Algorithms and Uncertainty, Summer 2020 Lecture 10 (5 pages)

Adaptivity Gap

Thomas Kesselheim Last Update: May 31, 2020

Consider the following problem: There are n boxes. Each box contains a prize, which we only get to know when we open it. Before, we only know the probability distribution the prize is drawn from (which might be different for different boxes). We are allowed to open k boxes and we will keep *the highest* prize that we see in these boxes. The question is which boxes to open in which order. More precisely, we are allowed to open a box and then, depending on the actual prize in the box, choose which box to open next and so on.

It is easy to model this problem as a Markov decision process. In the state space, we have to keep track of which boxes were opened so far and which was the highest prize in these boxes. Letting $X_1, \dots, X_n$ denote the (random) prizes in the boxes and Opened $\subseteq \{1, \dots, n\}$ the (possibly random) set of boxes that are opened, the reward is given by

$$\text{reward} = \max_{i \in \text{Opened}} X_i \ .$$

In advance, we know the probability distributions of all X_i and we assume that they have finite support. We write $f_{i,v}$ for $\mathbf{Pr}[X_i = v]$.

Example 10.1. *We have three boxes; so $n = 3$. The first box contains a prize of 24 with probability $\frac{1}{2}$ and 0 otherwise. The second box contains a prize of 30 with probability $\frac{1}{3}$ and 0 otherwise. The third box contains a prize of 12 with probability 1. We are allowed to open two boxes, i.e., $k = 2$.*

Let us first consider the policies that in advance fix which boxes to open. The expected rewards are depicted in the following table.

Boxes opened	*Expected prize*
1, 2	$\mathbf{E}[\max\{X_1, X_2\}] = \frac{1}{3} \cdot 30 + \frac{2}{3} \cdot \frac{1}{2} \cdot 24 = 18$
1, 3	$\mathbf{E}[\max\{X_1, X_3\}] = \frac{1}{2} \cdot 24 + \frac{1}{2} \cdot 12 = 18$
2, 3	$\mathbf{E}[\max\{X_2, X_3\}] = \frac{1}{3} \cdot 30 + \frac{2}{3} \cdot 12 = 18$

So the highest expected prize we can achieve using one of these policies is 18. However, we can do better than this: Open the first box. If it contains a prize, we have 24 for sure. So, opening the third box does not make any sense at this point and we continue with the second one. If, however, the first box is empty, we continue with the third box. The expected prize this way is $\frac{1}{2}(\frac{1}{3} \cdot 30 + \frac{2}{3} \cdot 24) + \frac{1}{2} \cdot 12 = 19$.

As we realize in this example, the choice in the second step depends on what we found in the first box. That is, the optimal policy is *adaptive*. Adaptive policies are generally complicated: We need a huge decision tree to represent them. Even if each X_i can only take two values, this tree has 2^k nodes.

Our question today is: What if one uses a simpler policy instead? How much worse is a *non-adaptive* policy, which will simply open a suitably chosen set of boxes and not adapt the choices based on the values seen?

For this and similar problems, one can quantify the loss by the so-called *adaptivity gap*, which is defined as

$$\frac{\max_{\text{any policy } \pi^*} V(s_1, \pi^*, T)}{\max_{\text{non-adaptive policy } \pi} V(s_1, \pi, T)} \ .$$

So, we compare how much more expected reward an adaptive policy can obtain in comparison to a non-adaptive policy.

We have already seen that the adaptivity gap in the example is at least $\frac{19}{18} \approx 1.056$. Our goal today will be to show that it is at most 8.[1] Our proof will be constructive. We will design an algorithm to compute a non-adaptive policy π and we will show that no policy can obtain more than 8-times the reward, adaptive or not.

1 An LP Relaxation

As a first step, we will devise a linear program (LP) such that the expected reward of any (adaptive) policy is upper-bounded by the optimal solution to the LP. In the following step, we will then construct a non-adaptive policy from the optimal LP solution. The loss that we incur in this second step is clearly an upper bound to the adaptivity gap.

To derive the LP, fix any policy π and observe its execution. We define random variables Y_i and $Z_{i,v}$ as follows. Let $Y_i = 1$ if box i is opened, 0 otherwise. Let $Z_{i,v} = 1$ if box i contains a prize of v and is selected. Based on this, define $y_i = \mathbf{E}[Y_i]$ and $z_{i,v} = \mathbf{E}[Z_{i,v}]$. Note that now y_i denotes the probability that box i is opened and that $z_{i,v}$ is the probability that box i contains prize v and it is selected (i.e., this is the prize that is kept eventually).

Example 10.2. *Consider the adaptive policy from Example 10.1. The first box is always opened, therefore $y_1 = 1$. The other boxes are opened each with probability $\frac{1}{2}$, so $y_2 = y_3 = \frac{1}{2}$.*

The value of $z_{1,24}$ is determined as follows. Given that the first box contains prize 24, we open the second box. With probability $\frac{2}{3}$, it is empty, and we select the 24. So the overall probability of this happening is $z_{1,24} = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$.

For $z_{3,12}$, we observe that a prize of 12 from the third box is always selected if this box is opened. This happens with probability $\frac{1}{2}$. So, $z_{3,12} = \frac{1}{2}$.

Finally, for $z_{2,30}$, we use that a prize of 30 from the second box is selected only if this box is opened and if it contains the respective prize. This happens with probability $\frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$. So, $z_{2,30} = \frac{1}{6}$.

We now observe some properties of y_i and $z_{i,v}$.

First, observe that the expected reward of the policy is

$$V(s_1, \pi, T) = \mathbf{E}\left[\sum_{i,v} v \cdot Z_{i,v}\right] = \sum_{i,v} v \cdot \mathbf{E}[Z_{i,v}] = \sum_{i,v} v z_{i,v} \tag{1}$$

by linearity of expectation.

Furthermore, the policy opens at most k boxes, regardless of the random outcomes. Therefore, $\sum_i Y_i \leq k$ with probability 1. This inequality still holds if we take the expectation on both sides, giving us

$$\sum_i y_i = \sum_i \mathbf{E}[Y_i] = \mathbf{E}\left[\sum_i Y_i\right] \leq k \ . \tag{2}$$

By definition, eventually only a single prize of one box is selected. Therefore $\sum_{i,v} Z_{i,v} \leq 1$ with probability 1. This gives us

$$\sum_{i,v} z_{i,v} = \sum_{i,v} \mathbf{E}[Z_{i,v}] = \mathbf{E}\left[\sum_{i,v} Z_{i,v}\right] \leq 1 \ . \tag{3}$$

[1] With more careful analyses better bounds can be obtained. The techniques, however, are similar.

Finally, recall that $Z_{i,v} = 1$ if and only if $Y_i = 1$, box i contains prize v, and v is the highest prize in any opened box. Ignoring this last condition, we get

$$\mathbf{Pr}\left[Z_{i,v} = 1\right] \leq \mathbf{Pr}\left[Y_i = 1 \text{ and box } i \text{ contains prize } v\right] \ .$$

Note that the two events if box i gets opened and if it contains some prize have to be independent. Therefore

$$\mathbf{Pr}\left[Y_i = 1 \text{ and box } i \text{ contains prize } v\right] = \mathbf{Pr}\left[Y_i = 1\right] \cdot \mathbf{Pr}\left[\text{box } i \text{ contains prize } v\right] = f_{i,v} y_i \ .$$

So

$$z_{i,v} = \mathbf{E}\left[Z_{i,v}\right] = \mathbf{Pr}\left[Z_{i,v} = 1\right] \leq \mathbf{Pr}\left[Y_i = 1 \text{ and box } i \text{ contains prize } v\right] = f_{i,v} y_i \ . \tag{4}$$

All of the above expressions are linear in y_i and $z_{i,v}$. Therefore, we may also use them as variables in an LP as follows.

$$\text{maximize} \quad \sum_{i,v} v \cdot z_{i,v} \tag{5}$$

$$\text{subject to} \quad \sum_{i} y_i \leq k \tag{6}$$

$$\sum_{i,v} z_{i,v} \leq 1 \tag{7}$$

$$z_{i,v} \leq f_{i,v} \cdot y_i \qquad \text{for all } i,\ v \tag{8}$$

$$z_{i,v} \geq 0 \qquad \text{for all } i,\ v \tag{9}$$

Lemma 10.3. *The expected reward of any adaptive policy is upper-bounded by the value of the optimal LP solution.*

Proof. We observe that every policy corresponds to an LP solution. The objective function (5) is the expected reward of the policy by (1). It is feasible because (6) is fulfilled due to (2), (7) due to (3), and (8) due to (4). So the optimal LP solution can only be better than the expected reward of the optimal policy. □

Note that not every feasible LP solution necessarily corresponds to a feasible policy.

Example 10.4. *Consider the case of two boxes. The first one contains a prize of* 2 *with probability* $\frac{1}{2}$ *and is empty otherwise. The second one contains a prize of* 1 *with probability* $\frac{1}{2}$ *and is empty otherwise. We are allowed to open two boxes. This means, we do not actually have a choice to make because we can open all boxes. The expected prize is* $\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot \frac{1}{2} \cdot 1 = 1.25$.

However, it is a feasible LP solution to set $y_1 = y_2 = 1$ *and* $z_{1,2} = z_{2,1} = \frac{1}{2}$. *The value is* $2z_{1,2} + 1z_{2,1} = 1.5$. *The reason is that Constraint* (7) *only requires us to take not more than one prize* in expectation. *The policy described by this LP solution sometimes takes both prizes and sometimes none. This is not allowed but our LP has no constraint to enforce it.*

2 From LP Solutions to Policies

Despite the fact that not all LP solutions correspond to feasible policies, we can derive feasible ones from them. Clearly, there has to be a loss in this step. Moreover, the policy that we derive will be non-adaptive. It will only open a (random) set of boxes.

- Solve the LP, get optimal solution (y, z)

- For i from 1 to n, as long as less than k have been opened
  - Open box i with probability $\frac{y_i}{4}$
- Keep the highest prize

To analyze this policy, we use the following one, which clearly has no larger expected reward.

- Solve the LP, get optimal solution (y, z)
- For i from 1 to n, as long as less than k have been opened
  - Open box i with probability $\frac{y_i}{4}$
  - Observe prize v in this box, select it with probability $\frac{z_{i,v}}{f_{i,v} \cdot y_i}$ without looking at further boxes

So, this policy is even stronger than it would need to be. Immediately after seeing the prize in a box, it decides whether this is the final prize to keep. Nonetheless, we can show the following.

Theorem 10.5. *The immediate-decision policy has expected reward at least* $\frac{1}{8} \sum_{i,v} v z_{i,v}$.

So, as any policy corresponds to a feasible LP solution, this bounds the adaptivity gap by 8.

Proof. Let us again define an indicator random variable $Z_{i,v}$ by setting $Z_{i,v} = 1$ if box i is opened, contains value v, and is selected and $Z_{i,v} = 0$ otherwise. It can happen that the for loop does not reach iteration i. In these cases $Z_{i,v} = 0$. Otherwise, for $Z_{i,v}$, we first have to open the box and then select the prize inside. Note that reaching iteration i, opening the box, and selecting it are three independent events: The first one only depends on what happens in iterations $1, \dots, i-1$, the second one only on the random coin flip if we open the box, and the third one only on the prize inside the box, which was irrelevant up to this point. Therefore, we have

$$\begin{aligned} &\mathbf{Pr}\left[Z_{i,v} = 1 \mid \text{the for loop reaches iteration } i\right] \\ &= \frac{y_i}{4} \cdot f_{i,v} \cdot \frac{z_{i,v}}{f_{i,v} \cdot y_i} = \frac{z_{i,v}}{4} \ . \end{aligned}$$

We will show that $\mathbf{Pr}\left[\text{the for loop reaches iteration } i\right] \geq \frac{1}{2}$. This then implies

$$\mathbf{E}\left[\sum_{i,v} v Z_{i,v}\right] = \sum_{i,v} v \mathbf{E}\left[Z_{i,v}\right] \geq \sum_{i,v} v \frac{1}{2} \frac{z_{i,v}}{4} = \frac{1}{8} \sum_{i,v} v z_{i,v} \ ,$$

which proves the claim.

To bound the probability that the for loop reaches iteration i, we use two standard tools from the analysis of randomized algorithms.

Lemma 10.6 (Markov's inequality)**.** *For any non-negative random variable X and any $\alpha > 0$, we have*

$$\mathbf{Pr}\left[X \geq \alpha\right] \leq \frac{\mathbf{E}\left[X\right]}{\alpha} \ .$$

Lemma 10.7 (Union Bound)**.** *For any sequence of not necessarily disjoint events $\mathcal{E}_1, \mathcal{E}_2 \dots$, we have*

$$\mathbf{Pr}\left[\mathcal{E}_1 \cup \mathcal{E}_2 \cup \dots\right] \leq \mathbf{Pr}\left[\mathcal{E}_1\right] + \mathbf{Pr}\left[\mathcal{E}_2\right] + \dots \ .$$

We only have to show that $\mathbf{Pr}$ [the for loop does not reach iteration i] $\leq \frac{1}{2}$. We split this up into the two events that too many boxes are opened or one box is selected before. By union bound, we have

$$\begin{aligned}&\mathbf{Pr}\left[\text{the for loop does not reach iteration } i\right]\\ &\leq \mathbf{Pr}\left[k \text{ boxes are opened in iterations } 1,\dots i-1\right]\\ &\quad + \mathbf{Pr}\left[\text{a box is selected in iterations } 1,\dots,i-1\right]\end{aligned}$$

We will show that both probabilities are upper-bounded by $\frac{1}{4}$.

Again, let $Y_{i'} = 1$ if box i' is opened, 0 otherwise. The expected number of boxes of $1,\dots,i-1$ that are opened is

$$\mathbf{E}\left[\sum_{i'<i} Y_{i'}\right] \leq \sum_{i'<i} \frac{y_{i'}}{4} \leq \frac{k}{4} \ .$$

So, by Markov's inequality, we have

$$\mathbf{Pr}\left[\sum_{i'<i} Y_{i'} \geq k\right] \leq \frac{\mathbf{E}\left[\sum_{i'<i} Y_{i'}\right]}{k} \leq \frac{1}{4} \ .$$

This shows the first bound.

The expected number of times one of the boxes $1,\dots,i-1$ is selected is

$$\mathbf{E}\left[\sum_{i'<i,v} Z_{i',v}\right] = \sum_{i'<i,v} \mathbf{E}\left[Z_{i',v}\right] \leq \sum_{i'<i,v} \frac{z_{i',v}}{4} \leq \frac{1}{4} \ .$$

Markov's inequality gives us

$$\mathbf{Pr}\left[\sum_{i'<i,v} Z_{i',v} \geq 1\right] \leq \frac{\mathbf{E}\left[\sum_{i'<i,v} Z_{i,v}\right]}{1} \leq \frac{1}{4} \ .$$

This shows the second bound and completes the proof. □