P8160 Group Project 1: Evaluation of Three Survival Models Using Simulation

Wenhan Bao | Tianchuan Gao | Jialiang Hua | Qihang Wu | Paula Wu

2/20/2022

Objective

In this project, we firstly designed a simulation study to compare the accuracy and efficiency of three different survival models (Exponential, Weibull, and Cox proportional-hazards models) where they are used to estimate the treatment effects (β). Meanwhile, we also evaluated their robustness to misspecified baseline hazard functions.

Background & Statistical Methods

Background

Survival analysis is a type of statistical method used to analyze the expected duration of time until one event occurs[1]. The outcome is usually measured by two variables: **survival time (T)** and **outcome status (S)**. The **survival function S(t)** describes the probability of one subject living longer than time t. The **Kaplan-Meier (K-M) estimator** is used to approach the true survival function when the sample size is large enough.

Another two related concepts are **hazard rate (hazard function)** (h) and **censoring**. Hazard rate measures the probability of surviving beyond time t. And censoring is a type of missing data problem and might be due to some reasons such as loss to follow-up or termination of the experiment before the outcome state of all subjects is observed. In our simulation, we assume that all data are uncensored, that is, the survival time equals the duration from the start of the study to the death of every subject, assuming subjects are mortal.

K-M model and K-M plots are usually used to represent the actual situation of the sample collected. However, to estimate the survival function, the K-M models cannot be perfectly applied to cases where the predictor variables are quantitative. We instead use a class of survival models, the **proportional hazards models**, to estimate the time duration before some events occur. These models assume that a one-unit change is associated with a multiplicative change in the hazard rate and consists of two parts: a **baseline hazard function** ($h_0(t)$) and the **effect parameters** (β). In this study, we evaluated three baseline hazard functions: Exponential, Weibull, and Cox. To be noted, the Cox model assumes that the baseline hazard does not follow a particular form. Therefore it is the most general one compared to the other two models. While the Exponential distribution is a special case of Weibull distribution and thus the most restrictive one among all these three models.

Statistical Methods

Assume T ($\in [0, \infty]$) is the time to a event of interest. We define a **survival function S(t)** as

$$S(t) = Pr(T > t) = \int_t^\infty f(t)dt = 1 - F(t),$$

where $f(t)$ is the density function of T and $F(t)$ can be described as the cumulative density function (CDF) of the survive time. The hazard function $h(t)$ is defined as

$$h(t) = \lim_{\triangle t \to 0} \frac{Pr(T \in (t, t + \triangle t)|T > t)}{\triangle t} = \frac{f(t)}{S(t)}$$

From the above two formulae, we have

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

We then integrate both sides of the above equation from 0 to t, which yields $\int_0^t h(t)dt = -log[S(t)]$. Assuming $\int_0^t h(t)dt = H(t)$, we get

$$S(t) = e^{-H(t)}$$

Considering the proportional hazards model, it assumes that the hazard rate for the i-th patient at time t is $h_i(t) = h_0(t) \cdot e^{X\beta}$. Based on inverse transformation method, we can easily generate the survive data such that

$$T = F^{-1}(\mu) = H_0^{-1}(-\frac{log(1 - \mu)}{e^{X\beta}}),$$

where $U \sim Unif(0, 1)$. Since $1 - U$ follows the same distribution as U, we can simplify the above equation even more:

$$T = H_0^{-1}(-\frac{log(\mu)}{e^{X\beta}})$$

In this simulation task, we generate survival data based on three different baseline hazard functions: Exponential, Weibull, and Gompertz and they are described as follows:

The One-parameter Exponential Distribution: This distribution is obtained by take the baseline hazard function to be a constant, such that $h_0(t) = \lambda$ and $H_0(t) = \int_0^t \lambda dt$. Following steps given above, eventually we have:

$$T = -\frac{log(\mu)}{\lambda \cdot e^{X\beta}}$$

The Two-parameter Weibull Distribution: Weibull distribution is a more generalized form of the exponential distribution with hazard function $h_0(t) = \lambda \gamma t^{\gamma-1}$, where $\gamma > 0$. Similarly, once we know the $H_0(t) = \lambda \cdot t^\gamma$ from integration, we obtain:

$$T = (-\frac{log(\mu)}{\lambda \cdot e^{X\beta}})^{\frac{1}{\gamma}}$$

The Two-parameter Gompertz Distribution: In biological science and demography, this distribution is parametrized based on the Gompertz-Makehan law of mortality, which describes the human mortality more accurately in the age between 30 to 80[2]. In many investigations, there are also faster changes in survival rate over time especially in adults over 35[3]. Its hazard function usually characterized as $h_0(t) = \lambda e^{\alpha t}$, where $\alpha \in R^+$ and $\lambda \in R^+$. Therefore, the cumulative hazard function can be written as $H_0(t) = \frac{\lambda}{\alpha}(e^{\alpha t} - 1)$. Therefore, we have:

$$T = \frac{1}{\alpha} \cdot log(1 - \frac{\alpha \cdot log(\mu)}{\lambda e^{X\beta}})$$

Design of Simulation

Data Generation

Predictor X (Treatment Status): In this study, we only consider one parameter X and it represents a binary treatment indicator (coded 0 for the control group and 1 for the treatment group). Therefore, we assumed that this predictor follows an n times Bernoulli distribution with a probability equal to 0.5. Such

assumption ensures an experimental subject has an equal probability of being assigned to either the treatment group or the control group.

Coefficient β (Treatment Effects): It is the coefficient of interest. Following the equation like $h(t) = h_0(t) \cdot e^{X\beta}$, this coefficient is also called as the log hazard ratio. By applying the Monte Carlo simulation method, we can then obtain the survival data under different distributions as mentioned before.

Survival Data: In total, we generated five original different datasets by conducting the different baseline functions which are exponential, Weibull, Gompertz distribution respectively. As for the parameters for the simulation, we assume $\beta = 0.5, \lambda = 2$ for all baseline functions while Weibull takes $\gamma(0.5, 3)$ and Gompertz takes $\alpha(2, 4)$ for their own baseline functions with sample size(n) is 100. By conducting those parameters into the inverse functions which are derived above in the statistical methods part, we generated the survival data by using $Unif(0, 1)$ as a random number generator for final model training.

Censoring Time: We assumed the survival data is uncensored in this simulation, i.e. outcomes of all subjects were observed throughout the study period.

Model fitting

Each dataset generated above is used for fitting all three different proportional-hazards models which are the exponential, Weibull, and Cox proportional-hazards respectively to predict the true treatment effects (β).

Prediction Evaluation

To test the accuracy and efficiency of the models' evaluations, we use mean squared error(MSE) and the bias of the treatment effects (β) in 1000 simulations.

Results

To roughly and intuitively assess the model performance of different simulated datasets, we first plotted the survival probability against time using the Kaplan-Meier curve, a step function that represents the probability of survival in detail and the fitted model (Figure.1). We then imposed other survival models on top of the K-M curves for straightforward comparison.
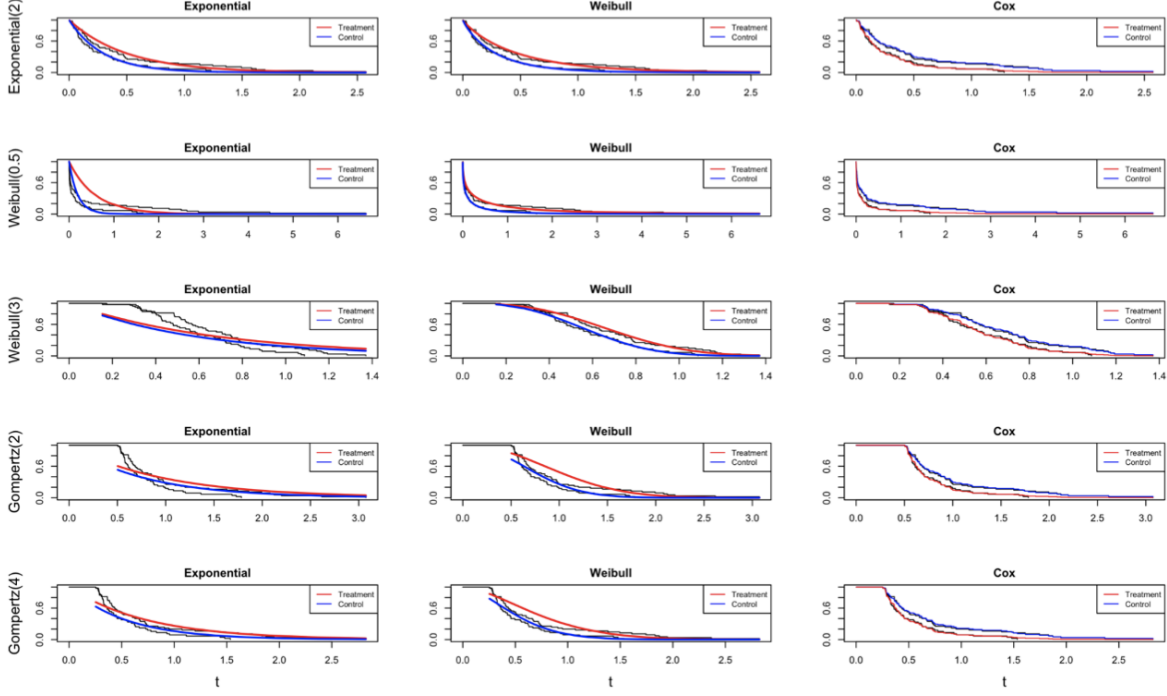
Figure 1. Different fitted models plotted versus Kaplan-Meier curve for different original datasets.
(Black Line: K-M curve, Blue Line: Fitted treatment model, Red Line: Fitted control model)

Comparing fitted models to the "true" population survival time represented by the K-M curve, we found that three survival models perform differently on the accuracy and efficiency of estimating the treatment effects β with different baseline hazard functions. When the baseline follows an exponential distribution, the performance of the three models is similar and roughly matches the K-M curve. On the other hand, when using Weibull as baseline distribution to get the simulated data, these three models showed a slight difference in their actual fit: the exponential model differs more from the K-M curve, while the Weibull and cox models overlap with the K-M curve more and thus fit better. Finally, we use another commonly used function - Gompertz distribution - as the baseline hazard function to generate data. The predictive ability of the three models appeared significantly different, where the overlap of exponential and Weibull model with K-M curve is very low, while Cox proportional-hazards model basically matches with K-M curve and has better prediction performance.

In order to further evaluate the accuracy and efficiency of the estimated treatment effect(β), we calculate the MSE of β(Table. 1). For the data generated by exponential, three models have relatively similar MSE. Then for the data simulated by Weibull, the exponential proportional-hazard model has a higher MSE while another two models have a lower MSE for both $\gamma(0.5, 3)$ value. When using the data generated by Gompertz, it's obvious to notice that the MSE of Cox proportional-hazard model is lower than others.

Model	Exponential	Weibull(0.5)	Weibull(3)	Gompertz(2)	Gompertz(4)
Exonential	0.0406	0.4516	0.1153	0.0865	0.0523
Weibull	0.0448	0.0433	0.0414	0.1155	0.0895
Cox	0.0465	0.0431	0.0426	0.0458	0.0444

Table1. MSEs of beta from three different proportional-hazard models using 1000 simulations

Model	Exponential	Weibull(0.5)	Weibull(3)	Gompertz(2)	Gompertz(4)
Exponential	0.501 (0.202)	0.999 (0.450)	0.168 (0.073)	0.219 (0.087)	0.306 (0.122)
Weibull	0.511 (0.212)	0.506 (0.208)	0.513 (0.203)	0.680 (0.288)	0.643 (0.263)
Cox	0.510 (0.215)	0.502 (0.208)	0.510 (0.206)	0.512 (0.214)	0.508 (0.211)

Table2. Mean (SD) of beta from three different proportional-hazard models using 1000 simulations

Model	Exponential	Weibull(0.5)	Weibull(3)	Gompertz(2)	Gompertz(4)
Exponential	0.001	0.499	-0.332	-0.281	-0.194
Weibull	0.011	0.006	0.013	0.180	0.143
Cox	0.010	0.002	0.010	0.012	0.008

Table3. Bias of beta from three different proportional-hazard models using 1000 simulations

When looking directly at the accuracy of predicting the estimated treatment effect(β), we provide the mean of estimated β with standard deviation by 1000 times simulation to compare with the true β we assumed as 0.5(Table. 2). Exponential proportional-hazard model only performs well when the original dataset generated by exponential distribution as baseline function. The Weibull proportional-hazard model could predict accurately when the training datasets are simulated by both exponential and Weibull distribution as baseline function in that the estimated β is close to 0.5. As for the Cox proportional-hazard model, it can predict the estimated β most accurately - all β are close to 0.5 regardless of the original training data based on different baseline functions. We also calculated the bias of the estimated β, which is defined as the difference between the estimated β and the true β. Results shown in Table.3 give the similar results as the previous tables.

Conclusions & Discussion

From the results above, we find that in general, survival models have the best performance under their corresponding baseline hazard functions. For example, from the leftmost column in Figure 1, we can see that the fitted exponential curves deviate from the K-M curves when the baseline is not exponentially distributed. This finding can also be confirmed numerically: the exponential survival model has a small MSE when the simulated dataset follows an exponential distribution, but the MSE grows large (worse performance) otherwise. In other words, the more generalized the model, the more robust it will be when the baseline hazard functions go misspecified. Combining our findings from the graph and the tables, we can see that the Cox model has small MSE, small bias with little variance across different underlying datasets. Given that MSE is the Thus, we concluded that the Cox proportional hazard model, in general, has the best performance among the three survival models in terms of accuracy, efficiency, and robustness to misspecified baseline hazard functions.

It should be noted that these three survival models all have their advantages and disadvantages. For Exponential and Weibull distribution, both of them can estimate survival function $S(t)$ and hazard ratio (HR). However, for the exponential distribution, it's not always realistic to assume that $h_0(t)$ is a constant. This problem went away in the Weibull distribution as it allows the hazard to increase or decrease proportionally with time. Nevertheless, since these two models are both parametric, which assumes that the underlying population distribution has been correctly specified. Thus, they lack robustness when the baseline hazard function went misspecified. As for the Cox proportional hazard model, it is more robust than the previous two, as we don't need to specify the baseline hazard function. However, this semi-parametric approach sacrifices its ability to estimate $S(t)$ and HR for its robustness.

There are still some limitations of our simulations. First of all, we didn't simulate censoring in our data and made the assumption that we're able to observe all the outcomes of each subject. This quickly becomes impossible in the realistic setting where losses to follow-ups are common. Second, out of numerous baseline hazard functions, we only choose three of them. Though we can still reach a relatively definite conclusion, we can still explore more in the future whether the Cox model is still robust to other baseline hazard functions.

Contributions

We contributed to this project evenly.

Appendix

For codes please click *here*

References

[1] http://www.stat.columbia.edu/~madigan/W2025/notes/survival.pdf

[2] Wikipedia. "Gompertz-Makeham law of mortality" https://en.wikipedia.org/wiki/Gompertz%E2%80%93Makeham_law_of_mortality

[3] Kalbfleisch, J.D. and Prentice, R.L. (2002). The Statistical Analysis of Failure Time Data (3rd ed.). Publisher: Wiley. DOI: 10.1002/9781118032985.