

# Online Learning & Game Theory

## A quick overview with recent results

---

*Vianney Perchet*

Laboratoire Probabilités et Modèles Aléatoires  
Univ. Paris-Diderot

---

Journées MAS 2014  
27 Août 2014



# Starting Examples



Toulouse peintre



vianney.

**Web** Images Maps Shopping Plus ▾ Outils de recherche

Environ 7 810 000 résultats (0,21 secondes)

Annonce relative à **Toulouse peintre** ⓘ

**Peintre Toulouse - Peinture, Tadelakt, Chaux etc**

[www.peinture-decoration31.com/Peintre](http://www.peinture-decoration31.com/Peintre)

Devis Gratuit, Contactez-Nous

**Henri de Toulouse-Lautrec - Wikipédia**

[fr.wikipedia.org/wiki/Henri\\_de\\_Toulouse-Lautrec](http://fr.wikipedia.org/wiki/Henri_de_Toulouse-Lautrec)

Aller à **Peintures**: Portrait de Henri de **Toulouse-Lautrec** par Giovanni Boldini. Monsieur Boileau (1893). Maxime Dethomas (1896). Au Moulin de la ...

Jeunesse - Problème de santé et infirmité - Vie parisienne - Décès

**Emploi Peintre - Toulouse (31) - Travail | Indeed.fr**

[www.indeed.fr/Toulouse-\(31\)-Emplois-Peintre](http://www.indeed.fr/Toulouse-(31)-Emplois-Peintre)

32 offres d'emploi **Peintre - Toulouse (31)** sur indeed.fr. un clic. tous les emplois.

**Site du peintre Françoise Toulouse**

[www.francoisetoulouse.com/](http://www.francoisetoulouse.com/)

**Roger TOULOUSE**

[www.roger-toulouse.com/](http://www.roger-toulouse.com/)

Roger **TOULOUSE** (Orléans 1918-1994) : **Peintre**, Sculpteur, Illustrateur et Poète - Site officiel bilingue français / anglais.

**Toulouse Peinture - Rénovations - Façades - Finitions**

[toulouse-peinture.com/](http://toulouse-peinture.com/) - Traduire cette page

Site en construction.

Annonces ⓘ

**Entreprise de Peinture**

[atout.peinture.tis.free.fr/](http://atout.peinture.tis.free.fr/)

Confiez vos travaux de décoration à un pro.Travail Soigné,devis gratuit

**Toulouse Lautrec Peintre**

[www.ask.com/Toulouse+Lautrec+Peintre](http://www.ask.com/Toulouse+Lautrec+Peintre)

Cherchez **Toulouse Lautrec Peintre**

Trouvez Résultats Immédiatement!

**Peintre à Toulouse**

[www.prest-travaux.com/](http://www.prest-travaux.com/)

Obtenez Un Devis Négocié Pour Un **Peintre Sur Toulouse**.

15 rue Jean Bouin, SAINT JEAN  
0 805 40 16 56 - Itinéraire

**Artiste Peintre Toulouse**

[www.zapmeta.fr/Artiste+Peintre+Toulouse](http://www.zapmeta.fr/Artiste+Peintre+Toulouse)

Recherche: Artiste **Peintre Toulouse**  
Résultats de 6 Moteurs de Recherche

**Achetez Affiches d'Art**

[www.allposters.fr/](http://www.allposters.fr/)

Choisissez parmi 500.000 Affiches  
Petits Prix - Encadrement Possible

# Starting Examples

Gmail



COMPOSE

Inbox

Spam (640)

arxiv (28)



8192

Search people...

- Alexander Rakhlin  
Philly
- Alexandre Cham...
- Axelle Ziegler
- Eilon Solan
- julie billet
- Laurent Menard
- Luis Briceño
- Matias Nunez
- Nicolas Mahler
- Stephane Bouche...
- Anaïs HERVOUET
- Arvind Singh
- Clément ESCRIH...
- Corentin Piwtorak
- Elad Hazan
- Emile RICHARD

Delete all spam messages now (messages that have been in Spam more than 30 days will be auto

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	LeLynx.fr par Guide Comp.	Votre assurance voiture au meilleur prix! - Cliquez ici pour la Version en ligne Vou
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Votre Supermarché	Votre coupon de 50EUR à valoir pour tout achat au supermarché de votre choi
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1Mutuelle	La Mutuelle à petit prix - Vous recevez cet e-mail car vous êtes inscrit sur Plandu
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Confirmation de gain NZ	Votre Volkswagen Golf vous attend - Si vous avez des difficultés pour visualiser c
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Auchan.fr par FWU	Les Essentiels Auchan : jusqu'à moins 50 pourcent - Si le message ne s'affiche
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Devis santé	RE: Le meilleur tarif mutuelle - Si ce message ne s'affiche pas correctement, cons
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	La NET agence	Ouvrez votre compte : 1 an de gratuite - Agence en ligne de BNP Paribas Bonjour
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Conforama by Exclusivité.	Record à battre : un PC portable à 349e - Vianney Vous recevez cet email car vou
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Fortuneo par Domosource	Anais, votre carte bancaire a 0 euro et 80 euros de prime - Offre valable jusqu'au
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Xtra (2)	Nous vous offrons ce bon de reduction - Si vous avez des difficultés pour visualis
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Caves d'Alsace	Trésors du Vignoble alsacien - Si cet email ne s'affiche pas correctement, vous po
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Spartoo special rentree	Jusqu'à moins 35 pour cent sur notre selection rentree - Livraison OFFERTE -
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Formations English Town	Demandez votre guide pour parler en Anglais sereinement - Profitez enfin de vo
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	vertbaudet par Guide Com.	Pour la rentrée c'est jusqu'à -30 pourcent*, profitez-en ! - Cliquez ici pour la Vers
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Fan d'Expresso	A ne pas manquer: vos 60 caspules Expresso à 1 Euro - Si vous n'arrivez pas à
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Maisons du Monde - BSY	La rentree c'est le moment ideal pour refaire votre deco - Pour visualiser correcte
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Coffee Box	Capsules pour Nespresso + tasse design pour 1 euro - Si ce message ne s'affich
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Super U	remportez un bon d'achat de 500 euros - Vous recevez cet e-mail car vous êtes in
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Moins 3 kilos en 6 jours	Vianney Profitez de votre cadeau mineur - Vianney Vous recevez cet email car v

# Outline

- 1 First, in a **stochastic** environment (i.i.d. processes)
- 2 Then, in an **adversarial** environment (or individual sequences)
- 3 Finally, some links with **game theory**

# First Part

**Stochastic** environment



# Estimation of Means

$K = 2$  discrete-time proc.:  $X_n^{(1)}, X_n^{(2)}$  in  $[0, 1]$

“The payoff of the ad 1/2 on query  $n$ ”

**Estimate the means**  $\mu^{(1)}, \mu^{(2)}$

## Hoeffding inequality: exponential decay

$$\left| \bar{X}_n^{(k)} - \mu^k \right| > \varepsilon \text{ with proba at most } 2 \exp(-2n\varepsilon^2).$$

Finite number of mistakes:

$$\mathbb{E} \sum_{n \in \mathbb{N}} \mathbb{1} \left\{ \left| \bar{X}_n^{(k)} - \mu^k \right| > \varepsilon \right\} \leq \frac{1}{\varepsilon^2}$$

# Regret Minimization

- Choose **one** ad to display  $k_n$ . Reward:  $X_n^{(k_n)}$

Maximize cumulative reward  $\sum_{m=1}^n X_m^{(k_m)}$  or  $\sum_{m=1}^n \mu^{(k_m)}$

## Minimize Regret [Hannan'56]

$$R_n = n\mu^* - \sum_{m=1}^n \mu^{(k_m)}, \quad \text{with } \mu^* = \max\{\mu^{(1)}, \mu^{(2)}\}$$

- Equivalent formulation with  $\Delta = \mu^* - \mu^k$ :

$$R_n = \Delta \sum_{m=1}^n \mathbb{1}\{k_m \neq \star\}$$

# Stochastic & Full Monitoring

- Full Monitoring: all values  $X_n^{(1)}, X_n^{(2)}$  observed.
- Optimal algorithm:  $k_n = \arg \max \bar{X}_n^{(k)}$ :

$$\mathbb{E}R_n \leq \frac{1}{\Delta} \quad \text{and for small } n, \mathbb{E}R_N \leq n\Delta$$

**Bounded** regret, **uniformly** in  $n$ !

- **Given**  $n$ , worst  $\Delta$  is  $1/\sqrt{n}$  and  $\mathbb{E}R_n \leq \sqrt{n}$
- But in the examples, **only**  $X_n^{(k_n)}$  is observed (**bandit** monitoring)!



# Stochastic & Bandit Monitoring

- $\bar{X}_n^{(k)} = \frac{1}{n} \sum_{m=1}^n X_m^{(k)}$  **not** available, **only**  $\hat{X}_n^{(k)} = \frac{\sum_{m:k_m=k} X_m^{(k)}}{\#\{m : k_m = k\}}$
- with  $k_n = \arg \max \hat{X}_n^{(k)}$ ,  $\mathbb{E}R_n = \Theta(n)$ .
- **Balance exploitation** (play arg max) and **exploration** (play arg min) to get information

## Upper Confidence Bound [Auer,Cesa-Bianchi,Fischer'02]

$$k_n = \arg \max \hat{X}_n^{(k)} + \sqrt{\frac{2 \log(n)}{\#\{m : k_m = k\}}}$$

$$\mathbb{E}R_n \leq \square \frac{\log(n)}{\Delta}$$

# New policy: Explore Then Commit [P,Rigollet '13]

– Finite horizon  $N \in \mathbb{N}$  given.

1) Play alternatively arm 1 and 2 as long as

$$\left| \widehat{X}_n^{(1)} - \widehat{X}_n^{(2)} \right| \leq 2\sqrt{2\frac{\log(4N/n)}{n}}$$

2) Then play for ever the best arm.

→  $\mathbb{E}R_N \leq \square \frac{\log(N\Delta^2)}{\Delta}$  vs  $\frac{1}{\Delta}$  with Full Info

→ Worst case  $\Delta \simeq \frac{1}{\sqrt{N}}$

Full Monit & ETC:  $\sqrt{N}$  vs UCB:  $\sqrt{N}\log(N)$

## Bandit vs Full Monitoring

Logarithmic vs bounded regret;

same worst case

# Bounded Regret ?

[Lai,Robbins'84],[Bubeck,P,Rigollet '13]

- Without additional assumption, **No**: lower bound in  $\log(n)/\Delta$
- With any given **intermediate value**  $\mu^\sharp \in (\mu^{(1)}, \mu^{(2)})$ , **yes**:

- If  $\widehat{X}_n^{(1)}$  or  $\widehat{X}_n^{(2)}$  above  $\mu^\sharp$ , then  $k_n = \arg \max \widehat{X}_n^{(k)}$
- Otherwise play alternatively both arms.

$\widehat{X}_n^* < \mu^\sharp$  on  $\frac{1}{(\mu^* - \mu^\sharp)^2}$  stages (same argument for other arm).

- If  $\mu^*$  and  $\Delta$  known:  $\mathbb{E}R_n \leq \square \frac{1}{\Delta}$  as with Full Monit.
- If **only**  $\mu^*$  known:  $\mathbb{E}R_n \leq \square \frac{\log(1/\Delta^2)}{\Delta}$

## More General Frameworks & Results

Results in worst case (“distribution independent bounds”)

- **Multi-armed bandit.** [Auer,Cesa-Bianchi,Freund,Schapire'02],[Audibert,Bubeck'09]

$K > 2$  arms,  $\mathbb{E}R_n \leq \square \sqrt{Kn}$

- **Continuous bandit.** [Kleinberg'08],[Bubeck,Munos,Stoltz,Szepesvari'11]

Infinite set of arms,  $x \in [0, 1]^d$  and  $\mu(\cdot)$  Lipschitz.  $\mathbb{E}R_n \leq \square n^{\frac{d+1}{d+2}}$

- **Linear bandit**[Dani,Hayes,Kakade'08],[Zinkevich'02],[Abernethy,Hazan,Rakhlin'08]

$x \in [0, 1]^d$  and  $\mu(\cdot)$  Linear.  $\mathbb{E}R_n \leq \square \sqrt{n}$

- **Bandit with covariates (cf Google Example)** [P,Rigollet'13],[Bull'14]

Covariates  $\omega \in [0, 1]^d$ ,  $\mathbb{E}[X^{(k)}|\omega] = \mu^{(k)}(\omega)$  1-Lip.  $\mathbb{E}R_n \leq \square n^{\frac{d+1}{d+2}}$

- **Higher order bounds/small losses/sparsity**[Hazan,Kale'10], [Gershinovitz'13],

[Cappé, Garivier, Maillard, Munos, Stoltz'13], [Gaillard, Stoltz, van Erven'14]

$$\sqrt{n} \text{ vs } \sqrt{\sum_{m=1}^n \left( X_m^{(k_m)} - \mu^{(k_m)} \right)^2}, \sqrt{\sum_{m=1}^n \sum_{k=1}^K p_n^{(k)} \left( X_n^{(k)} \right)^2}$$

## Second Part

### Adversarial environment

What we have learned so far:

- In **worst case analysis**
  - Regret minimization in  $\square \sqrt{\log(K)n}$  with full monet
  - Up to  $\sqrt{K}$ , learning **as fast with bandit monet.** than with full monet.
- In **distribution dependent** (not worst case)
  - Bounded regret in  $\square \sum \frac{1}{\Delta_k}$
  - Additional assumption required to learn as fast in bandit monet

# Adversarial World

- In the examples, data are **not i.i.d.**. Spam senders can even **adapt to spam filters**, that is:

The law of  $X_{n+1}^{(k)}$  can depend on  $X_1^{(1)}, \dots, X_n^{(1)}, X_1^{(K)}, \dots, X_n^{(K)}$  but **even on** the previous choices  $k_1, \dots, k_n$ .

The environment can **adapt** and choose rewards **strategically**.

- Same def of regret (except argmax changes with time)

$$R_n = \max_k \sum_{m=1}^n X_m^{(k)} - \sum_{m=1}^n X_m^{(k_m)}$$

- Goal: a policy with sublinear regret  $o(n)$  against **ANY** possible strategy of the environment (in particular any sequences  $X_n^{(k)}$ )

# A Popular Algorithm with Full Monitoring

- With  $k_n = \operatorname{argmax} \bar{X}_n^{(k)}$ ,  $\mathbb{E}R_n = \Theta(n)$ .
- With any **deterministic** policy,  $\mathbb{E}R_n = \Theta(n)$ .

$k$  with proba  $\frac{\exp\left(\eta \sum_{m=1}^n X_m^{(k)}\right)}{\sum_{j=1}^K \exp\left(\eta \sum_{m=1}^n X_m^{(j)}\right)}$ ; temperature  $\eta \simeq \sqrt{\log(K)n}$

- Regret of “exponential weights” [Auer,Cesa-Bianchi,Freund,Schapire'02]

$$\mathbb{E}R_n \leq \square \sqrt{\log(K)n}, \quad \forall n \in \mathbb{N}$$

- Same dependency in  $n$  as worst case i.i.d., optimal in  $K$ .

# Optimality and Bandit Monitoring

- **Optimality:**  $\mathbb{E}R_n \geq \Omega(\sqrt{\log(K)n})$  if  $X_n^{(k)} = \pm 1$  w.p.  $1/2$

$$\mathbb{E} \sum_{m=1}^n X_m^{(k_m)} = 0 \text{ but } \mathbb{E} \max_k \sum_{m=1}^n X_m^{(k)} = \Omega(\sqrt{\log(K)n})$$

- **Bandit Monit.:**  $\tilde{X}_n^{(k)} = X_n^{(k)} \frac{\mathbb{1}\{k_n = k\}}{\mathbb{P}_n\{k_n = k\}}$  unbiased estim. of  $X_n^{(k)}$

“Exponential weights” w.r.t.  $\tilde{X}_n^{(k)}$ :  $\mathbb{E}R_n \leq \Omega(\sqrt{K \log(K)n})$

Remark: Optimal bounds are  $\Omega(\sqrt{Kn})$



## Discrete/Continuous Time

$$- \frac{\exp\left(\eta \sum_{m=1}^n X_m^{(k)}\right)}{\sum_{j=1}^K \exp\left(\eta \sum_{m=1}^n X_m^{(j)}\right)} = \nabla \Phi(V_n) := \frac{1}{\eta} \log \left( \sum_{k=1}^K \exp(\eta V_n^{(k)}) \right)$$

with  $V_n^{(k)} = \sum_{m=1}^n X_n^{(k)} - X_m^{(k)}$

Deterministic continuous approx. of stochastic discrete proc.

[Benaïm, Hofbauer, Sorin'06], [Benaïm, Faure'13]

$$- \mathbb{E}[V_{n+1}] - V_n = \left( X_{n+1}^{(k)} - \langle \nabla \Phi(V_n), X_{n+1} \rangle \right)_{k=1, \dots, K}$$

Stochastic Approx of  $\dot{V} \in F(V) := \left\{ U - \langle \nabla \Phi(V), U \rangle \vec{\mathbf{1}}; U \in R^K \right\}$

- Differential inclusion with Lyapounov function  $\Phi(V)$ :

$$\Phi(V)' = \langle \dot{V}, \Phi(V) \rangle = \left\langle U - \langle U, \nabla \Phi(V) \rangle \vec{\mathbf{1}}, \nabla \Phi(V) \right\rangle = 0$$

$$- \lim R_n \leq \lim V_n = V(+\infty) = V(0) = \log(d)/\eta$$

## Refined Regret: Internal-Swap-

- **Regret:** “As well as the best constant strategy”
- **Internal:** “On the stages where  $k_n = k$ ,  $k$  was the best choice”

[Foster, Vohra'99]

$$R_n^{\text{int}} = \max_k \left\{ \max_j \sum_{m:k_m=k} X_m^{(j)} - X_m^{(k)} \right\}$$

- **Swap:** “As well as  $\phi(k)$  instead of  $k$ ,  $\phi : [K] \rightarrow [K]$ ” [Blum, Mansour'07]

$$R_n^{\text{swap}} = \max_{\phi:[K] \rightarrow [K]} \sum_{m=1}^n X_m^{(\phi(k_m))} - X_m^{(k_m)}$$

# General regret

- **Regret:** “As well as the best constant strategy”
- **General:** “As well as  $\xi(k_1, \dots, k_n)$  instead of  $k_n$ ,  $\xi \in \Xi$ ” [Lehrer’02]

$$R_n^{\text{gen}} = \max_{\xi \in \Xi} \left\{ \max_j \sum_{m=1}^n X_m^{(\xi(k_1, \dots, k_m))} - X_m^{(k_m)} \right\}$$

- Generalized version of “exponential weights” [P’14]

$$\mathbb{E}R_n^{\text{gen}} \leq \square \sqrt{\log(|\Xi|)n}$$

- Internal regret  $\leq \square \sqrt{\log(K)n}$ , Swap regret  $\leq \square \sqrt{K \log(K)n}$

## Third Part

### Links with **Game Theory**

What we have learned in the previous section:

- In **worst case analysis**
  - Learning is **as fast** in adversarial than stochastic environment
- In the adversarial framework
  - **Refined notions** of regret can be minimized

## Against Opponents - Game Theory

$X_n^{(k)}$  **not** arbitrary, but induced by choices of another player

- **TWO players**, simultaneous actions in  $\{1, \dots, K\}$  and  $\{1, \dots, L\}$
- Payoffs are defined by **two matrices**  $A \in \mathbb{R}^{K \times L}$  and  $B \in \mathbb{R}^{K \times L}$ .
  - Player 1 picks **row**  $k \in \{1, \dots, K\}$  and Player 2 **column**  $\ell \in \{1, \dots, L\}$
  - **Player 1 gets**  $A_{k,\ell}$  and **Player 2 gets**  $B_{k,\ell}$
- Choices can be **random**  $p \in \Delta([K])$  and  $q \in \Delta([L])$ 
  - Player 1 gets  $\sum_{k,\ell} p_k q_\ell A_{k,\ell} = p^T A q$ ; P2 gets  $p^T B q$
- Online **learning**:  $X_n^{(k)} = A_{k,\ell_n}$  and  $Y_n^{(\ell)} = B_{k_n,\ell}$ .

Assume both players minimize regret independently.

Do they “**learn a solution concept**” from game theory ?

# Nash Equilibria

“A Nash equilibria is a situation where no player has interest to change his action” [Nash'50], [Nash'51]

- A Nash equilibria is a pair  $(p^*, q^*) \in \Delta([K]) \times \Delta([L])$  such that

- Player 1 has no interest to change given  $q^*$ :

$$(p^*)^T A q^* \geq p^T A q^*, \quad \forall p \in \Delta([K])$$

- Player 2 has no interest to change given  $p^*$ :

$$(p^*)^T A q^* \geq (p^*)^T A q, \quad \forall q \in \Delta([L])$$

- There **always exist** Nash equilibria; generically an **odd** number

[Nash'50], [Nash'51], [Shapley'74]

# Are Nash Equilibria Learnable?

- Both players minimize their regret independently.

$$k_n \sim p_n \in \Delta([K]), \quad \ell_n \sim q_n \in \Delta([L])$$

## Learning Nash equilibria could mean:

- $(p_n, q_n) \in \Delta([K]) \times \Delta([L])$  cv to a NE, or to set of NE.
- $(\frac{1}{n} \sum_{m=1}^n \delta_{k_m}, \frac{1}{n} \sum_{m=1}^n \delta_{\ell_m}) \in \Delta([K]) \times \Delta([L])$  cv to a NE, or to set of NE
- $(\frac{1}{n} \sum_{m=1}^n \delta_{k_m, \ell_m}) \in \Delta([K] \times [L])$  cv to a NE, or to set of NE

- Nash equilibria **are not learnable** (independently): [Hart, Mas-Colell'04]

There always exists a game s.t. none of the convergence occur

- What is Learnable?

**correlated** eq, **Minmax-Value**, **Potential** eq [Coucheney, **Gaujal**, Mertikopolous]

# Correlated Equilibria

“Players use an external device to **correlate** (as traffic lights); when they are told to take an action (as stop or go), **it is optimal**”

- A correlated equilibrium is a distribution  $\pi \in \Delta([K] \times [L])$ .  
 $(k^*, \ell^*) \sim \pi$ ; P1 is told **secretly** to play  $k^*$ , P2 to play  $\ell^*$ 
  - if P1 plays  $k^* \in [K]$ , he gets  $\sum_{\ell \in [L]} \pi_{k^*, \ell} A_{k^*, \ell}$ . If he plays  $j \in [K]$  instead, he would get  $\sum_{\ell \in [L]} \pi_{k^*, \ell} A_{j, \ell}$
- $\sum_{\ell \in [L]} \pi_{k^*, \ell} A_{k^*, \ell} \geq \sum_{\ell \in [L]} \pi_{k^*, \ell} A_{j, \ell}$ , for all  $k^*, j \in [K]$ 
  - Similar to no **internal regret** !

If both players minimize **internal regret**, empirical distribution of actions converge to the **set of correlated equilibria**. [Foster, Vohra'99]



# Minmax Theory

In zero-sum games, players have **optimal** strategies

- “zero-sum”:  $B = -A$ ; P1 maximizes and P2 minimizes  $p^T A q$
- Value =  $\max_{p \in \Delta([K])} \min_{q \in \Delta([L])} p^T A q = \min_{q \in \Delta([L])} \max_{p \in \Delta([K])} p^T A q$
- $p^*$  optimal if  $(p^*)^T A q \geq \text{Value}$  for all  $q \in \Delta([L])$ .

- $R_n \leq 0 \implies \frac{1}{n} \sum_{m=1}^n X_m^{(k_m)} \geq \text{Value}$
- $(\frac{1}{n} \sum_{m=1}^n \delta_{k_m}, \frac{1}{n} \sum_{m=1}^n \delta_{\ell_m})$  cv to optimal strat, i.e. to NE

- NE are **fast learnable** in zero-sum game, at  $O\left(\frac{1}{n}\right)$  [Harris'98]

# Conclusion

- In **worst case analysis**
  - With **full monitoring**, learning is **as fast** in adversarial than **stochastic** environment
  - Up to  $\sqrt{K}$ , learning is **as fast** with **bandit monit.** than with **full monit.**
- In **distribution dependent** (not worst case)
  - **Additional assumption** required to learn as fast in bandit than in full monitoring
- In **game theoretic framework**
  - Nash equilibria are **not learnable** in general
  - **Correlated equilibria** are learnable (by minimizing internal regret)
  - In **zero-sum and potential** games, equilibria are **learnable**.

**Fundamental textbook:** [Cesa-Bianchi, Lugosi'06]