# Towards optimal algorithms for prediction with expert advice

Nick Gravin, Yuval Peres, Balasubramanian Sivan
Microsoft Research

# Prediction with expert advice

- Sequential decision process in adversarial setting
- For each time $t = 1$ to stopping time:
  - Player picks one of $k$ experts to follow, say $J = J(t)$
  - Adversary sets gain $g_{it} \in [0,1]$ for each expert $i$ (without knowing $J(t)$)
  - Player gains $g_{Jt}$; all gains are revealed to player

| | $1$ | $2$ | ..... | $t-1$ | $t$ |
|---|---|---|---|---|---|
|  | $g_{11}$ | $g_{12}$ | ..... | $g_{1,t-1}$ | $\overset{?}{g_{1t}}$ |
|  | $g_{21}$ | $g_{22}$ | ..... | $g_{2,t-1}$ | $\overset{?}{g_{2t}}$ |

# Finite and Geometric Horizons

- Finite horizon: Stopping time is $T$
- Geometric horizon: At each step, stop with probability $\delta$
  - Stopping time is geometric with mean $\frac{1}{\delta}$
  - Equivalent to time discounted future

# Strategies

- $\boldsymbol{g}_{[0,t-1]}$: Vector of gains for all steps before $t$

- $G_i(t) = \sum_{s=1}^{t} g_{is}$ (cumulative gains)

- Adversary strategy: A distribution $D_t$ for $\boldsymbol{g}_t \in [0,1]^k$ (depending on $\boldsymbol{g}_{[0,t-1]}$)
  - Binary adversaries are most powerful: Restrict to $\boldsymbol{g}_t \in \{0,1\}^k$

- Player strategy: A distribution $A_t$ for $J_t \in \{1, \dots, k\}$ (depending on $\boldsymbol{g}_{[0,t-1]}$)
  - Player's gain at time $t$ is $g_{J_t,t}$

Regret: $R_T(D, A) = \boldsymbol{E}[\max_{i \in [k]} G_i(T) - \sum_{t=1}^{T} g_{J_t,t}]$

# Minimax regret

- Worst-case regret for $A$: $\max_D R_T(D, A)$
- Minimax regret: $\min_A \max_D R_T(D, A)$
- von Neumann's Minimax theorem

$$\min_A \max_D R_T(D, A) = \max_D \min_A R_T(D, A)$$

- Randomization is crucial!
  - deterministic player will get -1 payoff
  - 50/50 randomization will get 0 payoff

| Row \ Column | Heads | Tails |
|---|---|---|
| Heads | (1, −1) | (−1, 1) |
| Tails | (−1, 1) | (1, −1) |

# What's known?

<u>Multiplicative weights algorithm</u>:

- Given cumulative gains $G_1(t-1), \ldots, G_k(t-1)$, follow expert $i$ at $t$ with probability $\frac{e^{\eta G_i(t-1)}}{\sum_j e^{\eta G_j(t-1)}}$, where $\eta = \sqrt{\frac{8 \ln k}{T}}$

- Multiplicative weights algorithm yields regret at most $\sqrt{\frac{T \ln k}{2}}$
- Cesa-Bianchi, Freund, Haussler, Helmbold, Schapire, Warmuth (1997)

- Asymptotically optimal as $T, k \to \infty$

# The question

For a constant number of experts:

1. What is the optimal algorithm?

2. What is the optimal adversary?

3. What is the optimal regret value?

# Two experts (Cover'1965)

- Optimal adversary: Advance expert 1 alone w.p. ½ and expert 2 alone w.p. ½ (experts always disagree)

| | 1 | 2 | … | $t-1$ | $t$ | |
|---|---|---|---|---|---|---|
|  | 1 | 0 | … | 0 | 1 | |
|  | 0 | 1 | … | 1 | 0 | |

- Optimal regret (finite horizon): Optimal regret $R_T$ is precisely half the expected distance travelled by a simple random walk in $T$ steps.
  - As $T \to \infty$, the optimal regret $R_T \sim \sqrt{\frac{T}{2\pi}}$
- Optimal regret (geometric horizon): Optimal regret $R_\delta$ is $\frac{1-\delta}{2\sqrt{1-(1-\delta)^2}}$
  - As $\delta \to 0$, the optimal regret $R_\delta \sim \frac{1}{2}\sqrt{\frac{1}{2\delta}}$

# Our results (2 experts, geometric horizon)

Convention: Number experts in descending order of cumulative gains

$\xi = P$[Reaching 0, starting at 1]

Uni ... erts (geometric horizon):

1. Initialize $\xi = \frac{1-\sqrt{1-(1-\delta)^2}}{1-\delta} \sim 1-\sqrt{2\delta}$ as $\delta \to 0$
2. For each $t$ till the game stops, do:
   - Compute cumulative gains $G_1(t-1)$ and $G_2(t-1)$
   - Let $d = G_1(t-1) - G_2(t-1)$
   - Follow the leading expert with probability $p_1(d) = 1 - \frac{1}{2}\xi^d$
   - Follow the lagging expert with probability $p_2(d) = \frac{1}{2}\xi^d$

$p_2(d) = \boldsymbol{P}$[Lagging expert finishes as leading expert]

# 2 experts, geometric horizon

Unique optimal algorithm for two experts (geometric horizon):

For each $t$ till the game stops, do:

- Follow laggard with the probability he finishes as leader
  - Depends on $d$
- Follow leader with remaining probability

# 2 experts, finite horizon

Unique optimal algorithm for two experts (finite horizon):

For each $t = 1 \dots T$, do:

- Follow laggard with the probability he finishes as leader
  - Depends on $d$ and $T - t$
- Follow leader with remaining probability

# Comparison with multiplicative weights

Optimal algorithm for two experts (geometric horizon):

- Follow the leading expert with probability $p_1(d) = 1 - \frac{1}{2}\xi^d$
- Follow the lagging expert with probability $p_2(d) = \frac{1}{2}\xi^d$

Multiplicative weights algorithm for two experts (geometric horizon):

- Follow the leading expert with probability $p_1(d) = \frac{e^{\eta d}}{e^{\eta d}+1}$
- Follow the lagging expert with probability $p_2(d) = \frac{1}{e^{\eta d}+1}$

- Optimal algorithm cannot be expressed as a MWA
- MWA's known regret of $\sqrt{\frac{T \ln 2}{2}}$ is 47.5% larger (prove a tight lower bound)

# Lower bound for multiplicative weights

$k$ = Number of experts

Finite horizon: $T$ = No. of steps

Geometric horizon: $\delta = P$[Stopping in any given round]

- What was known: As $T \to \infty, k \to \infty$, MWA regret $\sim \sqrt{\frac{T \ln k}{2}}$

  As $\delta \to 0, k \to \infty$, MWA regret $\sim \sqrt{\frac{\ln k}{2\delta}}$

- We show:
  - As $T \to \infty$, MWA regret $\geq \frac{1}{2}\sqrt{\frac{T \ln k}{2}}$ for every $k$
  - As $\delta \to 0$, MWA regret $\geq \frac{1}{2}\sqrt{\frac{\ln k}{2\delta}}$ for every $k$
  - MWA's regret for $k = 2$ is more than 10% larger than optimal regret

# Our results (three experts)

- Optimal adversary (geometric horizon, $\delta \to 0$):
  - Advance experts 1 and 3 (leading and lagging) together w.p. $\frac{1}{2}$
  - Advance expert 2 (middle) w.p. $\frac{1}{2}$

| | 1 | 2 | … | $t-1$ | $t$ | |
|---|---|---|---|---|---|---|
|  | 1 | 0 | | 0 | | |
|  | 0 | 1 | | 1 | | |
|  | 1 | 0 | | 0 | | |

# Our results (three experts)

- Optimal adversary (geometric horizon, $\delta \to 0$):
  - Advance experts 1 and 3 (leading and lagging) together w.p. $\frac{1}{2}$
  - Advance expert 2 (middle) w.p. $\frac{1}{2}$

- Optimal regret (geometric horizon): Optimal regret $R_\delta$ is $\frac{2}{3}\frac{1-\delta}{\sqrt{1-(1-\delta)^2}}$
  - As $\delta \to 0$, the optimal regret $R_\delta \sim \frac{2}{3}\sqrt{\frac{1}{2\delta}}$

# Our results (three experts)

$\xi = \boldsymbol{P}$[Reaching 0, starting at 1]

Op eometric horizon):

1. Initialize $\xi = \frac{1-\sqrt{1-(1-\delta)^2}}{1-\delta}$
2. For each $t$ till the game stops, do:
   - Compute cumulative gains $G_i(t-1)$ for $i = 1, 2, 3$
   - Let $d_{ij} = G_i - G_j$
   - Follow leading expert w.p. $p_1(\boldsymbol{d}) = 1 - \frac{1}{2}\xi^{d_{12}} - \frac{1}{2}p_3(\boldsymbol{d})$
   - Follow middle expert w.p. $p_2(\boldsymbol{d}) = \frac{1}{2}\xi^{d_{12}} - \frac{1}{2}p_3(\boldsymbol{d})$
   - Follow lagging expert w.p. $p_3(\boldsymbol{d}) = \frac{1}{3}\xi^{d_{13}+d_{23}}$

$p_3(\boldsymbol{d}) = \boldsymbol{P}$[Lagging expert finishes as leading expert]

# Conjectured algorithm for $k$ experts

Optimal algorithm for $k$ experts;

For each $t = 1 \ldots$ stopping-time, do:

- Follow expert $i$ w.p. $\boldsymbol{P}$ [Expert $i$ finishes as leader]

Q: How to compute $\boldsymbol{P}$ [Expert $i$ finishes as leader]?

A: We need to know how the optimal adversary sets expert gains

Q: What is the optimal adversary?

A: Coming soon

# Connections between finite and geometric horizons

# Finite and geometric horizons

- Q: Suppose in the geometric horizon model, the precise stopping time is revealed to both player and adversary. Who benefits from this?
  - a) adversary benefits: $R_\delta \leq \sum_{T=0}^{\infty} \delta(1-\delta)^T R_T$ ? ✓
  - b) player benefits: $\geq$ ?
  - c) neither benefits: $=$ ?

- Conjecture: As $\delta \to 0$, neither benefits

$$R_\delta \sim \sum_{T=0}^{\infty} \delta(1-\delta)^T R_T \text{ as } \delta \to 0$$

- True for $k = 2$ for all $\delta$; Supported by simulations for larger $k$
- If true, $R_\delta \sim R_T \frac{\sqrt{\pi}}{2}$ as $\delta = \frac{1}{T} \to 0$

# Computing the optimal adversary

# Optimal adversary is balanced

Balanced adversary: The distribution $D_t$ advances all experts equally in expectation, irrespective of the history of cumulative gains

Proof:

1) If an optimal adversary is not balanced at $t$, the best-response algorithm for this adversary will follow the expert with the largest expected gain

2) At time $t$, increase the expected gains of all other experts to match the largest expected gain

3) This doesn't increase gain of the best-response algorithm ✗

4) This can potentially help the adversary because the "max expert"'s gain can only increase ✓

# Balanced adversary

- We saw: Optimal player forces optimal adversary to be balanced
- Against balanced adversary, regret is independent of player algorithm
- Computational device: Simple-minded algorithm that follows each expert w.p. $\frac{1}{k}$
- Regret: $R_T = \boldsymbol{E}\left[\max_{i\in[k]} G_i(T) - \frac{1}{k}\sum_{i=1}^{k} G_i(T)\right]$

max − average

# Why does adversary alone benefit?

- Q: Suppose in the geometric horizon model, the precise stopping time is revealed to both player and adversary. Who benefits from this?
  - a) adversary benefits: $R_\delta \leq \sum_{T=0}^{\infty} \delta(1-\delta)^T R_T$ ✓
  - b) player benefits: $\geq$
  - c) neither benefits: $=$

# Optimal adversary for $k = 2$

- Regret: $R_T = \boldsymbol{E}\left[\max(\mathrm{G_1(T), G_2(T)}) - \frac{G_1(T)+G_2(T)}{2}\right]$

$$= \boldsymbol{E}\left[\frac{|G_1(T)-G_2(T)|}{2}\right]$$

$$= \frac{1}{2}\boldsymbol{E}\left|\sum_{i=1}^{T} X_t\right|$$

where $X_t = g_{1t} - g_{2t} \in \{-1,0,1\}$

- $\boldsymbol{E}[X_t] = 0$ (Balanced adversary)

- Only choice: what is the probability of $X_t = 0$?

- To maximize expected absolute distance from origin this probability is 0

# Optimal adversary for $k = 2$ (Cover'65)

- Adversary's available actions: $\{1\}, \{2\}, \{12\}, \{\}$
- Goal: Construct a balanced distribution <u>at every step</u> over these 4 actions to optimize regret
- Actions $\{12\}$ and $\{\}$ result in $g_{1t} - g_{2t} = 0$, and should receive 0 probability

Optimal adversary:

- w.p. ½ advance expert 1 alone: $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$
- w.p. ½ advance expert 2 alone: $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$

| | 1 | 2 | … | $t-1$ | $t$ | |
|---|---|---|---|---|---|---|
|  | 1 | 0 | … | 0 | 1 | |
|  | 0 | 1 | … | 1 | 0 | |

# Optimal adversary for general $k$

- Adversary's available actions: All subsets of $\{1,2,\ldots,k\}$

- Goal: Construct a balanced distribution at every step over these $2^k$ actions to optimize regret

- At every step, we have a convex polytope of balanced distributions to pick from

- Exponentially many vertices for this polytope

# Optimal Adversary + Algorithm

# Optimal alg ($k = 2$, geometric horizon)

- Normalize cumulative gain of leading expert to be 0;
- Lagging expert's gain is $x \leq 0$
- Optimal algorithm's probabilities: $p_1(x), p_2(x)$
- $f(x)$: Max regret starting at $(0, x)$

# Optimal alg ($k = 2$, geometric horizon)

$$f(x) = \delta.0 + (1-\delta).\max \begin{cases} f(x-1) + 1 - p_1(x) & \text{when } \{1\} \\ f(x+1) - p_2(x) & \text{when } \{2\} \end{cases}$$

- Optimal alg makes the optimal adversary indifferent between {1} and {2}
- So $f(x) = (1-\delta)[\frac{f(x-1)+f(x+1)}{2}]$
- Solve the 2-step recurrence: $p_2(x) = \frac{1}{2}\xi^{-x} = 1 - p_1(x)$
- where $\xi \in [0,1]$ satisfies $\xi^2 - \frac{2}{1-\delta}\xi + 1 = 0$

# Optimal adversary for $k = 3$

- Adversary maximizes: $\mathbf{E}[\text{Max} - \text{average}]$
- $\mathbf{E}[\text{Max} - \text{average}]$ grows only when top $2$ experts collide
- How to maximize such collisions?
- Push expert $2$ up by collisions with expert $3$

# Optimal adversary for $k = 3$

- Adversary's available actions: $\{1\}$, $\{2\}$, $\{3\}$, $\{12\}$, $\{13\}$, $\{23\}$, $\{123\}$, $\{\}$
- Goal: Construct a balanced distribution at every step over these 8 actions to optimize regret

Optimal adversary (as $\delta \to 0$):

- Advance experts 1 and 3 w.p. 1/2
- Advance expert 2 w.p. 1/2

| | 1 | 2 | … | $t-1$ | $t$ | |
|---|---|---|---|---|---|---|
|  | 1 | 0 | | 1 | | |
|  | 0 | 1 | | 0 | | |
|  | 1 | 0 | | 1 | | |

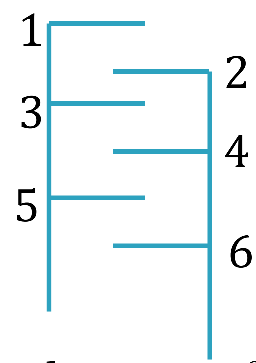
# Optimal adversary conj. for general $k$

- Adversary's available actions: all subsets of $\{1,2,\ldots,k\}$
- Goal: Construct a balanced distribution at every step over these $2^k$ actions to optimize regret

Optimal adversary conjecture (as $\delta \to 0$):

- Advance experts $\{1,3,5,\ldots\}$ w.p. $1/2$
- Advance experts $\{2,4,6,\ldots\}$ w.p. $1/2$

- "Comb" adversary:




- Generalizes the optimal adversary for $k = 2,3$
- Optimality supported by computer simulations

# Computing regret for comb adversary

# Our results (four experts)

- We compute the regret for the comb adversary for $k = 4$

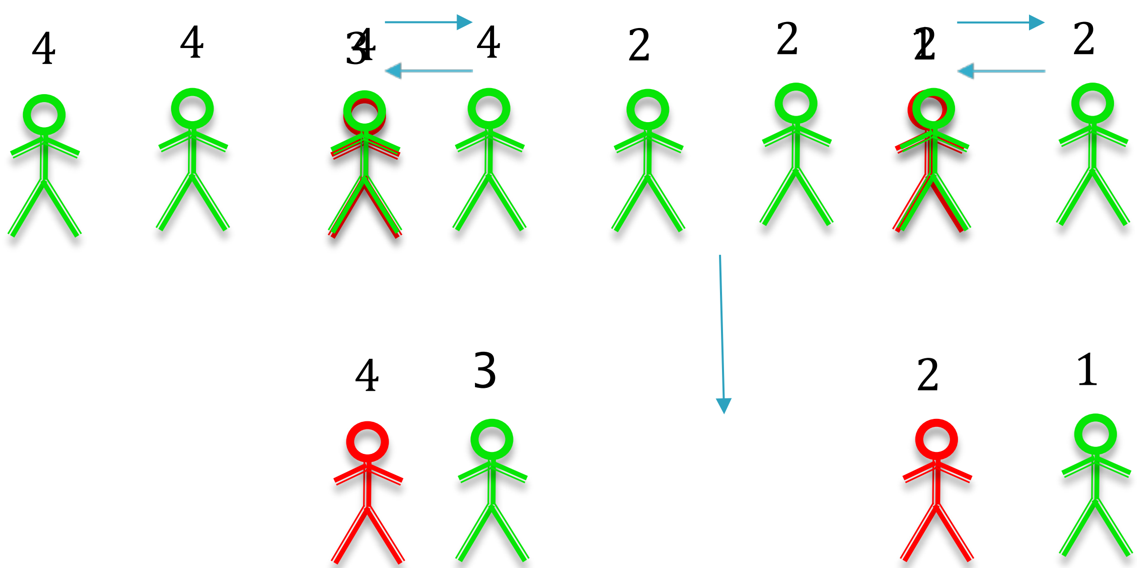
Comb adversary:
- Advance experts $\{1,3\}$ w.p. 1/2
- Advance experts $\{2,4\}$ w.p. 1/2

- Comb adversary's regret $\sim \frac{\pi}{4}\frac{1}{\sqrt{2\delta}}$
  - Conjecture: $\frac{\pi}{4}\frac{1}{\sqrt{2\delta}}$ is the asymptotically minimax regret

# Comb adversary for $k = 4$

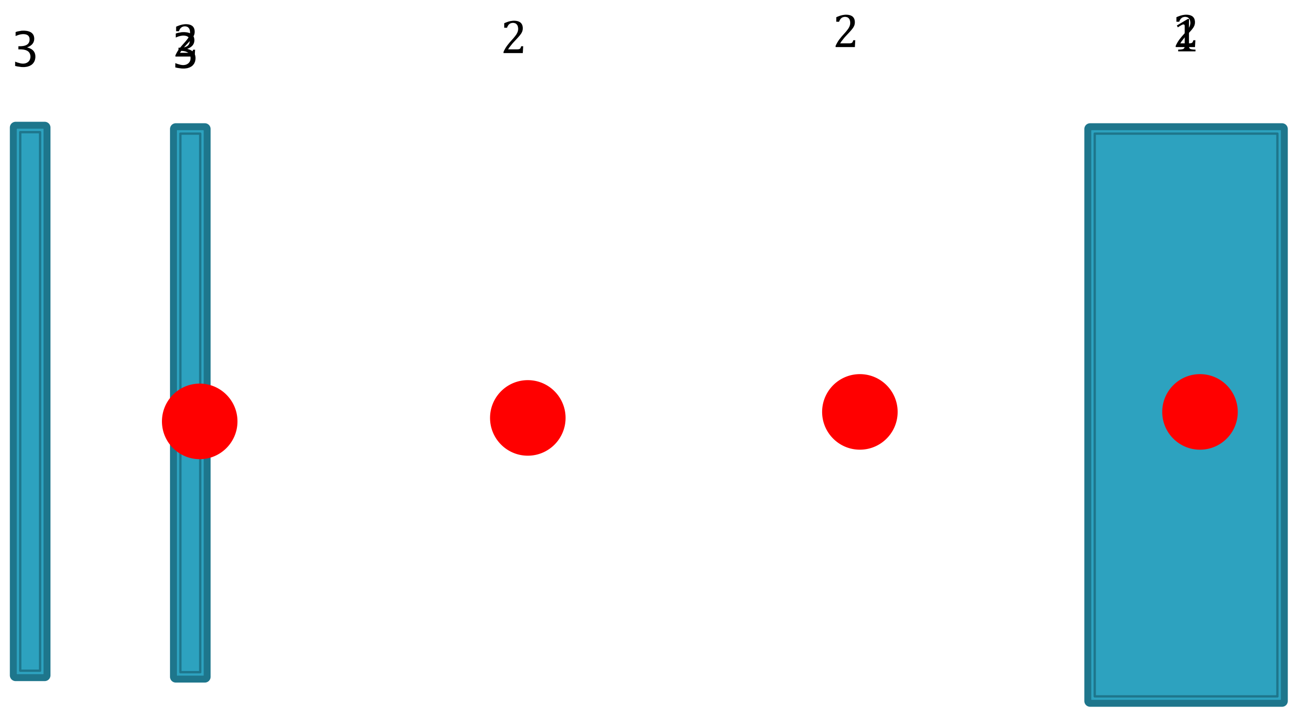
Comb adversary:

- Advance experts $\{1,3\}$ w.p. $1/2$
- Advance experts $\{2,4\}$ w.p. $1/2$



# Comb adversary for $k = 4$

- Distance between 1 and 3, is same as distance between 2 and 4
- Keep track of just 1, 2 and 3
- Mapping to random walk of a particle between a fixed reflecting wall and movable wall



# Comb adversary's regret

- Simulations: $R_\delta \sim R_\delta(\text{comb}) \sim \frac{0.785}{\sqrt{2\delta}}$
- Max−average increases in expectation by $\frac{1}{2}$ exactly when $W_1$ (fixed wall) and $P_2$ (the particle) coincide
- To compute: $\mathrm{R}_\delta(\text{comb}) = \frac{1}{2}\mathrm{E}[\#\text{ of visits of } P_2 \text{ to } W_1]$
- Let $\ell = \mathrm{dist}(W_1, W_3)$
- Random walk magic:

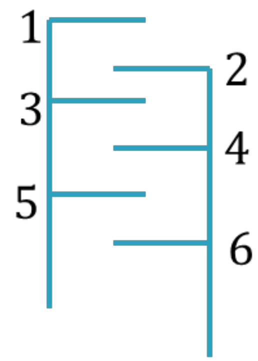
$$\mathrm{R}_\delta(\text{comb}) \sim \sum_{\ell=1}^{\infty} \frac{1}{\cosh^3(\ell\sqrt{2\delta})} \sim \frac{1}{\sqrt{2\delta}} \int_0^\infty \frac{dx}{\cosh^3(x)} = \frac{\pi}{4} \frac{1}{\sqrt{2\delta}}$$

- $\frac{\pi}{4} = 0.785\ldots$

# Conjectured optimal algorithm+adversary

Optimal adversary: Comb adversary?

- Advance experts $\{1,3,5,\ldots\}$ w.p. $1/2$
- Advance experts $\{2,4,6,\ldots\}$ w.p. $1/2$




Optimal algorithm: Probability matching?

For each $t$ till the game stops, do:

- Follow expert $i$ w.p. $\boldsymbol{P}_{\text{comb}}$ [Expert $i$ finishes as leader]

# Thanks for listening!