Online Learning
Lecture 4

Nicolò Cesa-Bianchi

Università degli Studi di Milano

Online convex optimization in the simplex

- Let $\mathbb{V}$ be the d-dimensional simplex Δ_d

Online convex optimization in the simplex

- Let $\mathbb{V}$ be the d-dimensional simplex Δ_d
- The loss at time t of $\boldsymbol{p}_t \in \Delta_d$ is $\boldsymbol{\ell}_t^\top \boldsymbol{p}_t = \mathbb{E}[\ell_t(I_t)]$ for $I_t \sim \boldsymbol{p}_t$

Online convex optimization in the simplex

- Let $\mathbb{V}$ be the d-dimensional simplex Δ_d
- The loss at time t of $\boldsymbol{p}_t \in \Delta_d$ is $\boldsymbol{\ell}_t^\top \boldsymbol{p}_t = \mathbb{E}[\ell_t(I_t)]$ for $I_t \sim \boldsymbol{p}_t$
- This is a linear loss with bounded coefficients $\ell_t(i) \in [0,1]$ for $i = 1, \ldots, d$

Prediction with expert advice

A sequential decision problem

- d actions
- Unknown deterministic assignment of losses to actions $\boldsymbol{\ell}_t = (\ell_t(1), \dots, \ell_t(d)) \in [0,1]^d$ for each time step t

For $t = 1, 2, \dots$

Prediction with expert advice

A sequential decision problem

- d actions
- Unknown deterministic assignment of losses to actions $\boldsymbol{\ell}_t = (\ell_t(1), \ldots, \ell_t(d)) \in [0,1]^d$ for each time step t

For $t = 1, 2, \ldots$

1. Player picks an action I_t (possibly using randomization) and incurs loss $\ell_t(I_t)$

Prediction with expert advice

A sequential decision problem

- d actions
- Unknown deterministic assignment of losses to actions $\boldsymbol{\ell}_t = (\ell_t(1), \dots, \ell_t(d)) \in [0,1]^d$ for each time step t

For $t = 1, 2, \dots$

1. Player picks an action I_t (possibly using randomization) and incurs loss $\ell_t(I_t)$
2. Player gets feedback information: $\ell_t(1), \dots, \ell_t(d)$

Regret

$$R_T = \sum_{t=1}^{T} \boldsymbol{\ell}_t^\top \boldsymbol{p}_t - \min_{\boldsymbol{p} \in \Delta_d} \sum_{t=1}^{T} \boldsymbol{\ell}_t^\top \boldsymbol{p}$$

Regret

$$R_T = \sum_{t=1}^{T} \boldsymbol{\ell}_t^\top \boldsymbol{p}_t - \min_{\boldsymbol{p}\in\Delta_d} \sum_{t=1}^{T} \boldsymbol{\ell}_t^\top \boldsymbol{p} = \mathbb{E}\left[\sum_{t=1}^{T} \ell_t(I_t)\right] - \min_{i=1,\dots,d} \sum_{t=1}^{T} \ell_t(i)$$

Regret

$$R_T = \sum_{t=1}^{T} \boldsymbol{\ell}_t^\top \boldsymbol{p}_t - \min_{\boldsymbol{p}\in\Delta_d} \sum_{t=1}^{T} \boldsymbol{\ell}_t^\top \boldsymbol{p} = \mathbb{E}\left[\sum_{t=1}^{T} \ell_t(I_t)\right] - \min_{i=1,\dots,d} \sum_{t=1}^{T} \ell_t(i)$$

Recall lower bound for the simplex: $R_T = \Omega\left(\sqrt{T \ln d}\right)$

Exponentially weighted forecaster (Hedge)

- Linear losses $\ell_t(\boldsymbol{p}_t) = \boldsymbol{\ell}_t^\top \boldsymbol{p}_t = \mathbb{E}[\ell_t(I_t)]$ for $I_t \sim \boldsymbol{p}_t$ and $\boldsymbol{\ell}_t \in [0,1]^d$

Exponentially weighted forecaster (Hedge)

- Linear losses $\ell_t(\boldsymbol{p}_t) = \boldsymbol{\ell}_t^\top \boldsymbol{p}_t = \mathbb{E}[\ell_t(I_t)]$ for $I_t \sim \boldsymbol{p}_t$ and $\boldsymbol{\ell}_t \in [0,1]^d$
- EG with linear losses is called the Hedge algorithm

$$p_{t+1}(i) = \frac{\exp\left(-\eta \sum_{s=1}^t \nabla \ell_s(\boldsymbol{p}_s)_i\right)}{\sum_{j=1}^d \exp\left(-\eta \sum_{s=1}^t \nabla \ell_s(\boldsymbol{p}_s)_j\right)} = \frac{\exp\left(-\eta \sum_{s=1}^t \ell_s(i)\right)}{\sum_{j=1}^d \exp\left(-\eta \sum_{s=1}^t \ell_s(j)\right)}$$

Exponentially weighted forecaster (Hedge)

- Linear losses $\ell_t(\boldsymbol{p}_t) = \boldsymbol{\ell}_t^\top \boldsymbol{p}_t = \mathbb{E}[\ell_t(I_t)]$ for $I_t \sim \boldsymbol{p}_t$ and $\boldsymbol{\ell}_t \in [0,1]^d$
- EG with linear losses is called the Hedge algorithm

$$p_{t+1}(i) = \frac{\exp\left(-\eta \sum_{s=1}^t \nabla \ell_s(\boldsymbol{p}_s)_i\right)}{\sum_{j=1}^d \exp\left(-\eta \sum_{s=1}^t \nabla \ell_s(\boldsymbol{p}_s)_j\right)} = \frac{\exp\left(-\eta \sum_{s=1}^t \ell_s(i)\right)}{\sum_{j=1}^d \exp\left(-\eta \sum_{s=1}^t \ell_s(j)\right)}$$

- At time $t+1$ draw action I_{t+1} from $\boldsymbol{p}_{t+1}$

Regret bound for Hedge

- If $\eta = \sqrt{\frac{\ln d}{8T}}$ then $R_T \le \sqrt{\frac{T \ln d}{2}}$

Regret bound for Hedge

- If $\eta = \sqrt{\frac{\ln d}{8T}}$ then $R_T \le \sqrt{\frac{T \ln d}{2}}$
- This matches the asymptotic lower bound, including constants

Regret bound for Hedge

- If $\eta = \sqrt{\frac{\ln d}{8T}}$ then $R_T \le \sqrt{\frac{T \ln d}{2}}$
- This matches the asymptotic lower bound, including constants
- We prove this later in a more general setting

The bandit problem: playing an unknown game

- d actions
- Unknown deterministic assignment of losses to actions $\boldsymbol{\ell}_t = (\ell_t(1), \dots, \ell_t(d)) \in [0,1]^d$ for each time step t

? ? ? ? ? ? ? ? ? ?

For $t = 1, 2, \dots$

The bandit problem: playing an unknown game

- d actions
- Unknown deterministic assignment of losses to actions $\boldsymbol{\ell}_t = (\ell_t(1), \ldots, \ell_t(d)) \in [0,1]^d$ for each time step t

For $t = 1, 2, \ldots$

1. Player picks an action I_t (possibly using randomization) and incurs loss $\ell_t(I_t)$

The bandit problem: playing an unknown game

- d actions
- Unknown deterministic assignment of losses to actions $\boldsymbol{\ell}_t = (\ell_t(1), \dots, \ell_t(d)) \in [0,1]^d$ for each time step t

? .3 ? ? ? ? ? ? ? ?

For $t = 1, 2, \dots$

1. Player picks an action I_t (possibly using randomization) and incurs loss $\ell_t(I_t)$
2. Player gets feedback information: Only $\ell_t(I_t)$ is revealed

A growing range of applications

- Ad placement

A growing range of applications

- Ad placement
- Dynamic content/layout optimization

A growing range of applications

- Ad placement
- Dynamic content/layout optimization
- Real time bidding

A growing range of applications

- Ad placement
- Dynamic content/layout optimization
- Real time bidding
- Recommender systems

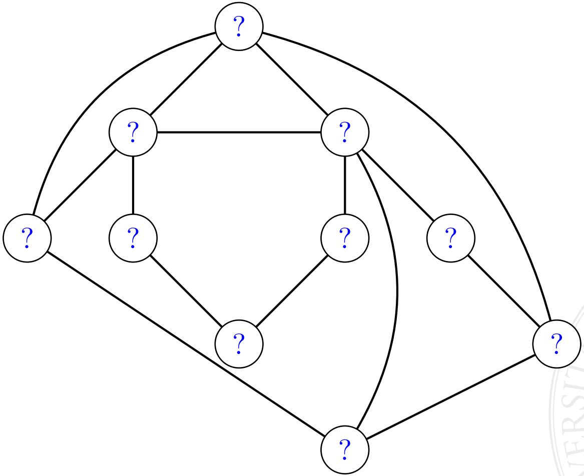
A growing range of applications

- Ad placement
- Dynamic content/layout optimization
- Real time bidding
- Recommender systems
- Clinical trials

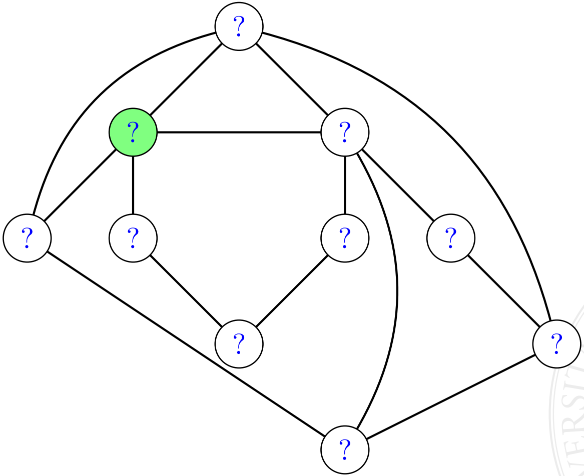
A growing range of applications

- Ad placement
- Dynamic content/layout optimization
- Real time bidding
- Recommender systems
- Clinical trials
- Network protocol optimization

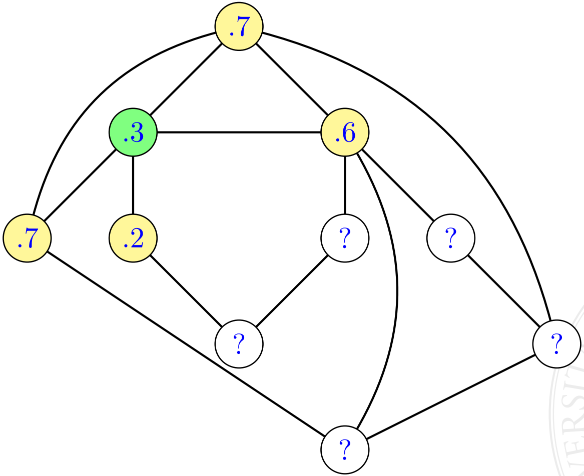
A feedback graph over actions



A feedback graph over actions



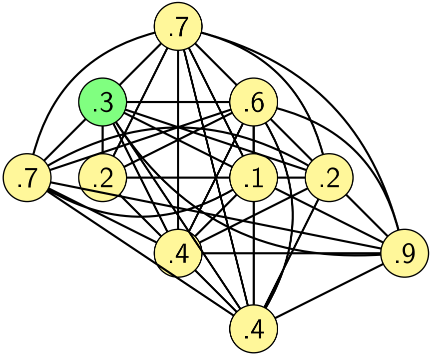
A feedback graph over actions



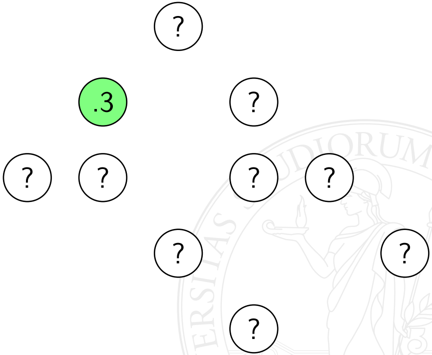
$\ell_t(i)$ is observed iff $I_t \in \{i\} \cup \mathcal{N}_G(i)$

Recovering expert and bandit settings

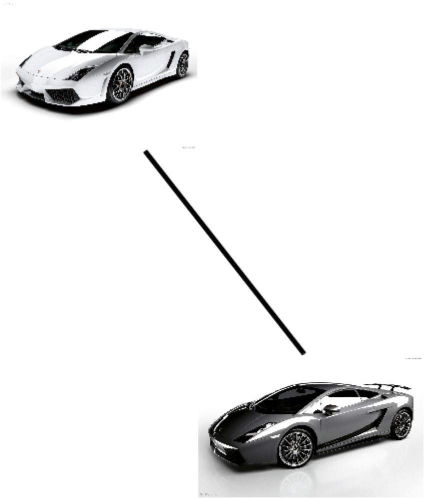
Experts: clique

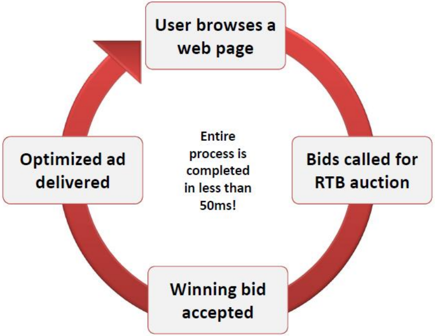


Bandits: edgeless graph



Relationships between actions





Hedge revisited on a feedback graph G

Player's strategy must use loss estimates

- $p_t(i) \quad \propto \quad \exp\left(-\eta \sum_{s=1}^{t-1} \widehat{\ell}_s(i)\right) \qquad i = 1, \ldots, d$

Hedge revisited on a feedback graph G

Player's strategy must use loss estimates

- $p_t(i) \quad \propto \quad \exp\left(-\eta \sum_{s=1}^{t-1} \widehat{\ell}_s(i)\right) \qquad i = 1, \dots, d$
- $\widehat{\ell}_t(i) = \begin{cases} \dfrac{\ell_t(i)}{\mathbb{P}_t\big(\ell_t(i) \text{ observed}\big)} & \text{if } \ell_t(i) \text{ is observed because } I_t \in \{i\} \cup \mathcal{N}_G(i) \\ 0 & \text{otherwise} \end{cases}$

Hedge revisited on a feedback graph G

Player's strategy must use loss estimates

- $p_t(i) \quad \propto \quad \exp\left(-\eta \sum_{s=1}^{t-1} \widehat{\ell}_s(i)\right) \qquad i = 1, \dots, d$
- $\widehat{\ell}_t(i) = \begin{cases} \dfrac{\ell_t(i)}{\mathbb{P}_t\big(\ell_t(i) \text{ observed}\big)} & \text{if } \ell_t(i) \text{ is observed because } I_t \in \{i\} \cup \mathcal{N}_G(i) \\ 0 & \text{otherwise} \end{cases}$

Importance sampling estimator

Hedge revisited on a feedback graph G

Player's strategy must use loss estimates

- $p_t(i) \quad \propto \quad \exp\left(-\eta \sum_{s=1}^{t-1} \widehat{\ell}_s(i)\right) \qquad i = 1, \dots, d$
- $\widehat{\ell}_t(i) = \begin{cases} \dfrac{\ell_t(i)}{\mathbb{P}_t(\ell_t(i) \text{ observed})} & \text{if } \ell_t(i) \text{ is observed because } I_t \in \{i\} \cup \mathcal{N}_G(i) \\ 0 & \text{otherwise} \end{cases}$

Importance sampling estimator

$$\mathbb{E}_t\left[\widehat{\ell}_t(i)\right] = \frac{\ell_t(i)}{\mathbb{P}_t(\ell_t(i) \text{ observed})} \times \mathbb{P}_t(\ell_t(i) \text{ observed}) + 0 = \ell_t(i)$$

$$\mathbb{E}_t\left[\widehat{\ell}_t(i)^2\right] = \frac{\ell_t(i)^2}{\mathbb{P}_t(\ell_t(i) \text{ observed})^2} \times \mathbb{P}_t(\ell_t(i) \text{ observed}) + 0 = \frac{\ell_t(i)^2}{\mathbb{P}_t(\ell_t(i) \text{ observed})}$$

Hedge revisited on a feedback graph G

Player's strategy must use loss estimates

- $p_t(i) \quad \propto \quad \exp\left(-\eta \sum_{s=1}^{t-1} \widehat{\ell}_s(i)\right) \qquad i = 1, \dots, d$
- $\widehat{\ell}_t(i) = \begin{cases} \dfrac{\ell_t(i)}{\mathbb{P}_t(\ell_t(i) \text{ observed})} & \text{if } \ell_t(i) \text{ is observed because } I_t \in \{i\} \cup \mathcal{N}_G(i) \\ 0 & \text{otherwise} \end{cases}$

Importance sampling estimator

$$\mathbb{E}_t\left[\widehat{\ell}_t(i)\right] = \frac{\ell_t(i)}{\mathbb{P}_t(\ell_t(i) \text{ observed})} \times \mathbb{P}_t(\ell_t(i) \text{ observed}) + 0 = \ell_t(i)$$

$$\mathbb{E}_t\left[\widehat{\ell}_t(i)^2\right] = \frac{\ell_t(i)^2}{\mathbb{P}_t(\ell_t(i) \text{ observed})^2} \times \mathbb{P}_t(\ell_t(i) \text{ observed}) + 0 \le \frac{1}{\mathbb{P}_t(\ell_t(i) \text{ observed})}$$

Regret analysis

$$\frac{W_{t+1}}{W_t} = \sum_{i=1}^{d} \frac{w_{t+1}(i)}{W_t} \qquad p_t(i) = \frac{1}{W_t} \exp\left(-\eta \sum_{s=1}^{t-1} \widehat{\ell}_s(i)\right) = \frac{w_t(i)}{W_t}$$

is a r.v.!

Regret analysis

$$\frac{W_{t+1}}{W_t} = \sum_{i=1}^{d} \frac{w_{t+1}(i)}{W_t} \qquad p_t(i) = \frac{1}{W_t}\exp\left(-\eta\sum_{s=1}^{t-1}\widehat{\ell}_s(i)\right) = \frac{w_t(i)}{W_t} \quad \text{is a r.v.!}$$

$$= \sum_{i=1}^{d} \frac{w_t(i)}{W_t}\exp\left(-\eta\,\widehat{\ell}_t(i)\right) \qquad \left(\text{because } w_{t+1}(i) = e^{-\eta\sum_{s=1}^{t-1}\hat{\ell}_s(i) - \eta\hat{\ell}_t(i)}\right)$$

Regret analysis

$$\frac{W_{t+1}}{W_t} = \sum_{i=1}^{d} \frac{w_{t+1}(i)}{W_t} \qquad p_t(i) = \frac{1}{W_t} \exp\left(-\eta \sum_{s=1}^{t-1} \widehat{\ell}_s(i)\right) = \frac{w_t(i)}{W_t} \quad \text{is a r.v.!}$$

$$= \sum_{i=1}^{d} \frac{w_t(i)}{W_t} \exp\left(-\eta\, \widehat{\ell}_t(i)\right) \qquad \left(\text{because } w_{t+1}(i) = e^{-\eta \sum_{s=1}^{t-1} \hat{\ell}_s(i) - \eta \hat{\ell}_t(i)}\right)$$

$$= \sum_{i=1}^{d} p_t(i) \exp\left(-\eta\, \widehat{\ell}_t(i)\right)$$

Regret analysis

$$\frac{W_{t+1}}{W_t} = \sum_{i=1}^{d} \frac{w_{t+1}(i)}{W_t} \qquad p_t(i) = \frac{1}{W_t} \exp\left(-\eta \sum_{s=1}^{t-1} \widehat{\ell}_s(i)\right) = \frac{w_t(i)}{W_t} \quad \text{is a r.v.!}$$

$$= \sum_{i=1}^{d} \frac{w_t(i)}{W_t} \exp\left(-\eta\, \widehat{\ell}_t(i)\right) \qquad \left(\text{because } w_{t+1}(i) = e^{-\eta \sum_{s=1}^{t-1} \hat{\ell}_s(i) - \eta \hat{\ell}_t(i)}\right)$$

$$= \sum_{i=1}^{d} p_t(i) \exp\left(-\eta\, \widehat{\ell}_t(i)\right)$$

$$\leq \sum_{i=0}^{d} p_t(i) \left(1 - \eta\, \widehat{\ell}_t(i) + \frac{\left(\eta\, \widehat{\ell}_t(i)\right)^2}{2}\right) \qquad \left(\text{using } e^{-x} \leq 1 - x + x^2/2 \text{ for all } x \geq 0\right)$$

Regret analysis

$$\frac{W_{t+1}}{W_t} = \sum_{i=1}^{d} \frac{w_{t+1}(i)}{W_t} \qquad p_t(i) = \frac{1}{W_t} \exp\left(-\eta \sum_{s=1}^{t-1} \widehat{\ell}_s(i)\right) = \frac{w_t(i)}{W_t} \quad \text{is a r.v.!}$$

$$= \sum_{i=1}^{d} \frac{w_t(i)}{W_t} \exp\left(-\eta\, \widehat{\ell}_t(i)\right) \qquad \left(\text{because } w_{t+1}(i) = e^{-\eta \sum_{s=1}^{t-1} \hat{\ell}_s(i) - \eta \hat{\ell}_t(i)}\right)$$

$$= \sum_{i=1}^{d} p_t(i) \exp\left(-\eta\, \widehat{\ell}_t(i)\right)$$

$$\le \sum_{i=0}^{d} p_t(i) \left(1 - \eta\, \widehat{\ell}_t(i) + \frac{\left(\eta\, \widehat{\ell}_t(i)\right)^2}{2}\right) \qquad \left(\text{using } e^{-x} \le 1 - x + x^2/2 \text{ for all } x \ge 0\right)$$

$$\le 1 - \eta \sum_{i=1}^{d} p_t(i) \widehat{\ell}_t(i) + \frac{\eta^2}{2} \sum_{i=1}^{d} p_t(i) \widehat{\ell}_t(i)^2$$

Regret analysis (cont.)

Taking logs, using $\sum_{t=1}^{T} \ln \frac{W_{t+1}}{W_t} = \ln \frac{W_{T+1}}{W_1}$ and $\ln(1+x) \le x$ yields

$$\ln \frac{W_{T+1}}{W_1} \le -\eta \sum_{t=1}^{T} \sum_{i=1}^{d} p_t(i) \widehat{\ell}_t(i) + \frac{\eta^2}{2} \sum_{t=1}^{T} \sum_{i=1}^{d} p_t(i) \widehat{\ell}_t(i)^2$$

Regret analysis (cont.)

Taking logs, using $\sum_{t=1}^{T} \ln \frac{W_{t+1}}{W_t} = \ln \frac{W_{T+1}}{W_1}$ and $\ln(1+x) \le x$ yields

$$\ln \frac{W_{T+1}}{W_1} \le -\eta \sum_{t=1}^{T} \sum_{i=1}^{d} p_t(i) \widehat{\ell}_t(i) + \frac{\eta^2}{2} \sum_{t=1}^{T} \sum_{i=1}^{d} p_t(i) \widehat{\ell}_t(i)^2$$

Moreover, for any fixed action k, we also have

$$\ln \frac{W_{T+1}}{W_1} \ge \ln \frac{w_{T+1}(k)}{W_1} = -\eta \sum_{t=1}^{T} \widehat{\ell}_t(k) - \ln d$$

Regret analysis (cont.)

Taking logs, using $\displaystyle\sum_{t=1}^{T} \ln \frac{W_{t+1}}{W_t} = \ln \frac{W_{T+1}}{W_1}$ and $\ln(1+x) \le x$ yields

$$\ln \frac{W_{T+1}}{W_1} \le -\eta \sum_{t=1}^{T} \sum_{i=1}^{d} p_t(i)\widehat{\ell}_t(i) + \frac{\eta^2}{2} \sum_{t=1}^{T} \sum_{i=1}^{d} p_t(i)\widehat{\ell}_t(i)^2$$

Moreover, for any fixed action k, we also have

$$\ln \frac{W_{T+1}}{W_1} \ge \ln \frac{w_{T+1}(k)}{W_1} = -\eta \sum_{t=1}^{T} \widehat{\ell}_t(k) - \ln d$$

Putting together and dividing both sides by $\eta > 0$ gives

$$\sum_{t=1}^{T} \sum_{i=1}^{d} p_t(i)\widehat{\ell}_t(i) - \sum_{t=1}^{T} \widehat{\ell}_t(k) \le \frac{\ln d}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \sum_{i=1}^{d} p_t(i)\widehat{\ell}_t(i)^2$$

Regret analysis (cont.)

Recall where we were:

$$\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\widehat{\ell}_t(i) - \sum_{t=1}^{T}\widehat{\ell}_t(k) \le \frac{\ln d}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\widehat{\ell}_t(i)^2$$

Regret analysis (cont.)

Recall where we were:

$$\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\widehat{\ell}_t(i) - \sum_{t=1}^{T}\widehat{\ell}_t(k) \le \frac{\ln d}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\widehat{\ell}_t(i)^2$$

Take expectation w.r.t. $I_1, \ldots, I_T$

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\mathbb{E}_t\left[\widehat{\ell}_t(i)\right] - \sum_{t=1}^{T}\mathbb{E}_t\left[\widehat{\ell}_t(k)\right]\right] \le \frac{\ln d}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\mathbb{E}_t\left[\widehat{\ell}_t(i)^2\right]\right]$$

Regret analysis (cont.)

Recall where we were:

$$\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\widehat{\ell}_t(i) - \sum_{t=1}^{T}\widehat{\ell}_t(k) \le \frac{\ln d}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\widehat{\ell}_t(i)^2$$

Take expectation w.r.t. $I_1, \dots, I_T$

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\mathbb{E}_t\left[\widehat{\ell}_t(i)\right] - \sum_{t=1}^{T}\mathbb{E}_t\left[\widehat{\ell}_t(k)\right]\right] \le \frac{\ln d}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\mathbb{E}_t\left[\widehat{\ell}_t(i)^2\right]\right]$$

Loss estimates are unbiased:

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\ell_t(i) - \sum_{t=1}^{T}\ell_t(k)\right] \le \frac{\ln d}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\mathbb{E}_t\left[\widehat{\ell}_t(i)^2\right]\right]$$

Regret analysis (cont.)

Recall where we were:

$$\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\widehat{\ell}_t(i) - \sum_{t=1}^{T}\widehat{\ell}_t(k) \le \frac{\ln d}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\widehat{\ell}_t(i)^2$$

Take expectation w.r.t. $I_1, \dots, I_T$

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\mathbb{E}_t\left[\widehat{\ell}_t(i)\right] - \sum_{t=1}^{T}\mathbb{E}_t\left[\widehat{\ell}_t(k)\right]\right] \le \frac{\ln d}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\mathbb{E}_t\left[\widehat{\ell}_t(i)^2\right]\right]$$

This is just the regret

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\ell_t(i) - \sum_{t=1}^{T}\ell_t(k)\right] \le \frac{\ln d}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\mathbb{E}_t\left[\widehat{\ell}_t(i)^2\right]\right]$$

Regret analysis (cont.)

$$R_T \le \frac{\ln d}{\eta} + \frac{\eta}{2}\,\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\mathbb{E}_t\left[\widehat{\ell}_t(i)^2\right]\right]$$

Regret analysis (cont.)

$$
\begin{aligned}
R_T &\le \frac{\ln d}{\eta} + \frac{\eta}{2}\,\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\mathbb{E}_t\left[\widehat{\ell}_t(i)^2\right]\right] \\
&\le \frac{\ln d}{\eta} + \frac{\eta}{2}\,\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} \frac{p_t(i)}{\mathbb{P}_t\big(\ell_t(i) \text{ is observed}\big)}\right] \qquad \text{(variance bound)}
\end{aligned}
$$

Regret analysis (cont.)

$$
\begin{aligned}
R_T &\le \frac{\ln d}{\eta} + \frac{\eta}{2}\,\mathbb{E}\left[\sum_{t=1}^T \sum_{i=1}^d p_t(i)\mathbb{E}_t\left[\widehat{\ell}_t(i)^2\right]\right] \\
&\le \frac{\ln d}{\eta} + \frac{\eta}{2}\,\mathbb{E}\left[\sum_{t=1}^T \sum_{i=1}^d \frac{p_t(i)}{\mathbb{P}_t\big(\ell_t(i) \text{ is observed}\big)}\right] && \text{(variance bound)} \\
&= \frac{\ln d}{\eta} + \frac{\eta}{2}\,\mathbb{E}\left[\sum_{t=1}^T \sum_{i=1}^d \frac{p_t(i)}{p_t(i) + \sum_{j\in\mathcal{N}_G(i)} p_t(j)}\right] && \text{(def. of feedback graph)}
\end{aligned}
$$

Regret analysis (cont.)

$$R_T \le \frac{\ln d}{\eta} + \frac{\eta}{2}\,\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} p_t(i)\mathbb{E}_t\left[\widehat{\ell}_t(i)^2\right]\right]$$

$$\le \frac{\ln d}{\eta} + \frac{\eta}{2}\,\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} \frac{p_t(i)}{\mathbb{P}_t(\ell_t(i) \text{ is observed})}\right] \qquad \text{(variance bound)}$$

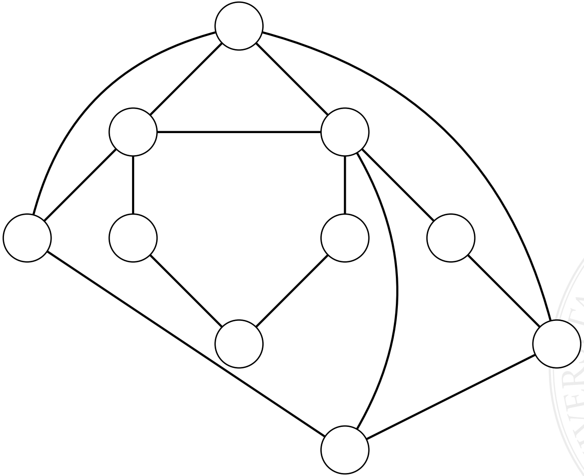
$$= \frac{\ln d}{\eta} + \frac{\eta}{2}\,\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} \frac{p_t(i)}{p_t(i) + \sum_{j\in\mathcal{N}_G(i)} p_t(j)}\right] \qquad \text{(def. of feedback graph)}$$

$$\le \frac{\ln d}{\eta} + \frac{\eta}{2} T\,\alpha(G) \qquad \text{(cool graph-theoretic fact)}$$

$\alpha(G)$ is the independence number of G

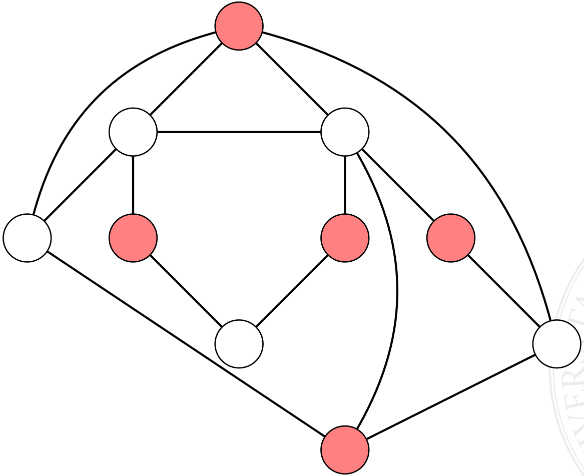
Independence number $\alpha(G)$

The size of the largest independent set in G



Independence number $\alpha(G)$

The size of the largest independent set in G



Regret bound

$$R_T \leq \frac{\ln d}{\eta} + \frac{\eta}{2} T \alpha(G)$$

Regret bound

$$R_T \leq \frac{\ln d}{\eta} + \frac{\eta}{2} T\alpha(G) = \sqrt{T\alpha(G)\ln d}$$

Regret bound

$$R_T \leq \frac{\ln d}{\eta} + \frac{\eta}{2} T \alpha(G) = \sqrt{T \alpha(G) \ln d}$$

Note: This bound is tight for all G (up to logarithmic factors)

Regret bound

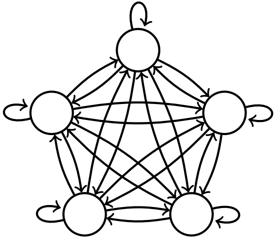
$$R_T \le \frac{\ln d}{\eta} + \frac{\eta}{2} T\alpha(G) = \sqrt{T\alpha(G)\ln d}$$

Note: This bound is tight for all G (up to logarithmic factors)

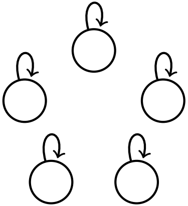
Special cases

Experts (clique):	$\alpha(G) = 1$	$R_T \le \sqrt{T \ln d}$	Hedge algorithm
Bandits (edgeless graph):	$\alpha(G) = d$	$R_T \le \sqrt{T d \ln d}$	Exp3 algorithm

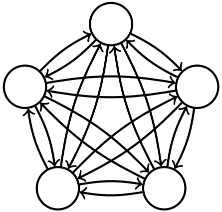
More general feedback models



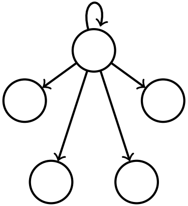
Experts



Bandits



Cops & Robbers



Revealing Action

Partial monitoring: not observing your own loss

Dynamic pricing: Perform as the best fixed price

1. Post a T-shirt price
2. Observe if next customer buys or not
3. Adjust price

Feedback does not reveal the player's loss

	1	2	3	4	5
1	0	1	2	3	4
2	c	0	1	2	3
3	c	c	0	1	2
4	c	c	c	0	1
5	c	c	c	c	0

Loss

	1	2	3	4	5
1	1	1	1	1	1
2	0	1	1	1	1
3	0	0	1	1	1
4	0	0	0	1	1
5	0	0	0	0	1

Feedback

A general gap theorem

- A constructive characterization of the minimax regret for any partial monitoring game

A general gap theorem

- A constructive characterization of the minimax regret for any partial monitoring game
- Only three possible rates for nontrivial games:

A general gap theorem

- A constructive characterization of the minimax regret for any partial monitoring game
- Only three possible rates for nontrivial games:
  1. Easy games (e.g., experts, bandits, cops & robbers): $\Theta(\sqrt{T})$

A general gap theorem

- A constructive characterization of the minimax regret for any partial monitoring game
- Only three possible rates for nontrivial games:
  1. Easy games (e.g., experts, bandits, cops & robbers): $\Theta(\sqrt{T})$
  2. Hard games (e.g., revealing action, dynamic pricing): $\Theta(T^{2/3})$

A general gap theorem

- A constructive characterization of the minimax regret for any partial monitoring game
- Only three possible rates for nontrivial games:
  1. Easy games (e.g., experts, bandits, cops & robbers): $\Theta(\sqrt{T})$
  2. Hard games (e.g., revealing action, dynamic pricing): $\Theta(T^{2/3})$
  3. Impossible games: $\Theta(T)$