$\forall d \in \mathcal{D},$ $$\sum_{t=1}^{n} Q(\hat{d}_t, z_t) - \sum_{t=1}^{n} Q(d, z_t) = o(n),$$

alors des hyp. d'absolue continuité de Q et de séparabilité de $\mathcal{D}$ suffisent.

Partie 2 : INFORMATION IMPARFAITE.

0. Résumé de l'épisode précédent (de la partie 1), pour le cas de l'agrégation linéaire.

Le statisticien a un nombre fini d'actions $\{1, \dots N\}$ et il est autorisé à utiliser des actions randomisées : il choisit à chaque tour $\hat{p}_t \in \mathcal{D}$ (qui est le simplexe $\mathcal{X}$ de $\mathbb{R}^N$) puis tire son action $I_t \in \{1, \dots N\}$ selon $\hat{p}_t$.

Jeu de prédiction : Pour tout $t = 1, 2, \dots$

- le statisticien choisit $\hat{p}_t \in \mathcal{X}$ (en fonction des infos des tours passés);
- le diable choisit $\underline{\ell}_t = (\ell_{jt})_{j=1,\dots N} \in \mathbb{R}^N$, vecteur des pertes (en fonction des infos des tours passés et même en fonction de $\hat{p}_t$);
- le statisticien tire $I_t \sim \hat{p}_t$ et subit la perte $\ell_{I_t, t}$;
- le statisticien et le diable observent I_t et $\underline{\ell}_t$.

But : Avoir un regret (inconditionnel) faible :

$$R_n = \sum_{t=1}^{n} \ell_{I_t, t} - \min_{j=1,\dots N} \sum_{t=1}^{n} \ell_{jt}$$
$$= \sum_{t=1}^{n} \ell_{I_t, t} - \min_{p \in \mathcal{X}} \sum_{t=1}^{n} p \cdot \underline{\ell}_t$$

Moyen : Considérer le regret conditionnel : $\overline{R}_n = \sum_{t=1}^{n} \hat{p}_t \cdot \underline{\ell}_t - \min_{p \in \mathcal{X}} \sum_{t=1}^{m} p \cdot \underline{\ell}_t$

En effet : Soit $\mathbb{E}_t$ l'espérance conditionnelle par rapport à l'info disponible au début du tour t (ie, par rapport aux $(\hat{p}_s, I_s, \underline{\ell}_s)_{s=1,\dots t-1}$). $\mathbb{E}_t$ fixe $\hat{p}_t$ et $\underline{\ell}_t$, le seul aléa est alors le choix aléatoire $\hat{p}_t \sim I_t$.

Ainsi, $$\mathbb{E}_t[\ell_{I_t,t}] = \sum_{j=1}^{N} \hat{p}_{jt}\,\ell_{jt} \hat{=} \hat{p}_t \cdot \underline{\ell}_t.$$

Il s'agit donc d'étudier la proximité de $\sum_{t=1}^{n} \ell_{I_t,t}$ à $\sum_{t=1}^{n} \mathbb{E}_t[\ell_{I_t,t}]$ ce qui se fait par inégalités de concentration pour les martingales.

P.ex., notant M_n l'étendue max. aux tours 1 jusque n,
$$M_n = \max_{t \leq n} \left(\max_{j=1,\dots N} \ell_{jt} - \min_{i=1,\dots N} \ell_{it} \right),$$
l'inégalité de Hoeffding-Azuma donne : Avec proba. au moins $1-\delta$,
$$\sum_{t=1}^{n} \ell_{I_t,t} \leq \sum_{t=1}^{n} \mathbb{E}_t[\ell_{I_t,t}] + M_n\sqrt{\frac{n}{2}\ln\frac{1}{\delta}}$$
$$\text{soit} \quad R_n \leq \overline{R}_n + M_n\sqrt{\frac{n}{2}\ln\frac{1}{\delta}}.$$

Rq: en fait, Hoeffding-Azuma est une inégalité maximale :

Avec proba. au moins $1-\delta$, $\quad \forall t \leq n, \quad R_t \leq \overline{R}_t + M_n\sqrt{\frac{n}{2}\ln 1/\delta}$.

<u>Stratégie</u> : Pondération par des poids exponentiels,
$$\hat{p}_{jt} = \exp\left(-\eta_t \sum_{s=1}^{t-1} \ell_{js}\right) \Big/ \sum_{i=1}^{N} \exp\left(-\eta_t \sum_{s=1}^{t-1} \ell_{is}\right)$$
où les $\eta_1, \eta_2, \dots$ forment une suite décroissante choisie sur les données ($(\eta_t)_t$ est donc elle aussi aléatoire, à cause du choix randomisé des I_t).

<u>Bornes</u> : (1) * la plus simple (qui valait pour une borne $M \geq M_n$ connue et finie) est :

Pour <u>toute stratégie</u> du diable, $\quad \overline{R}_n \leq \square\, M\sqrt{n \ln N}$.

pour tout n,

et utilise $\eta_t \sim \frac{1}{M}\sqrt{\frac{\ln N}{t}}$.

En réinjectant, on a donc qu'avec proba. au moins $1-\delta$,

$$\forall t \le n, \qquad R_t \le \square M \sqrt{n \ln(N/\delta)}$$

Par Borel-Cantelli (via les choix $n = 2^r$, $r = 1, 2, \dots$ et $\delta = 2^{-r}$) il vient alors

$$\underline{ps} \quad \overline{\lim_{n \to +\infty}} \frac{R_n}{\square M \sqrt{n \ln(N \ln n)}} = \overline{\lim_{n \to +\infty}} \frac{R_n}{\square M \sqrt{n \ln \ln n}} \le 1$$

et en particulier, $\overline{\lim} \frac{R_n}{n} \le 0$ $\underline{ps}$.

(2) * une borne plus précise, adaptative en n et M_n, est possible (via un meilleur choix des η_t) :

Pour toute stratégie du diable, $\overline{R}_n \le \square \sqrt{\left(\sum_{t=1}^{n} \sum_{i=1}^{N} \hat{p}_{it}\, \ell_{it}^2\right) \ln N} + \cdots$

même : $\sum_{i=1}^{N} \hat{p}_{it} \left(\ell_{it} - \hat{p}_t \cdot \ell_t\right)^2$

Cette fois-ci, on ne concentre plus avec Hoeffding-Azuma (sinon, on a un terme $M_n \sqrt{n \ln N}$ qui annule tous nos efforts pour trouver une borne plus précise !) mais avec l'inégalité de Bernstein-Freedman, sous une hyp. de bornitude par le haut, p.ex., $\ell_{jt} \le B$ pour tous j et t :

{ Avec proba au moins $1-\delta$ et contre toute stratégie du diable,

$$R_n \le \overline{R}_n + \square \sqrt{\left(\sum_{t=1}^{n} \underbrace{\sum_{i=1}^{N} \hat{p}_{it}\, \ell_{it}^2}_{\ge \mathrm{Var}_t\, \ell_{I_t,t}}\right) \ln \frac{1}{\delta}} + \cdots$$

variance conditionnelle de $\ell_{I_t,t}$

On a également une version maximale ; au final :

Avec proba. au moins $1-\delta$, pour tout $t' \le n$,

$$R_{t'} \le \square \sqrt{\left(\sum_{t=1,\dots,n} \sum_{i=1,\dots,N} \hat{p}_{it}\, \ell_{it}^2\right) \ln \frac{N}{\delta}} + \cdots$$

Cette borne implique (cf. pages 7-8) :

$$\text{Avec proba } 1-\delta, \quad \forall t' \le n, \qquad R_{t'} \le \square \sqrt{M_n \left(\min_{j=1,\dots,N} \sum_{t=1}^{n} \ell_{jt}\right) \ln \frac{N}{\delta}} + \cdots$$

et on peut appliquer les mêmes traitements avec Borel-Cantelli.

Rappels : * Pour avoir des bornes en espérance (cf. § inég. oracle), on prend l'espérance de $\overline{R}_n$, pas celle de R_n

* Le cadre simple proposé ici englobe la prédiction avec avis d'experts, cf. $\ell_{j,t} = \ell'_t(f_{j,t})$ si l'avis de l'expert j au tour t est $f_{j,t}$ et que les performances sont mesurées par ℓ'_t.

1. Bandits à N bras.

Origine du nom : Une machine à sous est aussi appelée " bandit manchot".

On suppose qu'on a N bandits manchots (ou, de manière équivalente, un bandit à N bras). A chaque tour, on choisit une machine et joue avec elle. On n'observe que le gain procuré par cette machine et pas le gain qu'on aurait obtenu en en choisissant une autre.

Jeu : Pour tout $t = 1, 2 \dots$

- le statisticien choisit $\hat{p}_t$ proba. sur $\{1, \dots N\}$ (en fonctions des infos des tours passés : $\left(\ell_{I_s, s}\right)_{s=1, \dots t-1}$) ;
- le diable choisit $(\ell_{j,t})_{j=1, \dots N} = \underline{\ell}_t$ (en fonction des infos des tours passés : $(I_s, \hat{p}_s)_{s=1, \dots t-1}$ et de $\hat{p}_t$) ; $\underline{\ell}_t \in (\mathbb{R}^+)^N$ pour simplifier.
- le statisticien tire $I_t \sim \hat{p}_t$ et subit la perte $\ell_{I_t, t}$; il n'observe pas $\underline{\ell}_t$ mais seulement $\ell_{I_t, t}$, sa propre perte ;
- le diable observe I_t (il avait déjà observé $\hat{p}_t$).

But (rappel) : regret $R_n = \sum_{t=1}^{n} \ell_{I_t, t} - \min_{j=1, \dots N} \sum_{t=1}^{n} \ell_{j,t}$ à minimiser

typiquement : $\leq \square M \sqrt{nN \ln(N/\delta)}$ avec proba. $1-\delta$

(notez le facteur $\sqrt{N}$ supplémentaire par rapport au cas d'information parfaite).

<u>Moyen</u>: Estimer les pertes non observées, et utiliser l'algo de pondération par poids exponentiels sur ces pertes estimées :

$$\tilde{\ell}_{jt} = \frac{\ell_{I_t,t}}{\hat{p}_{jt}} \mathbb{1}_{[I_t = j]} \quad \text{est bien un estimateur ;}$$

il est conditionnellement sans biais (E_t est l'espérance conditionnelle par rapport à l'info procurée par les tours $1, 2 \ldots t-1$: ça fixe $\hat{p}_t$ et $\underline{\ell}_t$, le seul aléa est le choix $I_t \sim \hat{p}_t$) :

$$E_t[\tilde{\ell}_{jt}] = \hat{p}_{jt} \cdot \frac{\ell_{jt}}{\hat{p}_{jt}} + 0 = \ell_{jt}.$$

<u>Algo</u>:
$$\hat{p}_{jt} = \exp\left(-\eta_t \sum_{s=1}^{t-1} \tilde{\ell}_{js}\right) \Big/ \sum_{i=1}^{N} \exp\left(-\eta_t \sum_{s=1}^{t-1} \tilde{\ell}_{is}\right)$$

où
$$\eta_t \approx \square \sqrt{\frac{\ln N}{\sum_{s=1}^{t-1} \sum_{i=1}^{N} \hat{p}_{is} (\tilde{\ell}_{is})^2}}$$
(choix adaptatif en tous les paramètres)

<u>Borne</u>: P. ex. en appliquant (2) page 17, qui valait pour toute suite de pertes :

$$\stackrel{ps}{=} \quad \sum_{t=1}^{n} \sum_{j=1}^{N} \hat{p}_{jt} \tilde{\ell}_{jt} - \min_{i=1,\ldots N} \sum_{t=1}^{n} \tilde{\ell}_{it}$$
$$\leqslant \square \sqrt{\left(\sum_{t=1}^{n} \sum_{i=1}^{N} \hat{p}_{it} (\tilde{\ell}_{it})^2\right) \ln N} + \cdots$$

or, $\sum_{j=1}^{N} \hat{p}_{jt} \tilde{\ell}_{jt} = \ell_{I_t,t}$ par déf. des $\tilde{\ell}_{jt}$

et aussi $\sum_{i=1}^{N} \hat{p}_{it} (\tilde{\ell}_{it})^2 = \sum_{i=1}^{N} \ell_{it} \cdot \tilde{\ell}_{it}$

soit
$$\stackrel{ps}{=} \quad \sum_{t=1}^{n} \ell_{I_t,t} - \min_{i=1,\ldots N} \sum_{t=1}^{n} \tilde{\ell}_{it} \leqslant \square \sqrt{\left(\sum_{t=1}^{n} \sum_{i=1}^{N} \ell_{it} \tilde{\ell}_{it}\right) \ln N} + \cdots$$

* Borne en <u>espérance</u> du regret : il suffit de prendre l'espérance dans les 2 membres et d'utiliser l'inégalité de Jensen,

$$\mathbb{E}\left[\sum_{t=1}^{n} \ell_{I_t,t} \right] - \min_{j=1,\dots,N} \mathbb{E}\left[\sum_{t=1}^{n} \ell_{jt} \right]$$

$$\leqslant \mathbb{E}\left[\sum_{t=1}^{n} \ell_{I_t,t} - \min_{j=1,\dots,N} \sum_{t=1}^{n} \tilde{\ell}_{jt} \right]$$

$$\leqslant \mathbb{E}\left[\square \sqrt{\left(\sum_{t=1}^{n} \sum_{i=1}^{N} \ell_{it}\, \tilde{\ell}_{it}\right) \ln N} + \dots \right]$$

$$\underset{\text{(Jensen)}}{\leqslant} \square \sqrt{\left(\sum_{t=1}^{n} \sum_{i=1}^{N} \mathbb{E}\left[(\ell_{it})^2\right]\right) \ln N} + \dots$$

(cf. $\mathbb{E}[\ell_{it}\, \tilde{\ell}_{it}] = \mathbb{E}[\ell_{it}\, \mathbb{E}_t[\tilde{\ell}_{it}]] = \mathbb{E}[(\ell_{it})^2]$)

→ Sous l'hyp. (typique des suites individuelles) que les $\underline{\ell}_t$ sont bornées par M, on a une borne $\leqslant \square M \sqrt{nN \ln N}$ sur l'espérance.

→ Juditsky, Nazin, Tsybakov & Vayatis '07 ont été les premiers à remarquer qu'une hyp. du second ordre pouvait être très judicieuse ici : $\mathbb{E}[(\ell_{it})^2] \leqslant \sigma^2$ pour tous i et t (c'est une hyp. qui se comprend bien dans un cadre stochastique, pas tellement pour des suites individuelles). On a :

$$\mathbb{E}\left[\sum_{t=1}^{n} \ell_{I_t,t} \right] - \min_{j=1,\dots,N} \mathbb{E}\left[\sum_{t=1}^{n} \ell_{jt} \right] \leqslant \square\, \sigma \sqrt{nN \ln N}.$$

Réf. : → Dans Auer, Cesa-Bianchi, Freund & Schapire '95 puis '02, l'algo utilise un mélange avec l'uniforme dont on n'a en fait pas besoin (selon une remarque faite postérieurement) pour le cas de l'espérance du regret (mais on en a besoin pour borner le regret avec grande proba). Voir, p. ex., Stoltz '05 (thèse, Th. 2.7 page 40) : c'est le chapitre des rappels, je doute que le résultat soit de moi (mais le moment où je l'ai connu pour la première fois m'échappe). Ça, c'est pour la borne $\square M \sqrt{nN \ln N}$.

→ La borne améliorée en $\sigma \sqrt{nN \ln N}$ n'apparaît <u>pas</u> dans Cesa-Bianchi, Mansour et Stoltz '05, '07, qui donnent en revanche l'ingrédient essentiel, à savoir l'inég. (2) p. 17 ; l'application d'une inég. d'info parfaite à des pertes estimées est faite p.ex. dans Cesa-Bianchi, Lugosi et Stoltz '05 ; mais au final, les calculs précis des pages 19-20 sous cette forme sont postérieurs à Juditsky, Nazin, Tsybakov & Vayatis '07 et ont été menés pour cet exposé. (Il restait même dans Cesa-Bianchi, Mansour et Stoltz '05, '07 comme question ouverte non-écrite autre part que dans des emails entre les auteurs à trouver une borne du 2nd ordre pour les bandits à N bras.)

Rq1 : Borne avec $\sqrt{N^2 \left(\min_{i=1,\dots N} \mathbb{E}\left[\sum_{t=1}^{n} \ell_{it} \right] \right) \ln N}$ dans Cesa-Bianchi, Mansour et Stoltz '05, version COLT uniquement.

Obtenue en revenant à la majoration $\frac{2 \ln N}{\eta_{n+1}} + \sum_{t=1}^{n} \eta_t \, \overline{\Phi}(\eta_t, \hat{p}_t, (\tilde{\ell}_{jt})_j)$ et en traitant les $\overline{\Phi}(\dots)$ convenablement.

Permet d'obtenir des vitesses rapides.

Rq2 : Borne sur le regret avec grande proba :

Ici, on ne passe plus par $\overline{R}_n$, ... On part de l'inég. de la page 19,

$$\sum_{t=1}^{n} \ell_{I_t,t} - \min_{i=1,\dots N} \sum_{t=1}^{n} \tilde{\ell}_{it} \leqslant \dots$$

et il s'agit de montrer (entre autres) que $\sum_{t=1}^{n} \tilde{\ell}_{it} \leqslant \sum_{t=1}^{n} \ell_{it} + \Delta_n$ avec grande proba. Δ_n dépend (entre autres) des $\mathrm{Var}_t \, \tilde{\ell}_{it}$ (cf. Bernstein)

$$\leqslant \mathbb{E}_t \left(\tilde{\ell}_{it}\right)^2 = \frac{(\ell_{it})^2}{\hat{p}_{it}}$$

Cela suggère qu'il faut minorer les $\hat{p}_{it}$ uniformément, c'est pourquoi on a recours à un mélange avec la loi uniforme :

$$\hat{p}_{it} = \underbrace{(1-\gamma) \frac{\exp\left(-\eta_t \sum_{s=1}^{t-1} \tilde{\ell}_{is}\right)}{\sum_{j=1}^{N} \exp\left(-\eta_t \sum_{s=1}^{t-1} \tilde{\ell}_{js}\right)}}_{\text{terme d'exploitation}} + \underbrace{\frac{\gamma}{N}}_{\text{terme d'exploration}}$$

γ est le compromis entre exploration et exploitation.

En fait, juste cette modif. conduit à des vitesses $n^{2/3}$ (borne de '95 de Cesa-Bianchi et al.). Plus de travail (conversion pertes → gains, translation par facteur de variance) donne (Cesa-Bianchi et al. '02, Cesa-Bianchi et Lugosi '06) :

__Algo__ Exp3.P (P = proba, Exp3 = exponential-weight algorithm for exploration and exploitation)

Paramètres : β, η, γ ; pertes $\underline{\ell}_t \in [0,1]^N$

Init. : $w_{i0} = 1$ et $\hat{p}_{i1} = 1/N$ pour tout $i = 1, \dots N$

Pour $t = 1, 2, \dots$

1. choisir $I_t \sim \hat{p}_t$;
2. estimer les gains $\tilde{g}_{it} = \dfrac{1-\ell_{it}}{\hat{p}_{it}} \mathbb{1}_{[I_t = i]} + \dfrac{\beta}{\hat{p}_{it}}$;
3. mettre à jour les poids selon $w_{it} = w_{i,t-1} \cdot e^{-\eta \tilde{g}_{it}}$;
4. définir $\hat{p}_{\cdot t+1}$ selon

$$\hat{p}_{i,t+1} = (1-\gamma) \frac{w_{it}}{\sum_{j=1}^{N} w_{jt}} + \frac{\gamma}{N}, \qquad i = 1, \dots N.$$

__Borne :__ Sous l'hyp. que $\underline{\ell}_t \in [0,1]^N$ pour tout t,
pour tout $\delta \in]0,1[$, pour tout $n \geq \frac{1}{N} \ln \frac{N}{\delta}$,
et pour les choix

$$\beta = \sqrt{\frac{\ln(N/\delta)}{nN}}, \qquad \gamma = \frac{6N\beta}{4+\beta}$$

et $\eta = \dfrac{\gamma}{3N}$, on a avec proba. $1-\delta$ contre toute stratégie du diable :

$$\sum_{t=1}^{n} \ell_{I_t,t} - \min_{j=1,\dots N} \sum_{t=1}^{n} \ell_{jt} \leq 6\sqrt{nN \ln \frac{N}{\delta}} + \frac{\ln N}{2}$$

Rq : Evidemment, on peut utiliser $\beta_t, \eta_t, \gamma_t$ pour traiter l'adaptation en les paramètres (y compris M pour $\underline{\ell}_t \in [0,M]^N$ avec M inconnu) ; ou rester aux param. fixes à l'intérieur d'un régime et couper le temps en ≠ts régimes (comme en page 14).

<u>Optimalité</u> : $\sqrt{nN}$ inévitable, même dans le cas stochastique :

pour $\ell : \mathbb{N} \times [0,1] \longrightarrow \{0,1\}$ définie par $\ell(j,y) = j^{\text{ème}}$ comp. du développement dyadique de y, on a : pour toute strat. du statisticien,

$$\forall N \geq 2, \quad \forall n \geq 2, \qquad \sup_{y_1, \dots y_n \in [0,1]} \mathbb{E}\left[\sum_{t=1}^{n} \ell(I_t, y_t) \right] - \min_{i=1,\dots N} \sum_{t=1}^{n} \ell(i, y_t)$$

$$\geq \frac{1}{20} \min \left\{ \sqrt{nN}, n \right\}$$

preuve : dans le cas où les y_j sont les réalisations de $Y_1, Y_2 \dots$ iid selon une loi bien choisie. Utilise Pinsker (pas Fano).

↳ facteur $\sqrt{\ln N}$ manquant ; améliorer borne inf. ou borne sup. ?

2. Prédiction économe en observations.

(Référence = Cesa-Bianchi, Lugosi, et Stoltz '05.)

<u>Jeu de prédiction</u> : Paramètres : décisions $\{1, 2, \dots N\}$ pour le statisticien

le diable prend des vecteurs de pertes dans $\mathbb{R}^N$

fonction $\mu : \mathbb{N} \to \mathbb{N}$ donnant le nombre max. d'observations

(croissante)

Pour $t = 1, 2, \dots$

1. le statisticien choisit $\hat{p}_t \in \mathcal{X}$ (proba. sur $\{1, \dots N\}$, en fonction des infos des tours passés)
2. le diable choisit $\underline{\ell}_t$ (en fonction des infos des tours passés, à savoir $(I_s, \hat{p}_s)_{s \leq t-1}$ et de $\hat{p}_t$)

   ($\underline{\ell}_t \in [0,1]^N$ pour simplifier)
3. le statisticien tire $I_t \sim \hat{p}_t$ et subit $\ell_{I_t, t}$ comme perte (non obs.)
4. si strictement moins de $\mu(t)$ vecteurs $\underline{\ell}_s$ ont été observés jusque là, le statisticien peut choisir de voir $\underline{\ell}_t$; sans quoi, il en demeure dans l'ignorance.

<u>But</u> : regret $R_n = \sum_{t=1}^{n} \ell_{I_t, t} - \min_{j=1,\dots N} \sum_{t=1}^{n} \ell_{j,t}$ à minimiser

Cas simple. Horizon n fixé, pertes $\underline{\ell}_t \in [0,1]^N$ et budget μ tq. pour un m fixé et connu : $\mu(1) = \dots = \mu(n) = m$.

Algo: Paramètres: $\eta > 0$ et $\varepsilon \in]0,1[$ (valeurs données ci-dessous).

Initialisation : $w_1 = (1, \dots 1)$

Pour $t = 1, 2, \dots n$:

1. choisir $I_t \sim \hat{p}_t$ où
$$\hat{p}_{jt} = \frac{w_{jt}}{\sum_{i=1,\dots N} w_{it}} \quad \text{pour } j = 1, \dots N ;$$
2. tirer $Z_t \sim Ber(\varepsilon)$;
3. si $Z_t = 1$, demander à voir $\underline{\ell}_t$ (si c'est possible!) et mettre à jour $w_{i,t+1} = w_{it}\, e^{-\eta \ell_{it}/\varepsilon}$ pour tout i ; sinon, lorsque $Z_t = 0$, $w_{i,t+1} = w_{it}$ pour tout i.

Analyse: * $Z_1, \dots Z_n$ sont des randomisations auxiliaires indépendants de tout le reste ; $\mathbb{E}_t$ est l'espérance conditionnelle par rapport aux tours passés (fixe $\hat{p}_t$ et $\underline{\ell}_t$, les seuls aléas sont dans $I_t \sim \hat{p}_t$ et $Z_t \sim Ber(\varepsilon)$).

On note $\tilde{\ell}_{jt} = \frac{\ell_{jt}}{\varepsilon} Z_t$: ce sont des estimateurs conditionnellement sans biais :
$$\mathbb{E}_t [\tilde{\ell}_{jt}] = \ell_{jt}$$
et même (on en aura besoin) :
$$\mathbb{E}_t \Big[\sum_{j=1,\dots N} \hat{p}_{jt}\, \tilde{\ell}_{jt} \Big] = \hat{p}_t \cdot \underline{\ell}_t$$

* On veut contrôler $Z_1 + \dots Z_n$ ($\le m$?): par inégalité de Bernstein.

Lm (Bernstein - Freedman) : $X_1, \dots X_n$ suite d'accroissements de martingale, par rapport à $(\mathcal{F}_t)_{1 \le t \le n}$, bornés supérieurement : $X_j \le K$ ps.

V_n somme des variances conditionnelles : $V_n = \sum_{t=1}^{n} \mathbb{E}[X_t^2 \mid \mathcal{F}_t]$.

Alors $\mathbb{P}\left(\max_{t=1,\dots n} X_1 + \dots X_t > \sqrt{2vx} + \frac{\sqrt{2}}{3} Kx \text{ et } V_n \le v \right) \le e^{-x}$.

version maximale à nouveau

Application à $X_t = Z_t - \varepsilon$: $V_n \le n\varepsilon = v$, $K = 1$,

soit avec proba. au moins $1-\delta'$, pour tout $t = 1, \dots n$:

$$Z_1 + \dots Z_t \le \underbrace{n\varepsilon + \sqrt{2n\varepsilon \ln 1/\delta'} + \frac{\sqrt{2}}{3} \ln 1/\delta'}_{\le m}$$

$\le m$ p. ex. pour :

$$\varepsilon = \left(\frac{m}{n} - \left(\frac{\sqrt{2m \ln 1/\delta'}}{n} + \frac{\sqrt{2}}{3n} \ln 1/\delta' \right) \right)^+ \le \frac{m}{n}.$$

* Contrôle du regret maintenant : on revient à la borne la plus simple (voir p.6)

$$\sum_{t=1}^{n} \sum_{j=1}^{N} \hat{p}_{jt} \tilde{\ell}_{jt} - \min_{i=1,\dots N} \sum_{t=1}^{n} \tilde{\ell}_{it}$$

$$\le \frac{\ln N}{\eta} + \eta \sum_{t=1}^{n} \sum_{j=1}^{N} \hat{p}_{jt} \left(\tilde{\ell}_{jt}\right)^2$$

cf. $\underline{\ell}_t \in [0,1]^N$ pour simplifier, voir p.23

et on concentre : $X_t = \tilde{\ell}_{it} - \ell_{it}$ $K = 1/\varepsilon$,

$$\mathrm{Var}_t\, X_t \le \mathbb{E}_t\left[\left(\tilde{\ell}_{it}\right)^2\right] \le \frac{(\ell_{it})^2}{\varepsilon^2} \varepsilon < 1/\varepsilon$$

soit $v \le n/\varepsilon$

et au final, avec proba. δ' :

$$\sum_{t=1}^{n} \tilde{\ell}_{it} \le \sum_{t=1}^{n} \ell_{it} + \sqrt{\frac{2n}{\varepsilon} \ln 1/\delta'} + \frac{\sqrt{2}}{3\varepsilon} \ln 1/\delta'$$

· de même pour $\sum_{j=1}^{N} \hat{p}_{jt} \ell_{jt} - \sum_{j=1}^{N} \hat{p}_{jt} \tilde{\ell}_{jt}$ (et même une borne plus petite, ici $K = 1$)

· et pour le membre de droite :

$$\sum_{t=1}^{n} \sum_{j=1}^{N} \hat{p}_{jt} \left(\tilde{\ell}_{jt}\right)^2 \leq \frac{n}{\varepsilon} + \sqrt{2(n/\varepsilon^3) \ln 1/\delta'} + \frac{\sqrt{2}}{3\varepsilon^2} \ln 1/\delta'$$

cf. $\mathbb{E}_t \left(\tilde{\ell}_{jt}\right)^2 \leq \frac{1}{\varepsilon}$ et $\begin{cases} \left(\tilde{\ell}_{jt}\right)^2 \leq \frac{1}{\varepsilon^2} \\ \mathrm{Var}_t \left(\tilde{\ell}_{jt}\right)^2 \leq \frac{1}{\varepsilon^3} \end{cases}$

- Au final : Avec proba. $1 - 3\delta'$,

$$R_n = \sum_{t=1}^{n} \sum_{j=1}^{N} \hat{p}_{jt} \ell_{jt} - \min_{i=1,\dots,N} \sum_{t=1}^{n} \ell_{it}$$

$$\leq \square \left(\sqrt{\frac{n}{\varepsilon} \ln 1/\delta'} + \frac{\ln 1/\delta'}{\varepsilon} \right) + \frac{\ln N}{\eta}$$

$$+ \square \eta \left(\frac{n}{\varepsilon} + \sqrt{\frac{n}{\varepsilon^3} \ln 1/\delta'} + \frac{\ln 1/\delta'}{\varepsilon^2} \right)$$

$\left(\varepsilon \sim \frac{m}{n} \right)$ $$\sim \square \frac{n}{\sqrt{m}} \sqrt{\ln 1/\delta'} + \square \frac{n}{m} \ln 1/\delta'$$

$$+ \frac{\ln N}{\eta} + \square \eta \left(\frac{n^2}{m} + \underbrace{\frac{n^2}{\sqrt{m^3}} \sqrt{\ln 1/\delta'} + \frac{n^2}{m^2} \sqrt{\ln 1/\delta'}}_{\leq \frac{n^2}{m} \text{ pour } m \geq \ln 1/\delta'} \right)$$

$\left(\begin{matrix} \text{choix} \\ \eta \sim \frac{1}{n} \sqrt{m \ln N} \end{matrix} \right)$ $\bar{R}_n \leq \square \left(\underbrace{\frac{n}{\sqrt{m}} \sqrt{\ln 1/\delta'}}_{} + \square \frac{n}{m} \ln 1/\delta' \right)$.

et celui-là est alors plus petit!

ce terme nous permet de supposer $m \geq \square \ln 1/\delta'$ sinon, borne triviale n est meilleure !

Un dernier coup d'Hoeffding-Azuma pour relier R_n à $\bar{R}_n$ (à $\sqrt{n \ln 1/\delta'}$ près) et on a (avec $\delta' = \delta/5$) :

<u>Th</u>: $\begin{cases} R_n \leq \square \frac{n}{\sqrt{m}} \sqrt{\ln \frac{5}{\delta}} \\ Z_1 + \dots Z_n \leq m \end{cases}$ avec proba $1-\delta$

<u>Rq</u>: Pour $m = n$, on retrouve la borne habituelle.

<u>Rq</u>: Hoeffding-Azuma ne suffit pas pour cette preuve ; si on l'utilise on a de plus mauvaises vitesses. (Ex : $\sum \tilde{\ell}_{it} \leq \sum \ell_{it} + \square \varepsilon \sqrt{n \ln 1/\delta'}$ ...).

Cas d'une fonction μ générale $\mathbb{N} \to \mathbb{N}$:

Cor : si μ est croissante tq. $\lim\limits_{n \to +\infty} \dfrac{\mu(n)}{(\log n) \log\log n} = +\infty$

alors il existe une stratégie pour le statisticien tq.

$$\overline{\lim} \ \frac{R_n}{n} \leq 0 \ \text{ps.}$$

tout en respectant les contraintes données par μ.

Preuve (éléments) : Cette stratégie découpe le temps en régimes $r = 1, 2, \dots$ et choisit dans chaque régime sa longueur n_r, le nombre max. d'obs. m_r (ainsi qu'un niveau de confiance δ_r, tq. $\Sigma \delta_r < +\infty$ pour appliquer Borel-C.).

Détails dans Cesa-Bianchi, Lugosi & Stoltz '05.

Optimalité (à n et m fixés).

Th : Pour $\ell : \mathbb{N} \times [0,1] \to \{0,1\}$ (définie comme d'habitude selon $\ell(j,y)$ = $j$ème élément du dvpt dyadique de $y \in [0,1]$)

et pour $c > 0$ (qu'on peut calculer précisément),

pour tout $N \geqslant 2$ et $n \geqslant m \geqslant \square \ln N$,

toute stratégie du statisticien utilisant au plus m observations a un regret tq.

$$\sup_{y_1, \dots, y_n \in [0,1]} \mathbb{E}\Big[\sum_{t=1}^{n} \ell(I_t, y_t) \Big] - \min_{i=1,\dots,N} \sum_{t=1}^{n} \ell(i, y_t) \geqslant c \ m \sqrt{\frac{\ln N}{m}} .$$

preuve : on remplace le $\sup\limits_{y_1, \dots, y_n \in [0,1]}$ par N probabilités-produits sur $[0,1]^n$, chacune favorisant une action i

à ε près :

· Q_i est la loi de $\quad \mathrm{Ber}(\tfrac12-\varepsilon)\times 2^{-i} + \sum_{\substack{k=1,\dots N\\ k\neq i}} 2^{-k}\,\mathrm{Ber}(\tfrac12) + 2^{-(N+1)}\,\mathcal{U}_{[0,1]}$

Sous $Q_i^{\otimes n}$, les

$$\ell(k,Y_t), \qquad k=1,\dots N, \quad t=1,\dots n \quad \text{sont}$$

indépendants, avec $\quad \ell(k,Y_t)\sim \mathrm{Ber}(\tfrac12)$ si $k\neq i$

$\qquad \ell(i,Y_t)\sim \mathrm{Ber}(\tfrac12-\varepsilon)$

· puis $\displaystyle \sup_{y_1,\dots y_n\in[0,1]} \max_{i=1,\dots N} \mathbb{E}\Big[\sum_{t=1}^n \ell(I_t,y_t)\Big] - \sum_{t=1}^n \ell(i,y_t)$

$$\geqslant \max_{i=1,\dots N} \mathbb{E}_{Q_i^{\otimes n}}\Big[\underbrace{\underbrace{\mathbb{E}_A}_{\text{random auxiliaire}}\Big[\sum_{t=1}^n \ell(I_t,Y_t)\Big]}_{=\frac{n}{2}-\varepsilon\sum_{t=1}^n Q_i^{\otimes n}\otimes \mathbb{P}_A(I_t=i)} - \underbrace{\sum_{t=1}^n \ell(i,Y_t)}_{=n/2-n\varepsilon}\Big]$$

$$= \max_{i=1,\dots N} n\varepsilon\Big(1-\frac1n\sum_{t=1}^n Q_i^{\otimes n}\otimes \mathbb{P}_A(I_t=i)\Big)$$

$$= n\varepsilon\Big(1-\underbrace{\min_{i=1,\dots N}\frac1n\sum_{t=1}^n Q_i^{\otimes n}\otimes \mathbb{P}_A(I_t=i)}_{\text{par FANO généralisé (voir plus bas)}}\Big)$$

ceci est

$$\leqslant \max\Big\{\frac{\overline{K}}{\ln N},\ \square\Big\} \quad \text{où } \overline{K}=\frac{1}{Nn}\sum_{\substack{t=1,\dots n\\ i=2,\dots N}} \mathrm{KL}\Big(\big(Q_i^{\otimes n}\otimes \mathbb{P}_A\big)^{I_t}, \big(Q_1^{\otimes n}\otimes \mathbb{P}_A\big)^{I_t}\Big)$$

or le choix de I_t dépend au plus de m observations : $\quad \mathrm{KL}(\ ,\ )\leqslant \mathrm{KL}(Q_i^{\otimes m},Q_1^{\otimes m})$

$\leqslant m\,\mathrm{KL}(Q_i,Q_1)$

$\leqslant m\big(\mathrm{KL}(\mathrm{Ber}(\tfrac12),\mathrm{Ber}(\tfrac12-\varepsilon)) + \mathrm{KL}(\mathrm{Ber}(\tfrac12-\varepsilon),\mathrm{Ber}(\tfrac12))\big)$

$\simeq \square\, m\varepsilon^2$

Ici il faut en fait considérer d'abord 1 strat. dét. et étendre aux strat. randomisées, voir CB, L, S '05 pour 1 vraie preuve.

· Au final, $\sup_{y_1,\dots y_n} \mathbb{E}_{\mathbb{A}}[R_n] \geqslant \max_{i=1,\dots N} \mathbb{E}_{\mathbb{Q}_i^{\otimes n} \otimes \mathbb{A}}[R_n]$

$$\geqslant n\varepsilon_1 \left(1 - \max\left\{\frac{\square m\varepsilon^2}{\ln N}, \square\right\}\right)$$

$$\geqslant \square \frac{n}{\sqrt{m}} \sqrt{\ln N} \qquad \text{pour le choix } \varepsilon \sim \sqrt{\frac{\ln N}{m}}.$$

Reste juste à voir le lemme de Fano pour des combinaisons convexes ...

<u>Lm</u> (Fano, version Birgé – Massart) : $\{A_0, \dots A_N\}$ partition de Ω,

$\mathbb{P}_0, \dots \mathbb{P}_N$ N probabilités sur Ω,

alors $\min_{j=1\dots N} \mathbb{P}_j(A_j) \leqslant \max\left\{\frac{\overline{K}}{\ln(N+1)}, \frac{2e}{2e+1}\right\}$

où $\overline{K} = \frac{1}{N}\sum_{j=1}^{N} KL(\mathbb{P}_j, \mathbb{P}_0)$.

<u>preuve</u> (très courte) :

- rappels : $\begin{cases} (\mathbb{P}, \mathbb{Q}) \mapsto KL(\mathbb{P},\mathbb{Q}) \text{ est convexe} \\ \text{pour toute v.a. } X, \quad KL(\mathbb{P}^X, \mathbb{Q}^X) \leqslant KL(\mathbb{P},\mathbb{Q}) \end{cases}$

- ainsi, $\forall\, \mathbb{P}_i, \mathbb{Q}_i, A_i$, sans conditions pour l'instant :

$$KL(\mathbb{P}_i, \mathbb{Q}_i) \geqslant KL\left(\mathbb{P}_i^{\mathbb{1}_{A_i}}, \mathbb{Q}_i^{\mathbb{1}_{A_i}}\right)$$
$$= KL\left(Ber(\mathbb{P}_i(A_i)), Ber(\mathbb{Q}_i(A_i))\right)$$

Par convexité, $\overline{K} = \frac{1}{N}\sum_{i=1}^{N} KL(\mathbb{P}_i, \mathbb{Q}_i) \geqslant \frac{1}{N}\sum_{i=1}^{N} KL\left(Ber(\mathbb{P}_i(A_i)), Ber(\mathbb{Q}_i(A_i))\right)$

$$\geqslant KL\left(Ber(p), Ber(q)\right)$$

où $p = \frac{1}{N}\sum_{i=1}^{N} \mathbb{P}_i(A_i)$, $q = \frac{1}{N}\sum_{i=1}^{N} \mathbb{Q}_i(A_i)$

$$(\ast\ast) \quad = p \log\frac{p}{q} + \underbrace{(1-p)\log\frac{1-p}{1-q}}$$

$\log\frac{1}{1-q} \geqslant 0$,

$(1-p)\log(1-p) \geqslant -1/e$

d'où $\quad \overline{K} \geqslant \left(p \log \frac{p}{q}\right) - 1/e$.

- On en déduit le lemme : on prend $Q_i = \mathbb{P}_0$,

$$a = \min_{j=0,\dots N} \mathbb{P}_j(A_j) \leqslant p$$

c'st là qu'on utilise l'hyp. de partition !

et $\quad q = \frac{1}{N}\left(1 - \mathbb{P}_0(A_0)\right) \geqslant \frac{1}{N}(1-a)$

on peut supposer $a \geqslant 1/e$ (et alors, $p \geqslant 1/e$) :

cf. on peut supp. $a \geqslant \frac{2e}{2e+1} \geqslant 1/e$, sinon le rés. est acquis.

puis : $x \mapsto x \log x$ croissante sur $[1/e, 1]$

et donc $\quad p \log p \geqslant a \log a$

$$\log q \geqslant \log \frac{1}{N}(1-a)$$

et donc $\quad p \log\left(\frac{p}{q}\right) - 1/e \geqslant a \log\left(\frac{Na}{1-a}\right) - 1/e$

$$\underset{\left(\text{car } a \geqslant 1/e\right)}{\geqslant} a \log\left(\frac{Na}{e(1-a)}\right)$$

$$\overline{K} \geqslant a \log\underbrace{\left(\frac{Na}{e(1-a)}\right)} \geqslant a \log(N+1) \quad \text{pour } a \geqslant \frac{2e}{2e+1}.$$

$a \mapsto \frac{a}{e(1-a)}$ fonction croissante de a, $\xrightarrow[a \to 1]{} +\infty$

donc pour $a \geqslant \square$,

c'st $\geqslant 1 + 1/2 \geqslant 1 + 1/N$

(P.ex., $\square = \frac{2e}{2e+1}$, proposition de Pascal Massart, fait que c'st $\geqslant 2$)

<u>Réf:</u> Birgé 2005 IEEE IT

Massart 2003 St-Flour

plus, notamment pour arriver à l'inég. clé $\overline{K} \geqslant KL(\mathcal{B}er(p), \mathcal{B}er(q))$, une simplification personnelle (non publiée)... qui d'ailleurs

nous permet d'arriver à un résultat partiel avec des couples $(\mathbb{P}_i, \mathbb{Q}_i)$;
ça pourrait être utile pour .... ?

$\longrightarrow$ <u>APPEL</u> aux bonnes idées !

Voyons l'extension aux combinaisons convexes :
Suffit de reprendre une couche de combinaisons convexes en $(\overset{**}{*})$:

$$\overline{K} = \frac{1}{N} \sum_{i=1}^{N} \sum_{s=1}^{S} \alpha_s \, KL(\mathbb{P}_i^s, \mathbb{Q}_i^s)$$

$$\geq KL(\mathcal{B}er(p), \mathcal{B}er(q))$$

où $$p = \frac{1}{N} \sum_{i=1}^{N} \sum_{s=1}^{S} \alpha_s \, \mathbb{P}_i^s(A_i^s)$$

$$q = \frac{1}{N} \sum_{i=1}^{N} \sum_{s=1}^{S} \alpha_s \, \mathbb{Q}_i^s(A_i^s)$$

On conclut comme en p 30 pour peu que pour tout s, $(A_0^s, \dots A_N^s)$ soit une partition de Ω. On a obtenu :

<u>Lm</u> (FANO pour des combinaisons convexes) :
$\{A_0^s, \dots A_N^s\}$ $(s = 1, \dots S)$ sont S partitions de Ω,
$(\mathbb{Q}_0^s, \dots \mathbb{Q}_N^s)$ sont S $(N+1)$-uplets de probabilités de Ω;
$\alpha_1, \dots \alpha_S \geq 0$ tq. $\alpha_1 + \dots \alpha_S = 1$;

alors $$\min_{j=0,\dots N} \sum_{s=1}^{S} \alpha_s \, \mathbb{Q}_j^s(A_j^s) \leq \max\left\{ \frac{2e}{2e+1}, \frac{\overline{K}}{\ln(N+1)} \right\}$$

où $$\overline{K} = \frac{1}{N-1} \sum_{j=2}^{N} \sum_{s=1}^{S} \alpha_s \, KL(\mathbb{Q}_j^s, \mathbb{Q}_0^s).$$