MVA ENS Cachan Paris Saclay

Prediction for Individual Sequences
Notes de Cours

Cours donné par Vianney Perchet
Note prises par Adrien Lina

9 mars 2018

1 Lecture 6 : Modèle en full monitoring

1.1 Pseudo-regret

Nous avons vu EXP4 au dernier cours : à chaque temps t

— le joueur observe les avis d'experts $\xi_t^{(1)}, ..., \xi_t^{(N)} \in [K]$

— le joueur choisit le bras $\pi_t = \xi_t^{(i)}$ avec proba $q_t^{(i)}$ où $q_t \in \Delta_N$. C'est équivalent à $\pi_t = k$ avec proba $p_t^{(k)}$ où $p_t = \mathbb{E}_{i \sim q_t}[\delta_{\xi_t^{(i)}}]$ (et donc $p_t^{(k)} = \sum\limits_{i=1}^{N} q_t^{(i)} \mathbb{1}_{\{\xi_t^{(i)} = k\}}$).

— on estime les rewards des bras : $\widehat{X_t^{(k)}} = \frac{X_t^{(k)} - 1}{p_t^{(k)}} \{\mathbb{k} = \pi_{\mathbb{t}}\}$

— on estime le reward des experts : $\widehat{Y_t^{(i)}} = \widehat{X_t^{(\xi_t^{(i)})}}$

— mise à jour des poids :

$$q_t^{(i)} = \frac{e^{\eta \widehat{R_t^{(i)}}}}{\sum\limits_{j=1}^{N} e^{\eta \widehat{R_t^{(j)}}}} \qquad \text{où} \qquad \widehat{R_t^{(i)}} = \sum_{s=1}^{t-1} \widehat{Y_s^{(i)}} - \widehat{X_s^{(i)}}$$

Théorème 1.1. *pour η bien choisit :*

$$\max_{i=1,\dots,N} \mathbb{E}\left[\sum_{t=1}^{T} X_t^{(\xi_t^{(j)})} - X_t^{(\pi_t)}\right] \leq 2\sqrt{TK \log N}$$

Remarque 1.1. En faisant les mêmes astuces que pour EXP3.P, on peut avoir le même résultat sur le regret et non le pseudo-regret.

Remarque 1.2. On peut remarquer (c'est utile pour la démonstration)

$$\begin{aligned}
\mathbb{E}_{i\sim q_t}\left[Z_t^{(\xi_t^{(i)})}\right] &= \sum_{i=1}^{N} q_t^{(i)} Z_t^{(\xi_t^{(i)})} \\
&= \sum_{i=1}^{N} q_t^{(i)} \overbrace{\sum_{k=1}^{K} \mathbb{1}\{k=\xi_t^{(i)}\} Z_t^k}^{Z_t^{(\xi_t^{(i)})}} \\
&= \sum_{k=1}^{K} \underbrace{\sum_{i=1}^{N} \mathbb{1}_{\{k=\xi_t^{(i)}\}}}_{p_t^{(k)}} Z_t^k \\
&= \mathbb{E}_{q\sim p_t}\left[Z_t^{(k)}\right]
\end{aligned}$$

Démonstration. en suivant la preuve de EXP avec q_t au lieu de p_t et l'esnemble des experts à la place des bras, on a (*faible perte* ou *small notice* en anglais) :

$$\widehat{R_t^{(i)}} \leq \frac{\log N}{\eta} + \eta \sum_{t=1}^{T} \mathbb{E}_{i\sim q_t}\left[\left(\widehat{Y_t^{(i)}}\right)^2\right] \tag{1}$$

où

$$\widehat{R_t^{(i)}} = \sum_{s=1}^{t-1} \widehat{Y_s^{(i)}} - \mathbb{E}_{k\sim p_t}\left[\widehat{X_s^{(k)}}\right] = \sum_{s=1}^{t-1} \widehat{X_s^{(\xi_t^{(i)})}} - \mathbb{E}_{k\sim p_t}\left[\widehat{X_s^{(k)}}\right]$$

car $\eta\widehat{Y_t^{(i)}} \leq 1 \ \forall(t,i)$

On calcule les espérences :

$$\begin{aligned}
\mathbb{E}\left[\widehat{Y_t^{(i)}}\right] &= \mathbb{E}\left[\widehat{X_t^{(\xi_t^{(i)})}}\right] \\
&= \mathbb{E}\left[\mathbb{E}_{\pi_t\sim p_t}\left[\frac{X_t^{(\xi_t^{(i)})}-1}{p_t^{(\xi_t^{(i)})}} \mathbb{1}_{\{\xi_t^{(i)}=\pi_t\}}\right]\right] \\
&= \mathbb{E}\left[X_t^{(\xi_t^{(i)})}-1\right]
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}_{k\sim p_t}\left[\widehat{X_s^{(k)}}\right]\right] &= \mathbb{E}\left[\sum_{k=1}^{K} p_t^{(k)} \widehat{X_t^{(k)}}\right] \\
&= \mathbb{E}\left[\sum_{k=1}^{K} p_t^{(k)} \frac{X_t^{(k)}-1}{p_t^{(k)}} \mathbb{1}_{\{k=\pi_t\}}\right] \\
&= \mathbb{E}\left[X_t^{(\pi_t)}-1\right]
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}_{i\sim q_t}\left[\left(\widehat{Y_t^{(i)}}\right)^2\right] &= \mathbb{E}_{i\sim q_t}\left[\left(\widehat{X_t^{(\xi_t^{(i)})}}\right)^2\right] \\
&= \mathbb{E}_{k\sim p_t}\left[\left(\widehat{X_t^{(k)}}\right)^2\right] \\
&= \mathbb{E}_{k\sim p_t}\left[\left(\frac{X_t^{(k)}-1}{p_t^{(k)}}\mathbb{1}_{\{k=\pi_t\}}\right)^2\right] \\
&= \mathbb{E}_{k\sim p_t}\left[\left(\frac{X_t^{(k)}-1}{p_t^{(k)}}\right)^2\mathbb{1}_{\{k=\pi_t\}}\right] \\
&= \frac{\left(X_t^{(\pi_t)}-1\right)^2}{p_t^{(\pi_t)}}
\end{aligned}
$$

Enfin

$$
\begin{aligned}
\mathbb{E}\left[\mathbb{E}_{i\sim q_t}\left[\left(\widehat{Y_t^{(i)}}\right)^2\right]\right] &\leq \mathbb{E}\left[\frac{1}{p_t^{(\pi_t)}}\right] \\
&\leq \mathbb{E}\left[\mathbb{E}_{\pi_t\sim p_t}\left[\frac{1}{p_t^{(\pi_t^{(\pi_t)})}}\right]\right] \\
&\leq K
\end{aligned}
$$

On prend l'espérance de (1) et on remplace :

$$
\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T} X_t^{(\xi_t^{(i)})} - 1 - (X_t^{(\pi_t)} - 1)\right] &\leq \frac{\log N}{\eta} + \eta TK \\
&\leq 2\sqrt{TK\log N} \quad \text{où } \eta = \sqrt{\frac{\log N}{TK}}
\end{aligned}
$$

□

1.2 Regret interne

Définition 1.1 (Regret interne)**.**

$$
\begin{aligned}
R_t^{(i\to j)} &= \sum_{t=1}^{T}\left(X_t^{(j)} - X_t^{(\pi_t)}\mathbb{1}_{\{\pi_t=i\}}\right) \\
R_t^{(interne)} &= \max_{(i,j)} R_T^{(i\to j)}
\end{aligned}
$$

Remarque 1.3. Si on sait minimiser le regret interne, alors $R_T^{(j)} = \sum_{i=1}^{K} R_t^{(i\to j)}$, donc $R_T \leq K R_T^{(\text{interne})}$ et donc on sait controler le regret externe.

Pour controler le regret interne, on définit les K^2 experts (on note la loi $q_t^{(k)}$) :

$$\xi_t^{(i\to j)} = \begin{cases} i & \text{avec proba } 0 \\ j & \text{avec proba } p_t^{(i)} + p_t^{(j)} \\ k & \text{avec proba } p_t^{(k)} \end{cases}$$

Définition valide car p_t ne dépend que de l'info jusqu'à $t-1$.

Utiliser l'algo précédet sur ces experts minimise le regret intere :

$$\mathbb{E}\left[\sum_{t=1}^{T} X_t^{(\xi_t^{(i\to j)})} - X_t^{(\pi_t)}\right] \leq 2\sqrt{2TK\log K}$$

où

$$\begin{aligned}
\mathbb{E}\left[X_t^{(\xi_t^{(i\to j)})} - X_t^{(\pi_t)}\right] &= \mathbb{E}\left[\sum_{k=1}^{K} q_t^{(k)} X_t^{(k)} - p_t^{(k)} X_t^{(k)}\right] \\
&= \mathbb{E}\left[p_t^{(i)} X_t^{(i)} - p_t^{(j)} X_t^{(j)}\right] \\
&= \mathbb{E}\left[p_t^{(i)} (X_t^{(j)} - X_t^{(i)})\right] \\
&= \mathbb{E}\left[(X_t^{(j)} - X_t^{(i)}) \mathbb{1}_{\{i=\pi_t\}}\right]
\end{aligned}$$

Donc

$$\mathbb{E}[R_T^{(i\to j)}] \leq 2\sqrt{2TK\log K}$$

Remarque 1.4. L'inverse n'est pas vrai.

1.3 Ensemble de bras continu

Définition 1.2. *À chaque temps t :*

- *Le joueur prend l'action $a_t \in \mathcal{A} \subseteq \mathbb{R}^d$;*
- *L'adversaire choisit $g_t : \mathcal{A} \to [0,1]$;*
- *Le joeur observe*
  - *$g_t(a) \forall a \in \mathcal{A}$: info complète*
  - *$g_t(a_t)$: bandit*

But *: minimiser le regret*

$$R_t^{(a)} = \sum_{t=1}^{T} g_t(a) - g_t(a_t) \forall a \in \mathcal{A}$$

1.3.1 Réduction du carde précédent en discrétisant

Si les fonctiosn g_t sont β-Holder : $|g_t(a_1) - g_t(a_2)| \leq c||a_1 - a_2||_2^{\beta}$.

On peut approximer $\mathcal{A}$ par une grille $\widehat{\mathcal{A}}_\epsilon$ telle que $\mathrm{Card}(\widehat{\mathcal{A}}_\epsilon) \lesssim \left(\frac{||\mathcal{A}||}{\epsilon}\right)^d$ et

$$\forall a \in \mathcal{A}, \exists \hat{a} \in \widehat{\mathcal{A}}_\epsilon | ||a - \hat{a}|| \ \leq \epsilon$$

Si on applique les algos précédents à $\widehat{\mathcal{A}}_\epsilon$ on a :

— Info complète : $\forall a \in \mathcal{A}, \exists \hat{a} \in \widehat{\mathcal{A}}_\epsilon | ||a - \hat{a}|| \ \leq \epsilon$:

$$\begin{aligned}
R_t^{(a)} &= \sum_{t=1}^{T} g_t(a) - g_t(a_t) \\
&= \sum_{t=1}^{T} g_t(a) - g_t(\hat{a}) + \sum_{t=1}^{T} g_t(\hat{a}) g_t(a_t) \\
&\leq cT ||a - \hat{a}||_2^\beta + 2\sqrt{T \log(\mathrm{Card}(\widehat{\mathcal{A}}_\epsilon))} \\
&\leq cT\epsilon^\beta + 2\sqrt{Td \log\left(\frac{||\mathcal{A}||}{\epsilon}\right)} \\
&\leq c + 2\sqrt{Td \log\left(||\mathcal{A}|| T^{\frac{1}{\beta}}\right)} \qquad \text{avec } \epsilon = \frac{1}{T}
\end{aligned}$$

On est content car $\max_a R_T^{(a)} \leq (O)(T)$, mais :

— complexité horrible $\simeq \frac{1}{\epsilon^d} \simeq T^{\frac{d}{\beta}}$
— dépendance en d indésirable
— choix de ϵ

— Bandits :

$$\begin{aligned}
R_t^{(a)} &\leq cT\epsilon^\beta + 2\sqrt{T \mathrm{Card}(\widehat{\mathcal{A}}_\epsilon) \log\left(\mathrm{Card}(\widehat{\mathcal{A}}_\epsilon)\right)} \\
&\leq cT\epsilon^\beta + 2\sqrt{Td \frac{||\mathcal{A}||}{\epsilon} \log\left(\frac{||\mathcal{A}||}{\epsilon}\right)}
\end{aligned}$$

On optimise en ϵ : $T\epsilon^\beta \simeq \sqrt{\frac{T}{\epsilon^\beta}} \Rightarrow \epsilon^{\beta + \frac{d}{2}} = T^{-\frac{1}{2}} \Rightarrow \epsilon = T^{-\frac{1}{2\beta+\alpha}}$

$$R_T^{(a)} \lesssim \sqrt{d} T^{1 - \frac{\beta}{2\beta+d}} \simeq \sqrt{d} T^{\frac{\beta+d}{2\beta+d}}$$

On a encore un regret $\mathcal{O}(T)$ mais il tend vers un regret linéaire quand $d \to +\infty$ ou $\beta \to 0$.

Remarque 1.5. La différence entre le regret pour un bandit et le regret pour un système d'expert :

$$\begin{aligned}
\sum_{t=1}^{T} X_t^{(\pi_t)} &= \underbrace{\sum_{t=1}^{T} X_t^{(\pi_t)} - X_t^{(k)}}_{\geq -\sqrt{TK \log K}} + \sum_{t=1}^{T} X_t^{(k)} \\
&= \underbrace{\sum_{t=1}^{T} X_t^{(\pi_t)} - X_t^{(\xi_t^{(i)})}}_{\geq -\sqrt{TK \log N}} + \sum_{t=1}^{T} X_t^{(\xi_t^{(i)})}
\end{aligned}$$

1.3.2 Descente de gradients sequentielle (Online Gradient Descent)

On se place en fonction complête.

Remarque 1.6. On peut penser en fonction de perte plutôt qu'en gain en posant $l_t = -g_t$.

Si $\mathcal{A}$ est convexe et les g_t sont concaves et G-Lipszchitz : $||\nabla g_t|| \leq G$ et différentiables, alors on peut utiliser l'algorithme de descente de gradient projeté.

Définition 1.3 (Descente de Gradient Projeté)**.**

$$a_{t+1} = \Pi_{\mathcal{A}}\Big(a_t + \eta \nabla g_t(a_t)\Big)$$

Théorème 1.2.

$$\max_{a\in\mathcal{A}} R_T^{(a)} \leq 2DG\sqrt{T}$$

où $D \geq \max_{a\in\mathcal{A}} ||a||$ *si* η *est bien choisit.*

Démonstration.

$$\begin{aligned} R_T^{(a)} &= \sum_{t=1}^{T} g_t(a) - g_t(a_t) \\ &\leq \sum_{t=1}^{T} \nabla g_t(a_t)^T(a - a_t) \end{aligned}$$

En notant $b_{t+1} = a_t + \eta\nabla g_t(a_t)$ de sorte que $a_{t+1} = \Pi_{\mathcal{A}}(b_t)$. On a :

$$R_T^{(a)} \leq \frac{1}{\eta}\sum_{t=1}^{T} \underbrace{(b_{t+1} - a_t)^T}_{x^T}\underbrace{(a - a_t)}_{y}$$

On utilise que $\langle x, y\rangle = \frac{||x||^2+||y||^2-||x-y||^2}{2}$:

$$R_T^{(a)} \leq \frac{1}{2\eta}\sum_{t=1}^{T}\Big(\underbrace{||b_{t+1} - a_t||^2}_{\eta^2||\nabla g_t(a_t)||^2 \leq \eta^2 G^2} + ||a - a_t||^2 - ||b_{t+1} - a||^2\Big)$$

Or $a_t = \Pi_{\mathcal{A}}(b_t)$ donc $||a - a_t||^2 \leq ||a - b_t||^2 \ \forall\, a \in \mathcal{A}$ car $\mathcal{A}$ est convexe.

$$\begin{aligned} R_T^{(a)} &\leq \frac{\eta G^2 T}{2} + \frac{1}{2\eta}\sum_{t=1}^{T}\Big(||a - b_t||^2 - ||a - b_{t+1}||^2\Big) \\ &\leq \frac{\eta G^2 T}{2} + \frac{||a - b_1||^2}{2\eta} \end{aligned}$$

On peut choisir $b_1 = 0$ et $\eta = \frac{D}{G\sqrt{T}}$ et donc :

$$R_T^{(a)} \leq \frac{\eta T G2}{2} + \frac{D^2}{2\eta} \leq GD\sqrt{T}$$

□

Remarque 1.7. On peut modifier la preuve pour que

$$\sum_{t=1}^{T} g_t(a_t^*) - g_t(a_t) \leq cG\sqrt{T\left(\sum_{t=1}^{T} ||a_t^* - a_{t+1}^*||^2 + ||a_1^*||^2\right)}$$

pour toute suite $a_1^*, ..., a_T^* \in \mathcal{A}$.

Remarque 1.8. Si $-g_t$ est fortement convexe, on peut avoir un regret de lordre de $\mathcal{O}(1)$.

1.3.3 Bandits linéaires

Définition 1.4 (Bandits linéaires)**.** *Correspond à $g_t(a_t) = a_t^T z_t$ où $z_t \in \mathcal{Z} \subseteq \mathbb{R}^d$ choisit par l'adversaire.*

Remarque 1.9. En info complète, on peut utiliser la descente de gradient : $G = \max_{z \in \mathcal{Z} ||z_t||^2}$:

$$R_T \leq GD\sqrt{T}$$

En bandit, on ne peut pas utiliser la descente de gradient car on observe juste $g_t(a_t) = a_t^T z_t$ et non z_t. On peut s'en sortir en discretisant l'espace $\mathcal{A}$ en $\widehat{\mathcal{A}_\epsilon}$ et en utilisant EXP sur $\widehat{\mathcal{A}_\epsilon}$.

$$\forall a \in \widehat{\mathcal{A}_\epsilon}, \quad \widehat{X_t^{(a)}} = a_t^T \widehat{z_t}$$

Comment estimer $\widehat{z_t}$? On remarque que $z_t = \mathbb{E}_{a \sim p_t}[aa^T]^{-1}\mathbb{E}_{a \sim p_t}[aa^T z_t]$. On peut choisir l'estimateur

$$\widehat{z_t} = \mathbb{E}_{a \sim p_t}[aa^T]^{-1} a_t \times \underbrace{a_t^T z_t}_{\text{reward observé}}$$

On a bien $\mathbb{E}_{a \sim p_t}[\widehat{z_t}] = z_t$ dès que $\mathbb{E}_{a \sim p_t}[aa^T]$ est inversible.

Définition 1.5. *pour $\epsilon > 0$, $\eta > 0$, $\gamma \in [0, 1]$*

- *On approche $\mathcal{A}$ par $\widehat{\mathcal{A}_\epsilon}$ de taille $K_\epsilon = \left(\frac{||\mathcal{A}||}{\epsilon}\right)^d$;*
- *On initialise p_1 uniforme sur $\widehat{\mathcal{A}_\epsilon}$.*
- *A chaque $t \geq 1$:*
  - *on choisit $a_t \in \widehat{\mathcal{A}_\epsilon}$ avec proba p_t ;*
  - *on observe $g_t(a_t) = a_t^T z_t$*
  - *on estime $\widehat{z_t} = \mathbb{E}_{a \sim p_t}[aa^T]^{-1} a_t \times a_t^T z_t$*

— *on met à jour :*

$$p_t^{(a)} = (1-\gamma)\frac{e^{\eta \widehat{R_{t-1}^{(a)}}}}{\sum\limits_{a' \in \widehat{\mathcal{A}_\epsilon}} e^{\eta \widehat{R_{t-1}^{(a')}}}} + \frac{\gamma}{d}\sum_{i=1}^{d} \mathbb{1}_{\{a\delta e_i\}}$$

où $\widehat{R_{t-1}^{(a)}} = \sum_{s=1}^{t-1} a^T \widehat{z}_t - a_t^T \widehat{z}_t$

Remarque 1.10. On a supposé $0 \in \mathcal{A}$ et $\delta e_1, ..., \delta e_d \in \widehat{\mathcal{A}_\epsilon} \subseteq \mathcal{A}$ et δ est le rayon de la plus grand boule de norme 2 centrée en 0 et incluse dans $\mathcal{A}$.

Théorème 1.3. *Pour η, γ, ϵ bien choisis, on a*

$$\max_a \mathbb{E}[R_t^{(a)}] \lesssim d\sqrt{T \log T}$$

Remarque 1.11. Avant, on avait $R_T^{(a)} \leq \mathcal{O}(T^{\frac{\beta+d}{2\beta+d}}) = \mathcal{O}(T^{\frac{1+d}{2+d}})$

Remarque 1.12. Inconvénient : mauvaise complexité, des algos efficaces sont possibles à partir de généralisation de la descente de gradient.