A Neural EM Algorithm

for Semi-Competing Risk Prediction

Stephen Salerno and Yi Li

Department of Biostatistics

Prospective Student Day

November 5, 2022

Motivation: Lung Cancer Prognosis

Lung cancer ***prognostication*** is a complex task, particularly when considering the unique risk factors and health events in a given patient's ***clinical course***

- One of the leading causes of cancer-related deaths to date, with a ***5-year survival rate*** of approximately ***1 in 5***
- Prognosis varies greatly and depends on several ***individualized risk factors*** including smoking status, genetic variants, and other comorbid conditions

Patients diagnosed with lung cancer may experience a disease ***progression***, go into remission, or have a recurrence ***prior to death***

In ***survival analysis***, the outcome is the time until the occurrence of a specific event, such as cancer progression or death

- What distinguishes survival outcomes is that the event of interest may not be observed for all subjects; i.e., subjects can be ***censored***
- Many survival processes involve a non-terminal (e.g., progression) and a terminal (e.g., death) event, which form a ***semi-competing*** relationship [3]

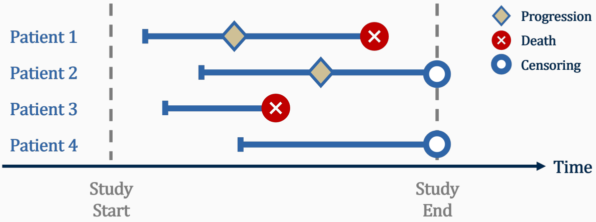


Figure: Schematic of four example patients with semi-competing risks. Diamonds indicate non-terminal events, crosses indicate terminal events, and open circles indicate censoring.

We base our approach on the ***illness-death model***, a compartment-type model for the ***hazards/transition rates*** between event states:

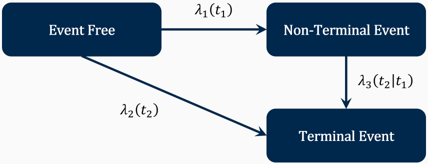


Figure: Illness-death model framework

$$\lambda_1(t_1 \mid \gamma_i, x_i) = \gamma_i \lambda_{01}(t_1) \exp\{h_1(x_i)\}; \quad t_1 > 0 \tag{1}$$
$$\lambda_2(t_2 \mid \gamma_i, x_i) = \gamma_i \lambda_{02}(t_2) \exp\{h_2(x_i)\}; \quad t_2 > 0 \tag{2}$$
$$\lambda_3(t_2 \mid t_1, \gamma_i, x_i) = \gamma_i \lambda_{03}(t_2 \mid t_1) \exp\{h_3(x_i)\}; \quad 0 < t_1 < t_2 \tag{3}$$

Specifying the Hazard Functions

$$\underbrace{\lambda_1\left(t_1 \mid \gamma_i, x_i\right)}_{\text{Hazard Function}} = \underbrace{\gamma_i}_{\text{Frailty}} \times \underbrace{\lambda_{01}\left(t_1\right)}_{\text{Baseline Hazard}} \times \underbrace{\exp\left\{h_1(x_i)\right\}}_{\text{Risk Function}}$$

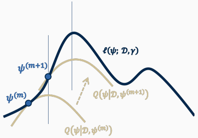
Here, we parameterize the ***hazards*** for transitioning between disease states based on three components:

1. A subject-specific random effect, or ***frailty***
2. The ***baseline hazards*** for the state transition
3. The effect of ***risk factors*** (covariates)

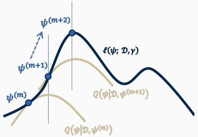
Estimating the Hazard Functions

The ***expectation-maximization (EM) algorithm*** provides a numerically stable approach for estimation, especially for large sample sizes[1]

- ***Expectation (E) Step***: Patient-specific ***frailties*** are ***estimated*** given the data and current values for the baseline hazard functions



- ***Maximization (M) Step***: The ***baseline hazards*** are ***maximized*** given the current estimates for the frailties



But how do we estimate the effect of potentially high-dimensional ***risk factors*** with complex relationships?

[1] The Hessian matrix for alternatives like the Newton-Raphson algorithm is not sparse, and its size increases in n

How to do Prediction?

Deep learning has emerged as a powerful tool for survival prediction; however, limited work has been done on multi-state outcomes, let alone semi-competing

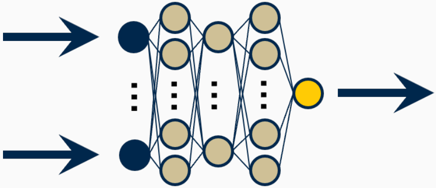


Figure: A fully-connected, feed-forward deep neural network with an input layer (blue), hidden layers (tan) and an output layer (maize)

Artificial neural networks try to mirror how the human brain functions, wherein ***nodes*** (or neurons) are connected in a network as a weighted sum of inputs through a series of ***affine transformations*** and ***nonlinear activations*** [1]

Our Extension: Neural EM Algorithm

We propose a new ***neural expectation-maximization algorithm*** which utilizes this deep learning framework and applies it semi-competing outcomes

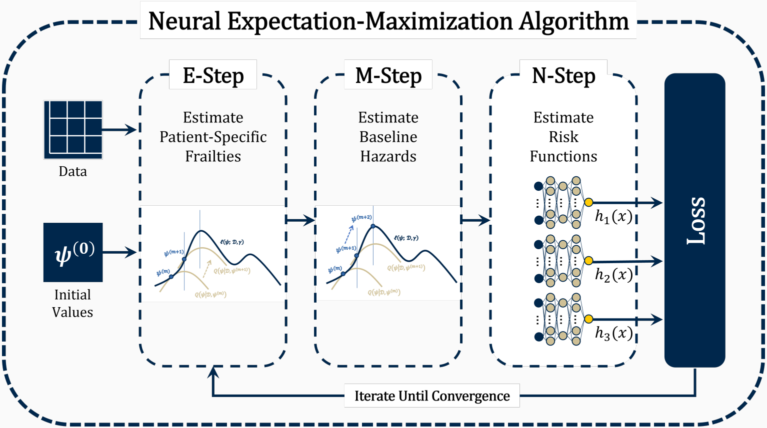


Figure: Overview of our proposed neural expectation-maximization algorithm

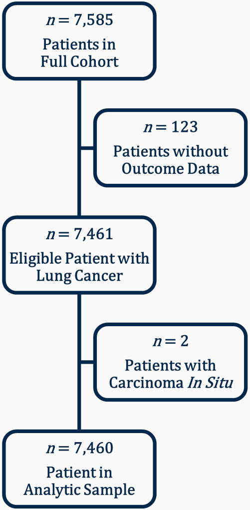
Boston Lung Cancer Study Cohort

Our study includes ***7,460 patients*** with lung cancer, diagnosed between June 1983 and October 2021 [2]

We investigated ***time to disease progression and death***, where progression might be censored by death or the study endpoint

Table: Observed Outcomes in the BLCS Cohort

	Progression	Censored
Death	143 (2%)	2,720 (36%)
Censored	295 (4%)	4,302 (58%)



There seems to exist a ***nonlinear*** effect of age that ***differs*** by type of event transition, cancer stage, and sex

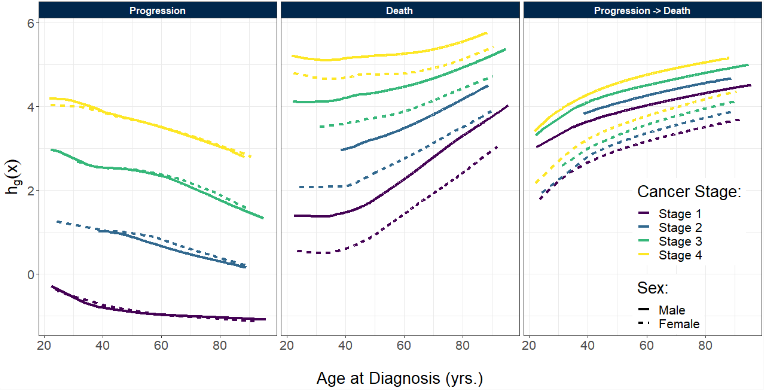


Figure: Log-risk functions of age at diagnosis on each state transition, stratified by sex (solid versus dashed lines) and initial cancer stage (line color)

Conclusions

- We have proposed a novel deep learning approach in the presence of semi-competing risks, a currently unexplored area
- Our method can recover non-linear relationships and potentially higher order interactions between disease progression, survival, and high-dimensional risk factors
- Utilizing existing paradigms for machine learning in R, we implement our method in a user-friendly workflow

Some Other Areas of Research

- Composite Quality Measures for Healthcare Reporting (Star Ratings)
- Reliability Testing for Scientific Acceptability
- Impact of COVID-19 on Patients with End-Stage Renal Disease
- Understanding radiomic features from COVID-19 chest x-rays
- Causal inference in complex survey designs

Questions?

salernos@umich.edu

BIOSTATISTICS

[1] B. Bauer and M. Kohler.
On deep learning as a remedy for the curse of dimensionality in nonparametric regression.
The Annals of Statistics, 47(4):2261–2285, 2019.

[2] D. C. Christiani.
The Boston lung cancer survival cohort.
`http://grantome.com/grant/NIH/U01-CA209414-01A1`, 2017.
[Online; accessed November 27, 2018].

[3] J. P. Fine, H. Jiang, and R. Chappell.
On semi-competing risks data.
Biometrika, 88(4):907–919, 2001.

[4] E. Fix and J. Neyman.
A simple stochastic model of recovery, relapse, death and loss of patients.
Human Biology, 23(3):205–241, 1951.

[5] S. Haneuse and K. H. Lee.
Semi-competing risks data analysis: accounting for death as a competing risk when the outcome of interest is nonterminal.
Circulation: Cardiovascular Quality and Outcomes, 9(3):322–331, 2016.

[6] J. D. Kalbfleisch and R. L. Prentice.
***The statistical analysis of failure time data*, volume 360.**
John Wiley & Sons, 2011.

[7] E. Sverdrup.
Estimates and test procedures in connection with stochastic models for deaths, recoveries and transfers between different states of health.
Scandinavian Actuarial Journal, 1965(3-4):184–211, 1965.

[8] J. Xu, J. D. Kalbfleisch, and B. Tai.
Statistical analysis of illness–death processes and semicompeting risks data.
Biometrics, 66(3):716–725, 2010.