

Partiel 2025/2026 – Durée 1h30

Vous serez évalués sur la démarche et la maîtrise des concepts utilisés. Par conséquent, peu d'importance sera accordée aux constantes numériques.

RÉGRESSION QUANTILE

On se donne $D_n = ((X_1, Y_1), \dots, (X_n, Y_n))$ un échantillon i.i.d., de loi commune (X, Y) sur $]0, 1[\times \mathbb{R}$, déterminée par les éléments suivants

- X admet une densité $0 < c_X \leq p_X \leq C_X < +\infty$ sur $]0, 1[$.
- $Y = \beta^* X + E$, où E est indépendante de X , **symétrique**, de densité f .

On suppose par ailleurs que la variable E vérifie :

- $\mathbb{E}(|E|) \leq 1$,
- $\mathbb{P}(|E| \leq t) \leq C_E t$, pour tout $t > 0$,
- pour tout $0 < t \leq 1$, $\mathbb{P}(|E| \leq t) \geq c_E t$.

GÉNÉRALITÉS

1. Soit Z variable aléatoire réelle vérifiant $\mathbb{E}(|Z|) < +\infty$. On définit $\text{Med}(Z) = \{z \mid \mathbb{P}(Z \geq z) \geq 1/2, \mathbb{P}(Z \leq z) \geq 1/2\}$. Montrer que

$$m \in \arg \min_{z \in \mathbb{R}} \mathbb{E}(|Z - z|) \quad \Leftrightarrow \quad m \in \text{Med}(Z).$$

2. On cherche à prédire Y à partir de X . De quel type de problème (classification/régression) s'agit-il ? Pour $f : x \mapsto (\mathbb{1}_{|x| \leq 1} + |x|^{-3} \mathbb{1}_{|x| > 1})/3$, montrer que E vérifie les hypothèses de l'énoncé.
3. Montrer que pour la fonction de coût quadratique, et pour le E de la question du dessus, le risque de Bayes vaut $+\infty$.
4. On se donne la fonction de coût $c(y, y') = |y - y'|$ (et on note R le risque associé). Montrer que g est un prédicteur de Bayes si et seulement si $g(X) \in \text{Med}(Y|X)$, P_X presque sûrement. **Dans toute la suite on notera $g^*(X)$ un prédicteur de Bayes.**
5. Que vaut $g^*(X)$ si $Y = \beta^* X + E$ (modèle de l'énoncé) ? En déduire une classe de prédicteur pertinente (au vu de ce modèle), on la notera $\mathcal{G}$.
6. Donner l'expression d'un minimiseur $\hat{g}_n$ de risque empirique sur $\mathcal{G}$. Dans le cas où (X, Y) ne suit pas forcément le modèle de l'énoncé, décomposer l'excès de risque de $\hat{g}_n$ en termes d'erreur d'approximation et d'estimation. **Dans toute la suite on supposera que (X, Y) suit le modèle de l'énoncé.**

APPROCHE DIRECTE (ET PEU SUBTILE)

Dans cette partie on supposera de plus la connaissance de $|\beta^*| \leq M$, pour un M connu, et on ajustera la classe $\mathcal{G}$ en fonction.

7. Montrer que

$$\mathbb{E}(R(\hat{g}_n) - R^*) \leq 2\mathbb{E}_{D_n}\mathbb{E}_\varepsilon \sup_{|\beta| \leq M} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i (|Y_i - \beta X_i| - |Y_i - \beta^* X_i|) \right|,$$

où les ε_i sont des variables de Rademacher i.i.d..

8. En déduire que

$$\mathbb{E}(R(\hat{g}_n) - R^*) \leq 8M\mathbb{E}_{D_n}\mathbb{E}_\varepsilon \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i X_i \right|.$$

9. En déduire que

$$\mathbb{E}(R(\hat{g}_n) - R^*) \leq \frac{8M}{\sqrt{n}}.$$

Quelle vitesse 'maximale' pouviez-vous espérer ?

APPROCHE PLUS SUBTILE

Dans cette partie on ne suppose plus la connaissance de $|\beta^*| \leq M$.

10. Montrer que $\hat{\beta}_n$ (celui qui correspond à l'ERM $\hat{g}_n$) est une médiane pour la loi $\tilde{P}_n$ définie sur $\{(Y_i/X_i)_{i=1,\dots,n}\}$ par

$$\tilde{P}_n(\{Y_i/X_i\}) = \frac{X_i}{\sum_{i=1}^n X_i}.$$

11. En déduire que, pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(\hat{\beta}_n \leq \beta^* - \varepsilon\right) \leq \mathbb{P}\left(\sum_{i=1}^n Z_i \geq 0\right),$$

avec $Z_i = X_i (\mathbb{1}_{Y_i \leq (\beta^* - \varepsilon)X_i} - 1/2)$.

12. En déduire une majoration de $\mathbb{E}(|\hat{\beta}_n - \beta^*|)$.

13. En utilisant ce qui précède, construire une majoration de $\mathbb{E}(R(\hat{g}_n) - R^*)$, plus fine que la précédente.

14. Essayer de commenter les mérites respectifs des deux approches (en termes d'hypothèses nécessaires, extensions possibles, vitesses d'apprentissage, etc.).

SOLUTION

1. Notons $G : u \mapsto \mathbb{E}(|Z - u|)$. G étant convexe (moyenne de fonctions convexes), propre ($G(u) \rightarrow +\infty$ lorsque $|u| \rightarrow +\infty$, en utilisant l'inégalité triangulaire), on a $m \in \arg \min G$ si et seulement si $G'_+(m) \geq 0$ et $G'_-(m) \leq 0$ (dérivées à gauche et à droite).

Pour un m quelconque, on a

$$\begin{aligned} G(m) &= \mathbb{E}((m - Z)\mathbb{1}_{Z < m}) + \mathbb{E}((Z - m)\mathbb{1}_{m < Z}) \\ &= m - \mathbb{E}(Z) + 2\mathbb{E}((Z - m)\mathbb{1}_{Z < m}). \end{aligned}$$

Pour $\delta > 0$, on a alors

$$\begin{aligned} G(m + \delta) - G(m) &= \delta + 2\mathbb{E}((Z - m)\mathbb{1}_{m < Z < m + \delta}) - 2\delta\mathbb{P}(Z > m + \delta) \\ &= \delta + o(\delta) - 2\delta\mathbb{P}(Z > m) + o(\delta). \end{aligned}$$

On en déduit

$$G'_+(m) = 1 - 2\mathbb{P}(Z > m) = 2(\mathbb{P}(Z \leq m) - 1/2).$$

Similairement, on a

$$G'_-(m) = 2(1/2 - \mathbb{P}(Z \geq m)).$$

2. Y étant à valeurs non discrètes, il s'agit d'un problème de régression. La fonction $f : x \mapsto (\mathbb{1}_{|x| \leq 1} + |x|^{-3}\mathbb{1}_{|x| > 1})/3$ est bien une densité (mesurable, positive, intègre à 1). Elle est par ailleurs symétrique. On vérifie ensuite
- $\mathbb{E}(E) = (2/3) \left(\int_0^1 x dx + \int_1^{+\infty} x^{-2} dx \right) = (2/3)(3/2) = 1$,
 - pour $t > 0$, $\mathbb{P}(|E| \leq t) \leq (2/3)t$.
 - pour $t \leq 1$, $\mathbb{P}(|E| \leq t) = 2/3t$.
3. Pour le E de la question du dessus, on a $\mathbb{E}(E^2) = +\infty$. Si g est un prédicteur, on a

$$\mathbb{E}((g(X) - Y)^2 | X) \geq \frac{1}{2}\mathbb{E}(E^2 | X) - 2(g(X) - \beta^*X)^2 = +\infty.$$

On en déduit donc que $\mathbb{E}((g(X) - Y)^2) = +\infty$.

4. Soit g un prédicteur. On a

$$\begin{aligned} \mathbb{E}(c(g(X), Y)) &= \mathbb{E}(\mathbb{E}(|Y - g(X)| | X)) \\ &\geq \mathbb{E}(\mathbb{E}(|Y - g^*(X)| | X)) \quad (\text{d'après la Q1, en prenant } g^*(X) \in \text{Med}(Y | X)), \end{aligned}$$

avec égalité si et seulement si $g(X) \in \text{Med}(Y | X)$ P_X presque sûrement.

5. Si $Y = \beta^*X + E$, avec $E \perp\!\!\!\perp X$ et E symétrique, on a $\text{Med}(Y | X) = \beta^*X + \text{Med}(E)$ (au sens ensembles).

Pour $e > 0$, $\mathbb{P}(E \geq e) \leq \mathbb{P}(E \geq e \wedge 1) \leq 1/2(1 - c_E(e \wedge 1)) < 1/2$, en utilisant la symétrie de E et la minoration de sa masse autour de 0. Pareillement, si $e < 0$, on a $\mathbb{P}(E \leq e) < 1/2$. On en déduit que $\text{Med}(E) = \{0\}$, et donc que $g^*(X) = \beta^*X$ P_X p.s.. La classe $\mathcal{G}$ correspondante est naturellement $\mathcal{G} = \{x \mapsto \beta x\}$, pour $\beta \in \mathbb{R}$.

6. Par définition, on a

$$\begin{aligned} \hat{g}_n &\in \arg \min_{g \in \mathcal{G}} \frac{1}{n} \sum_{i=1}^n c(g(X_i), Y_i) \\ &\in \arg \min_{g \in \mathcal{G}} \frac{1}{n} \sum_{i=1}^n |Y_i - g(X_i)|. \end{aligned}$$

En assimilant $g = x \mapsto \beta x$ à β , on a

$$\hat{g}_n \in \arg \min_{\beta \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^n |Y_i - \beta X_i| := R_n(\beta).$$

Dans un modèle général, on aurait

$$R(\hat{\beta}_n) - R^* = \left[R(\hat{\beta}_n) - \inf_{\beta \in \mathbb{R}} \mathbb{E}(|Y - \beta X|) \right] + \left[\inf_{\beta \in \mathbb{R}} \mathbb{E}(|Y - \beta X|) - \mathbb{E}(|Y - m(X)|) \right],$$

avec $m(X) \in \text{Med}(Y | X)$ P_X p.s.. Dans le cadre du modèle de l'énoncé, le deuxième terme (approximation) est nul.

7. On a

$$\begin{aligned} R(\hat{g}_n) - R^* &= (R - R_n)(\hat{\beta}_n - \beta^*) + R_n(\hat{\beta}_n) - R_n(\beta^*) \\ &\leq (R - R_n)(\hat{\beta}_n - \beta^*) \\ &= \frac{1}{n} \sum_{i=1}^n \left(|Y_i - \hat{\beta}_n X_i| - |Y_i - \beta^* X_i| - \mathbb{E}_{X,Y} \left(|Y - \hat{\beta}_n X| - |Y - \beta^* X| \right) \right) \\ &\leq \sup_{|\beta| \leq M} \frac{1}{n} \sum_{i=1}^n (|Y_i - \beta X_i| - |Y_i - \beta^* X_i| - \mathbb{E}_{X,Y} (|Y - \beta X| - |Y - \beta^* X|)) \end{aligned}$$

Le Lemme de symétrisation donne alors

$$\mathbb{E}(R(\hat{g}_n) - R^*) \leq 2\mathbb{E}_{D_n} \mathbb{E}_\varepsilon \sup_{|\beta| \leq M} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i (|Y_i - \beta X_i| - |Y_i - \beta^* X_i|) \right|.$$

8. Comme

$$||Y_i - \beta X_i| - |Y_i - \beta^* X_i|| \leq |\beta - \beta^*| |X_i|,$$

le Lemme de contraction donne

$$\mathbb{E}_\varepsilon \sup_{|\beta| \leq M} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i (|Y_i - \beta X_i| - |Y_i - \beta^* X_i|) \right| \leq 2 \mathbb{E}_\varepsilon \sup_{|\beta| \leq M} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i |\beta - \beta^*| |X_i| \right|.$$

Or

$$\mathbb{E}_\varepsilon \sup_{|\beta| \leq M} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i |\beta - \beta^*| |X_i| \right| \leq 2M \mathbb{E}_\varepsilon \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i X_i \right| \quad (\text{on rappelle que } X_i \in]0, 1[),$$

d'où le résultat.

9. L'inégalité de Jensen donne

$$\begin{aligned} \mathbb{E}_\varepsilon \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i X_i \right| &\leq \frac{1}{n} \sqrt{\mathbb{E}_\varepsilon \left(\sum_{i=1}^n \varepsilon_i X_i \right)^2} \\ &\leq \frac{1}{n} \sqrt{\sum_{i=1}^n X_i^2} \leq \frac{1}{\sqrt{n}}, \end{aligned}$$

ce qui permet de conclure. On pourrait espérer du d/n , avec ici $d = 1$. Pour ce genre de résultat, la convexité de R en β^* est souvent requise, on la vérifiera par la suite.

10. On peut réécrire, pour un β quelconque

$$\begin{aligned} R_n(\beta) &= \frac{1}{n} \sum_{i=1}^n |Y_i - \beta X_i| \\ &= \frac{1}{n} \left(\sum_{i=1}^n X_i \right) \sum_{i=1}^n \frac{X_i}{\sum_{i=1}^n X_i} \left| \frac{Y_i}{X_i} - \beta \right| \\ &= \left(\sum_{i=1}^n X_i \right) \int \tilde{P}_n(du) |u - \beta|. \end{aligned}$$

On en déduit que $\hat{\beta}_n \in \arg \min R_n(\beta)$ minimise $\int \tilde{P}_n(du) |u - \beta|$, et est donc une médiane de $\tilde{P}_n$ (cf Q1).

11. Soit $\varepsilon > 0$. On a

$$\begin{aligned} \mathbb{P} \left(\hat{\beta}_n \leq \beta^* - \varepsilon \right) &\leq \mathbb{P} \left(\tilde{P}_n([-\infty, \beta^* - \varepsilon]) \geq 1/2 \right) \\ &= \mathbb{P} \left(\sum_{i=1}^n \frac{X_i}{\sum_{i=1}^n X_i} \mathbb{1}_{(Y_i/X_i) \leq \beta^* - \varepsilon} \geq 1/2 \right) \\ &= \mathbb{P} \left(\sum_{i=1}^n X_i \mathbb{1}_{Y_i \leq (\beta^* - \varepsilon) X_i} \geq \left(\sum_{i=1}^n X_i \right) / 2 \right) \\ &= \mathbb{P} \left(\sum_{i=1}^n Z_i \geq 0 \right). \end{aligned}$$

12. On remarque que les Z_i sont i.i.d., et comprises dans $[-1/2; 1/2]$. On va donc pouvoir régler la concentration en utilisant l'inégalité de Hoeffding. Reste à regarder l'espérance. On a

$$\begin{aligned}\mathbb{E}(Z_i) &= \mathbb{E}\left(X\left(\mathbb{1}_{\beta^*X+E \leq (\beta^*-\varepsilon)X} - (1/2)\right)\right) \\ &= \mathbb{E}\left(X\left(\mathbb{P}(E \leq -\varepsilon X \mid X) - (1/2)\right)\right).\end{aligned}$$

Comme $E \perp\!\!\!\perp X$, on a

$$\begin{aligned}\mathbb{P}(E \leq -\varepsilon X \mid X) &= \frac{1}{2}\mathbb{P}(|E| \geq \varepsilon X) \quad (\text{symétrie}) \\ &\leq \frac{1}{2}(1 - c_E \varepsilon X) \mathbb{1}_{\varepsilon X < 1} + \frac{1}{2}(1 - c_E) \mathbb{1}_{\varepsilon X \geq 1} \\ &\quad (\text{propriété de } E, \text{ en remarquant au passage que nécessairement } c_E \leq 1) \\ &= \frac{1}{2}(1 - c_E(\varepsilon X \wedge 1)).\end{aligned}$$

On en déduit

$$\begin{aligned}\mathbb{E}(Z_i) &\leq \mathbb{E}\left(X \frac{1}{2}(1 - c_E \varepsilon X \wedge 1)\right) - (1/2)\mathbb{E}(X) \\ &= -\frac{c_E}{2}\mathbb{E}(X(\varepsilon X \wedge 1)).\end{aligned}$$

Or,

$$\begin{aligned}\mathbb{E}(X(\varepsilon X \wedge 1)) &= \int_0^1 (\varepsilon u^2 \mathbb{1}_{\varepsilon u \leq 1} + u \mathbb{1}_{\varepsilon u > 1}) p_X(u) du \\ &\geq c_X \left[\varepsilon \frac{1}{3} \mathbb{1}_{\varepsilon \leq 1} + \left(\varepsilon \frac{\varepsilon^{-3}}{3} + \frac{(1 - \varepsilon^{-1})^2}{2} \right) \mathbb{1}_{\varepsilon > 1} \right] \\ &\geq \frac{\varepsilon c_X}{3}.\end{aligned}$$

On en déduit $\mathbb{E}(Z_i) \leq \frac{-c_E c_X \varepsilon}{6}$, et il s'ensuit

$$\begin{aligned}\mathbb{P}\left(\hat{\beta}_n \leq \beta^* - \varepsilon\right) &\leq \mathbb{P}\left(\sum_{i=1}^n Z_i - \mathbb{E}(Z_i) \geq \frac{n c_E c_X \varepsilon}{6}\right) \\ &\leq \exp\left(-2n(c_E c_X \varepsilon/6)^2\right) = \exp\left(-n(c_E c_X)^2 \varepsilon^2/18\right) \quad (\text{Hoeffding}).\end{aligned}$$

De la même manière, on montre que

$$\mathbb{P}\left(\hat{\beta}_n \geq \beta^* + \varepsilon\right) \leq \exp\left(-n(c_E c_X)^2 \varepsilon^2/18\right).$$

On en déduit enfin que

$$\begin{aligned}\mathbb{E}(|\hat{\beta}_n - \beta^*|) &= \int_0^{+\infty} \mathbb{P}(|\hat{\beta}_n - \beta^*| > t) dt \\ &\leq 2 \int_0^{+\infty} \exp(-n(c_E c_X)^2 t^2 / 18) \\ &= \frac{3\sqrt{2\pi}}{c_E c_X \sqrt{n}}.\end{aligned}$$

13. Pour conclure, il faut relier excès de risque et écart entre paramètres. Notons $\delta = (\beta - \beta^*)$, et supposons $\delta > 0$. On a, en jouant avec la symétrie de f

$$\begin{aligned}R(\beta) - R(\beta^*) &= \int_0^1 p_X(x) dx \int_{\mathbb{R}} f(e) de (|-\delta x + e| - |e|) \\ &\leq C_X \left(\int_0^1 dx \int_{-\infty}^0 f(e) de ((\delta x - e) + e) + \int_0^1 dx \int_0^{\delta x} f(e) de ((\delta x - e) - e) \right) \\ &\quad + C_X \int_0^1 dx \int_{\delta x}^{+\infty} -\delta x f(e) de \\ &= C_X \left(\int_0^1 dx (\delta x) \int_0^{\delta x} f(e) de + \int_0^1 dx \int_0^{\delta x} (\delta x - 2e) f(e) de \right) \\ &= 2C_X \int_0^1 dx \int_0^{\delta x} (\delta x - e) f(e) de \\ &= 2C_X \int_0^{+\infty} f(e) de \int_{e/\delta}^1 (\delta x - e) dx \\ &= 2C_X \int_0^{+\infty} f(e) \delta^{-1} \frac{(\delta - e)^2}{2} \mathbb{1}_{e \leq \delta} de \\ &\leq \frac{C_X \delta}{2} \mathbb{P}(|E| \leq \delta) \\ &\leq \frac{C_X C_E \delta^2}{2} \quad (\text{propriété de } E).\end{aligned}$$

Par symétrie de E , $\mathbb{E}(|(\beta^* - \beta)X + E|) = \mathbb{E}(|(\beta - \beta^*)X + E|)$, on en déduit alors que, pour tout $\beta \in \mathbb{R}$,

$$R(\beta) - R(\beta^*) \leq \frac{C_X C_E (\beta - \beta^*)^2}{2}.$$

Pour conclure, il faut repasser par les déviations. On montre comme en question précédente que, pour tout $t > 0$,

$$\begin{aligned}\mathbb{P}(R(\beta) - R(\beta^*) > t) &\leq \mathbb{P}\left(|\hat{\beta}_n - \beta^*| > \sqrt{2t/(C_X C_E)}\right) \\ &\leq 2 \exp(-n(c_E c_X)^2 t / (9C_E C_X)).\end{aligned}$$

Comme précédemment, on en déduit que

$$\begin{aligned}\mathbb{E}(R(\hat{\beta}_n) - R(\beta^*)) &= \int_0^{+\infty} \mathbb{P}(R(\hat{\beta}_n) - R(\beta^*) > t) dt \\ &\leq \frac{18C_X C_E}{n(c_E c_X)^2}.\end{aligned}$$

14. La méthode 'brute force' (première) a un gros avantage : aucune structure sur X n'est requise (notamment la majoration/minoration de la densité). En ce sens, la vitesse en $1/\sqrt{n}$ est uniforme sur l'ensemble des lois qui s'écrivent sous la forme $Y = \beta^* X + E$, avec $\text{Med}(E|X) = 0$. La restriction $|\beta^*| \leq M$ est artificielle, on aurait pu s'en débarrasser techniquement en découpant suivant $|\hat{\beta}_n - \beta^*| \geq M_0$ et $|\hat{\beta}_n - \beta^*| \leq M_0$, pour un M_0 bien choisi.

La deuxième méthode donne des vitesses plus rapides, mais requiert des hypothèses structurelles supplémentaires sur (X, E) (quasi-uniformité de X , suffisamment de masse autour de 0 pour E , indépendance des 2). On pourrait se passer de la symétrie et de l'indépendance de E , ainsi que de la quasi-uniformité de X en travaillant conditionnellement à X , avec des hypothèses portant sur le remplissage de $E | X$ autour de 0 qui elles sont cruciales. Sous ces hypothèses, la vitesse d'apprentissage est en $1/n$. Pour se convaincre qu'on ne peut pas étendre cette vitesse rapide à l'ensemble des lois pour lesquelles la première approche est valide, on peut regarder l'ensemble de lois suivantes : $\beta^* = 0$, $X = 0$, $E_{-1} = (1/2 + c/\sqrt{n})\delta_{-1} + (1/2 - c/\sqrt{n})\delta_{+1}$, et $E_1 \sim -E_{-1}$ et faire un peu de Bayésien comme d'habitude.