Survival Models

Lecture IV. Regression models. Cox proportional hazards model. Accelerated failure time model

Notation. Formulation of the problem

Let us denote

- T – failure time
- $\underline{x} = (x^{(1)}, \ldots, x^{(m)})$ – vector of available covariates (regression variables/regressors/factors/explanatory variables)

Main question

Do any subsets of the m covariates help to explain the survival time? If so, how and by what estimated quantity?

Exponential regression model (1)

Let us recall that for $T \sim Exp(\lambda)$

- expectation $ET = \frac{1}{\lambda}$
- hazard function $h(t) = \lambda$

We model the hazard for an individual as a function of the covariate vector $\underline{x}$:

$$h(t|\underline{x}) = h_0(t) \cdot k(\underline{x}'\underline{\beta}) = \lambda \cdot k(\underline{x}'\underline{\beta}) = \lambda \cdot k(\beta_1 x^{(1)} + \ldots + \beta_m x^{(m)}),$$

where

- $\underline{\beta} = (\beta_1, \ldots, \beta_m)'$ is a vector of regression parameters (coefficients),
- $\lambda > 0$ is a constant,
- k is a specificed link function (NB! From GLM point of view it is usually called the response function),
- $h_0(t)$ is the baseline hazard (the value of hazard function when the covariate vector $\underline{x} = 0$ or $\underline{\beta} = 0$)

Note that this hazard function is constant w.r.t. time t, but depends on $\underline{x}$

Exponential regression model (2)

For suitable choice of link function, one can consider

- $k(\eta) = \exp(\eta)$ – implies that covariates act multiplicatively on hazard rate
- $k(\eta) = 1 + \eta$ – implies that hazard is a linear function of $\underline{x}$
- $k(\eta) = \frac{1}{1+\eta}$ – implies that the mean $E(T|\underline{x})$ is a linear function of $\underline{x}$

Although all these link functions have nice interpretations, the most natural choice is exponential function $\exp(\eta)$ since its value is always positive no matter what the $\underline{\beta}$ and $\underline{x}$ are.

Exponential regression model (3)

Now, using exponential link function we obtain the following model

$$
\begin{aligned}
h(t|\underline{x}) &= \lambda \cdot \exp(\underline{x}'\underline{\beta}) = \lambda \cdot \exp(\beta_1 x^{(1)} + \ldots + \beta_m x^{(m)}) \\
&= \lambda \cdot \exp(\beta_1 x^{(1)}) \cdot \ldots \cdot \exp(\beta_m x^{(m)}) = \tilde{\lambda}
\end{aligned}
$$

or, equivalently

$$
\ln(h(t|\underline{x})) = \ln(\lambda) + \underline{x}'\underline{\beta} = \ln(\lambda) + \beta_1 x^{(1)} + \ldots + \beta_m x^{(m)}
$$

Thus

- the covariates act multiplicatively on the hazard rate
- the covariates act additively on the log hazard rate

Exponential regression model (4)

Recall that if T is exponentially distributed, $Y = \ln(T)$ is distributed as extreme (minimum) value distribution with mean $\mu = -\ln(\lambda)$ and $\sigma = 1$.
The mean $\tilde{\mu}$ (given $\underline{x}$) is thus

$$\tilde{\mu} = -\ln(\tilde{\lambda}) = -\ln(\lambda \exp(\underline{x}'\underline{\beta})) = -\ln(\lambda) - \underline{x}'\underline{\beta}$$

Now, given $\underline{x}$ we can also write

$$Y = \ln(T) = \tilde{\mu} + \sigma Z = \beta_0^* + \underline{x}'\underline{\beta}^* + Z = \underline{x}'\underline{\beta}^* + Z^*,$$

where

- $\beta_0^* = -\ln(\lambda)$,
- $\underline{\beta}^* = -\underline{\beta}$,
- Z has standard extreme (min) value distribution: $f_Z(z) = \exp(z - e^z)$,
- $Z^* = \beta_0^* + Z$

Exponential regression model (5)

In summary,

- $h(t|\underline{x}) = \lambda \exp(\underline{x}'\underline{\beta})$ is a log-linear model for the failure rate
- the model transforms into a linear model for $Y = \ln(T)$ (the covariates act additively on Y)

Weibull regression model (1)

Let us now apply similar model to Weibull distributed T. Recall that the hazard function is $h(t) = \alpha\lambda^\alpha t^{\alpha-1}$

To include the covariate vector $\underline{x}$, we write the hazard given $\underline{x}$ as

$$\begin{aligned} h(t|\underline{x}) &= h_0(t) \cdot \exp(\underline{x}'\underline{\beta}) \\ &= \alpha\lambda^\alpha t^{\alpha-1} \cdot \exp(\underline{x}'\underline{\beta}) = \alpha \left(\lambda \cdot (\exp(\underline{x}'\underline{\beta}))^{\frac{1}{\alpha}}\right)^\alpha t^{\alpha-1} \\ &= \alpha\tilde{\lambda}^\alpha t^{\alpha-1}, \end{aligned}$$

where $\tilde{\lambda} = \lambda \cdot (\exp(\underline{x}'\underline{\beta}))^{\frac{1}{\alpha}}$

Weibull regression model (2)

Similarly to exponential model, we get the following result in log-scale:

$$\begin{aligned}\ln(h(t|\underline{x})) &= \ln(\alpha) + \alpha\ln(\tilde{\lambda}) + (\alpha-1)\ln(t)\\ &= \ln(\alpha) + \alpha\ln(\lambda) + \underline{x}'\underline{\beta} + (\alpha-1)\ln(t)\end{aligned}$$

Recall that if T has Weibull distribution, $Y = \ln(T)$ is distributed as extreme (minimum) value distribution with mean $\mu = -\ln(\lambda)$ and $\sigma = \frac{1}{\alpha}$.

The mean $\tilde{\mu}$ (given $\underline{x}$) is thus

$$\tilde{\mu} = -\ln(\tilde{\lambda}) = -\ln(\lambda\cdot(\exp(\underline{x}'\underline{\beta}))^{\frac{1}{\alpha}}) = -\ln(\lambda) - \frac{1}{\alpha}\underline{x}'\underline{\beta}$$

Weibull regression model (3)

Now, given $\underline{x}$ we can also write

$$Y = \ln(T) = \tilde{\mu} + \sigma Z = \beta_0^* + \underline{x}'\underline{\beta}^* + \sigma Z = \underline{x}'\underline{\beta}^* + Z^*,$$

where

- $\beta_0^* = -\ln(\lambda)$,
- $\underline{\beta}^* = -\sigma\underline{\beta}$,
- Z has standard extreme (min) value distribution: $f_Z(z) = \exp(z - e^z)$,
- $Z^* = \beta_0^* + \sigma Z$

Weibull regression model (4)

Since T given $\underline{x}$ is assumed to follow Weibull distribution with parameters $\tilde{\lambda}$ and α, we can also easily derive its survivor function $S(t|\underline{x})$

$$S(t|\underline{x}) = \exp\left(-(\tilde{\lambda}t)^{\alpha}\right)$$

and cumulative hazard function $H(t|\underline{x})$

$$H(t|\underline{x}) = -\ln(S(t)) = (\tilde{\lambda}t)^{\alpha} = \lambda t^{\alpha}\exp(\underline{x}'\underline{\beta}) = H_0(t)\exp(\underline{x}'\underline{\beta})$$

Weibull regression model (5)

In summary

- the effects of the covariates $\underline{x}$ act multiplicatively on the hazard function $h(t|\underline{x})$
- the model is log-linear for T, i.e. it transforms into a linear model for $Y = \ln(T)$. In other words, the covariates $\underline{x}$ also act additively on $\ln(T)$ (multiplicatively on T)

A remark on survival modelling

Remark

A statistical goal of survival analysis (similarly to linear and logistic ression) is to obtain some measure of effect that will describe the relationship between a predictor variable of interest and time to failure, after adjusting for the other variables we have identified in the study and included in the model.

In different analyses, different measures of effect are used:

- linear regression – regression coefficient
- logistic regression – odds ratio
- survival analysis – hazards ratio (HR)

Cox proportional hazards (PH) model (1)

For Cox PH model, the hazard function is

$$h(t|\underline{x}) = h_0(t) \cdot \exp(\underline{x}'\underline{\beta}),$$

where $h_0(t)$ is a baseline hazard function (does not depend on the covariates $\underline{x}$)

Obviously, exponential and Weibull models are special cases of this model

Definition (Proportional hazards property)

For two different observations $\underline{x}_1$ and $\underline{x}_2$, the hazard ratio

$$\frac{h(t|\underline{x}_1)}{h(t|\underline{x}_2)} = \frac{\exp(\underline{x}_1'\underline{\beta})}{\exp(\underline{x}_2'\underline{\beta})} = \exp\left((\underline{x}_1' - \underline{x}_2')\underline{\beta}\right)$$

is constant with respect to time t.

Cox proportional hazards (PH) model (2)

For any PH model, the survivor function of T given $\underline{x}$ is

$$S(t|\underline{x}) = \exp\left(-\int_0^t h(u|\underline{x})du\right) = \exp\left(-\exp(\underline{x}'\underline{\beta})\int_0^t h_0(u)du\right)$$
$$= \left(\exp\left(-\int_0^t h_0(u)du\right)\right)^{\exp(\underline{x}'\underline{\beta})} = (S_0(t))^{\exp(\underline{x}'\underline{\beta})},$$

with $S_0(t)$ being the baseline survivor function

The pdf of T given $\underline{x}$ is thus

$$f(t|\underline{x}) = h(t|\underline{x})S(t|\underline{x}) = h_0(t)\exp(\underline{x}'\underline{\beta})(S_0(t))^{\exp(\underline{x}'\underline{\beta})}$$

Cox PH model. Concluding remarks (1)

Remark

For Cox PH model, the (exact) values of observations are not important, the model (parameter estimates) is determined by the order of observations.

Let us have

- $g(\cdot)$ – a monotonely increasing function
- $g^{-1}(\cdot)$ – inverse of $g(\cdot)$, also monotonely increasing
- T – time to failure (initial)
- T^* – time to failure (transformed), $T^* = g(T)$

Now, the hazard function for T^* is

$$h_{T^*}(t) = \frac{f_{T^*}(t)}{\mathbb{P}(g(T) \geqslant t)} = \frac{-[S_T(g^{-1}(t))]'}{S_T(g^{-1}(t))} = h_T[(g^{-1}(t))](g^{-1})'(t)$$

and, assuming Cox PH model holds for T:

$$h_{T^*}(t|\underline{x}) = h_T(g^{-1}(t)|\underline{x})(g^{-1})'(t) = h_0(g^{-1}(t))\exp(\underline{x}'\underline{\beta})(g^{-1})'(t) = h_0^*(t)\exp(\underline{x}'\underline{\beta}),$$

i.e. the Cox PH model also holds for T^*

Cox PH model. Concluding remarks (2)

Remark

A general rule about applicability of PH model: if the hazard functions cross over time, the PH assumption is violated.

Possible generalizations of the Cox PH model:

- The baseline hazard $h_0(t)$ can be allowed to vary in specified subsets of the data – stratified Cox PH model
- The regression variables $\underline{x}$ can be alloved to depend on time ($\underline{x} = \underline{x}(t)$) – extended Cox model

Cox PH model. A visual test for assumptions

Let us note that if PH assumption holds, we have

$$S(t|\underline{x}) = (S_0(t))^{\exp(\underline{x}'\underline{\beta})},$$

or, equivalently,

$$\ln(-\ln S(t|\underline{x})) = \underline{x}'\underline{\beta} + \ln(-\ln S_0(t)),$$

which can be used to construct a simple visual test for the validity of the PH assumption:

- for all combinations of covariates $\underline{x}$, plot the described double logarithms of K-M curves
- for any pair $\underline{x}_1$ and $\underline{x}_2$, the distance between K-M curves should remain constant in time

Home assignment

Home assignment 5.

1. Show that for a Cox PH model, the hazard ratio is constant in time for any two observations.
2. Why isn't the intercept needed in Cox PH model?
3. Assume that the time to failure in the baseline group has Weibull distribution. What would be the distribution in other groups assuming the Cox PH model? Prove it.

Accelerated failure time (AFT) model (1)

Let us consider a log-linear regression model for T, i.e. we model $Y = \ln(T)$ as a linear function of the covariate $\underline{x}$:

$$Y = \underline{x}'\underline{\beta}^* + Z^*,$$

where Z^* has a certain distribution. Then

$$T = \exp(Y) = \exp(\underline{x}'\underline{\beta}^*) \cdot \exp(Z^*) = \exp(\underline{x}'\underline{\beta}^*) \cdot T^*,$$

where $T^* = \exp(Z^*)$

We can see that the covariate $\underline{x}$ acts multiplicatively on the survival time T

Accelerated failure time (AFT) model (2)

Suppose further that T^* has hazard function $h_0^*(t^*)$, which is independent of $\underline{\beta}^*$ (free of covariate vector $\underline{x}$). Then the hazard function of T given $\underline{x}$ can be written

$$h(t|\underline{x}) = h_0^*(\exp(-\underline{x}'\underline{\beta}^*)t) \cdot \exp(-\underline{x}'\underline{\beta}^*) \quad (1)$$

The proof of this result is fairly technical and thus omitted. See Kim&Tableman (2004, pp 103–104) for details.

Definition (Accelerated failure time (AFT) model)

A log-linear model for failure time T is called an AFT model if it has property (1).

Notice also that the following regression models belong to the class of AFT models:

- exponential
- Weibull
- log-logistic
- log-normal

Log-logistic regression model as AFT model (1)

Log-logistic regression model is a log-linear model

$$Y = \ln(T) = \beta_0^* + \underline{x}'\underline{\beta}^* + \sigma Z,$$

where

- T has log-logistic distribution
- Z has standard logistic distribution

This model is equivalent to

$$Y = \underline{x}'\underline{\beta}^* + Z^*,$$

where $Z^* = \beta_0^* + \sigma Z$ (logistic with mean β_0^* and scale σ)

Log-logistic regression model as AFT model (2)

Now, by the properties of log-logistic distribution, the baseline hazard for $T^* = \exp(Z^*)$ is

$$h_0^*(t^*) = \frac{\lambda\alpha(\lambda t^*)^{\alpha-1}}{1+(\lambda t^*)^\alpha} = \frac{\alpha\lambda^\alpha (t^*)^{\alpha-1}}{1+\lambda^\alpha (t^*)^\alpha},$$

where $\beta_0^* = -\ln(\lambda)$ and $\sigma = \frac{1}{\alpha}$
As this baseline hazard is free of $\underline{\beta}^*$, this log-linear model is indeed an AFT model.

The hazard function for T given $\underline{x}$ can also be easily derived:

$$h(t|\underline{x}) = h_0^*(\exp(-\underline{x}'\underline{\beta}^*)t)\cdot\exp(-\underline{x}'\underline{\beta}^*) = \frac{\alpha\lambda^\alpha(\exp(-\underline{x}'\underline{\beta}^*)t)^{\alpha-1}}{1+\lambda^\alpha(\exp(-\underline{x}'\underline{\beta}^*)t)^\alpha}\cdot\exp(-\underline{x}'\underline{\beta}^*)$$
$$= \frac{\alpha\tilde{\lambda}^\alpha t^{\alpha-1}}{1+\tilde{\lambda}^\alpha t^\alpha},$$

where $\tilde{\lambda} = \lambda\cdot\exp(-\underline{x}'\underline{\beta}^*)$

Thus, the log-logistic model, although a log-linear model, is not a PH model!

Alternative definition for AFT model

We can also derive the survivor function of T given $\underline{x}$:

$$S(t|\underline{x}) = \exp\left(-\int_0^t h_0^*(\exp(-\underline{x}'\underline{\beta}^*)u)\exp(-\underline{x}'\underline{\beta}^*)du\right)$$

Changing the integration variable to $v = \exp(-\underline{x}'\underline{\beta}^*)u$
(which means $dv = \exp(-\underline{x}'\underline{\beta}^*)du$ and $0 < v < \exp(-\underline{x}'\underline{\beta}^*)t$) implies

$$S(t|\underline{x}) = \exp\left(-\int_0^{\exp(-\underline{x}'\underline{\beta}^*)t} h_0^*(v)dv\right) = S_0^*(\exp(-\underline{x}'\underline{\beta}^*)t) = S_0^*(t^*),$$

where $S_0^*(t)$ denotes the baseline survivor function.

Remark

This relation depicts the essence of the AFT model and can be considered as an alternative (and most common) definition.

The essence of AFT: acceleration/deceleration of time

So, we established that an AFT model can be characterized by

$$S(t|\underline{x}) = S_0^*(\exp(-\underline{x}'\underline{\beta}^*)t) = S_0^*(t^*),$$

The scale-changing role of the covariate $\underline{x}$ is the key property of an AFT model:

- if $\underline{x}'\underline{\beta}^*$ decreases, then $S_0^*(t^*)$ decreases, which means the time to failure accelerates – accelerated failure time model
- if $\underline{x}'\underline{\beta}^*$ increases, then $S_0^*(t^*)$ increases, which means the time to failure decelerates – decelerated failure time model

Structure of studied regression models

Let Z be either a standard extreme value, standard logistic, or standard normal random variable ($\mu = 0$, $\sigma = 1$).

Then

$$Y = \log(T) = \tilde{\mu} + \sigma Z = \beta_0^* + \underline{x}'\underline{\beta}^* + \sigma Z$$

accelerated failure time model
log-linear model

$\nearrow\swarrow$	$\uparrow\downarrow$	$\searrow\nwarrow$
T	T	T
Weibull	log-logistic	log-normal
$\downarrow$	$\downarrow$	
PH property	proportional odds property	

Comparison between Cox PH and AFT models

Cox PH model

- is a multiplicative model for hazard
- is semi-parametric (the baseline hazard is not estimated)
- is more robust, less sensitive to outliers
- does not use all the information (only the order of failure times matters)

AFT model

- is a multiplicative model for failure time
- is parametric
- less robust, more sensitive to outliers
- uses all the information from data