Opportunistic Strategies for Generalized No-Regret Problems

Andrey Bernstein, Shie Mannor, and Nahum Shimkin

Technion – Israel Institute of Technology

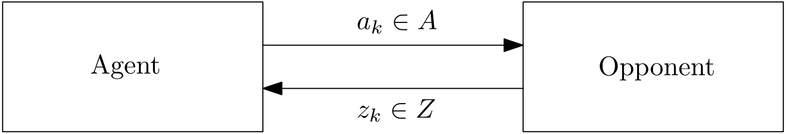
COLT 2013 / June 13, 2013

Outline

Outline

Generalized No-Regret Problems

A repeated matrix game between two players, the agent and the opponent.



$\bar{R}_n = \frac{1}{n}\sum_k^n r(a_k, z_k)$ $\bar{q}_n = \frac{1}{n}\sum_{k=1}^n \mathbf{1}(z_k) \in \Delta(Z)$

$$r(p,q) = \sum_{a,z} p(a)q(z)r(a,z), \quad p \in \Delta(A), q \in \Delta(Z)$$

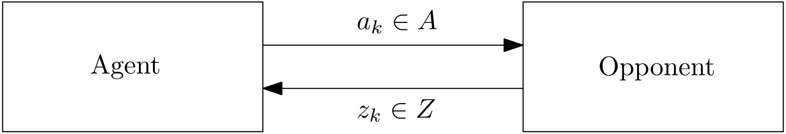
Vector-valued payoff function $r(a,z) \in \mathbb{R}^K$.

For each mixed action q of the opponent, the agent has:

- A desired payoff set $R^*(q) \subset \mathbb{R}^K$.
- A response $p = p^*(q)$ that satisfies $r(p,q) \in R^*(q)$.

Generalized No-Regret Problems

A repeated matrix game between two players, the agent and the opponent.



$\bar{R}_n = \frac{1}{n} \sum_k^n r(a_k, z_k)$ $\quad$ $\bar{q}_n = \frac{1}{n} \sum_{k=1}^n \mathbf{1}(z_k) \in \Delta(Z)$

$$r(p, q) = \sum_{a,z} p(a) q(z) r(a, z), \quad p \in \Delta(A), q \in \Delta(Z)$$

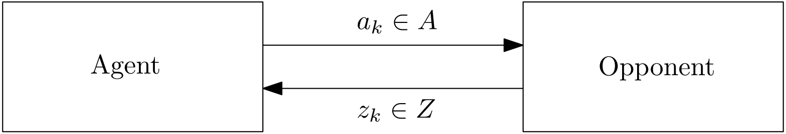
Vector-valued payoff function $r(a, z) \in \mathbb{R}^K$.

For each mixed action q of the opponent, the agent has:

- A desired payoff set $R^*(q) \subset \mathbb{R}^K$.
- A response $p = p^*(q)$ that satisfies $r(p, q) \in R^*(q)$.

Generalized No-Regret Problems

A repeated matrix game between two players, the agent and the opponent.



$\bar{R}_n = \frac{1}{n} \sum_k^n r(a_k, z_k)$ $\qquad$ $\bar{q}_n = \frac{1}{n} \sum_{k=1}^n \mathbf{1}(z_k) \in \Delta(Z)$

$$r(p, q) = \sum_{a,z} p(a) q(z) r(a, z), \quad p \in \Delta(A), q \in \Delta(Z)$$

Vector-valued payoff function $r(a, z) \in \mathbb{R}^K$.

For each mixed action q of the opponent, the agent has:

- A desired payoff set $R^*(q) \subset \mathbb{R}^K$.
- A response $p = p^*(q)$ that satisfies $r(p, q) \in R^*(q)$.

Definition (Generalized No-Regret)

An algorithm has no-regret with respect to $R^*(q)$ if

$$\lim_{n \to \infty} d(\bar{R}_n, R^*(\bar{q}_n)) = 0$$

(almost surely) for any strategy of the opponent. Here, d is the Euclidean distance.

- For short, we say that $\bar{R}_n$ approaches $R^*(\bar{q}_n)$.
- Standard no-regret is a special case with:
  - Scalar rewards $r(a, z)$,
  - $R^*(q) = \{r \in \mathbb{R} : r \geq r^*(q)\}$, where $r^*(q) \triangleq \max_p r(p, q)$.
  - The proof of existence relies on the convexity of the multifunction $q \mapsto R^*(q)$.
- Main interest here: cases when $q \mapsto R^*(q)$ is not convex.

Generalized No-Regret Algorithms

Definition (Generalized No-Regret)

An algorithm has no-regret with respect to $R^*(q)$ if

$$\lim_{n \to \infty} d(\bar{R}_n, R^*(\bar{q}_n)) = 0$$

(almost surely) for any strategy of the opponent. Here, d is the Euclidean distance.

- For short, we say that $\bar{R}_n$ approaches $R^*(\bar{q}_n)$.
- Standard no-regret is a special case with:
  - Scalar rewards $r(a, z)$,
  - $R^*(q) = \{r \in \mathbb{R} : r \geq r^*(q)\}$, where $r^*(q) \triangleq \max_p r(p, q)$.
  - The proof of existence relies on the convexity of the multifunction $q \mapsto R^*(q)$.
- Main interest here: cases when $q \mapsto R^*(q)$ is not convex.

Generalized No-Regret Algorithms

Definition (Generalized No-Regret)

An algorithm has no-regret with respect to $R^*(q)$ if

$$\lim_{n \to \infty} d(\bar{R}_n, R^*(\bar{q}_n)) = 0$$

(almost surely) for any strategy of the opponent. Here, d is the Euclidean distance.

- For short, we say that $\bar{R}_n$ approaches $R^*(\bar{q}_n)$.
- Standard no-regret is a special case with:
  - Scalar rewards $r(a, z)$,
  - $R^*(q) = \{r \in \mathbb{R} : r \geq r^*(q)\}$, where $r^*(q) \triangleq \max_p r(p, q)$.
  - The proof of existence relies on the convexity of the multifunction $q \mapsto R^*(q)$.
- Main interest here: cases when $q \mapsto R^*(q)$ is not convex.

Generalized No-Regret Algorithms

Definition (Generalized No-Regret)

An algorithm has no-regret with respect to $R^*(q)$ if

$$\lim_{n \to \infty} d(\bar{R}_n, R^*(\bar{q}_n)) = 0$$

(almost surely) for any strategy of the opponent. Here, d is the Euclidean distance.

- For short, we say that $\bar{R}_n$ approaches $R^*(\bar{q}_n)$.
- Standard no-regret is a special case with:
  - Scalar rewards $r(a, z)$,
  - $R^*(q) = \{r \in \mathbb{R} : r \geq r^*(q)\}$, where $r^*(q) \triangleq \max_p r(p, q)$.
  - The proof of existence relies on the convexity of the multifunction $q \mapsto R^*(q)$.
- Main interest here: cases when $q \mapsto R^*(q)$ is not convex.

Example: Constrained Regret Minimization

Maximize the average $\bar{U}_n$ of a scalar reward $u(a,z)$, subject to a long-term average cost constraint $\bar{C}_n \leq \gamma + o(1)$, where $\bar{C}_n$ is the average of a cost function $c(a,z)$.

Let

$$u^*_\gamma(q) = \max_{p \in \Delta(\mathcal{A})} \{u(p,q) : c(p,q) \leq \gamma\}.$$

The desired set of the pairs (reward, cost) and the response:

$$R^*(q) = \{r = (u,c) : u \geq u^*_\gamma(q), c \leq \gamma\},$$

$$p^*(q) \in \operatorname*{argmax}_{p \in \Delta(\mathcal{A})} \{u(p,q) : \ c(p,q) \leq \gamma\},$$

$$r(p^*(q),q) = (u(p^*(q),q), c(p^*(q),q)) \in R^*(q).$$

- $q \mapsto R^*(q)$ is non-convex due to the non-convexity of $u^*_\gamma(q)$,
- $R^*(\bar{q}_n)$ cannot be approached in general (Mannor et al., 2009).

Example: Constrained Regret Minimization

Maximize the average $\bar{U}_n$ of a scalar reward $u(a, z)$, subject to a long-term average cost constraint $\bar{C}_n \leq \gamma + o(1)$, where $\bar{C}_n$ is the average of a cost function $c(a, z)$.

Let

$$u^*_\gamma(q) = \max_{p \in \Delta(\mathcal{A})} \{u(p, q) : c(p, q) \leq \gamma\}.$$

The desired set of the pairs (reward, cost) and the response:

$$R^*(q) = \{r = (u, c) : u \geq u^*_\gamma(q), c \leq \gamma\},$$

$$p^*(q) \in \operatorname*{argmax}_{p \in \Delta(\mathcal{A})} \{u(p, q) : \ c(p, q) \leq \gamma\},$$

$$r(p^*(q), q) = (u(p^*(q), q), c(p^*(q), q)) \in R^*(q).$$

- $q \mapsto R^*(q)$ is non-convex due to the non-convexity of $u^*_\gamma(q)$,
- $R^*(\bar{q}_n)$ cannot be approached in general (Mannor et al., 2009).

Example: Constrained Regret Minimization

Maximize the average $\bar{U}_n$ of a scalar reward $u(a, z)$, subject to a long-term average cost constraint $\bar{C}_n \leq \gamma + o(1)$, where $\bar{C}_n$ is the average of a cost function $c(a, z)$.

Let

$$u^*_\gamma(q) = \max_{p \in \Delta(\mathcal{A})} \{u(p, q) : c(p, q) \leq \gamma\}.$$

The desired set of the pairs (reward, cost) and the response:

$$R^*(q) = \{r = (u, c) : u \geq u^*_\gamma(q), c \leq \gamma\},$$

$$p^*(q) \in \operatorname*{argmax}_{p \in \Delta(\mathcal{A})} \{u(p, q) : \; c(p, q) \leq \gamma\},$$

$$r(p^*(q), q) = (u(p^*(q), q), c(p^*(q), q)) \in R^*(q).$$

- $q \mapsto R^*(q)$ is non-convex due to the non-convexity of $u^*_\gamma(q)$,
- $R^*(\bar{q}_n)$ cannot be approached in general (Mannor et al., 2009).

Example: Constrained Regret Minimization

Maximize the average $\bar{U}_n$ of a scalar reward $u(a, z)$, subject to a long-term average cost constraint $\bar{C}_n \leq \gamma + o(1)$, where $\bar{C}_n$ is the average of a cost function $c(a, z)$.

Let

$$u^*_\gamma(q) = \max_{p \in \Delta(\mathcal{A})} \{u(p, q) : c(p, q) \leq \gamma\}.$$

The desired set of the pairs (reward, cost) and the response:

$$R^*(q) = \{r = (u, c) : u \geq u^*_\gamma(q), c \leq \gamma\},$$

$$p^*(q) \in \operatorname*{argmax}_{p \in \Delta(\mathcal{A})} \{u(p, q) : \ c(p, q) \leq \gamma\},$$

$$r(p^*(q), q) = (u(p^*(q), q), c(p^*(q), q)) \in R^*(q).$$

- $q \mapsto R^*(q)$ is non-convex due to the non-convexity of $u^*_\gamma(q)$,
- $R^*(\bar{q}_n)$ cannot be approached in general (Mannor et al., 2009).

Outline

Approachability Problem

A generalization of the general no-regret problem.

Approach S, namely bring the average payoff vector to a given target set S. If this can be done no matter what the opponent does, S is approachable.

- The generalized no-regret is equivalent to approachability of

$$S \triangleq \{v \triangleq (r, q) : r \in R^*(q)\}$$

with "lifted" vector-valued payoff function

$$v(a, z) = (r(a, z), 1(z)) \in \mathbb{R}^K \times \Delta(\mathcal{Z}),$$

$$v(p, q) = (r(p, q), q).$$

- S satisfies dual (necessary) Blackwell's condition for approachability: $\forall q \in \Delta(\mathcal{Z})\ \exists p \in \Delta(\mathcal{A}) : v(p, q) \in S$.
  - Note that this $p = p^*(q)$ is the one that satisfies $r(p, q) \in R^*(q)$.
  - We refer to such sets as D-sets.

Approachability Problem

A generalization of the general no-regret problem.

Approach S, namely bring the average payoff vector to a given target set S. If this can be done no matter what the opponent does, S is approachable.

- The generalized no-regret is equivalent to approachability of

$$S \triangleq \{v \triangleq (r, q) : r \in R^*(q)\}$$

with "lifted" vector-valued payoff function

$$v(a, z) = (r(a, z), 1(z)) \in \mathbb{R}^K \times \Delta(\mathcal{Z}),$$

$$v(p, q) = (r(p, q), q).$$

- S satisfies dual (necessary) Blackwell's condition for approachability: $\forall q \in \Delta(\mathcal{Z})\ \exists p \in \Delta(\mathcal{A}) : v(p, q) \in S$.
  - Note that this $p = p^*(q)$ is the one that satisfies $r(p, q) \in R^*(q)$.
  - We refer to such sets as D-sets.

Approachability Problem

A generalization of the general no-regret problem.

Approach S, namely bring the average payoff vector to a given target set S. If this can be done no matter what the opponent does, S is approachable.

- The generalized no-regret is equivalent to approachability of

$$S \triangleq \{v \triangleq (r, q) : r \in R^*(q)\}$$

with "lifted" vector-valued payoff function

$$v(a, z) = (r(a, z), 1(z)) \in \mathbb{R}^K \times \Delta(\mathcal{Z}),$$

$$v(p, q) = (r(p, q), q).$$

- S satisfies dual (necessary) Blackwell's condition for approachability: $\forall q \in \Delta(\mathcal{Z})\ \exists p \in \Delta(\mathcal{A}) : v(p, q) \in S$.
  - Note that this $p = p^*(q)$ is the one that satisfies $r(p, q) \in R^*(q)$.
  - We refer to such sets as D-sets.

Approachability Conditions

By standard results:

A convex set S is approachable if and only if it is a D-set.
Hence, the convex hull of a D-set S is approachable.

For the generalized no-regret problem:

$$\text{conv}(S) = \{v = (r, q) : r \in R^c(q)\}.$$

- $R^c(\cdot)$ is the convex hull of the multifunction $R^*(\cdot)$, namely the smallest convex multifunction that contains it.

More can be achieved if the opponent plays "regularly".

- E.g. uses a stationary strategy $q_n \equiv q_0$.

Online algorithm should capitalize on this regularity, approaching the original $R^*(q_0)$ (rather than the larger $R^c(q_0)$).
Standard approachability does not do this!

Approachability Conditions

By standard results:

A convex set S is approachable if and only if it is a D-set.
Hence, the convex hull of a D-set S is approachable.

For the generalized no-regret problem:

$$\mathrm{conv}(S) = \{v = (r, q) : r \in R^c(q)\}.$$

- $R^c(\cdot)$ is the convex hull of the multifunction $R^*(\cdot)$, namely the smallest convex multifunction that contains it.

More can be achieved if the opponent plays "regularly".

- E.g. uses a stationary strategy $q_n \equiv q_0$.

Online algorithm should capitalize on this regularity, approaching the original $R^*(q_0)$ (rather than the larger $R^c(q_0)$).
Standard approachability does not do this!

Approachability Conditions

By standard results:

A convex set S is approachable if and only if it is a D-set.
Hence, the convex hull of a D-set S is approachable.

For the generalized no-regret problem:

$$\text{conv}(S) = \{v = (r, q) : r \in R^c(q)\}.$$

- $R^c(\cdot)$ is the convex hull of the multifunction $R^*(\cdot)$, namely the smallest convex multifunction that contains it.

More can be achieved if the opponent plays "regularly".

- E.g. uses a stationary strategy $q_n \equiv q_0$.

Online algorithm should capitalize on this regularity, approaching the original $R^*(q_0)$ (rather than the larger $R^c(q_0)$).
Standard approachability does **not** do this!

Outline

Calibrated Forecasts

- A forecaster specifies a probabilistic forecast $y_n \in \Delta(\mathcal{Z})$ of the opponent's action z_n.

Definition (Calibrated Forecaster)

A forecaster is calibrated if for every Borel measurable set $Q \subseteq \Delta(\mathcal{Z})$ and every strategy of the opponent, it holds that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}\{y_k \in Q\}(1(z_k) - y_k) = 0, \quad \text{a.s.}$$

- A randomized forecaster specifies a probability measure η_n over $\Delta(\mathcal{Z})$. Then, $y_n \sim \eta_n$.
- Only randomized forecasters can be calibrated for all possible sequences. (Dawid, 1985)

Calibrated Forecasts

- A forecaster specifies a probabilistic forecast $y_n \in \Delta(\mathcal{Z})$ of the opponent's action z_n.

Definition (Calibrated Forecaster)

A forecaster is calibrated if for every Borel measurable set $Q \subseteq \Delta(\mathcal{Z})$ and every strategy of the opponent, it holds that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}\{y_k \in Q\}(1(z_k) - y_k) = 0, \quad \text{a.s.}$$

- A randomized forecaster specifies a probability measure η_n over $\Delta(\mathcal{Z})$. Then, $y_n \sim \eta_n$.
- Only randomized forecasters can be calibrated for all possible sequences. (Dawid, 1985)

Calibrated Forecasts

- A forecaster specifies a probabilistic forecast $y_n \in \Delta(\mathcal{Z})$ of the opponent's action z_n.

Definition (Calibrated Forecaster)

A forecaster is calibrated if for every Borel measurable set $Q \subseteq \Delta(\mathcal{Z})$ and every strategy of the opponent, it holds that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}\{y_k \in Q\}(1(z_k) - y_k) = 0, \quad \text{a.s.}$$

- A randomized forecaster specifies a probability measure η_n over $\Delta(\mathcal{Z})$. Then, $y_n \sim \eta_n$.
- Only randomized forecasters can be calibrated for all possible sequences. (Dawid, 1985)

The Algorithm

Recall: we have a response function $p^*(q)$ such that

$$r^*(q) \triangleq r(p^*(q), q) \in R^*(q).$$

The Calibrated Approachability Algorithm

At each time step n use the mixed action p_n which is specified by

$$p_n = p^*(y_n),$$

where y_n is the calibrated forecast at time n.

Previously proposed by Perchet (2009), but was not analyzed in the context of opportunistic approachability.

$$\bar{R}_n \to \frac{1}{n}\sum_{k=1}^{n} r^*(y_k), \quad \bar{q}_n \to \frac{1}{n}\sum_{k=1}^{n} y_k \quad \textbf{(from calibration)}$$

$$\Rightarrow \quad \frac{1}{n}\sum_{k=1}^{n} r^*(y_k) \to R^c(\bar{q}_n) \quad \Rightarrow \quad \bar{R}_n \to R^c(\bar{q}_n)$$

The Algorithm

Recall: we have a response function $p^*(q)$ such that

$$r^*(q) \triangleq r(p^*(q), q) \in R^*(q).$$

The Calibrated Approachability Algorithm

At each time step n use the mixed action p_n which is specified by

$$p_n = p^*(y_n),$$

where y_n is the calibrated forecast at time n.

Previously proposed by Perchet (2009), but was not analyzed in the context of opportunistic approachability.

$$\bar{R}_n \to \frac{1}{n}\sum_{k=1}^{n} r^*(y_k), \quad \bar{q}_n \to \frac{1}{n}\sum_{k=1}^{n} y_k \quad \textbf{(from calibration)}$$

$$\Rightarrow \quad \frac{1}{n}\sum_{k=1}^{n} r^*(y_k) \to R^c(\bar{q}_n) \quad \Rightarrow \quad \bar{R}_n \to R^c(\bar{q}_n)$$

The Algorithm

Recall: we have a response function $p^*(q)$ such that

$$r^*(q) \triangleq r(p^*(q), q) \in R^*(q).$$

The Calibrated Approachability Algorithm

At each time step n use the mixed action p_n which is specified by

$$p_n = p^*(y_n),$$

where y_n is the calibrated forecast at time n.

Previously proposed by Perchet (2009), but was not analyzed in the context of opportunistic approachability.

$$\bar{R}_n \to \frac{1}{n}\sum_{k=1}^{n} r^*(y_k), \quad \bar{q}_n \to \frac{1}{n}\sum_{k=1}^{n} y_k \quad \textbf{(from calibration)}$$

$$\Rightarrow \quad \frac{1}{n}\sum_{k=1}^{n} r^*(y_k) \to R^c(\bar{q}_n) \quad \Rightarrow \quad \bar{R}_n \to R^c(\bar{q}_n)$$

Outline

Desiderata for Opportunistic No-Regret Algorithm

Achieve the following two goals, simultaneously:

1. The convex hull of $R^*(\cdot)$, $R^c(\cdot)$, is approached, for any strategy of the opponent;
2. If the empirical frequencies of the opponent's actions are restricted (in the sense defined below), approach a strict subset of $R^c(\cdot)$. In particular, if the opponent is stationary, $R^*(\cdot)$ itself is approached.

Our Main Result

The Calibrated Approachability Algorithm does this! And more...

Desiderata for Opportunistic No-Regret Algorithm

Achieve the following two goals, simultaneously:

1. The convex hull of $R^*(\cdot)$, $R^c(\cdot)$, is approached, for any strategy of the opponent;
2. If the empirical frequencies of the opponent's actions are restricted (in the sense defined below), approach a strict subset of $R^c(\cdot)$. In particular, if the opponent is stationary, $R^*(\cdot)$ itself is approached.

Our Main Result

The Calibrated Approachability Algorithm does this! And more...

Desiderata for Opportunistic No-Regret Algorithm

Achieve the following two goals, simultaneously:

1. The convex hull of $R^*(\cdot)$, $R^c(\cdot)$, is approached, for any strategy of the opponent;
2. If the empirical frequencies of the opponent's actions are restricted (in the sense defined below), approach a strict subset of $R^c(\cdot)$. In particular, if the opponent is stationary, $R^*(\cdot)$ itself is approached.

Our Main Result

The Calibrated Approachability Algorithm does this! And more...

Desiderata for Opportunistic No-Regret Algorithm

Achieve the following two goals, simultaneously:

1. The convex hull of $R^*(\cdot)$, $R^c(\cdot)$, is approached, for any strategy of the opponent;
2. If the empirical frequencies of the opponent's actions are restricted (in the sense defined below), approach a strict subset of $R^c(\cdot)$. In particular, if the opponent is stationary, $R^*(\cdot)$ itself is approached.

Our Main Result

The Calibrated Approachability Algorithm does this! And more...

Restricted Opponent's Play

Statistically Q-Restricted Play

There exists a convex $Q \subseteq \Delta(\mathcal{Z})$ so that the sequence $\{q_n\}$ of the mixed actions of the opponent satisfies, for the given sample path,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} d\left(q_k, Q\right) = 0.$$

- A weakness in considering the mixed actions of the opponent:
  - They are not generally revealed (when its strategy is not known).
  - They may be meaningless (e.g., when the opponent is Nature).
- Need a weaker notion, in terms of the empirical frequencies of the pure actions.

Restricted Opponent's Play

Statistically Q-Restricted Play

There exists a convex $Q \subseteq \Delta(\mathcal{Z})$ so that the sequence $\{q_n\}$ of the mixed actions of the opponent satisfies, for the given sample path,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} d\left(q_k, Q\right) = 0.$$

- A weakness in considering the mixed actions of the opponent:
  - They are not generally revealed (when its strategy is not known).
  - They may be meaningless (e.g., when the opponent is Nature).
- Need a weaker notion, in terms of the empirical frequencies of the pure actions.

Restricted Opponent's Play

Statistically Q-Restricted Play

There exists a convex $Q \subseteq \Delta(\mathcal{Z})$ so that the sequence $\{q_n\}$ of the mixed actions of the opponent satisfies, for the given sample path,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} d\left(q_k, Q\right) = 0.$$

- A weakness in considering the mixed actions of the opponent:
  - They are not generally revealed (when its strategy is not known).
  - They may be meaningless (e.g., when the opponent is Nature).
- Need a weaker notion, in terms of the empirical frequencies of the pure actions.

Restricted Opponent's Play

Partition of the time axis into blocks of length τ_m, $m = 1, 2, \ldots$.

$$n_M = \sum_{m=1}^{M} \tau_m$$

Empirical distribution of the opponent's actions at block m $\hat{q}_m \in \Delta(\mathcal{Z})$:

$$\hat{q}_m(z) = \frac{1}{\tau_m} \sum_{k=n_{m-1}+1}^{n_m} \mathbb{I}\{z_k = z\}.$$

Empirically Q-Restricted Play w.r.t a Partition $\{\tau_m\}$

There exists a convex $Q \subseteq \Delta(\mathcal{Z})$ so that, for the given sample path,

$$\lim_{M\to\infty} \frac{1}{n_M} \sum_{m=1}^{M} \tau_m d\left(\hat{q}_m, Q\right) = 0.$$

Restricted Opponent's Play

Partition of the time axis into blocks of length τ_m, $m = 1, 2, \ldots$.

$$n_M = \sum_{m=1}^{M} \tau_m$$

Empirical distribution of the opponent's actions at block m $\hat{q}_m \in \Delta(\mathcal{Z})$:

$$\hat{q}_m(z) = \frac{1}{\tau_m} \sum_{k=n_{m-1}+1}^{n_m} \mathbb{I}\{z_k = z\}.$$

Empirically Q-Restricted Play w.r.t a Partition $\{\tau_m\}$

There exists a convex $Q \subseteq \Delta(\mathcal{Z})$ so that, for the given sample path,

$$\lim_{M \to \infty} \frac{1}{n_M} \sum_{m=1}^{M} \tau_m d\left(\hat{q}_m, Q\right) = 0.$$

Examples and Properties

- Consider binary sequences, and a singleton $Q = \{(0.5, 0.5)\}$.
  - 0101... is empirically Q-restricted w.r.t. any partition with fixed even blocks, or with strictly increasing blocks.
  - 01001100001111... is:
    - Empirically Q-restricted w.r.t. a partition with exponentially increasing blocks $\tau_m = 2^m$,
    - Not restricted on any partition with sub-exponentially increasing blocks,
    - $\bar{q}_n$ does not converge to Q.
- Lemma: If a sequence is statistically Q-restricted, it is also empirically Q-restricted w.r.t. any partition with super-logarithmical blocks.
- Lemma: If a sequence is empirically Q-restricted w.r.t. a partition with sub-exponential blocks, then $\bar{q}_n$ converges to Q.

Restricted Target Set

Let

$$r^*(q) = r(p^*(q), q) \in R^*(q)$$

denote the goal function under a given response function p^*.

Against Q-restricted play, we essentially require convergence of $\bar{R}_n$ to a restriction of $R^c(\bar{q}_n)$ to Q:

$$R_Q(\bar{q}_n) \triangleq \left\{ r = \sum_i \alpha_i r^*(q_i) : q_i \in Q, \sum_i \alpha_i q_i = \bar{q}_n, \sum_i \alpha_i = 1, \alpha_i \geq 0 \right\}.$$

Note that:

$$R_{\{q_0\}}(q) = \{r^*(q_0)\} \quad \Rightarrow \quad \bar{R}_n \to r^*(q_0) \in R^*(q_0),$$

$$R_{\Delta(\mathcal{Z})}(q) \subseteq R^c(q) \quad \Rightarrow \quad \bar{R}_n \to R^c(\bar{q}_n).$$

Restricted Target Set

Let

$$r^*(q) = r(p^*(q), q) \in R^*(q)$$

denote the goal function under a given response function p^*.

Against Q-restricted play, we essentially require convergence of $\bar{R}_n$ to a restriction of $R^c(\bar{q}_n)$ to Q:

$$R_Q(\bar{q}_n) \triangleq \left\{ r = \sum_i \alpha_i r^*(q_i) : q_i \in Q, \sum_i \alpha_i q_i = \bar{q}_n, \sum_i \alpha_i = 1, \alpha_i \geq 0 \right\}.$$

Note that:

$$R_{\{q_0\}}(q) = \{r^*(q_0)\} \quad \Rightarrow \quad \bar{R}_n \to r^*(q_0) \in R^*(q_0),$$

$$R_{\Delta(\mathcal{Z})}(q) \subseteq R^c(q) \quad \Rightarrow \quad \bar{R}_n \to R^c(\bar{q}_n).$$

Restricted Target Set

Let

$$r^*(q) = r(p^*(q), q) \in R^*(q)$$

denote the goal function under a given response function p^*.

Against Q-restricted play, we essentially require convergence of $\bar{R}_n$ to a restriction of $R^c(\bar{q}_n)$ to Q:

$$R_Q(\bar{q}_n) \triangleq \left\{ r = \sum_i \alpha_i r^*(q_i) : q_i \in Q, \sum_i \alpha_i q_i = \bar{q}_n, \sum_i \alpha_i = 1, \alpha_i \geq 0 \right\}.$$

Note that:

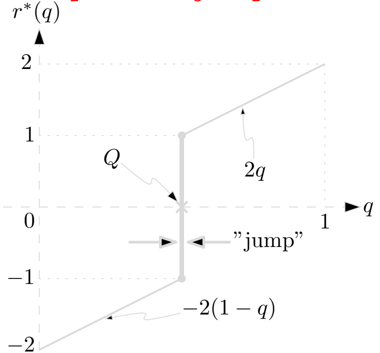
$$R_{\{q_0\}}(q) = \{r^*(q_0)\} \quad \Rightarrow \quad \bar{R}_n \to r^*(q_0) \in R^*(q_0),$$
$$R_{\Delta(\mathcal{Z})}(q) \subseteq R^c(q) \quad \Rightarrow \quad \bar{R}_n \to R^c(\bar{q}_n).$$

Restricted Target Set

Due to possible discontinuities in r^* on the boundary of Q, we need to slightly expand that definition, requiring convergence to

$$R_Q^+(\bar{q}_n) \triangleq \bigcap_{\epsilon>0} \left\{ r = \sum_i \alpha_i r^*(q_i) : d(q_i, Q) \leq \epsilon, \sum_i \alpha_i q_i = \bar{q}_n \right\}.$$

Contains the convex hull of all the points of $r^*(q)$, $q \in Q$, **together with possible jumps** in r^* on the boundary of Q.



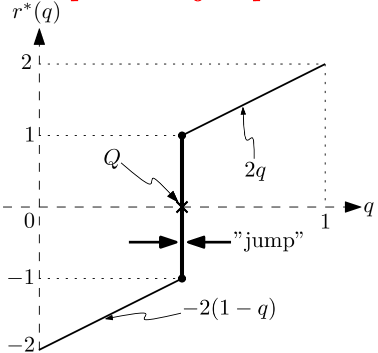
$$R_{\{q_0\}}^+(q) = \begin{cases} r^*(q_0), & q \neq 0.5 \\ [-1, 1], & q_0 = 0.5. \end{cases}$$

Restricted Target Set

Due to possible discontinuities in r^* on the boundary of Q, we need to slightly expand that definition, requiring convergence to

$$R_Q^+(\bar{q}_n) \triangleq \bigcap_{\epsilon>0} \left\{ r = \sum_i \alpha_i r^*(q_i) : d(q_i, Q) \le \epsilon, \sum_i \alpha_i q_i = \bar{q}_n \right\}.$$

Contains the convex hull of all the points of $r^*(q)$, $q \in Q$, **together with possible jumps** in r^* on the boundary of Q.



$$R^+_{\{q_0\}}(q) = \begin{cases} r^*(q_0), & q \neq 0.5 \\ [-1, 1], & q_0 = 0.5. \end{cases}$$

Opportunistic Strategies

Opportunistic strategy has no-regret with respect to $R_Q^+(q)$.

Statistically Opportunistic Approachability

A strategy is statistically opportunistic if

$$\lim_{n \to \infty} d\left(\bar{R}_n, R_Q^+(\bar{q}_n)\right) = 0$$

whenever the play of the opponent is statistically Q-restricted.

Empirically Opportunistic Approachability

A strategy is empirically opportunistic w.r.t. a partition $\{\tau_m\}$ if

$$\lim_{n \to \infty} d\left(\bar{R}_n, R_Q^+(\bar{q}_n)\right) = 0$$

whenever the play of the opponent is empirically Q-restricted w.r.t. $\{\tau_m\}$.

Opportunistic Strategies

Opportunistic strategy has no-regret with respect to $R_Q^+(q)$.

Statistically Opportunistic Approachability

A strategy is statistically opportunistic if

$$\lim_{n \to \infty} d\left(\bar{R}_n, R_Q^+(\bar{q}_n)\right) = 0$$

whenever the play of the opponent is statistically Q-restricted.

Empirically Opportunistic Approachability

A strategy is empirically opportunistic w.r.t. a partition $\{\tau_m\}$ if

$$\lim_{n \to \infty} d\left(\bar{R}_n, R_Q^+(\bar{q}_n)\right) = 0$$

whenever the play of the opponent is empirically Q-restricted w.r.t. $\{\tau_m\}$.

Main Results

Recall the Calibrated Approachability Algorithm (CAA):

$$p_n = p^*(y_n),$$

$y_n \sim \eta_n$ is the calibrated forecast at time n.

Theorem

(i) *CAA is statistically opportunistic.*

(ii) *CAA is empirically opportunistic if the probability distribution η_n of the employed calibrated forecast is changing slowly:*

$$\exists n_0 < \infty, \xi > 0, C < \infty : \quad \|\eta_n - \eta_{n-1}\|_{TV} \leq \frac{C}{n^{\xi}}, \quad n \geq n_0.$$

Theorem

There exist slowly varying calibration forecasters. In particular, the one that is based on internal regret minimization.

Main Results

Recall the Calibrated Approachability Algorithm (CAA):

$$p_n = p^*(y_n),$$

$y_n \sim \eta_n$ is the calibrated forecast at time n.

Theorem

(i) *CAA is statistically opportunistic.*

(ii) *CAA is empirically opportunistic if the probability distribution η_n of the employed calibrated forecast is changing slowly:*

$$\exists n_0 < \infty, \xi > 0, C < \infty : \quad \|\eta_n - \eta_{n-1}\|_{TV} \leq \frac{C}{n^{\xi}}, \quad n \geq n_0.$$

Theorem

There exist slowly varying calibration forecasters. In particular, the one that is based on internal regret minimization.

Main Results

Recall the Calibrated Approachability Algorithm (CAA):

$$p_n = p^*(y_n),$$

$y_n \sim \eta_n$ is the calibrated forecast at time n.

Theorem

(i) *CAA is statistically opportunistic.*

(ii) *CAA is empirically opportunistic if the probability distribution η_n of the employed calibrated forecast is changing slowly:*

$$\exists n_0 < \infty, \xi > 0, C < \infty : \quad \|\eta_n - \eta_{n-1}\|_{TV} \leq \frac{C}{n^\xi}, \quad n \geq n_0.$$

Theorem

There exist slowly varying calibration forecasters. In particular, the one that is based on internal regret minimization.

Proof Idea

Part (i) follows by showing that whenever the play of the opponent is statistically Q-restricted, so does the sequence of calibrated forecasts.

$$\bar{R}_n \to \frac{1}{n}\sum_{k=1}^{n} r^*(y_k), \quad \bar{q}_n \to \frac{1}{n}\sum_{k=1}^{n} y_k, \quad \frac{1}{n}\sum_{k=1}^{n} d(y_k, Q) \to 0$$

$$\Rightarrow \quad \frac{1}{n}\sum_{k=1}^{n} r^*(y_k) \to R_Q^+(\bar{q}_n) \quad \Rightarrow \quad \bar{R}_n \to R_Q^+(\bar{q}_n),$$

Part (ii) is obtained by showing that the calibration property implies a similar property in terms of the empirical frequencies of the actions provided that the distributions of calibrated forecasts are changing slowly, and by showing that

$$\bar{R}_{n_M} \to \frac{1}{n_M}\sum_{m=1}^{M} \tau_m r^*(\hat{q}_m)$$

Proof Idea

Part (i) follows by showing that whenever the play of the opponent is statistically Q-restricted, so does the sequence of calibrated forecasts.

$$\bar{R}_n \to \frac{1}{n}\sum_{k=1}^{n} r^*(y_k), \quad \bar{q}_n \to \frac{1}{n}\sum_{k=1}^{n} y_k, \quad \frac{1}{n}\sum_{k=1}^{n} d(y_k, Q) \to 0$$

$$\Rightarrow \quad \frac{1}{n}\sum_{k=1}^{n} r^*(y_k) \to R_Q^+(\bar{q}_n) \quad \Rightarrow \quad \bar{R}_n \to R_Q^+(\bar{q}_n),$$

Part (ii) is obtained by showing that the calibration property implies a similar property in terms of the empirical frequencies of the actions provided that the distributions of calibrated forecasts are changing slowly, and by showing that

$$\bar{R}_{n_M} \to \frac{1}{n_M}\sum_{m=1}^{M} \tau_m r^*(\hat{q}_m)$$

Proof Idea

Part (i) follows by showing that whenever the play of the opponent is statistically Q-restricted, so does the sequence of calibrated forecasts.

$$\bar{R}_n \to \frac{1}{n}\sum_{k=1}^{n} r^*(y_k), \quad \bar{q}_n \to \frac{1}{n}\sum_{k=1}^{n} y_k, \quad \frac{1}{n}\sum_{k=1}^{n} d(y_k, Q) \to 0$$

$$\Rightarrow \quad \frac{1}{n}\sum_{k=1}^{n} r^*(y_k) \to R_Q^+(\bar{q}_n) \quad \Rightarrow \quad \bar{R}_n \to R_Q^+(\bar{q}_n),$$

Part (ii) is obtained by showing that the calibration property implies a similar property in terms of the empirical frequencies of the actions provided that the distributions of calibrated forecasts are changing slowly, and by showing that

$$\bar{R}_{n_M} \to \frac{1}{n_M}\sum_{m=1}^{M} \tau_m r^*(\hat{q}_m)$$

Proof Idea

Part (i) follows by showing that whenever the play of the opponent is statistically Q-restricted, so does the sequence of calibrated forecasts.

$$\bar{R}_n \to \frac{1}{n}\sum_{k=1}^{n} r^*(y_k), \quad \bar{q}_n \to \frac{1}{n}\sum_{k=1}^{n} y_k, \quad \frac{1}{n}\sum_{k=1}^{n} d(y_k, Q) \to 0$$

$$\Rightarrow \quad \frac{1}{n}\sum_{k=1}^{n} r^*(y_k) \to R_Q^+(\bar{q}_n) \quad \Rightarrow \quad \bar{R}_n \to R_Q^+(\bar{q}_n),$$

Part (ii) is obtained by showing that the calibration property implies a similar property in terms of the empirical frequencies of the actions provided that the distributions of calibrated forecasts are changing slowly, and by showing that

$$\bar{R}_{n_M} \to \frac{1}{n_M}\sum_{m=1}^{M} \tau_m r^*(\hat{q}_m)$$

Computational Issues and Convergence Rates

- Calibration is hard! (Hazan & Kakade, 2012)
- ϵ-calibration (hence, ϵ-approachability) leads to $O(1/\sqrt{n})$ convergence rates.
- Efficiency was not the main point, but formulating the concept of opportunistic no-regret.
- Alternative that is efficient but less natural (Bernstein & Shimkin, 2013):
  - Partition the time into blocks.
  - At each block, run a standard no-regret algorithm with properly projected scalar reward function.

Computational Issues and Convergence Rates

- Calibration is hard! (Hazan & Kakade, 2012)
- ϵ-calibration (hence, ϵ-approachability) leads to $O(1/\sqrt{n})$ convergence rates.
- Efficiency was not the main point, but formulating the concept of opportunistic no-regret.
- Alternative that is efficient but less natural (Bernstein & Shimkin, 2013):
  - Partition the time into blocks.
  - At each block, run a standard no-regret algorithm with properly projected scalar reward function.

Computational Issues and Convergence Rates

- Calibration is hard! (Hazan & Kakade, 2012)
- ϵ-calibration (hence, ϵ-approachability) leads to $O(1/\sqrt{n})$ convergence rates.
- Efficiency was not the main point, but formulating the concept of opportunistic no-regret.
- Alternative that is efficient but less natural (Bernstein & Shimkin, 2013):
  - Partition the time into blocks.
  - At each block, run a standard no-regret algorithm with properly projected scalar reward function.

Computational Issues and Convergence Rates

- Calibration is hard! (Hazan & Kakade, 2012)
- ϵ-calibration (hence, ϵ-approachability) leads to $O(1/\sqrt{n})$ convergence rates.
- Efficiency was not the main point, but formulating the concept of opportunistic no-regret.
- Alternative that is efficient but less natural (Bernstein & Shimkin, 2013):
  - Partition the time into blocks.
  - At each block, run a standard no-regret algorithm with properly projected scalar reward function.

Main Result

(i) CAA is statistically opportunistic.

(ii) Suppose that the probability distribution η_n of the employed calibrated forecast is changing slowly. Namely, there exists $n_0 < \infty$ such that for all $n \geq n_0$,

$$\|\eta_n - \eta_{n-1}\|_{TV} \leq \frac{C}{n^{\xi}},$$

for some $\xi > 0$ and $C < \infty$. Then, the Calibrated Approachability Algorithm is empirically opportunistic, under either

(1) Bounded blocks $\tau_m \leq \bar{\tau} < \infty$, or
(2) Growing blocks $\tau_m = O(m^{\nu})$ with $\nu > 0$, under the condition that $\xi > \nu/(\nu + 1)$.

Back to Generalized No-Regret

- Vector valued rewards $r(a,z) \in \mathbb{R}^K$.
- For each mixed action q of the opponent, the agent has:
  - A desired reward set $R^*(q) \subset \mathbb{R}^K$.
  - A response $p = p^*(q)$ that satisfies $r(p,q) \in R^*(q)$.
  - Our main interest is in non-convex maps $q \mapsto R^*(q)$.
- If a strategy of the agent ensures that

$$\lim_{n\to\infty} d(\bar{R}_n, R^*(\bar{q}_n)) = 0$$

  (a.s.) for any strategy of the opponent, we say that it is a no-regret or regret minimizing strategy with respect to $R^*(q)$.
- Equivalent to the approachability of a non-convex D-set

$$S = \{v = (r,q) : r \in R^*(q)\}.$$

$$\mathrm{conv}(S) = \{v = (r,q) : r \in R^c(q)\}.$$

Constrained Regret Minimization

- Repeated game model as before, except that we are given:
  1. A scalar reward (or utility) function $u : \mathcal{A} \times \mathcal{Z} \to \mathbb{R}$,
  2. A vector-valued cost function $c : \mathcal{A} \times \mathcal{Z} \to \mathbb{R}^s$, and
  3. A convex constraint set $\Gamma \subseteq \mathbb{R}^s$. E.g., linear constraints:

$$\Gamma = \{c \in \mathbb{R}^s : \; c_i \leq \gamma_i, \; i = 1, ..., s\}.$$

- Assume that Γ is not excludable: $\forall q \exists p : c(p, q) \in \Gamma$.
- The best-reward-in-hindsight:

$$u_\Gamma^*(q) \triangleq \max_{p \in \Delta(\mathcal{A})} \{u(p, q) \; : \; c(p, q) \in \Gamma\}.$$

Definition (Constrained no-regret)

A strategy of the agent is a constrained no-regret strategy with respect to a function u_Γ^* if for every strategy of the opponent:
(i) $\limsup_{n\to\infty} \left(u_\Gamma^*(\bar{q}_n) - \bar{U}_n\right) \leq 0$; and
(ii) $\lim_{n\to\infty} d(\bar{C}_n, \Gamma) = 0$.
If such a strategy exists, we say that $u_\Gamma^*(\cdot)$ is attainable.

Formulation as Generalized No-Regret Problem

- $r(a,z) \triangleq (u(a,z), c(a,z)) \in \mathbb{R}^{1+s}$.
- The desired reward set:

$$R^*(q) = \{r = (u,c) \in \mathbb{R}^{1+s} : u \geq u_\Gamma^*(q), c \in \Gamma\}.$$

Attainability of $u_\Gamma^*(q)$
$\Leftrightarrow \lim_{n\to\infty} d(\bar{R}_n, R^*(\bar{q}_n)) = 0$
$\Leftrightarrow$ Approachability of

$$S = \{v = (r,q) \in \mathbb{R}^{1+s} \times \Delta(\mathcal{Z}) : r \in R^*(q)\}$$

with the vector-valued reward function $v(a,z) = (r(a,z), \mathbf{1}(z))$.

- Mannor et al. (2009): S is not approachable as $u_\Gamma^*(q)$ is not convex. A feasible (approachable) target set is $\text{conv}(S) = \left\{(r,q) \in \mathbb{R}^{s+1} \times \Delta(\mathcal{Z}) : \; r \in R^c(q)\right\}$.
  - $R^c(q) = \{r = (u,c) \in \mathbb{R}^{1+s} : u \geq \text{conv}\,(u_\Gamma^*)\,(q), c \in \Gamma\}$
  - $\text{conv}\,(u_\Gamma^*)$ is the lower convex hull of $u_\Gamma^*(\cdot)$.

Application of Calibrated Approachability Algorithm

- A response function: any choice of

$$p^*(q) \in \underset{p \in \Delta(\mathcal{A})}{\operatorname{argmax}} \{u(p,q): \ c(p,q) \in \Gamma\}.$$

- The goal function

$$v^*(q) = (u^*_\Gamma(q), c(p^*(q),q), q).$$

- Property: the closed convex image of singeltons satisfies $V^+(\{q\}) \subseteq S$ (rather than $\mathrm{conv}(S)$) for every $q \in \Delta(\mathcal{Z})$.
- Our results apply, in particular:
  - The algorithm approaches $\mathrm{conv}(S)$, hence attains the relaxed goal function $\mathrm{conv}(u^*_\Gamma)$.
  - $u^*_\Gamma(q_0)$ itself is attained whenever the opponent is either statistically or empirically restricted to a singleton $Q = \{q_0\}$.