EasyReg Module SURVIVAL2

1 Introduction

EasyReg module SURVIVAL2 estimates a proportional hazard model for a duration T, conditional on a vector of covariates, **without** unobserved heterogeneity.

The general form of the conditional survival function involved is:

$$P[T > t|X] = \exp\left(-\exp(\beta' X)\int_0^t \lambda(\tau|\alpha)d\tau\right),$$

where $\exp(\beta' X)$ is the systematic hazard, $\lambda(t|\alpha)$ is the baseline hazard function depending on a parameter (vector) α, and $X \in \mathbb{R}^k$ is a vector of covariates which should **not** include a constant.

Recall that the conditional hazard function $\widetilde{\lambda}(t|X) = \exp(\beta' X)\lambda(t|\alpha)$ is defined as

$$\widetilde{\lambda}(t|X) = \lim_{\delta \downarrow 0} \frac{P[T \in [t, t+\delta) \mid X, T \geq t]}{\delta}.$$

The survival function involved can be written as

$$S(t|\alpha, \beta' X) = \exp(-\exp(\beta' X)\Lambda(t|\alpha))$$

where

$$\Lambda(t|\alpha) = \int_0^t \lambda(\tau|\alpha)d\tau$$

is the integrated baseline hazard.

2 Interval censoring

It is assumed that the duration T is not observed directly, but only in the form of dummy variables corresponding to intervals: Let $0 \leq b_0 < b_1 < ... < b_M$. Then the dependent variables are M dummy variables

$$D_i = I(T \in (b_{i-1}, b_i]), i = 1, ..., M,$$

where $I(.)$ is the indicator function: $I(true) = 1$, $I(false) = 0$. The intervals $(b_{i-1}, b_i]$ may be replaced by $[b_{i-1}, b_i)$. However, these intervals should be disjoint. If the duration T is censored from above as well, the value of b_M should be such that for all observations, T is not censored if $T \in (0, b_M]$.

It is recommended to indicate the bracket values in the name of D_i, for example, use names like "Dummy T in (2,4]" or "I(T in (2,4])". EasyReg will then extract the lower and upper bounds of the bracket from the name.

Note that for $i = 1, 2, ..., M$,

$$P[D_i = 1|X] = S\left(b_{i-1}|\alpha, \beta' X\right) - S\left(b_i|\alpha, \beta' X\right)$$

and

$$P\left[\sum_{i=1}^{M} D_i = 0 \middle| X\right] = S\left(b_M|\beta' X\right).$$

3 Hazard function options

3.1 Piecewise linear integrated baseline hazard

Because these probabilities depend on the values of $\Lambda(b_i|\alpha)$ in M bracket points b_i only, we may without loss of generality parametrize $\Lambda(t|\alpha)$ as a piecewise linear function:

$$\begin{aligned}
\Lambda(t|\alpha) &= \Lambda(b_{i-1}|\alpha) + \alpha_i\left(t - b_{i-1}\right) \qquad (1)\\
&= \sum_{k=1}^{i-1} \alpha_k\left(b_k - b_{k-1}\right) + \alpha_i\left(t - b_{i-1}\right) \text{ for } t \in (b_{i-1}, b_i],\\
\alpha_i &> 0 \text{ for } i = 1, ..., M,\ \alpha = (\alpha_1, ..., \alpha_M)' \in \mathbb{R}^M.
\end{aligned}$$

There are other equivalent ways to specify $\Lambda(b_i|\alpha)$, but the advantage of the specification (1) is that the null hypothesis $\alpha_1 = ... = \alpha_M$ corresponds to the constant baseline hazard $\lambda(t|\alpha) = \alpha_1$. In that case $\exp\left(\beta' X\right)\Lambda(t|\alpha) = \exp\left(\ln(\alpha_1) + \beta' X\right)t$, so that $\ln(\alpha_1)$ acts as a constant term in the systematic hazard.

The specification (1) is adopted in EasyReg as the default option, and is called the piecewise linear integrated baseline hazard.

Next to this specifications, you have four other options for the baseline hazard:

3.2 Weibull baseline hazard

$$\lambda(t|\alpha) = \alpha_1\alpha_2 t^{\alpha_2 - 1}, \tag{2}$$
$$\alpha_1 > 0,\ \alpha_2 > 0,\ \alpha = (\alpha_1, \alpha_2)'.$$
$$\Lambda(t|\alpha) = \int_0^t \lambda(\tau|\alpha)d\tau = \alpha_1 t^{\alpha_2}.$$

The reason for the scale factor α_1 here and below is that X does not contain a constant, hence $\ln(\alpha_1)$ plays the role of constant: $\exp(\beta' X)\alpha_1 t^{\alpha_2} = \exp(\ln(\alpha_1) + \beta' X)t^{\alpha_2}$.

3.3 Generalized Weibull baseline hazard

$$\lambda(t|\alpha) = \alpha_1\alpha_2 \left(\alpha_3 + t\right)^{\alpha_2 - 1}, \tag{3}$$
$$\alpha_1 > 0,\ \alpha_2 > 0,\ \alpha_3 > 0,\ \alpha = (\alpha_1, \alpha_2, \alpha_3)'$$
$$\Lambda(t|\alpha) = \int_0^t \lambda(\tau|\alpha)d\tau = \alpha_1 \left(\left(\alpha_3 + t\right)^{\alpha_2} - \alpha_3^{\alpha_2}\right).$$

3.4 Unimodal baseline hazard

$$\lambda(t|\alpha) = \frac{2\alpha_1 t}{\alpha_2^2 + t^2},\ \alpha_2 = \arg\max_{t\geq 0} \lambda(t|\alpha), \tag{4}$$
$$\alpha_1 > 0,\ \alpha_2 > 0,\ \alpha = (\alpha_1, \alpha_2)'.$$
$$\Lambda(t|\alpha) = \int_0^t \lambda(\tau|\alpha)d\tau = \alpha_1. \ln\left(\frac{\alpha_2^2 + t^2}{\alpha_2^2}\right).$$

3.5 Generalized unimodal baseline hazard

$$\lambda(t|\alpha) = \frac{2\alpha_1 \left(\alpha_3 + t\right)}{\left(\alpha_2 + \alpha_3\right)^2 + \left(\alpha_3 + t\right)^2},\ \alpha_2 = \arg\max_{t\geq 0} \lambda(t|\alpha), \tag{5}$$
$$\alpha_1 > 0,\ \alpha_2 > 0,\ \alpha_3 > 0,\ \alpha = (\alpha_1, \alpha_2, \alpha_3)'.$$
$$\Lambda(t|\alpha) = \int_0^t \lambda(\tau|\alpha)d\tau = \alpha_1. \ln\left(\frac{\left(\alpha_2 + \alpha_3\right)^2 + \left(\alpha_3 + t\right)^2}{\left(\alpha_2 + \alpha_3\right)^2 + \alpha_3^2}\right).$$

4 Two-step ML estimation

EasyReg estimates the parameter vectors α and β in two steps. In first instance the parameters α_i are fixed to $\alpha_i = 1$, except that in the cases (4) and (5) $\alpha_2 = b_1$.

The quasi maximum likelihood estimator $\widetilde{\beta}_0$ of β in the first step will be used as starting values in the second step, together with the initial values of α_i. This step yields the maximum likelihood estimators $\widehat{\alpha}$ of α and $\widehat{\beta}$ of β.