Group Project 1

Huanyu Chen, Shaolei Ma, Ruiqi Xue

2024-02-17

Design a set of distributions/models under both proportional-hazard and non-proportional-hazard assumptions

(reference link: https://spia.uga.edu/faculty__pages/rbakker/pols8501/OxfordThreeNotes.pdf)

Proportional-Hazard Assumption

Under proportional-hazards assumption, the hazard function (Cox model) can be written as:

$$h(t|x) = h_0(t)exp(\beta' x)$$

where t is the time, x the vector of covariates, β the vector of regression coefficients, $h_0(t)$ is the baseline hazard function. Then, the survival function is

$$S(t|x) = exp[-H_0(t)exp(\beta' x)]$$

where

$$H_0(t) = \int_0^t h_0(u)du$$

Thus, the distribution function is

$$F(t|x) = 1 - exp[-H_0(t)exp(\beta' x)]$$

Let Y be a random variable with distribution function F, then $U = F(Y) \sim U[0,1]$, $(1-U) \sim U[0,1]$, i.e.

$$U = exp[-H_0(t)exp(\beta' x)] \sim U[0,1]$$

if $h_0(t) > 0$ for all t, then H_0 can be inverted and the survival time T of the model can be written as

$$T = H_0^{-1}[-log(U)exp(-\beta' x)]$$

where $U \sim U[0,1]$.

To simply the problem, here we only consider one covariate x, which indicates whether the sample belongs to the control arm ($x = 0$) or the treatment arm ($x = 1$), and set a negative β under the assumption that the treatment has a consistent positive effect.

Now, we only need to know H_0^{-1} to simulate the survival time. To do so, we consider two commonly used survival time distributions: exponential and Weibull distribution.

For exponential distribution with scale parameter $\lambda > 0$, the possibility density function is

$$f_0 = \lambda exp(-\lambda t)$$

Then,

$$F_0(t) = 1 - exp(-\lambda t)$$
$$S_0(t) = 1 - F_0(t) = exp(-\lambda t)$$
$$H_0(t) = -log(S_0(t)) = \lambda t$$
$$h(t) = H_0'(t) = \lambda > 0$$
$$H_0^{-1}(t) = \lambda^{-1} t$$

Thus,

$$T = -\lambda^{-1} log(U) exp(-\beta' x)$$

where $U \sim U[0, 1]$.

For Weibull distribution with the scale parameter λ, and is the shape parameter γ, the possibility density function is

$$f_0 = \lambda \gamma t^{\gamma - 1} exp(-\lambda t^{\gamma})$$

Then,

$$F_0(t) = 1 - exp(-\lambda t^{\gamma})$$
$$S_0(t) = 1 - F_0(t) = exp(-\lambda t^{\gamma})$$
$$H_0(t) = -log(S_0(t)) = \lambda t^{\gamma}$$
$$h(t) = H_0'(t) = \lambda \gamma t^{(\gamma - 1)} > 0$$
$$H_0^{-1}(t) = (\lambda^{-1} t)^{1/\gamma}$$

Thus,

$$T = (-\lambda^{-1} log(U) exp(-\beta' x))^{1/\gamma}$$

where $U \sim U[0, 1]$.

We can write the simulation process as follows:

```
simulate_func = function(n, baseline, params = list(), coveff)
{
  # Simulate treatment indicator variable
  x = rbinom(n = n, size = 1, prob = 0.5)
  # Draw from a U(0,1) random variable
  u = runif(n)
  # Simulate survival times depending on the baseline hazard
  if (baseline == "Exponential") {
    t = -log(u) / (params$lambda * exp(x * coveff))
  } else {
    t = (-log(u) / (params$lambda * exp(x * coveff)))^(1 / params$gamma)
  }
  # Make event indicator variable applying administrative censoring at t = 5
  d = as.numeric(t < 5)
  t = pmin(t, 5)
  # Return a data.frame
  data.frame(x = x, t = t, d = d, n = n, baseline = baseline)
}
```

```
dat = simulate_func(n = 100, baseline = "Exponential",
                    params = list(lambda = 0.5), coveff = 0.5)
logrank.test(dat$t, dat$d, dat$x, rho = 0, gamma = 0)
```

```
## Call:
## logrank.test(time = dat$t, event = dat$d, group = dat$x, rho = 0,
##     gamma = 0)
##
##    N Observed Expected (O-E)^2/E (O-E)^2/V
## 1 47       46       49     0.178     0.356
## 2 53       53       50     0.174     0.356
##
##  Chisq= 0.4  on 1 degrees of freedom, p= 0.6
##  rho   =  0 gamma =  0
```

```r
logrank.test(dat$t, dat$d, dat$x, rho = 1, gamma = 0)
```

```
## Call:
## logrank.test(time = dat$t, event = dat$d, group = dat$x, rho = 1,
##     gamma = 0)
##
##    N Observed Expected (O-E)^2/E (O-E)^2/V
## 1 47       46       49     0.178      1.03
## 2 53       53       50     0.174      1.03
##
##  Chisq= 0.2  on 1 degrees of freedom, p= 0.6
##  rho   =  1 gamma =  0
```

```r
logrank.test(dat$t, dat$d, dat$x, rho = 0, gamma = 1)
```

```
## Call:
## logrank.test(time = dat$t, event = dat$d, group = dat$x, rho = 0,
##     gamma = 1)
##
##    N Observed Expected (O-E)^2/E (O-E)^2/V
## 1 47       46       49     0.178      1.13
## 2 53       53       50     0.174      1.13
##
##  Chisq= 0.3  on 1 degrees of freedom, p= 0.6
##  rho   =  0 gamma =  1
```

Cox proportional hazards model

does not assume a particular baseline hazard function $h_0(t)$ but the proportional relation between groups and baseline.

$$h_i(t) = h_0(t)\exp[\beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}]$$

Weibull proportional hazards model

assumes a specific functional form for the hazard rate, which can either increase or decrease over time. Its shape parameter distinctly describes whether the hazard rate is increasing, decreasing, or constant.

$$h_i(t) = \lambda\gamma t^{\gamma-1}\exp[\beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}]$$

where λ is the scale parameter, and γ is the shape parameter.

Non-Proportional-Hazard Assumption

Weibull accelerated failure time model

Carry out a simulation study to compare performance of those hypothesis tests in those models.

References

Bender, R., Augustin, T., & Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in medicine*, 24(11), 1713–1723. https://doi.org/10.1002/sim.2059