EECS598: Prediction and Learning: It's Only a Game Fall 2013

Lecture 23: B.A.T Review and Calibrated Forecasting

Prof. Jacob Abernethy *Scribe: Lianli Liu*

Announcements

- One lecture remains.
- Project presentation coming soon.

23.1 Review of Blackwell Approachability

Given a biaffine function

$$r : X \times Y \to \mathbb{R}^d \tag{23.1}$$

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.

$$\frac{1}{T}\sum_{t=1}^{T} r(x_t, y_t) \to S, \quad \forall y_1, \cdots y_T \tag{23.2}$$

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

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 $\Leftrightarrow$ No internal regret in "expert" setting

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

For $t = 1, 2, \cdots T$

- player chooses $w^t \in \triangle_n$
- nature chooses $l^t \in [0,1]^n$

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

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

Question:$\forall l \in [0,1]^n, \exists w \in \triangle_n, r(w,l) \in R^n$, 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 ϵ 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

$$\left| \frac{1}{T} \sum p_t - \frac{1}{T} \sum y_t \right| \to 0 \tag{23.3}$$

Eq(23.3) may be too easy to achieve. Now consider calibrated forecaster. We say a Forecaster id ϵ calibrated if

$\forall p \in [0, 1]$, for large enough T

$$\left| \frac{\sum_{t=1}^{T} y_t \mathbf{1}[|p_t - p| \leq \epsilon]}{\sum_{t=1}^{T} \mathbf{1}[|p_t - p| \leq \epsilon]} - p \right| < c\epsilon \tag{23.4}$$

for some $c > 0$.

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

23.2.2 L1-Calibration Score

Definition: Assume $[q_1, \cdots, q_n]$ is an ϵ discretization of $[0, 1]$,

$$L1CS_T^{\epsilon} = \sum_{i=1}^{N} \left| \frac{1}{T} \sum_{t=1}^{T} (q_i - p_t) \mathbf{1}[|q_i - p_t| \leq \epsilon] \right|$$

If $\forall \epsilon, \exists T_0 : T > T_0, L1CS_T^{\epsilon} \leq c\epsilon$ is equivalent to the former definition about ϵ 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 $\sigma^t \in \triangle_N$ and $p_t = q_{I_t}$, where $I_t \sim \sigma^t$. Also, image 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 $\frac{1}{T} \sum_{t=1}^{T} r(\sigma^t, \alpha^t) = \langle \frac{1}{T} \sum_{t=1}^{T} (q_i - \alpha)\sigma_i^t \rangle = \mathbb{E}_{y_t \sim \alpha, p_t \sim \sigma^t} [\frac{1}{T} \sum_{t=1}^{T} (q_i - y^t) \mathbf{1}[p_t = q_i]]$

if the average payoff converges to $L1$ ball of radius $c\epsilon$, the we are calibrated, To show that ϵ-calibration $\Leftrightarrow$ Approachability of $B_1(c\epsilon)$, first we need to check $\forall \alpha \in [0,1], \exists \sigma \in \triangle_n$, s.t.

$$\langle (q_i - \alpha)\sigma_i \rangle_{i=1,\ldots,n} \in B_1(c\epsilon) \tag{23.5}$$

Set σ to put all weight on q_i^*, the nearest grid point to α,

$$\langle (q_i - \alpha)\sigma_i \rangle = \langle 0, \cdots, (q_i - \alpha)1, 0 \cdots, 0 \rangle \in B_1(c\epsilon) \tag{23.6}$$

we can approach $B_1(c\epsilon)$.

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 $t = 1, 2, \cdots$

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 $n_T^i := \sum_{t=1}^T 1[\hat{y}_t = q_i]$, that is, the number of times the forecaster predicted that $\hat{y}_t$ was the grid point q_i. Then we have

$$\begin{aligned}
\tfrac{1}{T}\textstyle\sum_{t=1}^T r(x_t, y_t) \quad &= \sum_{i=1}^N \left(\frac{1}{T} \sum_{t=1}^T r(x_t, y_t) 1[\hat{y}_t = q_i] \right) \\
&= \sum_{i=1}^N \left(\frac{1}{T} \sum_{t=1}^T r(x(q_i), y_t) 1[\hat{y}_t = q_i] \right) \\
\text{next we apply the calibration statement} \quad &= \sum_{i=1}^N r\left(x(q_i), \frac{1}{T} \sum_{t=1}^T y_t 1[\hat{y}_t = q_i] \right) \\
\text{where } u_i \text{ is } O(1)\text{-norm "error" vec} \quad &= \sum_{i=1}^N r\left(x(q_i), \frac{n_T^i}{T} (q_i + \epsilon u_i) \right) \\
\text{where } \bar{u} \text{ is } O(1)\text{-norm avg "error" vec} \quad &= \left(\sum_{i=1}^N \frac{n_T^i}{T} r\left(x(q_i), q_i \right) \right) + \epsilon \bar{u}
\end{aligned}$$

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..