Notes on survival analysis

Erwann Rogard

2022-04-04

Abstract

These are renovated notes from 2007. Some parts are commented out in the source file until I get a grasp of what I was up to.

1 Conventions

The most general framework that we will consider is that of staggered entry, static covariate and right censoring. Let's begin with the following definitions

Conventions		
t^{cal}		calendar time
t^{in}		entry time
t^{out}		exit time
$t^{\dagger}$		failure time
$t^{+}(t)$	$(t - t^{\text{in}})^{+}$	time on study
$t^{\text{c}}(t)$	$t^{\text{out}} \wedge t^{+}(t)$	censoring time
$t^{*}(t)$	$t^{\dagger} \wedge t^{\text{c}}(t)$	first event time
$\delta(t)$	$1\{t^{\dagger} < t^{\text{c}}(t)\}$	failure observed
$R(t,s)$	$1\{s \leq t^{*}(t)\}$	at risk
$N(t,s)$	$\delta(t)1\{t^{\dagger} \leq s\}$	observed failure prior to
$x^{\text{nuis}}(t)$		nuisance covariate
$z(t)$		treatment indicator
$x(t)$	$(z(t), x^{\text{nuis}}(t))$	covariate
$y(t)$	$(t^{*}(t), \delta(t))$	response
n		index for order of entry
$n_{*}(t)$		count of entries
$n_{\dagger}(t)$	$\sum_{\{n \leq n_{*}(t)\}} \delta_n(t)$	count of observed failures
$\mathbf{x}(t)$	$(x_1, ..., x_{n_{*}(t)})$	covariate data
$\mathbf{y}(t)$	$(y_1, ..., y_{n_{*}(t)})$	response data
$D(t)$	$D(t) = (\mathbf{x}(t), \mathbf{y}(t))$	data
β_{nuis}		effect of nuisance covariate
θ		treatment effect
β	$(\theta, \beta_{\text{nuis}})$	covariate effect
ϕ		baseline hazard parameter
ξ		all parameters

Note that t^{in} is measured on the same scale as t^{cal} whereas t^{out} and $t^{\dagger}$ are clocks that are started at t^{in}. In principle, we need to define a joint distribution for $(t^{\text{in}}, t^{\dagger}, t^{\text{out}})$, but in the definition of D we are implicitly treating $(t^{\text{in}}, t^{\text{out}})$ as ancillary variables. The staggered entry and static covariate assumptions imply $x(t) = x(t^{\text{in}}), t \geq t^{\text{in}}$, which justifies our definition of D. When the context specifies that we fix $t^{\text{cal}} = t$, we ommit t in all expressions that depend on it, e.g. we write D instead of $D(t)$.

1.1 Likelihood and derived quantities

We restrict the class of model, either parametric or semi–parametric, to the proportional hazard. It is standard convention that $f(.)$, $F(.)$, $S(.)$, $h(.)$ and $H(.)$ denote the PDF and CDF of time–to-event, survival, hazard and cumulative hazard functions, respectively. The relations between them are given by

$$S(t) = 1 - F(t) = \exp\left(-\int_0^t h(u)\right) du = \exp(-H(t)) \tag{1}$$

and $F(t) = \int_0^t f(u)du$. The proportional hazard assumption made at the beginning is given by $h(t|x) = h_0(t)\exp(x'\beta)$, which together with (1), implies $S(t) = S_0(t)^{\exp(x'\beta)}$. The name derives from the property that $h(t|x_0)/h(t|x_1)$, $x_0 \neq x_1$, is independent of t. For the particular case $x = z$, $\forall t$,

$$\begin{aligned} \theta &= \log(h(t|z=1)/h(t|z=0)) && (2) \\ &= \log\left(\log(S(t|z=1))/\log(S(t|z=0))\right) && (3) \end{aligned}$$

From the first equality θ is often referred to as the log hazard ratio. The second is useful in the Bayesian process of prior elicitation from expert knowledge.

Suppose we postulate a parametric family for $F(.)$ indexed by $\xi \in \Xi$. For example, in the case $t \sim \mathcal{W}(t|\alpha,\gamma)$, $F(t) = 1 - \exp(-\gamma t^{\alpha})$, and $h(t) = \alpha\gamma t^{\alpha-1}$, so that $\xi = (\alpha, \gamma)$. Furthermore, if $\gamma = \exp(x'\beta)$, then $h(.|\xi, x)$ is a proportional hazard with baseline $h_0(t|\phi) = \alpha t^{\alpha-1}$, $\phi \equiv \alpha$. Suppose we fix $t^{\text{cal}} = t$. The likelihood, in terms of ξ is

$$L(D|\xi) = \prod_{\{n:\delta_n=1\}} f(t_n^{\dagger}|\xi) \prod_{\{n:\delta_n=0\}} S(t_n^{\text{c}}|\xi) \tag{4}$$

The proportional hazard also permits an important semi–parameric formulaulation under which $h_0(.)$ is unspecified and the following partial likelihood[1] may be justified:

$$\begin{aligned} l_{\text{Cox}}(t|\beta) &= \sum_n \int_{[0,t]} \log\left(\frac{\exp(x_n'\beta)}{\sum_l R_l(t,s)\exp(x_l\beta)}\right) N_n(t, ds) && (5) \\ &= \sum_n \delta_n(t)\left(x_n'\beta - \log\left(\sum_l R_l(t, t_n^{\dagger})\exp(x_n'\beta)\right)\right) && (6) \end{aligned}$$

The term inside the parenthesis of (6) can be interpreted as the probability that failure is on individual n, conditional on failure time equating $t_n^\dagger$ and the risk set $\{l : R_l(t, t_n^\dagger) = 1\}$. The score and information matrix (adapted from [3], Chapter 7) are

$$
\begin{aligned}
U_{\text{Cox}}(t|\beta) &= \nabla_\beta l_{\text{Cox}}(t|\beta) && (7)\\
&= \sum_n \delta_n(t)(x_n - \bar{x}(t, t_n^\dagger|\beta)) && (8)\\
I_{\text{Cox}}(t|\beta) &= -\partial^2_{\beta,\beta'} l_{\text{Cox}}(t|\beta) && (9)\\
&= \sum_n \delta_n(t) \frac{\sum_l R_l(t, t_n^\dagger) \exp(x_l'\beta)(x_l - \bar{x}(t, t_n^\dagger|\beta))(x_l - \bar{x}(t, t_n^\dagger|\beta))'}{\sum_l R_l(t, t_n^\dagger) \exp(x_l'\beta)} && (10)
\end{aligned}
$$

where

$$\bar{x}(t, t_n^\dagger|\beta) = \frac{\sum_l R_l(t, t_n^\dagger) x_l \exp(x_l'\beta)}{\sum_l R_l(t, t_n^\dagger) \exp(x_l'\beta)} \tag{11}$$

For practical purposes, we may treat the partial likelihood as a standard likelihood so that for fixed t but sufficiently large n_*, the following approximations hold:

$$
\begin{aligned}
n^{-1/2} U_{\text{Cox}}(t|\beta) &\sim N(0, n^{-1} I_{\text{Cox}}(t|\beta)) && (12)\\
\hat{\beta}_{\text{Cox}}(t|\beta) &\sim N(\beta, I^{-1}_{\text{Cox}}(t|\hat{\beta}_{\text{Cox}})) && (13)
\end{aligned}
$$

where $\hat{\beta}_{\text{Cox}}$ solves the estimating equation $U_{\text{Cox}}(t|\hat{\beta}_{\text{Cox}}) = 0$. The second line follows from $U_{\text{Cox}}(t|\hat{\beta}_{\text{Cox}}) \approx U_{\text{Cox}}(t|\beta) + (\hat{\beta} - \beta) I_{\text{Cox}}(t|\beta)$.

For the particular case $x \equiv z$, $U_{\text{Cox}}(t|\theta = 0)$ is an estimate for the observed$-$expected number of events in a treatment group, and it is called the log–rank test statistic. Let $\tilde{U}_{\text{Cox}}(t|\theta) = I^{-1/2}_{\text{Cox}}(t|\theta) U_{\text{Cox}}(t|\theta)$. It underpins the popular test of equality between two lifetime distributions: reject if $|\tilde{U}_{\text{Cox}}(t|\theta = 0)| > c_\alpha$. Let $\tilde{U}_{\theta_*,\text{Cox}}(t|\theta) = I^{-1/2}_{\text{Cox}}(t|\theta_*) U_{\text{Cox}}(t|\theta)$. This quantity is useful for sequential analysis. Specifically, if $i = 1, ..., i_*$ indexes interim analyses, according to Chapter 9 of [4], for θ close to 0, $\tilde{U}_{\theta_*,\text{Cox}}(t_1, ..., t_{i_*}|\theta) = \{\tilde{U}_{\theta_*,\text{Cox}}(t_1|\theta), ..., \tilde{U}_{\theta_*,\text{Cox}}(t_{i_*}|\theta)\}$ is a normal vector such that

$$
\begin{aligned}
\tilde{U}_{\theta_*,\text{Cox}}(t_i|\theta) &\sim N(I^{1/2}_{\text{Cox}}(t_i|\theta_*)\theta, 1)) && (14)\\
\text{Cov}(\tilde{U}_{\theta_*,\text{Cox}}(t_i|\theta), \tilde{U}_{\theta_*,\text{Cox}}(t_{i+k}|\theta))|\theta_* &= I^{1/2}_{\text{Cox}}(t_i|\theta_*) I^{-1/2}_{\text{Cox}}(t_{i+k}|\theta_*) && (15)
\end{aligned}
$$

todo: make contiguity argument precise (Chapters 7&9[**Vaart1998**] and [2]) **todo**Although [4] use $\theta_* = 0$ would it not be better to take $\theta_* = \hat{\theta}$, the mle?

Formulations that are intermediary between the fully parameterized version of $L(.|\xi)$ and the partial likelihood $L_{\text{Cox}}(.|\beta)$ include the piecewise constant hazard model, such that $\xi = (\lambda, \beta)$ where λ is a vector of constants.

Revisions

The commits referred to are at `https://github.com/erwannr/statistics/survanal/survanal.tex`

Date	Comment	Commit
2007-05-14	First version; section of a larger document	`3811ec4`
2022-04-04	Added headers and bibliography; commented out some parts	

Bibliography

[1] T. Sellke and D. Siegmund. "Sequential analysis of the proportional hazards model". In: *Biometrika* 70.2 (Aug. 1983), pp. 315–326. ISSN: 0006-3444. DOI: 10.1093/biomet/70.2.315. eprint: https://academic.oup.com/biomet/article-pdf/70/2/315/632076/70-2-315.pdf.

[2] Yannis Bilias, Mingao Gu, and Zhiliang Ying. "Towards A General Asymptotic Theory For Cox Model With Staggered Entry". In: *Annals of Statistics* 25.2 (1997), pp. 662–682.

[3] Jerald F. Lawless. *Statistical Models and Methods for Lifetime Data, Second Edition*. Wiley Series in Probability and Statistics. Wiley, Nov. 2002. ISBN: 9781118033005. DOI: 10.1002/9781118033005.

[4] Joseph G. Ibrahim, Ming-Hui Chen, and Debajyoti Sinha. *Bayesian Survival Analysis*. Springer, Dec. 2004. ISBN: 9780387952772.