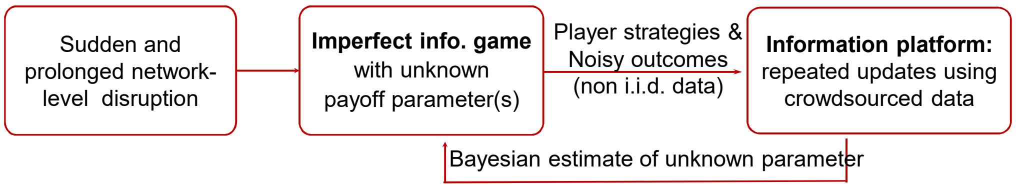
Learning in a strategic environment



Question: How do strategic players repeatedly adjust their strategies in a game, while learning an unknown payoff-relevant parameter?
(with M. Wu & A. Ozdaglar), arXiv:2010.09128

Applications: Strategic traffic routing, e-commerce, portals of financial returns

- ➢ Strategic decision making with imperfect information
- ➢ Data for parameter estimation is endogenous

Related literature:

[Learning in games] Fudenberg and Kreps (1993, 1995), Fudenberg and Levine (1993), Cominetti, Melo and Sorin (2010), Krichene et al. (2014)

[Learning in control] Tsitsiklis (1994), Benaim Hofbauer and Sorin (2005, 2006), Recht (2019)

[Social learning] Banerjee (1992), Bikhchandani, Hirshleifer, and Welch (1992), Gale and Kariv (2003), Acemoglu et al.(2011)

Stochastic learning dynamics

Bayesian belief update:

$$\theta^{k+1}(s) = \frac{\theta^k(s)\phi^s(y^k|q^k)}{\sum_{s'\in S}\theta^k(s')\phi^{s'}(y^k|q^k)}$$

Two types of strategy updates:

- ➢ Equilibrium strategy $g(\theta^{k+1}) \in EQ(\theta^{k+1})$

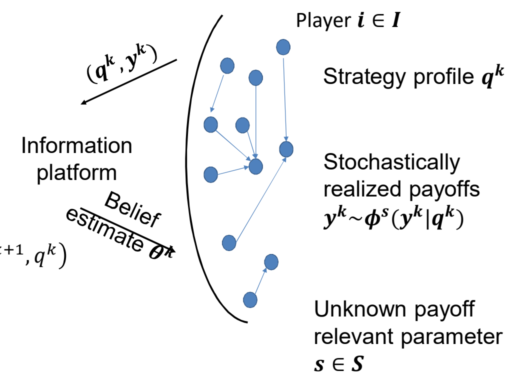
$$q_i^{k+1} = (1-a_i^k)q_i^k + a_i^k\, g_i(\theta^{k+1})$$

- ➢ Best response strategy $h(\theta^{k+1}, q^k) \in BR(\theta^{k+1}, q^k)$

$$q_i^{k+1} = (1-a_i^k)q_i^k + a_i^k\, h_i(\theta^{k+1}, q_{-i}^k)$$

Asynchronous step sizes $a_i^k \in [0,1]$:

- ➢ Relative speed of learning in games
- ➢ Constraints faced by players



Convergence and stability

Convergence

Under certain conditions, the states $(\theta^k, q^k)_{k=1}^{\infty}$ of both types of learning dynamics converge to a fixed point $(\bar\theta, \bar q)$ with probability 1. For any $s \in S \setminus S^*(q)$

$$\lim_{k\to\infty} \frac{1}{k} log\left(\theta^k(s)\right) = -D_{KL}(\phi^{s*}(y|\bar q)||\phi^s(y|\bar q))$$

Sufficient conditions for stability

Local stability	a. Parameters in $[\bar\theta]$ are payoff equivalent to s^* in local neighborhood of $\bar q$ b. Strategy update is robust to local perturbation of θ
Global stability	All fixed points have complete information of the unknown parameter

Complete learning

$[\bar{\boldsymbol\theta} = \boldsymbol\theta^*, \bar{\boldsymbol q} = \boldsymbol q^*]$ Any other beliefs include parameters that are non payoff equivalent given an eq. strategy

$[\bar{\boldsymbol\theta} \neq \boldsymbol\theta^*, \bar{\boldsymbol q} = \boldsymbol q^*]$ Condition (a) + concave payoff

Payoff equivalent parameters $S^*(q)$

$$D_{KL}(\phi^{s*}(y|\bar q)||\phi^s(y|\bar q)) = 0$$

Fixed point $(\bar\theta, \bar q)$:

- ➢ $[\bar\theta] \in S^*(\bar q)$
  $\bar\theta$ consistently estimates the payoff distribution given $\bar q$
- ➢ No incentive to deviate from $\bar q$
  $\bar q$ is an equilibrium w.r.t. $\bar\theta$

Complete info. Belief $\boldsymbol\theta^*$

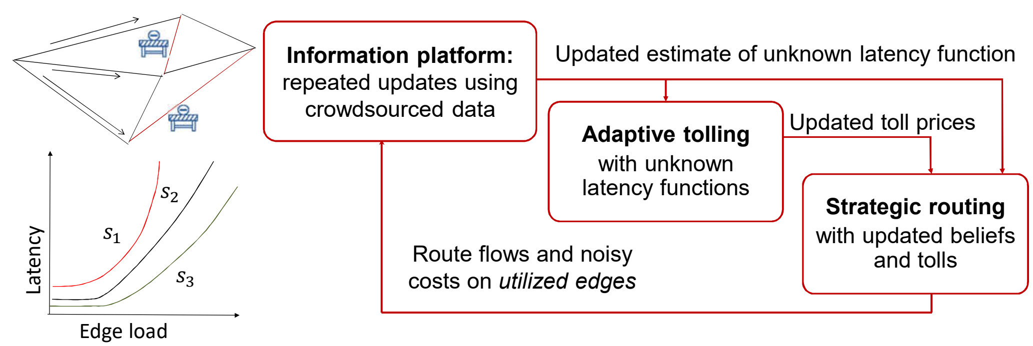
Complete info. Eq. $\boldsymbol q^*$

$(\bar\theta, \bar q) \equiv (\theta^*, q^*)$?

No!

- ➢ $\bar\theta$ may incorrectly estimate the payoff distribution given $q \neq \bar q$
- ➢ Players have incentive to deviate with complete info.

Adaptive tolling with unknown latency functions



Adaptive tolling mechanism: If all edges are initially taken, then

1. Learning leads to optimal tolling mechanism with complete info.
2. Adaptive tolling induces optimal routing with complete information.