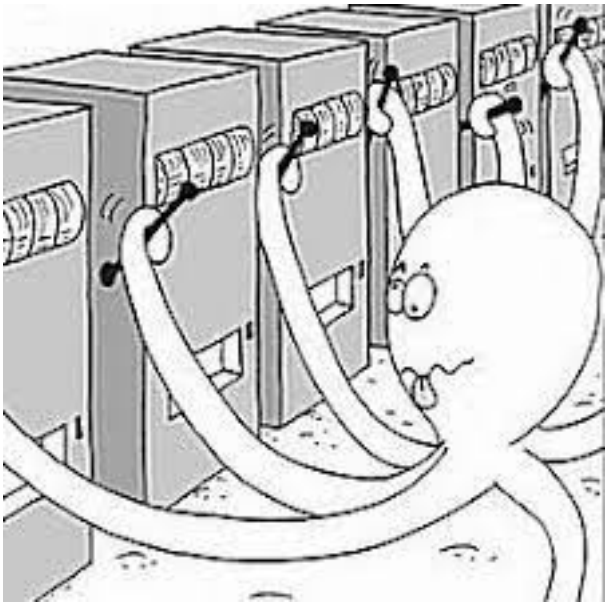




Ecole normale supérieure de Rennes

Stage de recherche M1

# Problème de bandit et clustering



*Auteur :*
Thuot Victor
ENS Rennes

*Encadrante :*
Alexandra Carpentier
Université Otto von Guericke
Magdebourg

2021

# Table des matières

## Introduction

On étudie les algorithmes de bandit depuis près d'un siècle. Le problème de bandit a été introduit par Thompson en 1933 [7]. Depuis quelques années, les questions liées à ce problème font l'objet de nombreuses publications. En outre, on utilise des algorithmes de type bandit dans le domaine médical, financier et pour la recommandation de contenu en ligne. Dans la totalité de ce rapport, nous considérons un joueur qui se retrouve face à plusieurs actions possibles (aussi appelés bras). Il peut jouer un nombre limité de fois. A chaque étape, le joueur choisit un bras et reçoit un gain qui est distribué selon une loi propre au bras. Le joueur ne sait pas de quelle manière sont répartis les gains.

Dans un premier temps, on va étudier les algorithmes de bandit qui cherchent à maximiser le gain total du joueur. Il s'agit d'un dilemme "exploration - exploitation". Après avoir décrit le cadre du problème, on s'intéresse à deux algorithmes : ETC et UCB. La performance de ces algorithmes est mesurée via leur regret et on décrira des bornes inférieures pour ce regret de manière minimax.

Ensuite, on va adopter un point de vue très différent. On supposera que les bras sont répartis en $K$ groupes à l'intérieur desquels les bras ont même moyenne. On cherche à identifier ces groupes. Si on ne peut observer qu'une fois chaque bras, on se retrouve face à un cadre non supervisé : le clustering. Mais ici, on supposera que pour partitionner les bras, on a le droit d'observer plusieurs fois chaque bras ce qui en un certain sens, rapproche le problème d'un problème d'apprentissage supervisé. On cherchera des méthodes pour résoudre le problème de "Clustering séquentiel".

# 1 Problème de bandit statistique : cadre et algorithmes de base.

## 1.1 Description du problème

Commençons par étudier le problème du bandit statistique. Dans le problème du bandit statistique, on considère un ensemble de bras $\mathcal{A}$ et un ensemble de distributions de probabilité sur $\mathbb{R}$, $\nu = (P_a : a \in \mathcal{A})$ qui définissent ensemble l'environnement. Chaque bras $a$ délivre un certain gain selon la distribution $P_a$. On note $n$ le nombre de tours appelé horizon. Pour chaque étape $i = 1, \ldots, n$, le joueur choisit un bras $A_t \in \mathcal{A}$ et obtient le gain $X_t$ où $X_t \sim P_{A_t}$. Le choix du joueur à chaque étape peut dépendre de tout le passé du jeux et la procédure qu'il adopte sera notée $\pi$.

Formellement, on construit une suite de variable aléatoire $A_1, X_1, A_2, X_2, \ldots, A_n, X_n$ et cette séquence doit satisfaire deux propriétés :

1. La loi conditionnelle de $X_t$ sachant $A_1, X_1, \ldots, X_{t_1}, A_t$ est $P_{A_t}$.
2. La loi conditionnelle de $A_t$ sachant $A_1, X_1, \ldots, A_{t-1}, X_{t-1}$ est $\pi_t(\cdot | A_1, X_1, \ldots, A_{t-1}, X_{t-1})$ où la suite $\pi_1, \pi_2, \ldots$ est une suite de noyau de probabilité qui définissent la stra-

tégie du joueur.

Le point 1 capture le fait que le gain $X_t$ est obtenu selon la loi $P_{A_t}$ à l'étape $t$. Le point 2 assure que le joueur ne peut pas utiliser le futur pour choisir le prochain bras. Il est très important de préciser que le joueur ne connaît pas les distributions de gain dans chaque bras et encore moins l'espérance de gain pour chacun de ces bras.

**Objectif :** L'objectif dans cette première partie est de maximiser le gain total $S_n = \sum_{t=1}^{n} X_t$.

◇ Remarque. Le caractère aléatoire de $A_t$ a une double origine, d'abord, le choix de $A_t$ par le joueur dépend des étapes de jeux précédentes, en particulier des bras choisis précédemment et des gains obtenus. Ensuite, le joueur peut adopter une stratégie aléatoire pour effectuer ses choix.

◇ Remarque. On parle de bandit statistique lorsque, comme introduit ci-dessus, les gains de chaque bras sont obtenu selon une distribution aléatoire qui ne dépend que du bras.

Le joueur se retrouve face à deux objectifs opposés :

– Apprentissage : on cherche à identifier le bras qui a la plus grande espérance de gain, pour cela, il est nécessaire d'explorer chacun des bras suffisamment pour obtenir une bonne estimation du bras le plus rentable.

– Exploitation : on doit également utiliser les données apprises pour exploiter les bras les plus rentables.

Il est nécessaire de jongler entre ces deux objectifs en choisissant le juste équilibre entre les deux. Si l'on exploite trop vite, on s'expose à une forte variance et au risque d'écarter à tort le bras le plus performant. Inversement, si l'on passe trop de temps à apprendre, on ne jouera pas assez sur le bras optimal et le gain total en sera limité.

### 1.2 Regret

Pour mesurer la performance d'une stratégie, on utilisera le regret. Si l'environnement est $\nu = (\mathcal{P}_a : a \in \mathcal{A})$, on définit le gain espéré sur le bras a par $\mu_a(\nu) := \int_{\mathbb{R}} x \mathrm{d}\mathcal{P}_a(x)$. On définit le gain optimal espéré, $\mu^*(\nu) := \max_{a\in\mathcal{A}} \mu_a(\nu)$. On suppose que toutes ces espérances sont finies et que le gain optimal espéré est atteint.

Définition 1. **Regret**
Le regret d'une procédure $\pi$ pour une instance de bandit $\nu$ est la perte de gain

entre cette stratégie et le gain optimale possible :

$$\mathcal{R}_n(\pi,\nu) = n\mu^*(\nu) - \mathbb{E}[\sum_{t=1}^{n} X_t].$$

Le regret permet de quantifier la qualité d'une procédure, minimiser le regret revient à maximiser l'espérance du gain total.

LEMME 2. Le regret est positif. Il peut être nul et il n'est nul que pour une stratégie qui choisit un des bras optimaux à chaque étape.

*Preuve* On réécrit le regret de la manière suivante : $\mathcal{R}_n = \sum_{t=1}^{n}(\mu^* - \mu_{A_t})$. Chaque terme est positif par définition de $\mu^*$. Si la somme est nulle, on a $\mu_{A_t} = \mu^*$ pour chaque $t$. □

DÉFINITION 3. Dans l'environnement $\nu = (P_a : a \in \mathcal{A})$, pour tout $a \in \mathcal{A}$, on va utiliser quelques notations.

1. $\mu^*(\nu) = \max_{a\in A} \mu_a(\nu)$ est le gain moyen optimal.
2. $\Delta_a(\nu) = \mu^*(\nu) - \mu_a(\nu)$ est la perte de gain espérée entre le bras $a$ et un bras optimal.
3. $T_a(t) = \sum_{s=1}^{t} \mathbb{1}\{A_s = a\}$ est la variable aléatoire donnant le nombre de fois où le bras $a$ est choisit avant le temps $t$
4. $\hat{\mu}_a(t) = \sum_{s=1}^{t} \frac{\mathbb{1}\{A_s = a\}X_s}{T_a(t)}$ la moyenne empirique des observations du bras $a$ avant le temps $t$.

THÉORÈME 4. **Décomposition du regret**
Si $\mathcal{A}$ est fini ou dénombrable et que l'horizon $n$ est fini :

$$\mathcal{R}_n(\pi,\nu) = \sum_{a\in\mathcal{A}} \Delta_a(\nu)\mathbb{E}[T_a(n)]$$

*Preuve*

$$\begin{aligned}
\mathcal{R}_n &= n\mu^* - \mathbb{E}[\sum_{t=1}^{n}\sum_{a\in\mathcal{A}} \mathbb{1}\{A_t = a\}X_t] \\
&= \mathbb{E}[\sum_{t=1}^{n}\sum_{a\in\mathcal{A}} \mathbb{1}\{A_t = a\}(\mu^* - X_t)] \\
&= \mathbb{E}\left[\mathbb{E}[\sum_{t=1}^{n}\sum_{a\in\mathcal{A}} \mathbb{1}\{A_t = a\}(\mu^* - X_t)|A_t]\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{a\in\mathcal{A}} \mathbb{1}\{A_t = a\}\mathbb{E}[\mu^* - X_t|A_t = a]\right] \\
&= \mathbb{E}[\sum_{t=1}^{n}\sum_{a\in\mathcal{A}} \mathbb{1}\{A_t = a\}\Delta_a] \\
&= \sum_{a\in\mathcal{A}} \Delta_a\mathbb{E}[T_a(n)] \text{ (Fubini-Tonelli)}
\end{aligned}$$

A la première ligne, on utilise $\sum_{a\in\mathcal{A}} \mathbb{1}\{A_t = a\} = 1$ puis on considère l'espérance comme l'espérance de l'espérance conditionnelle. A l'avant-dernière ligne, on remarque que $\mathbb{E}[\mu^* - X_t | A_t = a] = \Delta_a$. On conclut par Fubini-Tonelli. □

◇ REMARQUE. Le théorème de décomposition du regret, bien que très simple est très utile. En particulier, il permet d'obtenir des estimations et inégalités sur le regret à partir d'estimations des variables $T_a$.

### 1.3 Espace probabilisé sous-jacent

Dans la plupart des cas, l'espace de probabilité dans lequel se trouvent les variables $A_t$ et $X_t$ n'est pas précisé mais il est important de vérifier qu'un tel espace existe et que toutes les espérances soient bien définies. On propose ici un modèle canonique pour le bandit à $k$ bras et d'horizon $n$. L'environnement et la stratégie du joueur interagissent pour produire le tuple de variables aléatoires $H_n := (A_1,X_1,\dots,A_n,X_n)$. Pour tout $t$, on considère l'espace $\Omega_t := ([|1;k|]\times\mathbb{R})^t$ muni de la tribu $\mathcal{F}_t := \mathcal{B}(\Omega_t)$. La variable $X_i$ est alors les fonction mesurable de $(\Omega_n,\mathcal{F}_n)$ dans $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ qui à $(a_1,x_1,\dots,a_n,x_n)$ associe $x_i$. De même, $A_i := (a_1,x_1,\dots,a_n,x_n) \mapsto a_i$.

DÉFINITION 5. Une stratégie $\pi$ est une séquence $(\pi_t)_{t=1}^n$ où $\pi_t$ est un noyau de proba de $(\Omega_{t-1},\mathcal{F}_{t-1})$ dans $([|k|],2^k)$ muni de la tribu naturelle.

On considère $\nu = (P_i)_{i=1}^k$ des mesures de probabilité sur $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ et $\lambda$ une mesure $\sigma$-fini telle que $P_i \ll \lambda$ $(\forall i)$ (on note $p_i = dP_i/d\lambda$). On note aussi $\rho$ la mesure de comptage sur $[|1;n|]$.

THÉORÈME 6. La probabilité $\mathbb{P}_{\nu,\pi}$ est définit par la densité $p_{\nu,\pi} : \Omega_n \mapsto \mathbb{R}$ par rapport à $(\rho\times\lambda)^n$ :

$$p_{\nu,\pi}(a_1,x_1,\dots,a_n,x_n) = \prod_{t=1}^n \pi(a_t|a_1,x_1,\dots,a_{t-1},x_{t-1})p_{a_t}(x_t)$$

*Preuve* La preuve se fait par récurrence sur $n$ et en utilisant Fubini-Tonelli. On vérifie que dans cet espace, les variable $A_t$ et $X_t$ se comportent comme exigé lors de la description du problème. □

### 1.4 Cadre sous-gaussien

On énoncera un certain nombre de résultats dans le cas où les bras sont distribués selon des lois sous-gaussiennes. C'est un cadre commode et relativement général. Dans d'autres cadres, la principale différence sera l'inégalité de concentration choisie. Pire, on supposera souvent que les variables sont 1-sous-gaussienne pour s'économiser l'écriture du facteur $\sigma$ partout.

DÉFINITION 7. Une variable aléatoire est dite $\sigma$-sous-gaussienne si

$$\forall \lambda \in \mathbb{R} \ , \ \mathbb{E}[\exp(\lambda X)] \leqslant \exp\left(\frac{\sigma^2\lambda^2}{2}\right)$$

On utilisera très souvent cette inégalité de concentration, donnée sous forme d'un lemme.

LEMME 8. Si $X$ est $\sigma$-sous-gaussienne, alors pour tout $\varepsilon \geqslant 0$,

$$\mathbb{P}(X \geqslant \varepsilon) \leqslant \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right).$$

*Preuve* On utilise la méthode de Cramér-Chernoff. Soit $\lambda > 0$ :

$$\begin{aligned}\mathbb{P}(X \geqslant \varepsilon) = \mathbb{P}(\exp(\lambda X) \geqslant \exp(\lambda\varepsilon)) &\leqslant \frac{\mathbb{E}[\exp(\lambda X)]}{\exp(\lambda\varepsilon)} \text{ (Markov)} \\ &\leqslant \exp\left(\frac{\lambda^2\varepsilon^2}{2} - \lambda\varepsilon\right) \leqslant \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right)\end{aligned}$$

où la dernière ligne est obtenu en maximisant l'expression par rapport à $\lambda$. □

COROLLAIRE 9. Si $(X_1, \ldots, X_n)$ est un $n$-échantillon de variables aléatoires telles que $X_i - \mu$ soit $\sigma$-sous-gaussiennes. Alors pour $\varepsilon \geqslant 0$

$$\mathbb{P}(\hat{\mu} \geqslant \mu + \varepsilon) \leqslant \exp\left(-\frac{n\varepsilon^2}{2\sigma^2}\right) \text{ et } \mathbb{P}(\hat{\mu} \leqslant \mu - \varepsilon) \leqslant \exp\left(-\frac{n\varepsilon^2}{2\sigma^2}\right)$$

*Preuve* On montre que $\hat{\mu} - \mu$ est $\sigma/\sqrt{n}$-sous-gaussienne et on applique le théorème précédent. □

LEMME 10. **Hoeffding** Si $X$ est telle que $X \in [a,b]$ presque sûrement où $a < b$ alors $X$ est $(b-a)/2$-sous-gaussienne.

### 1.5 Algorithme ETC

On se place dans le cadre d'un bandit à $k$ bras, avec un horizon fini $n$ et des distributions de chaque bras 1-sous-gaussiennes.

#### (i) Principe

L'algorithme ETC : Explore then Commit, est caractérisé par un entier $m$. Lors de la première phase d'apprentissage, le joueur choisit chaque bras $m$ fois, puis il choisit toujours le bras ayant rapporté le plus lors de la phase d'apprentissage.

**Algorithme 1** ETC

**Entrée:** $m$

$\hat{\mu}_i = 0$ et $T_i = 0$ pour tout $i$

**pour** $t = 1, \dots, n$ **faire**

  **Choisir** le bras $A_t = \begin{cases} t \mod k & \text{si } t \leq mk \\ \underset{i=1,\dots,k}{\operatorname{argmax}}\, \hat{\mu}_i(mk) & \text{si } t > mk \end{cases}$

  **Obtenir** $X_t$ tel que $X_t \sim P_{A_t}$

  **Mettre à jour** $\begin{cases} T_{A_t} = T_{A_t} + 1 \\ \hat{\mu}_{A_t} = \frac{T_{A_t}-1}{T_{A_t}} \hat{\mu}_{A_t} + \frac{X_t}{T_{A_t}} \end{cases}$

**fin pour**

## (ii) Analyse de Regret

On se place dans le cas où $k = 2$ et où le bras 1 est optimal. On note $\Delta = \mu_2 - \mu_1$. Le théorème de décomposition assure que :

$$\mathcal{R}_n = \Delta \mathbb{E}[T_2(n)] = \Delta \mathbb{E}[T_2(n) \mathbb{1}_\xi] + \Delta \mathbb{E}[T_2(n) \mathbb{1}_{\xi^c}].$$

On cherche un ensemble $\xi$ tel que, sur $\xi$, on soit toujours dans le cas où $\hat{\mu}_1(mk) > \hat{\mu}_2(mk)$ et tel que $\xi$ soit très probable. L'ensemble $\xi = \{\hat{\mu}_1 \geqslant \mu_1 - \sqrt{\frac{2\log(2/\delta)}{m}}\} \cap \{\hat{\mu}_2 \leqslant \mu_2 + \sqrt{\frac{2\log(2/\delta)}{m}}\}$ est de probabilité au moins $1 - \delta$ dans le cadre 1-sous-gaussien. On souhaite que $\sqrt{\frac{2\log(2/\delta)}{m}} \leqslant \frac{\Delta}{2}$. En effet, si tel est le cas, on choisit toujours le bras otpimal sur $\xi$ et on a $\mathcal{R}_n \leqslant \Delta m + \Delta n \delta$. Si on choisit $\delta = 1/n$ et $m = \left\lceil \frac{8\log(2n)}{\Delta^2} \right\rceil$ alors on obtient :

$$\mathcal{R}_n \leqslant n\Delta \wedge (\frac{8}{\Delta} \log(2n) + \Delta + 1).$$

Enfin, si on majore par rapport à $\Delta$, on obtient que $\mathcal{R}_n = \mathcal{O}(\sqrt{n \log(n)})$.

Le gros problème que l'on souligne ici est que le choix de $m$ s'est fait en connaissance de $\Delta$ ce qui dans la plupart des cas est impossible. On généralise au cas où $k \neq 2$.

THÉORÈME 11. Si $mk \leqslant n$ et que les bras sont 1-sous-gaussien, le regret de *ETC* est majoré par :

$$\mathcal{R}_n \leqslant m \sum_{i=1}^{k} \Delta_i + (n - mk) \sum_{i=1}^{k} \Delta_i \exp\left(-\frac{m\Delta_i^2}{4}\right)$$

*Preuve* Par le théorème 4 de décomposition, il suffit de majorer $\mathbb{E}[T_i(n)]$. On suppose sans perte de généralité que le bras 1 est optimal. Le choix du bras $A_{mk+1}$ est important, il est choisit uniformément parmi les meilleurs bras au temps $mk$ et ensuite on s'y tient. $\mathbb{E}[T_i(n)] = m + (n - mk)\mathbb{P}(A_{mk+1} = i)$. Maintenant, calculons la probabilité d'estimer que le bras $i$ soit optimal pour $i \neq 1$.

$$\mathbb{P}(A_{mk+1} = i) \leqslant \mathbb{P}(\hat{\mu}_i(mk) \geqslant \max_{j \neq i} \hat{\mu}_j(mk)) \leqslant \mathbb{P}(\hat{\mu}_i(mk) \geqslant \hat{\mu}_1(mk))$$

Mais on a $\mathbb{P}(\hat{\mu}_i(mk) \geqslant \hat{\mu}_1(mk)) = \mathbb{P}(\hat{\mu}_i(mk) - \mu_1 - (\hat{\mu}_1(mk) - \mu_1) \geqslant \Delta_i)$ et $\hat{\mu}_i(mk) - \mu_1 - (\hat{\mu}_1(mk) - \mu_1)$ est $\sqrt{2/m}$-sous-gaussienne, ce qui conclue via le lemme de concentration 8. □

◇ Remarque. Cette borne illustre l'équilibre exploration-exploitation. Si on choisit $m$ trop grand, le premier terme croit et devient linéaire. Inversement, si $m$ est trop petit, la probabilité de se tromper de bras optimal après la phase d'apprentissage grandit. Le juste choix de $m$ permet d'obtenir une performance optimal (au vue du paragraphe dédié aux bornes de regret) mais ce choix est dans la plupart des situation impossible car dépendant de paramètres inconnus au joueur.

### 1.6 Algorithme UCB

## (i) Principe d'optimisme

L'algorithme UCB repose sur le principe d'optimisme qui consiste à supposer que l'environnement va agir de la meilleur façon plausible possible. A chaque étape, on se donne une borne supérieur pour chaque bras qui surestime très probablement sa moyenne. Cette borne se rapproche de la vraie valeur de la moyenne au fur et à mesure où l'on joue le bras en question. Si un bras présente une moyenne empirique plus faible que d'autre mais qu'il est toujours plausible que ce bras ait en fait une moyenne optimale, il sera joué. On se donne un degré de confiance $\delta$. D'abord, on définit cette borne supérieure qu'on appelle index, pour tout $i$,

$$\mathrm{UCB}_i(t-1,\delta) = \begin{cases} +\infty & \text{si } T_i(t-1) = 0 \\ \hat{\mu}_i(t-1) + \sqrt{\frac{2\log(1/\delta)}{T_i(t-1)}} & \text{sinon} \end{cases}$$

A chaque étape, on choisit le bras qui maximise cette index.

**Algorithme 2** UCB

**Entrée:** $\delta$,$k$

$UCB_i = +\infty$, $T_i = 0$ et $\hat{\mu}_i = 0$ pour tout $i$
**pour** $t = 1, \ldots, n$ **faire**
  **Choisir** le bras $A_t \in \underset{i=1,\ldots,k}{\operatorname{argmax}}\, UCB_i$
  **Obtenir** $X_t$ tel que $X_t \sim P_{A_t}$
  **Mettre à jour**
$$\begin{cases} T_{A_t} = T_{A_t} + 1 \\ \hat{\mu}_{A_t} = \frac{T_{A_t}-1}{T_{A_t}}\hat{\mu}_{A_t} + \frac{X_t}{T_{A_t}} \\ UCB_{A_t} = \hat{\mu}_{A_t} + \sqrt{\frac{2\log(1/\delta)}{T_{A_t}}} \end{cases}$$
**fin pour**

## (ii) Analyse de Regret

THÉORÈME 12. Si $\nu = (P_1, \dots, P_k)$ est un bandit à $k$ bras où $P_i$ est 1-sous-gaussien ($\forall i$), que l'horizon est $n$ et le degré de confiance est $\delta = 1/n^2$, alors :

$$\mathcal{R}_n \leqslant 3\sum_{i=1}^{k} \Delta_i + \sum_{i:\Delta_i>0} \frac{16\log(n)}{\Delta_i}$$

◇ REMARQUE. Dans la démonstration ci-dessous, on utilise un modèle commode pour construire les variables aléatoire $(A_t, X_t)_t$. On considère une collection $(Y_{t,i})_{t,i}$ de variables aléatoires indépendantes telles que pour tout $t = 1, \dots, n$ et $j = 1, \dots, k$, $Y_{t,i} \sim P_i$. On aura alors $X_t = Y_{t,A_t}$ et on note $\hat{\mu}_{i,u} = \sum_{s=1}^{u} \frac{Y_{s,i}}{u}$ la moyenne empirique des $u$ premières observations du bras $i$. Cette construction alternative au modèle canonique permet d'obtenir les mêmes lois pour les variables $(A_t, X_t)$.

*Preuve* Sans perte de généralité, on suppose que le premier bras est optimal, $\mu_1 = \mu^*$. On considère un bras non optimal $i$ et on cherche à estimer $T_i(n)$. Pour cela, on cherche un évènement $\xi_i$ de forte probabilité sur lequel on sait estimer $T_i$. Posons $\xi_i = \{\mu_1 < \min_{t\in[n]} \mathrm{UCB}_1(t,\delta)\} \cap \{\hat{\mu}_{i,u_i} + \sqrt{\frac{2}{u_i}\log(\frac{1}{\delta})} < \mu_1\}$. C'est l'ensemble sur lequel la moyenne $\mu_1$ n'est jamais sous-estimé par $\mathrm{UCB}_1$ et sur lequel la borne $\mathrm{UCB}_i$ après $u_i$ observations est inférieure à $\mu_1$. L'entier $u_i$ est à choisir plus tard.

*Étape 1 :* Sur $\xi_i$, on a $T_i(n) \leqslant u_i$.
Supposons que $T_i(n) > u_i$ sur $\xi_i$, à un certain temps $t$, on se retrouve dans la configuration où $A_t = i$ et $T_i(t-1) = u_i$ et par définition de l'argmax, on aurait alors $\mathrm{UCB}_i(t-1) \geqslant \mathrm{UCB}_1(t-1)$. Mais, sur $\xi_i$, on a $\mathrm{UCB}_1(t-1) > \mu_1 > \hat{\mu}_{i,u_i} + \sqrt{\frac{2}{u_i}\log(\frac{1}{\delta})} = \hat{\mu}_{i,T_i(t-1)} + \sqrt{\frac{2}{T_i(t-1)}\log(\frac{1}{\delta})} = \mathrm{UCB}_i(t-1)$ ce qui est absurde. Donc $T_i(n) \leqslant u_i$ sur $\xi_i$.

*Étape 2 :* Majorer le premier terme de $\mathbb{P}(G_i^c)$.

$$\begin{aligned}\{\mu_1 \geqslant \min_{t=1}^{n} \mathrm{UCB}_1(t,\delta)\} &\subset \left\{\mu_1 \geqslant \min_{t=1}^{n} \hat{\mu}_{1,t} + \sqrt{\frac{2\log(1/\delta)}{t}}\right\} \\ &= \bigcup_{t=1}^{n} \left\{\mu_1 \geqslant \hat{\mu}_{1,t} + \sqrt{\frac{2\log(1/\delta)}{t}}\right\}\end{aligned}$$

Maintenant, on utilise l'inégalité de concentration sur les variable 1-sous-gaussiennes, à savoir $\mathbb{P}(\mu_1 \geqslant \hat{\mu}_{1,s} + \sqrt{\frac{2\log(1/\delta)}{s}}) \leqslant \delta$. Et ainsi on a $\mathbb{P}(\mu_1 \geqslant \min_{t=1}^{n} \mathrm{UCB}_1(t,\delta)) \leqslant n\delta$.

*Étape 3 :* On minore la probabilité du deuxième ensemble dans $\xi_i^c$. On choisit $u_i$ assez large pour que $\Delta_i - \sqrt{\frac{\log(1/\delta)}{u_i}} \geqslant c\Delta_i$ pour un certain $c \in (0,1)$ et on

utilise encore le lemme de concentration 8.

$$\mathbb{P}(\hat{\mu}_{i,u_i} + \sqrt{\frac{2\log(1/\delta)}{u_i}} \geqslant \mu_i) = \mathbb{P}(\hat{\mu}_{i,u_i} - \mu_i \geqslant \Delta_i - \sqrt{\frac{2\log(1/\delta)}{u_i}})$$
$$\leqslant \mathbb{P}(\hat{\mu}_{i,u_i} - \mu_i \geqslant c\Delta_i) \leqslant \exp\left(-\frac{u_i c^2 \Delta_i^2}{2}\right)$$

*Étape 4 :* Il reste à choisir $u_i$ tel que $u_i \in [n]$ et $\Delta_i - \sqrt{\frac{2\log(1/\delta)}{u_i}} \geqslant c\Delta_i$. On prend le plus petit entier avec cette propriété soit $u_i = \left\lceil \frac{2\log(1/\delta)}{(1-c)^2\Delta_i^2} \right\rceil$. Maintenant :

$$\mathbb{E}[T_i(n)] = \mathbb{E}[T_i(n)\mathbb{1}_{\xi_i}] + \mathbb{E}[T_i(n)\mathbb{1}_{\xi_i^c}] \leqslant u_i + n\mathbb{P}(\xi_i^c) \leqslant u_i + n\left(n\delta + \exp(-\frac{u_i c^2\Delta_i^2}{2})\right).$$

On a pris $\delta = 1/n^2$ et on a $u_i \geqslant \frac{2\log(1/\delta)}{(1-c)^2\Delta_i^2}$ si bien que $\exp(-\frac{u_i c^2\Delta_i^2}{2}) \leqslant n^{-2c^2/(1-c)^2}$ et $\mathbb{E}[T_i(n)] \leqslant u_i + 1 + n^{1-2c^2/(1-c)^2}$. Si $c$ est choisit trop proche de 0, $u_i$ explose, de plus, si $c$ est trop proche de 1, le deuxième terme est linéaire en $n$. On prend $c = 1/2$ et alors :

$$\mathbb{E}[T_i(n)] \leqslant 3 + \frac{16\log(n)}{\Delta_i^2}$$

Le théorème 4 de décomposition du regret conclue. □

THÉORÈME 13. Sous les même hypothèses que le théorème précédent, on a :

$$\mathcal{R}_n \leqslant 8\sqrt{nk\log(n)} + 3\sum_{i=1}^{k}\Delta_i.$$

*Preuve* On prend $\Delta > 0$ une valeur seuil à identifier plus tard. Le théorème de décomposition du regret et la preuve du théorème précédent assure que :

$$\mathcal{R}_n = \sum_{i=1}^{k}\Delta_i\mathbb{E}[T_i(n)] = \sum_{i:\Delta_i<\Delta}\Delta_i\mathbb{E}[T_i(n)] + \sum_{i:\Delta_i\geqslant\Delta}\Delta_i\mathbb{E}[T_i(n)]$$
$$\leqslant n\Delta + \sum_{i:\Delta_i\geqslant\Delta}\left(3\Delta_i + \frac{16\log(n)}{\Delta_i}\right)$$
$$\leqslant n\Delta + \frac{16\log(n)}{\Delta} + 3\sum_i\Delta_i$$

Le résultat est obtenu en prenant $\Delta = \sqrt{16k\log(n)/n}$. □

◇ REMARQUE. On montrera plus tard qu'aucune stratégie sur des bras sous-Gaussien ne peut avec un regret meilleur qu'un $\mathcal{O}(\sqrt{nk})$, ce qui prouve que UCB est presque optimal sur cette classe de bandit.

### 1.7 UCB sans connaissance de l'horizon

#### (i) algorithme

Dans la version précédente de l'algorithme UCB, on a besoin de la connaissance de l'horizon $n$ pour choisir le facteur $\delta$, ce qui n'est pas toujours possible.

En utilisant le même principe d'optimisme avec quelques adaptations, on peut palier à ce problème.

**Algorithme 3** UCB asymptotiquement optimal

**Entrée:** $k$

1: **Choisir** chaque bras une fois.
2: **pour** $t = 1,2,\dots$ **faire**
3: **Choisir** le bras

$$A_t \in \operatorname{argmax}_i \left( \hat{\mu}_i(t-1) + \sqrt{\frac{2\log f(t)}{T_i(t-1)}} \right)$$

$$\text{où } f(t) = 1 + t\log^2(t)$$

4: **fin pour**

## (ii) Regret

Commençons par un lemme :

LEMME 14. Soit $X_1, \dots, X_n$ une séquence de varibale aléatoires 1-sous-gaussiennes. Soient $\hat{\mu}_t := \frac{1}{t}\sum_{s=1}^{t} X_s$, $\varepsilon > 0$, $a > 0$,

$$\kappa = \sum_{t=1}^{n} \mathbb{1}\{\hat{\mu}_t + \sqrt{\frac{2a}{t}} \geqslant \varepsilon\} \text{ et } \kappa' = u + \sum_{t=\lceil u \rceil}^{n} \mathbb{1}\{\hat{\mu}_t + \sqrt{\frac{2a}{t}} \geqslant \varepsilon\}$$

où $u = 2a\varepsilon^{-2}$. Alors on a :

$$\mathbb{E}[\kappa] \leqslant \mathbb{E}[\kappa'] \leqslant 1 + \frac{2}{\varepsilon^2}(a + \sqrt{\pi a} + 1)$$

*Preuve* On utilise encore et toujours l'inégalité de type Hoeffding pour les variables aléatoires sous-gaussiennes. $\mathbb{E}[\kappa] \leqslant \mathbb{E}[\kappa'] \leqslant u + \sum_{t=\lceil u \rceil}^{n} \exp\left( -\frac{t(\varepsilon - \sqrt{\frac{2a}{t}})^2}{2} \right)$.
La fonction sommée est décroissante sur $[\lceil u \rceil; +\infty]$, on utilise une comparaison série intégrale et le changement de variable $s = \varepsilon\sqrt{t} - \sqrt{2a}$.

$$\begin{aligned}
\sum_{t=\lceil u \rceil}^{n} \exp\left( -\frac{t(\varepsilon - \sqrt{\frac{2a}{t}})^2}{2} \right) &\leqslant 1 + \int_{t=u}^{+\infty} \exp\left( -\frac{t(\varepsilon - \sqrt{\frac{2a}{t}})^2}{2} \right) \mathrm{d}t \\
&= 1 + \int_{s=0}^{+\infty} \exp\left(-\frac{s^2}{2}\right) \frac{2}{\varepsilon^2}(s + \sqrt{2a})\,\mathrm{d}s \\
&= 1 + \frac{2}{\varepsilon^2}(\sqrt{\pi a} + 1)
\end{aligned}$$

□

THÉORÈME 15. Pour un bandit 1-sous-gaussien et avec l'algo précédent :

$$\mathcal{R}_n \leqslant \sum_{i:\Delta_i>0} \inf_{\varepsilon\in(0,\Delta_i)} \Delta_i \left(1 + \frac{5}{\varepsilon^2} + \frac{2(\log f(n) + \sqrt{\pi \log f(n)} + 1}{(\Delta_i - \varepsilon)^2}\right).$$

*Preuve* On note encore $\mathrm{UCB}_i(t-1) = \hat{\mu}_i(t-1) + \sqrt{\frac{2\log f(t)}{T_i(t-1)}}$ On va utiliser le théorème de décomposition du regret. Soit $i$ l'indice d'un bras non optimal. Il reste à borner $T_i(n)$ en espérence. On remarque que $\{A_t = i\} = \{A_t = i \text{ et } \mathrm{UCB}_i(t-1) \geqslant \mu_1 - \varepsilon\} \cup \{A_t = i \text{ et } \mathrm{UCB}_i(t-1) \leq \mu_1 - \varepsilon\} \subset \{A_t = i \text{ et } \mathrm{UCB}_i(t-1) \geqslant \mu_1 - \varepsilon\} \cup \{\mathrm{UCB}_1(t-1) \leqslant \mu_1 - \varepsilon\}$. Maintenant

$$\begin{aligned}\mathbb{E}[T_i(n)] = \sum_{t=1}^{n} \mathbb{P}(A_t = i) &\leqslant \sum_{t=1}^{n} \mathbb{P}(\mathrm{UCB}_1(t-1) \leqslant \mu_1 - \varepsilon) \\ &\quad + \sum_{t=1}^{n} \mathbb{P}(A_t = i \text{ et } \mathrm{UCB}_i(t-1) \geqslant \mu_1 - \varepsilon)\end{aligned}$$

*Borne du premier terme :*

$$\begin{aligned}\sum_{t=1}^{n} \mathbb{P}(\mathrm{UCB}_1(t-1) \leqslant \mu_1 - \varepsilon) &\leqslant \sum_{t=1}^{n}\sum_{s=1}^{n} \mathbb{P}\left(\hat{\mu}_{1,s} + \sqrt{\frac{2\log f(t)}{s}} \leqslant \mu_1 - \varepsilon\right) \\ &\leqslant \sum_{t=1}^{n}\sum_{s=1}^{n} \exp\left(-\frac{s(\sqrt{\frac{2\log f(t)}{s}} + \varepsilon)^2}{2}\right) \\ &\leqslant \sum_{t=1}^{n} \frac{1}{f(t)} \sum_{s=1}^{n} \exp(-\frac{s\varepsilon^2}{2}) \leqslant \frac{5}{\varepsilon^2}\end{aligned}$$

*Borne du deuxième terme :* (en utilisant le lemme 14)

$$\begin{aligned}\sum_{t=1}^{n} \mathbb{P}(A_t = i \text{ et } \mathrm{UCB}_i(t-1) \geqslant \mu_1 - \varepsilon) &\leqslant \mathbb{E}[\sum_{s=1}^{n} \mathbb{1}\{\hat{\mu}_{i,s} - \mu_i + \sqrt{\frac{2\log f(n)}{s}} \geqslant \Delta_i - \varepsilon\}] \\ &\leqslant 1 + \frac{2}{(\Delta_i + \varepsilon)^2}(\log f(n) + \sqrt{\pi \log f(n)} + 1)\end{aligned}$$

□

COROLLAIRE 16. Il existe $C > 0$ tel que :

$$\mathcal{R}_n \leqslant C\sum_{i=1}^{k} \Delta_i + 2\sqrt{Cnk\log(n)}$$

*Preuve* Prenons $\varepsilon = \delta_i/2$ dans le théorème précédent, on obtient

$$\mathcal{R}_n \leqslant \sum_{i:\Delta_i>0} (\Delta_i + \frac{1}{\Delta_i}(8\log f(n) + 8\sqrt{\log f(n)} + 28)$$

puis on factorise par $\log(n)$. Enfin on opère comme dans la démonstration du théorème 13 sur UCB. □

## 2 Bandit statistique : bornes inférieures de regret

On cherche dans cette partie des inégalités permettant de vérifier si une procédure est optimale. On peut par exemple trouver une borne inférieure du regret minimax, ou encore une borne valable avec grande probabilité. La dépendance au choix d'instance est aussi étudiée.

### 2.1 Incursion dans la théorie de l'information

#### (i) Entropie relative

DÉFINITION 17. Si $P$ et $Q$ sont deux mesures sur $[N]$, l'**entropie relative** ou divergence de Kullback-Leibler est définie par :

$$\mathrm{D}(P,Q) = \sum_{i\in[N]:p_i>0} p_i \log(p_i/q_i)$$

Dans le cas général, soit $(\Omega,\mathcal{F})$ un espace mesurable et $P,Q$ deux mesures sur cet espace. Alors, l'entropie relative est définit :

$$D(P,Q) = \begin{cases} \int \log\left(\frac{\mathrm{d}P}{\mathrm{d}Q}(\omega)\right) \mathrm{d}P(\omega) & \text{si } P \ll Q \\ \infty & \text{sinon} \end{cases}$$

▷ EXEMPLE. Un calcul direct donne :

$$D(\mathcal{N}(\mu_1,\sigma^2),\mathcal{N}(\mu_2,\sigma^2)) = \frac{(\mu_1-\mu_2)}{2\sigma^2}$$

THÉORÈME 18. **Inégalité de Bretagnolle-Huber** Soit $P$ et $Q$ deux mesures de probabilités sur $(\Omega,\mathcal{F})$ et soit $A \in \mathcal{F}$ un évènement quelconque. Alors :

$$P(A) + Q(A^c) \geqslant \frac{1}{2}\exp(-D(P,Q)).$$

*Preuve* On suppose que $P \ll Q$, sinon, on aurait $D(P,Q) = \infty$ et l'inégalité est claire. On pose donc $\nu = P + Q$, mesure qui vérifie $P,Q \ll \nu$ et on note $p := \frac{\mathrm{d}P}{\mathrm{d}\nu}$ et $q := \frac{\mathrm{d}P}{\mathrm{d}\nu}$. On a $D(P,Q) = \int p \log(\frac{p}{q})\,\mathrm{d}\nu$.

*Étape 1* On remarque que $pq = (p\wedge q)(p\vee q)$. Alors par Cauchy-Schwarz,

$$\left(\int \sqrt{pq}\,\mathrm{d}\nu\right)^2 = \left(\int \sqrt{p\wedge q}\sqrt{p\vee q}\,\mathrm{d}\nu\right)^2 \overset{CS}{\leqslant} \left(\int p\wedge q\,\mathrm{d}\nu\right)\left(\int p\vee q\,\mathrm{d}\nu\right)$$

Maintenant $p\wedge q + p\vee q = p+q$ donc $\int p\vee q = 2 - \int p\wedge q \leqslant 2$.

*Étape 2*

$$\begin{aligned}\left(\int \sqrt{pq}\right)^2 &= \exp\left(2\log\int\sqrt{pq}\right) = \exp\left(2\int_{p>0} p\sqrt{\frac{q}{p}}\right) \\ &\overset{Jen.}{\geqslant} \exp\left(2\int_{p>0}\frac{1}{2}\log\left(\frac{q}{p}\right)\right) = \exp\left(-\int p\log\left(\frac{p}{q}\right)\right) \\ &= \exp(-D(P,Q))\end{aligned}$$

*Étape 3*

$$P(A) + Q(A^c) = \int_A p + \int_{A^c} q \geqslant \int_A p \wedge q + \int_{A^c} p \wedge q = \int p \wedge q$$
$$\geqslant \frac{1}{2}(\int \sqrt{pq})^2 \geqslant \frac{1}{2}\exp(-D(P,Q))$$

□

◇ REMARQUE. Cette inégalité est très intéressante pour donner une borne inférieure sur le regret d'un bandit et plus généralement en statistique. Par exemple, si on a une gaussienne réduite de moyenne $\mu$ et qu'on souhaite tester l'hypothèse $\mu = \mu_1$ contre $\mu = \mu_2$, l'inégalité permet de minorer la somme des risques de première et de deuxième espèce.

## (ii) Entropie relative entre bandits

THÉORÈME 19. **Décomposition de la divergence**
Si $\nu = (P_1, \ldots, P_k)$ et $\nu' = (P'_1, \ldots, P'_k)$ sont deux vecteurs de distributions de gains et que $\pi$ est une stratégie. On note $\mathbb{P}_\nu = \mathbb{P}_{\nu,\pi}$ la probabilité engendrée par l'environnement $\nu$ et la stratégie $\pi$ donnée en théorème 6. Alors,

$$D(\mathbb{P}_\nu, \mathbb{P}_{\nu'}) = \sum_{i=1}^{k} \mathbb{E}_\nu[T_i(n)] D(P_i, P'_i)$$

*Preuve* On rappelle que la proba $\mathbb{P}_\nu$ (resp $\mathbb{P}_{\nu'}$) admet une densité par rapport à la mesure $(\rho \times \lambda)^n$ sur l'espace canonique $\Omega_n$ qui est donnée par :

$$p_\nu(a_1, 1, \ldots, a_n, x_n) = \prod_{t=1}^{n} \pi_t(a_t | a_1, x_1, \ldots, a_{t-1}, x_{t-1}) p_{a_t}(x_t).$$

où on remplace $p_{a_t}$ par $p'_{a_t}$ pour $\mathbb{P}_{\nu'}$. Ainsi :

$$\log \frac{\mathrm{d}\mathbb{P}_\nu}{\mathrm{d}\mathbb{P}_{\nu'}}(a_1, \ldots, x_n) = \sum_{t=1}^{n} \log \frac{p_{a_t}(x_t)}{p'_{a_t}(x_t)}.$$

Maintenant :

$$\begin{aligned} D(\mathbb{P}_\nu, \mathbb{P}_{\nu'}) &= \mathbb{E}_\nu \left[ \log \left( \frac{\mathrm{d}\mathbb{P}_\nu}{\mathrm{d}\mathbb{P}_{\nu'}} \right) (A_1, X_1, \ldots, A_n, X_n) \right] \\ &= \sum_{t=1}^{n} \mathbb{E}_\nu \left[ \log \frac{p_{A_t}(x_t)}{p'_{A_t}(x_t)} \right] = \sum_{t=1}^{n} \mathbb{E}_\nu \left[ \mathbb{E}_\nu \left[ \log \frac{p_{A_t}(x_t)}{p'_{A_t}(x_t)} | A_t \right] \right] \\ &= \sum_{t=1}^{n} \mathbb{E}_\nu \left[ D(P_{A_t}, P'_{A_t}) \right] = \sum_{i=1}^{k} \mathbb{E}_\nu [\sum_{t=1}^{n} \mathbb{1}\{A_t = i\} D(P_{A_t}, P'_{A_t})] \\ &= \sum_{i=1}^{k} \mathbb{E}_\nu[T_i(n)] D(P_i, P'_i) \end{aligned}$$

□

## 2.2 Borne inférieure

DÉFINITION 20. Pour une procédure $\pi$ et un ensemble d'instance de bandit $\mathscr{E}$, le regret maximal est :

$$\mathcal{R}_n(\pi, \mathscr{E}) = \sup_{\nu \in \mathscr{E}} \mathcal{R}_n(\pi,\nu)$$

Si on note $\Pi$ l'ensemble des procédures possibles, on définit le regret minimax.

DÉFINITION 21. Le **regret minimax** est :

$$\mathcal{R}_n^*(\mathscr{E}) = \inf_{\pi \in \Pi} \mathcal{R}_n(\pi, \mathscr{E}) = \inf_{\pi \in \Pi} \sup_{\nu \in \mathscr{E}} \mathcal{R}_n(\pi,\nu).$$

On dit qu'une procédure est minimax optimale si elle permet d'atteindre cette borne.

THÉORÈME 22. **Borne inférieur du regret**
Soit $k > 1$ et $n \geqslant k-1$. Alors, pour toute procédure $\pi$, il existe $\mu \in \mathbb{R}^k$ tel que :

$$\mathcal{R}_n(\pi,\nu_\mu) \geqslant \frac{1}{27}\sqrt{((k-1)n}$$

où $\nu_\mu$ est l'instance de bandit telle que le $i$ème bras soit distribué selon $\mathcal{N}(\mu_i,1)$.

*Preuve* On considère une procédure $\pi$, un horizon $n$ et un nombre de bras $k$ tels que $n \geqslant k-1$. On prend un bandit Gaussien de la forme $\nu_\mu$ où $\mu = (\Delta,0,\dots,0)$ et $\Delta \in [0,1/2]$. On cherche à construire à partir de cet environnement un autre environnement pour mettre en faute la procédure dans au moins l'un de ces deux environnements. L'idée est de faire exploser le gain dans le bras pour lequel la procédure pour $\mu$ est en défaut de connaissance. On choisit $i \in \operatorname{argmin}_{j>1}\mathbb{E}_\mu[T_j(n)]$ et on construit $\mu' = (\Delta,0,\dots,0,\underbrace{2\Delta}_{i\text{ème}},0,\dots,0)$.

$$\begin{aligned}\mathcal{R}_n(\pi,\nu_\mu) = \sum_{j=1}^k \Delta_k \mathbb{E}_\mu[T_j(n)] = \Delta \sum_{j=2}^k \mathbb{E}_\mu[T_j(n)] &= \Delta(n - \mathbb{E}_\mu[T_1(n)]) \\ &\geqslant \Delta(n - \mathbb{E}_\mu[\mathbb{1}_{\{T_1(n) \leqslant \frac{n}{2}\}} T_1(n)]) \\ &\geqslant \frac{n\Delta}{2}\mathbb{P}_\mu(T_1(n) \leqslant n/2)\end{aligned}$$

$$\begin{aligned}\mathcal{R}_n(\pi,\nu_{\mu'}) = \sum_{j=1}^k \Delta_k \mathbb{E}_\mu[T_j(n)] &\geqslant \Delta \mathbb{E}_{\mu'}[T_1(n)] \\ &\geqslant \Delta \mathbb{E}_{\mu'}[\mathbb{1}_{\{T_1(n) > \frac{n}{2}\}} T_1(n)] \\ &\geqslant \frac{n\Delta}{2}\mathbb{P}_\mu(T_1(n) > n/2)\end{aligned}$$

Maintenant, on s'est placé dans les conditions idéales pour utiliser l'inégalité

de Bretagnolle-Huber,

$$\begin{aligned}\mathcal{R}_n(\pi,\nu_\mu)+\mathcal{R}_n(\pi,\nu_{\mu'}) &\geqslant \frac{n\Delta}{2}\Big(\mathbb{P}_\mu(T_1(n)>n/2)+\mathbb{P}_\mu(T_1(n)\leqslant n/2)\Big)\\ &\geqslant \frac{n\Delta}{4}\exp\big(-\mathcal{D}(\mathbb{P}_\mu,\mathbb{P}_{\mu'})\big)\end{aligned}.$$

Comme $i\in\operatorname{argmin}_{j>2}\mathbb{E}_\mu[T_j(n)]$ et $\sum_{j=2}^n\mathbb{E}[T_1(n)]=n$, on a $n\geqslant(k-1)\mathbb{E}[T_i(n)]$. Avec le lemme de décomposition de la divergence, on a :

$$\begin{aligned}\mathrm{D}(\mathbb{P}_\nu,\mathbb{P}_{\nu'})=\sum_{i=1}^k\mathbb{E}_\nu[T_j(n)]\,\mathrm{D}(\mathcal{P}_i,\mathcal{P}_i') &= \mathbb{E}_\nu[T_i(n)]\,\mathrm{D}(\mathcal{N}(0,1),\mathcal{N}(2\Delta,1))\\ &= \mathbb{E}_\nu[T_i(n)]\frac{(2\Delta)^2}{2}\\ &\leqslant \frac{2\Delta^2 n}{k-1}\end{aligned}$$

Enfin, en utilisant $\max(a,b)\geqslant(a+b)/2$, et en choisissant $\Delta=\frac{1}{2}\sqrt{\frac{k-1}{n}}\leqslant 1/2$ on conclue :

$$\begin{aligned}\max(\mathcal{R}_n(\pi,\nu_\mu),\mathcal{R}_n(\pi,\nu_{\mu'})) &\geqslant \frac{1}{2}(\mathcal{R}_n(\pi,\nu_\mu)+\mathcal{R}_n(\pi,\nu_{\mu'}))\\ &\geqslant \frac{n\Delta}{8}\exp\left(-\mathrm{D}(\mathbb{P}_\mu,\mathbb{P}_{\mu'})\right)\\ &\geqslant \frac{n\Delta}{8}\exp\left(-\frac{2\Delta^2 n}{k-1}\right)\\ &= \frac{\sqrt{n(k-1)}}{16\exp(1/2)}\geqslant\frac{1}{27}\sqrt{((k-1)n}\end{aligned}$$

□

### 2.3 Borne inférieure dépendant de l'instance

Dans la section précédente, on a énoncé une borne inférieure du regret pour des bras distribuée par des mesures 1-sous-gaussiennes. On a prouvé qu'on pouvait toujours borner inférieurement le regret en $\sqrt{nk}$ via des bras distribués avec des gaussiennes. On va ici donner une borne inférieure qui dépend des instances de bandit choisies.

DÉFINITION 23. Une classe de bandit est ensemble $\mathscr{E}$ de distribution de gains pour les bras. On dit qu'elle est non structurée si la classe est de la forme $\mathscr{E}=\mathcal{M}_1,\mathcal{M}_2,\ldots,\mathcal{M}_k$ où $\mathcal{M}_i$ est un ensemble de distributions sur $\mathbb{R}$.

On va exiger pour une stratégie la propriété "raisonnable" suivante.

DÉFINITION 24. Une stratégie $\pi$ est appelée **consistante** sur une classe de

bandits $\mathcal{E}$ si pour tout $\nu \in \mathscr{E}$ et pour tout $p > 0$,

$$\lim_{n\to+\infty} \frac{\mathcal{R}_n(\pi,\nu)}{n^p} = 0.$$

▷ Exemple. La stratégie "Toujours choisir le premier bras" n'est consistante sur $\mathscr{E}$ que si le premier bras est optimal $\forall \nu \in \mathscr{E}$. L'algorithme UCB est consistant sur l'ensemble des distributions 1-sous-gaussiennes car alors $\frac{\mathcal{R}_n(\pi,\nu)}{n^p} \leqslant 3\sum^k \frac{\Delta_i}{n^p} + \sum_{i=1}^k \frac{\log(n)}{n^p} \to 0$.

Définition 25. Soit $\mathcal{M}$ un ensemble de distributions dans $L_1(\mathbb{R})$ et soit $\mu : \mathcal{M} \mapsto \mathbb{R}$ la fonction qui donne la moyenne d'une distribution. Soit $\mu^*$ et $P \in \mathcal{M}$ tel que $\mu(P) < \mu^*$. On définit alors :

$$d_{inf}(P,\mu^*,\mathcal{M}) = \inf_{P'\in\mathcal{M}}\{D(P,P') : \mu(P') > \mu^*\}$$

Théorème 26. Soit $\mathscr{E} = \mathcal{M}_1,\mathcal{M}_2,\ldots,\mathcal{M}_k$ et $\pi$ une stratégie consistante. Alors pour tout $\nu = (P_i)_{i=1}^k \in \mathscr{E}$, on a :

$$\liminf_{n\mapsto\infty} \frac{\mathcal{R}_n}{\log(n)} \geqslant c^*(\nu,\mathscr{E}) = \sum_{i:\Delta_i>0} \frac{\Delta_i}{d_{inf}(P_i,\mu^*,\mathcal{M}_\rangle)}$$

où $\mu^*$ est la moyenne du bras optimal et $\Delta_i$ est l'écart du bras $i$ à cette moyenne.

*Preuve* On va utiliser des outils similaires au théorème de la section précédente. Soit un bras sous-optimal $i$ de moyenne $\mu_i$ et on note $d_i := d_{inf}(P_i,\mu^*,\mathcal{M}_\rangle)$. On cherche à montrer que

$$\liminf_{n\to+\infty} \frac{\mathbb{E}_{\nu,\pi}[T_i(n)]}{\log(n)} \geqslant \frac{1}{d_i}$$

C'est clair si $d_i = \infty$, sinon, par définition de l'inf, il existe $P_i' \in \mathcal{M}_i$ tel que $\mu(P') > \mu^*$ et $D(P_i,P_i') \leqslant d_i + \varepsilon$ avec $> 0$. On construit $\nu'$ vecteur qui vaut $P_j$ si $j \neq i$ et vaut $P_i'$ en $i$.

D'après la décomposition de la divergence, on a :

$$D(\mathbb{P}_{\nu,\pi},\mathbb{P}_{\nu,\pi}) = \sum_{j=1}^n \mathbb{E}_{\nu,\pi}[T_j(n)]D(P_j,P_j') = \mathbb{E}_{\nu,\pi}[T_i(n)]D(P_i,P_i') \leqslant \mathbb{E}_{\nu,\pi}[T_i(n)](d_i+\varepsilon).$$

Ensuite, l'inégalité de Bretagnolle-Huber assure que pour tout évènement $A$,

$$\mathbb{P}_{\nu,\pi}(A) + \mathbb{P}_{\nu',\pi}(A^c) \geqslant \frac{1}{2}\exp(-D(\mathbb{P}_{\nu,\pi},\mathbb{P}_{\nu',\pi})) \geqslant \frac{1}{2}\exp(-\mathbb{E}_{\nu,\pi}[T_i(n)](d_i+\varepsilon))$$

Maintenant on choisit $A = \{T_i(n) > \frac{n}{2}\}$,

$$\begin{aligned}
\mathcal{R}_n + \mathcal{R}_n' &\geqslant \frac{n}{2}(\mathbb{P}_{\nu,\pi}(A)\Delta_i + \mathbb{P}_{\nu',\pi}(A^c)(\mu_i' - \mu^*)) \\
&\geqslant \frac{n}{2}\min\{\Delta_i,\mu_i'-\mu^*\}(\mathbb{P}_{\nu,\pi}(A) + \mathbb{P}_{\nu',\pi}(A^c)) \qquad . \\
&\geqslant \frac{n}{4}\min\{\Delta_i,\mu_i'-\mu^*\}\exp(-\mathbb{E}_{\nu,\pi}[T_i(n)](d_i+\varepsilon))
\end{aligned}$$

On réarrange et on prend la lim inf,

$$\begin{aligned}\liminf_{n\to\infty}\frac{\mathbb{E}_{\nu,\pi}[T_i(n)]}{\log(n)} &\geqslant \frac{1}{d_i+\varepsilon}\liminf_{n\to\infty}\frac{\log\left(\frac{n\min\{\Delta_i,\mu' i-\mu^*\}}{4(\mathcal{R}_n+\mathcal{R}'_n)}\right)}{\log(n)}\\ &= \frac{1}{d_i+\varepsilon}\left(1-\limsup_{n\to+\inf}\frac{\log(\mathcal{R}_n+\mathcal{R}'_n)}{\log(n)}\right)\geqslant\frac{1-p}{1+\varepsilon}\end{aligned}$$

Comme $\pi$ est consistante, on a pour $n$ assez grand $\mathcal{R}_n+\mathcal{R}'_n n^p$, d'où $\frac{\log(\mathcal{R}_n+\mathcal{R}'_n)}{\log(n)}\leqslant p$. Ensuite, on a $\varepsilon>0$ et $p>0$ quelconque ce qui conclue avec le théorème de décomposition du regret. □

### 2.4 Borne inférieure de forte probabilité

Pour parler de borne très probable, on n'utilise pas le regret espéré qui n'est pas aléatoire. On introduit le pseudo-regret

$$\overline{\mathcal{R}}_n=\sum_{i=1}^{n}T_i(n)\Delta_i.$$

On note $\mathscr{E}^k$ l'ensemble des bandits à $k$-bras, Gaussien et avec des écarts $\Delta_i\in[0;1]$.

THÉORÈME 27. Soit $n\geqslant 1$, $k\geqslant 2$, $B>0$ et $\pi$ une stratégie telle que pour tout $\nu\in\mathscr{E}^k$, $\mathcal{R}_n(\pi,\nu)\leqslant B\sqrt{(k-1)n}$. Soit $\delta\in(0,1)$. Alors il existe un bandit $\nu\in\mathscr{E}^k$ tel que :

$$\mathbb{P}\left(\overline{\mathcal{R}}_n(\pi,\nu)\geqslant\frac{1}{4}\min\left\{n,\frac{1}{B}\sqrt{(k-1)n}\log(\frac{1}{4\delta})\right\}\right)\geqslant\delta.$$

*Preuve* Soit $\Delta\in(0;1/2]$ à choisir. On construit deux bandits. D'abord $\nu=\nu_\mu$ avec $\mu_1=\Delta$ et $\mu_j=0$ pour $j>1$. On pose $i=\text{armin}_{j>1}\mathbb{E}[T_j(n)]$ et on construit le bandit $\nu'$ identique à $\nu$ sauf en $i$ où $\mu'_i=2\Delta$. Le théorème de décomposition du regret assure que $\mathbb{E}[T_i(n)]\leqslant\frac{\mathcal{R}_n}{\Delta(k-1)}\leqslant\frac{B}{\Delta}\sqrt{\frac{n}{k-1}}$. Ensuite, on utilise tous les outils introduit précédemment, Bretagnolle-Huber, le théroème de décomposition de la divergence.

$$\begin{aligned}\mathbb{P}(\overline{\mathcal{R}}_n\geqslant\frac{\Delta n}{2})+\mathbb{P}'(\overline{\mathcal{R}}'_n\geqslant\frac{\Delta n}{2}) &\geqslant \mathbb{P}(T_i(n)\geqslant\frac{n}{2})+\mathbb{P}'(T_i(n)\leq\frac{n}{2})\\ &\geqslant\frac{1}{2}\exp(-D(\mathbb{P},\mathbb{P}'))=\frac{1}{2}\exp(-\mathbb{E}[T_i(n)]D(\mathscr{P}_i,\mathscr{P}'_i))\\ &\geqslant\frac{1}{2}\exp(-\frac{B}{\Delta}\sqrt{\frac{n}{k-1}}\frac{(2\Delta)^2}{2})\geqslant 2\delta\end{aligned}$$

La dernière inégalité est obtenue en choisissant $\Delta=\min\left\{\frac{1}{2},\frac{1}{2B}\sqrt{\frac{k-1}{n}}\log(\frac{1}{4\delta})\right\}$. □

◇ REMARQUE. Pour clore cette partie, on peut voir que la stratégie UCB est presque optimale (à un facteur $\log(n)$ près. Des variantes de cet algorithme

permettent d'éliminer ce facteur. Pour ce qui est d'ETC, la performance est optimale si on connaît le $m$ optimal.

**Introduction au clustering**

Dans tout ce qui suit, on va s'intéresser à un autre autre objectif face à un bandit : partitionner les bras en groupes de même moyenne. On va considérer un bandit à $n$ bras tel que le gain du bras $i$ soit distribué selon une gaussienne $\mathcal{N}(\mu_i,\sigma^2 I_d)$. Les bras donnent un gain de dimension $d$ selon des gaussienne de même variance et peuvent être regroupés en $K$ groupes de même moyennes.
On cherche ainsi $G^* = \{G_1^*,\dots,G_M^*\}$ partition de $\{1,\dots,n\}$ en $K$ groupes tel que, si $i \in G_k^*$ alors le bras $i$ est distribué selon $\mathcal{N}(\mu_k,\sigma^2 I_d)$. On note $T$ le nb d'observations possibles. Pour $t = 1,\dots,T$, on choisit $A_t \in [|1;n|]$, on observe $X_t$ distribué selon $P_{A_t}$. L'objectif est de retrouver à partir de $X_1,\dots,X_T$ la partition $G^*$.

1. **Non supervisé :** On observe une fois chaque individu/bras et on infère à partir de ces observations ($T = n$).
2. **Supervisé :** On connaît la partition des $n$ premier bras, on cherche à découvrir où classer le $(n+1)^{ème}$
3. **Séquentiel :** on peut observer plusieurs fois chaque individu/bras que l'on souhaite clusterer, on a un crédit $T = cn$ où $c$ est un grand entier.

Dans la littérature, on observe que les performances du cas non supervisé et supervisé sont sensiblement équivalente avec $K = 2$ groupes alors même que le cas supervisé est clairement plus facile. En revanche, lorsque $K$ est plus grand que 2, il est plus facile d'être en supervisé qu'en non supervisé. La principale question est de savoir si le clustering séquentiel s'approche plus du cas supervisé ou du cas non supervisé.

Avant d'étudier un "clustering séquentiel", on s'intéresse à l'apprentissage supervisé qui sera une référence en terme de performance. On part d'un cas simple dont on va relâcher au fur et à mesure les hypothèses.

## 3 Apprentissage supervisé

### 3.1 Apprentissage supervisé avec deux groupes symétriques.

#### (i) Hypothèses et risque de Bayes

On suppose qu'on observe un $n$-échantillon $(X_i,Y_i)_i$ de variables aléatoires telles que :

$Y_i$ à valeurs dans $\{-1,1\}$ est distribué selon une Rademacher.
$X_i \sim \mathcal{N}(Y_i\Delta,\sigma^2 I_d)$

$\Delta \in \mathbb{R}^d$

On a donc deux groupes symétriques par rapport à l'origine et de même variance labellisés par $\pm 1$. On observe $X_{new}$ et on cherche $Y_{new}$. Il s'agit d'un cas élémentaire d'apprentissage supervisé.
On cherche un classifieur $f : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \mapsto \{-1,1\}$ mesurable et de risque minimal au sens de la perte $0-1$, ie tel que la probabilité d'erreur $R(f) = \mathbb{E}_{X,Y}[\mathbb{1}\{f(X) \neq Y\}]$ soit minimal. Un tel classifieur est dit de Bayes.

LEMME 28. Si on connaît $\Delta$, le classifieur de Bayes est :

$$f(X) = \text{sign}\left\{\mathbb{P}(Y=1|X) - \mathbb{P}(Y=-1|X)\right\}$$

*Preuve* Soit $f'$ un classifieur, par des manipulations directes, on a $\mathbb{P}(f'(X) \neq Y|X=x) = \mathbb{1}_{f'(X)=1}(1-2\mathbb{P}(Y=1|X=x)) + \mathbb{P}(Y=1|X=x)$. Donc,

$$\begin{aligned} R(f') - R(f) &= \mathbb{E}_X[\mathbb{P}(f'(x) \neq Y|X=x) - \mathbb{P}(f(X) \neq Y|X=x)] \\ &= \mathbb{E}_X[(\mathbb{1}_{f'(x)=1} - \mathbb{1}_{f(x)=1})(1-2\mathbb{P}(Y=1|X=x)] \geqslant 0 \end{aligned}$$

La dernière inégalité provient de la construction de $f$ et s'obtient par disjonction de cas. □

LEMME 29. Dans le cadre du mélange Gaussien avec deux groupes de moyennes symétriques, on a :

$$f(X) = \text{sign}\left(\langle X, \Delta \rangle\right)$$

*Preuve* Par Bayes, on a (en notant $g_{\Delta,\sigma^2 I_d}$ la densité de la loi $\mathcal{N}(\Delta,\sigma^2 I_d)$) : $\mathbb{P}(Y=1|X) = \frac{g_{\Delta,\sigma^2 I_d}(X)\times\frac{1}{2}}{g_{\Delta,\sigma^2 I_d}(X)\times\frac{1}{2} + g_{-\Delta,\sigma^2 I_d}(X)\times\frac{1}{2}}$. Ainsi :

$$\begin{aligned} f(X) &= \text{sign}\left(\mathbb{P}(Y=1|X) - \mathbb{P}(Y=-1|X)\right) \\ &= \text{sign}\left(\exp(-\frac{1}{2\sigma^2}||X-\Delta||^2) - \exp(-\frac{1}{2\sigma^2}||X+\Delta||^2)\right) \\ &= \text{sign}\left(\langle X, \Delta \rangle\right) \end{aligned}$$

□

Le problème est que $\Delta$ n'est pas connu, on doit donc l'estimer par exemple via l'estimateur de maximum de vraisemblance $\hat{\Delta} = \frac{1}{n}\sum_{i=1}^{n} Y_i X_i$. Ensuite, on construit un nouveau classifieur $\hat{f}(X) = \text{sign}\left(\langle X, \hat{\Delta} \rangle\right)$. On doit donc utiliser un modèle plus compliqué où $\Delta$ est inconnu, on a déjà vu que pour trouver un classifieur de Bayes en apprentissage supervisé, il faut maximiser la probabilité a posterio d'appartenir à une classe. On considère que $\Delta$ est distribué uniformément sur la sphère de rayon $\|\Delta\|$. On note $\mathcal{L} = (X_1, Y_1, \ldots, X_n, Y_n)$ les données apprises et $S = B(0, \|\Delta\|)$ tel que $\Delta \sim \mathcal{U}(S)$. Le lemme suivant permet d'affirmer que le classifieur "plug-in" précédent est optimal au sens minimax.

LEMME 30.

$$\hat{f}(x) = \text{sign}\left(\langle X, \hat{\Delta}\rangle\right) = \text{sign}(\mathbb{P}(Y=1|X=x,\mathcal{L}) - \mathbb{P}(Y=-1|X=x,\mathcal{L}))$$

*Preuve* Soit $\delta = \pm 1$ et $x \in \mathbb{R}^d$.

$$\mathbb{P}(Y=\delta|X=x,\mathcal{L}) = \int_S \mathbb{P}(Y=\delta|X=x,\mathcal{L},\Delta)\,\mathrm{d}\mathbb{P}(\Delta|X=x,\mathcal{L})$$

Calculons l'intégrante.

$$\mathbb{P}(Y=\delta|X=x,\mathcal{L},\Delta) = \mathbb{P}(Y=\delta|X=x,\Delta) = \frac{e^{-0.5\|\delta x-\Delta\|^2/\sigma^2}}{e^{-0.5\|x-\Delta\|^2/\sigma^2} + e^{-0.5\|x+\Delta\|^2/\sigma^2}}$$

$$\mathrm{d}\mathbb{P}(\Delta|X=x,\mathcal{L}) \propto (e^{-0.5\|x-\Delta\|^2/\sigma^2} + e^{-0.5\|x+\Delta\|^2/\sigma^2})e^{-0.5\sum_{i=1}^n \|X_iY_i-\Delta\|^2/\sigma^2}$$

Maintenant, en notant $\gamma$ la mesure uniforme sur $S$, on calcul

$$\mathbb{P}(Y=\delta|X=x,\mathcal{L}) = \frac{\int_S e^{-0.5\|\delta x-\Delta\|^2/\sigma^2} e^{-0.5\sum_{i=1}^n \|X_iY_i-\Delta\|^2/\sigma^2}\,\mathrm{d}\gamma(\Delta)}{\int_S (e^{-0.5\|x-\Delta\|^2/\sigma^2} + e^{-0.5\|x+\Delta\|^2/\sigma^2}) e^{-0.5\sum_{i=1}^n \|X_iY_i-\Delta\|^2/\sigma^2}\,\mathrm{d}\gamma(\Delta)}$$

$$\frac{\int_S e^{-\langle \delta x+\sum_i Y_iX_i,\Delta\rangle/\sigma^2}\,\mathrm{d}\gamma(\Delta)}{\int_S e^{-\langle x+\sum_i Y_iX_i,\Delta\rangle/\sigma^2}\,\mathrm{d}\gamma(\Delta) + \int_S e^{-\langle -x+\sum_i Y_iX_i,\Delta\rangle/\sigma^2}\,\mathrm{d}\gamma(\Delta)}$$

Maintenant, on remarque que la fonction $v \mapsto \int_S e^{\langle v,\Delta\rangle}\,\mathrm{d}\gamma(\Delta)$ ne dépend que de $\|v\|$ et est strictement croissante en $\|v\|$, si bien que :

$$\begin{aligned}\mathbb{P}(Z=1|X=x,\mathcal{L}) > \mathbb{P}(Z=-1|X=x,\mathcal{L}) &\iff \|x+\sum_i Y_iX_i\|^2 > \|-x+\sum_i Y_iX_i\|^2 \\ &\iff \langle x, \sum_i Y_iX_i\rangle > 0\end{aligned}$$

□

◇ REMARQUE. Le risque est lié au choix de la distribution a priori $\gamma$. Il est donc bon de justifier ce choix. Ce choix rend compte d'une certaine isotropie du problème. Si on choisissait une distribution non uniforme sur la sphère, le classifieur de Bayes aurait tendance à se spécialiser dans la direction la plus probable ce qui peut réduire les performances dans les autres directions.

### (ii) Heuristique, condition de séparation.

Il est très intéressant d'observer les conditions de séparation nécessaires pour qu'un clustering se déroule bien, on donner ici une heuristique. On cherche l'écartement nécessaire entre les moyennes pour pouvoir classer correctement les points.

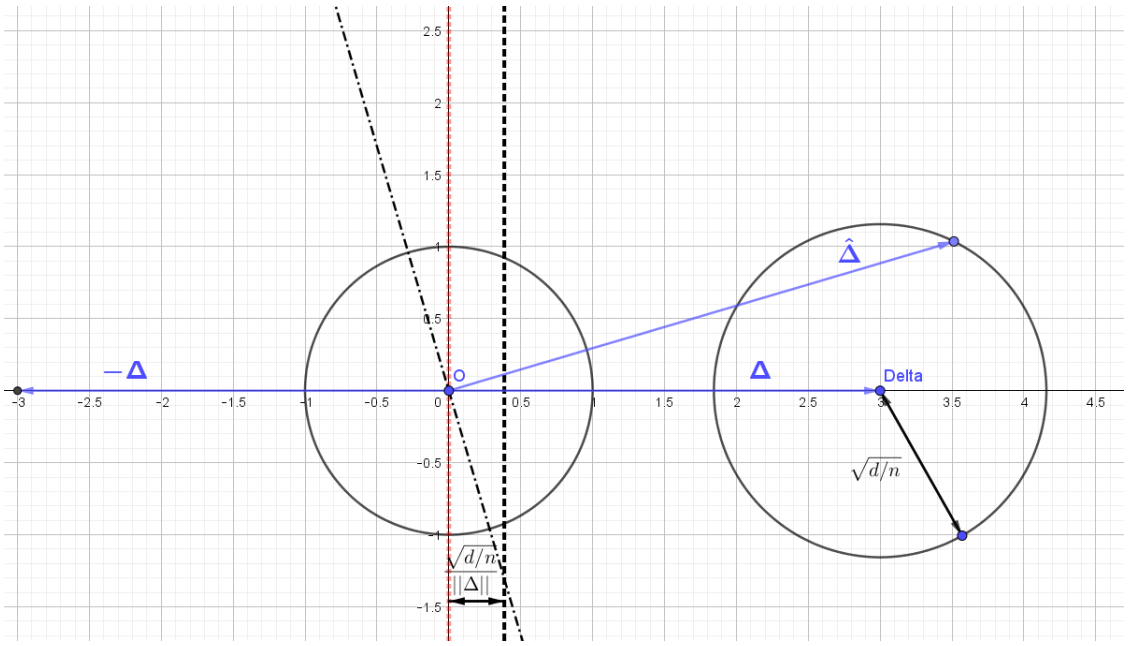
On remarque que $||\hat{\Delta}-\Delta||_2^2 \sim \frac{\sigma^2}{n}\chi^2_d$ et donc, en utilisant l'inégalité de concentration du $\chi^2$ démontrée en annexe, avec proba plus grande que $1-2\delta$ on a $\left|||\hat{\Delta}-\Delta||_2^2 - \frac{d\sigma^2}{n}\right| \leqslant \frac{2\sigma^2}{n}\sqrt{d\log(1/\delta)} + \frac{2\sigma^2}{n}\log(1/\delta)$. Maintenant, $\sup_{x\in S} |\langle x, \hat{\Delta} - \Delta\rangle| \leqslant \|\hat{\Delta}-\Delta\|_2^2 \lesssim \sqrt{\frac{d}{n}}\sigma$.

◇ Remarque. En grande dimension, le cas extrême où $\hat{\Delta}$ est placé de tel sorte que l'angle avec $\Delta$ soit maximal devient très probable. C'est pourquoi il est légitime de considérer ce cas extrême.

Lemme 31. $\forall x$, si $\left|\langle \frac{x}{\|x\|}, \frac{\Delta}{\|\Delta\|}\rangle\right| \geqslant \sqrt{\frac{d}{n}}\frac{\sigma}{\|\Delta\|}$ alors $\text{sign}\left(\langle x, \hat{\Delta}\rangle\right) \approx \text{sign}\left(\langle x, \Delta\rangle\right)$

*Preuve* $\langle x, \hat{\Delta}\rangle = \langle x, \Delta\rangle + \langle x, \hat{\Delta} - \Delta\rangle$. Si $\langle x, \Delta\rangle \geqslant 0$, on a par hypothèse $\langle \frac{x}{\|x\|}, \Delta\rangle \geqslant \sqrt{\frac{d}{n}}\sigma$ et $\langle \frac{x}{\|x\|}, \hat{\Delta} - \Delta\rangle \gtrsim -\sqrt{\frac{d}{n}}\sigma$ et $\langle x, \hat{\Delta}\rangle \gtrsim 0$. Idem cas négatif. □

◇ Remarque. La condition $\left|\langle \frac{x}{\|x\|}, \frac{\Delta}{\|\Delta\|}\rangle\right| \geqslant \sqrt{\frac{d}{n}}\frac{\sigma}{\|\Delta\|}$ n'est possible que si $\|\Delta\| \geqslant \sqrt{\frac{d}{n}}\sigma$, de plus, pour que l'estimateur de Bayes $f(x) = \langle x, \Delta\rangle$ soit performant, on a aussi besoin de $\|\Delta\| \geqslant \sigma$ si bien qu'on obtient la condition $\frac{\|\Delta\|}{\sigma} \geqslant 1 \vee \sqrt{\frac{d}{n}}$ qui est intrinsèque au problème.




## (iii) Risque de Bayes

Théorème 32.

$$\mathbb{P}(Y_{new} \neq \hat{f}(X_{new})) \leqslant 2\exp\left(-\frac{\|\Delta\|^2}{8\sigma^2} \wedge \frac{n\|\Delta\|^4}{16p\sigma^4}\right)$$

*Preuve* Par invariance par rotation et symétrie du problème, il suffit de supposer que $\Delta = [\|\Delta\|, 0, \ldots, 0]$ et que $X_{new} \sim \mathcal{N}(\Delta, \sigma I_d)$.

$$\mathbb{P}(Y_{new} \neq \hat{h}(X_{new})) = \mathbb{P}(\langle \hat{\Delta}, X_{new}\rangle < 0) = \mathbb{P}(\langle \Delta + \frac{\sigma}{\sqrt{n}}\varepsilon, \Delta + \sigma\varepsilon'\rangle < 0)$$

où $\varepsilon$ et $\varepsilon'$ suivent des $\mathcal{N}(0, I_d)$. On pose $W = -(\|\Delta\|\sqrt{1+\frac{1}{n}})^{-1}\langle \Delta, \frac{1}{\sqrt{n}}\varepsilon + \varepsilon'\rangle$ et

$Q = -\langle \varepsilon, \varepsilon' \rangle$. On réécrit alors,

$$\begin{aligned}\mathbb{P}(Y_{new} \neq \hat{f}(X_{new})) &= \mathbb{P}\left(\frac{\|\Delta\|^2}{\sigma^2} < \frac{\|\Delta\|}{\sigma}\sqrt{1+\frac{1}{n}}W + \frac{1}{\sqrt{n}}Q\right) \\ &\leqslant \mathbb{P}\left(W \geqslant \frac{\|\Delta\|}{2\sigma\sqrt{1+1/n}}\right) + \mathbb{P}\left(Q > \frac{\sqrt{n}\|\Delta\|^2}{2\sigma^2}\right)\end{aligned}$$

On estime chaque terme par une inégalité de concentration. On a $W \sim \mathcal{N}(0,1)$, donc $\mathbb{P}(W \geqslant \frac{\|\Delta\|}{2\sigma\sqrt{1+1/n}}) \leqslant \mathbb{P}(W \geqslant \frac{\|\Delta\|}{2\sigma}) \leqslant \exp(-\|\Delta\|^2/8\sigma^2)$. D'autre part, $\mathbb{P}(Q > \frac{\sqrt{n}\|\Delta\|^2}{2\sigma^2}) \leqslant \exp(-\frac{\sqrt{n}\|\Delta\|^2}{8\sigma^2} \wedge \frac{n\|\Delta\|^4}{16d\sigma^4})$ d'après le corollaire donné en annexe avec $A = -I_d$. On conclue en écrivant $a + b \leqslant 2a \vee b$. □

◇ Remarque. On retrouve les conditions de séparabilité obtenu par approximation au début de la section. En particulier, si $\frac{\|\Delta\|}{\sigma} \geqslant 1 \vee \sqrt{\frac{d}{n}}$, alors le risque est décroît exponentiellement avec $\|\Delta\|^2/\sigma^2$.

Théorème 33. Inversement, il existe $c,c'$ deux constantes telles que

$$\mathbb{P}(Y_{new} \neq \hat{f}(X_{new})) \geqslant c\exp\left(-c'\frac{\|\Delta\|^2}{8\sigma^2} \wedge \frac{n\|\Delta\|^4}{16p\sigma^4}\right)$$

### 3.2 Apprentissage supervisé avec deux groupes de moyennes quelconques.

On ne suppose plus que les deux groupes ont des moyennes opposées. On a donc $\Delta_1 \in \mathbb{R}^d$ et $\Delta_2 \in \mathbb{R}^d$ inconnus tels que $X \sim \mathbb{N}(\Delta_Y, \sigma^2 I_d)$ et $Y \sim \mathcal{U}(\{1,2\})$. Toujours en supervisé, on observe les données d'apprentissage $\mathcal{L} = ((X_1,Y_1),\ldots,(X_n,Y_n))$, on observe $X$ et on cherche $Y$. On note $N_1 = \sum_{i=1}^n \mathbb{1}_{Y_i=1}$ (resp $N_2$), $\hat{\Delta}_1 = \frac{1}{N_1}\sum_{i=1}^n X_i \mathbb{1}_{Y_i=1}$ (resp $\hat{\Delta}_2$) et enfin $\Delta = \frac{\|\Delta_1 - \Delta_2\|}{2}$. On construit le classifieur naturel $\hat{f}(X) = \underset{i=1,2}{\operatorname{argmin}} \|X - \hat{\Delta}_i\|^2$.

Théorème 34. Il existe $c,c'$ deux constantes telles que :

$$\mathbb{P}(\hat{f}(X) \neq Y) \leqslant c\exp\left(-c'\left(\frac{\Delta^2}{\sigma^2} \wedge \frac{n\Delta^4}{d\sigma^4}\right)\right)$$

*Preuve* On suppose par commodité que $N_1 = N_2$ et sans perte de généralité que $Y = 2$. On cherche la probabilité que $\hat{f}(X) = 1$. Or $\|X-\hat{\Delta}_1\|^2 - \|X-\hat{\Delta}_2\|^2 = 4\langle X - \frac{\hat{\Delta}_1+\hat{\Delta}_2}{2}, \frac{\hat{\Delta}_2-\hat{\Delta}_1}{2}\rangle$. Ainsi $\hat{f}(X) = 1 \Leftrightarrow \langle X - \frac{\hat{\Delta}_1+\hat{\Delta}_2}{2}, \frac{\hat{\Delta}_2-\hat{\Delta}_1}{2}\rangle < 0$. Par hypothèse, on a $\varepsilon,\varepsilon'$ et $\varepsilon''$ trois vecteur Gaussien $\mathcal{N}(0,I_d)$ tels que :

$$\begin{aligned}X &= \Delta_2 + \sigma\varepsilon \\ \hat{\Delta}_1 &= \Delta_1 + \frac{\sigma}{\sqrt{n/2}}\varepsilon' \\ \hat{\Delta}_2 &= \Delta_2 + \frac{\sigma}{\sqrt{n/2}}\varepsilon''\end{aligned}$$

Maintenant,

$$\begin{aligned}
&\left\langle X - \frac{\hat{\Delta}_1 + \hat{\Delta}_2}{2} \middle| \frac{\hat{\Delta}_2 - \hat{\Delta}_1}{2} \right\rangle \\
=&\left\langle \frac{\Delta_2 - \Delta_1}{2} + \sigma\varepsilon - \frac{\sigma}{\sqrt{2n}}\varepsilon' - \frac{\sigma}{\sqrt{2n}}\varepsilon'' \middle| \frac{\Delta_2 - \Delta_1}{2} + \frac{\sigma}{\sqrt{2n}}\varepsilon'' - \frac{\sigma}{\sqrt{2n}}\varepsilon' \right\rangle \\
=&\sigma^2 \left[ \frac{\Delta^2}{\sigma^2} + \left\langle \frac{\Delta_2 - \Delta_1}{2\sigma}, \varepsilon - \sqrt{\frac{2}{n}}\varepsilon' \right\rangle + \left\langle \varepsilon, \frac{1}{\sqrt{2n}}(\varepsilon'' - \varepsilon') \right\rangle + \frac{1}{\sqrt{2n}}(\|\varepsilon'\|^2 - \|\varepsilon''\|^2) \right]
\end{aligned}$$

On pose $W = \left(\frac{\Delta}{\sigma}\sqrt{1+\frac{2}{n}}\right)^{-1} \left\langle \frac{\Delta_2 - \Delta_1}{2\sigma}, \sqrt{\frac{2}{n}}\varepsilon' - \varepsilon \right\rangle \sim \mathcal{N}(0,1)$ et $Q = \langle \varepsilon, \theta$ avec $\theta = \frac{\varepsilon' - \varepsilon''}{\sqrt{2}} \sim \mathcal{N}(0,1)$. Alors, on écrit :

$$\begin{aligned}
\mathbb{P}(f(X) \neq Y) &= \mathbb{P}\left( \frac{\Delta^2}{\sigma^2} < \frac{\Delta}{\sigma}\sqrt{1+\frac{2}{n}}W + \frac{1}{\sqrt{n}}Q + \frac{1}{2n}(\|\varepsilon''\|^2 - \|\varepsilon'\|^2) \right) \\
&\leqslant \mathbb{P}\left( \frac{\Delta^2}{3\sigma^2} < \frac{\Delta}{\sigma}\sqrt{1+\frac{2}{n}}W \right) + \mathbb{P}\left( \frac{\Delta^2}{3\sigma^2} < \frac{1}{\sqrt{n}}Q \right) + \mathbb{P}\left( \frac{\Delta^2}{3\sigma^2} < \frac{1}{2n}(\|\varepsilon''\|^2 - \|\varepsilon'\|^2) \right)
\end{aligned}$$

Avec les inégalités de concentration données en annexe, on obtient :

$$\mathbb{P}(W > \frac{\Delta}{3\sigma\sqrt{1+2/n}}) \leqslant \exp\left(-\frac{\Delta^2}{18\sigma^2}\right)$$

$$\mathbb{P}(Q > \frac{\sqrt{n}\Delta^2}{3\sigma^2}) \leqslant \exp\left(-\frac{\sqrt{n}\Delta^2}{12\sigma^2} \wedge \frac{n\Delta^4}{36d\sigma^4}\right)$$

$$\begin{aligned}
\mathbb{P}(\|\varepsilon''\|^2 > \|\varepsilon'\|^2 + \frac{2n\Delta^2}{3\sigma^2}) &\leqslant \mathbb{P}(\|\varepsilon''\|^2 - d > \frac{n\Delta^2}{3\sigma^2}) + \mathbb{P}(d - \|\varepsilon'\|^2 > \frac{n\Delta^2}{3\sigma^2}) \\
&\leqslant 2\exp\left(-\frac{1}{8}\left(\frac{n\Delta^2}{3\sigma^2}\right) \wedge \frac{1}{d}\left(\frac{n\Delta^2}{3\sigma^2}\right)^2\right) \\
2 &\leqslant \exp\left(-\frac{1}{8}\left(\frac{n\Delta^2}{3\sigma^2} \wedge \frac{n^2\Delta^4}{9\sigma^4 d}\right)\right)
\end{aligned}$$

On conclue avec $a + b + c \leqslant 3(a \vee b \vee c)$. □

### 3.3 Apprentissage supervisé avec $K$ groupes équilibrés

#### (i) Moyennes connues

LEMME 35. S'il y a $K$ classes équilibrées de variance $\sigma^2 I_d$ et de moyennes connues $\Delta := \Delta_1, \ldots, \Delta_K$, le classifieur de Bayes est :

$$f(x) = \text{argmin}_{y=1,\ldots,K} \|x - \Delta_y\|^2$$

THÉORÈME 36. Pour "bien" partitionner $K$ avec le classifieur précédent, on a

besoin de la condition

$$\min_{\substack{i,i'\\ i\neq i'}} \frac{\|\Delta_i - \Delta_{i'}\|}{\sigma} \gtrsim \sqrt{\log K}$$

◇ REMARQUE. On se place dans le cas plus simple et visuel où $\Delta_k = re_k$ où $e_k$ est la base canonique, comme ces moyennes partagent une même norme, cet estimataeur se simplifie en $f(x) = \underset{i=1,\dots,K}{\text{argmax}} \langle x,\Delta_i\rangle = \underset{i=1,\dots,K}{\text{argmax}}\, x_i$. Dans ce cas, chaque moyenne est équidistante des autres.

*Preuve* Si $\Delta_k = re_k$ pour $k \leqslant K$, on observe $X = (X_1,\dots,X_d)$ et on estime $Y$ par $f(X)$ donné à la remarque ci-dessus. On suppose par exemple que $Y = j$ et $\Delta_j = (\Delta_1^j,\dots,\Delta_d^j)$. L'inégalité de Cramér pour les variables Gaussienne assure que pour $i \leqslant K$, avec probabilité plus grande que $1-\delta$, $|X_i - \Delta_i^j| \leqslant \sqrt{2\sigma^2\log(2/\delta)}$. Avec une borne d'union, on obtient qu'avec probabilité $1-\delta$, $\forall i \leqslant K$, $|X_i - \Delta_i^j| \leqslant \sqrt{2\sigma^2\log(2K/\delta)}$. Or, dans notre cas, on a $\Delta_j = (0,\dots,re_j,0\dots,0)$. Si on a la condition de séparation $r \geqslant 2\sqrt{2\sigma^2\log(2K/\delta)}$ alors avec probabilité plus grande que $1-\delta$, on a $\forall i \neq j$, $X_i \leqslant \sqrt{2\sigma^2\log(2K/\delta)} \leqslant r - \sqrt{2\sigma^2(2K/\delta)} \leqslant X_j$ et donc $X_j = \max_{i\leqslant K} X_i$, ie $f(X) = Y$. En reformulant, on a :

$$\mathbb{P}(f(X) \neq Y) \leqslant 2K\exp\left(-\frac{1}{8}\left(\frac{r}{\sigma}\right)^2\right)$$

□

THÉORÈME 37. Sous les hypothèses du lemme précédent, on a :

$$\mathbb{P}(f(X) \neq Y) \leqslant \delta$$

si la condition de séparation suivante est vérifiée :

$$\min_{\substack{i,j=1,\dots,K\\ i\neq j}} \frac{\|\Delta_i - \Delta_j\|}{\sigma} \geqslant \sqrt{8\log(2K/\delta)}$$

*Preuve* On ne suppose plus rien sur $\Delta_1,\dots,\Delta_K$ sauf que ces moyennes sont connues. Soit $X \in \mathbb{R}^d$, on suppose sans perte de généralité que $Y = j$, ainsi $\mathbb{E}[X] = \Delta_j$. On estime $Y$ via

$$f(X) = \text{argmin}_{i=1,\dots,K}\|X - \Delta_i\|^2.$$

On a $\|X-\Delta_i\|^2 - \|X-\Delta_j\|^2 = \|\Delta_i\|^2 - \|\Delta_j\|^2 + 2\langle X,\Delta_j - \Delta_i\rangle$. De plus, $\langle 2X,\Delta_j-\Delta_i\rangle \sim \mathcal{N}(2\langle\Delta_j,\Delta_j-\Delta_i\rangle, 4\sigma^2\|\Delta_j-\Delta_i\|^2)$. Avec Cramér et une borne d'union, on a avec probabilité $1-\delta$,

$$\forall 1 \leqslant i \leqslant K, |\langle 2X,\Delta_j-\Delta_i\rangle - \langle 2\Delta_j,\Delta_j-\Delta_i\rangle| \leqslant \sqrt{8\sigma^2\|\Delta_j-\Delta_i\|^2\log(2K/\delta)}.$$

Maintenant, si $\forall i \neq j$, $\frac{\|\Delta_i-\Delta_j\|}{\sigma} \leqslant \sqrt{8\log(2K/\delta)}$ alors avec probabilité supérieure à $1-\delta$ on a $\forall i \neq j$, $\langle 2X,\Delta_j-\Delta_i\rangle + \|\Delta_i\|^2 - \|\Delta_j\|^2 \geqslant \|\Delta_i-\Delta_j\|^2 -$

$\sqrt{8\sigma^2\|\Delta_i-\Delta_j\|^2\log(2K/\delta)} \geqslant 0$ et ainsi on a $f(X) = \text{argmax}_{i=1,\dots,K}\|X-\Delta_i\|^2 = j = Y$. □

### (ii) Moyennes inconnues

Toujours dans le cadre supervisé avec des points répartis uniformément dans $K$ groupes de moyennes quelconques inconnues. On observe les données d'apprentissage $\mathcal{L} = ((X_1,Y_1),\dots,(X_n,Y_n))$. On suppose qu'on a appris uniformément chaque groupe, c'est à dire que pour tout $i = 1,\dots,n$, $N_i := \sum_{j=1}^n \mathbb{1}\{Y_j = i\} = \frac{n}{k}$. Pour $i = 1,\dots,n$, on note $\hat{\Delta}_i = \frac{1}{N_i}\sum_{j=1}^n X_j\mathbb{1}\{Y_j = i\}$. On observe $X$ et on estime $Y$ via $\hat{f}(X) = \text{argmin}_{i=1,\dots,K}\|X-\hat{\Delta}_i\|^2$.

THÉORÈME 38. Il existe $c$,$c'$ deux constantes telles que :

$$\mathbb{P}(\hat{f}(X)\neq Y) \leqslant Kc\exp\left(-c'\left(\frac{\Delta^2}{\sigma^2}\wedge\frac{n\Delta^4}{Kd\sigma^4}\right)\right)$$

où $\Delta = \min_i \frac{\|\Delta_i-\Delta_j\|}{2}$

◇ REMARQUE. On trouve une condition de séparation de la forme

$$\frac{\Delta}{\sigma} \gtrsim \sqrt{\log(K)} \vee \sqrt{\frac{\log(K)dk}{n}}.$$

*Preuve* On suppose sans perte de généralité que $Y = 1$, alors l'évènement $\{\hat{f}(X)\neq 1\}$ s'écrit :

$$\left\{\hat{f}(X)\neq 1\right\} = \left\{\exists i\neq 1; \|X-\hat{\Delta}_i\|^2 < \|X-\hat{\Delta}_1\|^2\right\}.$$

Or $\|X-\hat{\Delta}_i\|^2 < \|X-\hat{\Delta}_1\|^2 \Leftrightarrow \langle X-\frac{\hat{\Delta}_1+\hat{\Delta}_i}{2}, \frac{\Delta_1-\Delta_i}{2}\rangle < 0$. D'où :

$$\begin{aligned}\mathbb{P}(\hat{f}(X)\neq 1) &= \mathbb{P}\left(\bigcup_{i\neq 1}\{\langle X-\frac{\hat{\Delta}_1+\hat{\Delta}_i}{2}, \frac{\Delta_1-\Delta_i}{2}\rangle < 0\}\right)\\ &\leqslant \sum_{i\neq 1}\mathbb{P}\left(\langle X-\frac{\hat{\Delta}_1+\hat{\Delta}_i}{2}, \frac{\Delta_1-\Delta_i}{2}\rangle < 0\right)\\ &\leqslant Kc\exp\left(-c'\left(\frac{\Delta^2}{\sigma^2}\wedge\frac{n\Delta^4}{kd\sigma^4}\right)\right)\end{aligned}$$

La dernière inégalité est obtenue en reprenant la démonstration du cas $k = 2$ (théorème 52) en notant que chaque moyenne est estimé à partir de $n/k$ observations plutôt que $n/2$ ce qui explique le $k$ apparaissant au dénominateur. □

## 4 Clustering séquentiel

On s'intéresse dans ce paragraphe à la version séquentiel du clustering. On a un crédit de $T = cn$ observations et on souhaite à l'issue de ces $T$ observa-

tions retrouver la partition des bras. La principale différence avec le clustering classique (batch) est que l'on s'autorise à observer plusieurs fois les individus (ici les bras) à partitionner. L'objectif est de retrouver à partir de $X_1,\dots,X_T$ la partition $G^*$.

### 4.1 Cas, deux groupes symétriques.

On se place dans le cas simplifié où il y a $K=2$ groupes de moyennes $+\Delta$ et $-\Delta$. Il existe $Y_1,\dots,Y_n \in \{-1,1\}$ variables aléatoires latentes telles que le $i$ème bras est distribué selon $P_i = \mathcal{N}(Y_i\Delta,\sigma^2 I_d)$ et $G^* = \{\{i : Y_i = +1\},\{i : Y_i = -1\}\}$. On a un budget $T = 2n$

**Algorithme 4** Clustering séquentiel 2 groupes symétriques

**pour** $t = 1,\dots,n+1$ **faire**
  **Tirer** le bras 1, $A_t = 1$
  **Obtenir** $X_t$ selon $P_1$
  **Estimer** $\hat{Y}_1 := 1$
**fin pour**
**pour** $t = n+2,\dots,2n$ **faire**
  **Tirer** le bras $t-n$, $A_t = t-n$
  **Obtenir** $X_t$ selon $P_{A_t}$
  **Estimer** $\hat{Y}_{t-n} = \text{sign}(\langle X_t, \frac{1}{n+1}\sum_{j=1}^{n+1} X_j\rangle)$
**fin pour**
**Sortie:** $\hat{G} = \{\{i : \hat{Y}_i = 1\},\{i : \hat{Y}_i = -1\}\}$

Analysons la performance de cet procédure, elle consiste d'abord à tirer $n+1$ fois le même bras 1 pour estimer $\hat{\Delta} = \frac{1}{n+1}\sum_{t=1}^{n+1} X_t$. On remarque que $\hat{\Delta} \sim \mathcal{N}(Y_1\Delta, \frac{\sigma^2}{n+1} I_d)$. On souhaite obtenir un bon partitionnement des bras, à permutation des indices des classes près.

Définition 39. Le taux d'erreur de $\hat{G}$ par rapport à $G^*$ est :

$$\text{err}(\hat{G},G^*) = \min_{\delta=\pm 1} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\hat{Y}_i \neq \delta Y_i\}$$

Lemme 40.

$$\mathbb{E}[\text{err}(\hat{G},G^*)] \leqslant 2\exp\left(-\frac{\|\Delta\|^2}{8\sigma^2} \wedge \frac{n\|\Delta\|^4}{16p\sigma^4}\right)$$

Et ainsi :

$$\mathbb{P}(\hat{G} = G^* \text{ ou } \hat{G} = -G^*) \geqslant 1 - 2n\exp\left(-\frac{\|\Delta\|^2}{8\sigma^2} \wedge \frac{n\|\Delta\|^4}{16p\sigma^4}\right)$$

*Preuve* Avec les mêmes calculs que le cas supervisé, on obtient que

$$\begin{aligned}\mathbb{E}[\text{err}(\hat{G},G^*)] = \min_{\delta=\pm 1}\frac{1}{n}\sum_{i=1}^{n}\mathbb{P}(\hat{Y}_i \neq \delta Y_i) &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{P}(\hat{Y}_i \neq Y_1 Y_i) \\ &\leqslant 2\exp\left(-\frac{\|\Delta\|^2}{8\sigma^2} \wedge \frac{n\|\Delta\|^4}{16p\sigma^4}\right)\end{aligned}$$

La variable $\mathrm{err}(\hat{G},G^*)$ est à valeur discrète, ne pas faire d'erreur revient à avoir strictement moins de $1/n$ proportion d'erreur. On utiliser ensuite Markov.

$$\begin{aligned}\mathbb{P}(\hat{G} = G^* \text{ ou } \hat{G} = -G^*) = \mathbb{P}(\mathrm{err}(\hat{G},G^*) < 1/n) &\geqslant n\mathbb{E}[\mathrm{err}(\hat{G},G^*)] \\ &\geqslant 1 - 2n\exp\left(-\frac{\|\Delta\|^2}{8\sigma^2} \wedge \frac{n\|\Delta\|^4}{16p\sigma^4}\right)\end{aligned}$$

□

◇ Remarque. La procédure précédente est optimale. En effet, on obtient la même performance (à une constante multiplicative près) que la classification supervisée. On a montré dans la section précédente, que cette performance est optimale pour le cas supervisé. Or, par réduction, on peut montrer que la performance d'un algorithme de clustering séquentiel est nécessairement moins bonne qu'en classification. On a donc atteint une performance optimale.

◇ Remarque. Dans la fin de ce paragraphe, on verra qu'il n'était pas gagné que la procédure soit optimal sans utiliser un argument de réduction. En effet, l'estimateur utilisé n'est pas l'estimateur de Bayes. Après avoir fait les observations $X_1,\dots,X_{2n}$ avec la procédure décrite dans l'algorithme 4, la meilleure estimation de $Y_1Y_i$ que l'on puisse faire est l'estimateur de Bayes $\mathrm{argmax}_{\delta=\pm1}\mathbb{P}(Y_i = \delta Y_1|X_1,\dots,X_{2n})$, le lemme suivant affirme que contrairement au cas supervisé, l'estimateur utilisé dans l'algorithme n'est pas l'estimateur de Bayes. Cela s'explique par le fait que les observations des $n-1$ bras de labels inconnus donnent tout de même des informations sur $\Delta$. En revanche, l'expression de l'estimateur de Bayes est compliquée et on a vu au dessus qu'une démarche naïve inspirée de la classification supervisée suffisait.

Lemme 41. Soit $i > 1$ et soit $\delta \in \{-1,1\}$. On note $\hat{\Delta}_1 := \frac{1}{n+1}\sum_{t=1}^{n+1} X_t$. Alors :

$$\hat{Y}_i = \mathrm{sign}(\langle X_{n+i},\hat{\Delta}_1\rangle) = \mathrm{argmax}_{\delta=\pm1}\mathbb{P}(Y_i = \delta Y_1|X_1,\dots,X_n;X_{n+i})$$

Si $i > 1$, on aurait pu estimer $Y_1Y_i$ par

$$\hat{Y}_i^{Bayes} := \mathrm{argmax}_{\delta=\pm1}\mathbb{P}(Y_i = \delta Y_1|X_1,\dots,X_{2n})$$

où on a :

$$\begin{aligned}&\mathbb{P}(Y_i = \delta Y_1|X_1,\dots,X_{2n}) \\ =&\mathrm{sign}\left[\int_S \prod_{\substack{j=n+2\\ j\neq n+i}}^{2n} \cosh(\langle X_i,\Delta\rangle)\cosh(\langle \delta X_{n+i} + \sum_{i=1}^{n+1} X_i,\Delta\rangle)\,\mathrm{d}\gamma(\Delta)\right]\end{aligned}$$

## 4.2 Clustering séquentiel à $K$ groupes

### (i) Algorithme à erreur $\delta$ près

On s'intéresse au problème de clustering séquentiel à $K$ groupes en dimension $d$. On rappel qu'on a $n$ bras répartis en $K$ groupes de même moyenne. On note $\mu_1,\dots,\mu_k$ ces moyennes de sorte que si le bras $a$ est dans le groupe $k$, sa loi est $\mathcal{N}(\mu_k,\sigma^2 I_d)$. On note $Y_a$ la variable aléatoire donnant la classe du bras $a$. On se donne un budget de $T = 3n$ samples autorisés. On va suivre l'algorithme suivant :

**Algorithme 5** Clustering séquentiel K groupes

**Entrée:** $\Delta$,$\delta$

1: Tirer aléatoirement $N_\delta$ bras où $N_\delta := \left\lceil \frac{32}{9} K \log\left(\frac{K}{\delta}\right)\right\rceil$. On note ces bras $(a_j)_{j=1,\dots,N_\delta}$.
2: **pour** $j = 1,\dots,N$ **faire**
3: Sampler $T_\delta = \frac{T}{3N_\delta}$ fois le bras $a_j$.
4: **Calculer** la moyenne empirique de ces samples de $a_j$, on note cette moyenne $\tilde{\mu}_{a_j}$.
5: **fin pour**
6: **pour** $k = 1,\dots,K$ **faire**
7: S'il existe, prendre un bras de la liste $a_1,\dots,a_N$ non éliminé et non identifié et le noter $a_k^*$.
8: Identifier ce bras $a_k^*$ comme représentant d'une nouvelle classe et noter sa moyenne empirique $\hat{\mu}_k^*$.
9: Éliminer tous les bras $j$ tel que $\|\tilde{\mu}_{a_j} - \hat{\mu}_k^*\|$ soit inférieur à $\sigma\frac{\Delta}{2}$.
10: Sampler le bras $a_i^*$ $T/3K$ fois puis calculer la moyenne empirique de ces sample notée $\hat{\mu}_{a_i^*}$.
11: **fin pour**
12: **pour** $i = 1,\dots,n$ **faire**
13: Sampler le bras $i$ une fois.
14: Classer le bras $i$ dans le groupe $j$ tel que $j = \mathrm{argmin}_{j=1,\dots,K}\|X_i - \hat{\mu}_{a_j^*}\|$.
15: **fin pour**

◇ Remarque. Pour appliquer l'algorithme, on doit connaître

$$\Delta := \min_{\substack{i,j=1,\dots,K \\ i\neq j}} \frac{\|\mu_i - \mu_j\|}{2}.$$

Théorème 42. Si on applique l'algorithme 8, avec en entrée une marge d'erreur $\delta$ et un écart $\Delta = \min_{\substack{i,j=1,\dots,K \\ i\neq j}} \frac{\|\mu_i-\mu_j\|}{2}$ et si de plus on a la condition de séparation suivante :

$$\frac{\Delta^2}{\sigma^2} \geqslant 4\left(\frac{d}{T_\delta} + \frac{\sqrt{8d\log(N_\delta/\delta)}}{T_\delta} \wedge \frac{8}{T_\delta}\log\left(\frac{N_\delta}{\delta}\right)\right)$$

où $N_\delta := \left\lceil \frac{32}{9} K \log\left(\frac{K}{\delta}\right)\right\rceil$ et $T_\delta = \frac{T}{3N_\delta}$ alors on arrive à bien classer les bras. En outre, on a une alors

$$\mathbb{P}(\hat{Y}_i \neq Y_i) \leqslant 2\delta + Kc\exp\left(-c'\left(\frac{\Delta^2}{\sigma^2} \wedge \frac{n\Delta^4}{Kd\sigma^4}\right)\right).$$

*Preuve Etape 1 :*

Pour une classe $k = 1, \dots, K$ et un bras $a = 1, \dots, n$, on note $B_{a,k} = \mathbb{1}_{\{Y_a = k\}}$. On commence par tirer $N_\delta := \left\lceil \frac{32}{9} K \log\left(\frac{K}{\delta}\right) \right\rceil$ bras que l'on note $a_1, \dots, a_{N_\delta}$. Soit $k = 1, \dots, K$, $\mathbb{P}(\sum_{j=1}^{N_\delta} B_{a_j,k} \leqslant \frac{1}{4}\frac{N_\delta}{K}) \leqslant \exp\left(-\frac{9}{32}\frac{N_\delta}{K}\right) = \frac{\delta}{K}$. En effet, on a pris $N_\delta = \frac{32}{9} K \log\left(\frac{K}{\delta}\right)$. On a utilisé ici le point 4 du lemme de concentration de Chernoff en annexe. Une borne d'union sur $k$ permet d'affirmer que avec probabilité supérieure à $1 - \delta$ on a $\forall k = 1, \dots, K$ :

$$\sum_{a=1}^{N_\delta} B_{a,k} \geqslant \frac{1}{4}\frac{N_\delta}{K}.$$

En particulier, sur un évènement de probabilité supérieure à $1 - \delta$, on a tiré au moins un bras de chaque classe parmi les $N_\delta$ tirés.

*Etape 2 :* On utilise un tiers du budget pour sampler uniformément les $N_\delta$ bras tirés. On tire donc chaque bras $T_\delta := \frac{T}{3N_\delta}$ fois. Pour $j = 1 \dots, N_\delta$, on note $\hat{\mu}_{a_j}$ la moyenne empirique du bras $a_j$ après $T_\delta$ samples. Observons les déviations de chacune des moyennes empiriques. L'inégalité de concentration du lemme 48 en annexe et une borne d'union assure qu'avec probabilité supérieure à $1 - \delta$, pour tout $i = 1, \dots, N_\delta$,

$$\frac{\|\hat{\mu}_{a_i} - \mu_{a_i}\|^2}{\sigma^2} \leqslant \frac{d}{T_\delta} + \frac{\sqrt{8d\log(N_\delta/\delta)}}{T_\delta} \wedge \frac{8}{T_\delta}\log\left(\frac{N_\delta}{\delta}\right).$$

*Etape 3 :* Maintenant, supposons que l'on ait la condition de séparation suivante :

$$\frac{\Delta^2}{\sigma^2} \geqslant 4\left(\frac{d}{T_\delta} + \frac{\sqrt{8d\log(N_\delta/\delta)}}{T_\delta} \wedge \frac{8}{T_\delta}\log\left(\frac{N_\delta}{\delta}\right)\right).$$

On se place sur l'intersection des événements de forte probabilité des étapes 1 et 2. Si $a_i$ et $a_j$ sont dans la même classe, on a grâce à la condition de séparation et aux inégalités de l'étape 2 : $\|\tilde{\mu}_{a_j} - \tilde{\mu}_{a_i}\| \leqslant \|\tilde{\mu}_{a_j} - \mu_{a_j}\| + \|\tilde{\mu}_{a_i} - \mu_{a_i}\| \leqslant \sigma\left(\frac{\Delta}{4} + \frac{\Delta}{4}\right) = \sigma\frac{\Delta}{2}$. En revanche, si $a_i$ et $a_j$ sont dans deux classes différentes, on a $\|\tilde{\mu}_{a_j} - \tilde{\mu}_{a_i}\| \geqslant \|\mu_{a_i} - \mu_{a_j}\| - \|\tilde{\mu}_{a_j} - \mu_{a_j}\| - \|\tilde{\mu}_{a_i} - \mu_{a_i}\| \geqslant \sigma\left(2\Delta - \frac{\Delta}{4} - \frac{\Delta}{4}\right) = \frac{3\sigma\Delta}{2}$. Ensuite, on a parmi les $a_1, \dots, a_{N_\delta}$ un représentant de chaque classe au moins. Si on prend un de ces bras $a_i$ et qu'on élimine tous les bras $a_j$ tels que $\|\tilde{\mu}_{a_j} - \tilde{\mu}_{a_i}\| \leqslant \sigma\frac{\Delta}{2}$, on élimine tous les bras de la même classe exactement. En répétant l'opération, on peut sélectionner un bras de chaque classe. On obtient ainsi des représentants $a_1^*, \dots, a_K^*$ avec un bras dans chaque classe.

*Etape 4 :* C'est alors presque gagné, on se ramène alors à une configuration supervisé, en effet, si les 3 premières étapes dédiées à la sélection d'un bras dans chaque groupe a bien marché, on rejoue $n/K$ fois chaque représentant et une fois chaque bras et on applique la méthode supervisée et le théorème 38 pour obtenir pour tout bras une estimation $\hat{Y}_i$ de sa classe $Y_i$. Maintenant, les risques d'erreur lors des phases 1 et 2 et le risque supervisé du théorème 38

s'additionnent et pour tout bras $i = 1, \dots, n$,

$$\mathbb{P}(\hat{Y}_i \neq Y_i) \leqslant 2\delta + Kc \exp\left(-c'\left(\frac{\Delta^2}{\sigma^2} \wedge \frac{n\Delta^4}{Kd\sigma^4}\right)\right).$$

□

◇ Remarque. On va maintenant chercher à interpréter la condition de séparation du théorème 46. Déjà, on

$$\begin{aligned}
&\left(\frac{d}{T_\delta} + \frac{\sqrt{8d\log(N_\delta/\delta)}}{T_\delta} \wedge \frac{8}{T_\delta}\log\left(\frac{N_\delta}{\delta}\right)\right) \\
=&\frac{32K\log(K/\delta)}{9n}\left(d + \sqrt{8d\log(\frac{32K}{9\delta}\log(K/\delta))} \bigwedge 8\log(\frac{32K}{9\delta}\log(K/\delta))\right) \\
\leqslant&\frac{32K\log(K/\delta)}{9n}\left(d + \sqrt{16d\log(\frac{32K}{9\delta})} \bigwedge 16\log(\frac{32K}{9\delta})\right)
\end{aligned}$$

Maintenant, en grande dimension, si $d \geqslant \log(K/\delta)$, il suffit de vérifier la condition

$$\frac{\Delta}{\sigma} \geqslant c''\sqrt{\frac{dK}{n}\log\left(\frac{K}{\delta}\right)}$$

où $c''$ est une constante universelle.

◇ Remarque. On peut par exemple choisir $\delta = \frac{1}{2n}$. C'est une exigence raisonnable. En effet, on a alors $\mathbb{E}[\text{nombre d'erreur}] \leqslant 2n\delta + np \leqslant 1 + np$. (On a noté $p$ la probabilité d'erreur supervisée du théorème 46).

# 5 Annexe

## 5.1 Concentration

Lemme 43. **Concentration du $\chi^2$**
Si $Z \sim \chi^2_d$, alors pour tout $x > 0$,

$$\mathbb{P}(Z - d \geqslant 2\sqrt{dx} + 2x) \leqslant \exp(-x)$$

*Preuve* On utilise la méthode de Cramér-Chernoff pour la variable aléatoire $Z - d$ après avoir prouvé qu'elle est sous-gamma. On va d'abord montrer que pour tout $0 < u < 1/2$ alors $\log \mathbb{E}[\exp(u(Z-d))] \leqslant \frac{Du^2}{1-2u}$. Si $Y \sim \mathcal{N}(0,1)$,

$$\begin{aligned}
\log \mathbb{E}[\exp(u(Y^2-1))] &= -u - \frac{1}{2}\log(1-2u) \\
&= 2u^2 \sum_{k\geqslant 0} \frac{(2u)^k}{k+2} \\
&\leqslant u^2 \sum_{k\geqslant 0} (2u)^k = \frac{u^2}{1-2u}
\end{aligned}$$

Maintenant $\log \mathbb{E}[\exp(u(Z-d))] = \sum_{i=1}^{d} \log \mathbb{E}[\exp(u(Y_i^2-1))] \leqslant \frac{Du^2}{1-2u}$.
Une étude de fonction assure que pour $x \geqslant 0$,

$$\max_{0<u<1/2}(xu - \frac{Du^2}{1-2u}) = \frac{D}{2}\left[1 + \frac{x}{D} - \sqrt{1+\frac{2x}{D}}\right].$$

L'inégalité de Cramér-Chernoff assure que $\mathbb{P}(Z-d \geqslant x) \leqslant \exp(-\frac{D}{2}[1+\frac{x}{D} - \sqrt{1+\frac{2x}{D}}])$. Il siffit désormais d'inverser la fonction la fonction $x \mapsto \frac{D}{2}[1+\frac{x}{D} - \sqrt{1+\frac{2x}{D}}]$ pour obtenir le lemme. □

LEMME 44. **Concentration d'un produit scalaire de vecteurs Gaussiens**
Soit $\varepsilon \sim \mathcal{N}(0,I_d)$ et $S$ une matrice symétrique réelle. Alors $\forall L \geqslant 0$,

$$\mathbb{P}\left(\varepsilon^T S\varepsilon - \mathrm{Tr}(S) > \sqrt{8\|S\|_F^2 L} \wedge 8|S|_{Op} L\right) \leqslant \exp(-L)$$

*Preuve* $S$ est symétrique réelle, elle est donc diagonalisable et il existe une base $\nu_1,\dots,\nu_d$ base de vecteurs propres orthogonale. $S = \sum_{k=1}^{d} \lambda_k \nu_k \nu_k^T$ et donc $\varepsilon^T S\varepsilon = \sum_{k=1}^{d} \lambda_k (\nu_k^T \varepsilon)^2$. La matrice des vecteurs propres $V = [\nu_1|\dots|\nu_d]$ est orthogonale, donc $V^T\varepsilon \sim \mathcal{N}(0,I_d)$ et ainsi les variables aléatoires $Z_k := \nu_k^T\varepsilon$ sont indépendantes et suivent une $\mathcal{N}(0,1)$. Maintenant, si $|s| \leqslant \frac{1}{4}$, avec des calculs fait précédemment, on a $\log \mathbb{E}[\exp(s(Z_k^2-1)] \underset{|s|\leqslant 0.5}{\leqslant} \frac{s^2}{1-2s} \underset{|s|\leqslant 1/4}{\leqslant} 2s^2$. On applique désormais la méthode de Cramér-Chernoff. Soit $t>0$ et $s$ tel que $|s| \leqslant \frac{1}{4|S|_{0p}}$

$$\begin{aligned}
\mathbb{P}(\varepsilon^T S\varepsilon - Tr(S) > t) &\overset{Markov}{\leqslant} e^{-st}\mathbb{E}\left[e^{s(\varepsilon^T S\varepsilon - Tr(S))}\right] \\
&\overset{\perp\!\!\!\perp}{=} e^{-st}\prod_{k=1}^{d}\mathbb{E}\left[\exp(s\lambda_k(Z_k^2-1))\right]. \\
&\overset{|s|\leqslant 1/4|S|_{Op}}{\leqslant} \exp\left(-st + 2s^2\|S\|_F^2\right)
\end{aligned}$$

Maintenant, la fonction $s \mapsto -st + 2s^2\|S\|_F^2$ est minimum sur $|s| \leqslant 1/4|S|_{Op}$ en $s^* = \frac{1}{4}(\frac{t}{\|S\|_F^2}) \wedge (\frac{1}{|S|_{Op}})$. Ainsi :

$$\begin{aligned}
\min_{|s|\leqslant(4|S|)^{-1}}(-st + 2\|S\|_F^2) &= \frac{-t^2}{8\|S\|_F^2}\mathbb{1}_{t|S|_{Op}\leqslant\|S\|_F^2} + \left(\frac{\|S\|_F^2}{8|S|^2} - \frac{1}{4|S|_{Op}}\right)\mathbb{1}_{t|S|_{Op}>\|S\|_F^2} \\
&\leqslant -\frac{1}{8}\left(\frac{t^2}{\|S\|_F^2} \wedge \frac{t}{|S|_{Op}}\right)
\end{aligned}$$

□

COROLLAIRE 45. Si $\varepsilon$,$\varepsilon'$ sont indépendant et de loi $\mathcal{N}(0,I_d)$ et $A$ matrice réelle, alors $\forall L \geqslant 0$,

$$\mathbb{P}(\varepsilon^T A\varepsilon' > \sqrt{4\|A\|_F^2 L} \wedge 4|A|_{0p} L) \leqslant \exp(-L)$$

*Preuve* On remarque que $\varepsilon^T A\varepsilon' = \begin{bmatrix}\varepsilon\\ \varepsilon'\end{bmatrix}^T S \begin{bmatrix}\varepsilon\\ \varepsilon'\end{bmatrix}$ où la matrice $S := \frac{1}{2}\begin{bmatrix}0 & A\\ A^T & 0\end{bmatrix}$ est symétrique réelle. On utilise enfin le lemme précédent et le fait que $|S|_{Op} = |A|_{Op}/2$ et $\|S\|_F^2 = \|A\|_F^2/2$. □

LEMME 46. **Inégalité de Chernoff**
Soit $(X_i)_{i=1}^n$ une suite de variables *iid* de loi $\mathcal{B}(p)$. On note $X = \sum_{i=1}^n X_i$ et $\mu = \mathbb{E}[X]$. Alors :

1. pour $\delta > 0$, $\mathbb{P}(X > (1+\delta)\mu) \leqslant \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu$
2. pour $\delta > 0$, $\mathbb{P}(X < (1-\delta)\mu) \leqslant \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^\mu$
3. il existe une constante $c > 0.7$ tel que pour $\delta > 1$, $\mathbb{P}(X > (1+\delta)\mu) \leqslant \exp(-\mu\delta\log(\delta)c)$
4. pour $0 < \delta < 1$, $\mathbb{P}(X < (1-\delta)\mu) \leqslant \exp\left(-\frac{\delta^2\mu}{2}\right)$

*Preuve* 1. On utilise la méthode de Crémér-Chernoff. Soit $t > 0$,

$$\begin{aligned}\mathbb{P}(X > (1+\delta)\mu) &\leqslant \frac{\mathbb{E}[\exp(Xt)]}{\exp((1+\delta)\mu t)} = \frac{\prod_{i=1}^n \mathbb{E}[\exp(X_i t)]}{\exp((1+\delta)\mu t)} = \frac{\prod_{i=1}^n (1+p_i(\exp(t)-1))}{\exp((1+\delta)\mu t)}\\ &\leqslant \frac{\prod_{i=1}^n \exp(p_i(\exp(t)-1))}{\exp((1+\delta)\mu t)} = \frac{\exp(\mu(\exp(t)-1))}{\exp((1+\delta)\mu t)} = \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu\end{aligned}$$

On a utilisé Markov à la première inégalité, puis l'inégalité $1 + x \leqslant \exp(x)$ valable pour $x > 0$ et la dernière égalité est obtenue en choisissant $t = \log(1+\delta)$.

2. Idem avec $-X$.

3. On écrit $\left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu = \exp\left(-\mu\delta\log(\delta)h(\delta)\right)$ où $h(\delta) = (1+\frac{1}{\delta})\frac{\log(1+\delta)}{\log(\delta)} - \frac{1}{\log(\delta)}$ et on montre que $h(\delta) \geqslant 0.7$ pour $\delta > 1$.

4. $\mathbb{P}(X \leqslant (1-\delta)\mu) \leqslant \exp(-\mu((1-\delta)\log(1-\delta)+\delta))$ et on remarque (par étude de fonction) que $(1-\delta)\log(1-\delta)+\delta \geqslant \frac{\delta^2}{2}$ sur (0,1).

□

## 5.2 Complément théorie de l'information

On a travailler sur la notion d'entropie relative, la notion d'entropie, bien que sans rapport avec les bandit est aussi intéressante.

Définition 47. **Entropie**
Si $P$ est une mesure sur $[|1;n|]$, l'entropie de $P$ est définit par

$$H(P) = \sum_{i\in[N]:p_i>0} p_i \log(1/p_i)$$

L'entropie est une grandeur fondamentale en théorie de l'information, avant de pouvoir expliquer en quoi, on s'intéresse au problème de codage optimale.

On considère $P$ une mesure de probabilité sur $[N]$ muni de la tribu naturelle et $X$ une variable aléatoire à valeur dans $[N]$ et de densité $P$. On cherche à transmettre la réalisation de cette variable aléatoire $X$ avec un code binaire en minimisant le nombre de bits à transmettre.

Définition 48. Un **code** est une fonction $c : \mathbb{N}^* \mapsto \{0,1\}^*$. On dit que $c$ est décodable de manière unique si $i_1,\dots,i_n \mapsto c(i_1)\cdots c(i_n)$ est injective (où la partie de droite est une concaténation des codes). Ce qui assure qu'une chaîne de caractère est décodable d'une unique manière en une suite de symboles.

Ici, on suppose une condition suffisante à cette unique décodabilité.

Définition 49. Un code est **prefixe** si aucun code n'est préfixe d'un autre code.

**Problème :** un objectif naturel est alors de chercher un code qui minimise la longueur de code espéré. (l'argmin est pris parmis les codes valides et $N$ est le nombre de symboles)

$$c^* = argmin_c \sum_{i=1}^{N} p_i |c(i)|$$

.

Théorème 50.

$$H_2(P) \leqslant \sum_{i=1}^{N} p_i |c^*(i)| \leqslant H_2(P) + 1$$

où $H_2(P)$ est l'entropie (en base 2) :

$$H(P) = \sum_{i\in[N]:p_i>0} p_i \log_2(1/p_i)$$

*Preuve* Au moins deux codes atteignant cette borne ont été énoncés, le code d'Huffman et de Shanonn. Exemple, avec la fréquence d'apparition des lettres dans l'alphabet français, le code de Huffman code $e$ en 2 bit et $w$ en 14 bits. □

◇ Remarque. le code naturel qui consiste à représenter en base 2 les entiers avec des mots de longueur $\lceil \log_2(N) \rceil$ n'est pas toujours optimal. Le théorème précédent assure qu'il est optimale si $P$ est uniforme mais si par exemple un symbole apparaît avec une très grande probabilité, on aura intérêt à représenter ce symbole avec un code court.

◇ REMARQUE. Si on utilise le code optimal pour la distribution $P$ sur un message $X$ qui est pourtant réparti selon $Q$, l'entropie relative entre $P$ et $Q$ est la longueur supplémentaire que l'on attend pour le message.

### 5.3 Complément : modèle de mélange Gaussien

Modèle de mélange Gaussien : on suppose que l'on observe $(Y_1,\dots,Y_n)$ variables aléatoires iid à valeur dans $\{1,\dots,K\}$ et de lois commune $\pi$. $\mathbb{P}(Y = y) = \pi_y \ \forall y \in \{1,\dots,K\}$ et $\sum_{y=1}^K \pi_y = 1$. On a une partition (aléatoire) $G^* = \{G_1^*,\dots,G_K^*\}$ où $G_k^* = \{i : Z_i = k\}$. Conditionnellement à $G^*$, les variables $X_1,\dots,X_n$ sont indépendantes et sont distribuées selon un modèle de sous-population Gausienne. $\forall i$, la loi de $X_i$ sachant $Y_i = k$ est $\mathcal{N}(\theta_k,\sigma_k^2 I_d)$

LEMME 51. La loi marginale de $X_i$ est $g = \sum_{i=1}^K \pi_i g_{\theta_i,\Sigma_i}$.

Par la formule des probabilités totales $\mathrm{d}\mathbb{P}(X = x) = \sum_{y=1}^K \mathrm{d}\mathbb{P}(X = x|Y = y)\mathbb{P}(Y = y) = g$. On cherche la loi de $Y$ sachant $X$. D'après la formule de Bayes, on a $\mathbb{P}(Y = y|X = x) = \frac{\mathrm{d}\mathbb{P}(X=x|Y=y)\mathbb{P}(y=y)}{\sum_{i=1}^K \mathrm{d}\mathbb{P}(X=x|Y=i)\mathbb{P}(Z=i)} = \frac{g_{\theta_y,\Sigma_y}(x)\pi_y}{\sum_{i=1}^K g_{\theta_i,\Sigma_i}(x)\pi_i}$.

LEMME 52.

$$\mathbb{P}(Y = y|X = x) = \frac{g_{\theta_y,\Sigma_y}(x)\pi_y}{\sum_{i=1}^K g_{\theta_i,\Sigma_i}(x)\pi_i}$$

Le classifieur de Bayes est donné par $t(x) = \mathrm{argmax}_{y=1}^K \mathbb{P}(Y = y|X = x) = \mathrm{argmax}_{y=1}^K \pi_y g_{\theta_y,\Sigma_y}(x)\pi_y$. C'est le classifieur qui minimise le risque d'erreur. Pour un classifieur $f$, $R(f) = \mathbb{P}(Y \neq f(X))$ et on peut montrer que $R(t) \leqslant R(f)$. Problème : on ne connaît pas $w := (\pi_1,\dots,\pi_M,\theta_1,\dots,\theta_M,\Sigma_1,\dots,\Sigma_M)$. On observe les données d'apprentissage $X,Y = (X_1,Y_1),\dots,(X_n,Y_n)$ et la variable aléatoire $X_{new}$ et on cherche à classer $X_{new}$ en inférant la valeur de $Y_{new}$ inconnue. Calculons l'estimateur de maximum de vraissemblance de $w$ pour l'observation $X,Y$.

LEMME 53. La vraissemblance est :

$$L_{X,Y}(w) = \prod_{k=1}^K \sum_{i:Y_i=k} \pi_k g_{\theta_k,\Sigma_k}(X_i)$$

LEMME 54. L'estimateur de maximum de vraissemblance de $w$ est :

$\hat{\pi}_k = \frac{N_k}{n}$ où $N_k = \#\{i : Z_i = k\}$

$\hat{\theta}_k = \frac{1}{N_k} \sum_{i:Z_i=k} X_i$

$\hat{\sigma^2}_k = \frac{1}{pN_k} \sum_{i:Z_i=k} ||X_i - \hat{\theta}_k||^2$

Finalement, on donne un classifieur par : $\hat{t}(x) = \text{argmax}_{y=1}^{K} \text{argmax}_{y=1}^{K} \hat{\pi}_y g_{\hat{\theta}_y, \hat{\Sigma}_y}(x) \pi_y$

# Bibliographie


[1] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2) :235–256, 2002.

[2] S. Bubeck and N. Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *arXiv preprint arXiv :1204.5721*, 2012.

[3] C. Giraud and N. Verzelen. Partial recovery bounds for clustering with the relaxed $k$-means. *Mathematical Statistics and Learning*, 1(3) :317–374, 2019.

[4] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning : data mining, inference, and prediction*. Springer Science & Business Media, 2009.

[5] T. Lattimore and C. Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

[6] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics*, pages 1302–1338, 2000.

[7] W. R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4) :285–294, 1933.