Part IIB: Parametric survival models

- Why survival regression models?
- Types of survival regressions models
  - Relative risk models (today)
  - Accelerated failure time models (later)
- Proportional hazards models
  - Parameterisation and properties
  - Exponential regression
  - Weibull regression
  - General proportional hazards model
- Survival likelihood

Survival regression models

- Regression models are used to investigate how different explanatory variables (covariates) affect the failure rates (hazards) or the failure times
- The two main types of regression models differ in whether the explanatory variables affect (multiplicatively) on hazards or failure times:
  - relative risk regression models
  - accelerated failure time models
- We will first consider proportional hazards models
  - The simplest of these, the exponential and Weibull regression models can be interpreted as accelerated failure time models, too
  - In general, however, the two classes of models are different

Survival regression models (2)

- Why survival regression models?
  - Why not model mean time-to-event as a function of covariates using linear regression?
    - Answer: This would ignore censoring and more complicated patterns of missing data
  - Why not model proportion of events using e.g. logistic regression?
    - Answer: This would ignore time
- Regression models are classified also according to how they are parameterised:
  - Parametric (this lecture)
  - Semiparametric (e.g. Cox regression model)
  - Nonparametric (e.g. Kaplan-Meier and Nelson-Aalen estimates)

Relative risk regression models

A family of relative risk regression models is written as

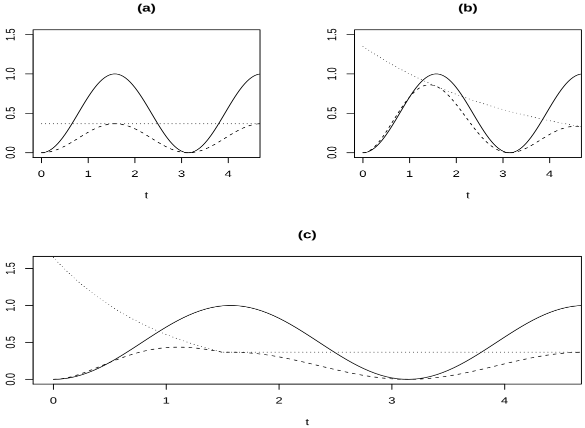
$$\lambda(t;z) = \lambda_0(t)r(t,z)$$

where

- $\lambda_0(t)$ is an arbitrary uspecified baseline hazard and
- the relative risk (ratio of the hazards or rate raio) $r(t,z)$ specifies the relationship between the covariate z and the failure rate.

The most common exponential form of the relative risk $r(t,x) = \exp\{z'(t)\beta\}$ where $z(t)$ are the possible time-dependent covariates. Note that the baseline hazard corresponds to $z(t) = (0,0,\cdots,0)$. This gives Cox regression model.

Relative risk regression models



Hazard and relative risk for two groups ($z = 0$, $z = 1$)
(a) proprtional hazards with RR $= \exp\{\beta z\}$ and $\beta = -1$;
(b) model with interaction between x and time RR $= \exp\{\beta_1 x + \beta_2 x t\}$, $\beta_1 = 0.3$, $\beta_2 = -0.3$;

(c) model with high initial risk RR $= \exp\{\beta_1 x + \beta_2 x(1.5 - t)I(t \leq 1.5)\}$, $\beta_1 = -1$, $\beta_2 = 1$.

Proportional hazards model

- Special case of relative risk regression model when covariates are fixed
- Consider comparing the hazard rates between two groups
  - the rate in the control group $\lambda_0(t)$
  - the rate in the treated group $\lambda_1(t)$
- The proportional hazards assumption says that the ratio of the two hazards is constant for *all* $t \geq 0$:

$$\frac{\lambda_1(t)}{\lambda_0(t)} = \theta = \text{ constant for all } t$$

- The model can thus be parameterised in terms of the baseline rate $\lambda_0(t)$ and a constant hazard ratio θ
- With a binary (0/1) explanatory variable Z, we can thus write

$$\lambda(t; \theta, Z) = \theta^Z \lambda_0(t)$$

Proportional hazards model (2)

- The hazard ratio (relative rate) needs to be positive, i.e. $\theta \geq 0$
- Technically, this is taken care by writing $\theta = \exp(\beta)$ where β is the logarithmic relative rate
  - Statistical programmes typically provide estimate of "beta" parameters, i.e. log-parameters so you might need to transform them back to the original scale
  - This means that once you have an estimate of β, the actual estimate you want to report is $\exp(\beta)$
- With $\beta = \exp(\theta)$, the hazard and logarithmic hazard can be written as follows:

$$
\begin{aligned}
\lambda(t; Z, \theta) &= \theta^Z \lambda_0(t) = \exp(\beta Z)\lambda_0(t), \\
\log(\lambda(t; Z, \theta)) &= \log(\lambda_0(t)) + \beta Z,
\end{aligned}
$$

Proportional hazards model (3)

- The hazard ratio between two individuals with covariates Z_1 and Z_2 is constant at any given time (as it should because that is the way the model was constructed):

$$\frac{\lambda_1(t; Z_1, \theta)}{\lambda_2(t; Z_2, \theta)} = \frac{\exp(\beta Z_1)\lambda_0(t)}{\exp(\beta Z_2)\lambda_0(t)} = \exp(\beta(Z_1 - Z_2))$$

Proportional hazards model (4)

- When one has more than one covariate (say, k of them), we write hazard ratio for individual i with covariates $Z_i = (Z_{i1}, \ldots, Z_{ik})$, as follows

$$\exp\left(\beta_1 Z_{i1} + \beta_2 Z_{i2} + \ldots \ldots \beta_k Z_{ik}\right) = \exp(\beta' Z_i)$$

- Under the proportional hazards assumption, the hazard for indviduals i is now written as

$$\lambda_i(t; Z_i, \theta) = \lambda_0(t) \exp(\beta' Z_i)$$

- The parameters θ of the model are now taken to include β (log hazard ratios) *and* the baseline rate $\lambda_0(t)$

Survival likelihood: the general case

- We can now rewrite the likelihood function (see part IIA) so that it explicitly shows the dependence of the hazard on the parameters and individual covariates, i.e we write hazard for individual i at observed time t_i as $\lambda(t_i; Z_i, \theta)$
- Survival data are presented as (t_i, d_i), $i = 1, ..., N$
- The likelihood function for $\lambda(t)$ is a product over individual contributions:

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{N} L_i(\theta) = \prod_{i=1}^{N} \lambda(t_i; Z_i, \theta)^{d_i} S(t_i; Z_i, \theta) \\
&= \prod_{i=1}^{N} \left[\lambda(t_i; Z_i, \theta)^{d_i)} \exp\left(-\int_0^{t_i} \lambda(u; Z_i, \theta) du\right)\right] \\
&= \prod_{i=1}^{N} \left[\lambda(t_i; Z_i, \theta)^{d_i)}\right] \times \exp\left(-\sum_{i=1}^{N} \int_0^{\infty} \lambda(u; Z_i, \theta) Y_i(u) du\right)
\end{aligned}
$$

where $Y_i(u) = \mathbf{1}(t_i \geq u)$ is an indicator for individual i still being in the risk set at time u

Exponential regression revisited (1)

- A proportional hazards exponential regression is defined by the following hazard rate

$$\lambda(t_i; Z_i, \theta) = \lambda_0(t; \theta) \exp(\beta' Z_i) = \lambda \exp(\beta' Z_i).$$

- The baseline hazard rate is constant over time, $\lambda_0(t) = \lambda$
- This is the same model as specified in part 2A, except that the dependence of the hazard on individuals covariates (explanatory variables) is now made explicit
- The likelihood function with individual-specific expressions for $\lambda(t; Z_i, \theta)$ is shown on the next page

Exponential regression model (2)

- The likelihood expression under the proportional hazards exponential regression model is

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} L_i(\theta) = \prod_{i=1}^{n} [\lambda \exp(\beta' Z_i)]^{d_i} \exp\{-\lambda \exp(\beta' Z_i) t_i\} \\
\log(L(\theta)) &= D \log(\lambda) + \sum_{i=1}^{n} d_i \beta' Z_i - \lambda \sum_{i=1}^{n} t_i \exp(\beta' Z_i),
\end{aligned}
$$

where $D = \sum_{i=1}^{n} d_i$, total number of failures.

Weibull regression model (1)

- When the baseline hazard is chosen as the Weibull hazard, we obtain proportional hazards Weibull regression:

$$\lambda(t_i; Z_i, \theta) = p\alpha^p t^{p-1} \exp(\beta' Z_i)$$

- When the shape parameter $p = 1$, the above reduces to the exponential model

Weibull regression model (2)

- The log-likelihood function is

$$\begin{aligned}\log(L(\theta)) &\propto Dp\log(\alpha) + D\log(p) \\ &+(p-1)\sum_{i=1}^{n} d_i \log(t_i) + \sum_{i=1}^{n} d_i \beta' Z_i \\ &-\sum_{i=1}^{n} (\alpha t_i)^p \exp(\beta' Z_i),\end{aligned}$$

where $D = \sum_{i=1}^{n} d_i$, total number of failures.

Weibull regression with stratification

- Assuming the same shape parameter but different scale parameter and the same effects of the covariates across the strata for the Weibull model, the stratum-specific hazard is defined as

$$\lambda_s(t_i; Z_i, \theta) = p\alpha_s^p t_i^{p-1} \exp(\beta' Z_i), s = 1, \ldots, S.$$