

Statistical methods

Renjie Wei

2/19/2022

Statistical Methods

Structure of Simulated Survival Data

In this study, we simulate right censored survival data with one binary treatment indicator x . Our response variable is a dichotomous variable, coded as 1 when event occurred or 0 when event did not occur during the 5-year observation period.

Follow up time is measured from time zero until the event occurs, the study ends or the participant is lost, whichever comes first.

Methods of Survival Analysis

Suppose $T \in [0, \infty)$ is the time to a event of interest, such as death, disease onset, device failure, etc. To analyze such data, we define a survival function S as

$$S(t) = \Pr(T > t) = \int_t^{\infty} f(s)ds$$

It measures the probability of survive beyond time t . If T is the time to death, then $S(t)$ is the probability of living longer than t . A closely-related concept, hazard function h , is defined as

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

where $f(t)$ is the density function of T . The hazard function measures the instantaneous risk of failure at time t giving that a patient has survived until time t .

Because of censoring in survival data, in our study, the proportional hazards models are used to investigate the efficacy of a treatment (X) on a survival time T through the hazard functions. The proportional hazards model (with a single binary treatment effect) is then given by

$$h_i(t) = h_0(t) \exp(x_i \beta),$$

where t is the time, $h_0(t)$ is the baseline hazard function, x_i is a binary treatment indicator variable coded 0 for control and 1 for the treatment, β is the parameter of interest, which is the log hazard ratio for the treatment effect. β measures the relative hazard reduction due to treatment in comparison to the control.

We fit the following three models with simulated data generated by using different distribution of time.

An exponential proportional-hazards model assumes the baseline hazard function is a constant

$$h_0(t) = \lambda$$

A Weibull proportional-hazards model assumes the baseline hazard function follows Weibull distribution, where

$$h_0(t) = \lambda \gamma t^{\gamma-1}$$

for $\gamma > 0$

A Cox proportional-hazards model leaves $h_0(t)$ unspecified.

Distribution of Time

Besides Exponential and Weibull distribution, we also used Gompertz distribution and its combination with exponential distribution as baseline hazard functions in order to evaluate the robustness of these three models, characteristics of these distributions is shown by the table below

Table 1: Characteristic of Exponential, Weibull and Gompertz distributions

	Exponential	Weibull	Gompertz
Parameter	$\lambda > 0$	$\lambda > 0, \gamma > 0$	$\lambda > 0, \alpha \in (-\infty, \infty)$
Hazard function	$h_0(t) = \lambda$	$h_0(t) = \lambda \gamma t^{\gamma-1}$	$h_0(t) = \lambda \exp(\alpha t)$
Cumulative hazard function	$H_0(t) = \lambda t$	$H_0(t) = \lambda t^\gamma$	$H_0(t) = \frac{\lambda}{\alpha} (\exp(\alpha t) - 1)$
Density function	$f_0(t) = \lambda \exp(-\lambda t)$	$f_0(t) = \lambda \gamma t^{\gamma-1} \exp(-\lambda t^\gamma)$	$f_0(t) = \lambda \exp(\alpha t) \exp(\frac{\lambda}{\alpha} (1 - \exp(\alpha t)))$
Survival function	$S_0(t) = \exp(-\lambda t)$	$S_0(t) = \exp(-\lambda t^\gamma)$	$f_0(t) = \exp(\frac{\lambda}{\alpha} (1 - \exp(\alpha t)))$

The exponential distribution with scale parameter $\lambda > 0$ has a constant hazard function for $t > 0$. In practice, the assumption of a constant hazard function is only rarely tenable. A more general form of the hazard function is given by the Weibull distribution, which is characterized by two positive parameters, scale parameter λ and the shape parameter γ . In the particular case where $\gamma = 1$ the hazard function reduces to that of the exponential distribution. For $\gamma > 1$, the hazard function increases from $0 \rightarrow \infty$ and for $0 < \gamma < 1$, it decreases monotonically from $\infty \rightarrow 0$.

Like the Weibull, the Gompertz distribution is characterized by two parameters. In the formulation shown in Table I, when $\alpha < 0$ (> 0) (we use γ instead of α in r codes for simplicity), the hazard function decreases (increases) from $\exp(\alpha)$, and when $\alpha = 0$, it reduces to the constant hazard function of an exponential distribution.