

Sélection d'estimateurs et tests multiples

Matthieu Lerasle

CNRS, Nice

29 mars 2016

Plan de la présentation

- 1 Le problème de sélection d'estimateurs linéaires de la densité.
 - Sélection par pénalisation.
 - Sélection par validation croisée
- 2 Estimation sous-Gaussienne.
- 3 Taux de séparation pour les tests multiples.

Exemples 1: estimateurs par projection

Soit S un sous-espace vectoriel de $L^2 \cap L^\infty$ de dimension finie D_S .
L'estimateur par projection sur S est défini par

$$\hat{f}_S = \arg \min_{t \in S} P_n \gamma(t) = \arg \min_{t \in S} \left\{ \|t\|^2 - \frac{2}{n} \sum_{i=1}^n t(X_i) \right\} .$$

C'est un estimateur linéaire de f , en effet, pour toute base orthonormée $\varphi_S = (\varphi_1, \dots, \varphi_{D_S})$ de S , on peut définir

$K_{\varphi_S}(x, y) = \sum_{i=1}^{D_S} \varphi_i(x) \varphi_i(y)$ et vérifier que

$$\hat{f}_S = \hat{f}_{K_{\varphi_S}} .$$

Exemple 2 : estimateurs à noyau

Supposons $\mathbb{X} \subset \mathbb{R}$, soit $k : \mathbb{R} \rightarrow \mathbb{R}$ une fonction paire de $L^1 \cap L^\infty$ et soit $h > 0$. L'estimateur à noyau de f associé au noyau k et la fenêtre de lissage h est défini par

$$\forall x \in \mathbb{X}, \quad \hat{f}_{k,h}(x) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x - X_i}{h}\right) .$$

C'est un estimateur linéaire de f , en effet, en définissant, pour tout $(x, y) \in \mathbb{X}^2$, $K_{k,h}(x, y) = \frac{1}{h} k\left(\frac{x-y}{h}\right)$, on a immédiatement

$$\hat{f}_{k,h} = \hat{f}_{K_{k,h}} .$$

Estimateurs de Pinsker-Stein

Soit $\mathbb{X} = [0, 1]$ et soit φ la base de Fourier, i.e. $\varphi_0 = 1$ et pour tout $k \geq 1$ et tout $x \in \mathbb{X}$,

$$\varphi_{1,k}(x) = \sqrt{2} \cos(2\pi kx), \quad \varphi_{2,k}(x) = \sqrt{2} \sin(2\pi kx) .$$

Soit $p \in \mathbb{N} \setminus \{0\}$ et soit $\omega = (\omega_n)_{n \in \mathbb{N}} \in [0, 1]^{\mathbb{N}}$ une suite décroissante, l'estimateur de Pinsker-Stein \hat{f}_ω de f est défini pour tout $x \in \mathbb{X}$ par

$$\hat{f}_\omega(x) = 1 + \sum_{\ell=1}^p \omega_\ell [(P_n \varphi_{1,\ell}) \varphi_{1,\ell}(x) + (P_n \varphi_{2,\ell}) \varphi_{2,\ell}(x)] ,$$

c'est l'estimateur linéaire associé à la fonction

$$K_\omega(x, y) = \varphi_0(x)\varphi_0(y) + \sum_{\ell=1}^p \omega_\ell [\varphi_{1,\ell}(x)\varphi_{1,\ell}(y) + \varphi_{2,\ell}(x)\varphi_{2,\ell}(y)] .$$

Sélection d'estimateurs linéaires

Soit $(\hat{f}_K)_{K \in \mathcal{K}}$ une collection finie d'estimateurs linéaires de f , on veut sélectionner l'un des noyaux qu'on notera \hat{K} , de manière à ce que

$$R(\hat{f}_{\hat{K}}) \leq L \inf_{K \in \mathcal{K}} R(\hat{f}_K) + \rho .$$

L : "leading constant".

ρ : terme de reste, typiquement $\approx \frac{(\log n)^\square}{n}$.

Sélection par pénalisation

Idée de Mallows :

On veut minimiser $\|\widehat{f}_K - f\|^2$ ou la perte $\ell(\widehat{f}_K) = P\gamma(\widehat{f}_K)$.

On ne connaît pas P , on estime $\ell(\widehat{f}_K)$ par $P_n\gamma(\widehat{f}_K)$.

On obtient un estimateur biaisé de $\ell(\widehat{f}_K)$, on corrige en ajoutant une pénalité $\text{pen} : \mathcal{K} \rightarrow \mathbb{R}$ vérifiant typiquement $\text{pen}(K) \geq \mathbb{E} [\text{pen}_{\text{id}}(K)]$ où la "pénalité idéale" [Arlot (2009)] $\text{pen}_{\text{id}}(K) = (P - P_n)\gamma(\widehat{f}_K)$.

On choisit finalement

$$\widehat{K} = \arg \min_{K \in \mathcal{K}} \left\{ P_n\gamma(\widehat{f}_K) + \text{pen}(K) \right\} .$$

Que doit satisfaire une pénalité?

Par construction, en notant $\mathcal{D}_{\text{pen}}(K) = \text{pen}(K) - \text{pen}_{\text{id}}(K)$, pour tout $K \in \mathcal{K}$, on a

$$P\gamma(\widehat{f}_{\widehat{K}}) \leq P\gamma(\widehat{f}_K) + \mathcal{D}_{\text{pen}}(K) - \mathcal{D}_{\text{pen}}(\widehat{K}) ,$$

d'où correction de l'heuristique de Mallows par Barron, Birgé et Massart (1999): pénalités déterministes telles que

$$\mathbb{E} \left[\inf_{K \in \mathcal{K}} \mathcal{D}_{\text{pen}}(K) \right] \geq 0 .$$

Alors

$$R(\widehat{f}_{\widehat{K}}) \leq \inf_{K \in \mathcal{K}} \left\{ R(\widehat{f}_K) + \mathbb{E} [\mathcal{D}_{\text{pen}}(K)] \right\} .$$

Objectif alternatif

En réalité, il suffit d'avoir [Arlot, Bartlett 2011, Arlot HDR]:

$$\mathbb{E} \left[\sup_{(K, K') \in \mathcal{K}^2} \left\{ |\mathcal{D}_{\text{pen}}(K) - \mathcal{D}_{\text{pen}}(K')| - \alpha R(K) - \beta R(K') \right\} \right] \leq \Gamma, \quad (1)$$

avec $\alpha \geq 0$, $\beta \in (0, 1)$ et $\Gamma \geq 0$. Alors, on a une inégalité oracle

$$R(\hat{f}_{\hat{K}}) \leq L \inf_{K \in \mathcal{K}} R(\hat{f}_K) + \rho, \quad \text{avec} \quad L = \frac{1 + \alpha}{1 - \beta}, \quad \rho = \frac{\Gamma}{1 - \beta}.$$

Stratégie

Pour atteindre (1), on utilise la décomposition suivante [Arlot, Massart 2009]

$$\text{pen}_{\text{id}}(K) = P(\gamma(\hat{f}_K) - \gamma(f_K)) + P_n(\gamma(f_K) - \gamma(\hat{f}_K)) + (P_n - P)\gamma(f_K) ,$$

où, pour tout $x \in \mathbb{X}$, $f_K(x) = \mathbb{E}[K(x, X)]$.

On montre alors

$$\mathbb{E} \left[\sup_{K \in \mathcal{K}} \left\{ |p_1(K) - \mathbb{E}[p_1(K)]| - \frac{\beta}{3} R(K) \right\} \right] \leq \Gamma_1 , \quad (2)$$

$$\mathbb{E} \left[\sup_{K \in \mathcal{K}} \left\{ |p_2(K) - \mathbb{E}[p_2(K)]| - \frac{\beta}{3} R(K) \right\} \right] \leq \Gamma_2 , \quad (3)$$

$$\mathbb{E} \left[\sup_{(K, K') \in \mathcal{K}^2} \left\{ |\delta(K, K')| - \frac{\beta}{3} (R(K) + R(K')) \right\} \right] \leq \Gamma_3 . \quad (4)$$

Alors (2),(3) et (4) impliquent (1) si $\text{pen}(K) = \mathbb{E}[p_1(K)] + \mathbb{E}[p_2(K)]$.

Formules closes

Proposition (L., Magalhães, Reynaud-Bouret HDP VII)

Il existe des fonctions Φ_1, Φ_2 bornées et des U -statistiques totalement dégénérées d'ordre 2 U_1, U_2 (explicités dans l'articles) telles que

$$p_1(K) - \mathbb{E}[p_1(K)] = (P_n - P)\Phi_1(K) + U_1(K)$$

$$p_2(K) - \mathbb{E}[p_2(K)] = (P_n - P)\Phi_2(K) + U_2(K)$$

$$\mathbb{E}[p_1(K)] = \frac{PA_K^\Delta}{n}, \quad \mathbb{E}[p_2(K)] = \frac{2PK^\Delta - PA_K^\Delta}{n},$$

avec $A_K(x, y) = \int_{\mathbb{X}} K(x, z)K(y, z)\mu(dz)$ et, pour tout $g : \mathbb{X}^2 \rightarrow \mathbb{R}$, $g^\Delta(x) = g(x, x)$.

Sélection optimale d'un estimateur à noyau

Avec Bernstein + concentration des U -statistiques totalement dégénérées d'ordre 2 [Giné, Zinn 2000, Houdré, Reynaud-Bouret 2003, Adamczack 2006, Giné, Nickl 2016], on déduit

Théorème (L., Magalhães, Reynaud-Bouret HDP VII)

Si $pen : \mathcal{K} \rightarrow \mathbb{R}$ satisfait

$$\forall K \in \mathcal{K}, \quad \left| pen(K) - \frac{2PK^\Delta}{n} \right| \leq \alpha R(K) + \square \frac{(\log n)^\square}{n},$$

pour un $\alpha \in (0, 1)$, alors, sous de bonnes hypothèses sur les $K \in \mathcal{K}$ et si $|\mathcal{K}| \leq e^{(\log n)^\gamma}$, pour tout $\epsilon \in (0, (1/4) \wedge (1 - \alpha))$,

$$R(\widehat{f}_{\widehat{K}}) \leq \frac{1 + \alpha + \epsilon}{1 - \alpha - \epsilon} \inf_{K \in \mathcal{K}} R(\widehat{f}_K) + \square \frac{(\log n)^\square}{(1 - \alpha)\epsilon^3 n}.$$

Pénalité minimale

Idée de [Arlot, Massart 2009] : si $\text{pen}(K) \approx \mathbb{E} [p_2(K)]$, le critère pénalisé se concentre autour de $P_\gamma(f_K) = \|f_K - f\|^2 - \|f\|^2$ donc le minimiseur \hat{K} d'un tel critère est typiquement un noyau "complexe" c'est à dire un noyau avec une petite fenêtre dans le cas des estimateurs à noyaux ou un histogramme de grande dimension.

Pénalité minimale

Idée de [Arlot, Massart 2009] : si $\text{pen}(K) \approx \mathbb{E}[p_2(K)]$, le critère pénalisé se concentre autour de $P_\gamma(f_K) = \|f_K - f\|^2 - \|f\|^2$ donc le minimiseur \hat{K} d'un tel critère est typiquement un noyau "complexe" c'est à dire un noyau avec une petite fenêtre dans le cas des estimateurs à noyaux ou un histogramme de grande dimension.

En revanche, dès qu'il existe $u > 0$ tel que de $\text{pen}(K) \approx \mathbb{E}[p_2(K)] + u\mathbb{E}[p_1(K)]$, le critère pénalisé commence à se concentrer autour de la somme de biais avec un terme de variance et le minimiseur \hat{K} de ce critère va vérifier une inégalité oracle.

$\mathbb{E}[p_2(K)]$ est appelée pénalité minimale en ce sens [Birgé, Massart 2007].

Pénalité minimale pour la sélection d'estimateurs linéaires

Théorème (L., Magalhães, Reynaud-Bouret HDP VII)

Supposons qu'il existe $0 \leq a < |b|$ tels que,

$$\forall K \in \mathcal{K}, \quad \left| \text{pen}(K) - \frac{2PK^\Delta - (1+b)PA_K^\Delta}{n} \right| \leq aR(K) + \square \frac{(\log n)^\square}{n} .$$

Si $b > 0$, alors $PA_K^\Delta + \square(\log n)^\square \geq C_{a,b} \max_{K \in \mathcal{K}} PA_K^\Delta$ et

$$R(\hat{f}_{\hat{K}}) \geq C_{a,b} \max_{K \in \mathcal{K}} R(\hat{f}_K) - \square \frac{(\log n)^\square}{n} .$$

Si $b < 0$, alors $PA_K^\Delta - \square(\log n)^\square \ll \max_{K \in \mathcal{K}} PA_K^\Delta$ et

$$R(\hat{f}_{\hat{K}}) \leq C_{a,b} \left(\min_{K \in \mathcal{K}} R(\hat{f}_K) + \square \frac{(\log n)^\square}{n} \right) .$$

Validation croisée

La validation croisée est une procédure générale d'estimation de quantités définies comme des fonctionnelles F connues de la loi P inconnue des observations et de la mesure empirique P_n de l'échantillon.

La première idée est de couper l'échantillon X_1, \dots, X_n en un bloc d'entraînement $(X_i)_{i \in T}$ et son complémentaire $(X_i)_{i \in T^c}$ dit bloc de validation. On estime alors $F(P, P_n)$ par son estimateur hold-out $F(P_{T^c}, P_T)$, où $P_A = \frac{1}{|A|} \sum_{i \in A} \delta_i$ est la mesure empirique définie sur l'échantillon $(X_i)_{i \in A}$.

La seconde idée est de stabiliser cette procédure en se donnant plusieurs échantillons d'entraînement \mathcal{T} et de construire l'estimateur

$$\hat{F}_{\mathcal{T}} = \frac{1}{|\mathcal{T}|} \sum_{T \in \mathcal{T}} F(P_{T^c}, P_T) .$$

Exemples de fonctionnelles d'intérêt

La première fonctionnelle étudiée pour la sélection d'estimateurs a été la perte [Burman 1989]

$$\ell(\hat{f}_K) = \left\| \hat{f}_K \right\|^2 - 2P\hat{f}_K .$$

Une autre fonctionnelle d'intérêt dans notre problème est la pénalité idéale [Arlot 2008, Celisse 2012, 2014].

$$\text{pen}_{\text{id}}(K) = p_1(K) + p_2(K) .$$

On peut enfin imaginer étudier séparément des estimateurs de p_1 et p_2 .

Exemples de collections \mathcal{T}

- ① Si $\mathcal{T} = \{T\}$, $\widehat{F}_{\mathcal{T}}$ est l'estimateur hold-out.
- ② Si, pour un $p \in \mathbb{N} \cap [2, n-1]$,
 $\mathcal{T}_p^{\text{lpo}} = \{T \subset \{1, \dots, n\}, |T| = n-p\}$, $\widehat{F}_{\mathcal{T}_p^{\text{lpo}}}$ est l'estimateur leave- p -out.
- ③ Soit V un diviseur de n et soit $\mathcal{B}_V = (B_1, \dots, B_V)$ une partition de $\{1, \dots, n\}$ en sous-ensembles de même taille $p = n/V$. Si
 $\mathcal{T}_V^{\text{vf}} = \{B^c, B \in \mathcal{B}_V\}$, $\widehat{F}_{\mathcal{T}_V^{\text{vf}}}$ est appelé estimateur V -fold.
- ④ Soit $\mathcal{T}_{p,B}^{\text{MC}}$ un échantillon de B sous-ensembles de $\mathcal{T}_p^{\text{lpo}}$ choisi indépendamment et de manière uniforme. L'estimateur $\widehat{F}_{\mathcal{T}_{p,B}^{\text{MC}}}$ est appelé estimateur Monte-Carlo V -fold.

Panorama des résultats

- 1 Obtention de formules closes et réduction de l'étude de tous ces schémas de validation croisés et de tous les critères obtenus à l'étude d'un critère générique pénalisé par une CL de moyennes empiriques et de U -statistiques totalement dégénérées.
- 2 Concentration de ces différents termes et calcul des espérances.
- 3 Obtention d'inégalités oracle pour un minimiseur du critère générique.

Quelques conséquences et limites

- 1 Obtention simultanée d'inégalités oracle pour les estimateurs sélectionnés par toutes les méthodes précédentes (par estimation de la perte ou de la pénalité idéale) quelque soit le schéma de validation croisée parmi ceux exposés.

Quelques conséquences et limites

- 1 Obtention simultanée d'inégalités oracle pour les estimateurs sélectionnés par toutes les méthodes précédentes (par estimation de la perte ou de la pénalité idéale) quelque soit le schéma de validation croisée parmi ceux exposés.
- 2 Seul le biais relatif de l'estimateur de validation croisée

$$\delta = \max_{K \in \mathcal{K}} \frac{\left| \mathbb{E} \left[\widehat{F}_{\mathcal{T}}(K) - F(K) \right] \right|}{\mathbb{E} [F(K)]}$$

permet sur ces résultats de différencier les \mathcal{T} à travers la "leading constant" dans l'inégalité oracle $L = \frac{1+\delta_+}{1-\delta_-}$.

Quelques conséquences et limites

- 1 Obtention simultanée d'inégalités oracle pour les estimateurs sélectionnés par toutes les méthodes précédentes (par estimation de la perte ou de la pénalité idéale) quelque soit le schéma de validation croisée parmi ceux exposés.
- 2 Seul le biais relatif de l'estimateur de validation croisée

$$\delta = \max_{K \in \mathcal{K}} \frac{\left| \mathbb{E} \left[\widehat{F}_{\mathcal{T}}(K) - F(K) \right] \right|}{\mathbb{E} [F(K)]}$$

permet sur ces résultats de différencier les \mathcal{T} à travers la "leading constant" dans l'inégalité oracle $L = \frac{1+\delta_+}{1-\delta_-}$.

- 3 A "biais" fixe, ou asymptotiquement fixe, pas de différence entre les schémas au premier ordre. Seuls le terme de reste ρ dans l'inégalité oracle semble se détériorer.

Différencier des critères de même biais

Des différences pratiques de performances sont nettes entre les estimateurs hold-out (rapides à calculer mais instables) et leave- p -out (incalculables en général mais très stables). Peut-on expliquer ces phénomènes théoriquement?

Différencier des critères de même biais

Des différences pratiques de performances sont nettes entre les estimateurs hold-out (rapides à calculer mais instables) et leave- p -out (incalculables en général mais très stables). Peut-on expliquer ces phénomènes théoriquement?

Heuristique [Arlot et L. 2016] : un critère $\mathcal{C}_1 : \mathcal{K} \rightarrow \mathbb{R}$ est "meilleur" que $\mathcal{C}_2 : \mathcal{K} \rightarrow \mathbb{R}$ si la probabilité qu'un minimiseur de \mathcal{C}_1 appartienne à l'ensemble

$$\mathcal{K}_r = \left\{ K' \in \mathcal{K}, R(K') \leq \inf_{K \in \mathcal{K}} R(K) + r \right\}$$

est plus grande que pour un minimiseur de \mathcal{C}_2 .

Obtention d'un critère de comparaison

Soient \mathcal{C}_1 et \mathcal{C}_2 deux critères sans biais. Soit $K_o \in \arg \min_{K \in \mathcal{K}} R(K)$, $\hat{K}_i \in \arg \min_{K \in \mathcal{K}} C_i(K)$ et supposons pour simplifier que ces minima soient uniques.

$$\begin{aligned} \mathbb{P} \left(\hat{K}_i \in \mathcal{K}_r \right) &= \mathbb{P} \left(\min_{K \notin \mathcal{K}_r} C_i(K) > \min_{K \in \mathcal{K}_r} C_i(K) \right) \\ &\approx \mathbb{P} \left(\min_{K \notin \mathcal{K}_r} C_i(K) > C_i(K_o) \right) \\ &\approx \mathbb{P} \left(\min_{K \notin \mathcal{K}_r} \{ \mathbb{E} [C_i(K) - C_i(K_o)] + N_K \} > 0 \right) , \end{aligned}$$

où $(N_K)_{K \in \mathcal{K}}$ est un vecteur Gaussien centré de matrice de covariance $\Sigma^{(i)}$ telle que

$$\Sigma_{K, K'}^{(i)} = \text{Cov} [C_i(K) - C_i(K_o), C_i(K') - C_i(K_o)] .$$

Heuristique finale et quelques conséquences

\mathcal{C}_1 est meilleur que \mathcal{C}_2 si $\Sigma^{(1)} \leq \Sigma^{(2)}$.

Heuristique finale et quelques conséquences

\mathcal{C}_1 est meilleur que \mathcal{C}_2 si $\Sigma^{(1)} \leq \Sigma^{(2)}$.

Important : [Arlot et L. 2016] on peut effectivement comparer les matrices de covariance des accroissements de critères de validation croisée pour différents schémas, dans le cas des estimateurs par projection. En particulier,

Heuristique finale et quelques conséquences

\mathcal{C}_1 est meilleur que \mathcal{C}_2 si $\Sigma^{(1)} \leq \Sigma^{(2)}$.

Important : [Arlot et L. 2016] on peut effectivement comparer les matrices de covariance des accroissements de critères de validation croisée pour différents schémas, dans le cas des estimateurs par projection. En particulier,

- ① Lorsque $p = n/V$ est fixé, le critère hold-out est le pire, leave- p -out le meilleur.
- ② Augmenter V dans les critères V -fold améliore les performances. L'amélioration est importante lorsque V passe de 2 à 5 ou 10, moins ensuite.
- ③ La variance du critère V -fold est plus petite, à coût algorithmique constant, que celle du Monte-Carlo V -fold.

Estimateurs sous-Gaussiens

Soient Y, Y_1, \dots, Y_n des variables aléatoires réelles, indépendantes et de même loi P de variance majorée par σ^2 finie. On va construire un estimateur \hat{E} à partir de Y_1, \dots, Y_n de $\mathbb{E}[Y]$ vérifiant la borne sous-Gaussienne suivante :

$$\forall x \leq \frac{n}{2} - C, \quad \mathbb{P} \left(\left| \hat{E} - \mathbb{E}[Y] \right| > L\sigma \sqrt{\frac{1+x}{n}} \right) \leq e^{-x},$$

avec L, C constantes absolues.

Médiane des moyennes

Soit x un diviseur de n et B_1, \dots, B_x une partition de $\{1, \dots, n\}$ en sous-ensembles de taille n/x . On considère $Z_k = \frac{x}{n} \sum_{j \in B_k} Y_j$ et l'estimateur $\hat{E}_x = \text{med}((Z_k)_{k=1, \dots, x})$.

Médiane des moyennes

Soit x un diviseur de n et B_1, \dots, B_x une partition de $\{1, \dots, n\}$ en sous-ensembles de taille n/x . On considère $Z_k = \frac{x}{n} \sum_{j \in B_k} Y_j$ et l'estimateur $\hat{E}_x = \text{med}((Z_k)_{k=1, \dots, x})$.

On a par Tchebycheff,

$$q = \mathbb{P} \left(|Z_k - \mathbb{E}[Y]| > 2e\sigma \sqrt{\frac{x}{n}} \right) \leq \frac{1}{(2e)^2} = p.$$

Donc \hat{E}_x vérifie

$$\begin{aligned} & \mathbb{P} \left(\left| \hat{E}_x - \mathbb{E}[Y] \right| > 2e\sigma \sqrt{\frac{x}{n}} \right) \\ & \leq \mathbb{P} \left(\left| \left\{ k, |Z_k - \mathbb{E}[Y]| > 2e\sigma \sqrt{\frac{x}{n}} \right\} \right| \geq \frac{x}{2} \right) \\ & = \mathbb{P} \left(\text{Bin}(x, q) \geq \frac{x}{2} \right) \leq \mathbb{P} \left(\text{Bin}(x, p) \geq \frac{x}{2} \right) \end{aligned}$$

Médiane des moyennes

$$\begin{aligned} \mathbb{P} \left(\left| \hat{E}_x - \mathbb{E}[Y] \right| > 2e\sigma \sqrt{\frac{x}{n}} \right) &\leq \mathbb{P} \left(\text{Bin}(x,p) \geq \frac{x}{2} \right) \\ &= \sum_{k=x/2}^x \binom{x}{k} p^k (1-p)^{x-k} \leq p^{x/2} 2^x = e^{-x} . \end{aligned}$$

Méthode de Lepski

Pour tout $k = 1, \dots, n/2$, on construit avec le principe de la médiane des moyennes des intervalles de confiance I_k au niveau e^{-k} de taille

$$L\sigma\sqrt{\frac{k}{n}}.$$

Méthode de Lepski

Pour tout $k = 1, \dots, n/2$, on construit avec le principe de la médiane des moyennes des intervalles de confiance I_k au niveau e^{-k} de taille $L\sigma\sqrt{\frac{k}{n}}$. On considère alors

$$\hat{k} = \min \{ k \in \{1, \dots, n\}, \cap_{j=k}^n I_j \neq \emptyset \} ,$$

Méthode de Lepski

Pour tout $k = 1, \dots, n/2$, on construit avec le principe de la médiane des moyennes des intervalles de confiance I_k au niveau e^{-k} de taille $L\sigma\sqrt{\frac{k}{n}}$. On considère alors

$$\hat{k} = \min \{ k \in \{1, \dots, n\}, \cap_{j=k}^n I_j \neq \emptyset \} ,$$

puis \hat{E} comme le milieu du segment $\cap_{j=\hat{k}}^n I_j$.

Méthode de Lepski

Pour tout $k = 1, \dots, n/2$, on construit avec le principe de la médiane des moyennes des intervalles de confiance I_k au niveau e^{-k} de taille $L\sigma\sqrt{\frac{k}{n}}$. On considère alors

$$\hat{k} = \min \{ k \in \{1, \dots, n\}, \cap_{j=k}^n I_j \neq \emptyset \} ,$$

puis \hat{E} comme le milieu du segment $\cap_{j=\hat{k}}^n I_j$. Soit alors $x \leq n/2$ et $k = [x] + 2$, l'événement $\mathbb{E}[Y] \in \cap_{j=k}^n I_j$ a une probabilité supérieure à $1 - e^{1-k} \geq 1 - e^{-x}$

Méthode de Lepski

Pour tout $k = 1, \dots, n/2$, on construit avec le principe de la médiane des moyennes des intervalles de confiance I_k au niveau e^{-k} de taille $L\sigma\sqrt{\frac{k}{n}}$. On considère alors

$$\hat{k} = \min \{ k \in \{1, \dots, n\}, \cap_{j=k}^n I_j \neq \emptyset \} ,$$

puis \hat{E} comme le milieu du segment $\cap_{j=\hat{k}}^n I_j$. Soit alors $x \leq n/2$ et $k = [x] + 2$, l'événement $\mathbb{E}[Y] \in \cap_{j=k}^n I_j$ a une probabilité supérieure à $1 - e^{1-k} \geq 1 - e^{-x}$ et, sur cet événement, $\hat{k} \leq k$, donc $\hat{E} \in I_k$, donc

$$|\hat{E} - \mathbb{E}[Y]| \leq L\sigma\sqrt{\frac{k}{n}} \leq L\sigma\sqrt{\frac{2+x}{n}} .$$

Présentation du problème

Soit \mathcal{F} un ensemble de densités sur \mathbb{X} uniformément bornées par σ^2 . Pour tout $f \in \mathcal{F}$, on note la loi d'un échantillon X_1, \dots, X_n de variables aléatoires indépendantes et de même densité f par $\mathbb{P}_f(\cdot)$ et l'espérance sous cette loi $\mathbb{E}_f[\cdot]$.

Présentation du problème

Soit \mathcal{F} un ensemble de densités sur \mathbb{X} uniformément bornées par σ^2 . Pour tout $f \in \mathcal{F}$, on note la loi d'un échantillon X_1, \dots, X_n de variables aléatoires indépendantes et de même densité f par $\mathbb{P}_f(\cdot)$ et l'espérance sous cette loi $\mathbb{E}_f[\cdot]$. Soit $(\varphi_i)_{i=\{1, \dots, p\}}$ un système orthonormé dans L^2 et, pour tout $i \in \{1, \dots, p\}$, soit H_i l'hypothèse

$$H_i : \mathbb{E}[\varphi_i(X)] = 0 .$$

Présentation du problème

Soit \mathcal{F} un ensemble de densités sur \mathbb{X} uniformément bornées par σ^2 . Pour tout $f \in \mathcal{F}$, on note la loi d'un échantillon X_1, \dots, X_n de variables aléatoires indépendantes et de même densité f par $\mathbb{P}_f(\cdot)$ et l'espérance sous cette loi $\mathbb{E}_f[\cdot]$. Soit $(\varphi_i)_{i=\{1, \dots, p\}}$ un système orthonormé dans L^2 et, pour tout $i \in \{1, \dots, p\}$, soit H_i l'hypothèse

$$H_i : \mathbb{E}[\varphi_i(X)] = 0 .$$

On souhaite tester simultanément toutes les hypothèses H_i , c'est à dire construire un ensemble $\mathcal{R} \subset \{1, \dots, p\}$ d'hypothèses rejetées.

Présentation du problème

Soit \mathcal{F} un ensemble de densités sur \mathbb{X} uniformément bornées par σ^2 . Pour tout $f \in \mathcal{F}$, on note la loi d'un échantillon X_1, \dots, X_n de variables aléatoires indépendantes et de même densité f par $\mathbb{P}_f(\cdot)$ et l'espérance sous cette loi $\mathbb{E}_f[\cdot]$. Soit $(\varphi_i)_{i=\{1, \dots, p\}}$ un système orthonormé dans L^2 et, pour tout $i \in \{1, \dots, p\}$, soit H_i l'hypothèse

$$H_i : \mathbb{E}[\varphi_i(X)] = 0 .$$

On souhaite tester simultanément toutes les hypothèses H_i , c'est à dire construire un ensemble $\mathcal{R} \subset \{1, \dots, p\}$ d'hypothèses rejetées. On veut garantir que ce test contrôle le taux d'erreur par famille par α , c'est à dire que, en notant $T(f) = \{i, \mathbb{E}_f[\varphi_i(X)] = 0\}$,

$$\text{FWER}(\mathcal{R}) = \sup_{f \in \mathcal{F}} \mathbb{P}_f(\mathcal{R} \cap T(f)) \leq \alpha .$$

Tests élémentaires

Construisons d'abord, pour tout i et α , un test simple $\phi_{i,\alpha} : \mathbb{R}^n \rightarrow \{0, 1\}$ de niveau α , i.e. satisfaisant

$$\mathbb{E}R(\phi_{i,\alpha}) = \sup_{f, H_i \in T(f)} \mathbb{P}_f(\phi_{i,\alpha}(X_1^n) = 1) \leq \alpha . \quad (5)$$

On peut utiliser les estimateurs précédents. Comme $\mathbb{E}_f[\varphi_i^2(\mathbf{X})] \leq \sigma^2$, il existe \hat{E}_i , estimateur de $\mathbb{E}_f[\varphi_i(\mathbf{X})]$ tel que $\forall f \in \mathcal{F}, \forall \alpha \geq e^{C-n/2}$,

$$\mathbb{P}_f \left(\left| \hat{E}_i - \mathbb{E}_f[\varphi_i(\mathbf{X})] \right| > L\sigma \sqrt{\frac{1 + \log(1/\alpha)}{n}} \right) \leq \alpha .$$

Tests élémentaires

Construisons d'abord, pour tout i et α , un test simple $\phi_{i,\alpha} : \mathbb{R}^n \rightarrow \{0, 1\}$ de niveau α , i.e. satisfaisant

$$\mathbb{E} \mathbb{R}(\phi_{i,\alpha}) = \sup_{f, H_i \in T(f)} \mathbb{P}_f(\phi_{i,\alpha}(X_1^n) = 1) \leq \alpha . \quad (5)$$

On peut utiliser les estimateurs précédents. Comme $\mathbb{E}_f[\varphi_i^2(\mathbf{X})] \leq \sigma^2$, il existe \hat{E}_i , estimateur de $\mathbb{E}_f[\varphi_i(\mathbf{X})]$ tel que $\forall f \in \mathcal{F}, \forall \alpha \geq e^{C-n/2}$,

$$\mathbb{P}_f \left(\left| \hat{E}_i - \mathbb{E}_f[\varphi_i(\mathbf{X})] \right| > L\sigma \sqrt{\frac{1 + \log(1/\alpha)}{n}} \right) \leq \alpha .$$

Le test $\phi_{i,\alpha} = \mathbf{1}_{|\hat{E}_i| > L\sigma \sqrt{\frac{1 + \log(1/\alpha)}{n}}}$ satisfait donc (5).

Principe du rejet séquentiel

Revenons maintenant au test multiple et définissons la fonction successeur de Holm:

$$\mathcal{N}_\alpha : \mathcal{P}(\{1, \dots, p\}) \rightarrow \mathcal{P}(\{1, \dots, p\})$$

$$\mathcal{H} \mapsto \left\{ i \notin \mathcal{H}, \phi_{i, \frac{\alpha}{p-|\mathcal{H}|}} = 1 \right\} \quad .$$

Principe du rejet séquentiel

Revenons maintenant au test multiple et définissons la fonction successeur de Holm:

$$\mathcal{N}_\alpha : \mathcal{P}(\{1, \dots, p\}) \rightarrow \mathcal{P}(\{1, \dots, p\})$$

$$\mathcal{H} \mapsto \left\{ i \notin \mathcal{H}, \phi_{i, \frac{\alpha}{p-|\mathcal{H}|}} = 1 \right\} .$$

La procédure de Holm rejette alors $\mathcal{R} = \mathcal{R}_p$ ou on a défini itérativement

$$\mathcal{R}_0 = \emptyset, \quad \mathcal{R}_{i+1} = \mathcal{R}_i \cup \mathcal{N}_\alpha(\mathcal{R}_i) .$$

Principe du rejet séquentiel

Revenons maintenant au test multiple et définissons la fonction successeur de Holm:

$$\mathcal{N}_\alpha : \mathcal{P}(\{1, \dots, p\}) \rightarrow \mathcal{P}(\{1, \dots, p\})$$

$$\mathcal{H} \mapsto \left\{ i \notin \mathcal{H}, \phi_{i, \frac{\alpha}{p-|\mathcal{H}|}} = 1 \right\} .$$

La procédure de Holm rejette alors $\mathcal{R} = \mathcal{R}_p$ ou on a défini itérativement

$$\mathcal{R}_0 = \emptyset, \quad \mathcal{R}_{i+1} = \mathcal{R}_i \cup \mathcal{N}_\alpha(\mathcal{R}_i) .$$

Proposition (Goeman, Solari 2010)

$$FWER(\mathcal{R}) \leq \alpha .$$

Erreur de 2ème espèce

On définit pour tout $r > 0$ et $f \in \mathcal{F}$ l'ensemble des hypothèses à distance r de f par

$$F_r(f) = \{i, |\mathbb{E}_f[\varphi_i(X)]| > r\} .$$

Erreur de 2ème espèce

On définit pour tout $r > 0$ et $f \in \mathcal{F}$ l'ensemble des hypothèses à distance r de f par

$$F_r(f) = \{i, |\mathbb{E}_f[\varphi_i(X)]| > r\} .$$

La vitesse de séparation d'un test multiple \mathcal{R} (taux d'erreur par famille) est ensuite définie pour tout $\beta \in (0, 1)$, par [Fromont, L., Reynaud-Bouret 2016]

$$\text{FWSR}^\beta(\mathcal{R}, \mathcal{F}) = \inf \left\{ r > 0, \inf_{f \in \mathcal{F}} \mathbb{P}_f(F_r(f) \subset \mathcal{R}) \geq 1 - \beta \right\} .$$

Erreur de 2ème espèce

On définit pour tout $r > 0$ et $f \in \mathcal{F}$ l'ensemble des hypothèses à distance r de f par

$$F_r(f) = \{i, |\mathbb{E}_f[\varphi_i(X)]| > r\} .$$

La vitesse de séparation d'un test multiple \mathcal{R} (taux d'erreur par famille) est ensuite définie pour tout $\beta \in (0, 1)$, par [Fromont, L., Reynaud-Bouret 2016]

$$\text{FWSR}^\beta(\mathcal{R}, \mathcal{F}) = \inf \left\{ r > 0, \inf_{f \in \mathcal{F}} \mathbb{P}_f(F_r(f) \subset \mathcal{R}) \geq 1 - \beta \right\} .$$

Le calcul de cette distance pour notre test peut se faire en utilisant la monotonie de $\text{FWSR}^\beta(\cdot, \mathcal{F})$ pour dire que

$\text{FWSR}^\beta(\mathcal{R}, \mathcal{F}) \geq \text{FWSR}^\beta(\mathcal{R}_1, \mathcal{F})$, puis en utilisant à nouveau la propriété de concentration sous-Gaussienne. On obtient

$$\text{FWSR}^\beta(\mathcal{R}, \mathcal{F}) \leq L\sigma \left(\sqrt{\frac{\log \frac{p}{\alpha}}{n}} + \sqrt{\frac{\log \frac{p}{\beta}}{n}} \right)$$

Optimalité minimax

La vitesse minimax de séparation est ensuite définie en suivant le principe de [Ingster 1993, Baraud 2002] par

$$\text{mFWSR}^{\alpha,\beta}(\mathcal{F}) = \inf_{\mathcal{R}, \text{FWER}(\mathcal{R}) \leq \alpha} \text{FWSR}^{\beta}(\mathcal{R}, \mathcal{F}) .$$

Optimalité minimax

La vitesse minimax de séparation est ensuite définie en suivant le principe de [Ingster 1993, Baraud 2002] par

$$mFWSR^{\alpha,\beta}(\mathcal{F}) = \inf_{\mathcal{R}, \text{FWER}(\mathcal{R}) \leq \alpha} FWSR^{\beta}(\mathcal{R}, \mathcal{F}) .$$

Pour minorer cette vitesse, on a le résultat général suivant

Théorème (Fromont, L. Reynaud-Bouret 2016)

$$mFWSR^{\alpha,\beta}(\mathcal{F}) \geq mSR^{\alpha,\beta}(\mathcal{F}, \cap_{i=1}^p H_i) .$$

Dans notre exemple

Soit $\mathbb{X} = [0, 1]$ et soit $(\varphi_i)_{i \in \mathbb{N}}$ la base de Fourier sur $[0, 1]$ ($\varphi_0 = 1$).

Dans notre exemple

Soit $\mathbb{X} = [0, 1]$ et soit $(\varphi_i)_{i \in \mathbb{N}}$ la base de Fourier sur $[0, 1]$ ($\varphi_0 = 1$).
Notons $H_i : \mathbb{E}[\varphi_i(X)] = 0$ pour tous les $i \in \{1, \dots, p\}$. Supposons
que \mathcal{F} contient la classe \mathcal{A} des fonctions $f = 1 + r\varphi_i$ avec $|r| < \frac{1}{\sqrt{2}}$ et
 $i \in \{1, \dots, p\}$. Notons enfin $H_u : f = 1$.

Dans notre exemple

Soit $\mathbb{X} = [0, 1]$ et soit $(\varphi_i)_{i \in \mathbb{N}}$ la base de Fourier sur $[0, 1]$ ($\varphi_0 = 1$).
Notons $H_i : \mathbb{E}[\varphi_i(X)] = 0$ pour tous les $i \in \{1, \dots, p\}$. Supposons
que \mathcal{F} contient la classe \mathcal{A} des fonctions $f = 1 + r\varphi_i$ avec $|r| < \frac{1}{\sqrt{2}}$ et
 $i \in \{1, \dots, p\}$. Notons enfin $H_u : f = 1$. Alors, on a

$$\text{mSR}^{\alpha, \beta}(\mathcal{F}, \cap_{i=1}^p H_i) \geq \text{mSR}^{\alpha, \beta}(\mathcal{A}, H_u) \geq \sqrt{\frac{\log(C_{\alpha, \beta} p)}{n}}.$$