

# Estimation en régression de Poisson

Stéphane Ivanoff

12 octobre 2012

Sous la direction de Vincent Rivoirard

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Présentation du modèle</b>	<b>2</b>
2.1	Régression de Poisson . . . . .	2
2.2	Le Lasso . . . . .	3
2.3	Estimateur de Dantzig . . . . .	3
<b>3</b>	<b>Choix des <math>\lambda_j</math></b>	<b>4</b>
3.1	Processus de Poisson . . . . .	4
3.2	Un théorème pour la calibration . . . . .	5
3.3	Inégalité oracle . . . . .	6
<b>4</b>	<b>Perspectives</b>	<b>7</b>
4.1	De meilleures inégalités oracles . . . . .	7
4.2	Group-Lasso et Group-Dantzig . . . . .	8

# 1 Introduction

En statistiques paramétriques, on cherche, de manière générale, à estimer un nombre  $p$  de **paramètres** à partir d'un nombre  $n \geq p$  d'**observations**. Typiquement, en régression linéaire gaussienne, la variable expliquée  $Y$  dépend linéairement de paramètres  $X_1, \dots, X_p$  selon des coefficients inconnus  $\beta_1, \dots, \beta_p$ , avec de plus un terme d'erreur  $\epsilon$  qu'on suppose gaussien. On observe  $n$  représentations indépendantes de ce modèle, d'où, en notant  $X_{ij}$  le  $j$ -ième paramètre de l'observation  $Y_i$ , le modèle :

$$Y = X\beta + \epsilon.$$

Un résultat classique affirme que le meilleur estimateur de  $\beta$  est la solution des **moindres carrés**  $\hat{\beta} = \operatorname{argmin} (\|Y - X\beta\|_2^2)$ , qui vaut  $\hat{\beta} = (X^*X)^{-1}X^*Y$ .

Cependant, que se passe-t-il lorsque  $p > n$  (ce que l'on appelle les **statistiques en grande dimension**) ? Le résultat précédent n'est plus applicable car la matrice  $X^*X$  n'est plus inversible. Cela traduit en fait un problème plus général : on n'a pas de solution unique à la minimisation de  $\|Y - X\beta\|_2^2$  (intuitivement, on peut comparer cette situation à celle d'un système de  $n$  équations linéaires à  $p$  inconnues). Il est donc nécessaire de développer d'autres méthodes pour l'estimation de  $\beta$ . Généralement, ces méthodes ont recours à la **pénalisation** pour effectuer un choix. L'exemple le plus classique est celui du Lasso, défini ainsi :

$$\hat{\beta}^L = \operatorname{argmin} \left\{ \|Y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\}.$$

Cet estimateur est une référence en statistiques en grande dimension. Nous reviendrons plus loin sur ces propriétés ; le lecteur intéressé par le Lasso et les statistiques en grandes dimensions en général pourra consulter [3]. Ce problème, loin d'être purement théorique, est au contraire au coeur de nombreuses problématiques modernes. Dans l'étude du génome humain, par exemple, on peut chercher à déterminer l'impact de milliers de gènes sur un phénomène particulier, alors qu'on ne dispose que de quelques dizaines d'observations.

## 2 Présentation du modèle

### 2.1 Régression de Poisson

Pour la suite de cet article, nous allons considérer non plus la régression linéaire gaussienne mais la **régression de Poisson**, qui s'écrit :

$$Y_i | X_i \sim \mathcal{P}(f_0(X_i)),$$

où les  $X_i$  peuvent être aléatoires ou déterministes. On les supposera dans  $[0, 1]$ .

En d'autres termes, on connaît les  $X_i$  et les  $Y_i$  et on suppose que  $Y_i$  suit une loi de Poisson dont le paramètre est une fonction déterministe inconnue  $f_0$  de  $X_i$ . Cette fonction étant strictement positive, on peut écrire  $f_0 = \exp(g_0)$  et on va supposer que  $g_0$  se décompose sur un **dictionnaire**, c'est-à-dire qu'il existe une famille  $\{\varphi_1, \dots, \varphi_p\}$  connue de fonctions telle que

$$g_0(x) = \sum_{j=1}^p \beta_{0j} \varphi_j(x).$$

Estimer  $f_0$  revient à estimer les coefficients  $\beta_{0j}$ . On cherchera donc un estimateur de  $f_0$  sous la forme

$$f_\beta(x) = \exp\left(\sum_{j=1}^p \beta_j \varphi_j(x)\right), \quad \beta \in \mathbb{R}^p.$$

Les  $\beta_j$  sont appelés **coefficients de régression**.

## 2.2 Le Lasso

On s'intéresse à la log-vraisemblance du modèle. La vraisemblance vaut :

$$L(\beta) = \prod_{i=1}^n \frac{e^{-f_\beta(X_i)} f_\beta(X_i)^{Y_i}}{Y_i!}$$

d'où on déduit la log-vraisemblance :

$$l(\beta) = \sum_{i=1}^n (Y_i \log f_\beta(X_i) - f_\beta(X_i))$$

à une constante près.

On définit  $C(\beta) = -l(\beta)$ . Maximiser la log-vraisemblance revient donc à minimiser  $C$ . Cependant, les solutions sont multiples à cause de l'hypothèse  $p > n$ , comme établi précédemment. On ajoute donc une pénalité sur les coefficients de  $\beta$  : on définit  $\text{pen}(\beta) = \sum_{j=1}^p \lambda_j |\beta_j|$  et l'estimateur suivant :

$$\hat{\beta}^L = \operatorname{argmin} \left\{ C(\beta) + \text{pen}(\beta) \right\}.$$

Les  $\lambda_j$  sont des constantes strictement positives laissées arbitraires pour le moment.

Cet estimateur est connu sous le nom de **Lasso** (Least Absolute Shrinkage and Selection Operator), introduit par Tibshirani en 1996. L'ajout de la norme  $\mathbb{L}_1$  pondérée dans la quantité à minimiser permet d'assurer l'existence d'une unique solution.

## 2.3 Estimateur de Dantzig

On définit maintenant un autre estimateur pour les grandes dimensions, l'**estimateur de Dantzig**. Il vient de l'approche suivante : en posant  $A_{ij} = \varphi_j(X_i)$ , on peut réécrire la log-vraisemblance ainsi :

$$l(\beta) = \sum_{j=1}^p \beta_j (A^* Y)_j - \sum_{i=1}^n \exp\left(\sum_{j=1}^p \beta_j A_{ij}\right).$$

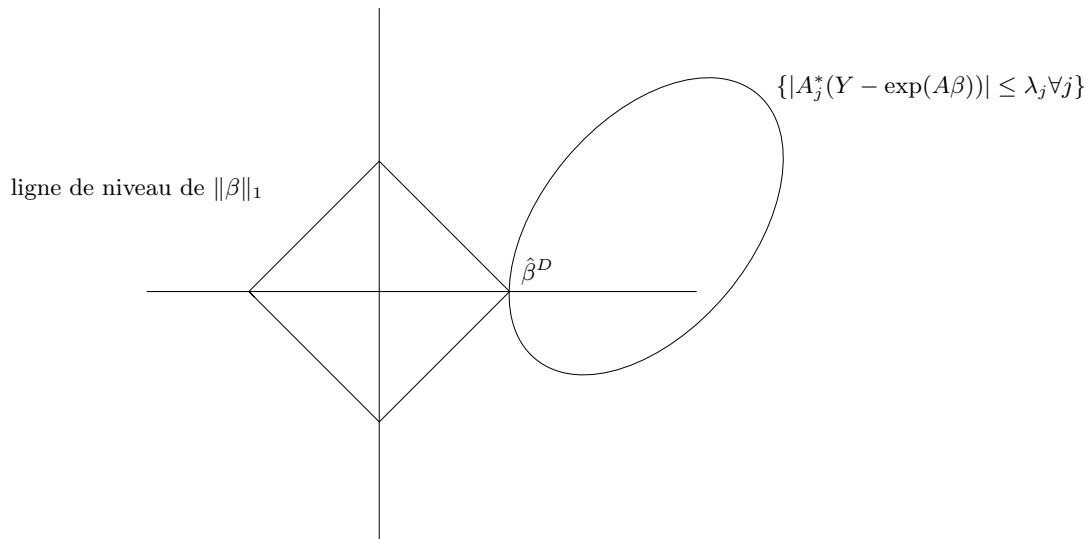
Ses dérivées partielles s'écrivent alors :

$$\frac{\partial l(\beta)}{\partial \beta_j} = A_j^* (Y - \exp(A\beta)),$$

où  $Y = (Y_1, \dots, Y_n)^*$  et  $\exp(A\beta) = (\exp((A\beta)_1), \dots, \exp((A\beta)_n))^*$ . La condition qu'on voudrait imposer est  $\frac{\partial l(\beta)}{\partial \beta_j} = 0$  mais c'est impossible. On relaxe donc cette condition en  $|A_j^* (Y - \exp(A\beta))| \leq \lambda_j$ , et on définit :

$$\hat{\beta}^D = \operatorname{argmin} \left\{ \|\beta\|_1 : \quad \forall j, |X^* (Y - \exp(A\beta))| \leq \lambda_j \right\}.$$

Il est intéressant de comparer ces deux estimateurs, d'autant plus qu'ils font intervenir les mêmes  $\lambda_j$ . L'estimateur de Dantzig est plus difficile à manipuler que le Lasso, mais d'un point de vue algorithmique, il est plus rapide à mettre en place. Tous deux présentent cependant l'avantage d'estimer une partie des coefficients précisément à zéro, à cause de la géométrie de la norme  $\mathbb{L}_1$ .



Cette propriété est utile en pratique. En effet, dans l'exemple de l'étude du génome, de nombreux gènes sont considérés dans le modèle mais n'ont en fait aucune influence sur le phénomène considéré. Il est d'ailleurs classique de supposer que le vecteur  $\beta_0$  est **S-sparse**, c'est-à-dire que seuls  $S$  de ses coefficients sont non nuls (typiquement,  $S$  est bien plus petit que  $p$ ). L'estimateur Lasso comme l'estimateur de Dantzig permettent ainsi de faire le tri entre les gènes influents (coefficients non nuls) et les autres ; on parle de **sélection de modèles**.

### 3 Choix des $\lambda_j$

Dans cette section, on cherche comment choisir les  $\lambda_j$  de manière optimale : on parle de **calibration**.

Cette calibration des  $\lambda_j$  fait intervenir la notion de processus de Poisson, qu'on rappelle ou présente ici.

#### 3.1 Processus de Poisson

Soit  $\mathbb{X}$  un ensemble mesurable. Un **processus ponctuel**  $N$  sur  $\mathbb{X}$  est un ensemble aléatoire au plus dénombrable de points de  $\mathbb{X}$ .

Pour  $A \subset \mathbb{X}$  mesurable, on note  $N_A$  le nombre de points de  $N$  dans  $A$ .

On dit alors que  $N$  est un **processus de Poisson de mesure moyenne  $\mu$**  si :

- $\forall n, \forall A_1, \dots, A_n \subset \mathbb{X}$  disjoints mesurables,  $N_{A_1}, \dots, N_{A_n}$  sont des variables aléatoires indépendantes.
- $\forall A \subset \mathbb{X}$  mesurable,  $N_A \sim \mathcal{P}(\mu(A))$ .

**Remarque.**  $\mu$  ne peut avoir d'atomes. En effet, si  $\mu(\{x\}) > 0$ , alors  $\mathbb{P}(N_{\{x\}} \geq 2) > 0$ , or il ne peut pas y avoir de répétition dans  $N$  car c'est un ensemble.

Supposons maintenant  $\mathbb{X} = \mathbb{R}_+$ . Alors  $d\mu = \lambda(x)dx$  et on appelle  $\lambda$  l'**intensité** du processus. Si  $\lambda$  est une constante, on est dans le cas plus simple d'un **processus de Poisson homogène** sur  $\mathbb{R}_+$ , qu'on peut concevoir, et construire, comme une suite de variables aléatoires indépendantes exponentielles de paramètre  $\lambda$  :

Soit  $(\tau_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires i.i.d.  $\sim \mathcal{E}(\lambda)$ . On pose  $T_i = \tau_1 + \dots + \tau_i$  et le processus ponctuel  $\{T_1, \dots, T_n, \dots\}$  est un processus de Poisson d'intensité  $\lambda$ .

Dans le cas plus général d'une fonction  $\lambda$  non constante, on définit la mesure ponctuelle  $dN = \sum_{t \in N} \delta_t$ , de sorte que pour toute fonction  $f$ ,

$$\int f(x)dN_x = \sum_{t \in N} f(t).$$

On a alors les formules suivantes sur l'espérance et la variance, qui lient le processus de Poisson  $N$  à sa mesure moyenne  $\mu$  :

$$\begin{aligned} \text{si } \int f d\mu < \infty, \quad \mathbb{E}\left(\int f dN\right) &= \int f d\mu. \\ \text{si } \int f^2 d\mu < \infty, \quad \text{Var}\left(\int f dN\right) &= \int f^2 d\mu. \end{aligned}$$

On dispose alors d'inégalités de concentrations. On présente la suivante, de "type" Bernstein :

**Lemme 1.** *Soit  $f$  une fonction bornée. Alors,  $\forall x > 0$ ,*

$$\mathbb{P}\left(\int f(dN - d\mu) \geq \sqrt{2vx} + \frac{\|f\|_\infty x}{3}\right) \leq e^{-x}, \quad (1)$$

où  $v = \text{Var}\left(\int f dN\right) = \int f^2 d\mu$ .

Ces notions, et en particulier ce lemme, permettent d'établir un théorème qui énonce le meilleur choix de  $\lambda_j$ .

### 3.2 Un théorème pour la calibration

Les  $\lambda_j$  doivent être choisis en fonction des données du problème. On va voir ici comment réaliser ce choix. Pour le moment, on va s'intéresser uniquement à l'estimateur de Dantzig.

On voudrait que  $\beta_0$  satisfasse à la **condition de Dantzig** :  $\forall j, |X_j^*(Y - \exp(A\beta))| \leq \lambda_j$ .

Comme  $\exp(A\beta_0) = \mathbb{E}[Y]$ , cela revient à chercher les  $\lambda_j$  tels que pour tout  $j$ ,

$$|A_j^*(Y - \mathbb{E}[Y])| \leq \lambda_j$$

avec forte probabilité.

$\lambda_j$  ne doit donc pas être pris trop petit. Cependant, s'il est pris trop grand, on risque d'estimer trop de paramètres à zéro. Un compromis est donc nécessaire.

La calibration se fait via le résultat suivant. On définit

$$V_j = \text{Var}(A_j^*Y) = \sum_{i=1}^n f_0(X_i)\varphi_j^2(X_i)$$

et son estimateur

$$\hat{V}_j = \sum_{i=1}^n Y_i \varphi_j^2(X_i).$$

**Théorème 1.** *Soit  $\gamma > 1$ . On pose*

$$\tilde{V}_j = \hat{V}_j + \sqrt{2\gamma(\log p)\|\varphi_j\|_{n,\infty}^2 \hat{V}_j} + 3\gamma(\log p)\|\varphi_j\|_{n,\infty}^2 \hat{V}_j$$

où  $\|\varphi_j\|_{n,\infty} = \max_{i \in \{1, \dots, n\}} |\varphi_j(X_i)|$ .

Alors

$$\mathbb{P}\left(|A_j^*(Y - \mathbb{E}[Y])| > \lambda_j\right) \leq \frac{3}{p^\gamma}, \quad (2)$$

où  $\lambda_j = \sqrt{2\gamma \log p \tilde{V}_j} + \frac{\gamma \log p}{3} \|\tilde{\varphi}_j\|_\infty$ .

$\gamma$  est un paramètre laissé libre. En pratique, on obtient de bons résultats en le choisissant proche de 1.

La preuve se fait en ramenant la quantité à contrôler  $|A_j^*(Y - \mathbb{E}[Y])|$  à une quantité faisant intervenir un processus de Poisson, de la manière suivante. On revient à notre modèle de régression de Poisson. Quitte à les renuméroter, on peut supposer que  $X_i < X_{i+1}$  pour tout  $i$ . On définit alors la fonction  $h$  constante par morceaux qui vaut

$$h(t) = \frac{f_0(X_i)}{X_{i+1} - X_i}$$

pour  $t \in [X_i, X_{i+1}[$ .

Soit  $N$  un processus de Poisson d'intensité  $h$ , donc de mesure moyenne  $\mu = h(t)dt$ . Des calculs simples permettent de lier la quantité à contrôler  $|A_j^*(Y - \mathbb{E}[Y])|$  au processus de Poisson  $N$  :

$$|A_j^*(Y - \mathbb{E}[Y])| = \left| \int \tilde{\varphi}_j(t)(dN(t) - h(t)dt) \right|.$$

Le lemme 1 permet de contrôler cette quantité en fonction de la variance  $V_j$ , puis de contrôler  $V_j$  par la quantité  $\tilde{V}_j$ . On obtient ainsi une inégalité valable avec probabilité  $1 - 3e^{-x}$ . Le choix  $x = \gamma \log p$  permet d'aboutir à (2).

### 3.3 Inégalité oracle

Il reste encore à montrer l'efficacité de l'estimateur de Dantzig pour les  $\lambda_j$  choisis. On construit pour cela une **inégalité oracle**.

Il nous faut d'abord définir certains coefficients. Soit  $X$  une matrice de taille  $n \times p$ . Pour tout ensemble  $T \subset \{1, \dots, p\}$ , on note  $X_T$  la matrice de taille  $n \times |T|$  obtenue en prenant les colonnes de  $X$  correspondant aux indices de l'ensemble  $T$ .

Pour tout  $S \leq p$ , on définit  $\delta_S^X$  comme étant la plus grande quantité telle que, pour tout ensemble  $T \subset \{1, \dots, p\}$  de taille  $|T| \leq S$  et tout vecteur  $c$  de longueur  $|T|$ ,

$$\|X_T c\|_2^2 \geq \delta_S^X \|c\|_2^2.$$

Pour tous  $S$  et  $S'$  tels que  $S + S' \leq p$ , on définit de plus  $\theta_{S,S'}^X$  comme la plus petite quantité telle que, pour tous sous-ensembles disjoints  $T$  et  $T'$  de  $\{1, \dots, p\}$  et tous vecteurs  $c$  et  $c'$  de longueurs respectives  $|T| \leq S$  et  $|T'| \leq S'$ ,

$$|(X_T c)^*(X_{T'} c')| \leq \theta_{S,S'}^X \|c\|_2 \|c'\|_2.$$

Soit  $\tilde{A}$  la matrice de terme général  $\tilde{A}_{ij} = \sqrt{f_0(X_i)} A_{ij} = \sqrt{f_0(X_i)} \varphi_j(X_i)$ . Dans la suite, on note  $\delta_S = \delta_S^A$  et  $\tilde{\delta}_S = \delta_S^{\tilde{A}}$ ,  $\theta_{S,S'} = \theta_{S,S'}^A$  et  $\tilde{\theta}_{S,S'} = \theta_{S,S'}^{\tilde{A}}$ .

On a alors le théorème suivant :

**Théorème 2.** *On suppose :*

- $\beta_0$  est  $S$ -sparse : l'ensemble  $T_0 = \{j / \beta_{0j} \neq 0\}$  est de taille  $S$  ;
- $3S \leq p$  ;
- $\tilde{\theta}_{S,2S} < \tilde{\delta}_{2S}$ .

Alors, en choisissant les  $\lambda_j$  comme dans la section précédente, on a, avec probabilité  $1 - O(\frac{1}{p^{\gamma-1}})$ ,

$$\|\hat{\beta}^D - \beta_0\|_2^2 \leq 256\gamma \log p \frac{\sum_{j \in T_{01}} \hat{V}_j}{(\tilde{\delta}_{2S} - \tilde{\theta}_{S,2S})^2},$$

où  $T_{01} = T_0 \cup T_1$  avec  $T_1$  l'ensemble des  $S$  indices  $j \notin T_0$  correspondant aux  $S$  plus grandes valeurs de  $\hat{\beta}$ .

**Remarque.** L'hypothèse  $\tilde{\theta}_{S,2S} < \tilde{\delta}_{2S}$  n'est pas vérifiable car elle porte sur la matrice  $\tilde{A}$  inconnue. Si on suppose  $f_0$  continue, il existe alors des constantes  $m$  et  $M$  strictement positives telles que  $m \leq f_0 \leq M$  sur  $[0, 1]$  et on peut remplacer notre hypothèse par celle-ci, plus forte :

$$M\theta_{S,2S} < m\delta_{2S}.$$

On remplace également  $\tilde{\delta}_{2S} - \tilde{\theta}_{S,2S}$  par  $m\delta_{2S} - M\theta_{S,2S}$  dans l'inégalité.

## 4 Perspectives

Il reste encore beaucoup à faire sur l'estimation en régression de Poisson, d'autant plus que c'est un modèle qui pourrait devenir assez actuel maintenant que certains phénomènes biologiques, jusqu'ici modélisés de manière gaussienne, sont mieux représentés par des processus de Poisson. Voici quelques perspectives pour l'avenir.

### 4.1 De meilleures inégalités oracles

Le théorème 2 n'est pas optimal. Il porte sur le carré de l'erreur  $\mathbb{L}_2$  d'estimation de  $\beta_0$ , là où on souhaiterait plutôt un résultat sur l'estimation de  $f_0$ .

De plus, pour être véritablement applicable, il faut supposer non seulement que  $f_0$  est bornée (ce qui n'est pas une grosse contrainte), mais aussi qu'on a des informations sur ses bornes, ce qui est bien plus fort en pratique.

Il est probablement nécessaire de poser des conditions plus précises sur le dictionnaire  $(\varphi_j)_{j \in \{1, \dots, p\}}$ . Une meilleure majoration peut être trouvée si on prend pour dictionnaire une base orthonormée classique (base de Haar, ondelettes...).

Une autre piste, à base de l'estimateur Lasso, permet d'obtenir une inégalité oracle plus propre. On rappelle que pour tout  $\beta$ , on note  $f_\beta(x) = \exp\left(\sum_{j=1}^p \beta_j \varphi_j(x)\right)$ . On peut alors définir la **divergence de Kullback** :

$$\begin{aligned} \tilde{K}_n(\beta_0, \beta) &= \mathbb{E}_{\mathbb{P}_{\beta_0}}[l(f_0) - l(f_\beta)] \\ &= \sum_{i=1}^n \left[ f_0(X_i) \left( \log f_0(X_i) - \log f_\beta(X_i) \right) - (f_0(X_i) - f_\beta(X_i)) \right]. \end{aligned}$$

Alors un résultat (de [6]) affirme que, pour tout  $\zeta > 0$ , on peut poser  $a_0 = 3 + \frac{4}{\zeta}$  et alors, sous l'hypothèse  $RE(s, a_0)$  qui demande essentiellement que la matrice  $\tilde{A}^* \tilde{A}$  vérifie une forme de "définitivité" restreinte aux vecteurs  $b$  vérifiant  $\|b_{J^c}\|_1 \leq a_0 \|b_J\|_1$  pour tout sous-ensemble  $J$  de  $\{1, \dots, p\}$  de taille inférieure ou égale à  $s$ , on a :

$$\tilde{K}_n(\beta_0, \hat{\beta}) \leq (1 + \zeta) \inf_{\beta \in \mathbb{R}^p, |S(\beta)| \leq s} \left\{ \tilde{K}_n(\beta_0, \beta) + C(\zeta, s) |S(\beta)| (\max_j \lambda_j)^2 \right\}.$$

Il s'agit ici d'une véritable inégalité oracle, qui affirme que l'estimateur  $\hat{\beta}$  approche  $\beta_0$  aussi bien que le meilleur  $\beta$ , à un terme additif proportionnel à la taille du support de  $\beta$ , et à un facteur  $(1 + \zeta)$  près.

Malheureusement, ce théorème demande pour l'instant une hypothèse supplémentaire trop forte (l'équation (12) de [6]). Il semble néanmoins possible d'améliorer cette hypothèse.

## 4.2 Group-Lasso et Group-Dantzig

Toujours dans l'exemple de l'étude du génome, il s'avère en réalité que les gènes n'agissent pas seuls, mais en groupe. On cherche donc non plus à sélectionner des gènes, mais des groupes indissociables de gènes.

On définit donc dans cette optique les estimateurs Group-Lasso et Group-Dantzig. Soit  $\{G_1, \dots, G_M\}$  une partition de  $\{1, \dots, p\}$ . On pose alors

$$\hat{\beta}^{gL} = \operatorname{argmin} \left\{ C(\beta) + \sum_{k=1}^M \lambda_k \|\beta_{G_k}\|_2 \right\}$$

et

$$\hat{\beta}^{gD} = \operatorname{argmin} \left\{ \sum_{k=1}^M \|\beta_{G_k}\|_2 : \forall k \in \{1, \dots, M\}, \|A_{G_k}^* (Y - \exp(A\beta))\|_2 \leq \lambda_k \right\}.$$

On remarque que si  $M = p$ , alors pour tout  $k$ , quitte à effectuer une permutation,  $G_k = k$ ,  $\|\beta_{G_k}\|_2 = |\beta_k|$  et on retrouve simplement les estimateurs Lasso et de Dantzig classiques.



Je remercie Vincent Rivoirard pour son aide et sa patience tout au long de mon stage, ainsi que Patricia Reynaud-Bouret pour son accueil chaleureux à l'université de Nice.

## Références

- [1] Antoniadis, A., Fryzlewicz, P. and Letué, F. (2010) *The Dantzig selector in Cox's proportional hazards model*. Scand. J. Stat. **37**(4), 531–552.
- [2] Bertin, K., Le Pennec, E. and Rivoirard, V. (2011) *Adaptive Dantzig density estimation*, Annales de l'Institut Henri Poincaré Probabilités et Statistiques, **47**(1), 43–74.
- [3] Bühlmann, P. and van de Geer, S. (2011) *Statistics for high-dimensional data*. Springer Series in Statistics. Springer, Heidelberg.
- [4] Candès, E. and Tao, T. (2007) *The Dantzig selector : statistical estimation when  $p$  is much larger than  $n$* , Ann. Statist. **35**(6), 2313–2351.
- [5] James, G.M. and Radchenko, P. (2009) *A generalized Dantzig selector with shrinkage tuning*. Biometrika **96**(2), 323–337.
- [6] Lemler, S. (2012) *Oracle inequalities for the Lasso for the conditional hazard rate in a high-dimensional setting*.
- [7] Kingman, J.F.C. (1993) *Poisson processes*. Oxford studies in Probabilites.
- [8] Reynaud-Bouret, P. and Rivoirard, V. (2010) *Near optimal thresholding estimation of a Poisson intensity on the real line*. Electronic Journal of Statistics, **4**, 172–238.