# Calibrated learning and games

Geelon So, agso@ucsd.edu
Forecasting/calibration reading group — December 5, 2023

# Online decision making

**Setting.** For $t = 1, 2, \ldots$

- (receive context/side-information)
- play action $a_t \in \mathcal{A}$
- incur loss $\ell_t(a_t)$ and receive feedback

**Types of feedback.**

- Full-information feedback: observe loss function $\ell_t : \mathcal{A} \to \mathbb{R}$
- Partial-information/bandit feedback: observe loss $\ell_t(a_t)$

# Examples of online decision making problems

- predicting an infinite bitstream
- online learning
- multi-armed bandit problem
- playing an infinitely repeated game

# Goals in online decision making

The cumulative loss at time $T$:

$$L_T = \sum_{t=1}^{T} \ell_t(a_t).$$

**Goal:** we would like to minimize the cumulative loss, but this is an extremely tall order

- regret analysis: compare against some restricted class of decision sequences

# Goals of online decision making

- **external regret**: I should've just invested all my money in an index fund
  - given a set of expert advice $(a'_t) \in \mathcal{E}$, perform just about as well as the best expert

$$R_t = \sup_{(a'_t) \in \mathcal{E}} \sum_{t=1}^{T} \ell_t(a_t) - \ell_t(a'_t)$$

- **internal/swap regret**: every time I bought stock A, I should have bought stock B
  - a modification rule $m : \mathcal{A} \to \mathcal{A}$ replaces an action $a$ by another $m(a)$; internal regret:

$$R_t = \sup_{m:\mathcal{A} \to \mathcal{A}} \sum_{t=1}^{T} \ell_t(a_t) - \ell_t(m(a_t))$$

# Decision making from predictions

A **forecasting approach** to online decision making:

- at each time step, predict the loss function $\hat{\ell}_t$
- make a decision by minimizing the predicted loss

$$a_t \leftarrow \arg\min_{a \in \mathcal{A}} \hat{\ell}_t(a).$$

**Example.** Each day, decide whether to bring an umbrella.

- forecast the chance of rain (equivalent to making prediction of $\hat{\ell}_t$)
- decide to carry an umbrella based on prediction

# Example of online decision making

**Decision problem.** Each day, decide whether to carry an umbrella where the cost is:

| umbrella? | sun | rain |
|---:|:---:|:---:|
| yes | 1 | 0 |
| no | 0 | 3 |

# Example of online decision making

Suppose that rain is the outcome of a biased coin flip. The expected cost of either action:

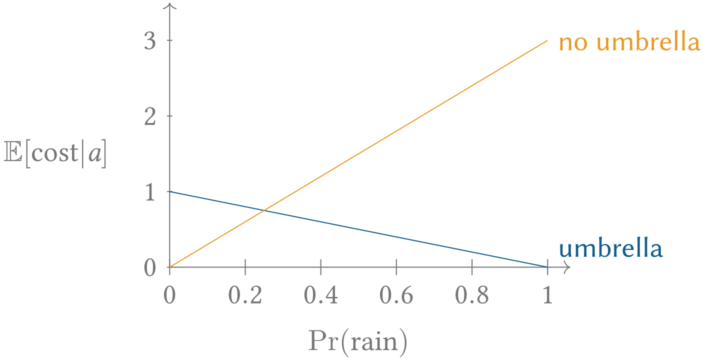



Figure 1: When $3p = 1 - p$, the agent minimizing expected cost is indifferent.

# Example of online decision making

- If the agent knew $p$, then the best-response action to $p$ is:

$$a_{\mathrm{BR}}(p) = \begin{cases} \text{bring umbrella} & p > 1/4 \\ \text{leave umbrella} & p < 1/4. \end{cases}$$

- In general, we make no probabilistic assumptions
  - i.e. $p$ might not even exist
  - still, we can give a prediction for $p_t$ using history
  - this is equivalent to giving estimate $\hat{\ell}_t$

$$\hat{\ell}_t(a) = \mathop{\mathbb{E}}_{p_t}\left[\mathrm{cost}(\mathrm{rain}) \middle| a\right]$$

# Calibration: motivation

**Question:** how do we make predictions that lead to decisions with low regret?

**Example.** Suppose we make discrete predictions for the chance of rain:

10% 20% 30% ... 70% 80% 90%

- the best-response play is to bring umbrella when prediction >25%
- the predictions are classically calibrated if for each prediction $p$ of rain:

$$\frac{\#\text{ times it rained when } p \text{ predicted}}{\#\text{ times } p \text{ was predicted}} \to p$$

- best-response play to calibrated predictions incur no internal regret (why?)

# Calibration

**Setting.**

- the domain $\mathcal{X} \subset \mathbb{R}^n$ is a compact and convex
- the prediction at time $t$ is denoted $p_t \in \mathcal{X}$
- the outcome at time $t$ is denoted $x_t \in \mathcal{X}$
- a (deterministic) forecasting procedure maps sequences of outcomes to predictions:

$$p_{t+1} \equiv p_{t+1}(h_t),$$

where $h_t = (x_1, \ldots, x_t)$ is the history up to time $t$.

# Calibration

## Definition (Continuously calibrated forecast)

*A forecasting procedure is* **continuously calibrated** *if for any continuous* $\phi : \mathcal{X} \to [0,1]$,

$$\lim_{T\to\infty} \left( \sup_{x_1,\ldots,x_T} \left\| \frac{1}{t} \sum_{t=1}^{T} \phi(p_t)(x_t - p_t) \right\| \right) = 0. \quad (1)$$

- Here, $(p_t)_t$ is generated by the forecasting procedure given the outcomes $(x_t)_t$.
- Let $\phi(p)$ be a continuous approximation to $\mathbf{1}\{p = p^*\}$. Intuitively, (1) holds when:
  - either $p^*$ is predicted very infrequently, or
  - the average of the outcomes whenever $p^*$ was predicted converges to $p^*$

# Existence of calibrated forecasting

**Theorem (Existence, Foster and Hart (2021))**

*Deterministic continuously-calibrated forecasting procedures exist.*

# Calibrated learning

# Game setting

**Setting.** Consider the following 2-player game with binary action.

| | G | S |
|---|---|---|
| g | (-10, -10) | (1, 0) |
| s | (0, 1) | (-1, -1) |

Table 1: Payoff matrix at a traffic intersection (go/stop).

- A (pure) Nash equilibrium is a joint action $a \in \{g, s\} \times \{G, S\}$ such that any player can only worsen their utility by unilaterally deviating.
- This game has two Nash equilibria: $(s, G)$ and $(g, S)$.

# Game setting

**General setting.** A finite game with $N$ players consists of:

- a finite set of actions/pure strategies $\mathcal{A}_i$ for each player
  - denote the set of joint pure strategies by $\mathcal{A} := \prod \mathcal{A}_i$
  - denote the set of joint, (product) mixed strategies by $\Delta(\mathcal{A}) := \prod \Delta(\mathcal{A}_i)$

$$p \equiv (p_1, \ldots, p_N) \in \Delta(\mathcal{A}) \qquad \Longleftrightarrow \qquad p(a_1, \ldots, a_n) = \prod p_i(a_i).$$

- a family of utilities $u_i : \mathcal{A} \to \mathbb{R}$

# Mixed Nash equilibria

## Definition

*A mixed strategy $p \in \Delta(\mathcal{A})$ is a **mixed Nash equilibrium** if for each player $i$ and alternative strategy $p' = (p_1, \ldots, p_{i-1}, p'_i, p_{i+1}, \ldots, p_N)$, the following holds:*

$$\mathop{\mathbb{E}}_{a \sim p}[u_i(a)] \geq \mathop{\mathbb{E}}_{a' \sim p'}[u_i(a')].$$

*We say that $p$ is an **$\varepsilon$-mixed Nash equilibrium** if:*

$$\mathop{\mathbb{E}}_{a \sim p}[u_i(a)] + \varepsilon \geq \mathop{\mathbb{E}}_{a' \sim p'}[u_i(a')].$$

# Continuously calibrated learning dynamics

**Definition ($\varepsilon$-calibrated learning)**

*Let $\varepsilon > 0$. An $\varepsilon$-continuously calibrated learning dynamic is a pair:*

(i) *a deterministically calibrated procedure on $\Delta(\mathcal{A})$*

(ii) *a continuous $\varepsilon$-best reply function $x_{\varepsilon\text{-BR}} : \Delta(\mathcal{A}) \to \Delta(\mathcal{A})$*

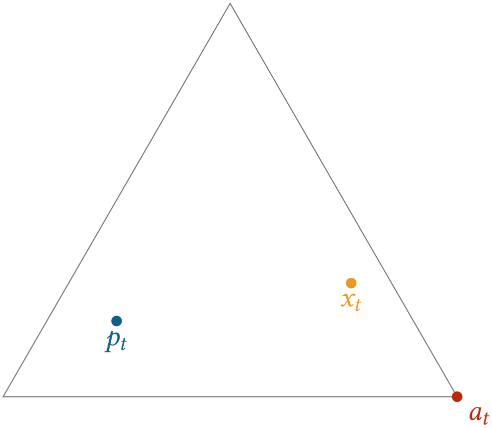
# Continuously calibrated learning dynamics

For $t = 1, 2, \ldots$

- Construct forecast $p_t \equiv p_t(h_{t-1})$ using history $h_{t-1}$
- Each player plays an $\varepsilon$-best response to $p_t$, jointly playing $x_t \equiv x_{\varepsilon\text{-BR}}(p_t)$
- A realization $a_t$ is drawn from $p_t$ and revealed; history is extended:

$$h_t = (a_1, \ldots, a_t).$$

# Continuously calibrated learning dynamics




**Setting.** The triangle is the set of all independent mixed strategies.

**Dynamics.** In round $t$,

- the forecast $p_t$ is made,
- an $\varepsilon$-best response $x_t$ is played,
- a realization $a_t$ is drawn.

# Continuous calibrated learning finds Nash equilibria

### Theorem

*Suppose players in an infinitely-repeated finite game follow an $\varepsilon$-continuously calibrated learning dynamic, where $(x_t)_t$ is the sequence of their mixed strategies. Then for all $\varepsilon' > \varepsilon$:*

$$\lim_{T\to\infty} \frac{\#\{t \in [T] : x_t \in \mathrm{NE}(\varepsilon')\}}{T} = 1 \qquad \text{a.s.}$$

# Proof sketch

- Consider the times $(t_k)_k$ when $p_{t_k}$ was close to some $p$.
- To be calibrated means that:

$$\lim_{K\to\infty}\left\|\frac{1}{t_K}\sum_{k=1}^{K}(a_{t_k}-p)\right\| = \lim_{K\to\infty}\frac{K}{t_K}\left\|\frac{1}{K}\sum_{k=1}^{K}(a_{t_k}-p)\right\| = 0.$$

  1. Either $p$ is eventually predicted very infrequently $t_K = \omega(K)$, or
  2. The average $\frac{1}{K}\sum a_{t_k}$ converges to $p$.
- The sequence $(a_t - x_t)_t$ is a martingale difference sequence.
  - The average $\frac{1}{K}\sum a_{t_k}$ also converges to $\frac{1}{K}\sum x_{t_k}$
  - Notice that $x_{t_k} \approx x_{\text{BR}}(p)$, so (2) implies an approximate Nash equilibrium condition:

$$x_{\text{BR}}(p) \approx p.$$

- Thus, if $(t_k)$ fairly dense, then $x_{t_k}$'s must be approximate Nash equilibria.

# Proof sketch

Almost done; but we need to rule out the following:

- **Bad case:** each $p$ is predicted very infrequently; i.e. we are trivially calibrated because the orange term goes to zero:

$$\lim_{K \to \infty} \frac{K}{t_K} \left\| \frac{1}{K} \sum_{k=1}^{K} \left( a_{t_k} - p \right) \right\| = 0.$$

    - This would require the set of possible predictions to be very large.
    - Ruled out by compactness of the space of mixed strategies for finite games and the uniform continuity of the approximate best response map.

□

# References

Dean P Foster and Sergiu Hart. Forecast hedging and calibration. *Journal of Political Economy*, 129(12): 3447–3490, 2021.