A Fast and Scalable Sparse Regression Method for Competing Risks Data

Eric S. Kawaguchi

University of California, Los Angeles
(Joint work with Drs. Marc A. Suchard, Jenny I. Shen, and Gang Li)
May 13, 2019

Outline

- Introduction to competing risks data and the Fine-Gray PSH model
- Computational issues
- Using the R package **fastcmprsk**
- Numerical studies
- Discussion

Introduction

- Competing risks data are commonly observed in biomedical studies.
  - Ex: Time until first kidney transplant for patients with dialysis
  - Right censoring: lost to followup or dropout

Introduction

- Competing risks data are commonly observed in biomedical studies.
  - Ex: Time until first kidney transplant for patients with dialysis
  - Right censoring: lost to followup or dropout
  - Competing risks: death, recovery of renal function, discontinuation of dialysis

Preliminaries: Data and model

Observe n i.i.d quadruplets $\{(X_i, \delta_i, \delta_i\epsilon_i, \mathbf{z}_i)\}_{i=1}^n$ where for subject i,

- $X_i = \min(T_i, C_i)$ is the observed event time;
- $T_i =$ event time, and $C_i =$ right-censoring time;
- $\delta_i = I(T_i \leq C_i)$ (right-censoring indicator);
- ϵ_i is the competing risk indicator (WLOG: $\epsilon = 1$ is the event of interest and $\epsilon = 2$ is the competing risk);
- $\mathbf{z}_i$ is a p_n-dimensional vector of time-independent covariates;
- Assume $T \perp C|\mathbf{z}$

Preliminaries: Data and model

- Cumulative Incidence Function (CIF) for event k:

$$F_k(t; \mathbf{z}) = \Pr(T \leq t, \epsilon = k | \mathbf{z})$$

- Subdistribution hazard (Gray, 1988):

$$\begin{aligned} h_1(t|\mathbf{z}) &= \lim_{\Delta t \to 0} \frac{\Pr\{t \leq T \leq t + \Delta t, \epsilon = 1 | T \geq t \cup (T \leq t \cap \epsilon \neq 1), \mathbf{z}\}}{\Delta t} \\ &= -\frac{d}{dt} \log\{1 - F_1(t; \mathbf{z})\} \end{aligned}$$

Preliminaries: Data and model

Proportional subdistribution hazards model (Fine and Gray, 1999):

$$h_1(t|\mathbf{z}) = h_{10}(t)\exp(\mathbf{z}'\boldsymbol{\beta}), \tag{1}$$

where

- $h_{10}(t)$ is a completely unspecified baseline hazard function;
- $\boldsymbol{\beta}$ is a $p_n \times 1$ vector of regression coefficients.

Estimation

The log-pseudolikelihood of (1) is defined as

$$l_n(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \left(\beta' \mathbf{z}_i - \log \left[\sum_j \hat{w}_j(s) Y_j(s) \exp(\beta' \mathbf{z}_j) \right] \right) \times \hat{w}_i(s) dN_i(s), \tag{2}$$

where

- $N_i(t) = I(T_i \leq t, \epsilon_i = 1)$, $Y_i(t) = 1 - N_i(t-)$;
- $\hat{w}_i(t)$ is a time-dependent weight based on the inverse probability of censoring weighting technique (Robins and Rotnitzky, 1992);
- $\hat{w}_i(t) = I(C_i \geq T_i \wedge t)\hat{G}(t)/\hat{G}(X_i \wedge t)$ (Fine and Gray, 1999)
  - $G(t) = \Pr(C \geq t)$ is the survival function of the censoring variable C;
  - $\hat{G}(t)$ is the Kaplan-Meier estimate for $G(t)$

Large-scale competing risks data

- A toy example: Over **26** hours for an USRDS data with $n = 125,000$ and $p = 63$ (on a system with an Intel Core i5 2.9 GHz processor and 16GB of memory);

Large-scale competing risks data

- A toy example: Over **26** hours for an USRDS data with $n = 125,000$ and $p = 63$ (on a system with an Intel Core i5 2.9 GHz processor and 16GB of memory);
- **Computational complexity =** $O(p \times n^2)$**;**
  E.g., the score function

$$\dot{l}_j(\boldsymbol{\beta}) = \sum_{i=1}^{n} I(\delta_i \epsilon_i = 1) z_{ij} - \sum_{i=1}^{n} I(\delta_i \epsilon_i = 1) \frac{\sum_{k \in R_i} z_{kj} \tilde{w}_{ik} \exp(\eta_k)}{\sum_{k \in R_i} \tilde{w}_{ik} \exp(\eta_k)},$$

  where

$$\tilde{w}_{ik} = \hat{w}_k(X_i) = \hat{G}(X_i)/\hat{G}(X_i \wedge X_k), \quad k \in R_i,$$

  $R_i = \{k : (X_k \geq X_i) \cup (X_k \leq X_i \cap \epsilon_k = 2)\}$ and $\eta_k = \mathbf{z}_k^T \boldsymbol{\beta}$.

Large-scale competing risks data

- A toy example: Over **26** hours for an USRDS data with $n = 125,000$ and $p = 63$ (on a system with an Intel Core i5 2.9 GHz processor and 16GB of memory);
- **Computational complexity =** $O(p \times n^2)$**;**
  E.g., the score function

$$\dot{l}_j(\boldsymbol{\beta}) = \sum_{i=1}^{n} I(\delta_i \epsilon_i = 1) z_{ij} - \sum_{i=1}^{n} I(\delta_i \epsilon_i = 1) \frac{\sum_{k \in R_i} z_{kj} \tilde{w}_{ik} \exp(\eta_k)}{\sum_{k \in R_i} \tilde{w}_{ik} \exp(\eta_k)},$$

  where

$$\tilde{w}_{ik} = \hat{w}_k(X_i) = \hat{G}(X_i)/\hat{G}(X_i \wedge X_k), \quad k \in R_i,$$

  $R_i = \{k : (X_k \geq X_i) \cup (X_k \leq X_i \cap \epsilon_k = 2)\}$ and $\eta_k = \mathbf{z}_k^T \boldsymbol{\beta}$.
- Selection of tuning parameters for penalized regression adds another layer of computational complexity.

The computational bottleneck

- Recall that the score function is

$$\dot{l}_j(\boldsymbol{\beta}) = \sum_{i=1}^{n} I(\delta_i \epsilon_i = 1) z_{ij} - \sum_{i=1}^{n} I(\delta_i \epsilon_i = 1) \frac{\sum_{k \in R_i} z_{kj} \tilde{w}_{ik} \exp(\eta_k)}{\sum_{k \in R_i} \tilde{w}_{ik} \exp(\eta_k)}, \quad (3)$$

where

$$\tilde{w}_{ik} = \hat{w}_k(X_i) = \hat{G}(X_i)/\hat{G}(X_i \wedge X_k), \quad k \in R_i,$$

$R_i = \{k : (X_k \geq X_i) \cup (X_k \leq X_i \cap \epsilon_k = 2)\}$ and $\eta_k = \mathbf{z}_k^T \boldsymbol{\beta}$.

- The main bottleneck arises from the second sum: $O(n^2)$.

Forward-backward scan algorithm

Lemma

Assume that no ties are present. Then

$$\sum_{k \in R_i} \tilde{w}_{ik} \exp(\eta_k) = \sum_{k \in R_i(1)} \exp(\eta_k) + \hat{G}(X_i) \sum_{k \in R_i(2)} \exp(\eta_k) / \hat{G}(X_k) \quad (4)$$

where $R_i(1) = \{y : (X_y \geq X_i)\}$ *and* $R_i(2) = \{y : (X_y < X_i \cap \epsilon_y = 2)\}$ *are distinct partitions of* R_i. *Furthermore,* $R_i(1)$ *is monotonically decreasing over time and* $R_i(2)$ *is monotonically increasing over time.*

Main takeaway: Reduces computation time to $O(n)$.

Functionalities in fastcmprsk

Function name	Basic description
Modeling functions	
fastCrr	Fits unpenalized Fine-Gray regression
fastCrrp	Fits penalized Fine-Gray regression
Utilities	
summary	Returns ANOVA table from **fastCrr** output
predict	Estimates CIF given a vector of covariates
plot	Plots output (object dependent)
varianceControl	Options for bootstrap variance
simulateTwoCauseFineGrayModel	Simulates two-cause competing risks data

fastcmprsk in action

Data Analysis: USRDS Dataset

- United States Renal Data System: National data system funded by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) that collects information about end-stage renal disease in the United States;
- We extract a subset of $n = 125,000$;
- Considered 63 demographic and clinical covariates;
- Event of interest: First kidney transplant for patients who were currently on dialysis;
- Competing risks: Death, renal function recovery, and discontinuation of dialysis;
- Lost to follow up or had no event by end of study period were considered as right censored.

	Timing comparison (seconds)	
Unpenalized	*crr*	*fastCrr*
w.o. variance	4,544	4
w. variance	96,120	380
Penalized	*crrp*	*fastCrrp*
LASSO	86,304	32
SCAD	92,591	35
MCP	102,585	33

Table: Timing comparison using a subset of the USRDS dataset. The first two rows correspond to unpenalized Fine-Gray regression with and without variance estimation using *crr* and *fastCrr*. The last three rows correspond to penalized Fine-Gray regression using *crrp* and *fastCrrp*.

Concluding remarks:

- We develop a package to allow for scalable Fine-Gray PSH regression (unpenalized and penalized);

Concluding remarks:

- We develop a package to allow for scalable Fine-Gray PSH regression (unpenalized and penalized);
- Future work
  - Can be extended to different Fine-Gray scenarios (time-dependent covariates, stratified/clustered analyses);
  - Allow parallel computing for bootstrap;
  - Extend methodology to tackle sparse high-dimensional massive sample size data.

Thank You!

ϵ_i	1	2	0	1	1	0	2	1	2	0	1	1
$i \backslash k$	12	11	10	9	8	7	6	5	4	3	2	1
12	$\hat{\gamma}_{12}$	$\hat{\gamma}_{11}\frac{\hat{G}(12)}{\hat{G}(11)}$	0	$\dots$	$\dots$	0	$\hat{\gamma}_6\frac{\hat{G}(12)}{\hat{G}(6)}$	0	$\hat{\gamma}_4\frac{\hat{G}(12)}{\hat{G}(4)}$	0	0	0
9	$\hat{\gamma}_{12}$	$\hat{\gamma}_{11}$	$\hat{\gamma}_{10}$	$\hat{\gamma}_9$	0	0	$\hat{\gamma}_6\frac{\hat{G}(9)}{\hat{G}(6)}$	0	$\hat{\gamma}_4\frac{\hat{G}(9)}{\hat{G}(4)}$	0	0	0
7	0	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	0
3	0	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	0
6	0	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	0
10	0	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	0
5	$\hat{\gamma}_{12}$	$\hat{\gamma}_{11}$	$\hat{\gamma}_{10}$	$\hat{\gamma}_9$	$\hat{\gamma}_8$	$\hat{\gamma}_7$	$\hat{\gamma}_6$	$\hat{\gamma}_5$	$\hat{\gamma}_4\frac{\hat{G}(5)}{\hat{G}(4)}$	0	0	0
4	0	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	0
8	$\hat{\gamma}_{12}$	$\hat{\gamma}_{11}$	$\hat{\gamma}_{10}$	$\hat{\gamma}_9$	$\hat{\gamma}_8$	0	$\hat{\gamma}_6\frac{\hat{G}(8)}{\hat{G}(6)}$	0	$\hat{\gamma}_4\frac{\hat{G}(8)}{\hat{G}(4)}$	0	0	0
2	$\hat{\gamma}_{12}$	$\hat{\gamma}_{11}$	$\hat{\gamma}_{10}$	$\hat{\gamma}_9$	$\hat{\gamma}_8$	$\hat{\gamma}_7$	$\hat{\gamma}_6$	$\hat{\gamma}_5$	$\hat{\gamma}_4$	$\hat{\gamma}_3$	$\hat{\gamma}_2$	0
11	0	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	0
1	$\hat{\gamma}_{12}$	$\hat{\gamma}_{11}$	$\hat{\gamma}_{10}$	$\hat{\gamma}_9$	$\hat{\gamma}_8$	$\hat{\gamma}_7$	$\hat{\gamma}_6$	$\hat{\gamma}_5$	$\hat{\gamma}_4$	$\hat{\gamma}_3$	$\hat{\gamma}_2$	$\hat{\gamma}_1$

Table: Risk set contribution for each subject. $\gamma_i = \exp(\eta_i)$

ϵ_i	1	2	0	1	1	0	2	1	2	0	1	1
$i \backslash k$	12	11	10	9	8	7	6	5	4	3	2	1
12	$\hat{\gamma}_{12}$	$\hat{\gamma}_{11}\frac{\hat{G}(12)}{\hat{G}(11)}$	0	$\dots$	$\dots$	0	$\hat{\gamma}_6\frac{\hat{G}(12)}{\hat{G}(6)}$	0	$\hat{\gamma}_4\frac{\hat{G}(12)}{\hat{G}(4)}$	0	0	0
11	0	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	0
10	0	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	0
9	$\hat{\gamma}_{12}$	$\hat{\gamma}_{11}$	$\hat{\gamma}_{10}$	$\hat{\gamma}_9$	0	0	$\hat{\gamma}_6\frac{\hat{G}(9)}{\hat{G}(6)}$	0	$\hat{\gamma}_4\frac{\hat{G}(9)}{\hat{G}(4)}$	0	0	0
8	$\hat{\gamma}_{12}$	$\hat{\gamma}_{11}$	$\hat{\gamma}_{10}$	$\hat{\gamma}_9$	$\hat{\gamma}_8$	0	$\hat{\gamma}_6\frac{\hat{G}(8)}{\hat{G}(6)}$	0	$\hat{\gamma}_4\frac{\hat{G}(8)}{\hat{G}(4)}$	0	0	0
7	0	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	0
6	0	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	0
5	$\hat{\gamma}_{12}$	$\hat{\gamma}_{11}$	$\hat{\gamma}_{10}$	$\hat{\gamma}_9$	$\hat{\gamma}_8$	$\hat{\gamma}_7$	$\hat{\gamma}_6$	$\hat{\gamma}_5$	$\hat{\gamma}_4\frac{\hat{G}(5)}{\hat{G}(4)}$	0	0	0
4	0	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	0
3	0	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	0
2	$\hat{\gamma}_{12}$	$\hat{\gamma}_{11}$	$\hat{\gamma}_{10}$	$\hat{\gamma}_9$	$\hat{\gamma}_8$	$\hat{\gamma}_7$	$\hat{\gamma}_6$	$\hat{\gamma}_5$	$\hat{\gamma}_4$	$\hat{\gamma}_3$	$\hat{\gamma}_2$	0
1	$\hat{\gamma}_{12}$	$\hat{\gamma}_{11}$	$\hat{\gamma}_{10}$	$\hat{\gamma}_9$	$\hat{\gamma}_8$	$\hat{\gamma}_7$	$\hat{\gamma}_6$	$\hat{\gamma}_5$	$\hat{\gamma}_4$	$\hat{\gamma}_3$	$\hat{\gamma}_2$	$\hat{\gamma}_1$

Table: Risk set contribution for each subject. $\gamma_i = \exp(\eta_i)$