Performances statistiques d'un estimateur Bayésien

Michaël CHICHIGNOUD

LATP
Université de Provence
Aix-Marseille 1

Le Mont Dore, Mai 2010

Plan

1. Modèle
   - Estimateur Bayésien
   - Condition sur la Densité g
2. Adaptation
   - Idée de la preuve
3. Extensions

Estimation de l'espérance d'une loi gaussienne

On observe $(Y_1, ..., Y_n)$ i.i.d. avec $Y \sim \mathcal{N}(\theta, \sigma^2)$. On cherche à estimer θ. Ici, on connait l'estimateur optimal (au sens des vitesses de convergences),

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i \xrightarrow[n\to\infty]{1/\sqrt{n}} \theta,$$

On peut écrire ce modèle comme

$$Y_i = \theta + \varepsilon_i, \quad i = 1, ..., n$$

avec $\varepsilon \sim \mathcal{N}(0, \sigma^2)$.

Estimation de l'espérance d'une loi gaussienne

On observe $(Y_1, ..., Y_n)$ i.i.d. avec $Y \sim \mathcal{N}(\theta, \sigma^2)$. On cherche à estimer θ. Ici, on connait l'estimateur optimal (au sens des vitesses de convergences),

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i \xrightarrow[n \to \infty]{1/\sqrt{n}} \theta,$$

On peut écrire ce modèle comme

$$Y_i = \theta + \varepsilon_i, \quad i = 1, ..., n$$

avec $\varepsilon \sim \mathcal{N}(0, \sigma^2)$.

Régression paramétrique avec bruit uniforme

$$Y_i = \theta + U_i, \quad i = 1, ..., n$$

avec $U_i \sim \mathcal{U}_{[-0.5,0.5]}$

- On peut construire un estimateur basé sur la "moyenne"

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow[n\to\infty]{1/\sqrt{n}} \theta$$

- On peut trouver un estimateur plus rapide

$$\hat{\theta} = \max_{i=\overline{1,n}} Y_i - \frac{1}{2} \xrightarrow[n\to\infty]{1/n} \theta$$

Régression paramétrique avec bruit uniforme

$$Y_i = \theta + U_i, \quad i = 1, \ldots, n$$

avec $U_i \sim \mathcal{U}_{[-0.5,0.5]}$ • On peut construire un estimateur basé sur la "moyenne"

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow[n\to\infty]{1/\sqrt{n}} \theta$$

• On peut trouver un estimateur plus rapide

$$\hat{\theta} = \max_{i=\overline{1,n}} Y_i - \frac{1}{2} \xrightarrow[n\to\infty]{1/n} \theta$$

Régression paramétrique avec bruit uniforme

$$Y_i = \theta + U_i, \quad i = 1, ..., n$$

avec $U_i \sim \mathcal{U}_{[-0.5,0.5]}$ • On peut construire un estimateur basé sur la "moyenne"

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i \xrightarrow[n\to\infty]{1/\sqrt{n}} \theta$$

- On peut trouver un estimateur plus rapide

$$\hat{\theta} = \max_{i=\overline{1,n}} Y_i - \frac{1}{2} \xrightarrow[n\to\infty]{1/n} \theta$$

Régression

On dispose d'un échantillon d'observations indépendantes

$$\big((X_1, Y_1), ..., (X_n, Y_n)\big)_{i=1,...,n}, \quad X \in [0,1]^d, \ Y \in \mathbb{R}$$

où

$$\mathbb{P}_{Y_i}(dx) = g\big(x, f(X_i)\big)\, dx, \quad i = 1, \ldots, n, .$$

où $f : [0,1]^d \to \mathbb{R}$ avec $f \in \mathbb{H}_d(\beta, L, M)$ isotrope et $X_i \in \left\{1/n^{1/d}, 2/n^{1/d}, \ldots, 1\right\}^d$ est le design régulier.

Exemples :

$$Y_i = f(X_i) + \varepsilon_i, \quad \varepsilon \sim \mathcal{N}(0, \sigma), \quad i = 1, ..., n$$

$$Y_i = f(X_i) + U_i, \quad U_i \sim \mathcal{U}_{[-0.5, 0.5]}, \quad i = 1, ..., n$$

$$Y_i = f(X_i) \times U_i, \quad U_i \sim \mathcal{U}_{[0,1]}, \quad i = 1, ..., n$$

Régression

On dispose d'un échantillon d'observations indépendantes

$$\big((X_1, Y_1), ..., (X_n, Y_n)\big)_{i=1,...,n}, \quad X \in [0,1]^d, \; Y \in \mathbb{R}$$

où

$$\mathbb{P}_{Y_i}(dx) = g\big(x, f(X_i)\big)dx, \quad i = 1, \ldots, n, .$$

où $f : [0,1]^d \to \mathbb{R}$ avec $f \in \mathbb{H}_d(\beta, L, M)$ isotrope et $X_i \in \left\{1/n^{1/d}, 2/n^{1/d}, \ldots, 1\right\}^d$ est le design régulier.

Exemples :

$$Y_i = f(X_i) + \varepsilon_i, \quad \varepsilon \sim \mathcal{N}(0, \sigma), \quad i = 1, ..., n$$
$$Y_i = f(X_i) + U_i, \quad U_i \sim \mathcal{U}_{[-0.5,0.5]}, \quad i = 1, ..., n$$
$$Y_i = f(X_i) \times U_i, \quad U_i \sim \mathcal{U}_{[0,1]}, \quad i = 1, ..., n$$

Approche localement polynomiale

On cherche à estimer f en un point $y \in [0,1]^d$.
Idée : Sur un voisinage de y de taille h^d, noté $V_y(h)$, on approche f par un polynôme f_θ, $\theta \in \Theta = [-M, M]^{D_\beta}$, tel que $\forall x \in V_y(h)$

$$\left| f(x) - f_\theta(x) \right| \leq L\, d\, h^\beta.$$

L'estimateur Bayésien nous permet d'estimer les coefficients θ.
Définissons la pseudo Vraisemblance pour $t \in \Theta$

$$L_h(t) = L_h(t, Y^{(n)}) = \prod_{X_i \in V_y(h)} g(Y_i, f_t(X_i))$$

Approche localement polynomiale

On cherche à estimer f en un point $y \in [0,1]^d$.
Idée : Sur un voisinage de y de taille h^d, noté $V_y(h)$, on approche f par un polynôme f_θ, $\theta \in \Theta = [-M, M]^{D_\beta}$, tel que $\forall x \in V_y(h)$

$$\left| f(x) - f_\theta(x) \right| \leq L\, d\, h^\beta .$$

L'estimateur Bayésien nous permet d'estimer les coefficients θ.
Définissons la pseudo Vraisemblance pour $t \in \Theta$

$$L_h(t) = L_h\left(t, Y^{(n)}\right) = \prod_{X_i \in V_y(h)} g\left(Y_i, f_t(X_i)\right)$$

Estimateur Bayésien

On definit

$$\hat{\theta} = \arg\min_{t\in\Theta} \pi_h(t), \quad \pi_h(t) = \int_\Theta \|t-u\|_1 \, \left(L_h(u)\right)^{1/m} du, \quad t \in \Theta$$

Pour tout $h \in (0,1)$, on peut définir l'estimateur Bayésien associé

$$\hat{f}^{(h)}(y) = f_{\hat{\theta}}(y),$$

But : A partir de cette famille d'estimateurs $\{\hat{f}^{(h)}(y)\}_{h\in H}$, on veut choisir le "meilleur" (la meilleure fenêtre).

- β, L sont connues, Approche minimax
- β, L inconnues, Adaptation au sens minimax.

Estimateur Bayésien

On definit

$$\hat{\theta} = \arg\min_{t\in\Theta} \pi_h(t), \quad \pi_h(t) = \int_\Theta \|t-u\|_1 \left(L_h(u)\right)^{1/m} du, \quad t \in \Theta$$

Pour tout $h \in (0,1)$, on peut définir l'estimateur Bayésien associé

$$\hat{f}^{(h)}(y) = f_{\hat{\theta}}(y),$$

But : A partir de cette famille d'estimateurs $\{\hat{f}^{(h)}(y)\}_{h\in H}$, on veut choisir le "meilleur" (la meilleure fenêtre).

- β, L sont connues, Approche minimax
- β, L inconnues, Adaptation au sens minimax.

Estimateur Bayésien

On definit

$$\hat{\theta} = \arg\min_{t \in \Theta} \pi_h(t), \quad \pi_h(t) = \int_\Theta \|t - u\|_1 \left(L_h(u)\right)^{1/m} du, \quad t \in \Theta$$

Pour tout $h \in (0,1)$, on peut définir l'estimateur Bayésien associé

$$\hat{f}^{(h)}(y) = f_{\hat{\theta}}(y),$$

But : A partir de cette famille d'estimateurs $\{\hat{f}^{(h)}(y)\}_{h \in H}$, on veut choisir le "meilleur" (la meilleure fenêtre).

- β, L sont connues, Approche minimax
- β, L inconnues, Adaptation au sens minimax.

Supposons que la vitesse minimax est $\varphi_{n,\gamma}(\beta) = n^{-\frac{\beta}{\gamma\beta+d}}$.
Pour tout $\theta \in \Theta$, soit $U_n = N_h\left[\Theta - \theta\right]$ avec $N_h = \left(nh^d\right)^{1/\gamma}$ où $\gamma > 0$ et $\forall u \in U_n$

$$Z_{n,\theta}(u) = \frac{L_h\left(\theta + u\, N_h^{-1}\right)}{L_h\left(\theta\right)},$$

Conditions 1 et 2

On suppose qu'il existe des constantes $\alpha > 0$, $m \geq 1$ et $c_1, c_2, c_3, C_1, C_2, s_1, s_2 > 0$ telle que $\forall n > 1$, $f \in \mathbb{H}_d(\beta, L, M)$, $\theta \in \Theta$ et $h \in \mathcal{H}_n$, on a

$$\textbf{(1)} \quad \mathbb{E}_f \left[1 - Z_{n,\theta}^{1/m}(u)\right]_+ \leq C_1 \, \mathcal{E}_h^{c_1} \|u\|_1^{\alpha}, \quad u \in U_n : \ \|u\|_1 < s_1,$$

$$\textbf{(2)} \quad \mathbb{E}_f Z_{n,\theta}^{1/m}(u) \leq C_2 \, \mathcal{E}_h^{c_2} \ \exp\left\{-c_3 \|u\|_1^{\gamma}\right\}, \ \forall u \in U_n : \ \|u\|_1 \geq s_2,$$

où $a_+ = \max(a, 0)$, $a \in \mathbb{R}$. On a besoin de connaître explicitement les constantes c_3, s_2, m, α.

Méthode de Lepski (Lepski and al. '97)

On se place dans le cas où M est connue et β, L inconnues.

On construit une grille $h^{(l)} = h_{\max} 2^{-l}$, $\forall l = 0, \ldots, M_n$.

A partir de la famille d'estimateurs Bayésiens $\{\hat{f}^l(y) = \hat{f}^{h^{(l)}}(y)\}_{l \in \mathbb{G}_n}$, on veut trouver un estimateur qui atteint la vitesse adaptative

$$\phi_{n,\gamma}(\beta) = \left(n^{-1} \rho_{n,\gamma}(\beta)\right)^{\frac{\beta}{\gamma\beta+d}}, \quad \rho_{n,\gamma}(\beta) = 1 + \ln \frac{\varphi_{n,\gamma}(\beta)}{\varphi_{n,\gamma}(b)} \sim (b-\beta) \ln n$$

pour tout $\beta \in]0, b]$ et $L \in [l_*, l^*]$.

Méthode de Lepski (Lepski and al. '97)

On se place dans le cas où M est connue et β, L inconnues.

On construit une grille $h^{(l)} = h_{\max} 2^{-l}$, $\forall l = 0, \dots, M_n$.
A partir de la famille d'estimateurs Bayésiens $\{\hat{f}^l(y) = \hat{f}^{h^{(l)}}(y)\}_{l \in \mathbb{G}_n}$, on veut trouver un estimateur qui atteint la vitesse adaptative

$$\phi_{n,\gamma}(\beta) = \left(n^{-1}\rho_{n,\gamma}(\beta)\right)^{\frac{\beta}{\gamma\beta+d}}, \quad \rho_{n,\gamma}(\beta) = 1 + \ln \frac{\varphi_{n,\gamma}(\beta)}{\varphi_{n,\gamma}(b)} \sim (b-\beta)\ln n$$

pour tout $\beta \in]0, b]$ et $L \in [l_*, l^*]$.

Procédure :

On définit la quantité suivante, $\forall l = 0, \ldots, M_n$:

$$S_n(l) = \left[\frac{1 + l \ln 2}{n\left(h^{(l)}\right)^d}\right]^{1/\gamma},$$

où $M_n \sim \ln n$. Soit

$$\rho = \left(s_2^\gamma \vee \frac{2^{|\nabla|+2\gamma+1}(2\gamma + 2dq)}{c_3 \gamma \left(1 \wedge \alpha D_b^{-1}\right)}\right)^{\frac{1}{\gamma}}$$

Définissons l'indice adaptatif

$$\hat{k} = \min\left\{k = 0, \ldots, M_n \;:\; \left|\hat{f}^k(y) - \hat{f}^l(y)\right| \leq \rho\, S_n(l)\right.$$
$$\left.\forall l = k+1, \ldots, M_n\right\}.$$

Procédure :

On définit la quantité suivante, $\forall l = 0, \ldots, M_n$:

$$S_n(l) = \left[\frac{1 + l \ln 2}{n\left(h^{(l)}\right)^d}\right]^{1/\gamma},$$

où $M_n \sim \ln n$. Soit

$$\rho = \left(s_2^\gamma \vee \frac{2^{|\nabla|+2\gamma+1}(2\gamma + 2dq)}{c_3\gamma\left(1 \wedge \alpha D_b^{-1}\right)}\right)^{\frac{1}{\gamma}}$$

Définissons l'indice adaptatif

$$\hat{k} = \min\left\{k = 0, \ldots, M_n : \left|\hat{f}^k(y) - \hat{f}^l(y)\right| \le \rho\, S_n(l) \quad \forall l = k+1, \ldots, M_n\right\}.$$

Procédure :

On définit la quantité suivante, $\forall l = 0, \ldots, M_n$:

$$S_n(l) = \left[\frac{1 + l \ln 2}{n \left(h^{(l)} \right)^d} \right]^{1/\gamma},$$

où $M_n \sim \ln n$. Soit

$$\rho = \left(s_2^{\gamma} \vee \frac{2^{|\nabla| + 2\gamma + 1} (2\gamma + 2dq)}{c_3 \gamma \left(1 \wedge \alpha D_b^{-1} \right)} \right)^{\frac{1}{\gamma}}$$

Définissons l'indice adaptatif

$$\hat{k} = \min \Big\{ k = 0, \ldots, M_n \ : \ \left| \hat{f}^k(y) - \hat{f}^l(y) \right| \leq \rho \, S_n(l) \quad \forall l = k+1, \ldots, M_n \Big\} .$$

Théorème

Pour $q \geq 1$ et $\hat{f}^{\hat{k}}(y)$ l'estimateur choisi par la procédure,
Alors $\forall n \in \mathbb{N}^*$ et $(\beta, L) \in J \times \mathcal{L}$:

$$\sup_{f \in \mathbb{H}_d(\beta,L,M)} \mathbb{E}_f \left| \hat{f}^{\hat{k}}(y) - f(y) \right|^q \leq C^* \phi_{n,\gamma}^q(\beta),$$

où $J =]0, b]$, $\mathcal{L} = [l_*, l^*]$ et $C^* = C^*(q, \beta, L, M, \rho)$ est une constante qui ne dépend pas de n.

Idée de la Preuve

Théorème

Soit $f \in \mathbb{H}_d(\vec{\beta}, L, M)$ *et le vecteur* $h \in (0,1)^d$.
Alors pour tout $\varepsilon \geq \varepsilon_0(h)$, $\forall n \in \mathbb{N}^*$ *:*

$$\mathbb{P}_f\left(\left|\hat{f}_h(y) - f(y)\right| \geq \frac{\varepsilon}{N_h}\right) \leq M_1 \varepsilon^D \exp\left\{-\varepsilon^\gamma\right\},$$

Ce théorème généralise un résultat disponible dans le livre de Ibragimov et Has'minskii ('83) *Statistical estimation*.

Idée de la Preuve

Théorème

Soit $f \in \mathbb{H}_d(\vec{\beta}, L, M)$ *et le vecteur* $h \in (0,1)^d$.
Alors pour tout $\varepsilon \geq \varepsilon_0(h)$, $\forall n \in \mathbb{N}^*$:

$$\mathbb{P}_f\left(\left|\hat{f}_h(y) - f(y)\right| \geq \frac{\varepsilon}{N_h}\right) \leq M_1 \varepsilon^D \exp\left\{-\varepsilon^\gamma\right\},$$

Ce théorème généralise un résultat disponible dans le livre de Ibragimov et Has'minskii ('83) *Statistical estimation*.

Idée de la Preuve

Théorème

Soit $f \in \mathbb{H}_d(\vec{\beta}, L, M)$ *et le vecteur* $h \in (0,1)^d$.
Alors pour tout $\varepsilon \geq \varepsilon_0(h)$, $\forall n \in \mathbb{N}^*$ *:*

$$\mathbb{P}_f\left(\left|\hat{f}_h(y) - f(y)\right| \geq \frac{\varepsilon}{N_h}\right) \leq M_1 \varepsilon^D \exp\left\{-\varepsilon^\gamma\right\},$$

Ce théorème généralise un résultat disponible dans le livre de Ibragimov et Has'minskii ('83) *Statistical estimation*.

$$\mathbb{P}_f\left(\left|\hat{f}_h(y) - f(y)\right| \geq \frac{\varepsilon}{N_h}\right) \leq \mathbb{P}_f\left(N_h \,\|\hat{\theta} - \theta\|_1 \geq \varepsilon\right) \leq \mathbb{P}_f\left(\inf_{N_h \,\|t-\theta\|_1 \geq \varepsilon} \pi_h(t) \leq \pi_h(\theta)\right).$$

Ainsi après quelques lignes, on obtient

$$\mathbb{P}_f\left(\left|\hat{f}_h(y) - f(y)\right| \geq \frac{\varepsilon}{N_h}\right) \leq \mathbb{P}_f\left\{\int_{U_n} Z_{n,\theta}^{1/m}(v)dv < \nu/2\right\} + \mathbb{P}_f\left\{\int_{U_n \cap \{\|u\|_1 > \varepsilon\}} \|u\|_1 Z_{n,\theta}^{1/m}(u)du > \nu\,\varepsilon/12\right\}.$$

$$
\begin{aligned}
\mathbb{P}_f\left(\left|\hat{f}_h(y) - f(y)\right| \geq \frac{\varepsilon}{N_h}\right) &\leq \mathbb{P}_f\left(N_h \, \|\hat{\theta} - \theta\|_1 \geq \varepsilon\right) \\
&\leq \mathbb{P}_f\left(\inf_{N_h \, \|t-\theta\|_1 \geq \varepsilon} \pi_h(t) \leq \pi_h(\theta)\right).
\end{aligned}
$$

Ainsi après quelques lignes, on obtient

$$
\begin{aligned}
&\mathbb{P}_f\left(\left|\hat{f}_h(y) - f(y)\right| \geq \frac{\varepsilon}{N_h}\right) \\
&\leq \mathbb{P}_f\left\{\int_{U_n} Z_{n,\theta}^{1/m}(v)dv < \nu/2\right\} \\
&\quad + \mathbb{P}_f\left\{\int_{U_n \cap \{\|u\|_1 > \varepsilon\}} \|u\|_1 Z_{n,\theta}^{1/m}(u)du > \nu\,\varepsilon/12\right\}.
\end{aligned}
$$

Problèmes ouverts et extensions

- Restrictions sur Hölder isotropes,
- Approximation de f par f_θ où $\theta \in Cp^D$,
- Conditions 1 et 2 : lien avec la distance d'Hellinger,
- $Z(u) = \frac{L_h(\theta + u\, N_h^{-1})}{L_h(\theta)}$,
- Estimation de M,
- Inégalité d'oracle,
- Hypohèse d'indépendance.

Problèmes ouverts et extensions

- Restrictions sur Hölder isotropes,
- Approximation de f par f_θ où $\theta \in Cp^D$,
- Conditions 1 et 2 : lien avec la distance d'Hellinger,
- $Z(u) = \frac{L_h(\theta + u\, N_h^{-1})}{L_h(\theta)}$,
- Estimation de M,
- Inégalité d'oracle,
- Hypohèse d'indépendance.

Problèmes ouverts et extensions

- Restrictions sur Hölder isotropes,
- Approximation de f par f_θ où $\theta \in Cp^D$,
- Conditions 1 et 2 : lien avec la distance d'Hellinger,
- $Z(u) = \frac{L_h(\theta + u\, N_h^{-1})}{L_h(\theta)}$,
- Estimation de M,
- Inégalité d'oracle,
- Hypohèse d'indépendance.

Problèmes ouverts et extensions

- Restrictions sur Hölder isotropes,
- Approximation de f par f_θ où $\theta \in Cp^D$,
- Conditions 1 et 2 : lien avec la distance d'Hellinger,
- $Z(u) = \frac{L_h(\theta + u\, N_h^{-1})}{L_h(\theta)}$,
- Estimation de M,
- Inégalité d'oracle,
- Hypohèse d'indépendance.

Problèmes ouverts et extensions

- Restrictions sur Hölder isotropes,
- Approximation de f par f_θ où $\theta \in Cp^D$,
- Conditions 1 et 2 : lien avec la distance d'Hellinger,
- $Z(u) = \frac{L_h(\theta + u\, N_h^{-1})}{L_h(\theta)}$,
- Estimation de M,
- Inégalité d'oracle,
- Hypohèse d'indépendance.

Problèmes ouverts et extensions

- Restrictions sur Hölder isotropes,
- Approximation de f par f_θ où $\theta \in Cp^D$,
- Conditions 1 et 2 : lien avec la distance d'Hellinger,
- $Z(u) = \frac{L_h(\theta + u\, N_h^{-1})}{L_h(\theta)}$,
- Estimation de M,
- Inégalité d'oracle,
- Hypohèse d'indépendance.

Problèmes ouverts et extensions

- Restrictions sur Hölder isotropes,
- Approximation de f par f_θ où $\theta \in Cp^D$,
- Conditions 1 et 2 : lien avec la distance d'Hellinger,
- $Z(u) = \frac{L_h(\theta + u\, N_h^{-1})}{L_h(\theta)}$,
- Estimation de M,
- Inégalité d'oracle,
- Hypohèse d'indépendance.

Problèmes ouverts et extensions

- Restrictions sur Hölder isotropes,
- Approximation de f par f_θ où $\theta \in Cp^D$,
- Conditions 1 et 2 : lien avec la distance d'Hellinger,
- $Z(u) = \frac{L_h(\theta + u\, N_h^{-1})}{L_h(\theta)}$,
- Estimation de M,
- Inégalité d'oracle,
- Hypohèse d'indépendance.

MERCI pour votre attention ! ! !