Sélection de modèles optimale par pénalité de rééchantillonnage pour des M-estimateurs à contraste régulier.

Adrien Saumard

Sélection de modèles optimale par pénalités de rééchantillonnage pour des M-estimateurs à contraste régulier.

Adrien Saumard

En thèse sous la direction de Philippe Berthet, Université Rennes 1, IRMAR.

Résumé

On se propose ici d'étudier l'efficacité, et en particulier l'optimalité de procédures de sélection de modèles par pénalités de rééchantillonnage pour une large classe de M-estimateurs. Nous donnerons la définition d'un contraste dit régulier, et dans ce cadre nous étudierons sous des hypothèses générales l'excès de risque du M-estimateur à modèle fixé, en donnant des bornes supérieures et inférieures en probabilité. Il nous faudra entre autre améliorer une inégalité due à Hoffman-Jorgensen pour des variables aléatoires à valeurs dans un Banach. Nous montrerons enfin comment utiliser ces informations pour dégager un cadre d'optimalité dans les procédures de sélection de modèles associés à ses estimateurs.

Abstract

Our goal is here to study the efficiency and in particular the optimality of resampling model selection procedures for a large class of M-estimators. We will thus define a regular contrast, and for M-estimators associated to those contrasts we will study upper and lower bounds in probability for their excess risk on a fixed model. To do this we will need an improvement of an inequality due to Hoffman-Jorgensen for random variables in Banach spaces. We then intend to show how to derive sharp oracle inequalities from such informations, and so to study general conditions of optimality.

Mots-clefs : sélection de modèles, rééchantillonnage, M-estimateurs, probabilité dans les espaces de Banach.

1 Introduction.

Il existe de nombreux travaux sur la qualité des procédures de sélection de modèles par pénalisation, comparant la qualité de l'estimateur sélectionné à celle dite de l'oracle, c'est-à-dire du meilleur estimateur parmi ceux construits sur chaque modèle. Une question subtile liée à l'optimalité des procédures a été étudiée dans les travaux pionniers de Lucien Birgé et Pascal Massart (2001 et 2006) dans un contexte gaussien à variance

connue : celle de l'existence de pénalité minimale et du phénomène de pente. En effet, ces auteurs ont montré que si la pénalité passe en-dessous d'une valeur critique appelée pénalité minimale, alors la procédure choisit systématiquement des modèles de dimension beaucoup trop grande, et de plus que la pénalité optimale, celle qui sélectionne un estimateur qui a asymptotiquement la même performance que l'oracle, est deux fois la pénalité minimale. Cette propriété dite de pente peut donc servir à calibrer numériquement la constante devant une pénalité "générique" en observant par exemple la dimension des modèles sélectionnés. Qu'en est-il de la généralité de ce phénomène ? Yannick Baraud, Christophe Giraud et Sylvie Huet (2007) ont montré l'existence de pénalités minimales dans le cas gaussien à variance inconnue. Puis Sylvain Arlot (2007) a montré le phénomène de pente dans un contexte beaucoup plus général de régression hétéroscédastique. Néanmoins, l'étude a été menée dans le cadre de la régression par histogrammes, et donc bien que le passage à la généralité ait été opéré sur le bruit, la structure des modèles sous-jacents est fortement rentrée en compte dans la démonstration des résultats. De plus, Sylvain Arlot (2007) a proposé des pénalités dites de rééchantillonnage qui sont sous certaines conditions asymptotiquement optimales. Le but du travail que nous présentons était donc de passer à une plus grande généralité et de se libérer des calculs spécifiques aux modèles par histogrammes. Nous allons donc définir le cadre statistique de notre étude.

2 M-estimation à contraste régulier.

Etant donnée une loi $P = P^Z$ sur un espace mesurable $(\mathcal{Z}, \mathcal{T})$ nous voulons estimer une fonctionnelle s_* de la loi de la forme

$$s_* = \arg\min_{s \in \mathcal{S}} P(Ks),$$

où $Pf = \mathbb{E}[f(Z)]$, $\mathcal{S}$ est un espace fonctionnel et K, appelé *contraste*, prend ses arguments dans $\mathcal{S}$ et vérifie $\forall s \in \mathcal{S}$, $|P(Ks)| < \infty$.

Par exemple, pour l'estimation d'une fonction de régression avec $Z = (X, Y)$, on a

$$s_* = \mathbb{E}[Y | X = .] = \arg\min_{s \in L_2(P^X)} P(Ks),$$

où $(Ks)(x, y) = (y - s(x))^2$.

De plus un contraste sera dit *régulier* s'il vérifie pour P-presque tout $z \in \mathcal{Z}$, et tout $s \in \mathcal{S}$,

$$(Ks)(z) - (Ks_*)(z) = \psi_0^s + \psi_1(z)(s - s_*)(z) + \psi_2^s(z)(s - s_*)(z)$$

où la fonction ψ_2^s et la constante ψ_0^s dépendent de s et $\psi_2^{s_*}(z) = 0$.

En reprenant le cas de la régression, avec $(Ks)(x, y) = (y - s(x))^2$ et $s(z) = s(x, y) = s(x)$, on a

$$\begin{aligned}(Ks)(z) - (Ks_*)(z) &= (s(x) - s_*(x))(s(x) - s_*(x) - 2(y - s_*(x))) \\ &= -2(y - s_*(x))(s(x) - s_*(x)) + (s(x) - s_*(x))^2.\end{aligned}$$

Donc K est un contraste régulier avec $\psi_0^s = 0$,

$$\psi_1(z) = -2(y - s_*(x))$$

et

$$\psi_2^s(z) = s(x) - s_*(x).$$

D'autres contrastes que celui des moindres carrés en régression vérifient cette propriété de régularité, c'est par exemple le cas en densité avec le contraste de Kullback-Leibler où celui des moindres carrés par rapport à une mesure de référence connue.

Ce développement du contraste va nous permettre un contrôle fin de la performance du M-estimateur sur un modèle fixé, étape d'étude préalable à celui de la sélection de modèles.

3 Excès de risque à modèle fixé.

Soit M un modèle, c'est-à-dire un sous-espace de dimension finie de $\mathcal{S}$ et soit $(Z_1, ..., Z_n)$ un échantillon i.i.d. de loi P. Nous définissons un M-estimateur s_n sur M associé au contraste K par

$$s_n = s_n(M) \in \arg\min_{s \in M} P_n(Ks),$$

où P_n est la mesure empirique associée à l'échantillon $(Z_1, ..., Z_n)$.

Nous cherchons à localiser l'excès de risque du M-estimateur, c'est-à-dire la quantité aléatoire

$$P(Ks_n) - P(Ks_*) (\geq 0).$$

Le contrôle se fait par bornes supérieures et inférieures en probabilité. Du côté des bornes supérieures, beaucoup de travaux ont été faits, comme par exemple celui de Pascal Massart et Elodie Nédélec (2006) où ils étudient l'influence de conditions de marge sur les M-estimateurs de manière très générale, ou encore celui d'Evarist Giné et Vladimir Koltchinskii (2006) qui généralise en certains aspects l'étude de P.Massart et E.Nédélec en proposant une technique différente pour aborder les processus empiriques renormalisés.

Néanmoins, il existe très peu d'études des bornes inférieures en probabilité, c'est-à-dire des bornes du type

$$\mathbb{P}[P(Ks_n) - P(Ks_*) \leq B_I] \leq \varepsilon.$$

Notre étude repose entre autre à ce niveau sur une amélioration d'une inégalité due à Hoffman-Jorgensen en théorie des probabilités dans les espaces de Banach. Cette inégalité stipule que si $(Y_1, ..., Y_n)$ sont des variables i.i.d. à valeurs dans un espace de Banach $(B, \|.\|)$ alors on a

$$\mathbb{E}^{\frac{1}{p}}\left[\left\|\sum_{i=1}^n Y_j\right\|^p\right] \leq K\frac{p}{\log p}\left(\mathbb{E}\left[\left\|\sum_{i=1}^n Y_j\right\|\right] + \mathbb{E}^{\frac{1}{p}}\left[\left(\max_{1\leq i\leq n}\|Y_j\|\right)^p\right]\right)$$

où $K \geq 1$ est une constante numérique. Cette inégalité majore donc le moment d'ordre deux de la somme des variables par le moment d'ordre 1, plus un terme résiduel, le tout à une constante multiplicative près. Quitte à grossir le terme résiduel nous aimerions changer la constante multiplicative par 1. Ceci est en fait possible avec des hypothèses assez souples sur les variables, le terme de reste nécessitant plus de notations.

4 Sélection de modèles.

Une fois le contrôle à modèle fixé établi, nous reprenons essentiellement l'algèbre de démonstration exposé par Sylvain Arlot (2007), afin de démontrer des inégalités oracles trajectorielles pour l'estimateur sélectionné, avec contrôle de la constante dans l'inégalité.

Bibliographie

[1] Arlot S. (2007) Rééchantillonnage et Sélection de modèles, mémoire de thèse.
[2] Baraud Y., Giraud C., Huet S. (2007) Gaussian model selection with unknown variance, To appear *Ann.Stat.*
[3] Birgé L. et Massart P. (2001) Gaussian model selection, *J. Eur. Math. Soc. (JEMS),* 3(3):203-268.
[4] Birgé L. et Massart P. (2006) Minimal penalties for model selection, *Probab. Theory Related Fields,* 134(3).
[5] Giné E. et Koltchinskii V. (2006) Concentration inequalities and asymptotic results for ratio type empirical processes, *Ann.Probab.,* 33:1143-1216.
[6] Massart P. et Nédélec E. (2006) Risks bounds for statistical learning, *Ann.Stat.,* 34(5),2326–2366.