Survival Models

Max Welz
welz@ese.eur.nl

Econometric Institute
Erasmus School of Economics

October 8, 2021

1 Essential Theory

1.1 Setup

Suppose we want to model the failure rate of a some statistical process. Concretely, let the random variable T denote the (non-negative) real-valued failure time of the process of interest. We assume that the process fails eventually and that its failure time T is observed (for now). Let F be the distribution function of T and f be its corresponding density. By definition, for some time $t \in [0, \infty]$, the distribution function

$$F(t) = \mathbb{P}[T \leq t] = \int_0^t f(s)\mathrm{d}s$$

measures the probability that the process fails before or at time t. We call F the *incidence function*. The survival function S is defined by

$$S(t) = 1 - F(t) = \mathbb{P}[T > t],$$

and denotes the probability that failure occurs *after* time t. An essential quantity in survival modeling is the *hazard function* $h : [0, \infty] \to [0, \infty)$, defined by

$$h(t) = \lim_{\Delta t \downarrow 0} \frac{\mathbb{P}[t \leq T < t + \Delta t | T \geq t]}{\Delta t}. \tag{1}$$

The hazard function $h(t)$ is interpreted as the instantaneous rate of failure if the process has not yet failed at time t. We emphasize that the hazard function is *not* a probability, which is a frequent misconception.

Proposition 1.1. *Let random variable T denote a failure time, let F and f its incidence function and corresponding density, respectively, and h its hazard function. It holds that*

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}.$$

Proof. Since f is the derivative of F, we can write

$$\begin{aligned} f(t) &= \lim_{\Delta t \downarrow 0} \frac{F(t+\Delta t) - F(t)}{\Delta t} = \lim_{\Delta t \downarrow 0} \frac{\mathbb{P}[T \leq t + \Delta t] - \mathbb{P}[T \leq t]}{\Delta t} \\ &= \lim_{\Delta t \downarrow 0} \frac{\mathbb{P}[t \leq < t + \Delta t]}{\Delta t}. \end{aligned}$$

Thereupon, we can write the hazard function h as

$$\begin{aligned} h(t) &= \lim_{\Delta t \downarrow 0} \frac{\mathbb{P}[t \leq T < t + \Delta t | T \geq t]}{\Delta t}. \\ &= \frac{1}{\mathbb{P}[T \geq t]} \lim_{\Delta t \downarrow 0} \frac{\mathbb{P}[t \leq < t + \Delta t]}{\Delta t}. \\ &= \frac{f(t)}{1 - F(t)}, \end{aligned}$$

where the second equality follows from Bayes' theorem. □

The *cumulative hazard function*, H, of hazard h is defined by

$$H(t) = \int_0^t h(s) \mathrm{d}s.$$

The hazard function h is of such importance in survival modeling because we can express the survival function (and thereby also the incidence function) in terms of the hazard. To see this, notice that $S'(t) = -f(t)$ and observe that

$$-\frac{\mathrm{d} \log(S(t))}{\mathrm{d}t} = \frac{f(t)}{S(t)} = h(t),$$

by Proposition 1.1. Taking integrals on both sides and rearranging yields the following useful expression of survival S:

$$S(t) = \exp\left(- \int_0^t h(s) \mathrm{d}s \right) = \exp\big(- H(t) \big).$$

1.2 Right-Censored Data

In practice, we may not observe the failure time T. This can happen because we may only observe the process until some (finite) point in time, and if the process has not failed by that time, we have no information on the time of its eventual failure. Suppose that the censoring (i.e. the point in time after which we stop to observe the process) takes place at some time C. The *observed* survival time (or time at risk), Y, is then defined by

$$Y = T \wedge C = \min\{T, C\},$$

which means that if the process fails before censoring time C, the failure time is observed via $Y = T$. Conversely, if the process has not yet failed at censoring time

C, the time at risk Y is equal to C and we have no information on the failure time T. We say that the observed time at risk Y is *right-censored.* Right-censoring can happen either by design (e.g. in a clinical trial which has a fixed ending time) or involuntarily due to losses to follow-up. Denote by $\delta = \mathbb{1}\{T \leq C\}$ an observed failure indicator which takes the value one if the process fails before the censoring time.

Since it is unrealistic to assume that we observe a process infinitely long until its eventual failure, survival model such as Cox proportional hazard models typically assume that the observed survival time is right-censored.

2 Ordinary Cox Proportional Hazard Modeling

2.1 Setup

A *Cox proportional hazard model* (Cox, 1972) attempts to explain the observed right-censored time at risk $Y = T \wedge C$ by some explanatory variables which are collected in a p-dimensional random vector $\mathbf{X}$. In proportional hazard modeling, we specify the hazard function h in (1) by using the semi-parametric specification

$$h(t; \mathbf{X}, \boldsymbol{\beta}) = h_0(t) \exp\left(\mathbf{X}^\top \boldsymbol{\beta}\right) \tag{2}$$

for fixed, but unknown coefficients $\boldsymbol{\beta} \in \mathbb{R}^p$. The function h_0 is also a hazard function (i.e. it satisfies the right-hand side of (1)), but it is completely unspecified and does not depend on anything but time t. We refer to h_0 as the *baseline hazard.* Hence, there exists an unspecified *baseline cumulative hazard function* $H_0(t) = \int_0^t h_0(s)\mathrm{d}s$ and baseline survival $S_0(t) = \exp(-H_0(t))$.

With the proportional hazard specification in (1) the cumulative hazard of h satisfies

$$\begin{aligned} H(t; \mathbf{X}, \boldsymbol{\beta}) &= \int_0^t h_0(s) \exp\left(\mathbf{X}^\top \boldsymbol{\beta}\right) \mathrm{d}s \\ &= \exp\big(\mathbf{X}^\top \boldsymbol{\beta}\big) H_0(t), \end{aligned}$$

and the associated survival function is given by

$$\begin{aligned} S(t; \mathbf{X}, \boldsymbol{\beta}) &= \exp\Big(- \exp\big(\mathbf{X}^\top \boldsymbol{\beta}\big) H_0(t)\Big) \\ &= \Big(\exp\big(- H_0(t)\big)\Big)^{\exp(\mathbf{X}^\top \boldsymbol{\beta})} \\ &= S_0(t)^{\exp(\mathbf{X}^\top \boldsymbol{\beta})}. \end{aligned} \tag{3}$$

2.2 Fitting a Proportional Hazards Model

Suppose we observe a random sample $\big\{(X_i, Y_i, \delta_i)\big\}_{i=1}^n$. Assume for now that for all failing individuals (i.e. individuals i for which $\delta_i = 1$), the failure times X_i are unique. The goal is to estimate the unknown coefficient vector $\boldsymbol{\beta} \in \mathbb{R}^p$ in the proportional hazards specification (2). For this purpose, we consider the partial likelihood function

$$\begin{aligned} L(\boldsymbol{\beta}) &= \prod_{\{i\in[n]:\delta_i=1\}} \frac{h(Y_i;\mathbf{X}_i,\boldsymbol{\beta})}{\sum_{\{j\in[n]:Y_j\geq Y_i\}} h(Y_i|\mathbf{X}_j,\boldsymbol{\beta})} \\ &= \prod_{\{i\in[n]:\delta_i=1\}} \frac{\exp(\mathbf{X}_i^\top\boldsymbol{\beta})}{\sum_{\{j\in[n]:Y_j\geq Y_i\}} \exp(\mathbf{X}_j^\top\boldsymbol{\beta})}, \end{aligned} \tag{4}$$

which is called *partial* because it is computed only on individuals who have failed before the censoring time. Observe that the partial likelihood does *not* depend on the baseline hazard h_0, which substantially facilitates the optimization task. We maximize the partial likelihood by solving

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\{ -\frac{2}{n}\log L(\boldsymbol{\beta}) + \lambda_n P(\boldsymbol{\beta}) \right\}, \tag{5}$$

where P is an optional regularization penalty on the size of the coefficients, $\lambda_n \geq 0$ is a tuning parameter, and the scaling factor $2/n$ has been added for mathematical convenience. If the penalty P is convex, the optimization problem is convex, meaning that it can be solved easily. The value of tuning parameter λ_n can be determined via cross-validation. For P the elastic net penalty (Zou and Hastie, 2005), numerical details are described in Simon et al. (2011).

2.3 Estimating Survival

Suppose we have obtained an estimator $\widehat{\boldsymbol{\beta}}$ of the coefficient vector $\boldsymbol{\beta}$ in (2) by solving the optimization problem in (5). The goal is to estimate the survival function $S(t,\mathbf{X},\boldsymbol{\beta}) = S_0(t)^{\exp(\mathbf{X}^\top\boldsymbol{\beta})}$ in (3). For this purpose, we need, in addition to $\widehat{\boldsymbol{\beta}}$, an estimate of the unspecified baseline survival function S_0. Estimating S_0 is typically done non-parametrically, for instance by using a Nelson-Aalen or Kaplan-Meier estimator, which are calculated using $\{(Y_i,\delta_i)\}_{i=1}^n$ [Add reference]. With an estimate $\widehat{S}_0$ of S_0, we can estimate the survival function S via

$$\widehat{S}(t,\mathbf{X},\widehat{\boldsymbol{\beta}}) = \widehat{S}_0(t)^{\exp(\mathbf{X}^\top\widehat{\boldsymbol{\beta}})}.$$

2.4 What if the Times at Risk are not Unique?

Consider a situation where, for individuals for which $\delta_i = 1$, some of the times at risk Y_i are not unique. Breslow (1975) and Efron (1977) propose two different approaches for this situation. In the following, we briefly discuss the approach of Breslow (1975).

Let the sets $\mathcal{D}_i = \{j \in [n] : Y_j = Y_i\}$ contain the observations whose times at risk are tied with the one of individual i. Then, the likelihood function L in (4) becomes

$$L(\beta) = \prod_{\{i\in[n]:\delta_i=1\}} \frac{\sum_{\{j\in\mathcal{D}_i\}} \exp(\mathbf{X}_j^\top\boldsymbol{\beta})}{\left(\sum_{\{j\in[n]:Y_j\geq Y_i\}} \exp(\mathbf{X}_j^\top\boldsymbol{\beta})\right)^{|\mathcal{D}_i|}}$$

and the optimization problem in (5) is adapted correspondingly.

3 Competing Risk Modeling

There might be several causes for an individual to fail. Suppose that there are K failure types/causes in total and that there exist variables ε_i that indicate the cause of failure of individual i. Without loss of generality, assume that ε_i have support on the set $\{1, \ldots, K\}$ and $\varepsilon_i = k$ means that individual i fails due to cause k. In practice, we observe the variables $\delta_i \varepsilon_i$. Hence, if individual i survives, we observe $\delta_i \varepsilon_i = 0$, whereas if individual i fails before the censoring time, we observe $\delta_i \varepsilon_i = \varepsilon_i$. Obviously, $\delta_i \varepsilon_i$ is supported on $\{0, 1, \ldots, K\}$. Models in which there are multiple causes of failure are referred to as *competing risk models*. In such models, we observe the random sample $\{(X_i, Y_i, \delta_i, \delta_i \varepsilon_i)\}_{i=1}^n$.

In situations with competing risk, one is typically only interested in one single cause of failure. For instance, in a trial on cardiovascular diseases, one is typically only interested in cardiovascular deaths and not in non-cardiovascular deaths (some individuals in the trial might die of causes other than cardiovascular diseases). A naive approach in such a situation is to artificially set the time at risk of all non-cardiovascular deaths equal to the censoring time, thereby effectively counting them as survivors. However, this approach may produce upward biased estimates of incidence function F, which will in turn lead to downwards biased estimates of the survival function S (e.g. Austin et al., 2016).

Approaches that provides accurate estimates of incidence and survival despite the presence of competing risks typically make use of *Cumulative Incidence Functions*. Unlike incidence functions F, cumulative incidence functions consider each failure type separately. Hence, if there are K types of failure, there are K cumulative incidence functions, denoted F_k, for $k = 1, \ldots, K$. Mathematically, for some time $t \in [0, \infty]$, the *cumulative incidence function of failure type* $k \in \{1, \ldots, K\}$ is defined by

$$F_k(t) = \mathbb{P}[T_1 \leq t, \varepsilon_1 = k].$$

This definition[1] gives rise to the following decomposition of incidence function F:

$$F(t) = \mathbb{P}[T_1 \leq t] = \sum_{k=1}^{K} \mathbb{P}[T_1 \leq t, \varepsilon_1 = k] = \sum_{k=1}^{K} F_k(t).$$

Hence, for the survival function S it holds that

$$S(t) = 1 - F(t) = 1 - \sum_{k=1}^{K} F_k(t).$$

We emphasize that for survivors (for which $\delta_i \varepsilon_i = 0$), no cumulative incidence function is considered.

[1] A cumulative incidence function F_k with $K > 1$ is not a distribution function, because $\lim_{t \to +\infty} F_k(t) \neq 1$.

There are two main ways of specifying competing risk models, *proportional cause-specific hazard* modedels. and *proportional subdistribution hazard* models.

3.1 Proportional Cause-Specific Hazard Models

Proportional cause-specific hazard models essentially fit K separate Cox proportional hazard models. Hence, for cause $k \in [K]$, the hazard function associated with this cause is given by

$$\begin{aligned} h^k_{\beta^k}(t|X_i) &= \lim_{\Delta \downarrow 0} \frac{\mathbb{P}[t \leq T_i < t + \Delta, \varepsilon_i = k | T_i \geq t]}{\Delta} \\ &= h^k_0(t) \exp(X^\top \beta^k), \end{aligned}$$

where h_0^k is a Breslow-type of the baseline hazard of individuals for which $\varepsilon_i = k$. We use the k-superscript in $\beta^k \in \mathbb{R}^p$ to remind us that the coefficients of each of the K. Consequently, we estimate β^k by solving the problem (**??**) on individuals for which $\varepsilon_i = k$. We can subsequently estimate the associated survival function $S(t) = S_{(\beta^1,\dots,\beta^K)}(t)$ by

$$\hat{S}(t) = S_{(\hat{\beta}^1,\dots,\hat{\beta}^K)}(t) = \prod_{k=1}^{K} S^k_0(t)^{\exp(X^\top \widehat{\beta}_k)}$$

make dependence on X explicit: we need info on t and X for prediction! We need to estimate beta K times

References

Austin, P. C., Lee, D. S., and Fine, J. P. (2016). Introduction to the analysis of survival data in the presence of competing risks. *Circulation*, 133(6):601–609.

Breslow, N. E. (1975). Analysis of survival data under the proportional hazards model. *International Statistical Review*, 43(1):45–57.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72(359):557–565.

Simon, N., Friedman, J. H., Hastie, T., and Tibshirani, R. (2011). Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *Journal of Statistical Software*, 39(5):1–13.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67:301–320.