

Problèmes de bandits en Apprentissage Séquentiel

Antoine BARRIER

doctorant à l'UMPA (ÉNS de Lyon) et au LMO (Université Paris-Saclay)

antoine.barrier@ens-lyon.fr – <http://perso.ens-lyon.fr/antoine.barrier/fr/>

UMPA

ENS
ENS DE LYON

m Mathématiques
Orsay

université
PARIS-SACLAY

15 décembre 2021, ÉNS de Lyon

2014 → 2016 CPGE MPSI/MP

Lycée Champollion, Grenoble

2016 → 2020 Études en mathématiques

ÉNS Paris-Saclay, Cachan

- licence de maths
- agrégation
- master MVA "Mathématiques, Vision et Apprentissage"

2020 → ... Doctorat en mathématiques

UMPA, Lyon

sous la direction d'Aurélien Garivier et de Gilles Stoltz

APPRENTISSAGE STATISTIQUE

Ensemble de techniques (algorithmes) permettant à une machine de comprendre des données dans le but d'optimiser une performance par rapport à un critère donné

Exemples : reconnaissance d'objets, prédiction de la météo, ...

Domaine très vaste : réseaux de neurones, apprentissage par renforcement, détection de communautés, ...

APPRENTISSAGE SÉQUENTIEL

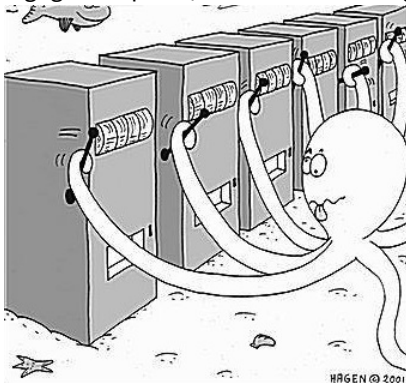
Un sous-ensemble des techniques d'apprentissage statistique pour lequel les données arrivent les unes après les autres et on choisit quelles futures données on souhaite observer.

Les problèmes de bandits

Contexte original : le casino

K machines à sous (bandits manchots) différentes

Quel bras tirer pour gagner le plus (perdre le moins ...) d'argent à long terme ?

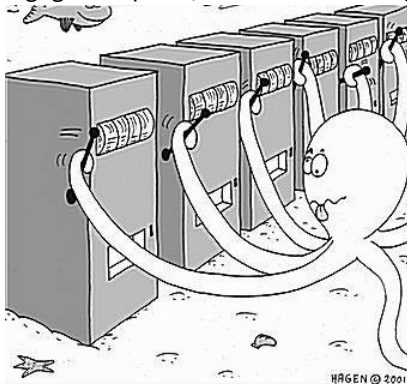


Les problèmes de bandits

Contexte original : le casino

K machines à sous (bandits manchots) différentes

Quel bras tirer pour gagner le plus (perdre le moins ...) d'argent à long terme ?



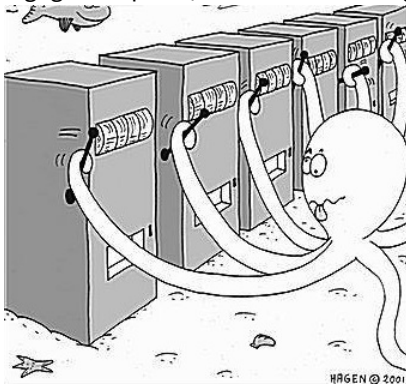
→ on tire chaque machine quelques fois puis selon les gains obtenus on peut être tenté de se focaliser sur une machine.
À raison ?

Les problèmes de bandits

Contexte original : le casino

K machines à sous (bandits manchots) différentes

Quel bras tirer pour gagner le plus (perdre le moins ...) d'argent à long terme ?



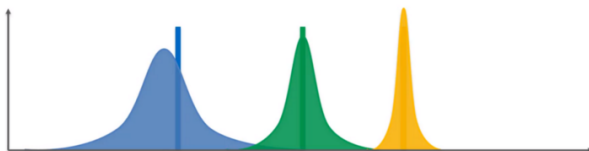
→ on tire chaque machine quelques fois puis selon les gains obtenus on peut être tenté de se focaliser sur une machine.

À raison ? **Dilemme exploration vs exploitation**

Les problèmes de bandits

Modélisation mathématique

- K groupes (bras). À chaque groupe $k \in \{1, \dots, K\}$ on associe une distribution de probabilité ν_k (inconnue!). On note $\mu_k = \mathbb{E}[\nu_k]$ son espérance. On supposera que $\mu_1 > \max_{k \in \{2, \dots, K\}} \mu_k$.



- À tout instant $t = 1, 2, \dots$, le joueur choisit un bras $A_t \in \{1, \dots, K\}$ en fonction des observations passées et reçoit une récompense X_t qui est une réalisation **aléatoire** de loi ν_{A_t} .

→ **apprentissage séquentiel : la manière dont le joueur choisit les bras à tirer influe sur les observations que l'on va faire**

Les problèmes de bandits

Un premier objectif : le regret

Dans le cadre des machines à sous, si on se fixe un nombre de tirages T :

- à l'instant t , en tirant A_t on a une perte en moyenne de $\mu_1 - \mu_{A_t}$ par rapport à la meilleure option qu'est le bras 1.
- en cumulant sur tous les temps on a une perte relative (un *regret*) de

$$R(T) = \sum_{t=1}^T \mu_1 - \mathbb{E}[\mu_{A_t}] = T\mu_1 - \sum_{t=1}^T \mathbb{E}[\mu_{A_t}] = \sum_{k \neq 1} (\mu_1 - \mu_k) \mathbb{E}[N_k(T)]$$

où $N_k(T)$ est le nombre de tirages du bras k après T étapes.

Les problèmes de bandits

Un premier objectif : le regret

Dans le cadre des machines à sous, si on se fixe un nombre de tirages T :

- à l'instant t , en tirant A_t on a une perte en moyenne de $\mu_1 - \mu_{A_t}$ par rapport à la meilleure option qu'est le bras 1.
- en cumulant sur tous les temps on a une perte relative (un *regret*) de

$$R(T) = \sum_{t=1}^T \mu_1 - \mathbb{E}[\mu_{A_t}] = T\mu_1 - \sum_{t=1}^T \mathbb{E}[\mu_{A_t}] = \sum_{k \neq 1} (\mu_1 - \mu_k) \mathbb{E}[N_k(T)]$$

où $N_k(T)$ est le nombre de tirages du bras k après T étapes.

Un bon algorithme doit minimiser ce regret :

- si $R(T) \simeq cT$ est linéaire en T : la proportion de tirages des mauvais bras ne diminue pas au cours du temps donc on n'apprend rien
- si $R(T) = o(T)$ est sous-linéaire en T : on tire de moins en moins les mauvais bras!

FOLLOW-THE-LEADER

On tire chaque bras une fois puis ensuite on choisit A_t le bras de meilleur moyenne empirique avant t .

Les problèmes de bandits

L'exemple de Follow-The-Leader

FOLLOW-THE-LEADER

On tire chaque bras une fois puis ensuite on choisit A_t le bras de meilleur moyenne empirique avant t .

Exemple : 2 bras, $\nu_1 = \mathcal{B}(0.6)$ et $\nu_2 = \mathcal{B}(0.4)$: on a $A_1 = 1$ et $A_2 = 2$

$$\mathbb{P}(X_1 = 1) = 0,6 \quad \mathbb{P}(X_1 = 0) = 0,4 \quad \mathbb{P}(X_2 = 1) = 0,4 \quad \mathbb{P}(X_2 = 0) = 0,6$$

$X_2 \backslash X_1$	1	0
1	$R(T) = ???$ proba 0.24	$R(T) = T - 1$ linéaire proba 0.16
0	$R(T) = 1$ proba 0.36	$R(T) = ???$ proba 0.24

→ regret linéaire avec probabilité assez forte : mauvais algorithme

Les problèmes de bandits

L'exemple de Follow-The-Leader

FOLLOW-THE-LEADER

On tire chaque bras une fois puis ensuite on choisit A_t le bras de meilleure moyenne empirique avant t .

Exemple : 2 bras, $\nu_1 = \mathcal{B}(0.6)$ et $\nu_2 = \mathcal{B}(0.4)$: on a $A_1 = 1$ et $A_2 = 2$

$$\mathbb{P}(X_1 = 1) = 0,6 \quad \mathbb{P}(X_1 = 0) = 0,4 \quad \mathbb{P}(X_2 = 1) = 0,4 \quad \mathbb{P}(X_2 = 0) = 0,6$$

$X_2 \backslash X_1$	1	0
1	$R(T) = ???$ proba 0.24	$R(T) = T - 1$ linéaire proba 0.16
0	$R(T) = 1$ proba 0.36	$R(T) = ???$ proba 0.24

→ regret linéaire avec probabilité assez forte : mauvais algorithme

Néanmoins la minimisation du regret est un problème bien compris aujourd'hui → beaucoup d'algorithmes performants

Les problèmes de bandits

D'autres cas d'applications

PUBLICITÉ, RECOMMANDATION DE FILMS, ...

K publicités différentes $\rightarrow$ laquelle intéresse le plus, fait cliquer le plus d'utilisateurs?

Récompense X_{A_t} : 1 si l'utilisateur clique, 0 sinon.

MÉDECINE

K vaccins $\rightarrow$ lequel est le plus efficace, immunise la proportion la plus forte de la population?

Récompense X_{A_t} (par exemple) : 1 si le patient est immunisé, 0 sinon.

Les problèmes de bandits

D'autres cas d'applications

PUBLICITÉ, RECOMMANDATION DE FILMS, ...

K publicités différentes → laquelle intéresse le plus, fait cliquer le plus d'utilisateurs?

Récompense X_{A_t} : 1 si l'utilisateur clique, 0 sinon.

MÉDECINE

K vaccins → lequel est le plus efficace, immunise la proportion la plus forte de la population?

Récompense X_{A_t} (par exemple) : 1 si le patient est immunisé, 0 sinon.

phase de test sur un panel puis mise en production

→ **changement d'objectif : on veut trouver le plus vite possible la meilleure pub / le meilleur vaccin**

Les problèmes de bandits

Un deuxième objectif : l'identification du meilleur bras

À cause de l'aléa, on ne peut pas être sûr d'identifier le meilleur bras. Ce que l'on peut faire à la place :

- 1 on fixe un niveau de confiance $\delta > 0$ (petit) et on cherche un algorithme de bandits qui finit par s'arrêter, i.e. prend une décision à un instant (aléatoire) τ et retourne un bras k_τ tel que $\mathbb{P}(k_\tau \neq 1) \leq \delta$: l'algorithme identifie le meilleur bras avec forte probabilité.

Qualité d'un algorithme : à quel point le nombre moyen de tirages $\mathbb{E}[\tau]$ est petit.

- 2 on fixe un budget, un nombre total de tirages T , et on cherche un algorithme de bandits qui après T étapes choisit un bras k_T qu'il estime être le meilleur.

Qualité d'un algorithme : on veut que la probabilité $\mathbb{P}(k_T \neq 1)$ soit la plus faible possible.

Les problèmes de bandits

Un deuxième objectif : l'identification du meilleur bras

À cause de l'aléa, on ne peut pas être sûr d'identifier le meilleur bras. Ce que l'on peut faire à la place :

- 1 on fixe un niveau de confiance $\delta > 0$ (petit) et on cherche un algorithme de bandits qui finit par s'arrêter, i.e. prend une décision à un instant (aléatoire) τ et retourne un bras k_τ tel que $\mathbb{P}(k_\tau \neq 1) \leq \delta$: l'algorithme identifie le meilleur bras avec forte probabilité.

Qualité d'un algorithme : à quel point le nombre moyen de tirages $\mathbb{E}[\tau]$ est petit.

- 2 on fixe un budget, un nombre total de tirages T , et on cherche un algorithme de bandits qui après T étapes choisit un bras k_T qu'il estime être le meilleur.

Qualité d'un algorithme : on veut que la probabilité $\mathbb{P}(k_T \neq 1)$ soit la plus faible possible.

→ **contexte de ma thèse : on cherche**

- des bornes inférieures : quelque soit l'algorithme il ne pourra pas faire mieux qu'une certaine borne,
- des algorithmes dont les garanties théoriques sont proches de la borne inférieure, donc optimaux en un certain sens.

PUBLICITÉ, RECOMMANDATION DE FILMS, ...

Chaque utilisateur est intéressé par différentes publicités!

En pratique si on a des informations sur l'utilisateur (we love cookies), on veut les utiliser pour lui donner la publicité la plus ciblée possible.

→ Comment inclure les informations utilisateurs dont on dispose dans notre modèle?

PUBLICITÉ, RECOMMANDATION DE FILMS, ...

Chaque utilisateur est intéressé par différentes publicités!

En pratique si on a des informations sur l'utilisateur (we love cookies), on veut les utiliser pour lui donner la publicité la plus ciblée possible.

→ Comment inclure les informations utilisateurs dont on dispose dans notre modèle?

→ bandits linéaires

PUBLICITÉ, RECOMMANDATION DE FILMS, ...

Chaque utilisateur est intéressé par différentes publicités!

En pratique si on a des informations sur l'utilisateur (we love cookies), on veut les utiliser pour lui donner la publicité la plus ciblée possible.

→ Comment inclure les informations utilisateurs dont on dispose dans notre modèle?

→ **bandits linéaires**

RENDEMENTS

Sur une parcelle de culture, K engrais différents : lequel donne les meilleurs rendements?

Problème : trouver l'engrais qui maximise le rendement moyen n'est pas suffisant : on veut éviter des années catastrophiques.

→ Comment adapter nos stratégies au changement d'objectif?

PUBLICITÉ, RECOMMANDATION DE FILMS, ...

Chaque utilisateur est intéressé par différentes publicités!

En pratique si on a des informations sur l'utilisateur (we love cookies), on veut les utiliser pour lui donner la publicité la plus ciblée possible.

→ Comment inclure les informations utilisateurs dont on dispose dans notre modèle?

→ **bandits linéaires**

RENDEMENTS

Sur une parcelle de culture, K engrais différents : lequel donne les meilleurs rendements?

Problème : trouver l'engrais qui maximise le rendement moyen n'est pas suffisant : on veut éviter des années catastrophiques.

→ Comment adapter nos stratégies au changement d'objectif?

→ **quantile bandits**

De futurs challenges : l'exemple de la posologie

Nombre infini de bras

Cadre médical : un médicament à tester mais dont on ne connaît pas la posologie idéale : il y en a potentiellement une infinité à tester ...

→ On peut voir l'ensemble des bras comme un intervalle $[0, L]$ et on a une fonction moyenne $m : [0, L] \rightarrow \mathbb{R}$ dont on cherche le maximum.

De futurs challenges : l'exemple de la posologie

Nombre infini de bras

Cadre médical : un médicament à tester mais dont on ne connaît pas la posologie idéale : il y en a potentiellement une infinité à tester ...

→ On peut voir l'ensemble des bras comme un intervalle $[0, L]$ et on a une fonction moyenne $m : [0, L] \rightarrow \mathbb{R}$ dont on cherche le maximum.

Idée : utiliser plusieurs grilles et affiner la recherche au cours du temps.
→ nécessite des hypothèses sur f : croissance, continuité, ...

De futurs challenges : l'exemple de la posologie

Un double objectif

Mais on a un problème : si on teste une trop grosse posologie, on peut avoir des effets secondaires non négligeables (séquelles, baisse de l'espérance de vie, décès) ... On ne veut pas tuer la moitié du panel juste pour comprendre quelle dose est optimale!

→ contrainte sur la manière dont on peut observer les bras = choix limités!

Combinaison d'objectifs à surmonter pour étendre le champ d'application des bandits

De nombreuses ressources de vulgarisation sont disponibles en ligne.

Pour les plus matheux d'entre vous, un livre assez complet :
Bandit Algorithms, Tor Lattimore et Csaba Szepesvári, Cambridge
University Press.

DES QUESTIONS?