Approachability in unknown games

Gilles Stoltz (CNRS — HEC Paris)

Joint work with

Shie Mannor (Technion)

Vianney Perchet (Univ. Paris-Diderot — INRIA)

Statement of the problem

Regret can be minimized whether the game is known or not.

Can also approachability theory be extended to unknown games?

Classical approachability theory (Blackwell, 1956)

Finite sets of actions $\mathcal{X}$ and $\mathcal{Y}$, actions taken $x_t \in \mathcal{X}$ and $y_t \in \mathcal{Y}$

Payoff function $r : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^d$

Aim: closed convex set $\mathcal{C}$, with first player to ensure

$$\frac{1}{T}\sum_{t=1}^{T} r(x_t, y_t) \longrightarrow \mathcal{C}$$

and second player to prevent this convergence

Characterization: $\quad \forall \mathbf{y} \in \Delta(\mathcal{Y}), \quad \exists\, \mathbf{x} \in \Delta(\mathcal{X}) : \; r(\mathbf{x}, \mathbf{y}) \in \mathcal{C}$

If this condition fails, the smallest approachable blow-up of $\mathcal{C}$ is

$$\mathcal{C}_{\alpha_{\text{unif}}} = \left\{ c' \in \mathbb{R}^d : \; \mathrm{d}_2(c', \mathcal{C}) \leqslant \alpha_{\text{unif}} \right\}$$

where $\alpha_{\text{unif}} = \max_{\mathbf{y} \in \Delta(\mathcal{Y})} \min_{\mathbf{x} \in \Delta(\mathcal{X})} \mathrm{d}_2\big(r(\mathbf{x}, \mathbf{y}), \mathcal{C}\big)$

Classical approachability theory (Blackwell, 1956)

Closed convex set $\mathcal{C}$ and $\quad \forall \mathbf{y} \in \Delta(\mathcal{Y}), \quad \exists \mathbf{x} \in \Delta(\mathcal{X}) : \quad r(\mathbf{x}, \mathbf{y}) \in \mathcal{C}$

Associated strategy

- Compute $\overline{c}_t = \Pi_{\mathcal{C}}\left(\overline{r}_t\right)$, the projection onto $\mathcal{C}$ of

$$\overline{r}_t = \frac{1}{t} \sum_{s=1}^{t} r(x_s, y_s)$$

- Draw $x_{t+1} \sim \mathbf{x}_{t+1}$ such that

$$\forall \mathbf{y} \in \Delta(\mathcal{Y}), \qquad \left\langle \overline{r}_t - \overline{c}_t, \; r(\mathbf{x}_{t+1}, \mathbf{y}) - \overline{c}_t \right\rangle \leqslant 0$$

What does the player need to know/observe?

- Bandit monitoring enough: observe $r(x_t, y_t)$, not necessarily y_t
- Game r needs to be known in general

Can the game be unknown?

Hope arises from the special case of regret minimization

Action sets $\mathcal{X}$ and $[0,1]^{\mathcal{X}}$

At each round and simultaneously,

- player 1 chooses $x_t \in \mathcal{X}$,
- player 2 picks $(g_{x',t})_{x' \in \mathcal{X}}$

Player 1 gets $g_{x_t,t}$ and ensures

$$\frac{1}{T}\sum_{t=1}^{T} g_{x_t,t} - \max_{x' \in \mathcal{X}} \frac{1}{T}\sum_{t=1}^{T} g_{x',t} \longrightarrow \mathbb{R}_-$$

No underlying game $g : \mathcal{X} \times \mathcal{Y} \to [0,1]$ needs to exist!

Can the game be unknown?

Hope arises from the special case of regret minimization

Player 1 ensures

$$\frac{1}{T}\sum_{t=1}^{T} g_{x_t,t} - \max_{x'\in\mathcal{X}} \frac{1}{T}\sum_{t=1}^{T} g_{x',t} \longrightarrow \mathbb{R}_-$$

Proof (full monitoring, known game)

Consider the vector payoff function

$$r(x_t, y_t) = \big(g(x_t, y_t) - g(x', y_t)\big)_{x'\in\mathcal{X}}$$

Convex set to approach $\mathcal{C} = \big(\mathbb{R}_-\big)^{\mathcal{X}}$

Solution: choose $\mathbf{x}_{t+1}$ proportional to $\big(\overline{r}_t\big)_+$, then

$$\forall\, \mathbf{y} \in \Delta(\mathcal{Y}), \qquad \big\langle \overline{r}_t - \overline{c}_t,\ r(\mathbf{x},\mathbf{y}) - \overline{c}_t \big\rangle = \Big\langle \big(\overline{r}_t\big)_+,\ r(\mathbf{x}_{t+1},\mathbf{y}) + \big(\overline{r}_t\big)_- \Big\rangle = 0$$

because of the definition of r as a vector of differences

Remark: Structure of the game not used, the game could be unknown

Can the game be unknown?

Hope arises from the special case of regret minimization

Player 1 ensures

$$\frac{1}{T}\sum_{t=1}^{T} g_{x_t,t} - \max_{x'\in\mathcal{X}} \frac{1}{T}\sum_{t=1}^{T} g_{x',t} \longrightarrow \mathbb{R}_-$$

Proof (bandit monitoring, unknown game)

Consider the unbiased estimates

$$\widetilde{g}_{x',t} = \frac{g_{x',t}}{\mathbf{x}_t(x')}\mathbb{I}_{\{x_t=x'\}}$$

and the associated vector payoff

$$\widetilde{r}_t = \left(\widetilde{g}_{x_t,t} - \widetilde{g}_{x',t}\right)_{x'\in\mathcal{X}}$$

Convex set to approach $\mathcal{C} = \left(\mathbb{R}_-\right)^{\mathcal{X}}$: again, doable even without knowing the structure

Approachability theory for unknown games

Statement of the problem

At each round and simultaneously,

- player 1 draws $x_t \in \mathcal{X}$ according to $\mathbf{x}_t \in \Delta(\mathcal{X})$
- player 2 picks $\mathbf{m}_t = (m_{x',t})_{x' \in \mathcal{X}} \in K$,

where $K \subset \left(\mathbb{R}^d\right)^{\mathcal{X}}$ is compact

Aim: force (player 1) or prevent (player 2) convergence of

$$\overline{r}_T = \frac{1}{T} \sum_{t=1}^{T} m_{x_t,t}$$

to some neighborhood of $\mathcal{C}$

To do:

- Indicate the targeted neighborhood
- Provide a strategy for player 1

Approachability theory for unknown games

First answers: probably not the end of the story...

Our COLT'14 paper — one solution to the problem

Recall that $\quad \overline{r}_T = \frac{1}{T} \sum_{t=1}^{T} m_{x_t,t}$

Target sets of the form $\mathcal{C}_{\varphi(\overline{\mathbf{m}}_T)}$ $\quad$ where $\overline{\mathbf{m}}_T = \frac{1}{T} \sum_{t=1}^{T} \mathbf{m}_t$

Ensure $\quad \mathrm{d}_2\Big(\overline{r}_T,\ \mathcal{C}_{\varphi(\overline{\mathbf{m}}_T)}\Big) \longrightarrow 0$

Why a function of the mean $\overline{\mathbf{m}}_T$ and not of the entire path?

Maybe too restrictive; and we will see that we anyway need to decompose $\overline{\mathbf{m}}_T$

What do we get by calibration?

On "average" good predictions $\widehat{\mathbf{m}}_t$ of the $\mathbf{m}_t$

(up to some grouping rounds according to the values of the $\widehat{\mathbf{m}}_t$)

Randomized strategy given by $\Psi : K \to \Delta(\mathcal{X})$

Average payoff close to

$$\frac{1}{T}\sum_{t=1}^{T} \Psi\big(\widehat{\mathbf{m}}_t\big) \odot \mathbf{m}_t$$

where $\mathbf{x} \odot \mathbf{m} = \mathbb{E}\big[m_X\big]$ when $X \sim \mathbf{x}$

But

- some grouping is needed (because of calibration)
- the guarantee needs to hold along the whole path (for all T)

What do we get by calibration?

Final guarantee in terms of convex decompositions:

$$\varphi(\overline{\mathbf{m}}_T) = \sup\left\{ \mathrm{d}_2\left(\sum_i \lambda_i \Psi\big(\mathbf{m}^{(i)}\big) \odot \mathbf{m}^{(i)},\ \mathcal{C}\right) : \quad \sum_i \lambda_i \mathbf{m}^{(i)} = \overline{\mathbf{m}}_T \right\}$$

Still a big problem to solve:

Which Ψ should be chosen?

There can be "compensations" and there are sometimes better choices than

$$\Psi(\mathbf{m}) \in \operatorname*{arg\,min}_{\mathbf{x}\in\Delta(\mathcal{X})} \mathrm{d}_2\big(\mathbf{x} \odot \mathbf{m},\, \mathcal{C}\big)$$

An efficient strategy

We tackled the efficiency issue and offer a strategy that

- minimizes some regret in rounds of lengths 1, 2, 3, ...;
- only calls Ψ once in a round;
- performs no projection;
- ensures $\mathrm{d}_2\Big(\overline{r}_T,\ \mathcal{C}_{\varphi(\overline{\mathbf{m}}_T)}\Big) = \mathcal{O}\big(T^{-1/4}\big)$

This strategy has consequences in classical approachability as well!

1. Classical approachability without projecting onto $\mathcal{C}$

I.e., with $\varphi \equiv 0$ as a target

Only by exploiting the dual condition

$$\Psi(\mathbf{m}) \odot \mathbf{m} \in \mathcal{C} \quad \text{where} \quad \Psi(\mathbf{m}) \in \operatorname*{arg\,min}_{\mathbf{x} \in \Delta(\mathcal{X})} \mathrm{d}_2\big(\mathbf{x} \odot \mathbf{m}, \mathcal{C}\big)$$

Note: By a clever trick using that $\mathcal{C}$ is approachable, Bernstein and Shimkin (2014) recover the classical $\mathcal{O}\big(T^{-1/2}\big)$ rate

2. Convergence to the smallest approachable expansion $\mathcal{C}_{\alpha_{\text{unif}}}$ of a known game

Without even knowing it! Same principle

Does not solve the NP-hard optimization problem of determining

$$\alpha_{\text{unif}} = \max_{\mathbf{y} \in \Delta(\mathcal{Y})} \min_{\mathbf{x} \in \Delta(\mathcal{X})} \mathrm{d}_2\big(r(\mathbf{x}, \mathbf{y}), \mathcal{C}\big)$$