design and generating data

Anyu Zhu

2/16/2022

Design of simulation settings

We conducted simulation studies to assess the performance of three survival models. In total, we created 7 simulation settings by mixing the event time generated from the three specified baseline hazard function: exponential, Weibull, and Gompertz hazard function. Then the generated data were fitted to exponential, Weibull, and Cox proportional hazard models. The parameters we applied in the models are constants: $\alpha = 0.5$, $\gamma = 1.5$, $\beta = -0.5$. Since the shape of Weibull distribution has great difference between $\gamma > 1$ and $\gamma < 1$, thus we simulated in scenarios where $\gamma = 0.5$ and $\gamma = 1.5$,

We simulated 500 data sets in each simulation setting. After running the models, a set of β was extracted and used to calculate the mean absolute errors, variance, and squared error. To evaluate the efficiency performances of models, we simulated data of different sample sizes: 20, 40, 60, 80, 100, 200, 400. Similarly, absolute errors, variance, MSE are calculated.

All the simulation processes were performed in R.

Methods for generating data

The survival dataset contains treatment assignment, status indicator, and observed time. Treatment assignment variable X_i is generated from a Bernoulli distribution with p = 0.5. By utilizing the inverse transformation method, we can obtain event time T:

$T = H_0^{-1}\left(\frac{-\log(u)}{e^{x^T\beta}}\right)$, where $U \sim U(0,1)$

The followings are specific baseline hazard functions we applied:

1. Survival time under Exponential distribution: $T = -\frac{\log(u)}{\lambda e^{x^T\beta}}$
2. Survival time under Weibull distribution: $T = \left(-\frac{\log(u)}{\lambda e^{x^T\beta}}\right)^{\frac{1}{\gamma}}$
3. Survival time under Gompertz distribution: $T = \frac{1}{\alpha}\log\left(1 - \frac{\alpha\log(u)}{\lambda e^{x^T\beta}}\right)$

We simulated survival data by the event time generated from the mixtures of three baseline distributions.

Mixture of Exponential and Weibull distribution:

$$T = p * (-\frac{\log(u)}{\lambda e^{x^T\beta}}) + (1-p) * (-\frac{\log(u)}{\lambda e^{x^T\beta}})^{1/\gamma}$$

We take values of p as 0, 0.5 and 1. When `p = 1`, event time is generated from exponential baseline; when `p = 0`, event time is generated from Weibull distribution.

Mixture of Exponential and Gompertz distribution:

$$T = p * (-\frac{\log(u)}{\lambda e^{x^T\beta}}) + (1-p) * (\frac{1}{\alpha}\log\left(1 - \frac{\alpha\log(u)}{\lambda e^{x^T\beta}}\right))$$

Similary, we take values of p as 0, 0.5 and 1. When `p = 1`, event time is generated from exponential baseline; when `p = 0`, event time is generated from Gompertz distribution. Finally, make event indicator variable by applying administrative censoring at $t = 5$.

By repeating each of the above simulation process 500 times, we get survival datasets with sample size ranging from 20 to 400.