

---

# Contrôle du FDR en sélection de modèles pour la régression linéaire gaussienne en grande dimension.

Perrine Lacroix<sup>\*1,2</sup>

<sup>1</sup>Institute of Plant Sciences - Paris Saclay (IPS2), INRA-CNRS-Université Paris Sud, Labex Saclay Plant Science, Orsay, UMR 518 INRA AgroParisTech Mathématiques et Informatique appliquées (MIA), Paris. France. – Institut national de la recherche agronomique (INRA) : UMR518 – France

<sup>2</sup>Laboratoire de Mathématiques d'Orsay – Université Paris-Saclay, Centre National de la Recherche Scientifique : UMR8628 – France

## Résumé

*Dans un contexte de grande dimension, une approche classique pour estimer le paramètre inconnu en régression linéaire gaussienne est de minimiser les moindres carrés pénalisés. Pour obtenir une inégalité oracle sur le risque prédictif, la théorie développée par (Birgé et Massart, 2001) fournit une fonction de pénalité connue à une constante multiplicative près. Cette constante est actuellement fixée à 2 via des considérations d'optimalité asymptotique sur le risque. Dans cet exposé, je présenterai brièvement le contexte biologique qui a motivé le sujet de thèse et je montrerai que dans ce cadre, minimiser uniquement le risque prédictif n'est plus suffisant. En effet, le nombre de variables sélectionnées à tort doit être réduit le plus possible. Notre approche consiste à modifier la constante multiplicative et d'étudier l'impact de cette variation sur le taux de fausses découvertes (false discovery rate (FDR)) en plus du risque prédictif.*

---

\*Intervenant