CS 683—Learning, Games, and Electronic Markets Spring 2007

Notes for Week 3: Learning in non-zero sum games

Instructor: Robert Kleinberg *February 05–09, 2007*
Notes by: Yogi Sharma

1 Announcements

The material in these lectures is mostly from Sections 7.4–7.8 of the book Prediction, Learning, and Games, by Cesa-Bianchi and Lugosi.

2 Preamble

Up until now in the course, we have been looking at zero-sum games and how the theory of learning can facilitate finding their Nash equilibria. This investigation culminated in the proof of von Neumann's Theorem using learning and some nice consequences of this proof, for example small support of equilibrium strategties (see notes from last week for details).

At the end of the last week, we also noticed that things become a little more subtle once we start looking at non-zero sum games. We gave an example of a non-zero sum game—called "Rock-Paper-Scissors with high stakes"—to demonstrate that the usual learning algorithm, which did well in zero-sum games, might not even converge to a mixed strategy. That gave us a sense that issues involved in the theory of learning in non-zero sum games are going to be more subtle, and probably more interesting.

This week, we are going to examine the theory of non-zero sum games. We will prove the famous theorem of Nash that there always exists a mixed Nash equilibrium in finite size normal-form games, and then we will go on to define an alternate notion of equilibrium, called the correlated equilibrium.

We will be looking at the concept of correlated equilibrium for a couple of different reasons. The first has to do with computational issues. As it turns out, although Nash equlibria always exist, computing a Nash equilibrium of a normal form game is believed to be computationally hard. This erodes the belief that Nash equilibria are a useful model of rationality of players. If players cannot compute an equilibrium, can we assume that they can play it, even if it is rational? Therefore, we will look at correlated equilibrium, a computationally feasible alternative notion of equilibrium.

Secondly, we want to look at how agents/players play a game in a distributed manner, and how they change their beliefs about others. This learning perspective also gives rise to correlated equilibria.

Once we define the notion of correlated equilibrium, we will look at two different processes through which players can be thought of as converging to correlated equilibrium. One of them derives a distribution of play converging to correlated equilibrium

by playing according to an algorithm having no *internal regret* (an alternate notion of regret which will be defined below). The second derivation uses a forecasting algorithm with a property called *calibration*; we will develop such an algorithm and prove that players who forecast their opponents' behavior using a calibrated rule, and then play a best response, will converge to the set of correlated equilibria. Underlying both of these constructions is a theorem about online multi-objective optimization called *Blackwell's approachability theorem*. In this whole process, we will also introduce the concept of *martingales*, which turn out to be a very powerful tool in analyzing the kind of stochastic processes we will be working with, i.e. processes with limited dependence between the past and future.

3 Existence of Nash equilibrium

We now state the main result we are going to prove in this section.

Theorem 1. *For every normal form game $\mathscr{G} = (\mathcal{I}, A_i, u_i)$ with $|\mathcal{I}|, |A| < \infty$ there exists a mixed strategy Nash equilibrium.*

This theorem will be proved using a famous result in topology, called the Brouwer Fixed Point Theorem, which we state below.

Theorem 2 (Brouwer Fixed Point Theorem). *If $X \subseteq \mathbb{R}^n$ is a compact (i.e. closed and bounded) convex set, then every continuous function $f : X \to X$ has a fixed point, i.e., an x such that $f(x) = x$.*

The proof of this theorem is beyond the scope of these lectures. We will just use this theorem as a "black box" to prove Theorem 1.

An attempt at the proof of Theorem 1. We somehow want to define a function f from the set of mixed stragegy profiles to itself such that Nash equilibria are its fixed points. Let us try a natural approach.

Let $k = |\mathcal{I}|$, and let $X = \{(p_1, p_2, \ldots, p_k) : p_i \in \mathbf{\Delta}(A_i)\}$, that is, X is the set of all mixed strategy profiles for the k players. Let us define the function f by

$$f(p_1, p_2, \ldots, p_k) = (q_1, q_2, \ldots, q_k) \quad \text{such that} \quad q_i \in \arg\max_q u_i(q, p_{-i}).$$

That is, q_i is a best response for player i when the current strategies are given by $(p_i)_{i=1}^k$. (Note that a point $\vec{p} = (p_1, \ldots, p_k)$ satisfying $f(\vec{p}) = \vec{p}$ is a mixed strategy Nash equilibrium.)

There is an immediate problem with this definition, which is that this function is not well defined. There could be many best reponses for a player i. In fact, if q_i and q_i' are both best reponses, then so is their convex combination $\lambda q_i + (1-\lambda)q_i'$ for $\lambda \in (0, 1)$. We could choose a unique best response from the set of all best responses, but the following example shows that when we do this, it is impossible to keep the function f continuous.

Example 1. Let us consider the matching pennies (or penalty-kick) game where the payoffs are given in Table 1. In this example, if the probability of goalie going right

		Striker (II)	
		Left (L)	Right (R)
Goalie (I)	Left (L)	$(+1, -1)$	$(-1, +1)$
	Right (R)	$(-1, +1)$	$(+1, -1)$

Table 1: Payoffs for the matching pennies game

is less than a half, then the best strategy for the striker is to go right with probability one. In the opposite case when the probability of goalie going right is more than a half, then the best strategy for the striker is to go left with probability one. The striker is indifferent between his/her strategies when goalie goes left and right with equal probability. See Figure 1 for an illustration. In this example, when the goalkeeper

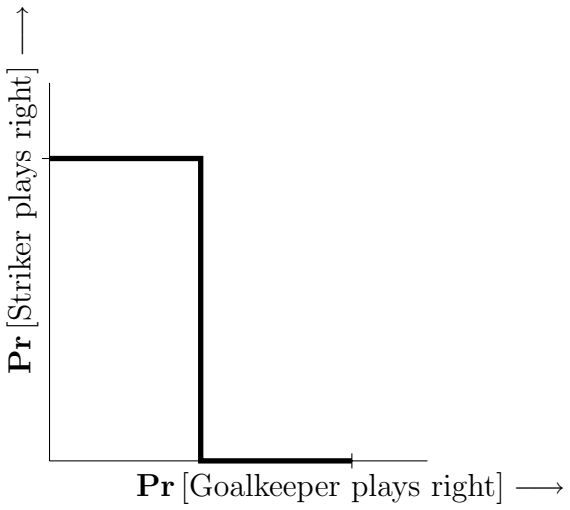


Figure 1: A figure showing the best response for Striker as a function of probability of Goalie going right.

changes from strategy $(L, R) = (0.5+\varepsilon, 0.5-\varepsilon)$ to strategy $(L, R) = (0.5-\varepsilon, 0.5+\varepsilon)$, the best response of the striker changes from $(L, R) = (1, 0)$ to $(L, R) = (0, 1)$. If we look at the limit $\varepsilon \to 0$, we deduce that the best response function is not continuous.

This example shows that the approach we were trying to take will not succeed.

So, we note that there is a problem if we define the function in the most natural way. Nash overcame this difficulty by defining a correspondence $F : X \to X$ which can map one point to many points. In other words, a correspondence is just a function

from X to 2^X. In topology, there is an analogue of Brouwer's Fixed Point Theorem which applies to correspondences; it is called Kakutani's Fixed Point Theorem and is considerably more obscure than Brouwer's Theorem. Using Kakutani's Fixed Point Theorem on the best-response correspondence, one can derive Theorem 1.

We will take an alternate route, and define f in a slightly different way from what we did above. Let

$$f(p_1, p_2, \ldots, p_k) = (q_1, q_2, \ldots, q_k) \quad \text{where } q_i \in \arg max_q \left(u_i(q, p_{-i}) - \|q - p_i\|_2^2\right).$$

We will use Brouwer's Fixed Point Theorem on this function to derive Theorem 1.

Proof. We need to prove two different things to apply Theorem 2 and derive Theorem 1.

f is well-defined and continuous. Let us first write the utility of player i resulting from a deviation.

$$\begin{aligned} u_i(q, p_{-i}) &= \sum_{a \in A_i} q(a) u_i(a, p_{-i}) \\ &= q \cdot v_i(p), \end{aligned}$$

where $v_i(p)$ is defined to be a vector whose a-th component is $u_i(a, p_{-i})$. We can now express the function in $\arg max$ in the definition above as

$$\begin{aligned} u_i(q, p_{-i}) - \|q - p_i\|_2^2 &= q \cdot v_i(p) - (q - p_i) \cdot (q - p_i) \\ &= -q \cdot q + q \cdot (v_i(p) + 2p_i) - p_i \cdot p_i \\ &= -\left\| q - \left(p_i + \frac{1}{2} v_i(p) \right) \right\|_2^2 + \text{constant}, \end{aligned}$$

where the constant is independent of q. Finding the maximum of the above expression is equivalent to finding the minimum of $\left\|q - \left(p_i + \frac{1}{2}v_i(p)\right)\right\|_2^2$. Since the set of all mixed strategy profiles q is a closed, bounded convex set, there exists a unique closest point in this set to the point $\left(p_i + \frac{1}{2}v_i(p)\right)$, by Theorem 3 below. This proves that the set $\arg\max_q(u_i(q, p_{-i}) - \|q - p_i\|_2^2)$ contains a unique point, i.e. f is a well-defined function. The continuity of f also follows from Theorem 3 below.

Every fixed point is a Nash equilibrium Let $(p_1, p_2, \ldots, p_k)$ be a fixed point of the function f, that is $f(p_1, p_2, \ldots, p_k) = (p_1, p_2, \ldots, p_k)$. Let $\hat{p}_i \neq p_i$ be any other mixed strategy of player i. Let us parametrize the line segment joining p_i to $\hat{p}_i$ by the function $q_i(t) = p_i + t(\hat{p}_i - p_i)$ for $t \in [0, 1]$. Since $q_i(0)$ maximizes the function $u_i(q_i(t), p_{-i}) - \|q_i(t) - p_i\|_2^2$ on the interval $0 \leq t \leq 1$, the derivative of this function at $t = 0$ should be nonpositive. We have

$$\frac{d}{dt}\left(u_i(q_i(t), p_{-i}) - \|q_i(t) - p_i\|_2^2\right)_{t=0} \leq 0.$$

Here, $-\|q_i(t)-p_i\|_2^2$ is just a quadratic function of t which achieves its global maximum at $t=0$, so its derivative at $t=0$ is zero. This can also be seen by

$$\|q_i(t)-p_i\|_2^2 = \|t(\hat{p}_i-p_i)\|_2^2 = \|(\hat{p}_i-p_i)\|_2^2 t^2,$$

and noting that its derivative at $t=0$ is $\|\hat{p}_i-p_i\|_2^2 \cdot 2t = 0$. The derivative of the first term can be written as

$$\begin{aligned}
\frac{d}{dt}u_i(q_i(t),p_{-i}) &= \frac{d}{dt}((p_i+t(\hat{p}_i-p_i))\cdot v_i(p)) \\
&= \frac{d}{dt}[p_i\cdot v_i(p)] + \frac{d}{dt}[t(\hat{p}_i-p_i)\cdot v_i(p)] \\
&= 0 + (\hat{p}_i-p_i)\cdot v_i(p) \qquad \text{(first term is independent of } t\text{)} \\
&= \hat{p}_i\cdot v_i(p) - p_i\cdot v_i(p) \\
&= u_i(\hat{p}_i,p_{-i}) - u_i(p_i,p_{-i}) \\
&\le 0 \qquad \text{(from the fact that derivative is at most zero)}
\end{aligned}$$

This shows that payoff for playing $\hat{p}_i$ is at most the payoff for playing p_i. Hence p_i is a best response for player i. So, $(p_1,p_2,\ldots,p_k)$ is a Nash equilibrium. □

To complete the proof of Theorem 1, we conclude with the following theorem which was used in deducing that f is well-defined and continuous.

Theorem 3. *If $A\subseteq\mathbb{R}^n$ is closed and convex, then for any $x\in\mathbb{R}^n$, there exists a unique point $y\in A$ closest to x. Moreover, y depends continuously on x.*

Proof. The function $d(z)=\|x-z\|_2$ is a continuous function of z, and A is a compact set, so the minimum value of d is achieved by at least one point y in A. First we will prove that this point y is unique. If $x\in A$ this is obvious, so assume that $x\notin A$. Translating, rotating, and rescaling if necessary, we may assume that x is at the origin and $y=(1,0,\ldots,0)$. (See Figure 2.) Let B denote the ball $\{z\in\mathbb{R}^n : \|z\|_2\le 1\}$. The assumption that y is the point of A closest to x implies that A does not intersect the interior of B. This, in turn, implies that A is contained in the halfspace consisting of all points of $\mathbb{R}^n$ whose first coordinate is greater than or equal to 1.[1] The distance from x to this halfspace is uniquely minimized at y, so the distance from x to A is also uniquely minimized at y.

It remains to show that y depends continuously on x. In other words, given $\varepsilon>0$ we must prove that there exists $\delta>0$ such that for all x' satisfying $\|x'-x\|_2<\delta$, if y' is the point of A closest to x', then $\|y'-y\|<\varepsilon$. So let x,y and x',y' be any pairs of points such that y is the point of A closest to x and y' is the point of A closest to x'.

[1] If $z=(z_1,\ldots,z_n)$ satisfies $z_1<1$ then for sufficiently small $t>0$, the convex combination $(1-t)y+tz$ is contained in the interior of B. Since A is disjoint from the interior of B, this implies that $(1-t)y+tz\notin A$ which in turn implies $z\notin A$.

(We allow the possibility that any two of these points may be equal.) If $\|x' - x\|_2 < \delta$ then

$$\begin{aligned}\|x - y'\|_2 &\leq \|x - x'\|_2 + \|x' - y'\|_2 \\ &\leq \|x - x'\|_2 + \|x' - y\|_2 \\ &\leq 2\|x' - x\|_2 + \|x - y\|_2 \\ &< 2\delta + \|x - y\|_2.\end{aligned}$$

Let $B(x, r)$ denote the (Euclidean) ball of radius r centered at x. We have seen that $y' \in B(x, r_0 + 2\delta)$, where $r_0 = \|x - y\|_2$. If we can prove that the diameter of $B(x, r_0 + 2\delta) \cap A$ tends to zero as $\delta \to 0$ then it follows that we may choose δ small enough that the diameter of $B(x, r_0 + 2\delta) \cap A$ is less than ε, from which it follows that $\|y' - y\|_2 < \varepsilon$ as desired.

To prove that the diameter of $B(x, r_0 + 2\delta) \cap A$ tends to zero with δ, we distinguish two cases. If $r_0 = 0$ then the diameter of $B(x, r_0 + 2\delta)$ is 4δ which tends to zero with δ. If $r_0 > 0$ then we may translate, rotate, and rescale, if necessary, so that $x = (0, \ldots, 0)$, $y = (1, 0, \ldots,)$, $r_0 = 1$. (The situation depicted in Figure 2, once again.) Recall that in this situation, A is contained in the halfspace H consisting of points $z \in \mathbb{R}^n$ whose first coordinate is greater than or equal to 1. So it suffices to prove that the diameter of $B(x, 1 + 2\delta) \cap H$ is less than ε, for δ sufficiently small. An easy calculation using the Pythagorean Theorem shows that for $\delta < \sqrt{1 + \varepsilon^2/4} - 1$, the set $B(x, 1 + 2\delta)$ has diameter less than ε. □

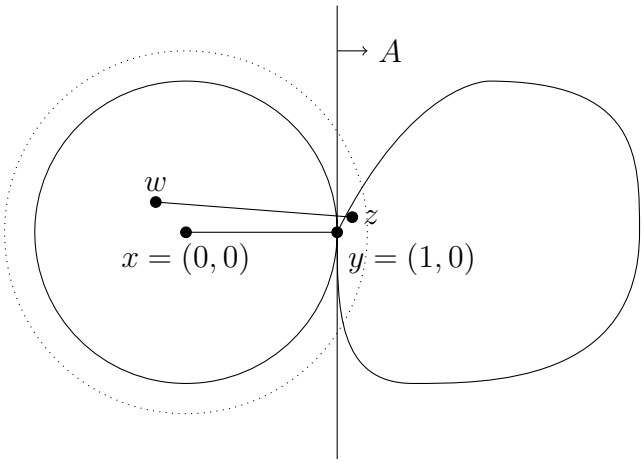


Figure 2: Illustration of the proof of Theorem 3

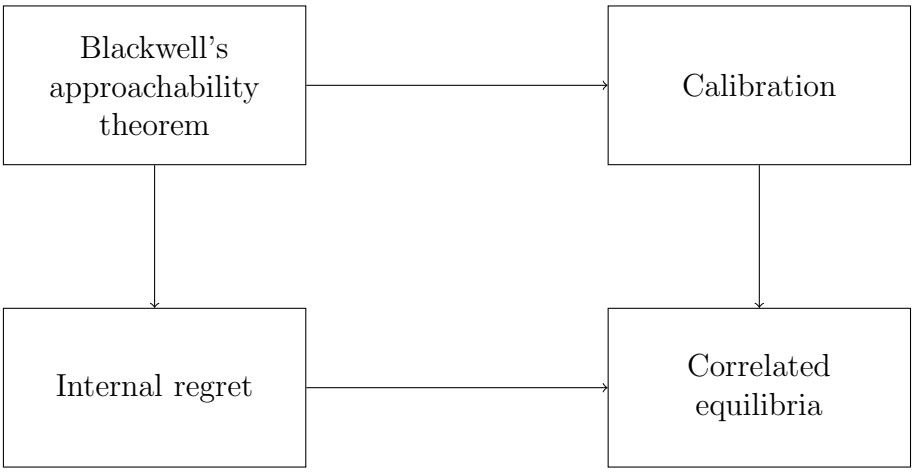
4 Plan for the next two weeks

Although Theorem 1 ensures that mixed strategy equilibria always exist, it turns out that the equilibrium concept most relevant to the theory of learning in games is a different and slightly subtler notion called *correlated equilibrium*, which generalizes Nash equilibrium. In the next two weeks, we will introduce correlated equilibrium and we will see two different ways that learning processes can converge into the set of correlated equilibria.

For two-player zero-sum games, we have seen that when players choose their mixed strategies using regret-minimizing learning algorithms, the time-averaged distribution of play converges into the set of Nash equilibria of the game. For non-zero-sum games, it turns out that the learners must satisfy a stronger condition called "no internal regret." We will define internal regret and show that when players choose their strategies using no-internal-regret learning algorithms, the time-averaged distribution of play converges into the set of correlated equilibria. The existence of learning algorithms with no internal regret will be derived from a general theorem called Blackwell's approachability theorem that concerns online algorithms for playing *games with vector payoffs*, a notion which we introduce below.

A second adaptive procedure converging to the set of correlated equilibria involves the notion of *calibrated forecasting*. This theory introduces algorithms for computing forecasts, i.e. predictions of the probabilities of future events. We will define a property of forecasting algorithms called "calibration" and we will use Blackwell's approachability theorem to prove that calibrated forecasting algorithms exist. Moreover, if players predict their opponents' strategies using a calibrated forecasting algorithm, and they play a best response to the predicted strategy profile of their opponents, we will see that the time-averaged distribution of play converges to the set of correlated equilibria.

The following diagram summarizes these two derivations of adaptive procedures which lead to correlated equilibrium.



5 Introduction to correlated equilibrium

We have been looking at models of rationality in game theory, and the concept of Nash equilibrium is a natural one. It also has a nice property: for normal-form games, a Nash equilibrium always exists. But there are some computational issues with this concept, since it is believed that computing a Nash equilibrium is computationally hard. To circumvent this problem, we move to another natural concept of equilibrium, called correlated equilibrium. While the definition of correlated equilibrium appears slightly less natural as a model of rational behavior, we will see that the algorithms which converge to correlated equilibria are actually much *more* natural than any known algorithms for players to converge to a Nash equilibrium.

5.1 Definition of correlated equlibrium

Let $\mathscr{G} = (\mathcal{I}, A_i, u_i)$ be a normal-form game with $|\mathcal{I}|, |A_i| < \infty$. A probability distribution p on $\prod_{j\in\mathcal{I}} A_j$ is a correlated equilibrium if for all $i \in \mathcal{I}$ and for all $\alpha, \beta \in A_i$ such that the probability of i playing α is positive, we have

$$\mathbf{E}\left[u_i(\alpha, a_{-i})|a_i = \alpha\right] \geq \mathbf{E}\left[u_i(\beta, a_{-i})|a_i = \alpha\right], \tag{1}$$

where a_i denote the i-th component of profile a, and a_{-i} denotes all compnents except the i-th one. This can also be rewritten as

$$\sum_{a_{-i}} (u_i(\alpha, a_{-i}) - u_i(\beta, a_{-i}))\mathbf{Pr}\left[\alpha, a_{-i}\right] \geq 0. \tag{2}$$

Inequality (2) is derived from (1) by multiplying the left and right sides by $\mathbf{Pr}\left[a_i = \alpha\right]$.

5.2 A story for correlated equilibrium

The story for correlated equilibrium goes as follows. In the game play, assume that there is a trusted random device who samples a pure strategy profile from the distribution dictated by p, and tells each player his/her component of the strategy profile. If all players other than i are following the strategy suggested by the random sampling device, than i does not have any incentive not to play what is suggested to her. The expected payoff of playing the suggested strategy is at least as large as the expected payoff of playing any other strategy.

Example 2 (Traffic light game). Consider the following game in which two players are driving and approaching an intersection from perpendicular directions. Each one has the options to stop or to go. The payoffs are shown in Table 2.
In this example, there are three Nash equilibria. Two of them are pure, and one is mixed. They are shown below.

0	1
0	0

0	0
1	0

1/4	1/4
1/4	1/4

		Driver 2	
		Stop	Go
Driver 1	Stop	$(4,4)$	$(1,5)$
	Go	$(5,1)$	$(0,0)$

Table 2: Payoffs for the traffic light game

Some examples of correlated equlibria are as follows.

0	1/2
1/2	0

1/3	1/3
1/3	0

The first correlated equilibrium shown (which randomizes between "driver 1 goes, driver 2 stops" and "driver 2 stops, driver 1 goes" with equal probability) resembles the equilibrium which would be implemented by installing a traffic light at the intersection which always tells one driver to go while telling the other one to stop. Assuming the drivers believe that the traffic light operates according to this specification, they have no incentive to disobey the advice of the traffic light.

Remark 1. The set of all correlated equilibria of a game is a convex set, while the set of all Nash equilibria is not. For example, in the traffic light game, we have seen that the set of Nash equilibria is a set of three isolated points. The set of correlated equilibria is convex simply because the inequalities defined by (2) — as i ranges over $\mathcal{I}$ and α, β range over A_i — constitute a system of linear inequalities. The constraint that p is a probability distribution on $\prod_{j\in\mathcal{I}} A_j$ is also expressed by a system of linear equations and inequalities stipulating that all probabilities are non-negative and their sum is equal to 1. The solution set of any system of linear equations and inequalities is a convex set.

6 Internal regret

We first define the notion of internal regret. In this section, a vector of size T will be denoted by drawing an arrow over it.

Definition 1 (Internal regret). Let A be a set of actions, $\vec{a} = a_1, a_2, \ldots, a_T$ be a sequence of choices of elements of A, and $g_1, g_2, \ldots, g_T : A \to \mathbb{R}$ be the set of payoff functions. For a function $f : A \to A$, we define

$$\hat{R}_f(\vec{a}, \vec{g}; T) = \frac{1}{T} \sum_{t=1}^{T} \left(g_t(f(a_t)) - g_t(a_t) \right). \tag{3}$$

Think of f as an *advisor* function, which gives an alternative strategy to play for any chosen strategy. Then the quantity $\hat{R}_f(\vec{a}, \vec{g}; T)$ can be thought of the regret the

algorithm has for not following the advice of this advisor function f. We now define the *internal regret* of the sequence of actions $\vec{a}$ by

$$\hat{R}_{\text{int}}(\vec{a}, \vec{g}; T) = \max_f \hat{R}_f(\vec{a}, \vec{g}; T). \tag{4}$$

It is also helpful to define a special case of Equation (3) which we will call *pairwise regret*. If $a, b \in A$ then

$$\hat{R}_{a,b}(\vec{a}, \vec{g}; T) = \frac{1}{T} \sum_{t=1}^{T} (g_t(b) - g_t(a)) \, \mathbf{1}[\vec{a}(t) = a]. \tag{5}$$

Here $\mathbf{1}[\vec{a}(t) = a]$ denote the indicator variable for the event that the t-th component of $\vec{a}$ is a. Formally,

$$\mathbf{1}[\vec{a}(t) = a] = \begin{cases} 1 & \text{if } t\text{-th component of } \vec{a} \text{ is } a \\ 0 & \text{otherwise.} \end{cases}$$

Pairwise regret is a special case of internal regret. This can be seen by defining the function f as follows.

$$f(c) = \begin{cases} b & \text{if } c = a \\ c & \text{otherwise.} \end{cases} \tag{6}$$

Intuitively, the pairwise regret $\hat{R}_{a,b}$ can be thought of as the algorithm's regret for not playing b whenever it played a.

We now define the notion of an algorithm having no internal regret.

Definition 2 (No internal regret). A randomized algorithm has *no internal regret* if for all adaptive adversaries given by payoff functions $\vec{g} = (g_1, g_2, \ldots, g_T)$, the algorithm computes $\vec{a} = (a_1, a_2, \ldots, a_T)$ such that

$$\lim_{T \to \infty} \mathbf{E}\left[\hat{R}_{\text{int}}(\vec{a}, \vec{g}; T)\right] = 0. \tag{7}$$

Remark 2. The term "no internal regret" is misleading. If an algorithm has no internal regret, it may mean that the expected value of $\hat{R}_{\text{int}}$ is positive at all finite times T; the definition of no-internal-regret algorithms only stipulates that this expected value must converge to zero as T tends to infinity. In other words, no-internal-regret algorithms are allowed to have internal regret, as long as the time-average of the internal regret tends to zero.

The following lemma and corollary give an easy criterion for detecting whether an algorithm has no internal regret by checking only the pairwise regret terms.

Lemma 4. *$\hat{R}_{int}(\vec{a}, \vec{g}; T)$ can be bounded as*

$$\max_{a,b} \hat{R}_{a,b}(\vec{a}, \vec{g}; T) \leq \hat{R}_{int}(\vec{a}, \vec{g}; T) \leq |A| \cdot \max_{a,b} \hat{R}_{a,b}(\vec{a}, \vec{g}; T).$$

Proof. The first inequality is an easy consequence of the fact that $\hat{R}_{a,b}(\vec{a}, \vec{g}; T)$ is a special case of $\hat{R}_f(\vec{a}, \vec{g}; T)$ when f is defined as above (in Equation (6)). This gives $\hat{R}_{a,b}(\vec{a}, \vec{g}; T) \leq \hat{R}_f(\vec{a}, \vec{g}; T)$ for all a, b.

The second inquality can be seen from the following.

$$\begin{aligned}
\hat{R}_f(\vec{a}, \vec{g}; T) &= \frac{1}{T} \sum_{t=1}^{T} (g_t(f(a_t)) - g_t(a_t)) \\
&= \frac{1}{T} \sum_{t=1}^{T} \sum_{a \in A} (g_t(f(a)) - g_t(a)) \mathbf{1}[a_t = a] \\
&= \sum_{a \in A} \frac{1}{T} \sum_{t=1}^{T} (g_t(f(a)) - g_t(a)) \mathbf{1}[a_t = a] \\
&= \sum_{a \in A} \hat{R}_{a,f(a)}(\vec{a}, \vec{g}; T) \\
&\leq |A| \cdot \max_{a,b} \hat{R}_{a,b}(\vec{a}, \vec{g}; T).
\end{aligned}$$

Here the third equality follows by changing the order of summation. □

Corollary 5. *An algorithm has no internal regret if and only if*

$$\lim_{T \to \infty} \left(\max_{a,b} \mathbf{E} \left[\hat{R}_{a,b}(\vec{a}, \vec{g}; T) \right] \right) = 0.$$

6.1 Internal regret and correlated equilibrium

We next see how players choosing strategies according to a no-regret algorithm can converge to the set of correlated equilibria.

Definition 3. If S is a subset of a finite-dimensional real vector space $\mathbb{R}^N$, and x is a point of $\mathbb{R}^N$, let

$$\operatorname{dist}(x, S) = \inf_{s \in S} \|x - s\|_2.$$

We will say that an infinite sequence $x_1, x_2, \ldots$ *converges to* S if $\lim_{n \to \infty} \operatorname{dist}(x_n, S) = 0$.

Theorem 6. *Let $\mathscr{G}$ be a normal-form game with $k < \infty$ players, and with finitely many strategies for each player. Suppose the players play $\mathscr{G}$ repeatedly, and that player i chooses its sequence of strategies $(a_i^t)_{t=1}^{\infty}$ using a no-regret learning algorithm with payoffs $g_i^t(a) = u_i(a, a_{-i}^t)$. Let $\mathcal{C}$ be the set of correlated equilibria of $\mathscr{G}$, and let $\bar{p}(T)$ be the uniform distribution on the multiset $\{(a_1^t, a_2^t, \ldots, a_k^t) : 1 \leq t \leq T\}$. Then the sequence $(\bar{p}(T))_{T=1}^{\infty}$ converges to $\mathcal{C}$.*

Proof. Before we begin the formal proof of the theorem, here's a sketch of the basic idea. If α, β are two strategies of player i, the pairwise regret $\hat{R}_{\alpha,\beta}$ at time T measures the average gain in payoff that player i would have obtained during the first T plays of the game, if it had used strategy β every time that it played α in the actual history of play. The assumption that players use no-regret learning algorithms to choose their strategies means that this average gain $\hat{R}_{\alpha,\beta}$ becomes negligible as $T \to \infty$. Now suppose that a third party samples a strategy profile $(a_1, \ldots, a_k)$ from distribution $\bar{p}(T)$ and suggests strategy a_i to player i. One can interpret $\hat{R}_{\alpha,\beta}$ as the expected amount that player i would gain if all other players followed the suggested strategy and player i adopted the policy of obeying the suggestion unless α is suggested, and playing β in that case. If we could say "player i gains nothing by following this policy" instead of "player i gains a negligible amount," this would be precisely the definition of correlated equilibrium. So it is reasonable to expect that no-regret learners will converge to the set of correlated equilibria.

To prove the theorem more formally, we will argue by contradiction. If the sequence does not converge to $\mathcal{C}$, then for some $\delta > 0$, there is an infinite subsequence which is contained in the set of probability distributions whose distance from $\mathcal{C}$ is greater than or equal to δ. This is a compact subset of the space of all probability distributions on $\prod_{j \in \mathcal{I}} A_j$, so there is an infinite subsequence which converges to some distribution p such that $\text{dist}(p, \mathcal{C}) \geq \delta$. Denote this subsequence by $\bar{p}(T_1), \bar{p}(T_2), \ldots$. The fact that $p \notin \mathcal{C}$ means that for some $\varepsilon > 0$, there exists a player i and two strategies $\alpha, \beta \in A_i$ such that

$$\sum_{a_{-i} \in \prod_{j \neq i} A_j} [u_i(\beta, a_{-i}) - u_i(\alpha, a_{-i})] p(\alpha, a_{-i}) = \varepsilon.$$

This means that whenever the trusted third party (in the definition of correlated equilibrium) suggested that player i should play α, she has the incentive to instead play β. Since p is a limit point of $\bar{p}(T_s)$, for sufficiently large s, we have that

$$\sum_{a_{-i} \in \prod_{j \neq i} A_j} [u_i(\beta, a_{-i}) - u_i(\alpha, a_{-i})] \bar{p}(T_s)(\alpha, a_{-i}) \geq \frac{\varepsilon}{2}.$$

Since $\bar{p}(T_s)(a)$ is equal to the number of times a was played in first T_s timesteps divided by T_s, we can write the above equation as

$$\frac{1}{T_s} \sum_{a_{-i} \in \prod_{j \neq i} A_j} [u_i(\beta, a_{-i}) - u_i(\alpha, a_{-i})] \left(\sum_{t=1}^{T_s} \mathbf{1}[a^t = a] \right) \geq \frac{\varepsilon}{2}.$$

By changing the order of summation, and using the fact that

$$\mathbf{1}[a^t = a] = \mathbf{1}[a^t_{-i} = a_{-i}] \cdot \mathbf{1}[a^t_i = a_i],$$

we have the following

$$\frac{1}{T_s}\sum_{t=1}^{T_s}\sum_{a_{-i}\in\prod_{j\neq i}A_j}[u_i(\beta,a_{-i})-u_i(\alpha,a_{-i})]\mathbf{1}[a_{-i}^t=a_{-i}]\cdot\mathbf{1}[a_i^t=\alpha]\geq\frac{\varepsilon}{2}$$

which can be further simplified to

$$\frac{1}{T_s}\sum_{t=1}^{T_s}(u_i(\beta,a_{-i}^t)-u_i(\alpha,a_{-i}^t))\mathbf{1}[a_i^t=\alpha]\geq\frac{\varepsilon}{2}. \tag{8}$$

But this term is just the pairwise regret $\hat{R}_{\alpha,\beta}$ for player i at time T_s, i.e. (8) can be rewritten as

$$\hat{R}_{\alpha,\beta}((a_i^1,a_i^2,\ldots,a_i^{T_s}),(g_i^1,g_i^2,\ldots,g_i^{T_s});T_s)\geq\frac{\varepsilon}{2}.$$

But this is a contradiction since player i is using a no-regret learning algorithm. This proves the theorem. □

7 Normal form games with vector payoffs

Definition 4 (Normal form games with vector payoffs). A normal-form game with vector payoffs is a game defined by $\vec{\mathscr{G}}=(\mathcal{I},A_i,\vec{u}_i)$ where

- $|\mathcal{I}|<\infty$,
- A_i is the set of strategies for player $i\in\mathcal{I}$, and
- $\vec{u}_i:\prod_{j\in\mathcal{I}}A_j\to\mathbb{R}^n$, i.e. payoffs are vectors in $\mathbb{R}^n$.

All of the vector-payoff games we will consider have two players. Therefore, unless stated otherwise, we will always assume that $\mathcal{I}=\{1,2\}$. Also, we will usually analyze these games from the perspective of player 1, and when doing so we will abbreviate $\vec{u}_1$ to $\vec{u}$.

7.1 The meaning of vector payoffs

At the first sight, it seems as if the notion of games with vector payoffs is contrived. If we conceptualize the "payoff" of player i as denoting how happy player i is with the outcome of the game, it seems much more natural to model this parameter as a scalar rather than a vector. But if we give it a little more thought, we see that games with vector payoffs are a convenient abstraction for *multi-objective optimization*, just as games with scalar payoffs were a useful abstraction for single-objective optimization.

Example 3 (Presidential elections.). Let us consider the example of presidential elections. In the primary elections, a candidate must get enough votes from the members of his or her own party to win that party's nomination. We then proceed to the general election, in which all voters participate and their opinion of the candidate is at least partly influenced by the candidate's strategy during the primaries. This can be modeled as a game with a two-dimensional vector payoff, where the first component of the vector represents the number of votes cast for the candidate in the primary election, and the second component represents the number of votes cast for the candidate in the general election. (Actually, there are primaries in all 50 states, and the electoral college system implies that the outcome of the general election also depends on the vote tallies in all 50 states and not just on the combined number of votes nationwide. Hence it may be even more sensible to model the game as having 100-dimensional vector payoffs.)

A skeptical reader could argue that Example 3 is really missing the point: in the end, a candidate's choice of strategy in the election will be guided by the sole objective of maximizing a single scalar value, namely the probability of winning the presidency. But regardless of whether games with vector payoffs have anything meaningful to say about the politics of presidential elections, it remains the case that they are a powerful abstraction for expressing optimization problems with multiple objectives. We will see several examples of the power of this abstraction in the upcoming lectures.

7.2 A note about adversaries and the dynamics of games

In the first week of class, we defined online algorithms and adversaries according to a paradigm in which the adversary and algorithm interact via an alternating sequence of decisions. First an adversary specifies an input at time 1, then the algorithm chooses an output at time 1, then the adversary specifies the input at time 2, and so on.

In the preceding lecture, and in the following ones, we are thinking about simultaneous move games. In this dynamics of the game, at time 1, the adversary and the algorithm both commit to their actions without knowing the other's choice. The actions of both parties are revealed simultaneously, and then the second time step begins.

All the different kinds of adversaries we have seen make sense in the simultaneous-move setting. The distinction between deterministic and randomized adversary is that the former one does not use any randomness while choosing its strategy for time step t, but it can of course use information about the strategies chosen by the algorithm (and itself) up to time step $t-1$. The distinction between an adaptive adversary and an oblivious adversary is that the former can make use of strategies of the algorithm and itself up to time $t-1$ in choosing its strategy in time period t, while later's strategy does not depend on the history (it is oblivious to the game play until the current time step).

8 Approachability

Definition 5 (Approachable sets). Let $\vec{\mathscr{G}}$ be a vector payoff game such that the range of $\vec{u}_i$ is a bounded subset of $\mathbb{R}^n$. A set $S \subseteq \mathbb{R}^n$ is called *approachable* (by player 1) if there exists a randomized algorithm to choose $i_1, i_2, \ldots \in A_1$ such that for all adaptive adversaries choosing $j_1, j_2, \ldots \in A_2$ we have that

$$\text{dist}\left(\frac{1}{T}\sum_{t=1}^{T}\vec{u}(i_t, j_t), S\right) \to 0 \qquad \text{as} \qquad T \to \infty.$$

In other words, a set is called approachable if no adversary can prevent player 1 from choosing strategies such that its average payoff vector converges to the set S as the time horizon, T, tends to infinity.

8.1 Approachability of halfspaces

To get a feel of what it means for a set to be approachable, let us consider the question of approachability of halfspaces. When is the set $S = \{\vec{u} : a^\top \vec{u} \geq b\}$ approachable?

Let us consider the "scalar payoff" zero-sum normal-form game $a^\top\vec{\mathscr{G}}$ which is the same as $\vec{\mathscr{G}}$ except that the payoff function is $a^\top\vec{u}$ instead of $\vec{u}$.

We now want to make the following claim.

Claim 7 (Half-space approachability). *The half-space S is approachable if and only if the value of game $a^\top\vec{\mathscr{G}}$ is at least b.*

Proof. The condition of S being approachable by player i means that the average payoff of player i in the game $\vec{a}^\top\vec{\mathscr{G}}$ is at least $b - o(1)$ at time horizon becomes large. Using von Neumann's Theorem, we can say that there exists a mixed strategy for player 1 such that its expected payoff is at least b if and only if the game value is at least b. To derive Claim 7 from von Neumann's Theorem, we need an additional step to ensure that the average of the payoffs actually received by player 1 converge, with probability 1, to its expected value as the time horizon tends to infinity. We will use Azuma's inequality to get this result. See Appendix A for background information on martingales and Azuma's inequality.

Since we are assuming that the payoffs vectors of $\vec{\mathscr{G}}$ belong to a bounded subset of $\mathbb{R}^n$, we may assume without loss of generality (rescaling the vector a and the scalar b, if necessary) that $-1 \leq a^\top\vec{u}(i,j) \leq 1$ for all payoff vectors $\vec{u}(i,j)$.

First let us assume that the value of the game is at least b. By von Neumann's Theorem, we know that there exists a mixed strategy for player 1 whose expected payoff is at least b. Let player 1 choose strategies $i_1, i_2, \ldots$ by independently sampling from this mixed strategy.

Let us define random variables $X_t = \sum_{s=1}^{t}[a^\top\vec{u}(i_s, j_s) - b]$. Since the expected value of $X_{t+1} - X_t$ is at least the game value minus b, it is at least 0. Therefore, the

sequence $X_0, X_1, \ldots, X_t$ forms a submartingale, as $\mathbf{E}[X_{t+1}|X_t, X_{t-1}, \ldots, X_1] \geq X_t$. (The payoffs are bounded by 1, so the condition of bounded difference is satisfied with $c_k = 1$.) Using Azuma's inequality,

$$\begin{aligned}\mathbf{Pr}\left[X_t \leq -\lambda\sqrt{t}\right] &\leq \mathbf{Pr}\left[X_t \leq \mathbf{E}[X_0] - \lambda\sqrt{t}\right] \qquad (\text{since } \mathbf{E}[X_0] = 0)\\ &\leq e^{-\frac{\lambda^2}{2}}.\end{aligned}$$

Take $\lambda = 2\sqrt{\log t}$, this gives

$$\begin{aligned}\mathbf{Pr}\left[X_t \leq -2\sqrt{t\log t}\right] &\leq e^{\frac{4\log t}{2}}\\ &= 1/t^2.\end{aligned}$$

Therefore, with high probability, $X_t \geq -2\sqrt{t\log t}$, which means that

$$\frac{1}{t}\sum_{s=1}^{t} a^\top \vec{u}(i_s, j_s) \geq b - 2\sqrt{\frac{\log t}{t}}.$$

The right side tends to b as $t \to \infty$. This shows that if the game value is at least b, then the halfspace $a^\top \vec{u} \geq b$ is approachable. The converse of this statement—that the half space is not approachable if game value is less than b—is also true, and can be proved by applying Azuma's inequality for supermartingales. This is left as an exercise for the reader. □

8.2 Blackwell's approachability theorem

In this section, we state the approachability theorem of Blackwell. We will defer the proof to next week.

Theorem 8 (Blackwell's Approachability Theorem). *A closed convex set S is approachable if and only if every halfspace containing S is approachable.*

Note that one direction in the proof is trivial. If S is approachable, then any halfspace containing S is trivially approachable. If the distance of average payoff in the long run comes arbitrarily close to S, it also comes arbitrarily close to the halfspace containing S. The surprising fact is that the converse is also true.

Remark 3. Continuing with our philosophy that games with vector payoffs are a metaphor for reasoning about multi-objective optimization, one can interpret Blackwell's approachability theorem as *a reduction from multi-objective optimization to single-objective optimization*, in the online setting. The precise form of this reduction is rather delicate (for example, the theorem says nothing about what happens if S is not convex) but the informal statement gives the correct intuition for how Blackwell's approachability theorem is generally applied.

8.3 An application of Blackwell's approachability theorem

Using Blackwell's approachability theorem, we will prove that there exists an algorithm for the best-expert problem described in the first week of class, whose average regret converges to 0 at $T \to \infty$.

In this game, let player 1 denote the algorithm, whose strategy set A_1 is $[n]$. (Strategy i corresponds to choosing expert i.) Player 2 can be thought of as the adversary who can determine the reward for each of the experts, that is $A_2 = \{g : [n] \to [0, 1]\}$. Let the payoff of player 1 be $\vec{u}_1(i, g)$ which is defined as follows.

$$\vec{u}(i, g) = \begin{pmatrix} g(1) - g(i) \\ g(2) - g(i) \\ \vdots \\ g(n) - g(i) \end{pmatrix}$$

Let us define *no external regret* analogously to the definition given earlier for the "no internal regret" property. In other words, a no-external-regret learning algorithm is one whose regret tends to zero as $T \to \infty$, i.e. for any adaptive adversary choosing a sequence of reward functions $g_1, g_2, \ldots$,

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbf{E}\left[\vec{u}(i_t, g_t)\right] \in S = \{(x_1, x_2, \ldots, x_n) : x_i \leq 0 \quad \text{for all } i\}.$$

This means that there exists a no-regret learning algorithm if and only if S is approachable.

We can easily show using Blackwell's approachability theorem that S is indeed approachable. We know that S is approachable if and only if every halfspace $H \supseteq S$ is approachable. For any halfspace H containing S, there is a nonzero vector a and a scalar b such that $a_i \geq 0$ for all i, $b \geq 0$, and $H = \{u : a^\top u \leq b\}$. We may assume without loss of generality that $\sum_{i=1}^{n} a_i = 1$. (If not, multiply both a and b by the scalar $\left(\sum_{i=1}^{n} a_i\right)^{-1}$. This doesn't change the solution set of $a^\top u \leq b$.) There is a simple algorithm that approaches this halfspace. First note that it suffices to show that $\{u : a^\top u \leq 0\}$ is approachable. This is because the latter halfspace is a subset of the earlier one, and if we can approach a subset then we can approach the whole set too. From Claim 7 it follows that $\{u : a^\top u \leq 0\}$ is approachable if and only if the value of the game $a^\top \vec{\mathscr{G}}$ is at most zero.

There is a simple strategy by which player player 1 can ensure nonnegative expected payoff in the game $a^\top \vec{\mathscr{G}}$. Let us say player 1 chooses strategy i with probability a_i In the game $a^\top \vec{\mathscr{G}}$, the payoff for player 1 is $a^\top \vec{u}(i, g) = a_1(g(1) - g(i)) + a_2(g(2) - g(i)) + \cdots + a_n(g(n) - g(i))$. So, the expected payoff for playing mixed strategy $(a_1, a_2, \ldots, a_n)$ can be expressed as

$$\mathbf{E}\left[a^\top \vec{u}(a, g)\right] = \sum_{i=1}^{n} a_i \cdot (a^\top \vec{u}(i, g))$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} a_i \sum_{j=1}^{n} a_j (g(j) - g(i)) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j (g(j) - g(i)) \\
&= 0.
\end{aligned}
$$

Therefore the expected payoff in each time step is zero, so by Claim 7, the halfspace H is approachable. As H was a generic halfspace containing S, we conclude that S is approachable. From the discussion above, we see that this proves there exists a no-regret learning algorithm for the best-expert problem with n experts.

A Martingales

Definition 6 (Martingale). A sequence of random variables $X_0, X_1, \ldots, X_n$ is a *martingale* if for all $k \in \{1, 2, \ldots, n\}$ the expected value of X_k given $X_0, X_1, \ldots, X_{k-1}$ is equal to X_{k-1}. That is

$$\mathbf{E}\left[X_k | X_{k-1}, X_{k-2}, \ldots, X_1, X_0\right] = X_{k-1} \quad \text{for all } k. \tag{Martingale}$$

Such a sequence is called a supermartingale if

$$\mathbf{E}\left[X_k | X_{k-1}, X_{k-2}, \ldots, X_1, X_0\right] \leq X_{k-1} \quad \text{for all } k, \tag{Supermartingale}$$

and a submartingale if

$$\mathbf{E}\left[X_k | X_{k-1}, X_{k-2}, \ldots, X_1, X_0\right] \geq X_{k-1} \quad \text{for all } k. \tag{Submartingale}$$

We make a straightforward observation about Martingales.

Lemma 9. *For a martingale sequence $X_0, X_1, \ldots, X_n$,*

$$\mathbf{E}\left[X_n\right] = \mathbf{E}\left[X_{n-1}\right] = \cdots = \mathbf{E}\left[X_0\right].$$

For a supermartingale, $\mathbf{E}\left[X_n\right] \leq \mathbf{E}\left[X_{n-1}\right] \leq \cdots \leq \mathbf{E}\left[X_0\right]$ and for a submartingale $\mathbf{E}\left[X_n\right] \geq \mathbf{E}\left[X_{n-1}\right] \geq \cdots \geq \mathbf{E}\left[X_0\right]$.

Proof. For each $k \in [n]$, we have

$$\mathbf{E}\left[X_k | X_{k-1}, \ldots, X_0\right] = X_{k-1}.$$

Taking expectation on both sides (over random choices of $X_{k-1}, \ldots, X_0$), we have

$$\mathbf{E}\left[\mathbf{E}\left[X_k | X_{k-1}, \ldots, X_0\right]\right] = \mathbf{E}\left[X_k\right] = \mathbf{E}\left[X_{k-1}\right].$$

The lemma now follows by induction. For submartingale and supermartingale, the extension is straightforward. □

An example of martingale that cannot be written as a sum of independent random variables is the following. The sequence of random variables $X_0, X_1, X_2, \ldots$ is defined as follows. X_0 is equal to 0. To get X_{i+1}, we add $+1$ or -1 to X_i with probability $\frac{1}{2}$ each. This continues until $X_n \in \{+100, -100\}$. Once the absolute value of the random variable becomes 100, the process stops. It is clear that $\mathbf{E}[X_k | X_{k-1}, \ldots, X_0] = X_{k-1}$ but because of the upper bound of 100, this cannot be written as sum of independent random variables.

We now proceed to state a very important inequality about martingales due to Kazuoki Azuma. Like the famous Chernoff bound for sums of independent random variables, this is an *exponential tail inequality*: it asserts that the probability of a certain random variable deviating from its expected value by a certain amount is exponentially small in the amount of the deviation.

Theorem 10 (Azuma's inequality). *If $X_0, X_1, X_2, \ldots, X_n$ is a supermartingale sequence, and $|X_k - X_{k-1}| \leq c_k$ for all k, then*

$$\mathbf{Pr}[X_n > X_0 + \lambda] < e^{-\frac{1}{2}\lambda^2 / \left(\sum_{i=1}^n c_i^2\right)},$$

for all $\lambda > 0$. Letting $\sigma = \sqrt{\sum_{i=1}^n c_i^2}$, the conclusion may be rewritten as

$$\mathbf{Pr}[X_n > X_0 + s\sigma] < e^{-\frac{1}{2}s^2}$$

for all $s > 0$. For a submartingale with σ defined as above,

$$\mathbf{Pr}[X_n < X_0 - s\sigma] < e^{-\frac{1}{2}s^2}.$$

For a martingale with σ defined as above,

$$\mathbf{Pr}[|X_n - X_0| > s\sigma] < 2e^{-\frac{1}{2}s^2}.$$

Proof. The bound for submartingales follows from the one for supermartingales by replacing the sequence $X_0, X_1, \ldots, X_n$ with the sequence $-X_0, -X_1, \ldots, -X_n$. The bound for martingales follows from the other two, using the union bound. Accordingly, for the remainder of the proof we will only work on establishing the bound for supermartingales. We will also assume $X_0 = 0$. This assumption is without loss of generality, since we can always replace the sequence $X_0, X_1, \ldots, X_n$ with $0, X_1 - X_0, X_2 - X_0, \ldots, X_n - X_0$ and this does not affect any of the hypotheses or conclusions of the theorem.

The idea behind the proof of Azuma's inequality is similar to the idea behind the proof of Chernoff's bound: we will compute the expectation of the random variable e^{tX_n} (for a parameter t) and then use Markov's inequality for that random variable to get the result. First let us define some terms. We define $Y_k = X_k - X_{k-1}$ for $k = 1, 2, \ldots, n$. Since the sequence of X's forms a supermartingale, it is clear that $\mathbf{E}[Y_k | X_{k-1}, \ldots, X_0] \leq 0$ and $|Y_k| \leq c_k$. We first claim that $\mathbf{E}\left[e^{tY_k} | X_{k-1}, X_{k-2}, \ldots, X_0\right] \leq e^{t^2 c_k^2/2}$.

Claim 11. *For a random variable Y with $\mathbf{E}[Y] \leq 0$ and $|Y| \leq c$, $\mathbf{E}\left[e^{tY}\right] \leq e^{\frac{t^2c^2}{2}}$.*

Proof of Claim 11. We acknowledge Ymir Vigfusson for supplying the short proof given here. For any real number p and $|X| \leq 1$,

$$\begin{aligned}
e^{pX} &= \sum_{n=0}^{\infty} \frac{(pX)^n}{n!} \\
&= \sum_{n=0}^{\infty} \frac{(pX)^{2n}}{(2n)!} + \sum_{n=0}^{\infty} \frac{(pX)^{2n+1}}{(2n+1)!} \\
&\leq \sum_{n=0}^{\infty} \frac{p^{2n}X^{2n}}{2^n n!} + X \cdot \sum_{n=0}^{\infty} \frac{p^{2n+1}X^{2n}}{(2n+1)!} \\
&\leq \sum_{n=0}^{\infty} \frac{(p^2/2)^n}{n!} + X \cdot \sum_{n=0}^{\infty} \frac{p^{2n+1}}{(2n+1)!} \\
&= e^{\frac{p^2}{2}} + X \cdot \sinh(p)
\end{aligned}$$

where the third line was derived from the second by observing that $2^n n!$ is the product of all positive even integers up to $2n$, while $(2n)!$ is the product of *all* positive integers up to $2n$. By plugging in $p = tc$ and $X = Y/c$, we conclude that

$$\mathbf{E}\left[e^{tY}\right] \leq e^{\frac{t^2c^2}{2}} + \mathbf{E}[Y] \frac{\sinh(tc)}{c} \leq e^{\frac{t^2c^2}{2}}.$$

□

We continue with the proof of Azuma's inequality for supermartingales.

$$\begin{aligned}
\mathbf{E}\left[e^{tX_n}\right] &= \mathbf{E}\left[e^{t(Y_n + X_{n-1})}\right] \\
&= \mathbf{E}\left[e^{tX_{n-1}}\right] \mathbf{E}\left[e^{tY_n} \mid X_{n-1}\right] && \text{(because } \mathbf{E}[YX] = \mathbf{E}[Y]\,\mathbf{E}[Y|X]\text{)} \\
&\leq e^{t^2c_n^2/2} \mathbf{E}\left[e^{tX_{n-1}}\right] && \text{(by Claim 11)}
\end{aligned}$$

By induction, we obtain

$$\begin{aligned}
\mathbf{E}\left[e^{tX_n}\right] &\leq \left(\prod_{i=1}^{n} e^{t^2c_i^2/2}\right) \mathbf{E}\left[e^{tX_0}\right] \\
&= \prod_{i=1}^{n} e^{t^2c_i^2/2} \\
&= e^{t^2\sigma^2/2}.
\end{aligned}$$

where the third line follows from our assumption that $X_0 = 0$.

Finally,

$$\begin{aligned}
\mathbf{Pr}\left[X_n \geq X_0 + \lambda\right] &= \mathbf{Pr}\left[X_n \geq \lambda\right] \qquad \text{(because } X_0 = 0\text{)} \\
&= \mathbf{Pr}\left[e^{tX_n} \geq e^{t\lambda}\right] \\
&\leq \mathbf{E}\left[e^{tX_n}\right] / e^{t\lambda} \qquad \text{(by using Markov's inequality)} \\
&\leq e^{\frac{t^2\sigma^2}{2}} \Big/ e^{t\lambda} \\
&= e^{\frac{t^2\sigma^2}{2} - t\lambda}.
\end{aligned}$$

By choosing $t = \frac{\lambda}{\sigma^2}$ so as to minimize the right hand side, we get

$$\mathbf{Pr}\left[X_n \geq X_0 + \lambda\right] \leq e^{-\frac{\lambda^2}{2\sigma^2}}.$$

□