I. Introduction

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 Quasi-Newton Monte Carlo 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
- 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

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

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

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

- Heterogeneity of the population: latent subgroups

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

- 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 $b^l \in \mathbb{R}^{r_l} \sim \mathcal{N}(0, D_{ll})$
- $\mathrm{Cov}[b^l, b^{l'}] = D_{ll'}$ 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

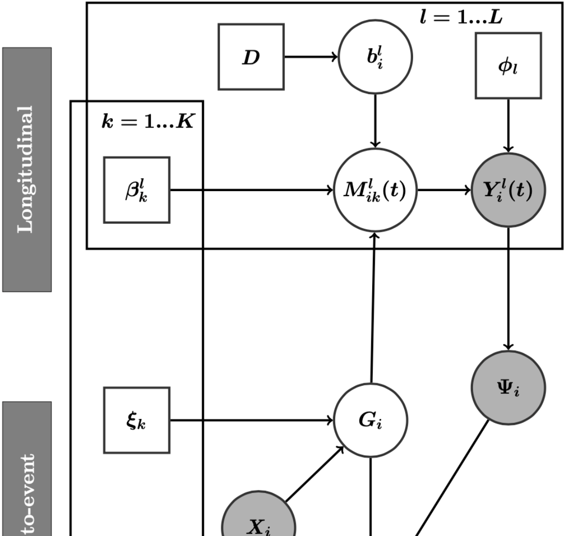
Group-specific risk of event

- $\lambda(t|G = k) = \lambda_0(t) \exp\left\{ \sum_{l=1}^{L} \sum_{a=1}^{\mathcal{A}} {\gamma_{k,a}^{l}}^{\top} \varphi_{k,a}(t, b^l) \right\}$
- Functionals $(\varphi_a)_{a \in \mathcal{A}}$

Description	$\varphi_{k,a}(t, b^l)$	$\imath_a$	Reference
Linear predictor	$m_k^l(t)$	1	Chi and Ibrahim [2]
Random effects	b^l	r_l	Hatfield et al. [3]
Time-dependent slope	$\frac{\mathrm{d}}{\mathrm{d}t} m_k^l(t)$	1	Rizopoulos and Ghosh [4]
Cumulative effect	$\int_0^t m_k^l(s)\mathrm{d}s$	1	Andrinopoulou et al. [1]

Submodels

- Graphical representation of a joint model of a time-to-event submodel and K-multivariate longitudinal outcomes submodel



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

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

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

Likelihood

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

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

- Then, the likelihood writes

$$\ell_n(\theta) = n^{-1}\sum_{i=1}^n \log \int_{\mathbb{R}^r} \sum_{k=0}^{K-1} \pi_{\xi_k}(x_i) f(t_i, \delta_i|b_i, G_i = k; \theta) \times f(y_i|b_i, G_i = k; \theta) f(b_i; \theta)\mathrm{d}b_i$$

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

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

- Resulting optimization problem

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

QNMCEM 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

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

for different functions g, where we denote

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

- Monte Carlo approximations used for untractable integrals

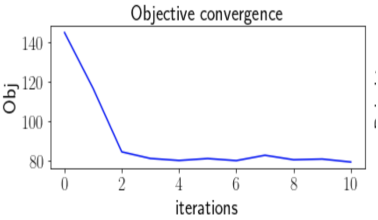
QNMCEM algorithm (2/2)

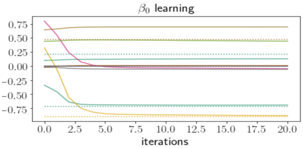
Quasi-Newton M-step

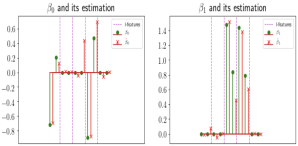
- $\theta^{(w+1)} \in \mathrm{argmin}_\theta\ \mathcal{Q}_n(\theta, \theta^{(w)}) + \sum_{k=0}^{K-1} \zeta_{1,k} \|\xi_k\|_{\mathrm{en},\eta} + \zeta_{2,k} \|\gamma_k\|_{\mathrm{sg}l_1,\tilde{\eta}}$
- $D^{(w+1)} = n^{-1} \sum_{i=1}^{n} \hat{\mathbb{E}}_{\theta^{(w)}}[b_i b_i^\top | t_i, \delta_i, y_i]$
- $R_{n,k}^{(w)}(\beta_k) =$
  $-n^{-1} \sum_{i=1}^{n} \hat{\pi}_{ik}^{\theta^{(w)}} \Big[(y_i \odot \Phi_i^{(w)})^\top \hat{\mathbb{E}}_{\theta^{(w)}}[M_{ik} | t_i, \delta_i, y_i] - \hat{\mathbb{E}}_{\theta^{(w)}}[c_{\phi^{(w)}}(M_{ik}) | t_i, \delta_i, y_i] \Big]$
- $\beta_k^{(w+1)} \in \mathrm{argmin}_{\beta_k \in \mathbb{R}^q}\ R_{n,k}^{(w)}(\beta_k)$
- $P_{n,k}^{(w)}(\xi_k) = -n^{-1} \sum_{i=1}^{n} \hat{\pi}_{ik}^{\theta^{(w)}} \log \pi_{\xi_k}(x_i)$
- $\xi_k^{(w+1)} \in \mathrm{argmin}_{\xi_k \in \mathbb{R}^p}\ P_{n,k}^{(w)}(\xi_k) + \zeta_{1,k} \|\xi_k\|_{\mathrm{en},\eta}$
- L-BFGS-B to solve the problem
- Proximal gradient method to estimate $\gamma_k^{(w+1)}$
- Predictive marker $\hat{\mathcal{R}}_{ik} = \dfrac{\pi_{\hat{\xi}_k}(x_i) \hat{f}(t_i^{max}, y_i | b_i, G_i = k; \hat{\theta})}{\sum_{k=0}^{K-1} \pi_{\hat{\xi}_k}(x_i) \hat{f}(t_i^{max}, y_i | b_i, G_i = k; \hat{\theta})}$, which is an estimate of $\mathbb{P}_\theta[G_i = k | T_i^\star > t_i^{max}, y_i]$

V. Evaluation

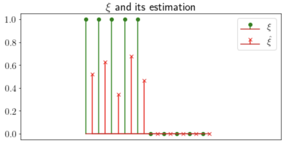
Experiments







figures/xi_learning.png



VI. 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
- Penalization of the likelihood in order to perform feature selection and to prevent overfitting
- New efficient estimation algorithm (QNMCEM) 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 on an arXiv paper.

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] 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.