Survival Analysis I (CHL5209H)

Olli Saarela

Dalla Lana School of Public Health
University of Toronto

olli.saarela@utoronto.ca

January 16, 2019

Parametrized hazard function

- Often we are interested in modeling the relationship between survival and some covariates X_i.
- This dependency is parametrized through a vector of parameters θ.
- The resulting individual-level hazard function may be defined as

$$\begin{aligned}\lambda_i(t;\theta) &\equiv \lim_{h\to 0} \frac{P(t \leq \tilde{T}_i < t + h \mid \tilde{T}_i \geq t, x_i; \theta)}{h} \\ &= \frac{f_i(t;\theta)}{S_i(t;\theta)},\end{aligned}$$

  where f_i and S_i are the density and survival functions characterizing the latent event time distribution.
- How to estimate θ in the presence of censoring?

Bring in the censoring times

- We can write further that

$$
\begin{aligned}
\lambda_i(t;\theta)\,\mathrm{d}t &= \frac{f_i(t;\theta)\,\mathrm{d}t}{S_i(t;\theta)} \\
&= \frac{P(t \leq \tilde{T}_i < t + \mathrm{d}t \mid x_i;\theta)}{P(\tilde{T}_i \geq t \mid x_i;\theta)} \\
&= \frac{P(C_i > t \mid x_i;\theta)P(t \leq \tilde{T}_i < t + \mathrm{d}t \mid x_i;\theta)}{P(C_i \geq t \mid x_i;\theta)P(\tilde{T}_i \geq t \mid x_i;\theta)}.
\end{aligned}
$$

Independent censoring

- Assume $\tilde{T}_i \perp\!\!\!\perp C_i \mid X_i$, that is, given the covariates X_i, the latent event and censoring times are independent:

$$
\begin{aligned}
\lambda_i(t;\theta)\,\mathrm{d}t &= \frac{P(C_i > t \mid x_i;\theta) P(t \leq \tilde{T}_i < t + \mathrm{d}t \mid x_i;\theta)}{P(C_i \geq t \mid x_i;\theta) P(\tilde{T}_i \geq t \mid x_i;\theta)} \\
&= \frac{P(t \leq \tilde{T}_i < t + \mathrm{d}t, C_i > t \mid x_i;\theta)}{P(\tilde{T}_i \geq t, C_i \geq t \mid x_i;\theta)} \\
&= \frac{P(t \leq T_i < t + \mathrm{d}t, E_i = 1 \mid x_i;\theta)}{P(T_i \geq t \mid x_i;\theta)} \\
&= P(t \leq T_i < t + \mathrm{d}t, E_i = 1 \mid T_i \geq t, x_i;\theta).
\end{aligned}
$$

- The last form characterizes a hazard function defined in terms of the observed times, rather than the latent event times.
- Interpretation: under random censoring, we can make inferences on θ based on the observed follow-up data.

Revisit the counting process notation

- How to express the same assumption usign the counting process framework?
- Introduce the observed counting process $\{N_i(t), t \geq 0\}$ defined through

$$N_i(t) = \mathbf{1}_{\{T_i \leq t, E_i = 1\}},$$

and the following notation for all observed data on individuals $i = 1, \ldots, n$ just before time t:

$$\mathcal{F}_{t^-} = \sigma(\{X_i, N_i(u), Y_i(u) : i = 1, \ldots, n, 0 \leq u < t\}).$$

- The independent censoring assumption can now be characterized through

$$P(\mathrm{d}N_i(t) = 1 \mid \mathcal{F}_{t^-}; \theta) = Y_i(t)\lambda_i(t; \theta)\,\mathrm{d}t.$$

Likelihood contributions

- The observed data consist of realizations of random vectors (T_i, E_i, X_i), $i = 1, \ldots, n$.
- Conditional on X_i, the likelihood contribution for a non-censored individual i with $E_i = 1$ is

$$\begin{aligned} & P(t_i \leq T_i < t_i + \mathrm{d}t, E_i = 1 \mid x_i; \theta) \\ & = P(C_i > t_i \mid x_i; \theta) P(t_i \leq \tilde{T}_i < t_i + \mathrm{d}t \mid x_i; \theta) \\ & = P(C_i > t_i \mid x_i; \theta) \lambda_i(t_i; \theta) \,\mathrm{d}t\, S_i(t_i; \theta). \end{aligned}$$

- Similarly, the likelihood contribution for a censored individual i with $E_i = 0$ is

$$\begin{aligned} & P(t_i \leq T_i < t_i + \mathrm{d}t, E_i = 0 \mid x_i; \theta) \\ & \qquad = P(t_i \leq C_i < t_i + \mathrm{d}t \mid x_i; \theta) P(\tilde{T}_i > t_i \mid x_i; \theta) \\ & \qquad = P(t_i \leq C_i < t_i + \mathrm{d}t \mid x_i; \theta) S_i(t_i; \theta). \end{aligned}$$

Non-informative censoring

- Assume further that the distribution of the latent censoring time does not involve θ.
- We can now write the full likelihood for θ as

$$
\begin{aligned}
&\prod_{i=1}^{n} P(t_i \leq T_i < t_i + \mathrm{d}t, E_i = e_i \mid x_i; \theta) \\
&\overset{\theta}{\propto} \prod_{i=1}^{n} \left[\lambda_i(t_i; \theta)^{e_i} S_i(t_i; \theta)\right] \\
&= \prod_{i=1}^{n} \left[\lambda_i(t_i; \theta)^{e_i} \exp\left\{-\int_0^{t_i} \lambda_i(u; \theta)\,\mathrm{d}u\right\}\right].
\end{aligned}
$$

- Note that with these assumptions, the likelihood is fully specified by the hazard function for the event of interest.
- Maximum likelihood criterion can now be applied as usual to estimate θ.

Interpretation?

- What did we have to assume to obtain the previous likelihood expression?
- Independent and non-informative censoring are rather abstract properties.
- Note that the independence of the censoring mechanism was conditional on the covariates X_i.
- Such independence is more believable if we can condition on all common determinants of the censoring events and the events of interest.

Piecewise constant hazard model

Parametric survival models

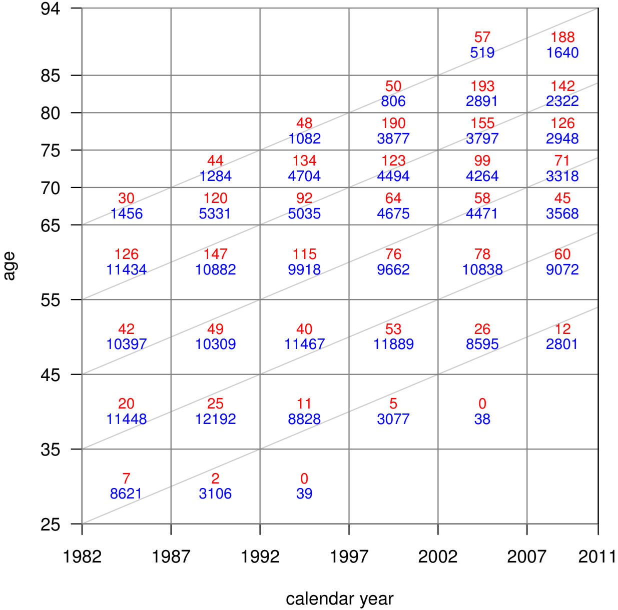
- We can parametrize the hazard function through a regression equation.
- For example,

$$\lambda_i(u;\theta) = \exp\{\alpha + \beta' X_i\},$$

where $\theta = (\alpha, \beta)$, would specify a Poisson regression model, with the baseline rate given by $\exp\{\alpha\}$ and the regression coefficients having interpretation as log-rate ratios (this is a special case of a proportional hazards model).

- Usually, we would not want to assume the hazard to be constant over time.
- A generalization of this model is obtained if we assume that hazard to be constant over pre-specified intervals.
- This also allows us to easily incorporate more than one time scale.

Lexis diagram



Notation for grouped follow-up data

- The Lexis diagram depicted the follow-up for total mortality of 9029 individuals recruited as a cross-sectional cohort in 1982 (then of age 25-65) until the end of year 2010.
- Assume that the mortality rate is constant within the agegroups $k = 1, \dots, 9$ in the Lexis diagram, and within one-year calendar time intervals $l = 1, \dots, 29$.
- Let $d_{ikl} \in \{0, 1\}$ denote whether individual i died at age k in year l.
- Let y_{ikl} denote the person-years individual i contributed in age group k and year l.
- If we have no other individual level information, the hazard rate of any individual i in age group k and year l is assumed to be λ_{kl}.
- This is why the model is called piecewise constant.

Likelihood

- The likelihood contribution of individual i is given by

$$\prod_{k=1}^{9}\prod_{l=1}^{29}\lambda_{kl}^{d_{ikl}}\exp\left\{-\sum_{k=1}^{9}\sum_{l=1}^{29}\int_{0}^{y_{ikl}}\lambda_{kl}\,\mathrm{d}t\right\}$$
$$=\prod_{k=1}^{9}\prod_{l=1}^{29}\left[\lambda_{kl}^{d_{ikl}}\exp\left\{-\lambda_{kl}y_{ikl}\right\}\right].$$

- The likelihood expression from n individuals is then

$$\prod_{i=1}^{n}\prod_{k=1}^{9}\prod_{l=1}^{29}\left[\lambda_{kl}^{d_{ikl}}\exp\left\{-\lambda_{kl}y_{ikl}\right\}\right].$$

Connection to Poisson model

- Since λ_{kl} does not depend on i, we get

$$\prod_{i=1}^{n}\prod_{k=1}^{9}\prod_{l=1}^{29}\left[\lambda_{kl}^{d_{ikl}}\exp\left\{-\lambda_{kl}y_{ikl}\right\}\right]$$
$$=\prod_{k=1}^{9}\prod_{l=1}^{29}\left[\lambda_{kl}^{\sum_{i=1}^{n}d_{ikl}}\exp\left\{-\lambda_{kl}\sum_{i=1}^{n}y_{ikl}\right\}\right].$$

- We would get the same likelihood expression if we assume the total number of deaths in each age group/year to be independently Poisson distributed as

$$\sum_{i=1}^{n}d_{ikl}\sim\mathrm{Poisson}\left(\lambda_{kl}\sum_{i=1}^{n}y_{ikl}\right).$$

- Thus, the model can be fitted using any available Poisson regression software such as the `glm` function in R.

More general models

- We can allow the piecewise constant mortality rates to further depend on individual-level covariates X_i, in which case the likelihood expression is of the form

$$\prod_{i=1}^{n} \prod_{k=1}^{9} \prod_{l=1}^{29} \left[\lambda_{ikl}^{d_{ikl}} \exp \{ -\lambda_{ikl} y_{ikl} \} \right].$$

- While there are no Poisson distributed counts here, the model can still be fitted as a Poisson regression. (Why?)

Model parametrization

- Even without further individual-level characteristics, the example model involved $9 \times 29 = 261$ mortality rate parameters λ_{kl}.
- Not all of these can be estimated, since some age group/year combinations have no observed deaths. (Why?)
- Estimating this many parameters would also be inefficient, and interpretation of the results would be difficult.
- Suppose that we are mainly interested in the calendar time trend in mortality, while removing the age effect.
- Further, we allow the mortality rate to depend covariates sex ($x_{i1} \in \{0, 1\}$, men/women) and region ($x_{i2} \in \{0, 1\}$, eastern/western Finland).
- A more parsimonious parametrization can now be specified through a regression equation.

Regression equation

- For example, we can specify the model as

$$\log(\lambda_{ikl}) = \alpha_k + \beta_l + \gamma_1 x_{i1} + \gamma_2 x_{i2}.$$

- This model involves only $9 + 29 + 2 = 40$ parameters.
- Now we are mainly interested in the calendar time effect parameters β_l, $l = 1, \ldots, 29$.
- Adjustment for age through the parameters α_k, $k = 1, \ldots, 9$ is needed to exctract the calendar time effect. (What would happen to the calendar time effect if we did not adjust for age?)
- Note that for this model to be identifiable, one parameter restriction, for example

$$\alpha_1 = 0$$

is needed.

Parametrization with an intercept term

- An alternative way to parametrize the same model would be
$$\log(\lambda_{ikl}) = \mu + \alpha_k + \beta_l + \gamma_1 x_{i1} + \gamma_2 x_{i2},$$
which includes a separate intercept term μ.
- Now two parameter restrictions are required, for example
$$\alpha_1 = 0 \text{ and } \beta_1 = 0.$$
- The interpretation of the calendar time effects β_l, $l = 2, \ldots, 29$ is now different, they represent log-rate ratios w.r.t. the first year.