Mechanisms for a No-Regret Agent: Beyond the Common Prior



Modibo Camara Jason Hartline Aleck Johnsen

Northwestern University

AEA/ASSA Annual Meeting

January 3-5, 2021

For more details, see full paper (arxiv.org/abs/2009.05518).

A High-Level Agenda in Game Theory

Can we replace prior knowledge with models of learning?

Yes? No-regret assumptions can discipline agents' behavior.

- E.g. Foster and Vohra (1997)
- E.g. Nekipelov, Syrgkanis, and Tardos (2015)

No? No-regret algorithms can behave in bizarre ways.

- E.g. Braverman, Mao, Schneider, and Weinberg (2018).

Can models of learning be adapted to strategic interactions?

This Paper (1/2)

We focus on Stackelberg games of incomplete information.

- Includes contract design, Bayesian persuasion, delegation, etc.

1. Principal decides on a policy, e.g. 

or 
2. Agent responds to that policy, e.g.  or 

3. Payoffs depend on hidden state, e.g.  or 

These models usually require common prior belief about the state.

- E.g. 40% chance policeman is at donut shop.

This Paper (2/2)

We replace common prior with adversarial online learning.

- Adversarial = no assumptions on the sequence of states.

Principal

Agent

State

We design low-regret mechanisms for the principal.

- Under permissive assumptions on the agent's behavior.
- We refine no-regret to counterfactual calibration.
- Agent must fully & consistently exploit her information.

Agent's Regret

No-regret:

- In hindsight, agent prefers her algorithm to any fixed action.

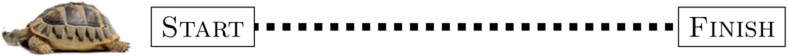
Calibration (no-internal-regret):

- No-regret with past behavior as context.

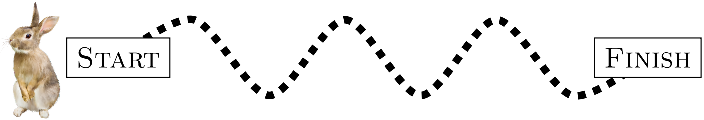
These restrictions do not rule out pathological behavior.

- The adversary can correlate agent's actions with sequence of states to make it appear as if agent has additional information.
- Informally, can think of an agent that has access to additional data or notices patterns that we missed.

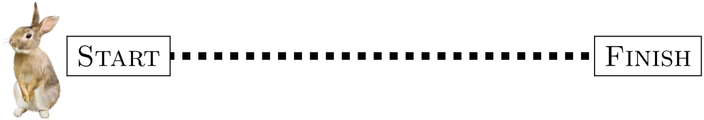
Tortoise travels 1km in 1h : uninformed agent satisfies no-regret.



Hare travels 1km in 1h : informed agent satisfies no-regret.



Hare travels 1km in 2min : informed agent satisfies no-regret conditioned on her information.



Counterfactual Calibration

Information is revealed by the agent's on- and off-path behavior.

- On-path: agent's behavior under the principal's mechanism.
- Off-path: agent's behavior under counterfactual mechanisms.

Counterfactual calibration:

- No-regret with on- and off-path behavior as context.

Agent must fully & consistently exploit her revealed information.

Online Mechanism Design

Principal's regret:

- How much he prefers the best-in-hindsight fixed policy to his mechanism.
- Takes into account how changes to mechanism affect agent's behavior.

Goal: design a mechanism that guarantees low principal's regret.

- For any sequence of states.
- For any counterfactually-calibrated behavior by the agent.

Ideally, we seek empirical analogs to the common prior policies.

Main Results (1/2)

Reduce online problem to robust version of common prior problem.

- Local robustness: agent nearly maximizes expected utility.

Theorem (Informal)

Assume: agent's behavior is counterfactually calibrated.

Assume: agent isn't better informed than us (non-negative regret).

Mechanism: use the locally-robust policy, replacing the common prior with a calibrated forecast.

Result: principal's regret vanishes.

Main Results (2/2)

Reduce online problem to robust version of common prior problem.

- Local robustness: agent nearly maximizes expected utility.
- Informational robustness: agent receives a private signal, but we do not know its quality.

Theorem (Informal)

Assume: agent's behavior is counterfactually calibrated.

Mechanism: use the locally- and informationally-robust policy, replacing the common prior with a calibrated forecast.

Result: principal's regret $\leq$ cost of informational robustness.

Conclusion

We replaced the common prior with online learning.

- For rich class of single-agent mechanism design problems.
- Counterfactual calibration $\approx$ non-Bayesian rationality.
- If the agent reveals useful information on one path, she must also exploit it on other paths.

Many open problems, e.g.

- Extension to bandit feedback – conceptually non-trivial!
- Mechanisms that learn about the agent's information.
- Computational tractability.

Thank you!

For more details, see full paper (arxiv.org/abs/2009.05518).