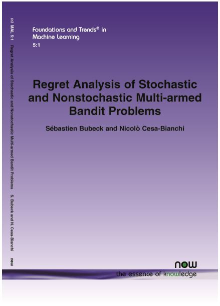
Lecture 1:
Introduction to regret analysis

Sébastien Bubeck
Machine Learning and Optimization group, MSR AI

Microsoft Research



Basic setting of online learning

Basic setting of online learning

Parameters: finite set of actions $[n]$ and number of rounds $T \geq n$.

Basic setting of online learning

Parameters: finite set of actions $[n]$ and number of rounds $T \geq n$.
Protocol: For each round $t \in [T]$, player chooses $i_t \in [n]$ and simultaneously adversary chooses a loss function $\ell_t : [n] \to [0, 1]$.

Basic setting of online learning

Parameters: finite set of actions $[n]$ and number of rounds $T \geq n$.
Protocol: For each round $t \in [T]$, player chooses $i_t \in [n]$ and simultaneously adversary chooses a loss function $\ell_t : [n] \to [0, 1]$.

Feedback model: In the *full information* game the player observes the complete loss function ℓ_t. In the *bandit* game the player only observes her own loss $\ell_t(i_t)$.

Basic setting of online learning

Parameters: finite set of actions $[n]$ and number of rounds $T \geq n$.
Protocol: For each round $t \in [T]$, player chooses $i_t \in [n]$ and simultaneously adversary chooses a loss function $\ell_t : [n] \to [0, 1]$.

Feedback model: In the *full information* game the player observes the complete loss function ℓ_t. In the *bandit* game the player only observes her own loss $\ell_t(i_t)$.

Performance measure: The regret is the difference between the player's accumulated loss and the minimum loss she could have obtained had she known all the adversary's choices:

$$R_T := \mathbb{E} \sum_{t=1}^{T} \ell_t(i_t) - \min_{i \in [n]} \mathbb{E} \sum_{t=1}^{T} \ell_t(i) =: L_T - \min_{i \in [n]} L_{i,T} .$$

Basic setting of online learning

Parameters: finite set of actions $[n]$ and number of rounds $T \geq n$.
Protocol: For each round $t \in [T]$, player chooses $i_t \in [n]$ and simultaneously adversary chooses a loss function $\ell_t : [n] \to [0,1]$.

Feedback model: In the *full information* game the player observes the complete loss function ℓ_t. In the *bandit* game the player only observes her own loss $\ell_t(i_t)$.

Performance measure: The regret is the difference between the player's accumulated loss and the minimum loss she could have obtained had she known all the adversary's choices:

$$R_T := \mathbb{E} \sum_{t=1}^{T} \ell_t(i_t) - \min_{i \in [n]} \mathbb{E} \sum_{t=1}^{T} \ell_t(i) =: L_T - \min_{i \in [n]} L_{i,T} .$$

What's it about? Full information game is about *hedging*, while bandit game also features the fundamental tension between *exploration* and *exploitation*.

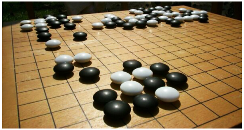
Applications

These challenges (scarce feedback, robustness to non i.i.d. data, exploration vs exploitation) are crucial components of many practical problems, hence the success of online learning and bandit theory!

Applications

These challenges (scarce feedback, robustness to non i.i.d. data, exploration vs exploitation) are crucial components of many practical problems, hence the success of online learning and bandit theory!

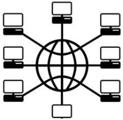
AI for games



Brain computer interface

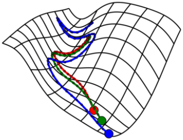
Medical trials

Packets routing



Ad placement

Hyperparameter opt



Hedging with multiplicative weights [Freund and Schapire 96, Littlestone and Warmuth 94, Vovk 90]

Assume for simplicity $\ell_t(i) \in \{0, 1\}$. MW keeps weights $w_{i,t}$ for each action, plays from normalized weights, and update as follows:

$$w_{i,t+1} = (1 - \eta\ell_t(i))w_{i,t} \,.$$

Hedging with multiplicative weights [Freund and Schapire 96, Littlestone and Warmuth 94, Vovk 90]

Assume for simplicity $\ell_t(i) \in \{0, 1\}$. MW keeps weights $w_{i,t}$ for each action, plays from normalized weights, and update as follows:

$$w_{i,t+1} = (1 - \eta \ell_t(i)) w_{i,t} \,.$$

Key insight: if i^* does not make a mistake on round t then we get "closer" to δ_{i^*} (i.e., we learn), and otherwise we might get confused but i^* had to pay for it.

Hedging with multiplicative weights [Freund and Schapire 96, Littlestone and Warmuth 94, Vovk 90]

Assume for simplicity $\ell_t(i) \in \{0, 1\}$. MW keeps weights $w_{i,t}$ for each action, plays from normalized weights, and update as follows:

$$w_{i,t+1} = (1 - \eta \ell_t(i)) w_{i,t} \,.$$

Key insight: if i^* does not make a mistake on round t then we get "closer" to δ_{i^*} (i.e., we learn), and otherwise we might get confused but i^* had to pay for it.

Theorem

For any $\eta \in [0, 1/2]$ *and* $i \in [n]$,

$$L_T \leq (1 + \eta) L_{i,T} + \frac{\log(n)}{\eta} \,.$$

By optimizing η *one gets* $R_T \leq 2\sqrt{T \log(n)}$.

Hedging with multiplicative weights [Freund and Schapire 96, Littlestone and Warmuth 94, Vovk 90]

Assume for simplicity $\ell_t(i) \in \{0, 1\}$. MW keeps weights $w_{i,t}$ for each action, plays from normalized weights, and update as follows:

$$w_{i,t+1} = (1 - \eta \ell_t(i)) w_{i,t} \,.$$

Key insight: if i^* does not make a mistake on round t then we get "closer" to δ_{i^*} (i.e., we learn), and otherwise we might get confused but i^* had to pay for it.

Theorem

For any $\eta \in [0, 1/2]$ and $i \in [n]$,

$$L_T \leq (1 + \eta) L_{i,T} + \frac{\log(n)}{\eta} \,.$$

By optimizing η one gets $R_T \leq 2\sqrt{T \log(n)}$.

Note that $\Omega(\sqrt{T \log(n)})$ is the best one could hope for.

Potential based analysis

Define $\psi(t) = \sum_{i=1}^n w_{i,t}$. One has:

$$\psi(t+1) = \sum_{i=1}^n (1 - \eta \ell_t(i)) w_{i,t} = \psi(t)(1 - \eta \langle p_t, \ell_t \rangle) \,,$$

Potential based analysis

Define $\psi(t) = \sum_{i=1}^{n} w_{i,t}$. One has:

$$\psi(t+1) = \sum_{i=1}^{n} (1 - \eta \ell_t(i)) w_{i,t} = \psi(t)(1 - \eta \langle p_t, \ell_t \rangle) \,,$$

so that (since $\psi(1) = n$):

$$\psi(T+1) = n \prod_{t=1}^{T} (1 - \eta \langle p_t, \ell_t \rangle) \leq n \exp(-\eta L_T) \,.$$

Potential based analysis

Define $\psi(t) = \sum_{i=1}^{n} w_{i,t}$. One has:

$$\psi(t+1) = \sum_{i=1}^{n} (1 - \eta \ell_t(i)) w_{i,t} = \psi(t)(1 - \eta \langle p_t, \ell_t \rangle) \,,$$

so that (since $\psi(1) = n$):

$$\psi(T+1) = n \prod_{t=1}^{T} (1 - \eta \langle p_t, \ell_t \rangle) \leq n \exp(-\eta L_T) \,.$$

On the other hand $\psi(T+1) \geq w_{i,T+1} = (1-\eta)^{L_{i,T}}$

Potential based analysis

Define $\psi(t) = \sum_{i=1}^{n} w_{i,t}$. One has:

$$\psi(t+1) = \sum_{i=1}^{n} (1 - \eta \ell_t(i)) w_{i,t} = \psi(t)(1 - \eta \langle p_t, \ell_t \rangle) ,$$

so that (since $\psi(1) = n$):

$$\psi(T+1) = n \prod_{t=1}^{T} (1 - \eta \langle p_t, \ell_t \rangle) \leq n \exp(-\eta L_T) .$$

On the other hand $\psi(T+1) \geq w_{i,T+1} = (1-\eta)^{L_{i,T}}$, and thus:

$$\eta L_T - \log\left(\frac{1}{1-\eta}\right) L_{i,T} \leq \log(n) ,$$

and the proof is concluded by $\log\left(\frac{1}{1-\eta}\right) \leq \eta + \eta^2$ for $\eta \in [0, 1/2]$.

Potential based analysis

Define $\psi(t) = \sum_{i=1}^{n} w_{i,t}$. One has:

$$\psi(t+1) = \sum_{i=1}^{n} (1 - \eta \ell_t(i)) w_{i,t} = \psi(t)(1 - \eta \langle p_t, \ell_t \rangle) \,,$$

so that (since $\psi(1) = n$):

$$\psi(T+1) = n \prod_{t=1}^{T} (1 - \eta \langle p_t, \ell_t \rangle) \leq n \exp(-\eta L_T) \,.$$

On the other hand $\psi(T+1) \geq w_{i,T+1} = (1-\eta)^{L_{i,T}}$, and thus:

$$\eta L_T - \log\left(\frac{1}{1-\eta}\right) L_{i,T} \leq \log(n) \,,$$

and the proof is concluded by $\log\left(\frac{1}{1-\eta}\right) \leq \eta + \eta^2$ for $\eta \in [0, 1/2]$.

The mirror descent framework (Lec. 2) will give a principled approach to derive both the MW algorithm and its analysis.

A principled game-theoretic approach to regret analysis

[Abernethy, Warmuth, Yellin 2008; Rakhlin, Sridharan, Tewari 2010; B., Dekel, Koren, Peres 2015]

Let us focus on an oblivious adversary, that is he chooses $\ell_1, \dots, \ell_T \in \mathcal{L}$ at the beginning of the game.

A principled game-theoretic approach to regret analysis

[Abernethy, Warmuth, Yellin 2008; Rakhlin, Sridharan, Tewari 2010; B., Dekel, Koren, Peres 2015]

Let us focus on an oblivious adversary, that is he chooses $\ell_1, \dots, \ell_T \in \mathcal{L}$ at the beginning of the game.

A deterministic player's strategy is specified by a sequence of operators $a_1, \dots, a_T$, where in the full information case $a_s : ([0,1]^n)^{s-1} \to \mathcal{K}$, and in the bandit case $a_s : \mathbb{R}^{s-1} \to \mathcal{K}$. Denote $\mathcal{A}$ the set of such sequences of operators.

A principled game-theoretic approach to regret analysis

[Abernethy, Warmuth, Yellin 2008; Rakhlin, Sridharan, Tewari 2010; B., Dekel, Koren, Peres 2015]

Let us focus on an oblivious adversary, that is he chooses $\ell_1, \dots, \ell_T \in \mathcal{L}$ at the beginning of the game.

A deterministic player's strategy is specified by a sequence of operators $a_1, \dots, a_T$, where in the full information case $a_s : ([0,1]^n)^{s-1} \to \mathcal{K}$, and in the bandit case $a_s : \mathbb{R}^{s-1} \to \mathcal{K}$. Denote $\mathcal{A}$ the set of such sequences of operators.

Write $R_T(\mathbf{a}, \ell)$ for the regret of playing strategy $\mathbf{a} \in \mathcal{A}$ against loss sequence $\ell \in \mathcal{L}^T$.

A principled game-theoretic approach to regret analysis

[Abernethy, Warmuth, Yellin 2008; Rakhlin, Sridharan, Tewari 2010; B., Dekel, Koren, Peres 2015]

Let us focus on an oblivious adversary, that is he chooses $\ell_1, \dots, \ell_T \in \mathcal{L}$ at the beginning of the game.

A deterministic player's strategy is specified by a sequence of operators $a_1, \dots, a_T$, where in the full information case $a_s : ([0,1]^n)^{s-1} \to \mathcal{K}$, and in the bandit case $a_s : \mathbb{R}^{s-1} \to \mathcal{K}$. Denote $\mathcal{A}$ the set of such sequences of operators.

Write $R_T(\mathbf{a}, \ell)$ for the regret of playing strategy $\mathbf{a} \in \mathcal{A}$ against loss sequence $\ell \in \mathcal{L}^T$. Now we are interested in:

$$\inf_{\mu \in \Delta(\mathcal{A})} \sup_{\ell \in \mathcal{L}^T} \mathbb{E}_{\mathbf{a} \sim \mu} R_T(\mathbf{a}, \ell) = \sup_{\nu \in \Delta(\mathcal{L}^T)} \inf_{\mu \in \Delta(\mathcal{A})} \mathbb{E}_{\ell \sim \nu, \mathbf{a} \sim \mu} R_T(\mathbf{a}, \ell),$$

where the swap of min and max comes from Sion's minimax theorem.

A principled game-theoretic approach to regret analysis

[Abernethy, Warmuth, Yellin 2008; Rakhlin, Sridharan, Tewari 2010; B., Dekel, Koren, Peres 2015]

Let us focus on an oblivious adversary, that is he chooses $\ell_1, \dots, \ell_T \in \mathcal{L}$ at the beginning of the game.

A deterministic player's strategy is specified by a sequence of operators $a_1, \dots, a_T$, where in the full information case $a_s : ([0,1]^n)^{s-1} \to \mathcal{K}$, and in the bandit case $a_s : \mathbb{R}^{s-1} \to \mathcal{K}$. Denote $\mathcal{A}$ the set of such sequences of operators.

Write $R_T(\mathbf{a}, \ell)$ for the regret of playing strategy $\mathbf{a} \in \mathcal{A}$ against loss sequence $\ell \in \mathcal{L}^T$. Now we are interested in:

$$\inf_{\mu \in \Delta(\mathcal{A})} \sup_{\ell \in \mathcal{L}^T} \mathbb{E}_{\mathbf{a} \sim \mu} R_T(\mathbf{a}, \ell) = \sup_{\nu \in \Delta(\mathcal{L}^T)} \inf_{\mu \in \Delta(\mathcal{A})} \mathbb{E}_{\ell \sim \nu, \mathbf{a} \sim \mu} R_T(\mathbf{a}, \ell) \,,$$

where the swap of min and max comes from Sion's minimax theorem.

In other words we can study the minimax regret by designing a strategy for a *Bayesian* scenario where $\ell \sim \nu$ and ν is known.

A Doob strategy [B., Dekel, Koren, Peres 2015]

Since we known ν, we also know the *distribution* of i^*. In fact as we make observations, we can update our knowledge of i^* with the *posterior distribution*. Denote $\mathbb{E}_t$ for this posterior distribution (e.g., in full information $\mathbb{E}_t := \mathbb{E}[\cdot|\ell_1, \dots, \ell_{t-1}]$).

A Doob strategy [B., Dekel, Koren, Peres 2015]

Since we known ν, we also know the *distribution* of i^*. In fact as we make observations, we can update our knowledge of i^* with the *posterior distribution*. Denote $\mathbb{E}_t$ for this posterior distribution (e.g., in full information $\mathbb{E}_t := \mathbb{E}[\cdot|\ell_1, \dots, \ell_{t-1}]$).
By convexity of $\Delta([n]) =: \Delta_n$ it is natural to consider playing from:

$$p_t := \mathbb{E}_t \; \delta_{i^*} \; .$$

In other words we are playing from the posterior distribution of the optimum, a kind of "probability matching". This is also called Thompson Sampling.

A Doob strategy [B., Dekel, Koren, Peres 2015]

Since we known ν, we also know the *distribution* of i^*. In fact as we make observations, we can update our knowledge of i^* with the *posterior distribution*. Denote $\mathbb{E}_t$ for this posterior distribution (e.g., in full information $\mathbb{E}_t := \mathbb{E}[\cdot|\ell_1, \dots, \ell_{t-1}]$).
By convexity of $\Delta([n]) =: \Delta_n$ it is natural to consider playing from:

$$p_t := \mathbb{E}_t \ \delta_{i^*} .$$

In other words we are playing from the posterior distribution of the optimum, a kind of "probability matching". This is also called Thompson Sampling.
The regret of this strategy can be controlled via the *movement* of this Doob martingale (recall $\|\ell_t\|_\infty \leq 1$)

$$\mathbb{E} \sum_{t=1}^{T} \langle p_t - \delta_{i^*}, \ell_t \rangle = \mathbb{E} \sum_{t=1}^{T} \langle p_t - p_{t+1}, \ell_t \rangle \leq \mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1 .$$

How stable is a martingale?

Question: is a martingale in Δ_n "stable"? Following famous inequality is a possible answer (proof on the next slide):

$$\mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1^2 \leq 2 \log(n) \, .$$

How stable is a martingale?

Question: is a martingale in Δ_n "stable"? Following famous inequality is a possible answer (proof on the next slide):

$$\mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1^2 \leq 2 \log(n)\,.$$

This yields by Cauchy-Schwarz:

$$\mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1 \leq \sqrt{T \times \mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1^2} \leq \sqrt{2 T \log(n)}\,.$$

How stable is a martingale?

Question: is a martingale in Δ_n "stable"? Following famous inequality is a possible answer (proof on the next slide):

$$\mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1^2 \leq 2 \log(n) \,.$$

This yields by Cauchy-Schwarz:

$$\mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1 \leq \sqrt{T \times \mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1^2} \leq \sqrt{2 T \log(n)} \,.$$

Thus we have recovered the regret bound of MW (in fact with an optimal constant) by a purely geometric argument!

Entropic proof of cotype for ℓ_1^n

$$\mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1^2 \leq 2 \log(n) \,.$$

Entropic proof of cotype for ℓ_1^n

$$\mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1^2 \leq 2 \log(n) \,.$$

By Pinsker's inequality:

$$\frac{1}{2} \|p_t - p_{t+1}\|_1^2 \leq \mathrm{Ent}(p_{t+1}; p_t) = \mathrm{Ent}_t(i^* | \ell_t; i^*) \,.$$

Entropic proof of cotype for ℓ_1^n

$$\mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1^2 \leq 2 \log(n) \,.$$

By Pinsker's inequality:

$$\frac{1}{2} \|p_t - p_{t+1}\|_1^2 \leq \mathrm{Ent}(p_{t+1}; p_t) = \mathrm{Ent}_t(i^* | \ell_t; i^*) \,.$$

Now essentially by definition one has (recall that $I(X, Y) = H(X) - H(X|Y) = \mathbb{E}_Y \mathrm{Ent}(p_{X|Y}; p_X)$):

$$\mathbb{E}_{\ell_t} \mathrm{Ent}_t(i^* | \ell_t; i^*) = H_t(i^*) - H_{t+1}(i^*) \,.$$

Entropic proof of cotype for ℓ_1^n

$$\mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1^2 \leq 2 \log(n) .$$

By Pinsker's inequality:

$$\frac{1}{2} \|p_t - p_{t+1}\|_1^2 \leq \mathrm{Ent}(p_{t+1}; p_t) = \mathrm{Ent}_t(i^* | \ell_t; i^*) .$$

Now essentially by definition one has (recall that $I(X, Y) = H(X) - H(X|Y) = \mathbb{E}_Y \mathrm{Ent}(p_{X|Y}; p_X)$):

$$\mathbb{E}_{\ell_t} \mathrm{Ent}_t(i^* | \ell_t; i^*) = H_t(i^*) - H_{t+1}(i^*) .$$

Proof concluded by telescopic sum and maximal entropy being $\log(n)$.

A more general story: M-cotype

Let us generalize the setting. In *online linear optimization*, the player plays $x_t \in K \subset \mathbb{R}^n$, and the adversary plays $\ell_t \in \mathcal{L} \subset \mathbb{R}^n$. We assume that there is a norm $\|\cdot\|$ such that $\|x_t\| \leq 1$ and $\|\ell_t\|^* \leq 1$.

A more general story: M-cotype

Let us generalize the setting. In *online linear optimization*, the player plays $x_t \in K \subset \mathbb{R}^n$, and the adversary plays $\ell_t \in \mathcal{L} \subset \mathbb{R}^n$. We assume that there is a norm $\|\cdot\|$ such that $\|x_t\| \leq 1$ and $\|\ell_t\|^* \leq 1$. The same game-theoretic argument goes through, and denoting $x^* = \mathrm{argmin}_{x \in K} \sum_{t=1}^T \langle \ell_t, x \rangle$, $x_t := \mathbb{E}_t x^*$, one has

$$\mathbb{E} \sum_{t=1}^{T} \langle \ell_t, x_t - x^* \rangle = \mathbb{E} \sum_{t=1}^{T} \langle \ell_t, x_t - x_{t+1} \rangle \leq \mathbb{E} \sum_{t=1}^{T} \|x_t - x_{t+1}\| \,.$$

A more general story: M-cotype

Let us generalize the setting. In *online linear optimization*, the player plays $x_t \in K \subset \mathbb{R}^n$, and the adversary plays $\ell_t \in \mathcal{L} \subset \mathbb{R}^n$. We assume that there is a norm $\|\cdot\|$ such that $\|x_t\| \leq 1$ and $\|\ell_t\|^* \leq 1$. The same game-theoretic argument goes through, and denoting $x^* = \mathrm{argmin}_{x \in K} \sum_{t=1}^{T} \langle \ell_t, x \rangle$, $x_t := \mathbb{E}_t x^*$, one has

$$\mathbb{E} \sum_{t=1}^{T} \langle \ell_t, x_t - x^* \rangle = \mathbb{E} \sum_{t=1}^{T} \langle \ell_t, x_t - x_{t+1} \rangle \leq \mathbb{E} \sum_{t=1}^{T} \|x_t - x_{t+1}\| .$$

The norm $\|\cdot\|$ has *M*-cotype (C, q) if for any martingale (x_t) one has:

$$\left(\mathbb{E} \sum_{t=1}^{T} \|x_t - x_{t+1}\|^q \right)^{1/q} \leq C \; \mathbb{E} \|x_{T+1}\| .$$

A more general story: M-cotype

Let us generalize the setting. In *online linear optimization*, the player plays $x_t \in K \subset \mathbb{R}^n$, and the adversary plays $\ell_t \in \mathcal{L} \subset \mathbb{R}^n$. We assume that there is a norm $\|\cdot\|$ such that $\|x_t\| \leq 1$ and $\|\ell_t\|^* \leq 1$. The same game-theoretic argument goes through, and denoting $x^* = \mathrm{argmin}_{x \in K} \sum_{t=1}^T \langle \ell_t, x \rangle$, $x_t := \mathbb{E}_t x^*$, one has

$$\mathbb{E} \sum_{t=1}^T \langle \ell_t, x_t - x^* \rangle = \mathbb{E} \sum_{t=1}^T \langle \ell_t, x_t - x_{t+1} \rangle \leq \mathbb{E} \sum_{t=1}^T \|x_t - x_{t+1}\| .$$

The norm $\|\cdot\|$ has M-cotype (C, q) if for any martingale (x_t) one has:

$$\left(\mathbb{E} \sum_{t=1}^T \|x_t - x_{t+1}\|^q \right)^{1/q} \leq C \ \mathbb{E} \|x_{T+1}\| .$$

In particular this gives a regret in $C \ T^{1-1/q}$.

A lower bound via M-type of the dual

Interestingly the analysis via cotype is tight in the following sense.

A lower bound via M-type of the dual

Interestingly the analysis via cotype is tight in the following sense. First if M-cotype (C, q) holds for $\|\cdot\|$, then so does M-type (C', p) for $\|\cdot\|_*$ (where p is the conjugate of q), i.e., for any martingale difference sequence (ℓ_t) one has

$$\mathbb{E}\left\|\sum_{t=1}^{T} \ell_t\right\|_* \leq C' \left(\mathbb{E}\sum_{t=1}^{T} \|\ell_t\|_*^p\right)^{1/p} .$$

A lower bound via M-type of the dual

Interestingly the analysis via cotype is tight in the following sense. First if M-cotype (C, q) holds for $\|\cdot\|$, then so does M-type (C', p) for $\|\cdot\|_*$ (where p is the conjugate of q), i.e., for any martingale difference sequence (ℓ_t) one has

$$\mathbb{E}\left\|\sum_{t=1}^{T} \ell_t\right\|_* \leq C' \left(\mathbb{E}\sum_{t=1}^{T} \|\ell_t\|_*^p\right)^{1/p} .$$

Moreover one can show that the violation of type/cotype can be witnessed by a martingale with unit norm increments. Thus if M-cotype (C, q) fails for $\|\cdot\|$, there must exist a martingale difference sequence (ℓ_t) with $\|\ell_t\|_* = 1$ that violates the above inequality.

A lower bound via M-type of the dual

Interestingly the analysis via cotype is tight in the following sense. First if M-cotype (C, q) holds for $\|\cdot\|$, then so does M-type (C', p) for $\|\cdot\|_*$ (where p is the conjugate of q), i.e., for any martingale difference sequence (ℓ_t) one has

$$\mathbb{E}\left\|\sum_{t=1}^{T} \ell_t\right\|_* \leq C' \left(\mathbb{E}\sum_{t=1}^{T} \|\ell_t\|_*^p\right)^{1/p} .$$

Moreover one can show that the violation of type/cotype can be witnessed by a martingale with unit norm increments. Thus if M-cotype (C, q) fails for $\|\cdot\|$, there must exist a martingale difference sequence (ℓ_t) with $\|\ell_t\|_* = 1$ that violates the above inequality. In particular:

$$\mathbb{E}\sum_{t=1}^{T}\langle \ell_t, x_t - x^*\rangle = \mathbb{E}\left\|\sum_{t=1}^{T} \ell_t\right\|_* \geq C' T^{1/p} = C' T^{1-1/q} .$$

A lower bound via M-type of the dual

Interestingly the analysis via cotype is tight in the following sense. First if M-cotype (C, q) holds for $\|\cdot\|$, then so does M-type (C', p) for $\|\cdot\|_*$ (where p is the conjugate of q), i.e., for any martingale difference sequence (ℓ_t) one has

$$\mathbb{E}\left\|\sum_{t=1}^{T} \ell_t\right\|_* \leq C' \left(\mathbb{E}\sum_{t=1}^{T} \|\ell_t\|_*^p\right)^{1/p} .$$

Moreover one can show that the violation of type/cotype can be witnessed by a martingale with unit norm increments. Thus if M-cotype (C, q) fails for $\|\cdot\|$, there must exist a martingale difference sequence (ℓ_t) with $\|\ell_t\|_* = 1$ that violates the above inequality. In particular:

$$\mathbb{E}\sum_{t=1}^{T} \langle \ell_t, x_t - x^* \rangle = \mathbb{E}\left\|\sum_{t=1}^{T} \ell_t\right\|_* \geq C' T^{1/p} = C' T^{1-1/q} .$$

Important: these are "dimension-free arguments", if one brings the dimension in the bounds then the story changes.

What about the bandit game? [Russo, Van Roy 2014]

So far we only talked about the *hedging* aspect of the problem. In particular for the full information game the "learning" part happens automatically. This is captured by the fact that the **variation in the posterior is lower bounded by the instantaneous regret**:

$$\mathbb{E}_t \langle p_t - \delta_{i^*}, \ell_t \rangle = \mathbb{E}_t \langle p_t - p_{t+1}, \ell_t \rangle \leq \mathbb{E}_t \| p_t - p_{t+1} \|_1 \, .$$

What about the bandit game? [Russo, Van Roy 2014]

So far we only talked about the *hedging* aspect of the problem. In particular for the full information game the "learning" part happens automatically. This is captured by the fact that the **variation in the posterior is lower bounded by the instantaneous regret**:

$$\mathbb{E}_t \langle p_t - \delta_{i^*}, \ell_t \rangle = \mathbb{E}_t \langle p_t - p_{t+1}, \ell_t \rangle \leq \mathbb{E}_t \|p_t - p_{t+1}\|_1 .$$

In the bandit game the first equality is not true anymore and thus the inequality does not hold a priori. In fact this is the whole difficulty: learning is now costly because of the tradeoff between exploration and exploitation.

What about the bandit game? [Russo, Van Roy 2014]

So far we only talked about the *hedging* aspect of the problem. In particular for the full information game the "learning" part happens automatically. This is captured by the fact that the **variation in the posterior is lower bounded by the instantaneous regret**:

$$\mathbb{E}_t \langle p_t - \delta_{i^*}, \ell_t \rangle = \mathbb{E}_t \langle p_t - p_{t+1}, \ell_t \rangle \leq \mathbb{E}_t \| p_t - p_{t+1} \|_1 .$$

In the bandit game the first equality is not true anymore and thus the inequality does not hold a priori. In fact this is the whole difficulty: learning is now costly because of the tradeoff between exploration and exploitation.

Importantly note that the cotype inequality for ℓ_1 is proved by relating the ℓ_1 variation squared to the mutual information between OPT and the feedback. Thus a weaker inequality that would suffice is:

$$\mathbb{E}_t \langle p_t - \delta_{i^*}, \ell_t \rangle \leq C \sqrt{I_t(i^*, (i_t, \ell_t(i_t)))},$$

which would lead to a regret in $C\sqrt{T \log(n)}$.

The Russo-Van Roy analysis

Let $\bar{\ell}_t(i) = \mathbb{E}_t \ell_t(i)$ and $\bar{\ell}_t(i,j) = \mathbb{E}_t(\ell_t(i)|i^* = j)$. Then

$$\mathbb{E}_t \langle p_t - \delta_{i^*}, \ell_t \rangle = \sum_i p_t(i)(\bar{\ell}_t(i) - \bar{\ell}_t(i,i)) \, ,$$

and

$$I_t((i_t, \ell_t(i_t)), i^*) = \sum_{i,j} p_t(i) p_t(j) \mathrm{Ent}(\mathcal{L}_t(\ell_t(i)|i^* = j) \| \mathcal{L}_t(\ell_t(i)))$$

The Russo-Van Roy analysis

Let $\bar{\ell}_t(i) = \mathbb{E}_t \ell_t(i)$ and $\bar{\ell}_t(i,j) = \mathbb{E}_t(\ell_t(i) | i^* = j)$. Then

$$\mathbb{E}_t \langle p_t - \delta_{i^*}, \ell_t \rangle = \sum_i p_t(i) (\bar{\ell}_t(i) - \bar{\ell}_t(i,i)) \,,$$

and

$$I_t((i_t, \ell_t(i_t)), i^*) = \sum_{i,j} p_t(i) p_t(j) \mathrm{Ent}(\mathcal{L}_t(\ell_t(i) | i^* = j) \| \mathcal{L}_t(\ell_t(i)))$$

Now using Cauchy-Schwarz the instantaneous regret is bounded by

$$\sqrt{n \sum_i p_t(i)^2 (\bar{\ell}_t(i) - \bar{\ell}_t(i,i))^2} \leq \sqrt{n \sum_{i,j} p_t(i) p_t(j) (\bar{\ell}_t(i) - \bar{\ell}_t(i,j))^2} \,.$$

The Russo-Van Roy analysis

Let $\bar{\ell}_t(i) = \mathbb{E}_t \ell_t(i)$ and $\bar{\ell}_t(i,j) = \mathbb{E}_t(\ell_t(i) | i^* = j)$. Then

$$\mathbb{E}_t \langle p_t - \delta_{i^*}, \ell_t \rangle = \sum_i p_t(i) (\bar{\ell}_t(i) - \bar{\ell}_t(i,i)) ,$$

and

$$I_t((i_t, \ell_t(i_t)), i^*) = \sum_{i,j} p_t(i) p_t(j) \mathrm{Ent}(\mathcal{L}_t(\ell_t(i) | i^* = j) \| \mathcal{L}_t(\ell_t(i)))$$

Now using Cauchy-Schwarz the instantaneous regret is bounded by

$$\sqrt{n \sum_i p_t(i)^2 (\bar{\ell}_t(i) - \bar{\ell}_t(i,i))^2} \leq \sqrt{n \sum_{i,j} p_t(i) p_t(j) (\bar{\ell}_t(i) - \bar{\ell}_t(i,j))^2} .$$

Pinsker's inequality gives (using $\|\ell_t\|_\infty \leq 1$):

$$(\bar{\ell}_t(i) - \bar{\ell}_t(i,j))^2 \leq \mathrm{Ent}(\mathcal{L}_t(\ell_t(i) | i^* = j) \| \mathcal{L}_t(\ell_t(i))) ,$$

The Russo-Van Roy analysis

Let $\bar{\ell}_t(i) = \mathbb{E}_t \ell_t(i)$ and $\bar{\ell}_t(i,j) = \mathbb{E}_t(\ell_t(i) | i^* = j)$. Then

$$\mathbb{E}_t \langle p_t - \delta_{i^*}, \ell_t \rangle = \sum_i p_t(i) (\bar{\ell}_t(i) - \bar{\ell}_t(i,i)) ,$$

and

$$I_t((i_t, \ell_t(i_t)), i^*) = \sum_{i,j} p_t(i) p_t(j) \mathrm{Ent}(\mathcal{L}_t(\ell_t(i) | i^* = j) \| \mathcal{L}_t(\ell_t(i)))$$

Now using Cauchy-Schwarz the instantaneous regret is bounded by

$$\sqrt{n \sum_i p_t(i)^2 (\bar{\ell}_t(i) - \bar{\ell}_t(i,i))^2} \leq \sqrt{n \sum_{i,j} p_t(i) p_t(j) (\bar{\ell}_t(i) - \bar{\ell}_t(i,j))^2} .$$

Pinsker's inequality gives (using $\|\ell_t\|_\infty \leq 1$):

$$(\bar{\ell}_t(i) - \bar{\ell}_t(i,j))^2 \leq \mathrm{Ent}(\mathcal{L}_t(\ell_t(i) | i^* = j) \| \mathcal{L}_t(\ell_t(i))) ,$$

Thus one obtains

$$\mathbb{E}_t \langle p_t - \delta_{i^*}, \ell_t \rangle \leq \sqrt{n \; I_t((i_t, \ell_t(i_t)), i^*)} .$$