MVA ENS Cachan Paris Saclay

Prediction for Individual Sequences
Notes de Cours

Cours donné par Vianney Perchet
Note prises par Adrien Lina

28 janvier 2018

1 Lecture 1 : Modèle en full monitoring

Notations :

— $[n] = \{1, ..., n\}$.

1.1 Construction du problème

Nous allons construire une modélisation de bandit manchot.
On sonsidère K processus stochastiques $(X_t^{(1)})_{t\in\mathbb{N}}, ..., (X_t^{(K)})_{t\in\mathbb{N}}$ iid
On peut supposer au choix :

— $X_t^{(k)} \in [0,1]$ et $\mathbb{E}[X_t^{(k)}] = \mu^{(k)}$;
— $X_t^{(k)} \sim \mathcal{N}(\mu^{(k)}, \sigma^{(k)})$;
— $X_t^{(k)}$ est sous-Gaussien de (proxy de) variance $\sigma^{(k)}$.

Définition 1.1 (σ^2 sous-gaussienne)**.** *X est σ^2 sous-gaussien si* $\forall \lambda \in \mathbb{R}$

$$\mathbb{E}\left[e^{\lambda(X-\mathbb{E}[X])}\right] \leq e^{\frac{\lambda^2\sigma^2}{2}}$$

Exemple 1.1 (Sous-gaussienne)*.*

— Une gaussienne de paramètre (μ, σ^2) est σ^2 sous-gaussienne ;
— Si $\text{supp}(X) \subset [a,b]$, alors X est $\frac{(b-a)^2}{4}$ sous-gaussienne.

On veur créer une règle/politique/algo d'échantillonage de ces processus ("tirage de bras")

$$\pi_k : (\text{observations } 1, ..., t) \to [K] = \{1, ..., K\}$$

de sorte à maximiser

$$\sum_{t=1}^{T} X_t^{(\pi_t)}$$

En "full monitoring", les observations auxquelles l'algo a accès sont

$$\begin{gathered} X_1^{(1)},\dots, X_T^{(1)} \\ \vdots \\ X_1^{(K)},\dots, X_T^{(K)} \\ \pi_1,\dots, \pi_T \end{gathered}$$

Dans les faits, on cherche plutôt à maximiser

$$\sum_{t=1}^{T} \mu^{(\pi_t)} \approx \sum_{t=1}^{T} X_t^{(\pi_t)}$$

Néanmoins cette valeur ne représente pas très bien à quelle point on est loin d'un choix optimum d'une règle. Pour évaluer la performance d'un algo, on définit donc son regret.

Définition 1.2 (Regret)**.** *Le regret R_T d'un algorithme $(\pi_t)_{t=1,\dots,T}$ est*

$$R_t = T \max_{k \in \{1,\dots,K\}} \mu^{(k)} - \sum_{t=1}^{T} \mu^{(\pi_t)}$$

On note $\max\limits_{k\in[K]} \mu^{(k)} = \mu^*$. Ainsi

$$\begin{aligned} R_T &= T\mu^* - \sum_{t=1}^{T} \mu^{(\pi_t)} \\ &= \sum_{t=1}^{T} \underbrace{(\mu^* - \mu^{(\pi_t)})}_{\Delta^{(\pi_t)}} \end{aligned}$$

Définition 1.3 (Gap)**.** $\Delta^{(k)} = \mu^* - \mu^{(k)}$ *est le gap du bras k.*

Plus les gaps sont grands, plus il est facile de déterminer quel est le meilleurs bras, mais plus une erreur de bras coûte cher. A l'invers, plus les gaps sont petits, moins les erreurs coûtent cher, mais plus il sera dur de trouver le meilleur bras, et donc plus cette erreur sera fréquente.

Continuons :

$$\begin{aligned} R_T &= \sum_{t=1}^{T} \Delta^{(\pi_t)} \\ &= \sum_{k=1}^{K} \Delta^{(k)} N_T^{(k)} \qquad\qquad \text{avec} \;\; N_T^{(k)} = \sum_{t=1}^{T} \mathbb{1}\{\pi_t = k\} \end{aligned}$$

$N_T^{(k)}$ est le nombre de fois que l'on a choisit le bras k jusqu'au temps T.

1.2 Un exemples d'algorithme

On peut avoir plusieurs idées d'algo : un très bon algo consiste à tirer

$$\pi_k = \underset{k\in[K]}{\text{argmax}}\, \overline{X_{t-1}^{(k)}} \qquad\qquad \text{avec}\; \overline{X_{t-1}^{(k)}} = \frac{1}{n}\sum_{t'=1}^{t-1} X_{t'}^{(k)}$$

Théorème 1.1. *Pour l'algorithme ci-dessus, dans le cas où $k = 2$, l'espérance du regret est bornée en*

$$\begin{aligned} \mathbb{E}[R_T] &\leq \frac{\Delta}{2} + \frac{4\sigma^2}{\Delta} \\ &\lesssim \Delta + \frac{\sigma^2}{\Delta} \end{aligned}$$

avec $\Delta = \max(\Delta^{(1)}, \Delta^{(2)}) = |\mu^{(1)} - \mu^{(2)}|$

Définition 1.4. *Nous notons $x \lesssim y$ s'il existe $\alpha \in \mathbb{R}_+$ tel que $x \leq \alpha y$.*

Afin de prouver Théorème 1.1, nous avons besoin du lemme suivant, qui n'est pas démontré ici.

Lemme 1.1 (Hoeffding)**.** *Si $X_1, ..., X_t$ sont iid et σ^2 sous-Gaussiens.*

On a :

$$\mathbb{P}\left(\overline{X_t} - \mathbb{E}[X] \geq \epsilon\right) \leq e^{-\frac{t\epsilon^2}{2\sigma^2}}$$

i.e. $\forall \delta \in [0, 1]$,

$$\mathbb{P}\left(\overline{X_t} - \mathbb{E}[X] \geq \sqrt{\frac{2\sigma^2 \log(1/t)}{t}}\right) \leq \delta$$

Théorème 1.1. Sans perte de généralité, on suppose que le meilleur bras est le bras 1 ($k^* = 1$).

On cherche à controler

$$N_t^{(2)} = \sum_{t=1}^{T} \mathbb{1}\{\pi_t = 2\}$$

Or

$$\{\pi_t = 2\} = \{\pi_1 = 2\} \cup \{\overline{X_{t-1}^{(1)}} \leq \overline{X_{t-1}^{(2)}}, t \geq 2\}$$

Or on a

- $\mathbb{P}(\pi_1 = 2) = \frac{1}{2}$
- Et

$$\begin{aligned}\mathbb{P}\left(\overline{X_{t-1}^{(1)}} \leq \overline{X_{t-1}^{(2)}}\right) &= \mathbb{P}\left(\overline{X_{t-1}^{(2)}} - \overline{X_{t-1}^{(1)}} \geq 0\right) \\ &= \mathbb{P}\left(\frac{1}{t}\sum_{t'=1}^{t}(X_{t'}^{(2)} - X_{t'}^{(1)}) \geq 0\right)\end{aligned}$$

Or on sait que $\mathbb{E}[X_{t'}^{(2)} - X_{t'}^{(1)}] = -\Delta$ et que $X_{t'}^{(2)} - X_{t'}^{(1)}$ est $2\sigma^2$ sous-Gaussienne.

Avec Lemme 1.1

$$\begin{aligned}\mathbb{P}(\overline{\Delta} \leq \epsilon) &= \mathbb{P}(\overline{\Delta_t} - (-\Delta) \geq \Delta) \\ &\leq e^{-\frac{t\Delta^2}{4\sigma^2}}\end{aligned}$$

Donc

$$\begin{aligned}E[N_T^{(2)}] &\leq \frac{1}{2} + \underbrace{\sum_{t=1}^{T} e^{\frac{t\Delta^2}{4\sigma^2}}}_{\xrightarrow[T\to+\infty]{}} \\ &\leq \frac{1}{2} + \frac{1}{e^{\frac{\Delta^2}{4\sigma^2}} - 1}\end{aligned}$$

Enfin, en utilisant $e^x \geq 1 + x$ on obtient :

$$E[N_T^{(2)}] \leq \frac{1}{2} + \frac{4\sigma^2}{\Delta^2}$$

□

Cet algorithme est "optimal" au sens où il n'existe pas d'algo avec un regret espéré *toujours* plus petit que $\frac{1}{\Delta}$ (si on oublie le premier tirage).

1.3 Borne inférieure pour tout algorithme

On peut prouver le résultat suivant, qui nous indique qu'il n'existe pas d'algorithme qui trouvera, quel que soit le bandit, le meilleur bras en moins qu'une certaine borne inférieure.

Théorème 1.2 (Borne inférieure)**.** *Pour tout algorithme il existe un choix de* $(\mu^{(1)}, \mu^{(2)})$ *tel que* $|\mu^{(1)} - \mu^{(2)}| = \Delta$ *et que*

$$\begin{aligned}\mathbb{E}[R_t] &\geq \frac{1}{4\Delta}\left(1 - e^{-T\Delta^2}\right) \\ &\geq \frac{1}{8\Delta} && \text{si } T \geq \frac{\log(2)}{\Delta^2}\end{aligned}$$

De manière plus précise, on peut choisir $X_t^{(1)} \sim \mathcal{N}(0,1)$ *et* $X_t^{(2)} \sim \mathcal{N}(\Delta,1)$ ***ou*** $X_t^{(1)} \sim \mathcal{N}(\Delta,1)$ *et* $X_t^{(2)} \sim \mathcal{N}(0,1)$.

Démonstration.

$$\max\left(\mathbb{E}_1[R_T], \mathbb{E}_1[R_T]\right) \geq \frac{\mathbb{E}_1[R_T] + \mathbb{E}_1[R_T]}{2}$$

où $\mathbb{E}_i[R_T]$ est le regret dans le monde i.

$$\mathbb{E}_1[R_T] + \mathbb{E}_1[R_T] = \Delta \sum_{t=1}^{T} \left[\mathbb{P}_1(\pi_t = 1)) + \mathbb{P}_2(\pi_t = 1)\right]$$

On peut appliquer Lemme 1.2 ici : à l'instant t

$$V_1 = (\mathcal{N}(\Delta, 1) \otimes \mathcal{N}(0, 1))^{\otimes t-1}$$
$$V_2 = (\mathcal{N}(0, 1) \otimes \mathcal{N}(\Delta, 1))^{\otimes t-1}$$

Où $\otimes$ dénote le produit entre distributions indépendantes.

$$\begin{aligned} KL(V_1, V_2) &= (t-1)\left[KL(\mathcal{N}(\Delta, 1), \mathcal{N}(0, 1)) + KL(\mathcal{N}(0, 1), \mathcal{N}(\Delta, 1))\right] \\ &= (t-1)\left[\frac{\Delta^2}{2} + \frac{\Delta^2}{2}\right] \\ &= (t-1)\Delta^2 \end{aligned}$$

Or ceci est le maximum de $\sum_{t=1}^{T} \mathbb{P}_1(\pi_t = 1)) + \mathbb{P}_2(\pi_t = 1)$, c'est donc la borne inf la plus fine du regret espéré. Exactement, cette borne inf est

$$\begin{aligned} \frac{\Delta}{2} \sum_{t=1}^{T} \frac{1}{2} e^{-(t-1)\Delta^2} &= \frac{\Delta}{4} \frac{1 - e^{-T\Delta^2}}{1 - e^{-\Delta^2}} \\ &\geq \frac{1 - e^{-T\Delta^2}}{4\Delta} \end{aligned}$$

□

Remarque 1.1. On peut donc se demander quel est le pire cas. On a vu que

$$\mathbb{E}[R_t] \leq \alpha\left(\left(\frac{1}{\Delta} + \Delta\right)\right) \wedge N)$$
$$\mathbb{E}[R_t] \geq \beta\left(\frac{1 - e^{-T\Delta^2}}{4\Delta}\right)$$

Quand T est fixé, le pire Δ possible est $\Delta = \frac{1}{\sqrt{T}}$ et on obient que

$$\alpha\sqrt{T} \leq \mathbb{E}[R_T] \leq \beta\sqrt{T}$$

Avec ce cas, l'aglo se plante parce qu'on a pas assez de données pour déterminer quel est le bon bras.

Lemme 1.2. *Étant donné $Z \in \mathbb{R}^d$ généré soit par une distribution V_1 sur $\mathbb{R}^d$ soit par une distribution V_2 et un test $\phi : \mathbb{R}^d \to \{1, 2\}$, alors*

$$\mathbb{P}_{V_1}(\phi = 2) + \mathbb{P}_{V_2}(\phi = 1) \geq \frac{1}{2} e^{-KL(V_1,V_2)}$$

où

$$KL(V_1, V_2) = \int_{\mathbb{R}^d} \log(\frac{dV_1}{dV_2}) dV_1$$

De plus, le ϕ^* qui minimise l'erreur de test est $\operatorname{argmax} dV_1(z)dV_2(z)$ *et vérifie*

$$\mathbb{P}_1(\phi^* = 2) + \mathbb{P}_2(\phi^* = 1) = \int_{\mathbb{R}^d} \min(dV_1, dV_2)$$

Lemme 1.2. On a :

$$\begin{aligned}\int \sqrt{dV_1 dV_2} &= \int \sqrt{\min(dV_1, dV_2) \max(dV_1, dV_2)} \\ &\leq \int \min(dV_1, dV_2) \int \max(dV_1, dV_2) \qquad \text{(Cauchy-Schwartz)}\end{aligned}$$

Mais

$$\int (\min(dV_1, dV_2) + \max(dV_1, dV_2)) = \int (dV_1 + dV_2) = 2$$

Donc

$$\left(\int \sqrt{dV_1 dV_2}\right)^2 \leq \int \min(dV_1, dV_2) \left(2 - \int \min(dV_1, dV_2)\right)$$

Ainsi :

$$\int_{\mathbb{R}^d} \min(dV_1, dV_2) \geq \frac{1}{2} \left(\int \sqrt{dV_1 dV_2}\right)^2$$

D'où

$$\begin{aligned}\mathbb{P}_1(\phi^* = 2) + \mathbb{P}_2(\phi^* = 1) &\geq \frac{1}{2} \left(\int \sqrt{dV_1 dV_2}\right)^2 \\ &\geq \frac{1}{2} \exp\left(2 \log\left(\int \sqrt{dV_1 dV_2}\right)\right) \\ &\geq \frac{1}{2} \exp\left(2 \log\left(\int \sqrt{\frac{dV_2}{dV_1}} dV_1\right)\right) \\ &\geq \frac{1}{2} \exp\left(2 \int \log\left(\sqrt{\frac{dV_2}{dV_1}}\right) dV_1\right) \\ &\geq \frac{1}{2} \exp\left(-KL(V_1, V_2)\right)\end{aligned}$$

Pour l'égalité supplémentaire, et sans détail concernant la première égalité du calcul (un dessin donne l'intuition) :

$$\begin{aligned}\mathbb{P}_1(\phi^* = 2) + \mathbb{P}_2(\phi^* = 1) &= \int_{\mathbb{R}^d \ | dV_1 < dV_2} dV_1 + \int_{\mathbb{R}^d \ | dV_1 > dV_2} dV_2 \\ &= \int_{\mathbb{R}^d} \mathbb{1}\{dV_1 < dV_2\} dV_1 + \int_{\mathbb{R}^d} \mathbb{1}\{dV_1 > dV_2\} dV_2 \\ &= \int_{\mathbb{R}^d} \min(dV_1, dV_2)\end{aligned}$$

□