Project 1: A simulation study examining three survival models

Adina Zhang

2/13/2020

Introduction

Methods

Data Generation

Generating survival data involves 5 components: actual event time, censoring time, status indicator, observed time, and covariates, if needed. In order to generate event time, we can utilize the relationship between the survival function, $S(t)$, and hazard function $h(t)$ to obtain a direct relationship between the survival function, $S(t)$, and the baseline hazard function, $h_0(t)$. Survival function is the probability of surviving beyond time t

$$S(t) = P(T > t) = 1 - F(t)$$

Hazard function is the instantaneous rate at which the event occurs at time t and can be defined as

$$h(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)} = -\frac{\partial}{\partial y} log[1 - F(t)] = -\frac{\partial}{\partial y} log[S(t)]$$

Isolating $S(t)$, we find

$$S(t) = e^{-H(t)}, where\ H(t) = \int_0^t h(t)dt$$

Additionally, we can find the connection between the cumulative hazard, $H(t)$, and the baseline hazard. Given that $h(t) = h_0(t)e^{x^T\beta}$,

$$H(t) = \int_0^t h_0(t)e^{x^T\beta}dt = e^{x^T\beta}H_0(t), where\ H_0(t) = \int_0^t h_0(t)dt$$

Putting it all together, we obtain

$$S(t) = e^{-H(t)} \qquad\qquad = e^{x^T\beta}H_0(t)$$

Finally, utilize the inverse transformation method to obtain T, event time

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

With a specific baseline hazard function, it is straightfoward to find generate time to event

1. Under exponential distribution:

$$\text{Baseline hazard function: } h_0(t) = \lambda, where\ \lambda > 0$$
$$\text{Cumulative hazard function: } H_0(t) = \lambda t$$
$$\text{Inverse cumulative hazard function: } H_0^{-1}(t) = \frac{t}{\lambda}$$
$$\text{Survival time: } T = -\frac{log(u)}{\lambda e^{x^T\beta}}$$

2. Under weibull distribution:

$$\text{Baseline hazard function: } h_0(t) = \lambda\gamma t^{\gamma-1}, where\ \lambda, \gamma > 0$$
$$\text{Cumulative hazard function: } H_0(t) = \lambda t^{\gamma}$$
$$\text{Inverse cumulative hazard function: } H_0^{-1}(t) = \bigg(\frac{t}{\lambda}\bigg)^{\frac{1}{\gamma}}$$
$$\text{Survival time: } T = \bigg(-\frac{log(u)}{\lambda e^{x^T\beta}}\bigg)^{\frac{1}{\gamma}}$$

3. Under gompertz distribution:

$$\text{Baseline hazard function: } h_0(t) = e^{\alpha t}, where\ \lambda > 0, -\infty < \alpha < \infty$$
$$\text{Cumulative hazard function: } H_0(t) = \frac{\lambda}{\alpha}\bigg(e^{\alpha t} - 1\bigg)$$
$$\text{Inverse cumulative hazard function: } H_0^{-1}(t) = \frac{1}{\alpha}log\bigg(\frac{\alpha}{\lambda}t + 1\bigg)$$
$$\text{Survival time: } T = \frac{1}{\alpha}log\bigg(1 - \frac{\alpha log(u)}{\lambda e^{x^T\beta}}\bigg)$$

Under the 3 scenarios, we use the following steps to generate the survival dataset

1. Randomly generate X_i, treatment assignment variable, from a bernoulli distribution with $p = 0.5$
2. Generate T_i, time to event, using X_i from step 1 and pre-specified β

$$T = H_0^{-1}\bigg(\frac{-\,log(u)}{e^{x^T\beta}}\bigg)$$

3. Randomly generate C_i, censoring time, from an exponential distribution
4. Determine the observe time, Y_i by comparing event and censoring time

$$Y_i = min(T_i, C_i)$$

5. Create the status indicator variable, where 1 represents if event is observed and 0 if event is censored

$$Status = \begin{cases} 1, & T_i \le C_i \\ 0, & T_i > C_i \end{cases}$$

Scenario Simiulation

Results

Discussion