Concentration d'excès de risque en régression linéaire pénalisée

Adrien Saumard [1]

[1] *CREST, ENSAI, Campus de Ker-Lann, Rue Blaise Pascal, BP 37203, 35172 Bruz, adrien.saumard@ensai.fr*

Résumé. Nous montrons, d'après P. Bellec et A. Tsybakov (2016), *Bounds on the prediction error of penalized least squares estimators with convex penalty*, comment des arguments d'optimisation convexe permettent de prouver la concentration de l'excès de risque de l'estimateur d'une fonction de régression, minimisant un critère convexe de moindres carrés pénalisés. Ce cadre théorique décrit en particulier le comportement des estimateurs LASSO, group LASSO et SLOPE. Ce résultat de concentration est essentiel pour obtenir de nouvelles inégalités oracles, intégrables car valident sur un évènement dont la probabilité ne dépend pas du niveau de pénalisation choisi.

Mots-clés. Excès de risque, inégalité de concentration, critère pénalisé, optimisation convexe.

Abstract. After P. Bellec and A. Tsybakov (2016), *Bounds on the prediction error of penalized least squares estimators with convex penalty*, we show how arguments from convex optimization allow to prove a concentration inequality for the estimator of regression function, defined as the minimizer of a convex penalized least-squares criterion. The framework thus includes estimators such as the LASSO, group LASSO and SLOPE. The concentration inequality is an essential ingredient for proving new oracles inequalities, that are integrable since they are valid on an event with probability independent of the penalization level.

Keywords. Excess risk, concentration inequality, penalized criterion, convex optimization.

L'excès de risque d'un M-estimateur (éventuellement pénalisé) est une quantité fondamentale en apprentissage statistique. Par conséquent, une théorie générale des vitesses de convergence a été développée dans les années 90 et au début des années 2000 (13, 11). Cependant, il a été récemment identifié que certaines descriptions théoriques des procédures d'apprentissage statistique demandent un contrôle plus fin que celui apporté par les bornes supérieures classiques pour l'excès de risque. En l'occurrence, la recherche d'inégalités de concentration pour l'excès de risque d'un M-estimateur constitue un nouvel axe général de recherche particulièrement porteur dans l'établissement d'inégalités oracles.

En effet, il a tout d'abord été remarqué que de telles inégalités de concentration permettent de discuter convenablement de l'optimalité des procédures de sélection de modèles (7, 3, 12, 16). Plus précisement, les inégalités de concentration pour l'excès de risque et l'excès de risque empirique (8) permettent d'accéder aux constantes optimales dans les inégalités oracles décrivant la qualité des procédures de sélection de modèles. De tels résultats ont permis de mettre en évidence d'optimalité de la méthode de l'heuristique de pente (7, 4) dans des cadres théoriques liés à la minimisation d'un risque quadratique (moindres carrés, mais aussi maximum de vraisemblance avec une régularité convenable des modèles). Sous des hypothèses analogues, il est aussi possible de discuter de l'optimalité des procédures de rééchantillonnage et de validation croisée (1, 2, 9, 15).
Dans le contexte de la grande dimension, les méthodes convexes permettent un calcul efficace des estimateurs. C'est pourquoi S. Chatterjee (10) s'est récemment intéressé à la question de l'estimation de la moyenne d'un vecteur Gaussien de grande dimension sous contraintes convexes. Moyennant l'obtention d'une inégalité de concentration pour l'excès de risque de l'estimateur des moindres carrés projeté, S. Chatterjee (10) montre l'admissibilité universelle de cet estimateur. Le résultat de concentration a ensuite été étendu aux estimateurs minimisant un critère empirique convexe pénalisé (14, 17), via l'utilisation de techniques d'analyse convexe.
Enfin, un point faible des résultats théoriques sur les estimateurs pénalisés est que les inégalités oracles classiques décrivent des estimateurs dont le niveau de pénalisation est choisi en fonction du niveau de confiance considéré pour l'inégalité oracle. Cela ne correspond pas à la pratique de ces estimateurs, où le paramètre de régularisation est typiquement fixé par validation croisée. Récemment, P. Bellec et A. Tsybakov (5) (voir aussi (6)) ont établi des inégalités oracles plus satisfaisantes, qui décrivent les estimateurs pénalisés avec un niveau de confiance indépendant du niveau de pénalisation, ce qui permet en particulier d'intégrer ces bornes. L'argument central pour obtenir de telles inégalités oracles est d'exhiber en premier lieu une inégalité de concentration pour l'excès de risque de l'estimateur considéré. L'approche développée par les auteurs se base en particulier fortement sur des arguments d'optimisation convexe que nous détaillerons.

Bibliographie

[1] S. Arlot. V-fold cross-validation improved: V-fold penalization, Feb. 2008. URL `http://hal.archives-ouvertes.fr/hal-00239182/en/`. arXiv:0802.0566v2.

[2] S. Arlot and M. Lerasle. Choice of V for V-fold cross-validation in least-squares density estimation. *J. Mach. Learn. Res.*, 2015. ISSN 1532-4435. To appear. arXiv:1210.5830.

[3] S. Arlot and P. Massart. Data-driven calibration of penalties for least-squares regression. *J. Mach. Learn. Res.*, 10:245–279 (electronic), 2009.

[4] J.-P. Baudry, C. Maugis, and B. Michel. Slope heuristics: overview and implementation. *Stat. Comput.*, 22(2):455–470, 2012.

[5] P. C. Bellec and A. B. Tsybakov. Bounds on the prediction error of penalized least squares estimators with convex penalty. *arXiv preprint arXiv:1609.06675*, 2016.

[6] P. C. Bellec, G. Lecué, and A. B. Tsybakov. Towards the study of least squares estimators with convex penalty. *arXiv preprint arXiv:1701.09120*, 2017.

[7] L. Birgé and P. Massart. Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields*, 138(1-2):33–73, 2007. ISSN 0178-8051.

[8] S. Boucheron and P. Massart. A high-dimensional Wilks phenomenon. *Probab. Theory Related Fields*, 150(3-4):405–433, 2011.

[9] A. Celisse. Optimal cross-validation in density estimation with the l^2-loss. *Ann. Statist.*, 42(5):1879–1910, 10 2014. doi: 10.1214/14-AOS1240. URL `http://dx.doi.org/10.1214/14-AOS1240`.

[10] S. Chatterjee. A new perspective on least squares under convex constraint. *Ann. Statist.*, 42(6):2340–2381, 12 2014. doi: 10.1214/14-AOS1254. URL `http://dx.doi.org/10.1214/14-AOS1254`.

[11] V. Koltchinskii. *Oracle inequalities in empirical risk minimization and sparse recovery problems*, volume 2033 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2011. Lectures from the 38th Probability Summer School held in Saint-Flour, 2008, École d'Été de Probabilités de Saint-Flour. [Saint-Flour Probability Summer School].

[12] M. Lerasle. Optimal model selection for density estimation of stationary data under various mixing conditions. *Ann. Statist.*, 39(4):1852–1877, 2011.

[13] P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.

[14] A. Muro and S. van de Geer. Concentration behavior of the penalized least squares estimator. *arXiv preprint arXiv:1511.08698*, 2015.

[15] F. Navarro and A. Saumard. Slope heuristics and v-fold model selection in heteroscedastic regression using strongly localized bases. *arXiv preprint arXiv:1505.05654*, 2015.

[16] A. Saumard. Optimal upper and lower bounds for the true and empirical excess risks in heteroscedastic least-squares regression. *Electron. J. Statist.*, 6(1-2):579–655, 2012. ISSN 1935-7524.

[17] S. van de Geer and M. Wainwright. On concentration for (regularized) empirical risk minimization. *arXiv preprint arXiv:1512.00677*, 2016.