HeKA Staff Meeting

Simon Bussy

simon.bussy@gmail.com

Lights: a generalized joint model for high-dimensional multivariate longitudinal data and censored durations

S. Bussy [1,2], V.T. Nguyen[1,2], A. Barbieri [3], S. Zohar [4], A.S. Jannot [4,5]

[1]LOPF, Califrais' Machine Learning Lab, Paris, France
[2]LPSM, UMR 8001, CNRS, Sorbonne University, Paris, France
[3]INSERM, UMR 1219, Bordeaux Population Health Research Center, Univ. Bordeaux, France
[4]INSERM, UMRS 1138, Centre de Recherche des Cordeliers, Paris, France
[5]Biomedical Informatics and Public Health Department, EGPH, APHP, Paris, France

I. Introduction

Overview

- Deal with the problem of joint modeling of longitudinal data and censored durations

Overview

- Deal with the problem of joint modeling of longitudinal data and censored durations
- Large number of both longitudinal and time-independent features are available

Overview

- Deal with the problem of joint modeling of longitudinal data and censored durations
- Large number of both longitudinal and time-independent features are available
- Flexibility in modeling the dependency between the longitudinal features and the event time with appropriate penalties

Overview

- Deal with the problem of joint modeling of longitudinal data and censored durations
- Large number of both longitudinal and time-independent features are available
- Flexibility in modeling the dependency between the longitudinal features and the event time with appropriate penalties
- Inference achieved using an efficient and novel Proximal Quasi-Newton Expectation Maximization algorithm

Use cases

- Predict the risk for an event of interest to occur quickly, taking into account simultaneously a huge number of longitudinal signals in a high-dimensional context

Use cases

- Predict the risk for an event of interest to occur quickly, taking into account simultaneously a huge number of longitudinal signals in a high-dimensional context
- Provides powerful interpretability by automatically pinpointing significant time-dependent and time-independent features

Use cases

- Predict the risk for an event of interest to occur quickly, taking into account simultaneously a huge number of longitudinal signals in a high-dimensional context
- Provides powerful interpretability by automatically pinpointing significant time-dependent and time-independent features

Real-time decision support

- Medical context $\rightarrow$ event of interest: survival time, re-hospitalization, relapse or disease progression ; longitudinal data: biomarkers or vital parameters measurements

Use cases

- Predict the risk for an event of interest to occur quickly, taking into account simultaneously a huge number of longitudinal signals in a high-dimensional context
- Provides powerful interpretability by automatically pinpointing significant time-dependent and time-independent features

Real-time decision support

- Medical context $\rightarrow$ event of interest: survival time, re-hospitalization, relapse or disease progression ; longitudinal data: biomarkers or vital parameters measurements
- Customer's satisfaction monitoring context $\rightarrow$ event of interest: time when a client churns ; longitudinal data: the client's activity recorded from account opening throughout the duration of the business relationship

High-dimensional framework

- Survival analysis

$$T = T^\star \wedge C \quad \text{and} \quad \Delta = \mathbb{1}_{\{T^\star \leq C\}}$$

High-dimensional framework

- Survival analysis

$$T = T^\star \wedge C \quad \text{and} \quad \Delta = \mathbb{1}_{\{T^\star \leq C\}}$$

- Time-independent features $X \in \mathbb{R}^p$ with $p \gg n$

High-dimensional framework

- Survival analysis

$$T = T^\star \wedge C \quad \text{and} \quad \Delta = \mathbb{1}_{\{T^\star \leq C\}}$$

- Time-independent features $X \in \mathbb{R}^p$ with $p \gg n$
- L longitudinal outcomes such that $L \gg n$ and

$$Y(t) = \big(Y^1(t), \ldots, Y^L(t)\big)^\top \in \mathbb{R}^L$$

High-dimensional framework

- Survival analysis

$$T = T^\star \wedge C \quad \text{and} \quad \Delta = \mathbb{1}_{\{T^\star \leq C\}}$$

- Time-independent features $X \in \mathbb{R}^p$ with $p \gg n$
- L longitudinal outcomes such that $L \gg n$ and

$$Y(t) = \big(Y^1(t), \dots, Y^L(t)\big)^\top \in \mathbb{R}^L$$

- Heterogeneity of the population: latent subgroups

$$G \in \{0, \dots, K-1\}$$

High-dimensional framework

- Survival analysis

$$T = T^\star \wedge C \quad \text{and} \quad \Delta = \mathbb{1}_{\{T^\star \leq C\}}$$

- Time-independent features $X \in \mathbb{R}^p$ with $p \gg n$
- L longitudinal outcomes such that $L \gg n$ and

$$Y(t) = \big(Y^1(t), \dots, Y^L(t)\big)^\top \in \mathbb{R}^L$$

- Heterogeneity of the population: latent subgroups

$$G \in \{0, \dots, K-1\}$$

- Softmax link function for the latent class membership probability given time-independent features

$$\pi_{\xi_k}(x) = \mathbb{P}[G = k | X = x] = \frac{e^{x^\top \xi_k}}{\sum_{k=0}^{K-1} e^{x^\top \xi_k}}$$

II. Model

Submodels

Group-specific marker trajectories

- $h_l\Big(\mathbb{E}[Y^l(t)|b^l, G = k]\Big) = m_k^l(t) = u^l(t)^\top \beta_k^l + v^l(t)^\top b^l$
  with fixed effect parameters $\beta_k^l \in \mathbb{R}^{q_l}$ and subject-and-longitudinal outcome specific random effects $b^l \in \mathbb{R}^{r_l} \sim \mathcal{N}(0, D_{ll})$

Submodels

Group-specific marker trajectories

- $h_l\big(\mathbb{E}[Y^l(t)|b^l, G=k]\big) = m_k^l(t) = u^l(t)^\top \beta_k^l + v^l(t)^\top b^l$
  with fixed effect parameters $\beta_k^l \in \mathbb{R}^{q_l}$ and subject-and-longitudinal outcome specific random effects $b^l \in \mathbb{R}^{r_l} \sim \mathcal{N}(0, D_{ll})$
- $\mathrm{Cov}[b^l, b^{l'}] = D_{ll'}$ and

$$D = \begin{bmatrix} D_{11} & \cdots & D_{1L} \\ \vdots & \ddots & \vdots \\ D_{1L}^\top & \cdots & D_{LL} \end{bmatrix}$$

  the global variance-covariance matrix

Submodels

Group-specific marker trajectories

- $h_l\big(\mathbb{E}[Y^l(t)|b^l, G=k]\big) = m_k^l(t) = u^l(t)^\top \beta_k^l + v^l(t)^\top b^l$ with fixed effect parameters $\beta_k^l \in \mathbb{R}^{q_l}$ and subject-and-longitudinal outcome specific random effects $b^l \in \mathbb{R}^{r_l} \sim \mathcal{N}(0, D_{ll})$
- $\mathrm{Cov}[b^l, b^{l'}] = D_{ll'}$

Group-specific risk of event

- $$\lambda(t|\mathcal{Y}(t), G=k) = \lambda_0(t) \exp\Bigg\{ \underbrace{\sum_{l=1}^{L} \sum_{a=1}^{\mathcal{A}} \gamma_{k,a}^l \Psi_a^l(t)}_{\gamma_k^\top \Psi(t)} \Bigg\}$$

  with $\mathcal{Y}(t) = \{Y(u), 0 \leq u < t\}$ and for each l-th longitudinal outcome, we consider $\mathcal{A} \in \mathbb{N}_+$ known functionals Ψ_a^l extracted from $\mathcal{Y}^l(t)$ through a given representation mapping

Submodels

Group-specific marker trajectories

- $h_l\Big(\mathbb{E}[Y^l(t)|b^l, G=k]\Big) = m_k^l(t) = u^l(t)^\top \beta_k^l + v^l(t)^\top b^l$
  with fixed effect parameters $\beta_k^l \in \mathbb{R}^{q_l}$ and subject-and-longitudinal outcome specific random effects $b^l \in \mathbb{R}^{r_l} \sim \mathcal{N}(0, D_{ll})$
- $\mathrm{Cov}[b^l, b^{l'}] = D_{ll'}$

Group-specific risk of event

- $\lambda(t|\mathcal{Y}(t), G=k) = \lambda_0(t)\exp\Big\{\underbrace{\sum_{l=1}^{L}\sum_{a=1}^{\mathcal{A}}\gamma_{k,a}^l \Psi_a^l(t)}_{\gamma_k^\top \Psi(t)}\Big\}$

  with $\mathcal{Y}(t) = \{Y(u), 0 \leq u < t\}$ and for each l-th longitudinal outcome, we consider $\mathcal{A} \in \mathbb{N}_+$ known functionals Ψ_a^l extracted from $\mathcal{Y}^l(t)$ through a given representation mapping
- $\Psi(t) \in \mathbb{R}^{L\mathcal{A}}$: high-dim representation features vector, highly flexible and independent of any modeling assumption !

Generalization of SREMs and JLCMs

- JLCMs [8] : homogeneous latent subgroups sharing the same marker trajectories and the same prognostic

Generalization of SREMs and JLCMs

- JLCMs [8] : homogeneous latent subgroups sharing the same marker trajectories and the same prognostic
- SREMs [9] : characteristics of the longitudinal processes (e.g. functions of the random effects) are included as features in the survival model

Generalization of SREMs and JLCMs

- JLCMs [8] : homogeneous latent subgroups sharing the same marker trajectories and the same prognostic
- SREMs [9] : characteristics of the longitudinal processes (e.g. functions of the random effects) are included as features in the survival model
- Usual functionals would be

Description	$\tilde{\Psi}_a^l(t, \beta_k^l, b^l)$	Reference
Linear predictor	$m_k^l(t)$	Chi and Ibrahim [2]
Random effects	b^l	Hatfield et al. [3]
Time-dependent slope	$\frac{\mathrm{d}}{\mathrm{d}t} m_k^l(t)$	Rizopoulos and Ghosh [6]
Cumulative effect	$\int_0^t m_k^l(s)\mathrm{d}s$	Andrinopoulou et al. [1]

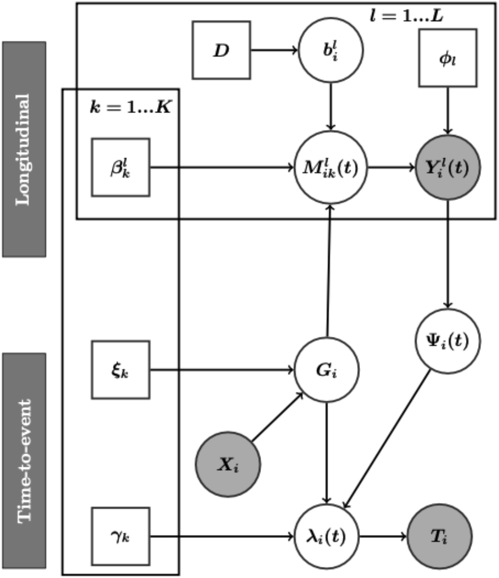
Generalization of SREMs and JLCMs

- JLCMs [8] : homogeneous latent subgroups sharing the same marker trajectories and the same prognostic
- SREMs [9] : characteristics of the longitudinal processes (e.g. functions of the random effects) are included as features in the survival model
- Usual functionals would be

Description	$\tilde{\Psi}_a^l(t, \beta_k^l, b^l)$	Reference
Linear predictor	$m_k^l(t)$	Chi and Ibrahim [2]
Random effects	b^l	Hatfield et al. [3]
Time-dependent slope	$\frac{\mathrm{d}}{\mathrm{d}t} m_k^l(t)$	Rizopoulos and Ghosh [6]
Cumulative effect	$\int_0^t m_k^l(s)\mathrm{d}s$	Andrinopoulou et al. [1]

- Depend on β_k (then on k) which leads to complicated updates, and on b which leads to untracktable integrals that requires approximation methods : computationally intensive and not scalable

Graphical model



Notations

- $\mathcal{D}_n = \big\{(x_1, y_1^1, \ldots, y_1^L, t_1, \delta_1), \ldots, (x_n, y_n^1, \ldots, y_n^L, t_n, \delta_n)\big\}$ with $y_i^l = (y_{i1}^l, \ldots, y_{in_i^l}^l)^\top \in \mathbb{R}^{n_i^l}$ and $y_{ij}^l = Y_i^l(t_{ij}^l)$

Notations

- $\mathcal{D}_n = \Big\{ (x_1, y_1^1, \ldots, y_1^L, t_1, \delta_1), \ldots, (x_n, y_n^1, \ldots, y_n^L, t_n, \delta_n) \Big\}$ with $y_i^l = (y_{i1}^l, \ldots, y_{in_i^l}^l)^\top \in \mathbb{R}^{n_i^l}$ and $y_{ij}^l = Y_i^l(t_{ij}^l)$
- $y_i = ({y_i^1}^\top \cdots {y_i^L}^\top)^\top \in \mathbb{R}^{n_i}$ with $n_i = \sum_{l=1}^L n_i^l$

Notations

- $\mathcal{D}_n = \Big\{(x_1, y_1^1, \dots, y_1^L, t_1, \delta_1), \dots, (x_n, y_n^1, \dots, y_n^L, t_n, \delta_n)\Big\}$ with $y_i^l = (y_{i1}^l, \dots, y_{in_i^l}^l)^\top \in \mathbb{R}^{n_i^l}$ and $y_{ij}^l = Y_i^l(t_{ij}^l)$
- $y_i = ({y_i^1}^\top \cdots {y_i^L}^\top)^\top \in \mathbb{R}^{n_i}$ with $n_i = \sum_{l=1}^L n_i^l$
- $b_i = ({b_i^1}^\top \cdots {b_i^L}^\top)^\top \in \mathbb{R}^{r}$ with $r = \sum_{l=1}^L r_l$

Notations

- $\mathcal{D}_n = \left\{(x_1, y_1^1, \ldots, y_1^L, t_1, \delta_1), \ldots, (x_n, y_n^1, \ldots, y_n^L, t_n, \delta_n)\right\}$ with $y_i^l = (y_{i1}^l, \ldots, y_{in_i^l}^l)^\top \in \mathbb{R}^{n_i^l}$ and $y_{ij}^l = Y_i^l(t_{ij}^l)$
- $y_i = ({y_i^1}^\top \cdots {y_i^L}^\top)^\top \in \mathbb{R}^{n_i}$ with $n_i = \sum_{l=1}^L n_i^l$
- $b_i = ({b_i^1}^\top \cdots {b_i^L}^\top)^\top \in \mathbb{R}^{r}$ with $r = \sum_{l=1}^L r_l$
- Design matrices

$$U_i = \begin{bmatrix} U_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & U_{iL} \end{bmatrix} \in \mathbb{R}^{n_i \times q} \text{ and } V_i = \begin{bmatrix} V_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V_{iL} \end{bmatrix} \in \mathbb{R}^{n_i \times r}$$

with $q = \sum_{l=1}^L q_l$ and where for all $l = 1, \ldots, L$, one writes

$$\begin{cases} U_{il} &= \left(u_i^l(t_{i1}^l)^\top \cdots u_i^l(t_{in_i^l}^l)^\top\right)^\top \in \mathbb{R}^{n_i^l \times q_l}, \\ V_{il} &= \left(v_i^l(t_{i1}^l)^\top \cdots v_i^l(t_{in_i^l}^l)^\top\right)^\top \in \mathbb{R}^{n_i^l \times r_l}. \end{cases}$$

Notations

- $\mathcal{D}_n = \left\{(x_1, y_1^1, \ldots, y_1^L, t_1, \delta_1), \ldots, (x_n, y_n^1, \ldots, y_n^L, t_n, \delta_n)\right\}$ with $y_i^l = (y_{i1}^l, \ldots, y_{in_i^l}^l)^\top \in \mathbb{R}^{n_i^l}$ and $y_{ij}^l = Y_i^l(t_{ij}^l)$
- $y_i = ({y_i^1}^\top \cdots {y_i^L}^\top)^\top \in \mathbb{R}^{n_i}$ with $n_i = \sum_{l=1}^L n_i^l$
- $b_i = ({b_i^1}^\top \cdots {b_i^L}^\top)^\top \in \mathbb{R}^r$ with $r = \sum_{l=1}^L r_l$
- Design matrices

$$U_i = \begin{bmatrix} U_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & U_{iL} \end{bmatrix} \in \mathbb{R}^{n_i \times q} \text{ and } V_i = \begin{bmatrix} V_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V_{iL} \end{bmatrix} \in \mathbb{R}^{n_i \times r}$$

with $q = \sum_{l=1}^L q_l$ and where for all $l = 1, \ldots, L$, one writes

$$\begin{cases} U_{il} &= \left(u_i^l(t_{i1}^l)^\top \cdots u_i^l(t_{in_i^l}^l)^\top\right)^\top \in \mathbb{R}^{n_i^l \times q_l}, \\ V_{il} &= \left(v_i^l(t_{i1}^l)^\top \cdots v_i^l(t_{in_i^l}^l)^\top\right)^\top \in \mathbb{R}^{n_i^l \times r_l}. \end{cases}$$

- $\beta_k = ({\beta_k^1}^\top \cdots {\beta_k^L}^\top)^\top \in \mathbb{R}^q$

Notations

- $\mathcal{D}_n = \big\{(x_1, y_1^1, \ldots, y_1^L, t_1, \delta_1), \ldots, (x_n, y_n^1, \ldots, y_n^L, t_n, \delta_n)\big\}$ with $y_i^l = (y_{i1}^l, \ldots, y_{in_i^l}^l)^\top \in \mathbb{R}^{n_i^l}$ and $y_{ij}^l = Y_i^l(t_{ij}^l)$
- $y_i = ({y_i^1}^\top \cdots {y_i^L}^\top)^\top \in \mathbb{R}^{n_i}$ with $n_i = \sum_{l=1}^L n_i^l$
- $b_i = ({b_i^1}^\top \cdots {b_i^L}^\top)^\top \in \mathbb{R}^{r}$ with $r = \sum_{l=1}^L r_l$
- Design matrices

$$U_i = \begin{bmatrix} U_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & U_{iL} \end{bmatrix} \in \mathbb{R}^{n_i \times q} \text{ and } V_i = \begin{bmatrix} V_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V_{iL} \end{bmatrix} \in \mathbb{R}^{n_i \times r}$$

with $q = \sum_{l=1}^L q_l$ and where for all $l = 1, \ldots, L$, one writes

$$\begin{cases} U_{il} &= \big(u_i^l(t_{i1}^l)^\top \cdots u_i^l(t_{in_i^l}^l)^\top\big)^\top \in \mathbb{R}^{n_i^l \times q_l}, \\ V_{il} &= \big(v_i^l(t_{i1}^l)^\top \cdots v_i^l(t_{in_i^l}^l)^\top\big)^\top \in \mathbb{R}^{n_i^l \times r_l}. \end{cases}$$

- $\beta_k = ({\beta_k^1}^\top \cdots {\beta_k^L}^\top)^\top \in \mathbb{R}^q$
- $M_{ik} = U_i\beta_k + V_i b_i \in \mathbb{R}^{n_i}$

Likelihood

- $\theta = \left(\xi_0^\top \cdots \xi_{K-1}^\top, \beta_0^\top \cdots \beta_{K-1}^\top, \phi^\top, \mathrm{vech}(D), \lambda_0^\top, \gamma_0^\top \cdots \gamma_{K-1}^\top\right) \in \mathbb{R}^\vartheta$

Likelihood

- $\theta = \left(\xi_0^\top \cdots \xi_{K-1}^\top, \beta_0^\top \cdots \beta_{K-1}^\top, \phi^\top, \text{vech}(D), \lambda_0^\top, \gamma_0^\top \cdots \gamma_{K-1}^\top\right) \in \mathbb{R}^\vartheta$
- $f(y_i|b_i, G_i = k) = \exp\left\{(y_i \odot \Phi_i)^\top M_{ik} - c_\phi(M_{ik}) + d_\phi(y_i)\right\}$ with $\Phi_i = (\phi_1^{-1}\mathbf{1}_{n_i^1}{}^\top \cdots \phi_L^{-1}\mathbf{1}_{n_i^L}{}^\top)^\top \in \mathbb{R}^{n_i}$

Likelihood

- $\theta = \left(\xi_0^\top \cdots \xi_{K-1}^\top, \beta_0^\top \cdots \beta_{K-1}^\top, \phi^\top, \mathrm{vech}(D), \lambda_0^\top, \gamma_0^\top \cdots \gamma_{K-1}^\top\right) \in \mathbb{R}^\vartheta$
- $f(y_i|b_i, G_i = k) = \exp\left\{(y_i \odot \Phi_i)^\top M_{ik} - c_\phi(M_{ik}) + d_\phi(y_i)\right\}$ with $\Phi_i = (\phi_1^{-1}\mathbf{1}_{n_i^1}{}^\top \cdots \phi_L^{-1}\mathbf{1}_{n_i^L}{}^\top)^\top \in \mathbb{R}^{n_i}$
- Survival part:

$$f(t_i, \delta_i | G_i = k; \theta) = \left[\lambda(t_i|\mathcal{Y}(t_i), G_i = k)\right]^{\delta_i} \times \exp\left\{-\int_0^{t_i} \lambda(s|\mathcal{Y}(s), G_i = k)\mathrm{d}s\right\}$$

Likelihood

- $\theta = \left(\xi_0^\top \cdots \xi_{K-1}^\top, \beta_0^\top \cdots \beta_{K-1}^\top, \phi^\top, \text{vech}(D), \lambda_0^\top, \gamma_0^\top \cdots \gamma_{K-1}^\top\right) \in \mathbb{R}^\vartheta$
- $f(y_i|b_i, G_i = k) = \exp\left\{(y_i \odot \Phi_i)^\top M_{ik} - c_\phi(M_{ik}) + d_\phi(y_i)\right\}$ with $\Phi_i = (\phi_1^{-1}\mathbf{1}_{n_i^1}{}^\top \cdots \phi_L^{-1}\mathbf{1}_{n_i^L}{}^\top)^\top \in \mathbb{R}^{n_i}$
- Survival part:

$$\begin{aligned} f(t_i, \delta_i | G_i = k; \theta) = &\Big[\lambda(t_i|\mathcal{Y}(t_i), G_i = k)\Big]^{\delta_i} \\ &\times \exp\left\{ - \int_0^{t_i} \lambda(s|\mathcal{Y}(s), G_i = k)\mathrm{d}s\right\} \end{aligned}$$

- Then, the log-likelihood (JLCMs type) writes

$$\ell_n(\theta) = n^{-1}\sum_{i=1}^n \log \sum_{k=0}^{K-1} \pi_{\xi_k}(x_i) f_\theta(t_i, \delta_i | G_i = k) f_\theta(y_i | G_i = k)\mathrm{d}b_i$$

III. Inference

Penalization

- Penalized objective

$$\ell_n^{\text{pen}}(\theta) = -\ell_n(\theta) + \sum_{k=0}^{K-1} \zeta_{1,k} \|\xi_k\|_{\text{en},\eta} + \zeta_{2,k} \|\gamma_k\|_{\text{sg}l_1,\tilde{\eta}}$$

with the elasticnet penalty

$$\|z\|_{\text{en},\eta} = (1-\eta)\|z\|_1 + \frac{\eta}{2}\|z\|_2^2$$

and the sparse group lasso penalty

$$\|z\|_{\text{sg}l_1,\tilde{\eta}} = (1-\tilde{\eta})\|z\|_1 + \tilde{\eta} \sum_{l=1}^{L} \|z^l\|_2$$

Penalization

- Penalized objective

$$\ell_n^{\text{pen}}(\theta) = -\ell_n(\theta) + \sum_{k=0}^{K-1} \zeta_{1,k} \|\xi_k\|_{\text{en},\eta} + \zeta_{2,k} \|\gamma_k\|_{\text{sg}l_1,\tilde{\eta}}$$

with the elasticnet penalty

$$\|z\|_{\text{en},\eta} = (1-\eta)\|z\|_1 + \frac{\eta}{2}\|z\|_2^2$$

and the sparse group lasso penalty

$$\|z\|_{\text{sg}l_1,\tilde{\eta}} = (1-\tilde{\eta})\|z\|_1 + \tilde{\eta} \sum_{l=1}^{L} \|z^l\|_2$$

- Resulting optimization problem

$$\hat{\theta} \in \text{argmin}_{\theta \in \mathbb{R}^\vartheta} \ \ell_n^{\text{pen}}(\theta)$$

prox-QNEM algorithm (1/2)

- $\ell_n^{\text{comp}}(\theta) = \ell_n^{\text{comp}}(\theta; \mathcal{D}_n, \boldsymbol{b}, \boldsymbol{G})$

prox-QNEM algorithm (1/2)

- $\ell_n^{\text{comp}}(\theta) = \ell_n^{\text{comp}}(\theta; \mathcal{D}_n, \boldsymbol{b}, \boldsymbol{G})$

Monte Carlo E-step

- $\mathcal{Q}_n(\theta, \theta^{(w)}) = \mathbb{E}_{\theta^{(w)}}[\ell_n^{\text{comp}}(\theta) | \mathcal{D}_n]$

prox-QNEM algorithm (1/2)

- $\ell_n^{\text{comp}}(\theta) = \ell_n^{\text{comp}}(\theta; \mathcal{D}_n, \boldsymbol{b}, \boldsymbol{G})$

Monte Carlo E-step

- $\mathcal{Q}_n(\theta, \theta^{(w)}) = \mathbb{E}_{\theta^{(w)}}[\ell_n^{\text{comp}}(\theta) | \mathcal{D}_n]$
- Requires to compute expectations of the form

$$\mathbb{E}_{\theta^{(w)}}[g(b_i, G_i) | t_i, \delta_i, y_i] = \sum_{k=0}^{K-1} \pi_{ik}^{\theta^{(w)}} \int_{\mathbb{R}^r} g(b_i, G_i) f_{\theta^{(w)}}(b_i | t_i, \delta_i, y_i) \mathrm{d} b_i$$

for different functions g, where we denote

$$\pi_{ik}^{\theta^{(w)}} = \mathbb{P}_{\theta^{(w)}}[G_i = k | t_i, \delta_i, y_i]$$

prox-QNEM algorithm (2/2)

Proximal Quasi-Newton M-step

- $\theta^{(w+1)} \in \text{argmin}_{\theta \in \mathbb{R}^{\vartheta}} \; \mathcal{Q}_n(\theta, \theta^{(w)}) + \sum_{k=0}^{K-1} \zeta_{1,k} \|\xi_k\|_{\text{en},\eta} + \zeta_{2,k} \|\gamma_k\|_{\text{sg}l_1, \tilde{\eta}}$

prox-QNEM algorithm (2/2)

Proximal Quasi-Newton M-step

- $\theta^{(w+1)} \in \text{argmin}_{\theta \in \mathbb{R}^\vartheta} \; \mathcal{Q}_n(\theta, \theta^{(w)}) + \sum_{k=0}^{K-1} \zeta_{1,k} \|\xi_k\|_{\text{en},\eta} + \zeta_{2,k} \|\gamma_k\|_{\text{sg}l_1,\tilde{\eta}}$
- $D^{(w+1)} = n^{-1} \sum_{i=1}^{n} \hat{\mathbb{E}}_{\theta^{(w)}}[b_i b_i^\top | t_i, \delta_i, y_i]$

prox-QNEM algorithm (2/2)

Proximal Quasi-Newton M-step

- $\theta^{(w+1)} \in \text{argmin}_{\theta \in \mathbb{R}^\vartheta} \; \mathcal{Q}_n(\theta, \theta^{(w)}) + \sum_{k=0}^{K-1} \zeta_{1,k} \|\xi_k\|_{\text{en},\eta} + \zeta_{2,k} \|\gamma_k\|_{\text{sg}l_1,\tilde{\eta}}$
- $D^{(w+1)} = n^{-1} \sum_{i=1}^{n} \hat{\mathbb{E}}_{\theta^{(w)}}[b_i b_i^\top | t_i, \delta_i, y_i]$
- $P_{n,k}^{(w)}(\xi_k) = -n^{-1} \sum_{i=1}^{n} \hat{\pi}_{ik}^{\theta^{(w)}} \log \pi_{\xi_k}(x_i)$

prox-QNEM algorithm (2/2)

Proximal Quasi-Newton M-step

- $\theta^{(w+1)} \in \text{argmin}_{\theta \in \mathbb{R}^{\vartheta}} \mathcal{Q}_n(\theta, \theta^{(w)}) + \sum_{k=0}^{K-1} \zeta_{1,k} \|\xi_k\|_{\text{en},\eta} + \zeta_{2,k} \|\gamma_k\|_{\text{sg}l_1,\tilde{\eta}}$
- $D^{(w+1)} = n^{-1} \sum_{i=1}^{n} \hat{\mathbb{E}}_{\theta^{(w)}}[b_i b_i^\top | t_i, \delta_i, y_i]$
- $P_{n,k}^{(w)}(\xi_k) = -n^{-1} \sum_{i=1}^{n} \hat{\pi}_{ik}^{\theta^{(w)}} \log \pi_{\xi_k}(x_i)$
- $\xi_k^{(w+1)} \in \text{argmin}_{\xi_k \in \mathbb{R}^p} P_{n,k}^{(w)}(\xi_k) + \zeta_{1,k} \|\xi_k\|_{\text{en},\eta}$

prox-QNEM algorithm (2/2)

Proximal Quasi-Newton M-step

- $\theta^{(w+1)} \in \text{argmin}_{\theta \in \mathbb{R}^\vartheta} \ \mathcal{Q}_n(\theta, \theta^{(w)}) + \sum_{k=0}^{K-1} \zeta_{1,k} \|\xi_k\|_{\text{en},\eta} + \zeta_{2,k} \|\gamma_k\|_{\text{sg}l_1,\tilde{\eta}}$
- $D^{(w+1)} = n^{-1} \sum_{i=1}^n \hat{\mathbb{E}}_{\theta^{(w)}}[b_i b_i^\top | t_i, \delta_i, y_i]$
- $P_{n,k}^{(w)}(\xi_k) = -n^{-1} \sum_{i=1}^n \hat{\pi}_{ik}^{\theta^{(w)}} \log \pi_{\xi_k}(x_i)$
- $\xi_k^{(w+1)} \in \text{argmin}_{\xi_k \in \mathbb{R}^p} P_{n,k}^{(w)}(\xi_k) + \zeta_{1,k} \|\xi_k\|_{\text{en},\eta}$
- L-BFGS-B to solve the problem

$$
\begin{aligned}
\text{minimize} \quad & P_{n,k}^{(w)}(\xi_k^+ - \xi_k^-) + \zeta_{1,k}\Big((1-\eta)\sum_{j=1}^{p}(\xi_{k,j}^+ + \xi_{k,j}^-) + \frac{\eta}{2}\|\xi_k{}^+ - \xi_k{}^-\|_2^2\Big) \\
\text{subject to} \quad & \xi_{k,j}^+ \geq 0 \text{ and } \xi_{k,j}^- \geq 0 \text{ for } j = 1, \ldots, p
\end{aligned}
$$

prox-QNEM algorithm (2/2)

Proximal Quasi-Newton M-step

- $\theta^{(w+1)} \in \text{argmin}_{\theta \in \mathbb{R}^{\vartheta}} \; \mathcal{Q}_n(\theta, \theta^{(w)}) + \sum_{k=0}^{K-1} \zeta_{1,k} \|\xi_k\|_{\text{en},\eta} + \zeta_{2,k} \|\gamma_k\|_{\text{sg}l_1,\tilde{\eta}}$
- $D^{(w+1)} = n^{-1} \sum_{i=1}^{n} \hat{\mathbb{E}}_{\theta^{(w)}}[b_i b_i^\top | t_i, \delta_i, y_i]$
- $P_{n,k}^{(w)}(\xi_k) = -n^{-1} \sum_{i=1}^{n} \hat{\pi}_{ik}^{\theta^{(w)}} \log \pi_{\xi_k}(x_i)$
- $\xi_k^{(w+1)} \in \text{argmin}_{\xi_k \in \mathbb{R}^p} P_{n,k}^{(w)}(\xi_k) + \zeta_{1,k} \|\xi_k\|_{\text{en},\eta}$
- L-BFGS-B to solve the problem

$$\begin{aligned} &\text{minimize} \quad && P_{n,k}^{(w)}(\xi_k^+ - \xi_k^-) + \zeta_{1,k}\Big((1-\eta)\sum_{j=1}^{p}(\xi_{k,j}^+ + \xi_{k,j}^-) + \frac{\eta}{2}\|\xi_k{}^+ - \xi_k{}^-\|_2^2\Big) \\ &\text{subject to} \quad && \xi_{k,j}^+ \geq 0 \text{ and } \xi_{k,j}^- \geq 0 \text{ for } j = 1, \dots, p \end{aligned}$$

- Closed-form update for $\beta_k^{(w+1)}$

prox-QNEM algorithm (2/2)

Proximal Quasi-Newton M-step

- $\theta^{(w+1)} \in \text{argmin}_{\theta \in \mathbb{R}^\vartheta}\ \mathcal{Q}_n(\theta, \theta^{(w)}) + \sum_{k=0}^{K-1} \zeta_{1,k} \|\xi_k\|_{\text{en},\eta} + \zeta_{2,k} \|\gamma_k\|_{\text{sg}l_1,\tilde\eta}$
- $D^{(w+1)} = n^{-1} \sum_{i=1}^n \hat{\mathbb{E}}_{\theta^{(w)}}[b_i b_i^\top | t_i, \delta_i, y_i]$
- $P_{n,k}^{(w)}(\xi_k) = -n^{-1} \sum_{i=1}^n \hat\pi_{ik}^{\theta^{(w)}} \log \pi_{\xi_k}(x_i)$
- $\xi_k^{(w+1)} \in \text{argmin}_{\xi_k \in \mathbb{R}^p} P_{n,k}^{(w)}(\xi_k) + \zeta_{1,k}\|\xi_k\|_{\text{en},\eta}$
- L-BFGS-B to solve the problem

$$\begin{aligned}
&\text{minimize} && P_{n,k}^{(w)}(\xi_k^+ - \xi_k^-) + \zeta_{1,k}\Big((1-\eta)\sum_{j=1}^p (\xi_{k,j}^+ + \xi_{k,j}^-) + \frac{\eta}{2}\|\xi_k{}^+ - \xi_k{}^-\|_2^2\Big) \\
&\text{subject to} && \xi_{k,j}^+ \geq 0 \text{ and } \xi_{k,j}^- \geq 0 \text{ for } j = 1, \ldots, p
\end{aligned}$$

- Closed-form update for $\beta_k^{(w+1)}$
- Proximal gradient (ISTA) for the $\gamma_k^{(w+1)}$ update, based on Lemma 1 that states $\text{prox}_{\text{sg}l_1,\tilde\eta,\zeta} = \text{prox}_{\zeta\tilde\eta \sum_{l=1}^L \|\cdot\|_{2,l}} \circ \text{prox}_{\zeta(1-\tilde\eta)\|\cdot\|_1}$

prox-QNEM algorithm (2/2)

Proximal Quasi-Newton M-step

- $\theta^{(w+1)} \in \text{argmin}_{\theta \in \mathbb{R}^\vartheta} \ \mathcal{Q}_n(\theta, \theta^{(w)}) + \sum_{k=0}^{K-1} \zeta_{1,k} \|\xi_k\|_{\text{en},\eta} + \zeta_{2,k} \|\gamma_k\|_{\text{sg}l_1,\tilde{\eta}}$
- $D^{(w+1)} = n^{-1} \sum_{i=1}^n \hat{\mathbb{E}}_{\theta^{(w)}}[b_i b_i^\top | t_i, \delta_i, y_i]$
- $P_{n,k}^{(w)}(\xi_k) = -n^{-1} \sum_{i=1}^n \hat{\pi}_{ik}^{\theta^{(w)}} \log \pi_{\xi_k}(x_i)$
- $\xi_k^{(w+1)} \in \text{argmin}_{\xi_k \in \mathbb{R}^p} P_{n,k}^{(w)}(\xi_k) + \zeta_{1,k} \|\xi_k\|_{\text{en},\eta}$
- L-BFGS-B to solve the problem

$$\begin{aligned} &\text{minimize} \quad && P_{n,k}^{(w)}(\xi_k^+ - \xi_k^-) + \zeta_{1,k}\Big((1-\eta)\sum_{j=1}^p (\xi_{k,j}^+ + \xi_{k,j}^-) + \frac{\eta}{2}\|\xi_k{}^+ - \xi_k{}^-\|_2^2\Big) \\ &\text{subject to} \quad && \xi_{k,j}^+ \geq 0 \text{ and } \xi_{k,j}^- \geq 0 \text{ for } j = 1, \ldots, p \end{aligned}$$

- Closed-form update for $\beta_k^{(w+1)}$
- Proximal gradient (ISTA) for the $\gamma_k^{(w+1)}$ update, based on Lemma 1 that states $\text{prox}_{\text{sg}l_1,\tilde{\eta},\zeta} = \text{prox}_{\zeta\tilde{\eta}\sum_{l=1}^L \|\cdot\|_{2,l}} \circ \text{prox}_{\zeta(1-\tilde{\eta})\|\cdot\|_1}$
- $\lambda_0^{(w+1)}(t) = \dfrac{\sum_{i=1}^n \delta_i \mathbb{1}_{\{t=t_i\}}}{\sum_{i=1}^n \sum_{k=0}^{K-1} \hat{\pi}_{ik}^{\theta^{(w)}} \exp\Big\{\gamma_k^{(w+1)\top} \Psi_i(t_i)\Big\} \mathbb{1}_{\{t_i \geq t\}}}$

Performance evaluation

Real-time prediction paradigm

- Practitioners need predictive prognostic tools to be used in real-time

Performance evaluation

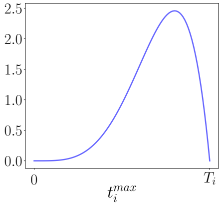
Real-time prediction paradigm

- Practitioners need predictive prognostic tools to be used in real-time
- Obtain $\hat{\theta}$ on training data

Performance evaluation

Real-time prediction paradigm

- Practitioners need predictive prognostic tools to be used in real-time
- Obtain $\hat{\theta}$ on training data
- On test data : sample $t_i^{max} \sim T_i\Big(1 - \text{Beta}(2,5)\Big)$ the time for subject i when one wants to perform the risk prediction – the "present" time



Performance evaluation

Real-time prediction paradigm

- Practitioners need predictive prognostic tools to be used in real-time
- Obtain $\hat{\theta}$ on training data
- On test data : sample $t_i^{max} \sim T_i\Big(1 - \text{Beta}(2,5)\Big)$ the time for subject i when one wants to perform the risk prediction – the "present" time
- Truncate y_i up to t_i^{max} to get $\tilde{y}_i$

Performance evaluation

Real-time prediction paradigm

- Practitioners need predictive prognostic tools to be used in real-time
- Obtain $\hat{\theta}$ on training data
- On test data : sample $t_i^{max} \sim T_i\Big(1 - \text{Beta}(2,5)\Big)$ the time for subject i when one wants to perform the risk prediction – the "present" time
- Truncate y_i up to t_i^{max} to get $\tilde{y}_i$

Predictive marker

- $\mathbb{P}_{\hat{\theta}}[G_i = k | T_i > t_i^{max}, \tilde{y}_i] = \dfrac{\pi_{\hat{\xi}_k}(x_i) f_{\hat{\theta}}(t_i^{max}, \delta_i = 0, \tilde{y}_i | G_i = k)}{\sum_{k=0}^{K-1} \pi_{\hat{\xi}_k}(x_i) f_{\hat{\theta}}(t_i^{max}, \delta_i = 0, \tilde{y}_i | G_i = k)}$

Performance evaluation

Real-time prediction paradigm

- Practitioners need predictive prognostic tools to be used in real-time
- Obtain $\hat{\theta}$ on training data
- On test data : sample $t_i^{max} \sim T_i\Big(1 - \text{Beta}(2,5)\Big)$ the time for subject i when one wants to perform the risk prediction – the "present" time
- Truncate y_i up to t_i^{max} to get $\tilde{y}_i$

Predictive marker

- $\mathbb{P}_{\hat{\theta}}[G_i = k | T_i > t_i^{max}, \tilde{y}_i] = \dfrac{\pi_{\hat{\xi}_k}(x_i) f_{\hat{\theta}}(t_i^{max}, \delta_i = 0, \tilde{y}_i | G_i = k)}{\sum_{k=0}^{K-1} \pi_{\hat{\xi}_k}(x_i) f_{\hat{\theta}}(t_i^{max}, \delta_i = 0, \tilde{y}_i | G_i = k)}$
- In practice $K = 2$

Performance evaluation

Real-time prediction paradigm

- Practitioners need predictive prognostic tools to be used in real-time
- Obtain $\hat{\theta}$ on training data
- On test data : sample $t_i^{max} \sim T_i\Big(1 - \text{Beta}(2,5)\Big)$ the time for subject i when one wants to perform the risk prediction – the "present" time
- Truncate y_i up to t_i^{max} to get $\tilde{y}_i$

Predictive marker

- $\mathbb{P}_{\hat{\theta}}[G_i = k | T_i > t_i^{max}, \tilde{y}_i] = \dfrac{\pi_{\hat{\xi}_k}(x_i) f_{\hat{\theta}}(t_i^{max}, \delta_i = 0, \tilde{y}_i | G_i = k)}{\sum_{k=0}^{K-1} \pi_{\hat{\xi}_k}(x_i) f_{\hat{\theta}}(t_i^{max}, \delta_i = 0, \tilde{y}_i | G_i = k)}$
- In practice $K = 2$
- C-index metric $\mathcal{C}_\tau = \mathbb{P}[\hat{\mathcal{R}}_i > \hat{\mathcal{R}}_j | T_i < T_j, T_i < \tau]$

Competing models

- Landmark Cox model : `survival` R package [10]

Competing models

- Landmark Cox model : `survival` R package [10]
- The time-dependent Cox model [7] : `survival` R package

Competing models

- Landmark Cox model : `survival` R package [10]
- The time-dependent Cox model [7] : `survival` R package
- Multivariate joint latent class model : R package `lcmm` [4]

Competing models

- Landmark Cox model : `survival` `R` package [10]
- The time-dependent Cox model [7] : `survival` `R` package
- Multivariate joint latent class model : `R` package `lcmm` [4]
- Multivariate shared random effect model : `JMbayes` package [5]

High-dimensional simulation study (1/2)

- High-risk subjects proportion $\pi_1 \in [0, 1]$

High-dimensional simulation study (1/2)

- High-risk subjects proportion $\pi_1 \in [0, 1]$
- $\mathcal{H} = \big\{ \lfloor \pi_1 n \rfloor$ random samples without replacement$\big\}$

High-dimensional simulation study (1/2)

- High-risk subjects proportion $\pi_1 \in [0, 1]$
- $\mathcal{H} = \big\{ \lfloor \pi_1 n \rfloor \text{ random samples without replacement} \big\}$
- $[x_{ij}] \in \mathbb{R}^{n \times p} \sim \mathcal{N}\big(0, \Sigma_1(\rho_1)\big)$

High-dimensional simulation study (1/2)

- High-risk subjects proportion $\pi_1 \in [0, 1]$
- $\mathcal{H} = \big\{ \lfloor \pi_1 n \rfloor$ random samples without replacement$\big\}$
- $[x_{ij}] \in \mathbb{R}^{n \times p} \sim \mathcal{N}\big(0, \Sigma_1(\rho_1)\big)$
- $x_{ij} \leftarrow x_{ij} + (-1)^{\mathbb{1}_{\{i \notin \mathcal{H}\}}} gap$ for $j = 1, \ldots, s$

High-dimensional simulation study (1/2)

- High-risk subjects proportion $\pi_1 \in [0,1]$
- $\mathcal{H} = \big\{ \lfloor \pi_1 n \rfloor \text{ random samples without replacement} \big\}$
- $[x_{ij}] \in \mathbb{R}^{n \times p} \sim \mathcal{N}\big(0, \Sigma_1(\rho_1)\big)$
- $x_{ij} \leftarrow x_{ij} + (-1)^{\mathbb{1}_{\{i \notin \mathcal{H}\}}} gap$ for $j = 1, \ldots, s$
- $\xi = (\underbrace{\varsigma_1, \ldots, \varsigma_1}_{s}, 0, \ldots, 0) \in \mathbb{R}^p$

High-dimensional simulation study (1/2)

- High-risk subjects proportion $\pi_1 \in [0, 1]$
- $\mathcal{H} = \big\{ \lfloor \pi_1 n \rfloor \text{ random samples without replacement} \big\}$
- $[x_{ij}] \in \mathbb{R}^{n \times p} \sim \mathcal{N}\big(0, \Sigma_1(\rho_1)\big)$
- $x_{ij} \leftarrow x_{ij} + (-1)^{\mathbb{1}_{\{i \notin \mathcal{H}\}}} gap$ for $j = 1, \ldots, s$
- $\xi = (\underbrace{\varsigma_1, \ldots, \varsigma_1}_{s}, 0, \ldots, 0) \in \mathbb{R}^p$
- $G_i \sim \mathcal{B}\big(\pi_\xi(x_i)\big)$

High-dimensional simulation study (1/2)

- High-risk subjects proportion $\pi_1 \in [0, 1]$
- $\mathcal{H} = \big\{ \lfloor \pi_1 n \rfloor$ random samples without replacement$\big\}$
- $[x_{ij}] \in \mathbb{R}^{n \times p} \sim \mathcal{N}\big(0, \Sigma_1(\rho_1)\big)$
- $x_{ij} \leftarrow x_{ij} + (-1)^{\mathbb{1}_{\{i \notin \mathcal{H}\}}} gap$ for $j = 1, \ldots, s$
- $\xi = (\underbrace{\varsigma_1, \ldots, \varsigma_1}_{s}, 0, \ldots, 0) \in \mathbb{R}^p$
- $G_i \sim \mathcal{B}\big(\pi_\xi(x_i)\big)$
- $\mathcal{S}_k = \Big\{ k \big\lfloor \frac{Lr_s}{K} \big\rfloor + 1, \ldots, (k+1) \big\lfloor \frac{Lr_s}{K} \big\rfloor \Big\}$

High-dimensional simulation study (1/2)

- High-risk subjects proportion $\pi_1 \in [0, 1]$
- $\mathcal{H} = \{\lfloor \pi_1 n \rfloor$ random samples without replacement$\}$
- $[x_{ij}] \in \mathbb{R}^{n \times p} \sim \mathcal{N}\big(0, \Sigma_1(\rho_1)\big)$
- $x_{ij} \leftarrow x_{ij} + (-1)^{\mathbb{1}_{\{i \notin \mathcal{H}\}}} gap$ for $j = 1, \ldots, s$
- $\xi = (\underbrace{\varsigma_1, \ldots, \varsigma_1}_{s}, 0, \ldots, 0) \in \mathbb{R}^p$
- $G_i \sim \mathcal{B}\big(\pi_\xi(x_i)\big)$
- $\mathcal{S}_k = \Big\{ k \big\lfloor \frac{L r_s}{K} \big\rfloor + 1, \ldots, (k+1) \big\lfloor \frac{L r_s}{K} \big\rfloor \Big\}$
- $\beta_k^l \sim \mathbb{1}_{\{l \in \mathcal{S}_k\}} \mathcal{N}\left(\mu_k, \begin{bmatrix} \rho_3 & 0 \\ 0 & \rho_3 \end{bmatrix}\right)$

High-dimensional simulation study (1/2)

- High-risk subjects proportion $\pi_1 \in [0, 1]$
- $\mathcal{H} = \big\{ \lfloor \pi_1 n \rfloor \text{ random samples without replacement} \big\}$
- $[x_{ij}] \in \mathbb{R}^{n \times p} \sim \mathcal{N}\big(0, \Sigma_1(\rho_1)\big)$
- $x_{ij} \leftarrow x_{ij} + (-1)^{\mathbb{1}_{\{i \notin \mathcal{H}\}}} gap$ for $j = 1, \ldots, s$
- $\xi = (\underbrace{\varsigma_1, \ldots, \varsigma_1}_{s}, 0, \ldots, 0) \in \mathbb{R}^p$
- $G_i \sim \mathcal{B}\big(\pi_\xi(x_i)\big)$
- $\mathcal{S}_k = \Big\{ k \Big\lfloor \dfrac{Lr_s}{K} \Big\rfloor + 1, \ldots, (k+1) \Big\lfloor \dfrac{Lr_s}{K} \Big\rfloor \Big\}$
- $\beta_k^l \sim \mathbb{1}_{\{l \in \mathcal{S}_k\}} \mathcal{N}\left(\mu_k, \begin{bmatrix} \rho_3 & 0 \\ 0 & \rho_3 \end{bmatrix}\right)$
- $Y_i^l(t) = \epsilon_i^l(t) + \mathbb{1}_{\{l \in \mathcal{S}_k\}} \sum_{k=0}^{K-1} \mathbb{1}_{\{G_i = k\}} \big((1, t)^\top \beta_k^l + (1, t)^\top b_i^l\big)$

High-dimensional simulation study (1/2)

- High-risk subjects proportion $\pi_1 \in [0, 1]$
- $\mathcal{H} = \big\{ \lfloor \pi_1 n \rfloor$ random samples without replacement$\big\}$
- $[x_{ij}] \in \mathbb{R}^{n \times p} \sim \mathcal{N}\big(0, \Sigma_1(\rho_1)\big)$
- $x_{ij} \leftarrow x_{ij} + (-1)^{\mathbb{1}_{\{i \notin \mathcal{H}\}}} gap$ for $j = 1, \ldots, s$
- $\xi = (\underbrace{\varsigma_1, \ldots, \varsigma_1}_{s}, 0, \ldots, 0) \in \mathbb{R}^p$
- $G_i \sim \mathcal{B}\big(\pi_\xi(x_i)\big)$
- $\mathcal{S}_k = \Big\{ k \big\lfloor \dfrac{Lr_s}{K} \big\rfloor + 1, \ldots, (k+1) \big\lfloor \dfrac{Lr_s}{K} \big\rfloor \Big\}$
- $\beta_k^l \sim \mathbb{1}_{\{l \in \mathcal{S}_k\}} \mathcal{N}\Bigg(\mu_k, \begin{bmatrix} \rho_3 & 0 \\ 0 & \rho_3 \end{bmatrix}\Bigg)$
- $Y_i^l(t) = \epsilon_i^l(t) + \mathbb{1}_{\{l \in \mathcal{S}_k\}} \sum_{k=0}^{K-1} \mathbb{1}_{\{G_i = k\}} \big((1, t)^\top \beta_k^l + (1, t)^\top b_i^l\big)$
- $\big(\tilde{\Psi}_{k,a}(t, \beta_k, b_i^l)\big)_{a \in \{1,2\}} = \big(\beta_{k,1}^l + \beta_{k,2}^l t + b_{i,1}^l + b_{i,2}^l t\ , {b_i^l}^\top\big)$

High-dimensional simulation study (1/2)

- High-risk subjects proportion $\pi_1 \in [0, 1]$
- $\mathcal{H} = \big\{ \lfloor \pi_1 n \rfloor$ random samples without replacement$\big\}$
- $[x_{ij}] \in \mathbb{R}^{n \times p} \sim \mathcal{N}\big(0, \Sigma_1(\rho_1)\big)$
- $x_{ij} \leftarrow x_{ij} + (-1)^{\mathbb{1}_{\{i \notin \mathcal{H}\}}} gap$ for $j = 1, \ldots, s$
- $\xi = (\underbrace{\varsigma_1, \ldots, \varsigma_1}_{s}, 0, \ldots, 0) \in \mathbb{R}^p$
- $G_i \sim \mathcal{B}\big(\pi_\xi(x_i)\big)$
- $\mathcal{S}_k = \Big\{ k \lfloor \dfrac{Lr_s}{K} \rfloor + 1, \ldots, (k+1) \lfloor \dfrac{Lr_s}{K} \rfloor \Big\}$
- $\beta_k^l \sim \mathbb{1}_{\{l \in \mathcal{S}_k\}} \mathcal{N}\left(\mu_k, \begin{bmatrix} \rho_3 & 0 \\ 0 & \rho_3 \end{bmatrix} \right)$
- $Y_i^l(t) = \epsilon_i^l(t) + \mathbb{1}_{\{l \in \mathcal{S}_k\}} \sum_{k=0}^{K-1} \mathbb{1}_{\{G_i = k\}} \big((1, t)^\top \beta_k^l + (1, t)^\top b_i^l \big)$
- $\big(\tilde{\Psi}_{k,a}(t, \beta_k, b_i^l) \big)_{a \in \{1,2\}} = \big(\beta_{k,1}^l + \beta_{k,2}^l t + b_{i,1}^l + b_{i,2}^l t \ , {b_i^l}^\top \big)$
- $\tilde{\gamma}_{k,a}^l = \varsigma_2 \mathbb{1}_{\{l \in \mathcal{S}_k, \ a \in \{1,2\}\}}$

High-dimensional simulation study (1/2)

- High-risk subjects proportion $\pi_1 \in [0,1]$
- $\mathcal{H} = \big\{ \lfloor \pi_1 n \rfloor \text{ random samples without replacement} \big\}$
- $[x_{ij}] \in \mathbb{R}^{n \times p} \sim \mathcal{N}\big(0, \Sigma_1(\rho_1)\big)$
- $x_{ij} \leftarrow x_{ij} + (-1)^{\mathbb{1}_{\{i \notin \mathcal{H}\}}} gap$ for $j = 1, \ldots, s$
- $\xi = (\underbrace{\varsigma_1, \ldots, \varsigma_1}_{s}, 0, \ldots, 0) \in \mathbb{R}^p$
- $G_i \sim \mathcal{B}\big(\pi_\xi(x_i)\big)$
- $\mathcal{S}_k = \Big\{ k \big\lfloor \frac{L r_s}{K} \big\rfloor + 1, \ldots, (k+1) \big\lfloor \frac{L r_s}{K} \big\rfloor \Big\}$
- $\beta_k^l \sim \mathbb{1}_{\{l \in \mathcal{S}_k\}} \mathcal{N}\Big(\mu_k, \begin{bmatrix} \rho_3 & 0 \\ 0 & \rho_3 \end{bmatrix}\Big)$
- $Y_i^l(t) = \epsilon_i^l(t) + \mathbb{1}_{\{l \in \mathcal{S}_k\}} \sum_{k=0}^{K-1} \mathbb{1}_{\{G_i = k\}} \big((1,t)^\top \beta_k^l + (1,t)^\top b_i^l\big)$
- $\big(\tilde{\Psi}_{k,a}(t, \beta_k, b_i^l)\big)_{a \in \{1,2\}} = \big(\beta_{k,1}^l + \beta_{k,2}^l t + b_{i,1}^l + b_{i,2}^l t \ , {b_i^l}^\top\big)$
- $\tilde{\gamma}_{k,a}^l = \varsigma_2 \mathbb{1}_{\{l \in \mathcal{S}_k, \ a \in \{1,2\}\}}$
- $\lambda_i(t | G_i = k) = \lambda_0(t) \exp\big\{ \iota_{i,k,1} + \iota_{i,k,2} t \big\}$

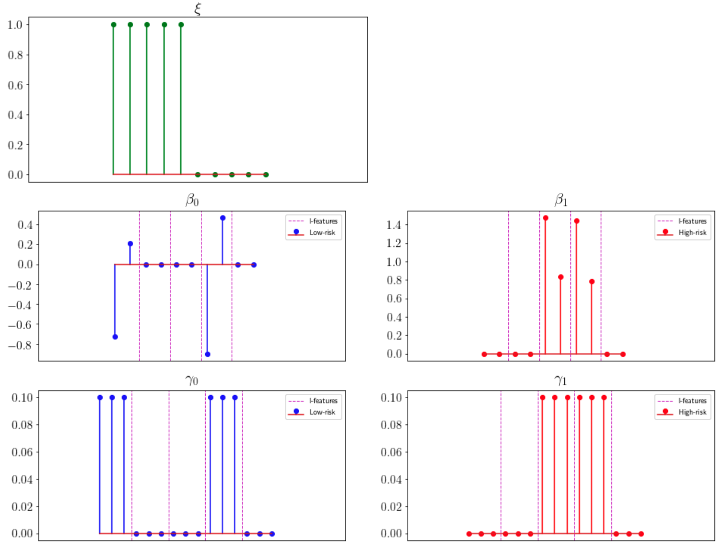
High-dimensional simulation study (1/2)

- High-risk subjects proportion $\pi_1 \in [0,1]$
- $\mathcal{H} = \big\{ \lfloor \pi_1 n \rfloor \text{ random samples without replacement} \big\}$
- $[x_{ij}] \in \mathbb{R}^{n \times p} \sim \mathcal{N}\big(0, \Sigma_1(\rho_1)\big)$
- $x_{ij} \leftarrow x_{ij} + (-1)^{\mathbb{1}_{\{i \notin \mathcal{H}\}}} gap$ for $j = 1, \ldots, s$
- $\xi = (\underbrace{\varsigma_1, \ldots, \varsigma_1}_{s}, 0, \ldots, 0) \in \mathbb{R}^p$
- $G_i \sim \mathcal{B}\big(\pi_\xi(x_i)\big)$
- $\mathcal{S}_k = \Big\{ k \big\lfloor \dfrac{Lr_s}{K} \big\rfloor + 1, \ldots, (k+1) \big\lfloor \dfrac{Lr_s}{K} \big\rfloor \Big\}$
- $\beta_k^l \sim \mathbb{1}_{\{l \in \mathcal{S}_k\}} \mathcal{N}\left(\mu_k, \begin{bmatrix} \rho_3 & 0 \\ 0 & \rho_3 \end{bmatrix} \right)$
- $Y_i^l(t) = \epsilon_i^l(t) + \mathbb{1}_{\{l \in \mathcal{S}_k\}} \sum_{k=0}^{K-1} \mathbb{1}_{\{G_i = k\}} \big((1,t)^\top \beta_k^l + (1,t)^\top b_i^l \big)$
- $\big(\tilde{\Psi}_{k,a}(t, \beta_k, b_i^l) \big)_{a \in \{1,2\}} = \big(\beta_{k,1}^l + \beta_{k,2}^l t + b_{i,1}^l + b_{i,2}^l t \ , b_i^l{}^\top \big)$
- $\tilde{\gamma}_{k,a}^l = \varsigma_2 \mathbb{1}_{\{l \in \mathcal{S}_k, \ a \in \{1,2\}\}}$
- $\lambda_i(t | G_i = k) = \lambda_0(t) \exp\big\{ \iota_{i,k,1} + \iota_{i,k,2} t \big\}$
- Gompertz baseline $\lambda_0(t) = \kappa_1 \kappa_2 \exp(\kappa_2 t)$

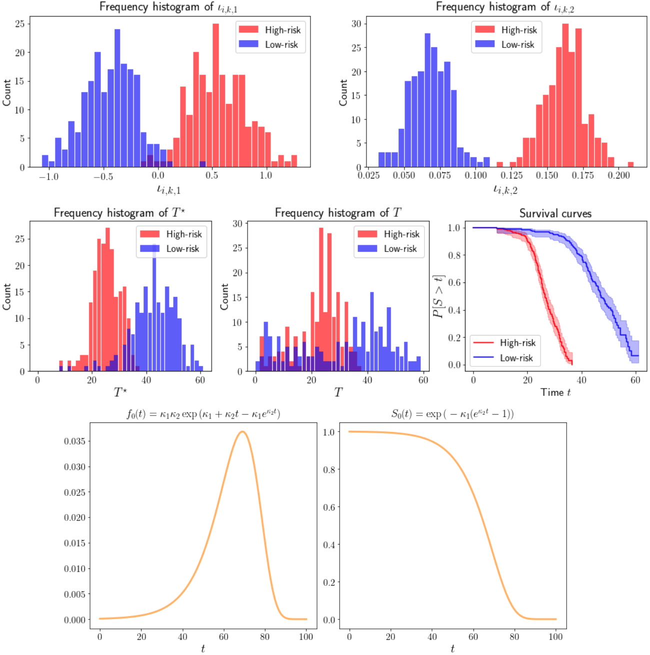
High-dimensional simulation study (1/2)

- High-risk subjects proportion $\pi_1 \in [0, 1]$
- $\mathcal{H} = \big\{ \lfloor \pi_1 n \rfloor \text{ random samples without replacement} \big\}$
- $[x_{ij}] \in \mathbb{R}^{n \times p} \sim \mathcal{N}\big(0, \Sigma_1(\rho_1)\big)$
- $x_{ij} \leftarrow x_{ij} + (-1)^{\mathbb{1}_{\{i \notin \mathcal{H}\}}} gap$ for $j = 1, \ldots, s$
- $\xi = (\underbrace{\varsigma_1, \ldots, \varsigma_1}_{s}, 0, \ldots, 0) \in \mathbb{R}^p$
- $G_i \sim \mathcal{B}\big(\pi_\xi(x_i)\big)$
- $\mathcal{S}_k = \Big\{ k \big\lfloor \frac{L r_s}{K} \big\rfloor + 1, \ldots, (k+1) \big\lfloor \frac{L r_s}{K} \big\rfloor \Big\}$
- $\beta_k^l \sim \mathbb{1}_{\{l \in \mathcal{S}_k\}} \mathcal{N}\Big(\mu_k, \begin{bmatrix} \rho_3 & 0 \\ 0 & \rho_3 \end{bmatrix} \Big)$
- $Y_i^l(t) = \epsilon_i^l(t) + \mathbb{1}_{\{l \in \mathcal{S}_k\}} \sum_{k=0}^{K-1} \mathbb{1}_{\{G_i = k\}} \big((1, t)^\top \beta_k^l + (1, t)^\top b_i^l \big)$
- $\big(\tilde{\Psi}_{k,a}(t, \beta_k, b_i^l) \big)_{a \in \{1,2\}} = \big(\beta_{k,1}^l + \beta_{k,2}^l t + b_{i,1}^l + b_{i,2}^l t \ , {b_i^l}^\top \big)$
- $\tilde{\gamma}_{k,a}^l = \varsigma_2 \mathbb{1}_{\{l \in \mathcal{S}_k, \ a \in \{1,2\}\}}$
- $\lambda_i(t | G_i = k) = \lambda_0(t) \exp \big\{ \iota_{i,k,1} + \iota_{i,k,2} t \big\}$
- Gompertz baseline $\lambda_0(t) = \kappa_1 \kappa_2 \exp(\kappa_2 t)$
- $T_i^\star | G_i = k \sim \dfrac{1}{\iota_{i,k,2} + \kappa_2} \log \Big(1 - \dfrac{(\iota_{i,k,2} + \kappa_2) \log U_i}{\kappa_1 \kappa_2 \exp \iota_{i,k,1}} \Big)$

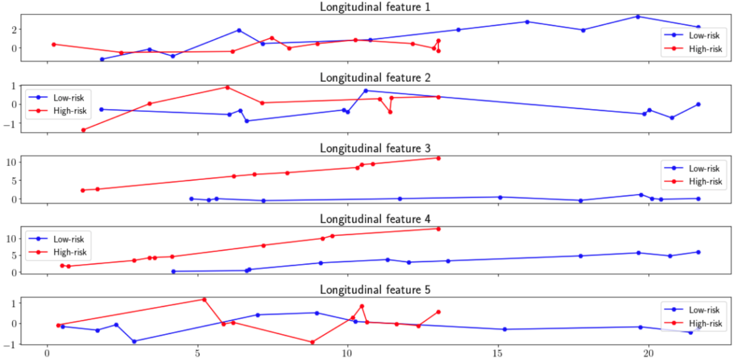
High-dimensional simulation study (2/2)



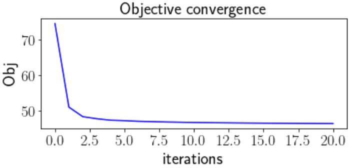
High-dimensional simulation study (2/2)

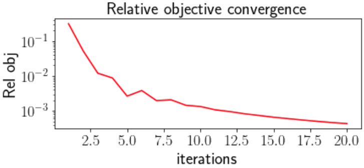


High-dimensional simulation study (2/2)

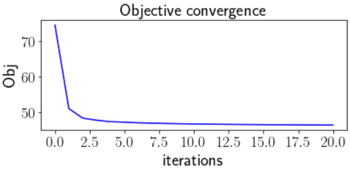


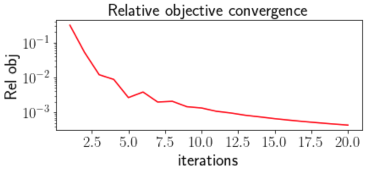
First results (1/2)

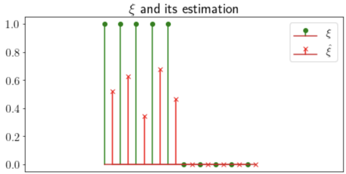




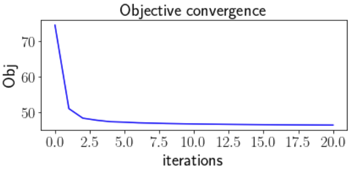
First results (1/2)

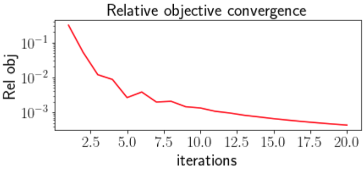


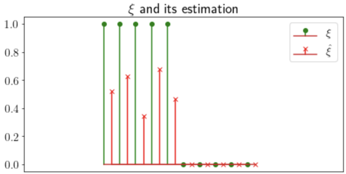


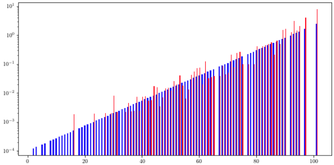


First results (1/2)

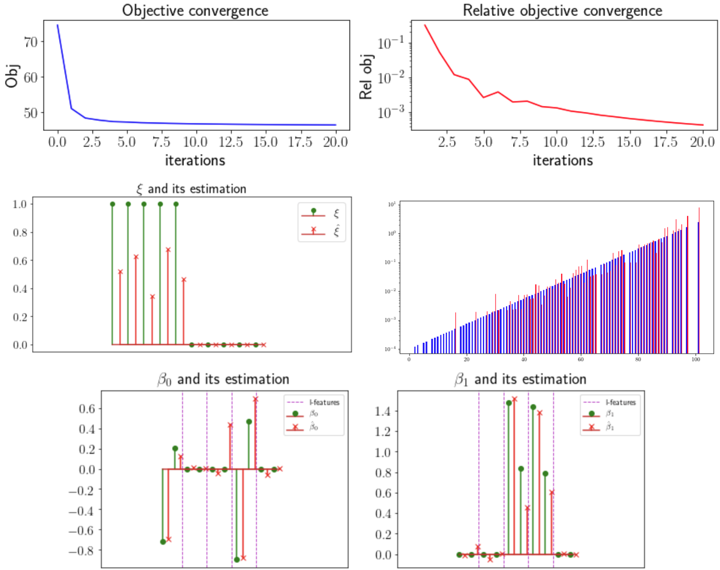




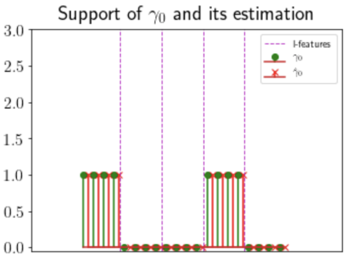


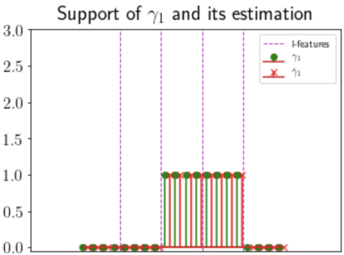


First results (1/2)

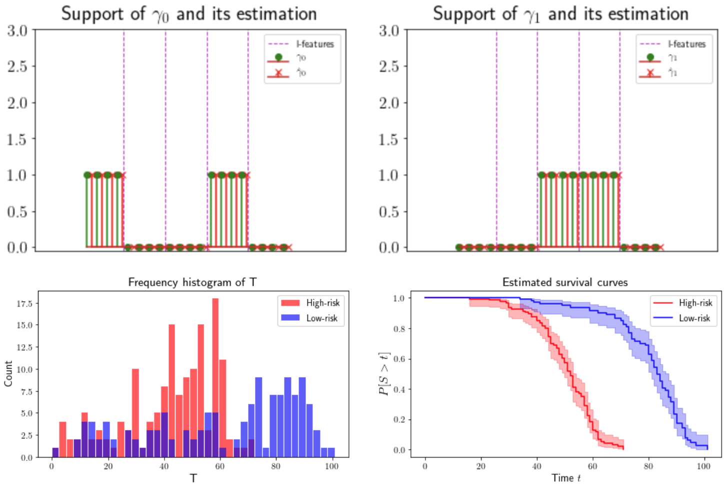


First results (2/2)





First results (2/2)



V. Conclusion

Conclusion

- Prognostic method called lights introduced to deal with the problem of joint modeling of longitudinal data and censored durations, where a large number of longitudinal features are available

Conclusion

- Prognostic method called lights introduced to deal with the problem of joint modeling of longitudinal data and censored durations, where a large number of longitudinal features are available
- Penalization of the likelihood in order to perform feature selection and to prevent overfitting

Conclusion

- Prognostic method called lights introduced to deal with the problem of joint modeling of longitudinal data and censored durations, where a large number of longitudinal features are available
- Penalization of the likelihood in order to perform feature selection and to prevent overfitting
- New efficient estimation algorithm (prox-QNEM) has been derived

Conclusion

- Prognostic method called lights introduced to deal with the problem of joint modeling of longitudinal data and censored durations, where a large number of longitudinal features are available
- Penalization of the likelihood in order to perform feature selection and to prevent overfitting
- New efficient estimation algorithm (prox-QNEM) has been derived
- Automatically determines significant prognostic longitudinal features

Conclusion

- Prognostic method called lights introduced to deal with the problem of joint modeling of longitudinal data and censored durations, where a large number of longitudinal features are available
- Penalization of the likelihood in order to perform feature selection and to prevent overfitting
- New efficient estimation algorithm (prox-QNEM) has been derived
- Automatically determines significant prognostic longitudinal features

Python 3 package

- Available at https://github.com/Califrais/lights

Conclusion

- Prognostic method called lights introduced to deal with the problem of joint modeling of longitudinal data and censored durations, where a large number of longitudinal features are available
- Penalization of the likelihood in order to perform feature selection and to prevent overfitting
- New efficient estimation algorithm (prox-QNEM) has been derived
- Automatically determines significant prognostic longitudinal features

Python 3 package

- Available at `https://github.com/Califrais/lights`
- Applications of the model available soon

Thank you!

References

[1] Eleni-Rosalina Andrinopoulou, Dimitris Rizopoulos, Johanna JM Takkenberg, and Emmanuel Lesaffre. Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. *Statistical methods in medical research*, 26(4):1787–1801, 2017.

[2] Yueh-Yun Chi and Joseph G Ibrahim. Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62(2):432–445, 2006.

[3] Laura A Hatfield, Mark E Boye, and Bradley P Carlin. Joint modeling of multiple longitudinal patient-reported outcomes and survival. *Journal of Biopharmaceutical Statistics*, 21(5):971–991, 2011.

[4] Cécile Proust-Lima, Viviane Philipps, and Benoit Liquet. Estimation of extended mixed models using latent classes and latent processes: The r package lcmm. *Journal of Statistical Software, Articles*, 78(2):1–56, 2017. ISSN 1548-7660. doi: 10.18637/jss.v078.i02. URL https://www.jstatsoft.org/v078/i02.

[5] Dimitris Rizopoulos. The r package jmbayes for fitting joint models for longitudinal and time-to-event data using mcmc. *Journal of Statistical Software, Articles*, 72(7):1–46, 2016. ISSN 1548-7660. doi: 10.18637/jss.v072.i07. URL https://www.jstatsoft.org/v072/i07.

[6] Dimitris Rizopoulos and Pulak Ghosh. A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in medicine*, 30(12):1366–1380, 2011.

[7] Glenn T Sueyoshi. Semiparametric proportional hazards estimation of competing risks models with time-varying covariates. *Journal of econometrics*, 51(1-2):25–58, 1992.

[8] Jeroen K Vermunt and Jay Magidson. Latent class models for classification. *Computational Statistics & Data Analysis*, 41(3-4):531–537, 2003.

[9] Michael S Wulfsohn and Anastasios A Tsiatis. A joint model for survival and longitudinal data measured with error. *Biometrics*, pages 330–339, 1997.

[10] Zhongheng Zhang, Jaakko Reinikainen, Kazeem Adedayo Adeleke, Marcel E Pieterse, and Catharina GM Groothuis-Oudshoorn. Time-varying covariates and coefficients in cox regression models. *Annals of translational medicine*, 6(7), 2018.