Le contrôle sous-gaussien sera utile ensuite. Mentionnons une conséquence :

Si $Z \leq 0$ p.s. et $\mathbb{E}[Z^2] = v$, alors

(*) $\forall \lambda > 0$, $\ln \mathbb{E}\left[e^{\lambda(Z - \mathbb{E}Z)}\right] \leq \frac{\lambda^2 v}{2}$ (passage à la limite $b \to 0$).

Application : majoration de δ_t de type Bennett

$$\delta_t = \frac{1}{\eta_t} \ln\left(\sum_{i=1}^K p_{i,t} e^{-\eta_t (l_{i,t} - p_t \cdot l_t)} \right) = \frac{1}{\eta_t} \ln \mathbb{E}_{i \sim p_t}\left[e^{-\eta_t (l_{i,t} - \mathbb{E}_{j \sim p_t} l_{j,t})} \right]$$

d'après (*) avec $\begin{cases} Z = -l_{i,t} \\ i \sim p_t \end{cases}$ $\leq \frac{\eta_t}{2} \sum_{i=1}^K p_{i,t} l_{i,t}^2$ car $l_{i,t} \geq 0$.

II / Information imparfaite : les bandits antagonistes

1) Formulation du problème

⚠ Désormais, le statisticien n'observe plus $l_{i,t}$ pour $i \neq I_t$.

Protocole de décision (bandits antagonistes à K bras) : pour chaque $t \in \mathbb{N}^*$,

1) Le statisticien choisit et révèle $p_t \in \mathcal{M}_1^+(\{1, \dots, K\})$ en fonction des données disponibles $(l_{I_s,s}, I_s)_{s \leq t-1}$.

2) Simultanément :
- Le statisticien tire $I_t \sim p_t$ (conditionnellement au passé)
- L'environnement choisit $l_t = (l_{i,t})_{1 \leq i \leq K} \in [0,1]^K$ en fonction des données disponibles $(I_s, p_s)_{s \leq t-1}$ et même p_t (mais pas I_t)

3) Le statisticien encourt et observe la perte $l_{I_t,t}$ (les $l_{j,t}$, $j \neq I_t$, restent cachés) ; l'environnement observe I_t.

Objectif : minimiser le regret $R_T = \sum_{t=1}^T l_{I_t,t} - \min_{1 \leq i \leq K} \sum_{t=1}^T l_{i,t}$

Plus précisément, on cherche des stratégies (choix séquentiels des p_t) telles que, pour toute suite de fonctions $l_t(\cdot)$ (i.e. pour tout adversaire),

- $\mathbb{E}[R_T] \leq o(T)$ (on obtiendra $\mathbb{E}[R_T] \leq \square \sqrt{TK \ln K}$)
- p.s., $\overline{\lim}_{T \to +\infty} \frac{R_T}{T} \leq 0$ (on obtiendra $\overline{\lim}_{T \to +\infty} \frac{R_T}{\sqrt{T \ln \ln T}} \leq 0$)

Dans un premier temps, on va s'intéresser à une quantité plus faible :

$$\mathbb{E}\left[\sum_{t=1}^T l_{I_t,t}\right] - \min_{1 \leq i \leq K} \mathbb{E}\left[\sum_{t=1}^T l_{i,t}\right] \leftarrow \text{le "pseudo-regret".}$$

2) Estimation des pertes et algorithme associé

On estime $\underline{\ell}_t$ par $\underline{\tilde{\ell}}_t$ défini par $\tilde{\ell}_{i,t} = \frac{\ell_{I_t,t}}{p_{i,t}} \mathbb{1}_{\{I_t = i\}}$

↑ ne dépend que des données observées jusqu'à la fin du tour t.

On a $\tilde{\ell}_{i,t} = \frac{\ell_{i,t}}{p_{i,t}} \mathbb{1}_{\{I_t=i\}}$ d'où $\mathbb{E}\left[\tilde{\ell}_{i,t} \,|\, \mathcal{F}_{t-1}\right] = \frac{\ell_{i,t}}{p_{i,t}} \underbrace{\mathbb{P}(I_t = i \,|\, \mathcal{F}_{t-1})}_{= p_{i,t}} = \ell_{i,t}$

Algorithme :

$$p_{i,t} = \frac{\exp\left(-\eta_t \sum_{s=1}^{t-1} \tilde{\ell}_{i,s}\right)}{\sum_{j=1}^{K} \exp\left(-\eta_t \sum_{s=1}^{t-1} \tilde{\ell}_{j,s}\right)}, \quad 1 \le i \le K \qquad \left(\text{N.B. } p_1 = \left(\tfrac{1}{K}, \dots, \tfrac{1}{K}\right)\right)$$

3) Majoration du pseudo-regret

Théorème : Calibré avec $\eta_t = \sqrt{\frac{\ln K}{tK}}$, l'algorithme Exp3 vérifie

$$\mathbb{E}\left[\sum_{t=1}^{T} \ell_{I_t,t}\right] - \min_{1 \le i \le K} \left[\sum_{t=1}^{T} \ell_{i,t}\right] \le (1+\sqrt{2}) \cdot \sqrt{TK \ln K}$$

[N.B. pertes $\ell_{i,t}$ dans $[0,1]$.]

Preuve : d'après prop 2 et la majoration de δ_t de type Bennett (page 1), on a :

$$(*) \quad \sum_{t=1}^{T} p_t \cdot \tilde{\ell}_t \le \min_{1 \le i \le K} \sum_{t=1}^{T} \tilde{\ell}_{i,t} + \frac{\ln K}{\eta_{T+1}} + \sum_{t=1}^{T} \frac{\eta_t}{2} \underbrace{\sum_{i=1}^{K} p_{i,t} \tilde{\ell}_{i,t}^2}_{= \ell^2_{I_t,t} / p_{I_t,t}} \qquad \text{P.S.}$$

En prenant l'espérance à gauche et à droite de l'inégalité, et en remarquant que $\mathbb{E}\left[\tilde{\ell}_{i,t}\right] = \mathbb{E}\left[\mathbb{E}\left[\tilde{\ell}_{i,t} \,|\, \mathcal{F}_{t-1}\right]\right] = \mathbb{E}\left[\ell_{i,t}\right]$ ainsi que $\mathbb{E}\left[p_t \cdot \underbrace{\tilde{\ell}_t}_{\mathcal{F}_{t-1}\text{-mesurable}}\right] = \mathbb{E}\left[p_t \cdot \ell_t\right]$, il vient :

$$\mathbb{E}\left[\sum_{t=1}^{T} p_t \cdot \ell_t\right] \overset{\text{Jensen}}{\le} \min_{1 \le i \le K} \mathbb{E}\left[\sum_{t=1}^{T} \ell_{i,t}\right] + \frac{\ln K}{\eta_{T+1}} + \sum_{t=1}^{T} \frac{\eta_t}{2} \mathbb{E}\left[\frac{\ell^2_{I_t,t}}{p_{I_t,t}}\right]$$

En remarquant que $\mathbb{E}\left[\frac{\ell^2_{I_t,t}}{p_{I_t,t}}\right] \overset{\ell_{i,t} \in [0,1]}{\le} \mathbb{E}\left[\frac{1}{p_{I_t,t}}\right] = \mathbb{E}\left[\sum_{i=1}^{K} p_{i,t} \frac{1}{p_{i,t}}\right] = K$

et en utilisant le choix $\eta_t = \sqrt{\frac{\ln K}{tK}}$, on obtient finalement :

$$\mathbb{E}\left[\sum_{t=1}^{T} \ell_{I_t,t}\right] - \min_{1\le i\le K} \mathbb{E}\left[\sum_{t=1}^{T} \ell_{i,t}\right] \le \sqrt{(T+1)K\ln K} + \frac{K}{2}\sum_{t=1}^{T}\sqrt{\frac{\ln K}{tK}}$$

$$\le (1+\sqrt{2})\sqrt{TK\ln K} \quad \text{car} \quad \sum_{t=1}^{T}\frac{1}{\sqrt{t}} \le 2\sqrt{T}. \quad \blacksquare$$

Remarque sur les vitesses :

- Information parfaite : $\mathbb{E}[R_T] \le \sqrt{\frac{T}{2}\ln K}$
- Bandits : pseudo-regret $\le (1+\sqrt{2})\sqrt{TK\ln K}$

Même vitesse en T ($\sqrt{T}$) mais complexité supérieure en K pour les bandits.
C'est naturel : en divisant par T, on a les majorations suivantes sur le (pseudo) regret :

- Info parfaite : $\square\sqrt{\frac{\ln K}{T}}$
- Bandits : $\square\sqrt{\frac{\ln K}{(T/K)}}$ ← K fois moins d'observations que dans le cas "info parfaite".

Autre remarque : pseudo-regret = $\mathbb{E}[\text{regret}]$ pour un adversaire aveugle

- Déf : un adversaire $(\underline{\ell}_t(\cdot))_{t\ge 1}$ est dit "aveugle" ("oblivious") si les fonctions $\underline{\ell}_t$ sont constantes : $\forall t\ge 1, \underbrace{\forall a,b, \underline{\ell}_t(a) = \underline{\ell}_t(b)}$
  pertes $\ell_{i,t}$ déterministes
  (l'adversaire ne se sert pas du passé)

- Pour un adversaire aveugle, on a :

$$\underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\ell_{I_t,t}\right] - \min_{1\le i\le K}\mathbb{E}\left[\sum_{t=1}^{T}\ell_{i,t}\right]}_{\text{pseudo-regret}} = \mathbb{E}\left[\underbrace{\sum_{t=1}^{T}\ell_{I_t,t} - \min_{1\le i\le K}\sum_{t=1}^{T}\ell_{i,t}}_{\text{regret}}\right]$$

($\min_{1\le i\le K}\sum_{t=1}^{T}\ell_{i,t}$ déterministe)

Le thm précédent fournit donc aussi une majoration du regret pour des adversaires aveugles, mais pas pour des adversaires antagonistes.

4) Amélioration : majoration du regret avec grande proba

On donne seulement quelques pistes.
On considère un adversaire quelconque (potentiellement antagoniste).

On souhaite obtenir des bornes de la forme :

$$\sum_{t=1}^{T} \ell_{I_t,t} \leq \min_{1\leq i\leq K} \sum_{t=1}^{T} \ell_{i,t} + \Delta_{T,K}(\delta) \quad \text{avec probabilité} \geq 1-\delta.$$

Puisque le choix précédent de $\tilde{\ell}_{i,t} = \frac{\ell_{i,t}}{p_{i,t}} \mathbb{1}_{\{I_t=i\}}$ conduit à $p_t \cdot \tilde{\ell}_t = \sum_{i=1}^{K} p_{i,t} \tilde{\ell}_{i,t} = \ell_{I_t,t}$, on pourrait vouloir partir de l'inégalité (*) page 6 et la combiner avec une majoration de la forme : $\min_{1\leq i\leq K} \sum_{t=1}^{T} \tilde{\ell}_{i,t} \leq \min_{1\leq i\leq K} \sum_{t=1}^{T} \ell_{i,t} + \dots$ avec gde proba

Problème : à chaque date t, la variance $\mathrm{Var}\left(\tilde{\ell}_{i,t} \mid \mathcal{F}_{t-1}\right) = \frac{\ell_{i,t}^2 (1-p_{i,t})}{p_{i,t}}$ peut être grande (de l'ordre de $1/p_{i,t}$). D'où plusieurs idées :

Solution n°1 (1ère partie du papier de Auer, Cesa-Bianchi, Freund et Schapire 2002)

On mélange l'EWA avec l'uniforme :

$$p_{i,t} = (1-\gamma) \frac{e^{-\eta \tilde{L}_{i,t-1}}}{\sum_{j=1}^{K} e^{-\eta \tilde{L}_{j,t-1}}} + \frac{\gamma}{K}, \quad 1\leq i\leq K$$

} algorithme Exp3 ("Exponential Weights for Exploration and Exploitation")

où $\tilde{L}_{i,t-1} = \sum_{s=1}^{t-1} \tilde{\ell}_{i,s}$ et $\gamma > 0$ paramètre à calibrer.

Malheureusement, les meilleurs choix de γ ne donnent qu'une vitesse

$\Delta_{T,K}(\delta) \approx T^{2/3}$ sous-optimale (pour un contrôle en grande proba).

Solution n°2 (2e partie du papier)

Algorithme Exp3.P : ("P" = "probability")

- Paramètres : $\eta > 0$ et $\beta, \gamma \in]0,1[$.
- Initialisation : $p_1 = \left(\frac{1}{K}, \dots, \frac{1}{K}\right)$
- A chaque date $t \geq 1$,
  - Tirer aléatoirement $I_t \sim p_t$ conditionnellement à $\mathcal{F}_{t-1} = \sigma(I_1, \dots, I_{t-1})$.
  - Observer $\ell_{I_t,t}$ et calculer $\tilde{\ell}_{i,t} = 1 - \frac{(1-\ell_{i,t})\mathbb{1}_{\{I_t=i\}} + \beta}{p_{i,t}}$, $i = 1, \dots, K$
  - Choisir le vecteur de poids p_t selon :

$$p_{i,t} = (1-\gamma) \frac{e^{-\eta \tilde{L}_{i,t-1}}}{\sum_{j=1}^{K} e^{-\eta \tilde{L}_{j,t-1}}} + \frac{\gamma}{K}, \quad 1\leq i\leq K.$$

Remarques :

- Les pertes ont été transformées en gains : $\ell_{i,t} \in [0,1] \rightsquigarrow g_{i,t} = 1 - \ell_{i,t} \in [0,1]$
  C'est utile pour contrôler les queues de distribution à droite des $\sum_{t=1}^{T} \tilde{\ell}_{i,t}$ (car $\tilde{\ell}_{i,t} \leq 1$

- On biaise volontairement les estimateurs des $\ell_{i,t}$:

$$\mathbb{E}\left[\tilde{\ell}_{i,t} \,|\, \mathcal{F}_{t-1}\right] = \ell_{i,t} - \frac{\beta}{p_{i,t}}$$

Cela permet de majorer uniformément en $i \in \{1,\dots,K\}$ les $\sum_{t=1}^{T} \tilde{\ell}_{i,t}$:

$$\forall i,\ \sum_{t=1}^{T} \tilde{\ell}_{i,t} \leq \sum_{t=1}^{T} \ell_{i,t} + \frac{1}{\beta} \ln\left(\frac{K}{\delta}\right),\ \text{avec proba} \geq 1-\delta$$

- On utilise toujours le mélange avec l'uniforme pour éviter que les pertes estimées $\tilde{\ell}_{i,t}$ soient trop négatives (utile pour contrôler les log-Laplace

$$\frac{1}{\eta} \ln \mathbb{E}_{i \sim w_t}\left[e^{-\eta\left(\tilde{\ell}_{i,t} - w_t \cdot \tilde{\ell}_t\right)}\right]$$

où $w_t = \dfrac{p_t - \gamma\left(\frac{1}{K},\dots,\frac{1}{K}\right)}{1-\gamma}$ = EWA).

Borne : pour des choix de la forme (avec $c_1, c_2, \dots, c_5$ constantes bien choisies)

$$\eta = c_1 \sqrt{\frac{\ln K}{TK}},\quad \beta = c_2 \sqrt{\frac{\ln K}{TK}} \quad \text{et} \quad \gamma = c_3 \sqrt{\frac{K \ln K}{T}},$$

$$\sum_{t=1}^{T} \ell_{I_t,t} \overset{\text{avec proba} \geq 1-\delta}{\leq} \min_{1 \leq i \leq K} \sum_{t=1}^{T} \ell_{i,t} + c_4 \sqrt{\frac{TK}{\ln K}} \ln\left(\frac{1}{\delta}\right) + c_5 \sqrt{TK \ln K}$$

Cf. l'article de survol de Bubeck et Cesa-Bianchi (2012) pour une preuve.

III – Bornes inférieures

1) Borne inférieure en information parfaite : $\sqrt{\frac{T}{2}\ln K}$ (pertes bornées $0 \leq \ell_{i,t} \leq 1$)

On va montrer que la borne supérieure $\sqrt{(T/2)\ln K}$ obtenue sur le regret de l'EWA en information parfaite est minimax-optimale quand $T, K \to +\infty$ au sens suivant :

Proposition : On pose $R_{K,T}^{\text{minimax}} = \inf_{(p_t(\cdot))_{t\geq 1}} \sup_{(\ell_t(\cdot))_{t\geq 1}} \mathbb{E}\left[\sum_{t=1}^{T} \ell_{I_t,t} - \min_{1\leq i\leq K} \sum_{t=1}^{T} \ell_{i,t}\right]$

Alors :

$$\lim_{K\to+\infty} \lim_{T\to+\infty} \frac{R_{K,T}^{\text{minimax}}}{\sqrt{\frac{T}{2}\ln K}} \geq 1.$$

N.B. : L'inf est pris sur toutes les stratégies (i.e. choix séq. des p_t en fonction du passé). Le sup s'étend sur tous les adversaires (i.e. choix séq des ℓ_t) tels que $0 \leq \ell_{i,t} \leq 1$.

Preuve (partielle) :

On se ramène d'abord à des pertes déterministes puis stochastiques i.i.d. :

$$\inf_{(p_t(\cdot))_{t\geq 1}} \sup_{(\ell_t(\cdot))_{t\geq 1}} \mathbb{E}_{\mathbb{P}_A}\left[\text{regret}(T)\right] \geq \inf_{(p_t(\cdot))_{t\geq 1}} \sup_{\underline{\ell}_1,\dots,\underline{\ell}_T \in [0,1]^K} \mathbb{E}_{\mathbb{P}_A}\left[\text{regret}(T)\right]$$

- $(\ell_t(\cdot))_{t\geq 1}$: adversaires potentiellement antagonistes
- $\mathbb{E}_{\mathbb{P}_A}$: espérance par rapport à la randomisation de l'algorithme.
- $\underline{\ell}_1,\dots,\underline{\ell}_T$: suites déterministes ($\underline{\ell}_t$ ne dépend pas des observations passées)

$$\geq \inf_{(p_t(\cdot))_{t\geq 1}} \int_{[0,1]^{KT}} \mathbb{E}_{\mathbb{P}_A}\left[\text{regret}(T)\right] d\mathbb{Q}(\underline{\ell}_1,\dots,\underline{\ell}_T)$$

pour toute proba $\mathbb{Q}$ sur $[0,1]^{KT}$.

$$= \inf_{(p_t(\cdot))_{t\geq 1}} \mathbb{E}_{\mathbb{P}_A\otimes\mathbb{Q}}\left[\text{regret}(T)\right] \text{ par Fubini.}$$

On choisit $\mathbb{Q}$ telle que $(\ell_{i,t})_{\substack{1\leq i\leq K\\ 1\leq t\leq T}} \overset{\text{iid}}{\sim}_{\mathbb{Q}} \mathcal{B}(1/2)$ $\left(\text{i.e. } \mathbb{Q} = \mathcal{B}(1/2)^{\otimes TK}\right)$.

Fixons une stratégie $(p_t(\cdot))_{t\geq 1}$

et remarquons :

$$\mathbb{E}_{\mathbb{P}_A\otimes\mathbb{Q}}\left[\text{regret}(T)\right] = \mathbb{E}_{\mathbb{P}_A\otimes\mathbb{Q}}\left[\sum_{t=1}^T \ell_{I_t,t} - \min_{1\leq i\leq K}\sum_{t=1}^T \ell_{i,t}\right]$$

$$= \frac{T}{2} - \mathbb{E}_{\mathbb{P}_A\otimes\mathbb{Q}}\left[\min_{1\leq i\leq K}\sum_{t=1}^T \ell_{i,t}\right] \quad \text{car } \mathbb{E}\left[\ell_{I_t,t}\right] = \frac{1}{2}$$

$$= \frac{1}{2}\mathbb{E}_{\mathbb{Q}}\left[\max_{1\leq i\leq K}\sum_{t=1}^T \underbrace{(1-2\ell_{i,t})}_{=:\varepsilon_{i,t}}\right]$$

N.B. $\varepsilon_{i,t}$ Rademacher iid ($\varepsilon_{i,t} = \pm 1$ avec proba $1/2$).

Donc $$\frac{\mathbb{E}_{\mathbb{P}_A\otimes\mathbb{Q}}\left[\text{regret}(T)\right]}{\sqrt{\frac{T}{2}\ln K}} = \frac{1}{\sqrt{2\ln K}}\underbrace{\mathbb{E}_{\mathbb{Q}}\left[\max_{1\leq i\leq K}\frac{1}{\sqrt{T}}\sum_{t=1}^T \varepsilon_{i,t}\right]}_{\xrightarrow[T\to+\infty]{} \mathbb{E}\left[\max_{1\leq i\leq K} Z_i\right] \text{ avec } Z_1,\dots,Z_K \overset{\text{iid}}{\sim} \mathcal{N}(0,1)}$$

car (essentiellement) $\frac{1}{\sqrt{T}}\sum_{t=1}^T \underline{\varepsilon}_t \xrightarrow[T\to+\infty]{\mathcal{L}} (Z_1,\dots,Z_K)$

Donc $$\varliminf_{K\to+\infty}\ \varliminf_{T\to+\infty} \frac{\mathbb{E}_{\mathbb{P}_A\otimes\mathbb{Q}}\left[\text{regret}(T)\right]}{\sqrt{\frac{T}{2}\ln K}} \geq \varliminf_{K\to+\infty} \frac{\mathbb{E}\left[\max_{1\leq i\leq K} Z_i\right]}{\sqrt{2\ln K}}$$

$$= 1.$$

D'où le résultat. ∎

Rem 1 : Contre toute attente, la stratégie $(\ell_t(\cdot))_{t\geq 1}$ de l'adversaire exibée dans la preuve ↑ est aveugle et même iid ! Le terme $\sqrt{T\ln K}$ provient de la différence entre $\min_{1\leq i\leq K} \mathbb{E}\left[\sum_{t=1}^T \ell_{i,t}\right]$ et $\mathbb{E}\left[\min_{1\leq i\leq K} \sum_{t=1}^T \ell_{i,t}\right]$. Dès lors, minimiser le regret (externe) en information parfaite est aussi simple face à un adversaire antagoniste que face à une suite iid.

Rem 2 : On peut aussi prouver une borne inférieure non-asymptotique de la forme : $\forall K\geq 1$, $\forall T \geq \square \ln K$, $R^{\text{minimax}}_{K,T} \geq \square\sqrt{T\ln K}$ [preuve via le lemme de Fano par ex].

2) Borne inférieure dans le cas bandits : $\sqrt{TK}$

Le problème de bandits est plus difficile qu'en information parfaite. On va en effet prouver une borne inférieure plus grande (de l'ordre de $\sqrt{TK}$ contre $\sqrt{T\ln K}$ tout à l'heure). On retrouve à un terme $\sqrt{\ln K}$ près la borne sup $\sqrt{TK\ln K}$ sur le (pseudo-)regret de l'EWA avec pertes estimées $\tilde{\ell}_{i,t}$.

Proposition : Pour tous $K\geq 2$ et $T\geq K/2$,

$$\inf_{(p_t(\cdot))_{t\geq 1}} \sup_{(\ell_t(\cdot))_{t\geq 1}} \mathbb{E}\left[\sum_{t=1}^T \ell_{I_t,t} - \min_{1\leq i\leq K} \sum_{t=1}^T \ell_{i,t}\right]$$

$$\geq \inf_{(p_t(\cdot))_{t\geq 1}} \sup_{\ell_1,\dots,\ell_T\in[0,1]^K} \left\{\mathbb{E}\left[\sum_{t=1}^T \ell_{I_t,t}\right] - \min_{1\leq i\leq K} \sum_{t=1}^T \ell_{i,t}\right\}$$

$$\geq \frac{1}{2\sqrt{12}}\sqrt{TK}$$

où le $\sup_{(\ell_t(\cdot))_{t\geq 1}}$ est pris sur tous les adversaires t.q. $0\leq \ell_{i,t}\leq 1$.

Preuve : Remarquons tout d'abord qu'on peut toujours supposer que

$\forall t\geq 1$, $I_t = \psi(p_t, U_t)$, où $U_1,\dots,U_T \overset{\text{iid}}{\sim} \mathcal{U}([0,1])$, sans changer la loi de la suite $(I_1,\dots,I_T)$.

(Il suffit par ex de définir $I_t = i$ si $U_t \in \left[\sum_{j=1}^{i-1} p_{j,t}, \sum_{j=1}^{i} p_{j,t}\right[$)

Ramenons-nous comme précédemment à des pertes déterministes puis stochastiques :

$$\inf_{(p_t(\cdot))_{t\geq 1}} \sup_{(\ell_t(\cdot))_{t\geq 1}} \mathbb{E}_{\mathbb{P}_A}\left[\text{regret}(T)\right] \geq \inf_{(p_t(\cdot))_{t\geq 1}} \sup_{\ell_1,\dots,\ell_T \in [0,1]^K} \mathbb{E}_{\mathbb{P}_A}\left[\text{regret}(T)\right]$$

$$\geq \inf_{(p_t(\cdot))_{t\geq 1}} \int_{[0,1]^{KT}} \mathbb{E}_{\mathbb{P}_A}\left[\text{regret}(T)\right] d\mathbb{Q}^*(\ell_1,\dots,\ell_T)$$

$\mathbb{Q}^*$ proba sur $[0,1]^{KT}$

$$\overset{\text{Fubini}}{=} \inf_{(p_t(\cdot))_{t\geq 1}} \int \mathbb{E}_{\mathbb{Q}^*}\left[\text{regret}(T)\right] d\mathbb{P}_A(u_1,\dots,u_T)$$

La randomisation du statisticien provient de la suite $(U_1,\dots,U_T) \sim \mathcal{U}([0,1])^{\otimes T} = \mathbb{P}_A$

$$\geq \inf_{(p_t(\cdot))_{t\geq 1}} \inf_{u_1,\dots,u_T \in [0,1]} \mathbb{E}_{\mathbb{Q}^*}\left[\sum_{t=1}^T \ell_{I_t,t} - \min_{1\leq i\leq K} \sum_{t=1}^T \ell_{i,t}\right]$$

$I_t = \psi(p_t, u_t)$ est uniquement fonction du passé (via p_t) si $(u_1,\dots,u_T)$ est fixé.

$$\geq \inf_{(I_t(\cdot))_{t\geq 1} \text{ prévisibles}} \mathbb{E}_{\mathbb{Q}^*}\left[\sum_{t=1}^T \ell_{I_t,t} - \min_{1\leq i\leq K} \sum_{t=1}^T \ell_{i,t}\right]$$

I_t fonction mesurable du passé $(\ell_{I_s,s})_{s\leq t-1}$

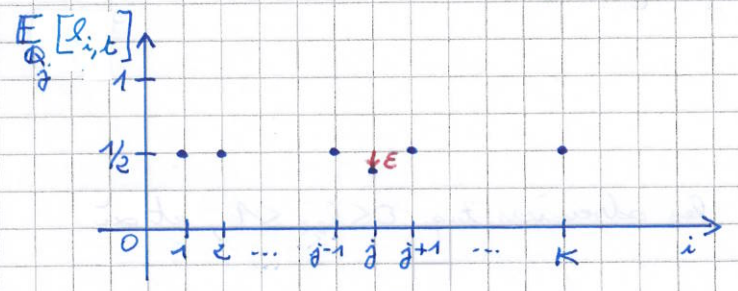
On s'est donc ramenés à des algorithmes déterministes. Soit $(I_t(\cdot))_{t\geq 1}$ une telle suite de fonctions.

On va prendre $\mathbb{Q}^* = \frac{1}{K}\sum_{j=1}^K \mathbb{Q}_j$ où $\mathbb{Q}_j \in \mathcal{M}_1^+([0,1]^{KT})$ est telle que :

sous $\mathbb{Q}_j$, • $\ell_{i,t}$ indépendants, $1\leq i\leq K$, $1\leq t\leq T$

• $\ell_{i,t} \sim \begin{cases} \mathcal{B}\left(\frac{1}{2}\right) & \text{si } i\neq j \\ \mathcal{B}\left(\frac{1}{2}-\varepsilon\right) & \text{si } i=j \end{cases}$, avec $\varepsilon>0$ à choisir ultérieurement.

(en d'autres termes : $\mathbb{Q}_j = \left(\bigotimes_{i=1}^K \mathcal{B}\left(\frac{1}{2}-\varepsilon \mathbb{1}_{i=j}\right)\right)^{\otimes T}$)



Sous $\mathbb{Q}_j$, le bras j est optimal et meilleur que les autres à ε près. On a :

$$\mathbb{E}_{\mathbb{Q}_j}\left[\sum_{t=1}^T \ell_{I_t,t}\right] = \sum_{t=1}^T \sum_{i=1}^K \mathbb{E}_{\mathbb{Q}_j}\left[\ell_{i,t}\mathbb{1}_{I_t=i}\right] = \sum_{t=1}^T \sum_{i=1}^K \mathbb{E}_{\mathbb{Q}_j}\left[\left(\frac{1}{2}-\varepsilon\mathbb{1}_{i=j}\right)\mathbb{1}_{I_t=i}\right]$$

conditionnement / $(\ell_s)_{s\leq t-1}$ car I_t prévisible.

D'où $\mathbb{E}_{\mathbb{Q}_j}\left[\sum_{t=1}^{T} \ell_{I_t,t}\right] = \frac{T}{2} - \varepsilon \sum_{t=1}^{T} \mathbb{Q}_j[I_t = j]$

Par ailleurs,

$$\mathbb{E}_{\mathbb{Q}_j}\left[\min_{1\le i\le K} \sum_{t=1}^{T} \ell_{i,t}\right] \le \min_{1\le i\le K} \mathbb{E}_{\mathbb{Q}_j}\left[\sum_{t=1}^{T} \ell_{i,t}\right] = \frac{T}{2} - T\varepsilon$$

Par conséquent,

$$\mathbb{E}_{\mathbb{Q}_j}\left[\sum_{t=1}^{T} \ell_{I_t,t} - \min_{1\le i\le K} \sum_{t=1}^{T} \ell_{i,t}\right] \ge \frac{T}{2} - \varepsilon \sum_{t=1}^{T} \mathbb{Q}_j(I_t = j) - \left(\frac{T}{2} - T\varepsilon\right)$$

$$= T\varepsilon\left(1 - \frac{1}{T}\sum_{t=1}^{T} \mathbb{Q}_j(I_t = j)\right)$$

D'où $\mathbb{E}_{\mathbb{Q}^*}\left[\sum_{t=1}^{T} \ell_{I_t,t} - \min_{1\le i\le K} \sum_{t=1}^{T} \ell_{i,t}\right] \ge T\varepsilon\left(1 - \frac{1}{T}\sum_{t=1}^{T}\frac{1}{K}\sum_{j=1}^{K} \mathbb{Q}_j(I_t = j)\right) \quad (*)$

Reste à choisir ε suffisamment petit pour que, à chaque date t, le statisticien ne soit pas capable (quel que soit son algo) de détecter le bras optimal j avec proba suffisamment proche de 1, en moyenne sur $j \in \{1,\dots,K\}$.

↳ problème de test d'hypothèses multiples

↳ ⚠ Il faut bien prendre en compte l'information de type "bandits".

Rappel : inégalité de Pinsker

Soit $\mathbb{P}$ et $\mathbb{Q}$ deux probabilités sur un même espace mesurable $(\Omega, \mathcal{A})$.

Alors : $\|\mathbb{P} - \mathbb{Q}\|_{TV} \le \sqrt{\frac{KL(\mathbb{P},\mathbb{Q})}{2}}$,

où $\|\mathbb{P} - \mathbb{Q}\|_{TV} := \sup_{A\in\mathcal{A}} |\mathbb{P}(A) - \mathbb{Q}(A)|$ et $KL(\mathbb{P},\mathbb{Q}) := \begin{cases} \int_\Omega d\mathbb{P} \log \frac{d\mathbb{P}}{d\mathbb{Q}} & \text{si } \mathbb{P} \ll \mathbb{Q} \\ +\infty & \text{sinon} \end{cases}$

Soit $t \in \{1,\dots,T\}$ fixé. D'après l'inégalité de Pinsker, on a, pour tout $j \in \{1,\dots,K\}$,

$\mathbb{Q}_j(I_t = j) \le \mathbb{Q}_0(I_t = j) + \sqrt{\frac{1}{2} KL\left(\mathbb{Q}_0^{I_t}, \mathbb{Q}_j^{I_t}\right)}$, où $\mathbb{Q}_0^{I_t}$ et $\mathbb{Q}_j^{I_t}$ désignent les lois de I_t sous $\mathbb{Q}_0$ et $\mathbb{Q}_j$ respectivement,

et où $\mathbb{Q}_0 = \mathcal{B}(1/2)^{\otimes TK}$ rend les $\ell_{i,t}$ iid $\mathcal{B}(1/2)$.

On en déduit :

$$\frac{1}{K}\sum_{j=1}^{K} \mathbb{Q}_j(I_t = j) \overset{\text{Jensen}}{\le} \frac{1}{K} + \sqrt{\frac{1}{2}\frac{1}{K}\sum_{j=1}^{K} KL\left(\mathbb{Q}_0^{I_t}, \mathbb{Q}_j^{I_t}\right)}$$

Majorons $KL\left(\mathbb{Q}_0^{I_t}, \mathbb{Q}_j^{I_t}\right)$:

$$KL\left(\mathbb{Q}_0^{I_t}, \mathbb{Q}_j^{I_t}\right) \leq KL\left(\mathbb{Q}_0^{(\ell_{I_s,s})_{s\leq t-1}}, \mathbb{Q}_j^{(\ell_{I_s,s})_{s\leq t-1}}\right)$$ car I_t est fonction mesurable de $(\ell_{I_s,s})_{1\leq s\leq t-1}$ ← bandits

"chain rule for relative entropy" →

$$= \sum_{s=1}^{t-1} \int d\mathbb{Q}_0^{(\ell_{I_\tau,\tau})_{\tau\leq s-1}} KL\left(\mathbb{Q}_0^{\ell_{I_s,s}\,|\,(\ell_{I_\tau,\tau})_{\tau\leq s-1}}, \mathbb{Q}_j^{\ell_{I_s,s}\,|\,(\ell_{I_\tau,\tau})_{\tau\leq s-1}}\right)$$

loi conditionnelle de $\ell_{I_s,s}$ sachant $(\ell_{I_\tau,\tau})_{\tau\leq s-1}$ sous $\mathbb{Q}_0$.

idem pour $\mathbb{Q}_j$

$$= \sum_{s=1}^{t-1} \int d\mathbb{Q}_0^{(\ell_{I_\tau,\tau})_{\tau\leq s-1}} KL\left(\mathcal{B}\left(\tfrac{1}{2}\right), \mathcal{B}\left(\tfrac{1}{2} - \varepsilon \mathbb{1}_{I_s=j}\right)\right)$$

car I_s est fonction mesurable de $(\ell_{I_\tau,\tau})_{\tau\leq s-1}$

$$\leq \sum_{s=1}^{t-1} \int d\mathbb{Q}_0^{(\ell_{I_\tau,\tau})_{\tau\leq s-1}} 6\varepsilon^2 \mathbb{1}_{I_s=j} \quad \boxed{\text{si } \varepsilon \leq 1/\sqrt{6}} \quad \text{(kl localement quadratique)}$$

$$= 6\varepsilon^2 \sum_{s=1}^{t-1} \mathbb{Q}_0\left(I_s = j\right)$$

D'où $$\frac{1}{K}\sum_{j=1}^{K} KL\left(\mathbb{Q}_0^{I_t}, \mathbb{Q}_j^{I_t}\right) \leq 6\varepsilon^2 \sum_{s=1}^{t-1} \frac{1}{K}\sum_{j=1}^{K} \mathbb{Q}_0\left(I_s = j\right) \leq \frac{6\varepsilon^2 T}{K}$$

$$\leq \frac{1}{2} \quad \text{pour } \varepsilon = \sqrt{\frac{K}{12T}} \quad \left(\underline{\text{N.B.}}\ \varepsilon \leq \frac{1}{\sqrt{6}} \text{ si } T \geq \frac{K}{2}\right)$$

En injectant cette majoration pour tout $t \in \{1, \dots, T\}$ dans (✳), il vient :

$$\mathbb{E}_{\mathbb{Q}^*}\left[\sum_{t=1}^{T} \ell_{I_t,t} - \min_{1\leq i\leq K} \sum_{t=1}^{T} \ell_{i,t}\right] \geq T\varepsilon\left(1 - \frac{1}{2}\right) \quad \text{dès que } T \geq \frac{K}{2}$$

$$= \frac{\sqrt{TK}}{2\sqrt{12}} \quad \blacksquare$$