# Estimation et sélection de variables dans un modèle joint de survie et de données longitudinales avec des effets aléatoires.

Antoine Caillebotte [1,2] & Estelle Kuhn [2] & Sarah Lemler [3]

[1] *Université Paris-Saclay, INRAE, UMR GQE-Moulon, France, caillebotte.antoine@inrae.fr,*
[2] *Université Paris-Saclay, INRAE, UR MaIAGE, France, estelle.kuhn@inrae.fr,*
[3] *Université Paris-Saclay, CentraleSupélec, Laboratoire MICS, France, sarah.lemler@centralesupelec.fr*

**Résumé.** Ce travail se concentre sur l'étude jointe d'un modèle de survie et d'un modèle à effets mixtes pour expliquer le temps de survie à partir de données longitudinales et de covariables de grande dimension. Les données longitudinales sont modélisées à l'aide d'un modèle non linéaire à effets mixtes, dans lequel la fonction de régression sert de fonction de lien incorporée dans un modèle de Cox en tant que covariable. De cette manière, les données longitudinales sont liées à la durée de survie à un moment donné. De plus, le modèle de Cox prend en compte l'inclusion de covariables de grande dimension. Les principaux objectifs de cette recherche sont doubles : premièrement, identifier les covariables pertinentes qui contribuent à expliquer le temps de survie, et deuxièmement, estimer tous les paramètres inconnus du modèle joint. Pour ce faire, nous considérons la maximisation de la vraisemblance pénalisée par le LASSO. Pour résoudre le problème d'optimisation, nous introduisons un gradient stochastique préconditionné adapté aux variables latentes du modèle non linéaire à effets mixtes, associé à un opérateur proximal qui permet de gérer la non-différentiabilité de la pénalité. Nous proposons une large étude de simulations afin de montrer les performances de la procédure proposée, à la fois en termes de sélection de variables et d'estimation des paramètres du modèle considéré.

**Mots-clés.** Statistique appliquée, grande dimension et réduction de dimension, Données de survie, données censurées, biostatistique, statistique computationnelle.

**Abstract.** This paper considers a joint survival and mixed-effects model to explain the survival time from longitudinal data and high-dimensional covariates. The longitudinal data is modeled using a nonlinear mixed effects model, where the regression function serves as a link function incorporated into a Cox model as a covariate. In that way, the longitudinal data is related to the survival time at a given time. Additionally, the Cox model takes into account the inclusion of high-dimensional covariates. The main objectives of this research are two-fold: first, to identify the relevant covariates that contribute to explaining survival time, and second, to estimate all unknown parameters of the joint model. For that purpose, we consider the maximization of a LASSO penalized likelihood. To tackle the optimization problem, we implement a pre-conditioned stochastic gradient to handle the latent variables of the nonlinear mixed-effects model associated with a proximal operator to manage the non-differentiability of the penalty. We provide an extensive simulation study showcasing the performance of the proposed variable selection and the parameter estimation method.

**Keywords.** Applied statistics, high dimension and dimension reduction, survival data, censored data , biostatistics, computational statistics.

# 1 Introduction

Une problématique très actuelle dans de nombreux domaines consiste à mieux comprendre les interactions entre des phénomènes dynamiques dépendants. On peut considérer, en médecine, la dynamique des tumeurs d'un patient en oncologie et les effets des traitements anticancéreux administrés au patient. Les phénomènes considérés sont souvent complexes, tant d'un point de vue de leurs modes d'interaction que de leur dynamique temporelle et spatiale. De plus, ces phénomènes sont souvent observés dans des populations d'individus hétérogènes ou structurés.

La modélisation mathématique s'est avérée être un outil puissant pour comprendre les interactions entre plusieurs phénomènes dynamiques. Elle permet également de prendre en compte la variabilité présente dans la population observée d'individus. La modélisation jointe de plusieurs phénomènes a en particulier démontré son efficacité dans plusieurs domaines, notamment la médecine, la pharmacologie et la biologie [Kerioui *et al.*, 2022]. Un cas particulier de modèles joints concerne la modélisation simultanée de données longitudinales et de données de survie observées sur le même individu. Dans ce type de modèle joint, les données longitudinales sont souvent modélisées par un modèle à effets mixtes [Davidian et Giltinan, 1995], et les données de survie par un modèle de Cox [Cox, 1972]. Ce dernier permet de modéliser le risque instantané de la variable de survie en fonction de covariables. La modélisation des données longitudinales intervient en tant que covariable dans le modèle de Cox via une fonction de lien. L'objectif est alors d'estimer les paramètres du modèle à partir des observations et de sélectionner les covariables pertinentes [Rizopoulos, 2012]. En raison de la présence de variables latentes dans le modèle à effets mixtes, l'inférence par maximum de vraisemblance est délicate à réaliser et nécessite d'être adaptée, par exemple via les algorithmes de type Expectation Maximization (EM) [Rizopoulos, 2012]. Les algorithmes de type EM, tels que le Stochastic Approximation Expectation Maximization (SAEM), sont les approches les plus classiques pour inférer les paramètres en présence de variables latentes. Ils sont particulièrement faciles à mettre en œuvre dans le contexte d'une famille exponentielle courbe basée sur des statistiques exhaustives du modèle. De plus, des résultats de convergence théoriques de l'algorithme ont été établis dans ce contexte. Cependant, lorsque le modèle n'appartient pas à la famille exponentielle, ce qui est le cas dans notre contexte, l'algorithme est difficile à mettre en œuvre et les garanties théoriques ne sont pas établies.

Les méthodes basées sur le gradient, souvent omises, mais pourtant adaptées à l'estimation des paramètres dans les modèles latents, ne nécessitent pas d'être dans un modèle de la famille exponentielle. Ainsi, [Baey *et al.*, 2023] a suggéré d'utiliser un algorithme de gradient stochastique préconditionné pour traiter l'estimation des paramètres en présence de variables latentes. À noter que des méthodes numériques bayésiennes ont également été proposées en parallèle [Rizopoulos, 2012], [Kerioui *et al.*, 2022].

Par ailleurs, dans de nombreuses applications, les moyens technologiques actuels permettent de collecter des covariables explicatives de grande dimension. Celles-ci peuvent être, par exemple, des marqueurs génétiques ou des données omiques. Au-delà de la richesse d'informations fournies par ces covariables, elles génèrent des difficultés dans l'analyse statistique des modèles, car il est nécessaire d'adapter les approches statistiques et numériques à leur grande dimension. Une approche possible est de considérer un estimateur pénalisé, tel que le LASSO [He *et al.*, 2015], et des méthodes numériques adaptées, telles que le gradient proximal stochastique [Achab, 2017, Fort *et al.*, 2017].

Nous considérons dans cette contribution un modèle joint qui combine, par le biais d'une fonction de lien, un modèle non linéaire à effets mixtes pour les données longitudinales et un modèle de Cox pour les temps de survie, incluant des covariables de grande dimension. Notre travail vise à sélectionner les variables pertinentes parmi les covariables de grande dimension dans la partie survie du modèle joint sur la base de l'ensemble des données et ensuite d'estimer les paramètres inconnus du modèle. À cette fin, nous utilisons un gradient proximal stochastique préconditionné pour traiter les variables latentes dans le modèle joint et la pénalisation LASSO. L'algorithme proposé est facile à mettre en œuvre dans des modèles joints généraux sans supposer que la densité du modèle appartient à la famille exponentielle courbe.

Le modèle joint est détaillé dans la section 2. Dans la section 3, nous présentons la méthode d'inférence proposée basée sur un estimateur pénalisé par LASSO et une procédure numérique basée sur un algorithme de gradient proximal stochastique. Enfin, nous illustrons la méthodologie à la section 4 par une étude de simulation.

# 2 Modèle joint pour données de survie et longitudinales

Nous considérons $N$ individus et étudions, pour chaque individu $i$, le temps de survie $\mathbf{T}_i$, correspondant à la durée jusqu'à la survenue d'un événement d'intérêt, et des données longitudinales, plus précisément des observations répétées $J$ fois notées $\mathbf{Y}_{i,j}$ avec $i \in \{1, \ldots, N\}$ et $j \in \{1, \ldots, J\}$. Soit $\mathcal{D}_i = (\mathbf{Y}_i, \mathbf{T}_i, \delta_i)$ les variables observées.

## 2.1 Modèle de Survie

Le temps de survie $\mathbf{T}_i$ de l'individu $i$ est le temps entre un instant initial et la survenue d'un événement d'intérêt et est modélisé par une variable aléatoire positive. Pour caractériser la distribution de $\mathbf{T}_i$, nous utilisons la fonction de risque définie par :

$$h_i(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}(t \leq \mathbf{T}_i < t + \Delta t | \mathbf{T}_i \geq t)}{\Delta t}; \forall t \geq 0. \tag{1}$$

Le modèle de Cox [Cox, 1972] est l'un des modèles les plus classiques en analyse de survie. Il relie la fonction de risque du temps de survie $\mathbf{T}_i$ aux covariables $U_i \in \mathbb{R}^p$, avec $p$ le nombre

de covariables. Le modèle de Cox pour l'individu $i$ est écrit comme suit :

$$h(t|U_i) = h_{\theta_b}(t)\exp(\beta^T U_i), \tag{2}$$

avec $\beta \in \mathbb{R}^p$ un paramètre de régression et $h_b$ la fonction de risque de base qui caractérise un comportement commun dans la population observée. Nous considérerons un risque de base paramétrique où $\theta_b$ est le vecteur de paramètres. Par conséquent, les paramètres inconnus du modèle de Cox sont $\beta$ et $\theta_b$.

En plus des covariables, nous souhaitons expliquer une partie de la variabilité du risque en utilisant la dynamique des données longitudinales, qui sera modélisée à l'aide d'un modèle à effets mixtes non linéaire. Nous présentons d'abord le modèle à effets mixtes avant d'expliquer l'intégration de cette nouvelle composante dans le modèle de Cox.

## 2.2 Modèle Non Linéaire à Effets Mixtes

Les données longitudinales sont observées $J$ fois pour chaque individu $i \in \{1, \dots, N\}$. Notons par $\mathbf{Y}_{i,j}$ la $j$-ème observation de l'individu $i$ pour $j \in \{1, \dots, J\}$ et $i \in \{1, \dots, N\}$. Nous modélisons cette observation longitudinale à l'aide d'une fonction non linéaire $m$ qui dépend des paramètres individuels représentés par la variable latente $Z_i$ comme suit :

$$\begin{cases} \mathbf{Y}_{i,j} & = m(t_j; Z_i) + \varepsilon_{i,j}, \\ Z_i & \underset{i.i.d.}{\sim} \mathcal{N}(\mu, \Gamma) \ ; \ \varepsilon_{i,j} \underset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2), \end{cases} \quad \forall 1 \leq i \leq N, 1 \leq j \leq J \tag{3}$$

où, $t_j$ est le temps de la $j$-ème observation, et $\varepsilon_{i,j}$ est un bruit additif supposé centré gaussien avec une variance inconnue $\sigma^2$. La variable latente $Z_i$ décrit la variabilité interindividuelle de la population. On suppose que $Z_i$ suit une distribution gaussienne avec une espérance inconnue $\mu$ et une variance $\Gamma$ inconnue. Les paramètres inconnus du modèle à effets mixtes non linéaire sont donc $\mu, \Gamma$, et $\sigma^2$.

Nous introduisons ensuite la fonction de lien, qui combine les deux modèles précédents en modélisant l'influence de la dynamique de l'observation longitudinale sur la fonction de risque.

## 2.3 Modèle Joint de Survie et Longitudinal à Effets Mixtes

Nous supposons que le risque du temps de survie est lié à la dynamique des données longitudinales à travers la fonction de lien au sein du modèle joint définie par :

$$\begin{cases} h(t|\mathcal{M}(t, Z_i), U_i) & = h_{\theta_b}(t)\exp(\beta^T U_i + \alpha m(t, Z_i)) \\ Y_{i,j} & = m(t_j; Z_i) + \varepsilon_{i,j} \\ Z_i \underset{i.i.d.}{\sim} \mathcal{N}(\mu, \Gamma) & \ ; \ \varepsilon_{i,j} \underset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2), \end{cases} \quad \forall 1 \leq i \leq N, 1 \leq j \leq J \tag{4}$$

où $\mathcal{M}(t; Zi) = \{m(s; Zi) | \forall s, 0 \leq s < t\}$ décrit les valeurs passées de la dynamique longitudinale jusqu'au temps $t$. Le paramètre $\alpha$ représente l'influence de la dynamique longitudinale sur les données de survie.

Les paramètres inconnus du modèle joint comprennent les paramètres du modèle de Cox et ceux du modèle à effets mixtes non linéaires, ainsi que le paramètre de fonction de lien du modèle joint. Nous notons $\theta = (\theta_b, \beta, \mu, \Gamma, \sigma^2, \alpha) \in \Theta$ le vecteur des paramètres inconnus avec $\Theta \subset \mathbb{R}^d$ l'espace des paramètres.

# 3 Inférence des paramètres

Dans cette section, nous proposons une méthode d'estimation des paramètres du modèle présenté ci-dessus.

## 3.1 Définition de la Vraisemblance Marginale

Nous considérons l'estimateur du maximum de vraisemblance pour estimer les paramètres du modèle joint. Dans le contexte des modèles à variables latentes, on considère la vraisemblance marginale, définie par :

$$\mathcal{L}_{marg}(\theta; \mathcal{D}) = \prod_{i=1}^{n} \int p_\theta(\mathcal{D}_i, Z_i) dZ_i \tag{5}$$

où $p_\theta(\mathcal{D}, Z)$ est la densité du couple $(\mathcal{D}, Z)$.

En raison de l'intégrale, il est difficile de calculer directement le maximum de la vraisemblance marginale, qui n'a pas de forme analytique dans ce modèle de variable latente. Par conséquent, nous utilisons des méthodes numériques pour résoudre ce problème de maximisation.

## 3.2 Définition de l'estimateur pénalisé pour la sélection de variables

Nous introduisons une pénalité et considérons un estimateur du maximum de vraisemblance pénalisé pour traiter la grande dimension des covariables. Notre objectif est de sélectionner les variables pertinentes parmi les covariables du modèle de survie. Nous utilisons la procédure LASSO (Least Absolute Shrinkage and Selection Operator) qui a été initialement développée pour les modèles de régression linéaire et le modèle de Cox [Tibshirani, 1997].

Nous choisissons de ne pénaliser que le vecteur de paramètres $\beta$ : $\text{pen}_{LASSO}(\theta) = \|\beta\|_1 = \sum_{k=1}^{p} |\beta_k|$. Nous définissons l'estimateur du maximum de vraisemblance pénalisé par :

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\max} \left\{ \log L_{marg}(\theta; \mathcal{D}) - \lambda \ \text{pen}_{LASSO}(\theta) \right\}, \tag{6}$$

où $\Theta$ représente l'espace des paramètres et où $\lambda$ est un paramètre positif appelé paramètre de régularisation. Plus la valeur de $\lambda$ est grande, plus $\beta$ sera contraint d'avoir des composantes nulles. Inversement, plus la valeur de $\lambda$ est petite, plus les composantes de $\beta$ seront libres.

Les méthodes classiques utilisées pour estimer les paramètres sont des algorithmes de type Expectation Maximization. Ces procédures sont bien adaptées aux modèles appartenant à la famille exponentielle, ce qui n'est pas le cas du modèle joint considéré. Récemment, [Baey *et al.*, 2023] a présenté une descente de gradient stochastique préconditionnée pour l'estimation dans des modèles à variables latentes. De plus, en raison de la non-différentiabilité de la pénalité considérée, nous utilisons un algorithme proximal tel que présenté par [Achab, 2017] et [Fort *et al.*, 2017]. Ainsi, nous intégrons un gradient proximal dans la procédure présentée dans [Baey *et al.*, 2023]. Au final nous mettons en œuvre un algorithme de gradient proximal stochastique préconditionné pour calculer l'estimateur.

## 3.3 Algorithme d'Estimation

Nous mettons en œuvre un gradient proximal stochastique préconditionné, appelé SPG-FIM dans la suite. L'algorithme est divisé en trois étapes : une réalisation des variables latentes est échantillonnée lors d'une première étape appelée *Simulation*, qui utilise un algorithme de Metropolis-Hastings. La deuxième étape est la descente classique du gradient sur la vraisemblance complète approximée, l'étape *Forward*. Suivant la procédure présentée dans [Baey *et al.*, 2023], nous avons choisi d'utiliser un préconditionnement du gradient avec une estimation de la matrice d'information de Fisher (FIM). Cette dernière est mise à jour au cours des itérations à l'aide de l'estimation présentée par [Delattre et Kuhn, 2023]. La dernière étape, appelée *Backward*, traite le terme de pénalité. Nous appliquons l'opérateur proximal classique, défini ci-dessous.

$$\text{Prox}_{\text{pen}}(\beta) = \underset{\beta' \in \mathbb{R}^p}{\arg\min} \left( \text{pen}(\beta') + \frac{1}{2} \left\| \beta - \beta' \right\|_2^2 \right). \tag{7}$$

Avec la pénalité LASSO, l'opérateur proximal a la forme explicite suivante :

$$(\text{Prox}_{LASSO}(\beta))_i = \begin{cases} 0 & \text{if } |\beta_i| < \lambda \\ \beta_i - \lambda & \text{if } \beta_i \geq \lambda \\ \beta_i + \lambda & \text{if } \beta_i \leq -\lambda \end{cases} ; \ \forall i \in \{1, ..., p\}. \tag{8}$$

L'étape *Backward* correspond à l'application de l'opérateur proximal sur le résultat de l'étape *Forward*. L'Algorithme 1 détaille les étapes de SPG-FIM. Il convient de noter qu'il est possible de différencier les deux suites de taille de pas impliquées dans l'approximation stochastique ou la descente de gradient. Toutefois, par souci de clarté, nous les avons notées de la même manière ici.

Comme la pénalité ne dépend que de $\beta$, l'opérateur proximal sélectionne les composantes de $\beta$ qui semblent les plus explicatives des données. Il calcule une solution parcimonieuse pour $\beta$ mais applique également un rétrécissement sur les composantes non nulles, de sorte que l'estimateur LASSO est biaisé. C'est pourquoi nous détaillons dans ce qui suit une méthode permettant d'obtenir un estimateur non biaisé.

**Algorithm 1:** Gradient proximal stochastique avec préconditionnement FIM (SPG-FIM)

**Require:** Nombre d'itérations $K \geq 1$ ; séquence de pas $(\gamma_k)_{k\geq 1}$
**Initialize** Point de départ $\theta_0 \in \mathbb{R}^d$, $\Delta_0$
**for** $k = 1$ **to** $K$ **do**
  • **Étape de Simulation :**
    **Obtenir** $Z^{(k)}$ **avec une étape de Metropolis-Hastings**
  • **Calcul du gradient :** $v_k = \frac{1}{N}\sum_{i=1}^{N} \nabla \log p_{\theta_k}(\mathcal{D}_i, Z_i^{(k)})$
  • **Calcul du FIM :**
    • **Calculer l'approximation stochastique**
    $\forall i \in \{1, ..., N\}, \Delta_i^{(k)} = (1-\gamma_k)\Delta_i^{(k-1)} + \gamma_k \nabla \log p_{\theta_k}(\mathcal{D}_i, Z_i^{(k)})$
    • **Calculer le FIM :** $FIM_k = \frac{1}{N}\sum_{i=1}^{N} \Delta_i^{(k)} \left(\Delta_i^{(k)}\right)^T$
  • **Descente de gradient :**
    • **Étape Forward :** $\omega_{k+1} = \theta_k - \gamma_k FIM_k^{-1} v_k$
    • **Étape Backward :** $\theta_{k+1} = \mathrm{Prox}_{\gamma_k \mathrm{pen}}(\omega_{k+1})$
**end**
**return** $\hat{\theta} = \theta_K$

## 3.4 Procédure d'Estimation

Comme expliqué ci-dessus, l'opérateur proximal (8) a un effet rétrécissant sur l'estimateur après son application, ce qui signifie que les valeurs trouvées pour $\beta$ sont plus petites que prévu, et donc l'estimateur de $\beta$ est biaisé. Pour débiaiser l'estimateur, nous allons procéder en deux étapes. Une première étape exploratoire nous permet de sélectionner le support du vecteur $\beta$ à l'aide d'une procédure Lasso,

$$\hat{\theta}_{\mathbf{LASSO}}(\lambda) = \arg\max_{\theta\in\Theta} \left\{\log \mathcal{L}_{marg}(\theta; \mathcal{D}) - \lambda \mathrm{pen}_{LASSO}(\theta)\right\}.$$

La seconde étape consiste à maximiser la vraisemblance non pénalisée en les paramètres sélectionnés à l'étape précédente. Nous devons également sélectionner une valeur bien équilibrée pour le paramètre de régularisation, il est d'usage de déterminer la valeur de $\lambda$ par validation croisée. Mais ici, sans contexte prédictif, le paramètre de régularisation de la procédure Lasso est sélectionné à l'aide du critère BIC défini de la façon suivante:

$$BIC(\lambda) = -2\log(\mathcal{L}_{marg}(\hat{\theta}_{\mathbf{LASSO}}(\lambda); \mathcal{D})) + k\ \log(N(1+J)).$$

où k est le nombre de composantes non nulles dans $\hat{\beta}_{LASSO}$.

# 4 Étude de Simulation

Dans cette section, nous proposons d'étudier les performances de la procédure que nous venons de présenter. Nous considérons le modèle conjoint défini par 4 avec la fonction logistique classique pour le modèle non linéaire à effets mixtes, définie par :

$$m : t \mapsto \frac{Z_1}{1 + \exp\left(\frac{Z_2 - t}{\mu_3}\right)}, \tag{9}$$

Nous modélisons pour chaque individu $i$ le paramètre individuel correspondant $Z_i \in \mathbb{R}^2$ par une variable aléatoire gaussienne avec une espérance $(\mu_1, \mu_2) \in \mathbb{R}^2$ et une variance diagonale $\Gamma = diag(\gamma_1^2, \gamma_2^2)$. $\mu_3$ est considéré comme un paramètre de population inconnu. Nous considérons le risque de base comme étant une Weibull définie comme $h_{a,b}(t) = ba^{-b}t^{b-1}$, où $a$ et $b$ sont connus.

## Configuration de la simulation

Nous avons généré 50 ensembles de données selon le modèle joint présenté précédemment dans l'équation 4. Pour chaque valeur différente de $p$, nous choisissons le vecteur $\beta$ de telle sorte que les quatre premières composantes soient égales à $(-2, -1, 1, 2)$ et que le reste soit égal à zéro. Nous générons également la matrice des covariables $U$ avec $N$ lignes et $p$ colonnes, suivant une distribution uniforme $U_{i,l} \sim \mathcal{U}([-1, 1]),\ \forall i \in \{1, ..., N\}, l \in 1, ..., p$. Toutes les valeurs des paramètres sont détaillées dans la Table 1.

Table 1: Valeurs réelles des paramètres utilisées pour la simulation

| Paramètres | $\mu_1$ | $\mu_2$ | $\mu_3$ | $\gamma_1^2$ | $\gamma_2^2$ | $\sigma^2$ | $\alpha$ |
|---|---|---|---|---|---|---|---|
| Valeur réelle | 0.3 | 90 | 7.5 | 2.5 $10^{-3}$ | 20 | $10^{-3}$ | 11.11 |
| Paramètres | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ | ... | $\beta_p$ |
| Valeur réelle | -2 | -1 | 1 | 2 | 0 | ... | 0 |

Notre objectif est de montrer la consistance numérique de l'estimateur 6, nous nous concentrons donc sur quatre scénarios où le nombre d'individus observés augmente $N \in \{50, 100, 200, 300\}$ lorsque le nombre de covariables et le nombre d'observations longitudinales sont fixes, $p = 200$, $J = 5$. Nous considérons séparément les erreurs quadratiques moyennes relatives ($rrmse$) des paramètres de grande dimension $\beta \in \mathbb{R}^p$ sélectionnés par la méthode et des autres, définis par $\nu \in \mathbb{R}^d$.

La Table 2 donne les erreurs relatives calculées sur 50 jeux de données pour les paramètres $\beta$ et $\nu$ séparément. On observe bien une diminution des erreurs relatives lorsque le nombre d'observations augmente.

Table 2: Erreurs d'estimation dans le modèle joint pour les scénarios $N \in \{50, 100, 200, 300\}$ et $p = 200, J = 5$.

| | Errors | |
|---|---|---|
| Scenarios | $rrmse(\beta)$ | $rrmse(\nu)$ |
| $N = 50$ | 0.848 | 0.122 |
| $N = 100$ | 0.575 | 0.106 |
| $N = 200$ | 0.171 | 0.042 |
| $N = 300$ | 0.124 | 0.020 |

On souhaite également à l'avenir obtenir des résultats sur la capacité de la méthode à sélectionner les variables les plus pertinentes dans un grand ensemble de covariables. Pour cela, on souhaite présenter des résultats de sensibilité et spécificité de l'algorithme que l'on propose. Pour cela, on fixera le nombre d'individus $N = 100$ et l'on étudiera les scénarios où le nombre de covariables augmente par exemple $P \in \{50, 150, 400\}$. On calculera alors la proportion de vrai positif et faux positif.

# 5 Discussions

Dans ce travail, nous avons considéré un modèle joint en couplant un modèle de survie avec un modèle non linéaire à effets mixtes via une fonction de lien. Nous avons traité la sélection de variables et l'estimation de paramètres dans ce modèle.

Une première perspective à ce travail serait d'introduire de la grande dimension également dans le modèle non linéaire à effets mixtes. Une seconde perspective intéressante consisterait à aborder la prédiction dans le contexte de la modélisation jointe. Il serait ainsi intéressant de mettre en place une méthode de prédiction du temps de survie à partir d'un début d'observation de données longitudinales.

**Remerciements.** Ce travail a été financé par le projet (Stat4Plant) ANR-20-CE45-0012.

# References

[Achab, 2017] Achab, M. (2017). Learning from sequences with point processes. Issue: 2017SACLX068.

[Baey *et al.*, 2023] Baey, Delattre, Kuhn, Leger et Lemler (2023). Efficient precondtionned stochastic gradient descent fro estimation in latent variables models. *ICML*.

[Cox, 1972] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220. Publisher: [Royal Statistical Society, Wiley].

[Davidian et Giltinan, 1995] Davidian, M. et Giltinan, D. (1995). *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

[Delattre et Kuhn, 2023] Delattre, M. et Kuhn, E. (2023). Computing an empirical Fisher information matrix estimate in latent variable models through stochastic approximation. *Computo*.

[Fort *et al.*, 2017] Fort, G., Ollier, E. et Samson, A. (2017). Stochastic proximal gradient algorithms for penalized mixed models.

[He *et al.*, 2015] He, Z., Tu, W., Wang, S., Fu, H. et Yu, Z. (2015). Simultaneous variable selection for joint models of longitudinal and survival outcomes: Variable selection in joint models. *Biometrics*, 71(1):178–187.

[Kerioui *et al.*, 2022] Kerioui, M., Bertrand, J., Bruno, R., Mercier, F., Guedj, J. et Desmée, S. (2022). Modelling the association between biomarkers and clinical outcome: An introduction to nonlinear joint models. *British Journal of Clinical Pharmacology*, 88(4):1452–1463.

[Rizopoulos, 2012] Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. CRC press.

[Tibshirani, 1997] Tibshirani, R. (1997). The lasso method for variable selection in the cox model. *Statistics in Medicine*, 16(4):385–395.