Announcements

- One lecture remains.
- Project presentation coming soon.

23.1 Review of Blackwell Approachability

Given a biaffine function

\[ r : X \times Y \rightarrow \mathbb{R}^d \]  

(23.1)

where \( X, Y \) are convex and \( r(x, y) \) is the “payoff vector”.

Denote \( S \) as some goal set, Blackwell Approachability states that

If \( \forall y \in Y, \exists x \in X \) s.t. \( r(x, y) \in S \), then there exists an adaptive strategy s.t.

\[
\frac{1}{T} \sum_{t=1}^{T} r(x_t, y_t) \rightarrow S, \quad \forall y_1, \ldots, y_T
\]

(23.2)

where \( x_t \) is computed via a strategy given \( y_1, \ldots, y_{t-1} \), i.e. \( x_t \leftarrow f(y_1, \ldots, y_{t-1}) \).

Last time we show that

- B.A.T \( \Rightarrow \) No external regret in “expert” setting
- B.A.T \( \Leftarrow \) No regret in OCO
- B.A.T \( \iff \) No internal regret in “expert” setting

We know that B.A.T \( \Rightarrow \) No regret for experts. Consider the following setting

For \( t = 1, 2, \ldots, T \)

- player chooses \( w^t \in \Delta_n \)
- nature chooses \( l^t \in [0, 1]^n \)

We want to guarantee that \( \frac{1}{T} \left( \sum_t w^t l^t - \min_i \sum_t l^t_i \right) = O(1) \).

Define the vector game \( r(w, l) = \langle (w \cdot l - l_1, \ldots, w \cdot l - l_n) \rangle \), if \( \frac{1}{T} \sum r(w^t, l^t) \rightarrow R^n \), we say there is no regret.

**Question:** \( \forall l \in [0, 1]^n, \exists w \in \Delta_n, r(w, l) \in R^n \), how to choose \( w \)?

**Answer:** Choose \( w(l) = e_{i^*} \), where \( i^* = \arg \min_i, l_i \).
23.2 Calibrated Forecasting

23.2.1 Forecast and $\epsilon$ Calibration

What does it mean to make correct forecast?
Repeats prediction for $t = 1, 2, \cdots$

- Forecaster says $p_t \in [0, 1]$
- Nature reveals $y_t \in \{0, 1\}$

Intuitively, what we would expect for $p_1, y_1, \cdots, p_T, y_T$ is
\[
\left| \frac{1}{T} \sum_{t=1}^T p_t - \frac{1}{T} \sum_{t=1}^T y_t \right| \to 0 \quad (23.3)
\]

Eq.(23.3) may be too easy to achieve. Now consider calibrated forecaster. We say a Forecaster is $\epsilon$ calibrated if
\[
\forall p \in [0, 1], \text{ for large enough } T
\left| \sum_{t=1}^T y_t 1[|p_t - p| \leq \epsilon] \right| \frac{1}{T} - p < c \epsilon \quad (23.4)
\]

for some $c > 0$.

**Problem with the definition above:** What if the forecaster never predicts $p$? We need to assume that $\liminf_{T \to \infty} \sum_{t=1}^T 1[|p_t - p| \leq \epsilon] > 0$.

23.2.2 $L_1$-Calibration Score

**Definition:** Assume $[q_1, \cdots, q_N]$ is an $\epsilon$ discretization of $[0, 1]$,
\[
L_1\text{CS}_T^\epsilon = \sum_{i=1}^N \left| \frac{1}{T} \sum_{t=1}^T (q_i - p_t) 1[q_i - p_t \leq \epsilon] \right|
\]

If $\forall \epsilon, \exists T_0: T > T_0, L_1\text{CS}_T^\epsilon \leq c \epsilon$ is equivalent to the former definition about $\epsilon$ calibrated.

23.2.3 Calibration Against an Adversary

It is difficult to calibrate against an adversary. For example, if forecaster says $p_t > 0.5$, adversary chooses $y_t = 0$ and if forecaster says $p_t \leq 0.5$, adversary chooses $y_t = 1$.

**Solution:** The forecaster must actually predict randomly!

Imagine that forecaster chooses $\sigma^t \in \Delta_N$ and $p_t = q_{I_t}$, where $I_t \sim \sigma^t$. Also, imagine adversary chooses $y_t \sim \alpha \in [0, 1]$.

Define vector game $r(\sigma, \alpha) = \langle (q_i - \alpha)\sigma_i \rangle$ for $i = 1, \cdots, N$. Towards using B.A.T., the average payoff is
\[
\frac{1}{T} \sum_{t=1}^T r(\sigma^t, \alpha^t) = \frac{1}{T} \sum_{i=1}^N (q_i - \alpha)\sigma_i^t = \mathbb{E}_{y_t \sim \alpha, p_t \sim \sigma^t}[\frac{1}{T} \sum_{t=1}^T (q_i - y_t) 1[p_t = q_i]]
\]

if the average payoff converges to $L1$ ball of radius $c \epsilon$, the we are calibrated,
To show that \( \epsilon \)-calibration \( \Leftrightarrow \) Approachability of \( \mathbb{B}_1(\epsilon) \), first we need to check \( \forall \alpha \in [0,1], \exists \sigma \in \Delta_n \), s.t.

\[
\langle (q_i - \alpha) \sigma_i \rangle_{i=1,\ldots,n} \in \mathbb{B}_1(\epsilon) \tag{23.5}
\]

Set \( \sigma \) to put all weight on \( q_i^* \), the nearest grid point to \( \alpha \),

\[
\langle (q_i - \alpha) \sigma_i \rangle = \langle 0, \ldots, (q_i - \alpha)1, 0 \ldots, 0 \rangle \in \mathbb{B}_1(\epsilon) \tag{23.6}
\]

we can approach \( \mathbb{B}_1(\epsilon) \).

Sketch proof on reverse reduction: Calibration \( \Rightarrow \) B.A.T.

Given \( r : X \times Y \to \mathbb{R}^d \), a convex set \( S \subset \mathbb{R}^d \). Assume that \( \forall y \in Y, \exists x \in X, r(x,y) \in S \) and we have a calibrated algorithm.

For \( t = 1, 2, \ldots \)

1. Player “guesses” opponent’s cation \( \hat{y}_t \in Y \). Let this be a “calibrated forecast” \( x(\hat{y}_t) \)
2. Player selects \( x_t \) s.t. \( r(x_t, \hat{y}_t) \)
3. Player observes true \( y_t \)

For the sake of the analysis, let \( u_T^i := \sum_{t=1}^T 1[\hat{y}_t = q_i] \), that is, the number of times the forecaster predicted that \( \hat{y}_t \) was the grid point \( q_i \). Then we have

\[
\frac{1}{T} \sum_{t=1}^T r(x_t, y_t) = \sum_{i=1}^N \left( \frac{1}{T} \sum_{t=1}^T r(x_t, y_t) 1[\hat{y}_t = q_i] \right)
\]

\[
= \sum_{i=1}^N \left( \frac{1}{T} \sum_{t=1}^T r(x(q_i), y_t) 1[\hat{y}_t = q_i] \right)
\]

next we apply the calibration statement

\[
= \sum_{i=1}^N r(x(q_i), \frac{1}{T} \sum_{t=1}^T y_t 1[\hat{y}_t = q_i])
\]

where \( u_t \) is \( O(1) \)-norm “error” vec

\[
= \sum_{i=1}^N r(x(q_i), \frac{n_i}{T} (q_i + \epsilon \bar{u}_i))
\]

where \( \bar{u} \) is \( O(1) \)-norm avg “error” vec

\[
= \left( \frac{\sum_{i=1}^n n_i}{T} r(x(q_i), q_i) \right) + \epsilon \bar{u}
\]

Notice that the first term in the final expression is an average of elements of \( S \) by construction, and the second term is a vector of norm \( O(\epsilon) \). Hence the final vector is \( O(\epsilon) \) close to \( S \) as desired..