MVA ENS Cachan Paris Saclay

Prediction for Individual Sequences
Notes de Cours

Cours donné par Vianney Perchet
Note prises par Adrien Lina

16 février 2018

1 Lecture 3 : Bandits contextuels

1.1 Rappels des lectures précédentes

— K bras de retours $X_t^k \in [0,1]$ avec $\mathbb{E}[X_t^k] = \mu_k$
— $R_t = \max\limits_{k\in[K]} T\mu^k - \sum_{t=1}^T \mu_{\pi_t} = \sum_{t=1}^T \Delta_{\pi_t}$
— UCB : $\pi_t = \operatorname{argmax}\limits_{k\in[K]} \overline{X_t^k} + \sqrt{\frac{2\log(t)}{N_t^k}}$. On obtient :

$$\mathbb{E}[R_t] \lesssim \max\left\{\sum_{k=2}^K \frac{\log(t)}{\Delta_k}, \sqrt{kT\log(T)}\right\}$$

— ETC : On alterne entre les bras. Si

$$\overline{X_t^i} + \sqrt{\frac{2\log(\frac{t}{N_t^k})}{N_t^i}} \leq \overline{X_t^j} - \sqrt{\frac{2\log(\frac{t}{N_t^j})}{N_t^j}}$$

alors on retire le bras k. On sait que cela arrive au temps où

$$\mu^k + \sqrt{\frac{2\log(\frac{t}{N_t^k})}{N_t^k}} \approx \mu^j - \sqrt{\frac{2\log(\frac{t}{N_t^j})}{N_t^j}}$$

donc on élimine k après $\approx \frac{\log(T\Delta_k^2)}{\Delta_k^2}$ samples.

Remarque 1.1. Pour l'ETC, on n'est donc pas obligé de regarder exactement à chaque étape que la condition est vérifiée : on peut la regarder que toutes les 2^m itérations.

On peut donc réduire le nombre de changement de bras en prenant sur une période $[2^m, 2^{m+1}]$: par exemple pour 2 bras, on prend le bras 1 pour les 2^m

itérations, puis le bras 2 pour les 2^m itérations suivante, et enfin faire le test à la fin.

Cela peut être intéressant pour un bandit manchot où le temps de retour de chaque bras est long. Par exemple, pour des essais cliniques où il faudrait 6 mois pour connaître le résultat, UCB demande à attendre 6 mois avant de donner le médicament au $2^{\text{ème}}$ patient, puis 6 autres mois pour le $3^{\text{ème}}$, etc. Avec la stratégie présentée ci-dessus, on peut faire des essais par blocs pour faire un plus grand nombre d'essais en 6 mois, mais toujours avec un regret intéressant.

On obtient

$$\mathbb{E}[R_T] \lesssim \sum_{k=2}^{K} \frac{\log(T\Delta_k^2)}{\Delta_k}$$
$$\Rightarrow \mathbb{E}[R_T] \lesssim \sqrt{TK\log(K)}$$

1.2 Variantes d'algo de bandits

1.2.1 Algorithme ϵ-greedy

$$\pi_k = \begin{cases} \sim \mathcal{U}([K]) & \text{avec proba } \epsilon \\ \operatorname{argmax} \overline{X_t^k} & \text{avec proba } 1-\epsilon \end{cases}$$

En général, $\epsilon_t \in \Omega(\frac{1}{t})$.

1.2.2 Variantes de UCB

— MOSS : $\sqrt{\frac{\log(\frac{T}{KN_t^k})}{N_t^k}}$, $\Rightarrow \mathbb{E}[R_T] \leq \sqrt{TK}$

— KL-UCB : si $X_t^k \sim Ber(\mu^k)$ On remplace $\overline{X_t^k} + \sqrt{\frac{2\log(t)}{N_t^k}}$ par

$$\max\left\{\mathbb{E}_Q[X]; KL\left(Q, \overline{X_t^k}\right) \leq f(t, N_t^k)\right\}$$

1.2.3 Thompson sampling

L'algorithme Thomson a une approche bayésienne :

— Prior p_0 sur $(\mu^1, ..., \mu^K)$.

— On obtient $X_1^{\pi_1}, ..., X_t^{\pi_t}$ du bandit.

— On calcule le posterior p_t sur $(\mu^1, ..., \mu^K)$

— On tire $(\mu_{t+1}^1, ..., \mu_{t+1}^K)$ suivant p_t.

— On choisit $\pi_{t+1} = \underset{k\in[K]}{\operatorname{argmax}}\, \mu_{t+1}^k$.

En théorie, cela fonctionne bien. En pratique cela fonctionne encore mieux, malgré le fait qu'il y ait beaucoup de calculs par étapes (le calcul des posteriors).

Le problème principal de cette méthode est le sampling par rapport à une distribution, ce qui est problématique en grande dimension.

1.2.4 Conclusion

En pratique, quand confronté à un problème de bandit, il faut se souvenir de 3 algorithmes :

— UCB ;
— ETC ;
— Thomson.

1.3 Bandits contextuels : contexte

On a un espace de contexte Ω qui nous permet de déterminer le "type" de contexte. A chaque étape, on observe $\omega_t \in \Omega$ qui nous permet de choisir $\pi_t \in [K]$ de sorte à maximiser notre revenu.

Par exemple, dans un contexte de pubs à choisir, le contexte est l'age et le sexe de la personne, l'heure de la journée, etc... Il nous faut alors déterminer quel publicité choisir au vu de ce contexte.

Notation : $\mathbb{E}[X^k|\omega] = \mu^k(\omega)$.

Hypothèse 1.1. *ω_t I.I.D. de loi $\mathcal{D}$ connue sur Ω* [1] ***ou** de densité bornée inférieurement et supérieurement.*

Pour simplifier les calculs, on va supposer ici $\Omega = [O, 1]^d$ et $\mathcal{D}$ est uniforme sur Ω.

La stratégie optimale est

$$\pi^*(\omega) = \underset{k\in[K]}{\operatorname{argmax}}\, \mu^k(\omega)$$

ce qui donne un regret

$$R_T = \sum_{t=1}^{T} \mu^{\pi^*(\omega_t)}(\omega_t) - \sum_{t=1}^{T} \mu^{\pi_t}(\omega_t)$$

Sans hypothèses de régularité sur les $\mu^k(.)$ on ne pourra pas trouver une stratégie efficace car on ne peut interpoler efficacement une fonction quelconque avec un nombre fini de points.

Hypothèse 1.2 (Hypothèse de Régularité sur les $\mu^k(.)$)**.** *Nous allons considérer le cadre où les $\mu^k(.)$ sont β-Holder où les paramètres L et β sont connus.*

$$|\mu^k(\omega_1) - \mu^k(\omega_2)| \leq L|\omega_1 - \omega_2|^\beta \quad \forall(\omega_1, \omega_2) \in \Omega^2$$

Remarque 1.2. On peut choisir en fonction du problème une continuité de type :

— Lipzshitz : $|f(x) - f(y)| \leq L|x - y|\ \forall(x, y)$
— Holder : $|f(x) - f(y)| \leq L|x - y|^k\ \forall(x, y)$
— Logistique : $\mu^t(\omega) = \frac{e^{\omega^T\theta_k}}{1+e^{\omega^T\theta_k}}$ avec $\theta_k \in \mathbb{R}^d$
— Quadratique
— Smooth
— *Etc.*

1. En pratique, la connaissance de $\mathcal{D}$ n'est pas un problème : on peut estimer la densité avec les observations de ω, car les ω_t sont indépendant des π_t.

1.4 Un algorithme contextuel

Proposition 1.1. *Sous ces hypothèses, il existe un algo Binned-ETC dont le regret est*

$$\mathbb{E}[R_t] \lesssim T\left(\frac{K\log(K)}{T}\right)^{\frac{\beta}{2\beta+d}}$$

Remarque 1.3. Et donc,

- si $d = 0$ (pas de contexte), $\mathbb{E}[R_t] \lesssim \sqrt{TK\log(K)}$ on retrouve le regret d'ETC normal ;
- si $d = +\infty$ (contexte infiniment grand), $\mathbb{E}[R_t] \lesssim T$, c'est-à-dire qu'on n'arrivera jamais à trouver une stratégie par rapport au contexte et qu'on choisira toujours un bras au hasard.

Remarque 1.4. Ce sont des vitesses "pire cas" (indépendant des distributions). On pourrait se demander quelle est la vitesse par rapport à une distribution.

Dans le cas "multi-bras", les paramètres de complexité d'un problème étaient $\Delta_1, \Delta_2, ..., \Delta_K$, c'est-à-dire les $\mu^* - \mu^k$. Ici, on cherche à identifier $\max \mu_k(\omega)$ à ω donné. La complexité réside dans la proximité de $\max \mu_k(\omega)$ avec le deuxième maximum $\mu^{\#}(\omega) = \max\{\mu^j(\omega); \mu^j(\omega) < \mu^*(\omega)\}$.

Hypothèse 1.3. *On va paramétrer le problème par une condition de marge $\alpha \in \mathbb{R}_+$ tel que,*

$$\forall \Delta \leq cste, \quad \mathbb{P}_{\mathcal{D}}(0 < \mu^*(\omega) - \mu^{\#}(\omega) \leq \Delta) \leq c\Delta^{\alpha}$$

Remarque 1.5. Ainsi, plus α est grand, plus le problème est facile.

Remarque 1.6. On a $\alpha\beta \leq d$ **ou** $\alpha = +\infty$ nécessairement.

Preuve (Proposition 1.1). On ne montre pas $\alpha \geq 1$, qui est nettement plus compliqué. On montre le résultat pour $\alpha \leq 1$, $k = 2$.

On va utiliser la technique du regressogramme. On partitionne $\Omega = [0,1]^d$ en carrés ("bins") de côtés $\epsilon \in [0,1]$ et on traite chaque carré de manière indépendante. On lance un ETC par bin.

À l'étape t, le contexte ω_t tombe dans la bin $b_t \in B$. C'est l'algo ETC_{b_t} qui est actif, i.e. c'est le seul des ETC qui choisit le bras et qui se met à jour.

On approche la fonction $\mu^k(\omega)$ par des fonctions constantes par morceau $\overline{\mu_b}^k$ sur le bin b où $\overline{\mu_b}^k = \mathbb{E}[\mu^k(\omega)|\omega \in b]$.

ETC_b va minimiser le regret

$$\max \sum_{t:w_t \in b} \overline{\mu_b}^k - \overline{\mu_b}^{\pi_t}$$

Par ailleurs, comme les $\mu^k(.)$ sont β Holder,

$$\begin{aligned} |\mu^k(\omega) - \overline{\mu_b}^k| &= \left| \int_{b\in B} \mu^k(\omega) - \mu^k(v)dv \right| \\ &\leq L\epsilon^{\beta} \end{aligned}$$

et, de la même manière,

$$|\mu^{\pi_t}(\omega) - \overline{\mu_b}^*| \leq L\epsilon^\beta$$

Donc,

$$\begin{aligned}
\mathbb{E}[R_t] &\leq \sum_{b\in B} \sqrt{N_t(b)K\log(K)} + L\epsilon^\beta T \\
&\leq \sqrt{\sum_{b\in B} 1}\sqrt{\sum_{b\in B} N_t(b)K\log(K)} + L\epsilon^\beta T \qquad \text{(Cauchy-Shwarz)} \\
&\lesssim \frac{1}{\epsilon^{d/2}}\sqrt{TK\log(K)} + \epsilon^\beta T
\end{aligned}$$

Et si on choisit

$$\frac{TK\log(K)}{\epsilon^d} = \epsilon^{2\beta}T^2 \quad \Rightarrow \epsilon^{2\beta+d} = \frac{K\log(K)}{T}$$

donc

$$\mathbb{E}[R_t] \lesssim T\left(\frac{K\log(K)}{T}\right)^{\frac{\beta}{2\beta+d}}$$

On a donc montré le résultat pour $\alpha = 0$. Montrons le avec la condition de marge $\alpha > 0$.

Le regret de ETC_b est inférieur à $\max\left\{\frac{\log(N_T(b)\Delta_b^2)}{\Delta_b}, \sqrt{N_T(b)K\log K}\right\}$.

Si

$$\sqrt{N_T(b)K\log K} \leq \frac{\log(N_T(b)\Delta_b^2)}{\Delta_b}$$

alors on a

$$\Delta_b \lesssim \frac{\log(K)}{\sqrt{N_T(b)K\log(K)}} = \sqrt{\frac{\log(k)}{KN_T(b)}}$$

Pour $K = 2$, on obtient $\Delta_b \lesssim \frac{1}{\sqrt{N_T(b)}}$ ce qui arrive avec petite proba (condition de marge).

Ainsi, on va décomposer l'espace B en 2 :

$$\begin{aligned}
B_1 &:= \{b \in B : \Delta_b \leq \underline{\Delta}\} \\
B_2 &:= \{b \in B : \Delta_b > \underline{\Delta}\}
\end{aligned}$$

Pour les bins dans B_2, on ordonne les bins par valeurs croissantes de leurs Δ_b :

$$\Delta_{(1)} \leq \Delta_{(2)} \leq ... \leq \Delta_{(1/\epsilon^d)}$$

On souhaite borner inférieurement $\Delta_{(i)}$. On a $\mathbb{P}(0 < \Delta(\omega) \leq \epsilon) \leq \epsilon^\alpha$, donc $\mathbb{P}(0 < \Delta(\omega) \leq \Delta_{(i)}) \leq \Delta_{(i)}^\alpha$.

Parce qu'on a ordonné les bins, on obtient $\mathbb{P}(0 < \Delta(\omega) \leq \Delta_{(j)}) \gtrsim j \times$ (taille d'un bin) $= j\epsilon^d$. Ainsi, on obtient donc que $\Delta_{(j)} \geq (j\epsilon^d)^{\frac{1}{\alpha}}$

Pour les bins dans B_1,

$$\mathbb{P}\left(w \in \bigcup_{b \in B_1} b\right) \leq P(\Delta(\omega) < \underline{\Delta}) \leq \underline{\Delta}^{\alpha}$$

et si $b \in B_2$, $\Delta_b \geq (j\epsilon^d)^{1/\alpha}$ si Δ_b est le $j^{\text{ème}}$ plus grand.

Donc au final,

$$\mathbb{E}[R_T] \leq (T \times \underline{\Delta}^{\alpha})\underline{\Delta} + \sum_{b \in B_2} \frac{\log(T\Delta_b^2)}{\Delta_b}$$

□

Remarque 1.7.

$$\underbrace{\Delta_{(1)} \leq ... \leq \Delta_{(j*-1)}}_{\in B_1} \leq \underline{\Delta} \leq \underbrace{\Delta_{(j*)} \leq ... \leq \Delta_{(\frac{1}{\epsilon^d})}}_{\in B_2}$$

$$\mathbb{E}[R_T] \leq T\underline{\Delta}^{1+\alpha} + \sum_{j=\frac{\underline{\Delta}^{\alpha}}{\epsilon^d}}^{1/\epsilon^d} \frac{\log(T\epsilon(j\epsilon^d)^{2/\alpha})}{(j\epsilon)^{1/\alpha}}$$