Partiel 2023/2024 – Durée 1h30

Vous serez évalués sur la démarche et la maîtrise des concepts utilisés. Par conséquent, peu d'importance sera accordée aux constantes numériques. Pour simplifier, ***on supposera que tous les ratios de quantités qui nous arrangent sont des entiers*** *(si cela vous perturbe, libre à vous d'utiliser des parties entières quand nécessaire).*

Prédiction par histogrammes

On se donne $D_n = ((X_1, Y_1), \ldots, (X_n, Y_n))$ un échantillon i.i.d. de loi P, où $X_1 \in [0,1]^2$ p.s. (carré unité de dimension 2) et $Y_1 \in \{0,1\}$. On cherche à construire $\hat{f} : [0,1]^2 \to \{0,1\}$ en se basant sur D_n, de telle sorte que $\hat{f}(X)$ prédise au mieux Y.

1. De quel type de probème (classification/régression) s'agit-il ? Quelle est la fonction de coût standard associée à ce problème ?
2. Définir la fonction de risque associée, et rappeler la forme d'un classifieur de Bayes.

On se donne $M \in \mathbb{N}^*$, on découpe le carré $[0,1]^2$ en $(A_j)_{j=1,\ldots,M^2}$ sous-carrés de longueur $1/M$ (formant alors une partition du carré initial), et on définit le classifieur par histogrammes $\hat{f}_M$ via

$$\hat{f}_M(x) = \sum_{j=1}^{M^2} \mathbb{1}_{A_j}(x) \mathbb{1}_{\sum_{i=1}^n (2Y_i - 1)\mathbb{1}_{A_j}(X_i) > 0},$$

pour $x \in [0,1]^2$.

3. Montrer que $\hat{f}_M$ est un classifieur de type plug-in, c'est à dire du type $\mathbb{1}_{\hat{\eta}_M(x) > (1/2)}$, où $\hat{\eta}_M$ est un estimateur de la fonction de régression $\eta(X) = \mathbb{E}(Y \mid X)$ que l'on précisera.
4. En notant R_P^* le risque de Bayes, montrer que
$$R_P(\hat{f}_M) - R_P^* \leq 2\sqrt{E_X((\hat{\eta}_M(X) - \eta(X))^2)}.$$
5. Pour $j \in [\![1, M^2]\!]$, on note $p_j = \mathbb{P}(X \in A_j)$, $\bar{\eta}_j = \mathbb{E}(\eta(X) \mid X \in A_j)$ (en prenant pour convention $\bar{\eta}_j = 0$ si $p_j = 0$), et
$$\bar{\eta}_M(x) = \sum_{j=1}^{M^2} \bar{\eta}_j \mathbb{1}_{A_j}(x).$$
Montrer que
$$E_X((\hat{\eta}_M(X) - \eta(X))^2) = E_X((\eta(X) - \bar{\eta}_M(X))^2) + E_X((\bar{\eta}_M(X) - \hat{\eta}_M(X))^2).$$

6. **Attention question technique**. En notant $N_j = |\{i \mid X_i \in A_j\}|$, montrer que

$$E_{D_n}\left[E_X((\bar{\eta}_M(X) - \hat{\eta}_M(X))^2)\right] \leq \sum_{j=1}^{M^2} p_j \left(\mathbb{E}\left(\frac{1}{2N_j}\mathbb{1}_{N_j \geq 1}\right) + \mathbb{P}(N_j = 0)\right).$$

7. On admet pour le moment les deux inégalités suivantes : si $X \sim \mathcal{B}(n,p)$, alors $\mathbb{E}((X^{-1})\mathbb{1}_{X>0}) \leq 2/(np)$, et, pour tout $u \in [0,1]$, $u(1-u)^n \leq 1/n$. En déduire que

$$E_{D_n}\left[E_X((\bar{\eta}_M(X) - \hat{\eta}_M(X))^2)\right] \leq \frac{2M^2}{n}.$$

8. On suppose maintenant que η est L-Lipschitz pour la norme infinie. Montrer que

$$E_{D_n}\left(R_P(\hat{f}_M) - R_P^*\right) \leq \frac{2L}{M} + \frac{2\sqrt{2}M}{\sqrt{n}}.$$

9. En choisissant judicieusement M, montrer

$$\inf_{\hat{f}} \sup_{\eta\ L\text{-Lip}} E_{D_n}\left(R_P(\hat{f}) - R_P^*\right) \leq c_0 \left(\frac{L}{\sqrt{n}}\right)^{\frac{1}{2}},$$

pour une constante absolue c_0.

Point de vue ERM

Pour un M donné et avec les notations précédentes, on note, pour $\sigma \in \{0,1\}^{M^2}$, f_σ le classifieur donné par

$$f_\sigma(x) = \sum_{j=1}^{M^2} \sigma_j \mathbb{1}_{A_j(x)},$$

et $\mathcal{F}_M = \{f_\sigma \mid \sigma \in \{0,1\}^{M^2}\}$ l'ensemble des tels classifieurs.

10. Montrer que $\hat{f}_M \in \arg\min_{f \in \mathcal{F}_M} R_n(f)$, où R_n est le risque empirique. En notant $\bar{f}_M = \mathbb{1}_{\bar{\eta}_M > (1/2)}$, montrer que $\bar{f}_M \in \arg\min_{f \in \mathcal{F}_M} R_P(f)$.
11. Soit D la dimension de Vapnik de $\mathcal{F}_M$. Montrer que $D = M^2$.
12. En déduire un résultat similaire à la question 7 :

$$E_{D_n}(R_P(\hat{f}_M) - R_P(\bar{f}_M)) \leq 2c_0\sqrt{\frac{M^2}{n}},$$

où c_0 est une constante absolue. **NB : Il n'est pas nécessaire de redémontrer les résultats de cours, préciser pourquoi ils s'appliquent suffira.**

Dans cette partie, le but est de choisir un M adapté au L que l'on ne connaît pas.

13. Pour un M donné, on note

$$Z_M = \sup_{f \in \mathcal{F}_M} |(R_n(f) - R(f)| \,.$$

Montrer que

$$\mathbb{P}\left(Z_M \geq c_0 \frac{M}{\sqrt{n}} + \sqrt{2vx}\right) \leq e^{-x},$$

avec $v = \frac{1}{4n}$.

14. Construire une famille $(M_j)_{j\in\mathbb{Z}}$ de "modèles" et mettre en oeuvre une stratégie de sélection de modèle pour que le prédicteur final, $\hat{f} = \hat{f}_{M_{\hat{j}}}$ vérifie une inégalité de type

$$E_{D_n}(R_P(\hat{f}) - R_P^*) \leq c_0 \left(\frac{L}{\sqrt{n}}\right)^{\frac{1}{2}} + u_n,$$

avec $u_n = o(n^{-1/4})$. Bien évidemment le prédicteur final ne peut pas prendre en compte le L inconnu dans sa conception.

15. Prouver les résultats admis de la question $7-$.

SOLUTION

1. Classification (encore que la suite montre que ce n'est pas si claire : dans la première partie on se base sur un problème de régression pour construire un classifieur). Fonction de coût standard : pour $y, y' \in \{0,1\}$, $c(y', y) = \mathbb{1}_{y\neq y'}$.
2. Fonction de risque : $R_P(f) = \mathbb{P}(f(X) \neq Y)$. Un classifieur de Bayes : $f^*(X) = \mathbb{1}_{\eta(X)>(1/2)}$, pour $\eta(X) = \mathbb{E}(Y \mid X)$. Par ailleurs un classifieur est de Bayes ssi il coïncide avec f^* sur $\{\eta(X) \neq (1/2)\}$ p.s..
3. On note $N_j = |\{i \mid X_i \in A_j\}|$. On remarque que, si $N_j \neq 0$,

$$\begin{aligned}\sum_{i=1}^{n} (2Y_i - 1)\mathbb{1}_{A_j}(X_i) > 0 &\Leftrightarrow 2\frac{\sum_{i=1}^n Y_i \mathbb{1}_{X_i \in A_j}}{N_j} > 1 \\ &\Leftrightarrow \hat{\eta}_j > (1/2),\end{aligned}$$

avec $\hat{\eta}_j = \frac{\sum_{i=1}^n Y_i \mathbb{1}_{X_i\in A_j}}{N_j}$. En posant $\hat{\eta}_j = 0$ si $N_j = 0$, et

$$\hat{\eta}_M(x) = \sum_{j=1}^{M^2} \hat{\eta}_j \mathbb{1}_{A_j}(x),$$

on a bien $\hat{f}_M = \mathbb{1}_{\hat{\eta}_M > (1/2)}$.

4. On a

$$\begin{aligned} R_P(\hat{f}_M) - R_P(f^*) &= 2E_X\left(|\eta(X) - (1/2)\mathbb{1}_{\hat{f}_M(X)\neq f^*(X)}\right) \\ &= 2E_X\left(|\eta(X) - (1/2)|\mathbb{1}_{\mathbb{1}_{\hat{\eta}_M(X)>(1/2)}\neq \mathbb{1}_{\eta^*(X)>(1/2)}}\right). \end{aligned}$$

Or, si $\mathbb{1}_{\hat{\eta}_M(X)>(1/2)} \neq \mathbb{1}_{\eta^*(X)>(1/2)}$, on a $|\eta(X) - (1/2)| \leq |\eta(X) - \hat{\eta}_M(X)|$ (il sont forcément de part et d'autre de $1/2$). On en déduit

$$\begin{aligned} R_P(\hat{f}_M) - R_P(f^*) &\leq 2E_X\left(|\hat{\eta}_M(X) - \eta(X)|\right) \\ &\leq 2\sqrt{E_X((\hat{\eta}_M(X) - \eta(X))^2)} \qquad \text{(Jensen)}. \end{aligned}$$

5. Pour un j fixé, on a

$$\begin{aligned} E_X((\hat{\eta}_M(X) - \eta(X))^2\mathbb{1}_{X\in A_j}) &= E_X((\hat{\eta}_j - \eta(X))^2\mathbb{1}_{X\in A_j}) \\ &= E_X\mathbb{E}\left[((\hat{\eta}_j - \eta(X))^2\mathbb{1}_{X\in A_j}) \mid X \in A_j\right] \\ &= E_X\left(\mathbb{E}\left[(\hat{\eta}_j - \bar{\eta}_j)^2 + (\eta(X) - \bar{\eta}_j)^2 \mid X \in A_j\right]\mathbb{1}_{X\in A_j}\right) \\ &= E_X((\hat{\eta}_M(X) - \bar{\eta}_M(X))^2\mathbb{1}_{X\in A_j}) + E_X((\eta(X) - \bar{\eta}_M(X))^2\mathbb{1}_{X\in A_j}). \end{aligned}$$

En sommant sur les j on obtient le résultat voulu.

6. On a

$$E_X((\bar{\eta}_M(X) - \hat{\eta}_M(X))^2) = \sum_{j|p_j\neq 0} p_j\left[\left(\bar{\eta}_j - \frac{\sum_{i=1}^n Y_i\mathbb{1}_{X_i\in A_j}}{N_j}\right)^2\mathbb{1}_{N_j\geq 1} + \bar{\eta}_j^2\mathbb{1}_{N_j=0}\right]. \tag{1}$$

Travaillons maintenant à j fixé, et commençons par remarquer que

$$E_{D_n}\left(\bar{\eta}_j^2\mathbb{1}_{N_j=0}\right) \leq \mathbb{P}(N_j = 0),$$

car $\bar{\eta}_j \in [0,1]$. Ensuite, occupons nous du terme en $\mathbb{1}_{N_j\geq 1}$. Conditionnellement à $X_1, \ldots X_n$, les $Y_i\mathbb{1}_{X_i\in A_j}$ sont indépendantes de loi $\mathbb{1}_{X_i\in A_j}\mathcal{B}(\eta(X_i))$. On peut alors écrire

$$\begin{aligned} &E_{D_n}\left[\left(\bar{\eta}_j - \frac{\sum_{i=1}^n Y_i\mathbb{1}_{X_i\in A_j}}{N_j}\right)^2\mathbb{1}_{N_j\geq 1} \mid X_1, \ldots, X_n\right] \\ &\quad = \left(\left(\bar{\eta}_j - \frac{\sum_{i=1}^n \eta(X_i)\mathbb{1}_{X_i\in A_j}}{N_j}\right)^2 + \frac{1}{N_j^2}\sum_{i=1}^n \eta(X_i)(1-\eta(X_i))\mathbb{1}_{X_i\in A_j}\right)\mathbb{1}_{N_j\geq 1} \\ &\quad \leq \left[\left(\bar{\eta}_j - \frac{\sum_{i=1}^n \eta(X_i)\mathbb{1}_{X_i\in A_j}}{N_j}\right)^2 + \frac{1}{4N_j}\right]\mathbb{1}_{N_j\geq 1}, \end{aligned}$$

en utilisant $\eta(X_i)(1-\eta(X_i) \leq 1/4$. Notons maintenant $Z_n^j = (\mathbb{1}_{X_i \in A_j})_{i=1,\dots,n}$. Conditionnellement à Z_n^j, les $\eta(X_i)$ restent indépendantes, et on a $\mathbb{E}(\eta(X_i)\mathbb{1}_{X_i\in A_j} \mid Z_n^j) = \bar{\eta}_j \mathbb{1}_{X_i \in A_j}$. Comme $\eta(X_i) \in [0,1]$, on a de plus $\mathrm{Var}(\eta(X_i) \mid Z_n^j) \leq 1/4$. On en déduit alors

$$\begin{aligned} E_{D_n}\left[\left(\bar{\eta}_j - \frac{\sum_{i=1}^n \eta(X_i)\mathbb{1}_{X_i\in A_j}}{N_j}\right)^2 \mathbb{1}_{N_j\geq 1} \mid Z_n^j\right] &= \mathrm{Var}\left(\frac{\sum_{i=1}^n \eta(X_i)\mathbb{1}_{X_i\in A_j}}{N_j} \mid Z_n^j\right)\mathbb{1}_{N_j\geq 1} \\ &= \frac{1}{N_j^2}\sum_{i=1}^n \mathrm{Var}(\eta(X_i) \mid Z_n^j)\mathbb{1}_{X_i\in A_j}\mathbb{1}_{N_j\geq 1} \\ &\leq \frac{1}{4N_j}\mathbb{1}_{N_j\geq 1}. \end{aligned}$$

En recollant tous les morceaux de (1), on obtient le résultat voulu.

7. En remarquant que $N_j \sim \mathcal{B}(n, p_j)$, on obtient $\mathbb{E}(N_j^{-1}\mathbb{1}_{N_j\geq 1}) \leq 2/(np_j)$. Par ailleurs, on a, à j fixé,

$$p_j\mathbb{P}(N_j = 0) = p_j(1-p_j)^n \leq \frac{1}{n}.$$

On en déduit alors

$$\begin{aligned} \sum_{j=1}^{M^2} p_j\left(E_{D_n}\left(\frac{1}{2N_j}\mathbb{1}_{N_j\geq 1}\right) + \mathbb{P}(N_j=0)\right) &\leq \sum_{j=1}^{M^2}\left(\frac{1}{n}+\frac{1}{n}\right) \\ &\leq \frac{2M^2}{n}, \end{aligned}$$

ce dont on déduit l'inégalité voulue.

8. Si η est L-Lipschitz, on a, pour tout j tel que $p_j > 0$, et $x \in A_j$,

$$|\bar{\eta}_j - \eta(x)| \leq \frac{L}{M}.$$

On en déduit alors

$$\begin{aligned} E_X((\eta(X) - \bar{\eta}_M(X))^2) &\leq \sup_{j=1,\dots,M^2} p_j \frac{L^2}{M^2} \\ &\leq \frac{L^2}{M^2}. \end{aligned}$$

En utilisant le résultat de la question $4-$, on déduit

$$E_{D_n}\left(R_P(\hat{f}_M) - R_P^*\right) \leq 2\left[\frac{L}{M} + \frac{M\sqrt{2}}{\sqrt{n}}\right].$$

9. Optimiser la somme revient à choisir $M = \frac{\sqrt{L}n^{1/4}}{2^{1/4}}$, ce qui donne

$$\inf_{\hat{f}} \sup_{\eta\ L\text{-Lip}} E_{D_n}\left(R_P(\hat{f}) - R_P^*\right) \leq \sup_{\eta\ L\text{-Lip}} E_{D_n}\left(R_P(\hat{f_M}) - R_P^*\right) \leq 4 \times 2^{1/4}\left(\frac{L}{\sqrt{n}}\right)^{\frac{1}{2}}.$$

   C'est bien la vitesse optimale (en terme de n et L). Toutefois le prédicteur proposé prend en compte la constante L. On aimerait trouver un prédicteur qui s'adapte au L inconnue : on parle de procédure adaptative.

10. Pour $\sigma \in \{0,1\}^{M^2}$, on peut écrire

$$R_n(f_\sigma) = \sum_{j=1}^{M^2} \sum_{i|X_i \in A_j} \mathbb{1}_{\sigma_j \neq Y_i}.$$

   Or, à j fixé,

$$\sum_{i|X_i \in A_j} \mathbb{1}_{\sigma_j \neq Y_i} = \mathbb{1}_{\sigma_j=0} \sum_{i=1}^{n} Y_i \mathbb{1}_{X_i \in A_j} + \mathbb{1}_{\sigma_j=1} \sum_{i=1}^{n} (1-Y_i)\mathbb{1}_{X_i \in A_j}$$

   est minimal pour $\sigma_j = \mathbb{1}_{\sum_{i=1}^{n}(2Y_i-1)\mathbb{1}_{A_j}(X_i)>0}$, et donc pour $f_\sigma = \hat{f}_M$. Pareillement, on peut écrire

$$R_P(f_\sigma) = \sum_{j=1}^{M^2} p_j \mathbb{E}(Y\mathbb{1}_{A_j}(X) \neq \sigma_j),$$

   et, à j fixé tel que $p_j \neq 0$ (ailleurs on peut mettre ce qu'on veut comme label),

$$\begin{aligned}
\mathbb{E}(\mathbb{1}_{Y\neq\sigma_j}\mathbb{1}_{A_j}(X)) &= \mathbb{1}_{\sigma_j=0} E_X(\eta(X)\mathbb{1}_{A_j}(X)) + \mathbb{1}_{\sigma_j=1} E_X((1-\eta(X))\mathbb{1}_{A_j}(X)) \\
&= \mathbb{1}_{\sigma_j=0} E_X(\bar{\eta}_j \mathbb{1}_{A_j}(X)) + \mathbb{1}_{\sigma_j=1} E_X((1-\bar{\eta}_j)\mathbb{1}_{A_j}(X)) \\
&= p_j\left(\mathbb{1}_{\sigma_j=0}\bar{\eta}_j + \mathbb{1}_{\sigma_j=1}(1-\bar{\eta}_j)\right),
\end{aligned}$$

   qui est minimal pour $\sigma_j = \mathbb{1}_{\bar{\eta}_j>(1/2)}$, et donc pour $f_\sigma = \bar{f}_M$.

11. En notant $D_M = M^2$. Si on se donne $(x_1, \dots, x_{D_M})$ tel que pour tout j $x_j \in A_j$, alors, pour tout $\sigma \in \{0,1\}^{D_M}$, on aura bien $f_\sigma(x_j) = \sigma_j$, et donc

$$\left|\{(f_\sigma(x_1), \dots, f_\sigma(x_{D_M})) \mid \sigma \in \{0,1\}^{D_M}\}\right| = 2^{D_M},$$

   ce dont on déduit $D \geq D_M$.
   Dans l'autre sens, soit $(x_1, \dots, x_{D_M+1}) \in ([0,1]^2)^{D_M+1}$. $(A_j)_{j=1,\dots,D_M}$ formant une partition de $[0,1]^2$, il existe $j_1 \neq j_2 \leq D_M+1$ et $j_0 \leq D_M$ tels que $x_{j_1}, x_{j_2} \in A_{j_0}$. Dans ce cas, si $\varepsilon \in \{0,1\}^{D_M+1}$ est tel que $\varepsilon_{j_1} \neq \varepsilon_{j_2}$, il ne peut exister de $\sigma \in \{0,1\}^{D_M}$ tel que $f_\sigma(x_{j_1}) = \varepsilon_{j_1}$ et $f_\sigma(x_{j_2}) = \varepsilon_{j_2}$ (rappelons que les f_σ sont constantes sur les A_j). On en déduit alors que

$$\forall (x_1, \dots, x_{D_M+1}) \quad \left|\{(f_\sigma(x_1), \dots, f_\sigma(x_{D_M+1})) \mid \sigma \in \{0,1\}^{D_M}\}\right| < 2^{D_M+1},$$

   et donc que $D < D_M + 1$.

12. En utilisant un résultat de cours sur les ERM (symétrisation puis inégalité entropique de Dudley), on a

$$\begin{aligned} E_{D_n}(R_P(\hat{f}_M) - R_P(\bar{f}_M)) &\leq \mathbb{E}\left(\sup_{f\in\mathcal{F}_M} |(R_n(f) - R_P(f)|\right) \\ &\leq 2E_{D_n}E_\varepsilon\left(\sup_{f\in\mathcal{F}_M}\frac{1}{n}\sum_{i=1}^n \varepsilon_i \mathbb{1}_{Y_i\neq f(X_i)}\right) \\ &\leq 2c_0\frac{\sqrt{D}}{\sqrt{n}}, \end{aligned}$$

où les $\varepsilon_1, \dots \varepsilon_n$ sont des variables de Rademacher i.i.d..

13. D'après la question précédente, on sait déjà que

$$\mathbb{E}(Z_M) \leq c_0\frac{M}{\sqrt{n}}.$$

Par ailleurs, si on pose $Z_M = g(V_1, \dots, V_n)$, où $V_i = (X_i, Y_i)$, on a, pour tout $i \in [\![1, n]\!]$,

$$|g(x_1, \dots, x_i, \dots, x_n) - g(x_1, \dots, x_i', \dots, x_n)| \leq 1/n.$$

L'inégalité de Mac Diarmid donne alors

$$\mathbb{P}\left(Z_M \geq E(Z_M) + \sqrt{2\frac{1}{4n}x}\right) \leq e^{-x},$$

correspondant à des queues sous-Gaussiennes de variance $v = \frac{1}{4n}$.

14. Posons $M_j = 2^j$, pour $j \in \mathbb{Z}$, et

$$\mathrm{pen}(j) = c_0\frac{M_j}{\sqrt{n}} + \sqrt{\frac{\log(2) + \log(\frac{\pi^2}{3}(j\vee 1)^2)}{2n}}.$$

On a alors

$$\mathbb{P}\left(\bigcup_{j\in\mathbb{Z}} Z_{M_j} \leq \mathrm{pen}(j) + \sqrt{x/2}\right) \geq 1 - \left(\frac{1}{2} + \frac{1}{2}\sum_{j\in\mathbb{Z}^*}\frac{1}{j^2}\times\frac{3}{\pi^2}\right)e^{-x} \geq 1 - e^{-x}.$$

Dans la suite on se place sur cet évènement. On choisit alors

$$\hat{j} \in \arg\min_{j\in\mathbb{Z}} R_n(\hat{f}_{M_j}) + \mathrm{pen}(j).$$

On peut alors écrire, pour un j quelconque

$$
\begin{aligned}
R_P(\hat{f}_{\hat{j}}) - R_P^* &\leq R_n(\hat{f}_{\hat{j}}) + \mathrm{pen}(\hat{j}) + \sqrt{x/(2n)} - R_P^* \\
&\leq R_n(\hat{f}_j) + \mathrm{pen}(j) + \sqrt{x/(2n)} - R_P^* \\
&\leq R_n(\bar{f}_j) + \mathrm{pen}(j) + \sqrt{x/(2n)} - R_P^* \\
&\leq R_P(\bar{f}_j) - R_P^* + 2\mathrm{pen}(j) + \sqrt{2x/n}.
\end{aligned}
$$

Essayons maintenant de déterminer $\inf_j R_P(\bar{f}_j) - R_P^* + 2\mathrm{pen}(j)$. En utilisant les bornes obtenues en question 8 et 4, on a

$$
R_P(\bar{f}_j) - R_P^* \leq \frac{2L}{M_j},
$$

et donc

$$
\begin{aligned}
R_P(\bar{f}_j) - R_P^* &\leq \frac{2L}{M_j} + 2\mathrm{pen}(j) \\
&\leq 2^{1-j}L + c_0\frac{2^{j+1}}{\sqrt{n}} + 2\sqrt{\frac{\log(2) + \log(\frac{\pi^2}{3}(j \vee 1)^2)}{2n}}.
\end{aligned}
$$

Posons maintenant $j^* = \lfloor \frac{1}{2}\log_2\left(\frac{\sqrt{n}L}{c_0}\right)\rfloor$, de sorte que

$$
\frac{1}{2}\left(\frac{\sqrt{n}L}{c_0}\right)^{1/2} \leq 2^{j^*} \leq \left(\frac{\sqrt{n}L}{c_0}\right)^{1/2}.
$$

On a d'une part

$$
2^{1-j^*}L + c_0\frac{2^{j^*+1}}{\sqrt{n}} \leq 6\sqrt{c_0}\left(\frac{L}{\sqrt{n}}\right)^{1/2},
$$

et d'autre part

$$
2\sqrt{\frac{\log(2) + \log(\frac{\pi^2}{3}(j^* \vee 1)^2)}{2n}} \leq \frac{\sqrt{2}}{\sqrt{n}}\sqrt{\log(2) + \log\left(\frac{\pi^2}{3}\left(\frac{1}{4}\log_2^2\left(\frac{\sqrt{n}L}{c_0}\right) \vee 1\right)\right)} := v_n,
$$

avec $v_n \sim \log(n)/\sqrt{n} = o(n^{-1/4})$. On en déduit alors que, en posant $\hat{f} = \hat{f}_{\hat{j}}$

$$
\mathbb{P}\left(R_P(\hat{f}) - R_P^* - 6\sqrt{c_0}\left(\frac{L}{\sqrt{n}}\right)^{\frac{1}{2}} - v_n \geq \sqrt{2x/n}\right) \leq e^{-x}.
$$

En posant $\Delta_n = R_P(\hat{f}) - R_P^* - 6\sqrt{c_0}\left(\frac{L}{\sqrt{n}}\right)^{\frac{1}{2}} - v_n$, on déduit

$$\begin{aligned} E_{D_n}\Delta_n \leq E_{D_n}\Delta_n^+ &= \int_0^{+\infty} \mathbb{P}(\Delta_n > t)dt \\ &\leq \int_0^{+\infty} e^{-\frac{nt^2}{2}} \\ &\leq \frac{1}{2}\sqrt{\frac{2\pi}{n}}. \end{aligned}$$

En posant $u_n = v_n + \sqrt{\frac{\pi}{2n}} = o(n^{-1/4})$, on a bien

$$E_{D_n}(R_P(\hat{f}) - R_P^*) \leq 6\sqrt{c_0}\left(\frac{L}{\sqrt{n}}\right)^{\frac{1}{2}} + u_n,$$

on a donc construit une procédure optimale ne prenant pas en compte la constante de Lispchitz inconnue de la fonction de régression cible.

15. Posons $g(u) = u(1-u)^n$, pour $u \in [0,1]$. On a $g'(u) = (1-u)^{n-1}((1-u) - nu)$. Le maximum est donc atteint en $u = 1/n + 1$, et est inférieur à $1/(n+1)$ (donc à $1/n$). Maintenant soit $X \sim \mathcal{B}(n,p)$. On a

$$\mathbb{E}(X^{-1}\mathbb{1}_{X\geq 1}) = \sum_{k=1}^{n} \frac{1}{k}\binom{n}{k} p^k (1-p)^{n-k}.$$

Or, pour $k \geq 1$,

$$\frac{1}{k}\binom{n}{k} \leq \frac{2}{k+1}\binom{n}{k} \leq \frac{2}{n+1}\binom{n+1}{k+1}.$$

On en déduit

$$\mathbb{E}(X^{-1}\mathbb{1}_{X\geq 1}) \leq \frac{2}{(n+1)p}\sum_{k=1}^{n}\binom{n+1}{k+1}p^{k+1}(1-p)^{(n+1)-(k+1)} \leq \frac{2}{(n+1)p} \leq \frac{2}{np}.$$