=1}^{n} \langle \widehat{y}_t, z_t \rangle ,$$

while the best cumulative cost in hindsight attained by any expert is

$$\min_{v\in\{\boldsymbol{e}_1,\ldots,\boldsymbol{e}_N\}} \sum_{t=1}^{n} \langle v, z_t \rangle$$

($\boldsymbol{e}_j$ are the standard basis vectors).

We can view this problem as a game: player I chooses actions in $\Delta(N)$, player II chooses actions in $[-1,1]^N$, and the benchmark is defined as

$$\min_{v\in\Delta(N)} \sum_{t=1}^{n} \langle v, z_t \rangle .$$

The minimization is over the simplex since the minimum of a linear function is achieved at the vertex.

To develop a more general machinery we shall consider a slightly different problem: we expand the set $\Delta(N)$ to be the unit ℓ_1 ball

$$B_1 = \left\{ v : \sum_{j=1}^{N} |v(j)| \leq 1 \right\} = \{v : \|v\|_1 \leq 1\} .$$

Clearly, B_1 includes $\Delta(N)$. Dual to ℓ_1 norm is the ℓ_∞ norm:

$$\|z\|_\infty = \max_j |z(j)|,$$

and the dual ball B_∞ is precisely $[-1,1]^N$. We arrived at the following online learning protocol:

For $t = 1, \dots, n$

Predict $\widehat{y}_t \in B_1$

Observe costs $z_t \in B_\infty$

One can check easily that

$$-\min_{v \in B_1} \sum_{t=1}^n \langle v, z_t \rangle = \max_{v \in B_1} \left\langle v, \sum_{t=1}^n -z_t \right\rangle = \left\| \sum_{t=1}^n z_t \right\|_\infty \tag{7}$$

and thus the (unnormalized by n) *regret*

$$\sum_{t=1}^n \langle \widehat{y}_t, z_t \rangle - \min_{v \in B_1} \sum_{t=1}^n \langle v, z_t \rangle \tag{8}$$

of the algorithm with respect to the best "signed" expert $\pm \boldsymbol{e}_j$ can be written as

$$\sum_{t=1}^n \langle \widehat{y}_t, z_t \rangle + \left\| \sum_{t=1}^n z_t \right\|_\infty . \tag{9}$$

Aside: The choice of the pair B_1/B_∞ is not arbitrary. As we will see shortly, the analysis remains unchanged for any pair of dual balls: e.g B_2/B_2, B_p/B_q with $1/p + 1/q = 1$, and so on. An abstract online linear optimization problem with dual balls B and B_* is defined as a repeated game where player I chooses from B, player II chooses from B_*, and regret is the difference of inner products of the choices relative to a single choice by player I.

2.1 Preparing for the magic

Unfortunately, in the problem of online linear optimization, the form of the relaxation is not in the form (1). The technique we describe here fixes this problem.

Lemma 2. *Let D be a uniform distribution on $\{\pm 1\}^N$. Then for any $v \in \mathbb{R}^N$ and any distribution p on B_∞ (which may be chosen based on v),*

$$\mathbb{E}_{z \sim p} \left\| v + (z - \mathbb{E}[z]) \right\|_\infty \le \mathbb{E}_{u \sim D} \left\| v + 6u \right\|_\infty \tag{10}$$

The proof of this lemma is simple and is left as an exercise. But what does the lemma say? Imagine that you know a vector v, and your task is to come up with a distribution on $\{\pm 1\}^N$ such that, on average over the draw of z from this distribution, the maximum coordinate of v when perturbed by $(z - \mathbb{E}[z])$ is as large as possible (this may or may not be the largest coordinate of v itself). Suppose there is some best distribution $p^*(v)$ to achieve this task. The lemma says that one can do no worse by taking a uniform distribution D (that does not depend on the particular v) and measuring the maximum coordinate of v when perturbed by 6 times the draw from this distribution. That is, up to a constant 6, *there is no gain in tailoring p to the particular value of v.*

A couple of remarks. First, the constant 6 may be improvable, but this value allows for a simple proof. Second, many other choices for D are possible. Of particular interest in a gaussian distribution with independent coordinates.

The lemma, while being very simple, allows us to relate martingales back to i.i.d. random variables. This can be seen in the proof in the next section. *Statements as in Lemma 2 imply equivalence of learning with i.i.d. and worst-case sequences.*

2.2 Non-algorithmic analysis of online linear optimization for B_1/B_∞

To find an upper bound on the optimal value of regret and to develop a near-minimax strategy, consider the last step:

$$\min_{\widehat{y}_n \in B_1} \max_{z_n \in B_\infty} \{\langle \widehat{y}_n, z_n \rangle + \mathbf{Rel}(z_{1:n})\} \tag{11}$$

with

$$\mathbf{Rel}(z_{1:n}) = \left\| \sum_{t=1}^{n} z_t \right\|_\infty .$$

Now that we learned about minimax basics, we can understand why the min and max are over the deterministic choices (can you see why?) The inner maximization is over the deterministic choice z_n since the inner player can be deterministic. How about the outer min? The B_1 ball is convex, and the objective inside the curly brackets is linear (convex) in $\widehat{y}_n$. Hence, we may appeal to the lemma from the previous lecture.

However, we cannot perform the minimax swap since the objective function is convex (rather than concave) in z_n (can you see why?). The solution is to linearize the choice by moving to the space of distributions. This is the first key step in our analysis, and one should always keep this possibility in mind when facing a minimax expression.

We can equivalently write (11) as

$$\min_{\widehat{y}_n \in B_1} \max_{p_n \in \Delta(B_\infty)} \mathbb{E}_{z_n \sim p_n} \left\{ \langle \widehat{y}_n, z_n \rangle + \left\| \sum_{t=1}^{n} z_t \right\|_\infty \right\}, \tag{12}$$

with the maximum over probability distributions on $z_n \in B_\infty$. The objective is now linear in $\widehat{y}_n$ and linear in p_n. Compactness and continuity required by the theorem from the previous lecture hold, and so the above expression is equal to

$$\max_{p_n \in \Delta(B_\infty)} \min_{\widehat{y}_n \in B_1} \mathbb{E}_{z_n \sim p_n} \left\{ \langle \widehat{y}_n, z_n \rangle + \left\| \sum_{t=1}^{n} z_t \right\|_\infty \right\}. \tag{13}$$

Now, the expectation distributes over the two terms (by linearity of expectation) and the minimum only applies to the first term. Hence, the preceding expression is equal to

$$\max_{p_n \in \Delta(B_\infty)} \left\{ \min_{\widehat{y}_n \in B_1} \langle \widehat{y}_n, \mathbb{E}_{z_n \sim p_n}[z_n] \rangle + \mathbb{E}_{z_n \sim p_n} \left\| \sum_{t=1}^{n} z_t \right\|_\infty \right\}. \tag{14}$$

But

$$\min_{\widehat{y}_n \in B_1} \langle \widehat{y}_n, \mathbb{E} z_n \rangle = - \|\mathbb{E} z_n\|_\infty , \tag{15}$$

and, hence, by triangle inequality, the minimax expression is upper bounded by

$$\max_{p_n \in \Delta(B_\infty)} \mathbb{E}_{z_n \sim p_n} \left\| \sum_{t=1}^{n-1} z_t + (z_n - \mathbb{E} z_n) \right\|_\infty . \tag{16}$$

We now appeal to Lemma 2 and upper bound the above expression by

$$\mathbb{E}_{u_n \sim D} \left\| \sum_{t=1}^{n-1} z_t + 6u_n \right\|_\infty \triangleq \mathbf{Rel}(z_{1:n-1}) \tag{17}$$

Continuing in this manner,

$$\mathbf{Rel}(z_{1:t}) = \mathbb{E}_{u_{t+1:n}\sim D}\left\|\sum_{j=1}^{t} z_j + 6\sum_{s=t+1}^{n} u_s\right\|_{\infty} \tag{18}$$

and

$$\mathbf{Rel}(\varnothing) = 6\mathbb{E}_{u_{1:n}\sim D}\left\|\sum_{s=1}^{n} u_s\right\|_{\infty} \tag{19}$$

where D is uniform on $\{\pm 1\}^n$. Let us examine the vector $U = \sum_{s=1}^{n} u_s$. Each coordinate i of U is a sum of n independent Rademacher random variables, and thus

$$\mathbb{E}\exp\{\lambda U(i)\} \le \exp\{n\lambda^2/2\}$$

(prove this using the inequality we already discussed). Using the soft-max argument from previous lectures,

$$\mathbb{E}_{u_{1:n}\sim D}\left\|\sum_{s=1}^{n} u_s\right\|_{\infty} \le 6\sqrt{2n\log N}. \tag{20}$$

Once again we proved the experts bound, but this time we did not specify the algorithm. The soft-max relaxation was only used at the very end.

In the next lecture, we ask what algorithm corresponds to the above relaxation. We will see that it is the famous Follow-the-Perturbed-Leader. This algorithm is always presented as an ad-hoc algorithm, but we see that it corresponds to a very particular sequence of relaxation steps. Once this is understood, the technique can be applied in a principled manner where no prediction method is presently known.