Apprentissage statistique

Chapitre 4 : Consistance des algorithmes par moyennage local

Lucie Le Briquer

19 février 2018

Table des matières

Convention $0/0 = 0$.

Définition 1 (règle d'apprentissage par moyennage local)

Une règle d'apprentissage par moyennage local est caractérisée par un tableau de fonctions mesurables $(\omega_{i,n})_{i\in\{1,\ldots,n\}}$, $\forall i, n$ $\omega_{i,n} : \mathbb{X}^n \times \mathbb{X} \longrightarrow \mathbb{R}_+$.
On définit la fonction de régression associée $\hat{\eta}$ définie pour tout $n \geqslant 1$, $(x_i, y_i)_{i\in\{1,\ldots,n\}}$ et $x \in \mathbb{X}$:

$$\hat{\eta}\big((x_i, y_i), x\big) = \sum_{i=1}^{n} \omega_{i,n}((x_i)_i, x) y_i$$

On définit alors la règle de classification $\hat{f}$ comme étant la règle plug-in associée à $\hat{\eta}$ définie pour tout $n \geqslant 1$ par :

$$\hat{f}\big((x_i, y_i), x\big) = \mathbb{1}_{\{\hat{\eta}((x_i,y_i),x) > \frac{1}{2}\}}$$

On rappelle la règle par partition : on se donne une suite de partition $(P_n)_{n\in\mathbb{N}}$ telle que :

1. P_n est de cardinal dénombrable.
2. Tout élément de P_n est mesurable.

La règle pour la régression par partition est donnée par :

$$\hat{\eta}((x_i, y_i), x) = \frac{\sum_{i=1}^{n} y_i \mathbb{1}_{P_n(x)}(x_i)}{\sum_{i=1}^{n} \mathbb{1}_{P_n(x)}(x_i)}$$

$P_n(x)$: élément de P_n qui contient x. C'est donc une règle de type voisinage avec les poids $(\omega_{i,n})$ définis par :

$$\omega_{i,n}((x_i), x) = \frac{\mathbb{1}_{P_n(x)}(x_i)}{\sum_{i=1}^{n} \mathbb{1}_{P_n(x)}(x_i)}$$

Remarques.

1. $(\omega_{i,n})$ sont appelés les poids.
2. En générale on note $\omega_{i,n}((x_i), x)$ par $\omega_{i,n}(x)$.
3. Pour l'algorithme de partition $\sum_{i=1}^{n} \omega_{i,n}(x) = \mathbb{1}\{\sum_{i=1}^{n} \mathbb{1}_{P_n(x)}(x_i) > 0\} \leqslant 1$.

1 Théorème de Stone

Théorème 1 (de Stone)

On considère $\mathbb{X} = \mathbb{R}^d$, $\mathbb{Y} = \{0,1\}$. Soit $\|.\|$ sur $\mathbb{R}^d$, (ω_{in}) suite de poids pour une règle d'apprentissage par voisinage local. Soit $\mathbb{P} \in \mathcal{P}$ et $(X_i, Y_i)_{i\in\mathbb{N}}$ i.i.d. de loi $\mathbb{P}$ et $(X,Y) \perp (X_i, Y_i)$ de même loi $\mathbb{P}$. On suppose :

1. $\sum_{i=1}^n \omega_{i,n}(X_{1:n}, X) \xrightarrow[n\to+\infty]{} 1$ dans $\mathcal{L}^1(\mathbb{P})$. Il existe $c_1 \geqslant 0$ tel que :
$$\sum_{i=1}^n \omega_{i,n}(X_{1:n}, X) \leqslant c_1 \quad \mathbb{P} - \text{p.s.}$$
2. Pour tout $a > 0$,
$$\lim_{n\to+\infty} \mathbb{E}\left[\sum_{i=1}^n \omega_{i,n}(X_{1:n}, X) \mathbb{1}\{\|X_i - X\| \geqslant a\}\right] = 0$$
3. $$\lim_{n\to+\infty} \mathbb{E}\left[\max_{1\leqslant i\leqslant n} \omega_{i,n}(X_{1:n}, X)\right] = 0$$
4. Il existe $c_2 \geqslant 0$ tel que pour toute fonction f telle que $\mathbb{E}[|f(X)|] < +\infty$ on ait :
$$\mathbb{E}\left[\sum_{i=1}^n \omega_{i,n}(X_{1:n}, X)|f(X_i)|\right] \leqslant c_2 \mathbb{E}[|f(X_1)|]$$

Alors la règle par voisinage local associée à $(\omega_{i,n})$ est faiblement consistante pour $\mathbb{P}$:
$$\lim_{n\to+\infty} \mathbb{E}\left[R_{\mathbb{P}}^{D_n}(\hat{f}_n)\right] = R_{\mathbb{P}}^*$$

Remarque. $\hat{f}\big((x_i, y_i)_{i\in\{1,\ldots,n\}}, x\big) = \hat{f}_n(x)$ pareil pour $\hat{\eta}$.

Preuve.
D'après le deuxième cours,
$$|R_{\mathbb{P}}^{D_n}(\hat{f}_n) - R_{\mathbb{P}}^*| \leqslant 2\mathbb{E}\left[|\eta(X) - \hat{\eta}_n(X) \mid D_n\right]$$
Donc :
$$\left|\mathbb{E}\left[R_{\mathbb{P}}^{D_n}(\hat{f}_n)\right] - R_{\mathbb{P}}^*\right| \leqslant 2\mathbb{E}[|\eta(X) - \hat{\eta}_n(X)|]$$
Donc on montre que $\lim \mathbb{E}[|\eta(X) - \hat{\eta}_n(X)|] = 0$.
On introduit $\tilde{\eta}\colon \mathbb{X}^n \times \mathbb{X} \longrightarrow [0,1]$ définit pour tout $x_1, \ldots, x_n, x \in \mathbb{X}$ par :
$$\tilde{\eta}((x_i), x) = \sum_{i=1}^n \omega_{i,n}((x_i), x)\eta(x_i)$$

Soit $n \geqslant 1$, $(X_i, Y_i) \sim \mathbb{P}$ i.i.d. $\perp$ $(X, Y) \sim \mathbb{P}$

$$\mathbb{E}[|\eta(X) - \hat{\eta}_n(X)|] \leqslant \mathbb{E}[|\eta(X) - \tilde{\eta}_n(X)|] + \mathbb{E}[|\tilde{\eta}_n(X) - \hat{\eta}_n(X)|] = A + B$$

Par Cauchy-Schwarz, on a :

$$\begin{aligned}
B^2 &\leqslant \mathbb{E}[(\tilde{\eta}_n(X) - \hat{\eta}_n(X))^2] \\
&= \sum_{i,j} \mathbb{E}\big[\omega_{i,n}(X)\omega_{j,n}(X)(Y_i - \eta(X_i))(Y_j - \eta(X_j))\big] \\
&= \sum_{i=0}^{n} \mathbb{E}\big[\omega_{i,n}(X)^2(Y_i - \eta(X_i))^2\big] + \sum_{i \neq j = 1}^{n} \mathbb{E}\big[\mathbb{E}[\dots | X_{1:n}, X]\big] \\
&= \sum_{i=0}^{n} \mathbb{E}\Big[\omega_{i,n}(X)^2(Y_i - \eta(X_i))^2\big] \\
&\quad + \sum_{i \neq j = 1}^{n} \mathbb{E}\big[\omega_{i,n}(X)\omega_{j,n}(X)\mathbb{E}[(Y_i - \eta(X_i))(Y_j - \eta(X_j)) \mid X_{1:n}, X]\Big] \\
&= \sum_{i=0}^{n} \mathbb{E}\big[\omega_{i,n}(X)^2(Y_i - \eta(X_i))^2\big] \\
&\quad + \sum_{i \neq j = 1}^{n} \mathbb{E}\big[\omega_{i,n}\omega_{j,n}\mathbb{E}[Y_i - \eta(X_i) | X_{1:n}, X]\mathbb{E}[Y_j - \eta(X_j) | X_{1:n}, X]\big] \\
&= \sum_{i=0}^{n} \mathbb{E}\big[\omega_{i,n}(X)^2(Y_i - \eta(X_i))^2\big] \\
&\leqslant 2\sum_{i=1}^{n} \mathbb{E}[\omega_{i,n}(X)^2] \\
&\leqslant 2\mathbb{E}\left[\max_{1 \leqslant j \leqslant n} \omega_{j,n}(X) \sum_{i=1}^{n} \omega_{i,n}(X)\right] \\
&\leqslant 2c_1\mathbb{E}\left[\max_{1 \leqslant j \leqslant n} \omega_{j,n}(X)\right]
\end{aligned}$$

On se donne η^ε pour $\varepsilon > 0$ fixé tel que :

1. η^ε est uniformément continue sur $\mathbb{R}^d$
2. $\mathbb{E}[|\eta(X) - \eta^\varepsilon(X)|] \leqslant \varepsilon$

On peut sans perte de généralité supposer que pour tout $\varepsilon > 0$ $\sup_{\mathbb{R}^d} |\eta^\varepsilon| \leqslant 2$ (à faire en exercice).
On définit aussi $\tilde{\eta}^\varepsilon$ par :

$$\tilde{\eta}_n^\varepsilon(x) = \sum_{i=1}^{n} \eta^\varepsilon(x_i)\omega_{i,n}((x_i), x)$$

$$\begin{aligned}
A &= \mathbb{E}\big[|\eta(X) \pm \eta^\varepsilon(X) \pm \tilde{\eta}_n^\varepsilon(X) - \tilde{\eta}_n(X)|\big] \\
&\leqslant \mathbb{E}[|\eta(X) - \eta^\varepsilon(X)|] + \mathbb{E}[|\tilde{\eta}_n^\varepsilon(X) - \tilde{\eta}_n(X)|] + \mathbb{E}[|\eta^\varepsilon(X) - \tilde{\eta}_n^\varepsilon(X)|] \\
&\leqslant \varepsilon + c_1\varepsilon + \underbrace{\mathbb{E}[|\tilde{\eta}_n^\varepsilon(X) - \tilde{\eta}_n(X)|]}_{C}
\end{aligned}$$

Pour $\varepsilon > 0$ fixé, il existe $a > 0$ tel que $\forall x, \tilde{x} \in \mathbb{R}^d$ $|\eta^\varepsilon(x) - \eta^\varepsilon(\tilde{x})| \leqslant \varepsilon$.

$$\begin{aligned}
C &= \mathbb{E}\left[|\eta^\varepsilon(X) - \sum \omega_{i,n}(X)\eta^\varepsilon(X_i)|\right] \\
&\leqslant \mathbb{E}\left[\left|1 - \sum_{i=1}^n \omega_{i,n}(X)\right| |\eta^\varepsilon(X)|\right] + E\left[\left|\sum_{i=1}^n \omega_{i,n}(X)(\eta^\varepsilon(X) - \eta^\varepsilon(X_i))\right|\right] \\
&\leqslant 2\mathbb{E}\left[\left|1 - \sum_{i=1}^n \omega_{i,n}(X)\right|\right] + \mathbb{E}\left[\left|\sum_{i=1}^n \omega_{i,n}(X)(\eta^\varepsilon(X) - \eta^\varepsilon(X_i))\left(\mathbb{1}_{\{\|X_i - X\| \geqslant a\}} + \mathbb{1}_{\{\|X_i - X\| \leqslant a\}}\right)\right|\right] \\
&\leqslant 2\mathbb{E}\left[\left|1 - \sum_{i=1}^n \omega_{i,n}(X)\right|\right] + 2\mathbb{E}\left[\left|\sum_{i=1}^n \omega_{i,n}(X)\mathbb{1}_{\{\|X_i - X\| \geqslant a\}}\right|\right] + \varepsilon c_1
\end{aligned}$$

Conclusion, pour tout $\varepsilon > 0$,

$$A + B \leqslant \sqrt{2c_1}\sqrt{\mathbb{E}[\max \omega_{i,n}]} + (1 + c_2)\varepsilon + c_1\varepsilon + \mathbb{E}\left[\left|1 - \sum_{i=1}^n \omega_{i,n}\right|\right] + 2\mathbb{E}\left[\left|\sum_{i=1}^n \omega_{i,n}\mathbb{1}_{\{\|X_i - X\| \geqslant a\}}\right|\right]$$

Donc $\limsup A + B \leqslant (1 + c_1 + c_2)\varepsilon \ \forall \varepsilon > 0$. □

2 Application du théorème de Stone aux algorithmes par partitions

Soit $(P_n)_n$ une suite de partitions mesurables et dénombrables de $\mathbb{R}^d$. On suppose :

$$(1) \qquad \lim_{n \to +\infty} \sup_{A \in P_n} \operatorname{diam}(A) = 0$$

$$(2) \qquad \lim_{n \to +\infty} \frac{\operatorname{Card}(A \in P_n | A \cap \mathcal{B}(0, r) \neq \emptyset)}{n} = 0 \qquad \text{pour tout } r > 0$$

Remarque. La norme que l'on prend pour définir le diamètre et $\mathcal{B}(0, r)$ est quelconque.

Proposition 1

Si (P_n) vérifie (1) et (2), alors la règle par partitions associée à (P_n) est universellement faiblement consistante.

Preuve.
On vérifie que les hypotèses du théorème de Stone sont satisfaites.

1. Par la remarque du début on a :

$$\sum_{i=1}^n \omega_{i,n}(x_{1:n}, x) = \sum_{i=1}^n \frac{\mathbb{1}_{P_n(x)}(x_i)}{\sum_{i=1}^n \mathbb{1}_{P_n(x)}(x_i)} = \mathbb{1}_{\{N_{P(x)}(x_{1:n}) > 0\}}$$

   où $N_A(x_{1:n}) = \sum_{i=1}^n \mathbb{1}_A(x_i) \ \forall A \subset \mathbb{R}^d$.

Soit $\mathbb{P}_X$ la loi de X sur $\mathbb{R}$. Pour tout $\varepsilon > 0$ il existe $r > 0$ tel que

$$\mathbb{P}_X(\mathcal{B}(0,r)) \geqslant 1-\varepsilon \;\Leftrightarrow\; \mathbb{P}_X(\mathcal{B}(0,r)^C) \leqslant \varepsilon$$

Soit $\varepsilon > 0$ et $r_\varepsilon > 0$ fixé. On considère :

$$P_n^\varepsilon = \{A \in P_n \;:\; A \cap \mathcal{B}(0,r_\varepsilon) \neq \emptyset\}$$

$$\begin{aligned}
\mathbb{E}\left[\left|1-\sum_{i=1}^n \omega_{i,n}(X)\right|\right] &= \mathbb{P}(N_{P(X)}(X_{1:n})=0)\\
&= \mathbb{P}\big(P(X) \in P_n^\varepsilon,\; N_{P(X)}(X_{1:n})=0\big) + \mathbb{P}\big(P(X) \notin P_n^\varepsilon,\; N_{P(X)}(X_{1:n})=0\big)\\
&= A+B
\end{aligned}$$

$$B \leqslant \mathbb{P}(P(X) \notin P_n^\varepsilon) \leqslant \mathbb{P}(X \notin \mathcal{B}(0,r)) \leqslant \varepsilon$$

Et,

$$A = \sum_{C \in P_n^\varepsilon} \mathbb{P}(P_n(X)=C,\; N_C(X_{1:n})=0)$$

Or $\forall C \in P_n^\varepsilon$, $N_C(X_{1:n}) = \sum_{i=1}^n \mathbb{1}_C(X_i) \sim \mathcal{B}(n, \mathbb{P}(X \in C))$. Alors,

$$\begin{aligned}
A &= \sum_{C \in P_n^\varepsilon} \mathbb{P}(X \in C)(1-\mathbb{P}(X \in C))^n\\
&\leqslant \left[\sup_{t \in [0,1]} t(1-t)^n\right] \mathrm{Card}(P_n^\varepsilon)\\
&\leqslant \frac{\mathrm{Card}(P_n^\varepsilon)}{n}
\end{aligned}$$

Par (2), on a $\forall \varepsilon > 0$,

$$\limsup_{n\to+\infty} \mathbb{E}\left[|1-\sum_{i=1}^n \omega_{1,n}|\right] \leqslant \varepsilon$$

On a donc montré la condition (1) du théorème de Stone.

2. Soit $a > 0$.

$$\omega_{i,n}(X)\mathbb{1}_{\{\|X_i-X\| \geqslant a\}} \leqslant \omega_{i,n}(X)\mathbb{1}_{\{\mathrm{diam}(P_n(X) \geqslant a\}}$$

Donc

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^n \omega_{i,n}(X_{i:n},X)\mathbb{1}_{\{\|X_i-X\|\geqslant a\}}\right] &\leqslant \mathbb{E}\left[\mathbb{1}_{\{\mathrm{diam}(P_n(X)\geqslant a)\}}\sum_{i=1}^n \omega_{i,n}\right]\\
&\underset{\text{Markov}}{\leqslant} \frac{\big(\sup_{A\in P_n}\mathrm{diam}(A)\big)}{a} \xrightarrow[n\to+\infty]{} 0 \text{ par hyp (1)}
\end{aligned}$$

3. Montrons la condition (3). Soit $\varepsilon > 0$ et r_ε tel que $\mathbb{P}(X \notin \mathcal{B}(0,r)) \leqslant \varepsilon$.

$$\begin{aligned}
\mathbb{E}[\max_i \omega_{i,n}] &\leqslant \mathbb{E}\left[\mathbb{1}_{P_n^\varepsilon}(P_n(X))\max \omega_{i,n}\right] + \varepsilon \quad \text{même raisonnement qu'avant}\\
&\leqslant \sum_{C\in P_n^\varepsilon} \mathbb{E}\left[\mathbb{1}_{\{P_n(X)=c\}} \max_i \frac{\mathbb{1}_C(X_i)}{\sum_j \mathbb{1}_C(X_j)}\right] + \varepsilon\\
&\leqslant \sum_{C\in P_n^\varepsilon} \mathbb{E}\left[\mathbb{1}_{\{P_n(X)=C\}} \frac{\mathbb{1}_{\{N_C(X_{1:n})>0\}}}{N_C(X_{1:n})}\right] + \varepsilon
\end{aligned}$$

Lemme : si $Z \sim \mathcal{B}(n,p)$:

$$\mathbb{E}\left[\frac{\mathbb{1}_{\{Z>0\}}}{Z}\right] \leqslant \frac{2}{p(n+1)}$$

Retour à la preuve :

$$\leqslant \sum_{C \in P_n^\varepsilon} \mathbb{P}(X \in C)\frac{2}{\mathbb{P}(X \in C)(n+1)} + \varepsilon$$

Comme $X \perp (X_i)$ et $N_C(X_{1:n}) \sim \mathcal{B}(n, \mathbb{P}(X \in C))$ + lemme. Finalement :

$$\mathbb{E}[\max \omega_{i:n}] \leqslant 2\frac{\mathrm{Card}(P_n^\varepsilon)}{n+1} + \varepsilon$$

Par l'hypothèse (2), $\forall \varepsilon > 0$ on a :

$$\limsup_{n \to +\infty} \mathbb{E}[\max \omega_{i,n}] \leqslant \varepsilon$$

Ce qui conclut.

4. Montrons que l'hypothèse (4) du théorème de Stone est vérifiée. Soit f tel que $\mathbb{E}[|f(X)|] < +\infty$.

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^n \omega_{i,n}(X)|f(X_i)|\right] &= \mathbb{E}\left[\sum_{i=1}^n \frac{\mathbb{1}_{P_n(X)}(X_i)}{N_{P_n(X)}(X_{1:n})}|f(X_i)|\right] \\
&= \sum_{i=1}^n \mathbb{E}\left[\frac{\mathbb{1}_{P_n(X)}(X_i)}{N_{P_n(X)}(X_{1:n})}|f(X_i)|\right] \\
&= \sum_{i=1}^n \mathbb{E}\left[\frac{\mathbb{1}_{P_n(X)}(X_i)}{N_{P_n(X_i)}(X_1, \ldots, X_{i-1}, X, X_{i+1}, \ldots, X_n)}|f(X_i)|\right]
\end{aligned}$$

Car les (X_i) i.i.d. $\perp$ X de même loi. Or, $\forall i$:

$$\frac{\mathbb{1}_{P_n(X_i)}(X)}{N_{P_n(X_i)}(X_1, \ldots, X, \ldots, X_n)} = \omega_{i,n}(X)$$

car $\mathbb{1}_{P_n(X)}(X_i) = 0$ ssi $\mathbb{1}_{P_n(X_i)}(X) = 0$ et si $\mathbb{1}_{P_n(X)}(X_i) = 1 = \mathbb{1}_{P_n(X_i)}(X)$ on a par définition :

$$N_{P_n(X)}(X_{1:n}) = N_{P_n(X_n)}(X_1, \ldots, X, \ldots, X_n)$$

Donc :

$$\mathbb{E}\left[\sum_{i=1}^n \omega_{i,n}|f(X_i)|\right] = \mathbb{E}\left[|f(X)|\sum_{i=1}^n \omega_{i,n}\right] \leqslant \mathbb{E}[|f(X)|]$$

Conclusion, on peut appliquer le théorème de Stone. □

Preuve. (du lemme utilisé)

$Z \sim \mathcal{B}(n,p)$

$$\mathbb{E}\left[\frac{\mathbb{1}_{\{Z>0\}}}{Z}\right] \leqslant \frac{2}{p(n+1)}$$

Comme $\frac{\mathbb{1}_{\{Z>0\}}}{Z} \leqslant \frac{2}{Z+1}$,

$$\mathbb{E}\left[\frac{\mathbb{1}_{\{Z>0\}}}{Z}\right] \leqslant 2\mathbb{E}\left[\frac{1}{Z+1}\right]$$

Or,

$$\begin{aligned}
\mathbb{E}[(Z+1)^{-1}] &= \sum_{k=0}^{n}(k+1)^{-1}\binom{n}{k}p^k(1-p)^{n-k} \\
&= \frac{1}{(n+1)p}\sum_{k=0}^{n}\frac{n+1}{k+1}\binom{n}{k}p^{k+1}(1-p)^{n-k} \\
&= \frac{1}{(n+1)p}\sum_{k=0}^{n}\binom{n+1}{k+1}p^{k+1}(1-p)^{n-k} \\
&= \frac{1}{(n+1)p}\sum_{k'=1}^{n+1}\binom{n+1}{k'}p^{k'}(1-p)^{n-k'} \\
&\leqslant \frac{1}{p(n+1)}
\end{aligned}$$

□

3 Algorithme des k plus proches voisins ($k-$ppv ou $k-$NN)

Soit $n \geqslant 1$, $x_{1:n} \in \mathbb{X}^n$ (on suppose que $\mathbb{X} = \mathbb{R}^d$ muni d'une certaine norme $\|.\|$). Soit $x \in \mathbb{X}$. On définit la suite d'applications mesurables $(i_1, \ldots, i_n) \colon \mathbb{X}^n \times \mathbb{X} \longrightarrow \{1, \ldots, n\}$ par récurrence comme suit :

$$i_1(x_{1:n}, x) = \min\left\{i \in \{1,\ldots,n\},\ d(x,x_i) \leqslant d(x,x_j)\ \forall j \in \{1,\ldots,n\}\right\}$$

$$i_2(x_{1:n}, x) = \min\left\{i \in \{1,\ldots,n\}\backslash\{i_1(x_{1:n},x)\},\ d(x,x_i) \leqslant d(x,x_j)\ \forall j \in \{1,\ldots,n\}\backslash\{i_1(x)\}\right\}$$

$i_n(x_{1:n}, x) =$l'unique élément de $\{1,\ldots,n\}\backslash\{i_1(x),\ldots,i_{n-1}(x)\}$.
La règle pour la régression associée au k ppv est la suivante :

$$\hat{\eta}((x_{1:n}, y_{1:n}), x) = \frac{1}{k_n}\sum_{j=1}^{k_n} y_{i_j}(x_{i:n}, x)$$

pour une suite d'entiers $(k_n)_n$ d'entiers > 0. On note $i_j(x_{1:n}, x)$: (j). Ainsi :

$$\hat{\eta}((x_{1:n}, y_{1:n}), x) = \frac{1}{k_n}\sum_{j=1}^{k_n} y(j)$$

La règle de classification $\hat{f}$ associée au k ppv est la règle plug-in associée à $\hat{\eta}$. L'algorithme de k ppv fait partie des algorithme par voisinage local avec :

$$\omega_{i,n}(x) = \frac{1}{k_n}\mathbb{1}_{\{x_i \in k-\mathrm{ppv}(x)\}} = \frac{1}{k_n}\mathbb{1}_{\left\{i \in \{i_1(x),\ldots,i_n(x)\}\right\}}$$

Remarque. $\sum_{i=1}^{n} \omega_{i,n} = 1$ donc la première hypothèse du théorème de Stone est vérifiée.

Théorème 2

Soit $\mathbb{X} = \mathbb{R}^d$, (k_n) une suite d'entiers > 0. On suppose $\mathbb{Y} = \{0,1\}$ et on prend le coût $0-1$. On suppose :

$$\lim_{n\to+\infty} k_n = +\infty \qquad \lim_{n\to+\infty} \frac{k_n}{n} = 0$$

Alors l'algorithme des k-ppv est universellement faiblement consistant.

Preuve.

Soit $\mathbb{P}$ une loi sur $\mathbb{X} \times \mathbb{Y}$ et $(X_i, Y_i)_i$ i.i.d de loi $\mathbb{P}$, $(X,Y) \perp (X_i, Y_i)$ de loi $\mathbb{P}$.

1. Il suffit de montrer les conditions 2,3,4 du théorème de Stone d'après la remarque.
2. Soit $a > 0$.

$$\mathbb{E}\left[\sum_{i=1}^{n} \omega_{i,n}(X) \mathbb{1}_{\{\|X_i - X\| \geqslant a\}}\right] = \mathbb{E}\left[\frac{1}{k_n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \in k-\text{ppv}(X)\}} \mathbb{1}_{\{\|X_i - X\| \geqslant a\}}\right]$$

Or,

$$\mathbb{1}_{\{X_i \in k-\text{ppv}(X)\}} \mathbb{1}_{\{\|X_i - X\| \geqslant a\}} \leqslant \mathbb{1}_{\{\|X_{k_n} - X\| \geqslant a\}} \mathbb{1}_{\{X_i \in k_n-\text{ppv}(X)\}}$$

$$\mathbb{E}\left[\sum_{i=1}^{n} \omega_{i,n}(X) \mathbb{1}_{\{\|X_i - X\| \geqslant a\}}\right] =\leqslant \mathbb{P}\big(\|X_{k_n} - X\| \geqslant a\big)$$

On va montrer en TD que :

$$\lim \mathbb{P}(\|X_{k_n} - X\| \geqslant a) = 0 \quad \text{si } \frac{k_n}{n} \to 0$$

3. $$\mathbb{E}[\max_i \omega_{i,n}] = \mathbb{E}\left[\max_i \frac{1}{k_n} \mathbb{1}_{\{X_i \in k-\text{ppv}(X)\}}\right] \leqslant \frac{1}{k_n} \xrightarrow[n\to+\infty]{} 0 \quad \text{par hypothèse}$$

4. c.f. lemme suivant

□

Lemme 1 (de Stone)

Soit (X_i) i.i.d ; de loi $\mathbb{P}$ et $X \perp (X_i)$ de loi $\mathbb{P}$. Soit $f \colon \mathbb{R}^d \longrightarrow \mathbb{R}$ tel que $\mathbb{E}[|f(X)|] < +\infty$. Alors il existe $\gamma_d > 0$ tel que :

$$\mathbb{E}\left[\frac{1}{k}\sum_{i=1}^{k} f(X_i)\right] \leqslant \gamma_d \mathbb{E}[|f(X)|] \quad \forall k \in \mathbb{N}^*$$

De plus,

$$\gamma_d \leqslant (1 - 2/\sqrt{2-\sqrt{3}})^d - 1$$

La preuve de ce résultat est basée sur deux lemmes géométriques.

Définition 2

Soit $x \in \mathbb{R}^d \backslash \{0\}$ et $\theta \in [0, \frac{\pi}{2}]$. On définit le cône de direction x et d'angle θ par :

$$C(x, \theta) = \left\{ \tilde{x} \in \mathbb{R}^d \mid \frac{\langle x, \tilde{x} \rangle}{\|x\| \|\tilde{x}\|} \geqslant \cos \theta \right\}$$

Lemme 2

Soit $x \in \mathbb{R}^d \backslash \{0\}$. Soit $\omega, z \in C(x, \frac{\pi}{6})$. Si $\|z\| \leqslant \|\omega\|$. Alors :

$$\|z - \omega\| \leqslant \|\omega\|$$

Lemme 3

Soit $\theta \in [0, \frac{\pi}{2}]$. Alors il existe une famille finie $\{x_1, \ldots, x_{N(\theta)}\}$ tel que :

$$\mathbb{R}^d = \bigcup_{i=1}^{N(\theta)} C(x_i, \theta)$$

Preuve.

La première observation : s'il existe $\{x_1, \ldots, x_{N(\theta)}$ tel que :

$$(*) \quad \mathbb{S}^d = \bigcup_{i=1}^{N(\theta)} C(x_i, \theta) \cap \mathbb{S}^d \quad \text{où } \mathbb{S}^d = \{x \ : \ \|x\| = 1\}$$

alors la démontstration est finie. Soit $z \in \mathbb{R}^d \backslash \{0\}$ alors $\frac{z}{\|z\|} \in \mathbb{S}^d$ et alors $\exists i$ et ω tels que $\omega = \frac{z}{\|z\|} \in C(x_i, \theta)$. $\|z\| \omega \in C(x_i, \theta)$, $z \in C(x_i, \theta)$.

On montre $(*)$. Pour cela on montre que pour tout $r \in [0, 1]$ il existe $x_1, \ldots, x_N \neq 0$ tels que :

$$\mathbb{S}^d = \bigcup_{i=1}^{N} \mathbb{S}^d \cap \mathcal{B}(x_i, r)$$

On construit $x_1, \ldots, x_N$ par récurrence. $x_1 = e_1$. Supposons $x_1, \ldots, x_k$ construits. S'il existe $\tilde{x}$ tel que :

$$\inf_{i \in \{1, \ldots, k\}} \|\tilde{x} - x_i\| \geqslant k$$

alors on pose $x_{k+1} = \tilde{x} / \|\tilde{x}\|$, et sinon on arrête.

On montre d'abord qu'on a un nombre fini de x_i. En effet, pour tout $i \neq j$,

$$\mathcal{B}(x_i, r/2) \cap \mathcal{B}(x_j, r/2) = \emptyset \qquad (1)$$

par définition. DE plus :

$$\bigcup_{i=1}^{n} \mathcal{B}(x_i, r/2) \subset \mathcal{B}(0, 1 + r/2) \backslash \mathcal{B}(0, r/2) \qquad (2)$$

$\Rightarrow$ on a donc :

$$n\text{Vol}(\mathcal{B}(0,1)) \left(\frac{r}{2}\right)^d \leqslant \text{Vol}(\mathcal{B}(0,1)) \left(\left(1 + \frac{r}{2}\right)^d - \left(\frac{r}{2}\right)^d\right)$$

Comme c'est valable pour tout n si (x_i) est infini oon a une contradiction. Donc on a une famille finie de (x_i) et $(1), (2)$ sont encore vérifiées. Donc si on note N_r le cardinal de (x_i) on a :

$$N_R \leqslant \left(\frac{2}{r} + 1\right)^d - 1$$

Pour en revenir au cône. $\forall \theta \in [0, \pi/2]$ et $x \in \mathbb{R}^d \backslash \{0\}$, $\|x\| = 1$. Alors :

$$C(x, \theta) \cap \mathbb{S}^d = \mathcal{B}(x, r_\theta) \cap \mathbb{S}^d \qquad (3)$$

avec $r_\theta = 2\sin(\theta/2)$ (à faire en exercice).
Donc on prend $x_1, \ldots, x_{N_{r_\theta}} \in \mathbb{S}^d$ tels que :

$$\mathbb{S}^d = \bigcup_{i=1}^{N_{r_\theta}} \mathbb{S}^d \cap \mathcal{B}(x_i, r_\theta)$$

D'après (3) on a $(*)$ et de plu :

$$N_{r_\theta} \leqslant \left(\frac{2}{r_\theta} + 1\right)^d - 1 = \left(\frac{1}{\sin\left(\frac{\theta}{2}\right)} + 1\right)^d - 1$$

Et pour $\theta = \frac{\pi}{6}$:

$$N_{r_\theta} \leqslant \left(\frac{2}{\sqrt{2 - \sqrt{3}}} + 1\right)^d - 1$$

□

Preuve. (du Lemme 1 de Stone)

$$\sum_{p=1}^{k} \mathbb{E}[f(X_p)] = \sum_{p=1}^{n} \mathbb{E}\big[f(X_p)\mathbb{1}_{\{X_p \in k-\text{ppv}(X)\}}\big]$$

Par hypothèse, les (X_i) sont i.i.d. $\perp X$ de même loi, donc :

$$= \sum_{p=1}^{n} \mathbb{E}\big[f(X)\mathbb{1}_{\{X \in k-\text{ppv}(X_p) \text{ parmi } (X_1, \ldots, X, \ldots, X_n)\}}\big]$$

On découpe $\mathbb{R}^d$ comme :

$$\mathbb{R}^d = x + \bigcup_{i=1}^{N_{r_{\pi/6}}} C(x_i, \pi/6)$$

Définissons :

$$x + C(x_i, \pi/6) = A_i(x)$$

$$\{X_p \in A_j(X)\} \cap \{X \in k - \mathrm{ppv}(X_p) \subset \{X_p \in A_j(X)\} \cap \{X_p \in k - \mathrm{ppv}(X) \text{ dans } A_j(X)\}$$

d'après le premier lemme. On en déduit que :

$$\sum_{p=1}^{k} \mathbb{E}[|f(X_p)|] \leqslant \sum_{p=1}^{k} \sum_{j=1}^{N_{\pi/6}} \mathbb{E}[|f(X)| \mathbb{1}_{X_p \in A_j(X)} \mathbb{1}_{\{X_p \in k - \mathrm{ppv}(X) \text{ dans } A_j(X)\}}]$$

Donc :

$$\sum_{p=1}^{k} \mathbb{E}[(X_p)] \leqslant k N_{\pi/6} \mathbb{E}[|f(X)|]$$

Conclusion :

$$\frac{1}{k} \mathbb{E}\left[\sum_{p=1}^{k} |f(X_p)|\right] \leqslant N_{\pi/6} \mathbb{E}[|f(X)|]$$

et on a vu une majoration de $N_{\pi/6}$. $\square$