

E x 1

1) La fonction \exp est convexe. Ainsi, $\forall z \in [a, b]$,

$$\exp(sz) \leq \frac{b-z}{b-a} \exp(sa) + \frac{z-a}{b-a} \exp(sb)$$

$$E[z] = 0.$$

$$\begin{matrix} 0 & \\ \curvearrowleft & \curvearrowright \\ 0 & \end{matrix}$$

$$\Rightarrow E[e^{sz}] \leq \frac{be^{sa} - E[ze^{sa}] + E[ze^{sb}] - ae^{sb}}{b-a}$$

$$= \frac{e^{sa}(b - e^{s(b-a)}a)}{b-a}$$

$$= e^{-s(b-a)p} (1-p + pe^{s(b-a)}), \quad p \triangleq \frac{a}{b-a}$$

$$= e^{-up} (1-p + pe^u), \quad u \triangleq s(b-a)$$

$$= \exp(-up + \ln(1-p + pe^u))$$

On pose $\Psi(x) = -xp + \ln(1-p + pe^x)$.

$$\text{On a } \Psi'(x) = -p + \frac{pe^x}{1-p + pe^x} = -p + \frac{p}{(1-p)e^{-x} + p}$$

$$\Psi''(x) = \frac{p(1-p)e^{-2x}}{((1-p)e^{-x} + p)^2} = t(1-t) \leq \frac{1}{4}$$

$$t = \frac{(1-p)e^{-x}}{(1-p)e^{-x} + p}$$

On utilise la formule de Taylor-Lagrange : $\exists \theta \in [0, u]$,

$$\begin{aligned} Y(u) &= Y(0) + u Y'(0) + \frac{u^2}{2!} Y''(\theta) \quad Y(0) = k'(0) = 0 \\ &\leq \frac{u^2}{2!} = \frac{s(b-a)^2}{8}. \end{aligned}$$

On a donc bien $E[e^{sz}] \leq \exp\left(\frac{s^2(b-a)^2}{8}\right)$

2) $(Z_i)_{\substack{i \in \mathbb{N} \\ i \leq n}}$ iid, $P(0 \leq Z_i \leq 1) = 1$.

On pose $X_i = Z_i - E[Z_i]$.

$$\begin{aligned} P(\bar{X}_n > t) &= P(e^{s\bar{X}_n} > e^{st}), \quad \forall s > 0 \\ &\leq e^{-st} E[e^{s\bar{X}_n}] \quad (\text{Markov}) \\ &= e^{-st} E\left[e^{\frac{s}{n} \sum_i X_i}\right] \\ &= e^{-st} \prod_i E\left[e^{\frac{s}{n} X_i}\right] \\ &= e^{-st} \prod_i E\left[e^{\frac{s}{n} X_i}\right] \quad ((X_i) \text{ iid}) \\ &\leq e^{-st} \prod_i \exp\left(\frac{s^2}{8n^2}\right) \quad (\text{Lemma avec } E[X_i], 1 - E[X_i]) \\ &= e^{\frac{s^2}{8n} - st} \end{aligned}$$

On cherche s qui minimise $f: s \mapsto \frac{s^2}{8n} - st$.

$$f'(s) = \frac{s}{hn} - t$$

s	0	ht	+oo

↓ ↗

On pose $s = hnt > 0$:

$$P(\bar{X}_n > t) \leq \exp\left(-h_n t^2 + \frac{\ln^2 t^2}{R_n}\right) = e^{-2n t^2}$$

On procéde de manière similaire avec $-\bar{X}_n = E[Z_i] - \bar{Z}_n$
et $[a,b] = [E(Z_1) - 1, E(Z_1)]$.

Ex 2)

On commence par prouver le lemme suivant :

Soit V et Z deux r.v. tq $E[V|Z] = 0$ et

$\exists f, c$ tq $f(Z) \leq V \leq f(Z) + c$.

$\forall s > 0 \quad E[e^{sV}|Z] \leq e^{sc^2/8}$.

Même preuve que pour le lemme de Hoeffding avec
 $E[V e^{f(Z)}|Z] = 0$, $E[V e^{s(f(Z)+c)}] = 0$, et $[a,b] = [f(Z), f(Z)+c]$

□

En utilisant l'inégalité de Markov, on a

$$P\left(\sum_{i=1}^n V_i > t\right) = P\left(e^{s \sum V_i} > e^{st}\right), \quad \forall s > 0$$

$$\leq e^{-st} \mathbb{E}[e^{s \sum V_i}] \quad (\text{Markov})$$

On développe l'espérance :

$$\mathbb{E}[e^{s \sum V_i}] = \mathbb{E}\left[\mathbb{E}\left[e^{s V_n} e^{s \sum_{i=1}^{n-1} V_i} \mid Z_1, \dots, Z_{n-1}\right]\right]$$

$$(V_i)_{1 \leq i \leq n-1} \text{ is a function of } Z_1, \dots, Z_{n-1} = \mathbb{E}\left[e^{s \sum_{i=1}^{n-1} V_i} \mathbb{E}[e^{s V_n} \mid Z_1, \dots, Z_{n-1}]\right]$$

En utilisant le lemme précédent avec $f(Z_n) = Z_n$ et $c = c_n$, on obtient

$$\leq \mathbb{E}\left[e^{s \sum_{i=1}^{n-1} V_i} e^{s^2 c_n^2 / 8}\right]$$

$$\leq e^{s^2 \sum c_i^2 / 8}.$$

Par récurrence,

$$\text{Ainsi, } P\left(\sum V_i > t\right) \leq e^{-st} e^{s^2 \sum c_i^2 / 8}$$

$$= e^{-st + s^2 \sum c_i^2 / 8}$$

$$= \exp\left(-\frac{2t^2}{\sum c_i^2}\right), \quad s \triangleq \frac{ht}{\sum c_i^2}.$$

□

On procède de manière analogue avec $(-V_i)_{1 \leq i \leq n}$ et

$$f(Z_n) = -Z_n - c_n.$$

$$2) \text{ On pose } V_i = \mathbb{E}[h(Z_1, \dots, Z_n) | Z_1, \dots, Z_{i-1}] - \mathbb{E}[h(Z_1, \dots, Z_n) | Z_1, \dots, Z_{n-1}]$$

$$\begin{aligned} \text{On a alors } \sum_{i=1}^n V_i &= \mathbb{E}[h | Z_1] - \mathbb{E}[h] \\ &\quad + \mathbb{E}[h | Z_1, Z_2] - \mathbb{E}[h | Z_1] \\ &\quad \dots \\ &= \underbrace{\mathbb{E}[h | Z_1, \dots, Z_n]}_{h \text{ est } \sigma(Z_1, \dots, Z_n)\text{-mesurable}} - \mathbb{E}[h] \\ &= h(Z_1, \dots, Z_n) - \mathbb{E}[h]. \end{aligned}$$

On pose $F_i \triangleq \sigma(Z_1, \dots, Z_i)$. Comme $F_{i-1} \subseteq F_i$, V_i est F_i -mesurable. De plus :

$$\begin{aligned} \mathbb{E}[V_i | F_{i-1}] &= \mathbb{E}[\mathbb{E}[h | F_i] | F_{i-1}] \\ &= \mathbb{E}[h | F_{i-1}] \\ &= \mathbb{E}[h | F_{i-1}] - \mathbb{E}[h | F_{i-1}] \end{aligned}$$

(Tower property)

$$\begin{aligned} \text{On pose } W_i &\triangleq \sup_{\mathbb{Z}} \mathbb{E}[h | Z_1, \dots, Z_i = z] - \mathbb{E}[h | F_{i-1}] \\ X_i &\triangleq \inf_{\mathbb{Z}'} \mathbb{E}[h | Z_1, \dots, Z_i = z'] - \mathbb{E}[h | F_{i-1}] \end{aligned}$$

On obtient donc $X_i \leq V_i \leq W_i$.

De plus,

$$W_i - X_i \leq \sup_z \sup_{z'} E[h(z_1, \dots, z_i = z)] - E[h(z_1, \dots, z_i = z')]$$
$$\left(= \sup_z \sup_{z'} \int_{Z_{i+1}, \dots, Z_n} h(z_1, \dots, z_i, z_{i+1}, \dots, z_n) - h(z_1, \dots, z'_i, z_{i+1}, z_n) dP \right)$$
$$\leq \sup_z \sup_{z'} \int_{Z_{i+1}, \dots, Z_n} c_i dP = c_i$$

On en conclut que $X_i \leq V_i \leq X_i + c_i$. Comme X_i est une fonction des Z_i , appliquer Azuma donne le résultat désiré.

Ex 3

$$\sup_{-\sum \epsilon_i} \sum \leq \sum \sup$$

$$\text{Soit } h(z_1, \dots, z_n) = \widehat{R}_n(F) = \mathbb{E} \left[\sup_{f \in F} \frac{1}{n} \sum \epsilon_i f(z_i) \mid D_n \right]$$

On a :

$$\sup_{z_1, \dots, z_n, z'_i} |h(z_1, \dots, z_n) - h(z_1, \dots, z'_i, \dots, z_n)|$$

$$= \sup_{(z_i), z'_i} \left| \mathbb{E} \left[\sup_{f \in F} \frac{1}{n} \sum \epsilon_i f(z_i) - \sup_{f \in F} \frac{1}{n} \sum_{k=1}^n \epsilon_k f(z'_k) \mid D_n \right] \right|$$

$z'_n = z_n, \forall n \neq i$

①

Notons f^* la fonction qui maximise le sup dans ①. On a alors

$$\frac{1}{n} \sum \epsilon_k f^*(z'_k) \leq \sup_{f \in F} \frac{1}{n} \sum \epsilon_k f(z'_k). \text{ Ainsi :}$$

$$\left(\sup_{(z_i), z'_i} \left| \mathbb{E} \left[\frac{1}{n} \sum \epsilon_i (f^*(z_i) - f^*(z'_i)) \mid D_n \right] \right| \right)$$

$$= \sup_{z_i, z'_i} \left| \mathbb{E} \left[\frac{1}{n} \epsilon_i (f^*(z_i) - f^*(z'_i)) \mid D_n \right] \right|$$

$\left(\frac{1}{n} (f^* \text{ est } \{0,1\}-\text{valuee.}) \right)$

On applique Mc Diarmid :

$$P(\widehat{R}_n(F) - R_n(F) < -t) \leq \exp \left(\frac{-2t^2}{\sum c_i^2} \right)$$

$$\Leftrightarrow \widehat{R}_n(F) - R_n(F) > -t, \text{ a.p.} \geq 1 - \exp \left(\frac{-2t^2}{\sum \frac{1}{n^2}} \right) = 1 - \exp(-2t^2/n^2)$$

$$\Leftrightarrow R_n(F) \leq \bar{R}_n(F) + t, \quad a.p. \geq -e^{-2t^2 n}.$$

On pose $\delta \triangleq e^{-2t^2 n}$ et on cherche t en fonction de δ :

$$2t_n^2 = \ln(\frac{1}{\delta}) \Rightarrow t = \left(\frac{\ln(\frac{1}{\delta})}{2n} \right)^{\frac{1}{2}}.$$

□

2) On pose $\phi(z_1, \dots, z_n) = \sup_{f \in F} (\mathbb{E}[f(z_1)] - \frac{1}{n} \sum f(z_i))$

$$|\phi(z_1, \dots, z_n) - \phi(z_1, \dots, z'_1, \dots)| \leq \sup_f \frac{1}{n} (-f(z_i) + f(z'_i)) \leq \frac{1}{n}.$$

On peut appliquer McDiarmid:

$$\phi(z_1, \dots, z_n) \leq \mathbb{E}[\phi(z_1, \dots, z_n)] + \sqrt{\frac{\ln(\frac{1}{\delta})}{2n}}, \quad a.p. \geq -\frac{\delta}{2}.$$

Majorons $\mathbb{E}[\phi]$:

$$\begin{aligned} \mathbb{E}[\phi] &= \mathbb{E}\left[\sup_f \left(\mathbb{E}[f(z_1)] - \frac{1}{n} \sum f(z_i)\right)\right] \\ &= \mathbb{E}\left[\sup_f \mathbb{E}_{z'} \left[\frac{1}{n} \sum f(z'_i) - \frac{1}{n} \sum f(z_i) \right]\right] \\ &\leq \mathbb{E}\left[\sup_f \frac{1}{n} \sum f(z'_i) - f(z_i)\right] \quad \sup \mathbb{E} \leq \mathbb{E} \sup \text{ (à)} \\ &\quad \xrightarrow{s: x \mapsto d} \mathbb{E}(f(x) - f(x')) \\ &\quad = \mathbb{E}(f(x) - f(x')). \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E}_{z, z', \varepsilon} \left[\sup_{f} \frac{1}{n} \sum \varepsilon_i (f(z'_i) - f(z_i)) \right] \\
&\leq \mathbb{E}_{\varepsilon, z} \left[\sup_f \frac{1}{n} \sum \varepsilon_i f(z'_i) \right] + \mathbb{E}_{\varepsilon, z} \left[\sup_f \frac{1}{n} \sum \varepsilon_i f(z_i) \right] \\
&= \mathbb{E}_{\varepsilon, z} \left[\sup_f \frac{1}{n} \sum \varepsilon_i f(z'_i) \right] + \mathbb{E}_{\varepsilon, z} \left[\sup_f \frac{1}{n} \sum \varepsilon_i f(z_i) \right] \\
&= 2 R_n(F).
\end{aligned}$$

Ainsi, on a $\phi(z_1, \dots, z_n) \leq 2R_n(F) + \sqrt{\frac{\ln(2/\delta)}{2n}}$, avec probabilité d'au moins $1 - \frac{\delta}{2}$.

De plus, avec proba $1 - \frac{\delta}{2}$, on a que $R_n(F) \leq \bar{R}_n(F) + \left(\frac{\ln(2/\delta)}{2n}\right)^{1/2}$.

En combinant ces 2 inégalités, on obtient

$$\phi(z_1, \dots, z_n) \leq 2\bar{R}_n(F) + 3\sqrt{\frac{\ln(2/\delta)}{2n}}.$$

Exh

$$\gamma(C, n) = \max_{\substack{X \\ |X|=n}} |C \cap X| = |\{c \in C \mid c \in X\}|$$

Lemme de Sauer :

$$\forall C, VC(C) = V \text{ ou } +\infty, \quad \forall n, \quad \gamma(C, n) \leq \sum_{i=0}^V \binom{n}{i}$$

Preuve :

On procède par récurrence sur $n+V$. Si $n=0$, alors

$$\gamma(C, n) = |\{c \in C \mid c \cap \emptyset \neq \emptyset\}| = |\{\emptyset\}| = 1 = \sum_{i=0}^V \binom{0}{i}$$

Si $V=1$, alors C ne peut contenir aucun point : $\gamma(C, n)=0$.

Supposons que la propriété soit vérifiée pour tout $n'+V \leq n+V$.

Posons C un ensemble d'ensembles de \mathbb{R}^d tq $VC(C) = V$.

Soit S un ensemble de n points. Posons C_1, C_2 tq :

$$C = C_1 \triangleq \{c \in C \mid s \notin c\} \cup \{c \cup \{s\} \mid c \in C\}$$

$$C_2 \triangleq \{c \cup \{s\} \in C \mid c \in C\}. \quad \text{X}$$

En posant $\gamma(C, S) = |C \cap S|$, on a que

$$\gamma(C, S) \leq \gamma(C_1, S) + \gamma(C_2, S).$$

De plus, $\Upsilon(C, S \setminus \{s\}) = \Upsilon(C_1, S)$

En effet, $\forall c \in C$, on veut que tout unique ensemble $c \cap S \setminus \{s\}$ soit associé à un unique ensemble $c \cap S$, $c' \in C_1$.

- Si $s \notin c$, alors $c \in C_1$ et $c \cap S \setminus \{s\} = c \cap S$.

- Si $s \in c$, alors

- Si $c \setminus \{s\} \in C$, alors $c \setminus \{s\} \in C_1$ et $c \cap S \setminus \{s\} = c \setminus \{s\} \cap S$

- Si $c \setminus \{s\} \notin C$, alors $c \in C_1$ on mappe $c \cap S \setminus \{s\} \mapsto c \cap S$

On a donc qu'il existe une bijection entre $C \cap S \setminus \{s\}$ et $C_1 \cap S$.

Notons C' l'ensemble des c qui constituent C_1 :

$$C' = \{c \in C \mid s \notin c, c \cup \{s\} \in C\}$$

De la même manière que précédemment, $\Upsilon(C_1, S) =$

$$\Upsilon(C', S \setminus \{s\})$$

Ainsi, si C' pulvérise un échantillon, $C(C_2)$ le pulvérise également, même si on lui rapporte $\{s\}$. On a alors :

$$VC(C') + 1 \leq VC(C) \Leftrightarrow VC(C') \leq V - 1.$$

Par hypothèse de récurrence, on a donc, $\forall S \text{ tq } |S|=n$:

$$\gamma(C, S) \leq \gamma(C_1, S) + \gamma(C_2, S) = \gamma(C, S \setminus \{s\}) + \gamma(C', S \setminus \{s\})$$

$$\begin{aligned} &\leq \sum_{i=0}^V \binom{n-1}{i} + \sum_{i=0}^{V-1} \binom{n-1}{i} \\ &= \sum_{i=0}^V \binom{n}{i} \end{aligned}$$

- ④ On découpe C en deux : l'ensemble des éléments de C qui ne contiennent pas s : $\{c \in C \mid s \notin c\}$ et l'ensemble des éléments de C qui contiennent s . On coupe le dernier en deux : $\{c \cup \{s\} \mid c \in C\}$
- $\{c \cup \{s\} \mid c \notin C\}$

Ex 5

On commence par chercher une borne inf à $V(A)$.

On choisit $d+1$ points tels que x_0 soit l'origine et $\{x_1, \dots, x_d\} = \{e_1, \dots, e_d\}$. On pose aussi $\{y_0, \dots, y_d\}$ un ensemble d'échantillons arbitraire associé à $\{x_0, \dots, x_d\}$.

Posons $w = (y_1, \dots, y_d)^T$. Ainsi, l'hyperplan d'égalation $wx + \frac{y_0}{2} = 0$ sépare le $\{x_0, \dots, x_d\}$ car, $\forall i \in \{1, \dots, d\}$:

$$\operatorname{sgn}(wx_i + \frac{y_0}{2}) = \operatorname{sgn}(y_i + \frac{y_0}{2}) = \operatorname{sgn}(y_i)$$

et $\operatorname{sgn}(wx_0 + \frac{y_0}{2}) = \operatorname{sgn}(\frac{y_0}{2}) = \operatorname{sgn}(y_0)$.

Donc $d+1 \leq V(A)$.

On cherche maintenant une borne supérieure en utilisant le théorème de Radon.

Tout ensemble X de $d+2$ points de \mathbb{R}^d peut être partitionné en 2 ensembles X_1, X_2 tels que

$$X_1 \cap X_2 = \emptyset, \quad X_1 \cup X_2 = X \text{ et } \operatorname{conv}(X_1) \cap \operatorname{conv}(X_2) \neq \emptyset.$$

Posons X un ensemble arbitraire de $d+2$ points. Par le théorème de Radon, on peut partitionner X en 2 ensembles

de manière à ce que leur enveloppe convexe s'intersecte). Or, si deux ensembles X_1 et X_2 peuvent être séparés par un hyperplan, alors leur enveloppe convexe peut aussi être séparée par ce même hyperplan^(*). Ainsi, la décomposition fournie par le théorème de Radon ne pourra pas être séparée par un hyperplan, donc X ne sera pas pulvrisé.

On en conclut $d+1 \leq V(A) \leq d+2 \Leftrightarrow V(A) = d+1$.

Si X , il existe une séparation (celle de Radon) qui n'est pas achievable par un hyperplan

(*) Si X_1 et X_2 sont séparés par un k -plan, alors, on peut utiliser cet k -plan pour construire l'enveloppe convexe. Ainsi, les deux enveloppes sont séparées par le même k -plan.

Ex 7)

$$r(g, n) = \max_{(x_i)_{i \in [n]} \in \mathbb{R}^d} |\{ (g(x_i))_{i \in [n]} \mid g \in \mathcal{G} \}|$$

Nombre maximum de manières de classer n points en utilisant des fonctions dans \mathcal{G}

$$1) \bar{R}_n(g_g) = \mathbb{E}_{D_n} \left[\mathbb{E} \left[\sup_{g \in \mathcal{G}} \frac{1}{n} \sum_{i \in [n]} \epsilon_i g(z_i) \mid D_n \right] \right].$$

Posons $\mathcal{G}_{|D_n}$ l'ensemble des vecteurs $(g(z_1), \dots, g(z_n))^T$ pour toute fonction dans $\mathcal{G} : \mathcal{G}_{|D_n} = \{ (g|_{D_n})^T = \{ (g(z_i))_{i \in [n]} \mid g \in \mathcal{G} \} \}$.

On obtient

$$\bar{R}_n(g_g) = \mathbb{E}_{D_n} \left[\mathbb{E} \left[\sup_{z \in \mathcal{G}_{|D_n}} \frac{1}{n} \sum \epsilon_i z_i \mid D_n \right] \right].$$

Comme $g \in \mathcal{G}$ est à valeurs dans $\{-1, 1\}$,

$$\max_{z \in \mathcal{G}_{|D_n}} \|z\|_2 \leq \sqrt{n}.$$

On peut donc appliquer le lemme de Massart :

$$\mathbb{E}_{D_n} \left[\mathbb{E} \left[\sup_{z_i \in \mathcal{G}_{D_n}} \frac{1}{n} \sum_i z_i | D_n \right] \right] \leq \mathbb{E}_{D_n} \left[\frac{\sqrt{n} \sqrt{2 \ln |\mathcal{G}_{D_n}|}}{n} \right].$$

Or, $\mathcal{G}_{D_n} \subseteq \{(g(x_i))_{i \leq n} \mid (x_i)_{i \leq n} \in \mathbb{R}^d, g \in \mathcal{G}\}$. Ainsi,

$$|\mathcal{G}_{D_n}| \leq \max_{(x_i)_{i \leq n} \in \mathbb{R}^d} \left| \{f(g(x_i))_{i \leq n} \mid g \in \mathcal{G}\} \right| = \mathcal{J}(g, n).$$

On en déduit

$$\begin{aligned} \mathbb{E}_{D_n} \left[\frac{\sqrt{n} \sqrt{2 \ln |\mathcal{G}_{D_n}|}}{n} \right] &\leq \frac{\sqrt{n} \sqrt{2 \ln \mathcal{J}(g, n)}}{n} \\ &= \left(\frac{2 \ln \mathcal{J}(g, n)}{n} \right)^{1/2}. \end{aligned}$$

On suppose $V(g) < +\infty$.

Si $n \leq V(g)$, $\mathcal{J}(g, n) = 2^n$. Ainsi

$$\mathbb{E}_n(g) \leq \left(\frac{2 \ln(2^n)}{n} \right)^{1/2} \leq \left(\frac{2 \log_2 2^n}{n} \right)^{1/2} = \sqrt{2}$$

Si $n > V(G)$:

$$\chi(G_{\frac{n}{V(G)}}) \leq \left(\frac{e_n}{\sqrt{V(G)}} \right)^{V(G)}. \text{ Posons } V = V(G)$$

On en déduit

$$\overline{\chi}_{\frac{n}{V}}(G) \leq \left(\frac{2 \ln \left(\frac{e_n}{\sqrt{V}} \right)^V}{n} \right)^{\frac{1}{2}}$$

$$= \left(\frac{2 \left(V + \sqrt{V} \ln \left(\frac{n}{V} \right) \right)}{n} \right)^{\frac{1}{2}}$$

$$= \left(\frac{2 V \left(\ln \frac{n}{V} + 1 \right)}{n} \right)^{\frac{1}{2}}$$

$$= \left(\frac{\sqrt{V} \ln \left(\frac{n}{V} \right)}{n} \left(2 + \frac{2}{\ln \left(\frac{n}{V} \right)} \right) \right)^{\frac{1}{2}}$$

$$= \left(\frac{\sqrt{V} \ln n}{n} \left(2 - 2 \frac{\ln V}{\ln n} + \frac{2}{\ln n} \right) \right)^{\frac{1}{2}}$$

$$= \left(\frac{\sqrt{V} \ln n}{n} \right)^{\frac{1}{2}} \left(2 - 2 \frac{\ln V}{\ln n} + \frac{2}{\ln n} \right)^{\frac{1}{2}}$$

$$\left(\log_2 \frac{\ln n}{\ln 2} \geq \frac{\ln n}{\ln 2} \geq \frac{\ln n}{\ln n} \right) \leq \left(\frac{\sqrt{V} \log_2 n}{n} \right)^{\frac{1}{2}} C$$