Clustered survival data

Advanced survival analysis 2021 day 6

1 Clustered survival data

In many survival studies there is a natural clustering such that failure times within the same cluster may be correlated, but are uncorrelated across clusters.

Examples

- *Multi-centre survival studies.* Clusters : centres.
- *Diabetic retinopathy.* Time to blindness in both eyes. Clusters : patients.
- *Twin studies.* Clusters: twin pairs.

Typically there are many clusters with few measurements per cluster. Even if we are not interested in characterizing the dependency within clusters, the clustering needs to be accounted for in the analysis.

2 Notation

For $i = 1, \ldots, n$, $j = 1, \ldots, k$, let T^*_{ij} and C_{ij} be the failure and censoring times for the jth individual in the ith cluster and X_{ij} the covariates.

- Right-censored failure time: $T_{ij} = T^*_{ij} \wedge C_{ij}$.
- Non-censoring indicator: $\Delta_{ij} = I(T^*_{ij} \leq C_{ij})$
- At-risk indicator: $Y_{ij}(t) = I(T_{ij} \geq t)$
- Counting process: $N_{ij}(t) = I(T_{ij} \leq t, \Delta_{ij} = 1)$

Individuals are assumed independent across, but not within, clusters.

3 Filtrations and intensities

The individual filtration (history) for subject ij contains the information (covariates, failures, censorings) gathered from observing this subject only,

$$\mathcal{F}^{ij}_t = \sigma\left(N_{ij}(s), Y_{ij}(s), X_{ij};\ 0 \leq s \leq t\right),$$

The cluster i observed filtration includes information from the whole cluster,

$$\mathcal{F}^{i}_t = \sigma\left(N_{ij}(s), Y_{ij}(s), X_{ij}; j = 1, \ldots, k,\ 0 \leq s \leq t\right).$$

The full observed filtration is $\mathcal{F}_t = \bigvee_{i=1}^{n} \mathcal{F}^i_t$.

Counting process intensities for individual ij (by the innovation theorem)

- The marginal intensity $\lambda_{ij}^{\mathcal{F}_{ij}}(t)dt = E[dN_{ij}(t)|\mathcal{F}_{t-}^{ij}]$
- The cluster intensity $\lambda_{ij}^{\mathcal{F}_i}(t)dt = E[dN_{ij}(t)|\mathcal{F}_{t-}^{i}]$
- The full intensity $\lambda_{ij}^{\mathcal{F}}(t)dt = E[dN_{ij}(t)|\mathcal{F}_{t-}]$

Note that due to independence between clusters, $\lambda_{ij}^{\mathcal{F}_i}(t) = \lambda_{ij}^{\mathcal{F}}(t)$, but $\lambda_{ij}^{\mathcal{F}_{ij}}(t)$ and $\lambda_{ij}^{\mathcal{F}}(t)$ are not the same.

4 Marginal (Cox) models

A marginal regression model is specified by assuming a model, such as the Cox model, for the marginal intensity, $\lambda_{ij}^{\mathcal{F}_{ij}}$.

- $\lambda_{ij}^{\mathcal{F}_{ij}}$ is not the intensity with respect to the observed filtration $\mathcal{F}_t$.
- The error terms

$$M_{ij}^{\mathcal{F}^{ij}}(t) = N_{ij}(t) - \int_0^t \lambda_{ij}^{ij}(s)ds$$

  are martingales with respect to $\mathcal{F}_t^{ij}$, but due to within-cluster dependency $M_{ij}^{\mathcal{F}^{ij}}$ are not martingales with respect to the joint filtration generated by all observed information t, $\mathcal{F}_t$.

We sketch the derivation of the large sample properties for the marginal Cox model. For a full proof see Spiekerman and Lin (1998). Estimate β using the Cox score (cf. day 2),

$$U(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{k}\int_0^\tau \left(X_{ij} - \frac{S^1(t,\beta)}{S^0(t,\beta)}\right) dN_{ij}(t)$$

where

$$S^l(t,\beta) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{k} Y_{ij}(t)X_{ij}^{\otimes j}\exp(X_{ij}\beta).$$

Let

$$s^l(t,\beta) = E\left(\sum_{j=1}^{k} Y_{ij}(t)X_{ij}^{\otimes l}\exp(X_{ij}\beta)\right).$$

Under standard regularity conditions, $S^*(t,\beta)$ converges uniformly to $s^l(t,\beta)$ in probability, and

$$U(\beta_0) = \sum_{i=1}^{n}\overbrace{\sum_{j=1}^{k}\int_0^\tau \left(X_{ij} - \frac{s^1(t,\beta_0)}{s^0(t,\beta_0)}\right) dM_{ij}^{\mathcal{F}^{ij}}(t)}^{\text{within-cluster residuals } \epsilon_i} + o_P(n^{1/2}) = \sum_{i=1}^{n}\epsilon_i + o_P(n^{1/2}),$$

where the ϵ_i's are i.i.d.. From the multivariate central limit theorem,

$$n^{-1/2}U(\beta_0) \xrightarrow{\mathcal{D}} N(0, B),$$

where $B = E(\epsilon_1^{\otimes 2})$.

By a Taylor expansion,

$$n^{1/2}\left(\hat{\beta} - \beta_0\right) = A^{-1}n^{-1/2}U(\beta_0) + o_P(1) \xrightarrow{\mathcal{D}} N\left(0, A^{-1}B(A^{-1})^T\right)$$

where $A = -E\left(\frac{\partial}{\partial\beta}U(\beta_0)\right)$. The asymptotic variance of $\hat{\beta}$ can be estimated by the sandwich estimator

$$\hat{A}^{-1}\hat{B}(\hat{A}^{-1})^T$$

where $\hat{A}$ and $\hat{B}$ are the empirical A and B. The estimator $\hat{\beta}$ is the same as with independent data, but the variance estimator accounts for the clustering.

5 Shared frailty models

Let Z_i be an unobserved cluster-specific random effect - the frailty. Individuals with a higher frailty can be thought of as being more frail, and therefore expected to die sooner than other individuals with the same measured covariates.

The **unobserved** cluster i conditional filtration is

$$\mathcal{H}_t^i = \sigma\left(N_{ij}(s), Y_{ij}(s), X_{ij}, Z_i; j = 1, \ldots, k,\ 0 \leq s \leq t\right),$$

and the full conditional filtration is $\mathcal{H}_t = \bigvee_{i=1}^n \mathcal{H}_t^i$. The frailty intensity (model) is specified with respect to $\mathcal{H}_t^i$

$$\lambda_{ij}^{\mathcal{H}}(t)dt = E[dN_{ij}(t)|\mathcal{H}_{t-}] = E[dN_{ij}(t)|\mathcal{H}_{t-}^i] = Z_i\lambda_{ij}^*(t)dt,$$

for some conditional *basic intensity* λ_{ij}^*.

- Given the frailty, the survival times are independent.
- $Z_1, \ldots, Z_n$ are i.i.d.
- Censoring is, conditional on covariates, assumed to be independent and noninformative on Z_i.
- Suggested frailty distributions: gamma, lognormal, positive stable, inverse Gaussian and compound Poisson.
- In practice mainly the gamma and lognormal distributions are used.

5.1 Cox shared gamma frailty model

By far the most popular choice of frailty distribution is the gamma distribution. Assume that $\lambda_{ij}^{\mathcal{H}}(t) = Y_{ij}(t)Z_i\alpha_0(t)e^{X_{ij}^T\beta}$ and that $Z_i \sim \text{Gamma}(1/\theta, \theta)$

- $E[Z_i] = 1$ and $\text{Var}(Z_i) = \theta$.
- $\theta = 0$ corresponds to independence
- A higher θ value indicates a stronger within-cluster correlation.
- Parameterized by β, $\alpha_0(\cdot)$, θ
- The model can be fitted in `coxph` by including a `+frailty(id)` term or with the package `frailtySurv`.
- Mathematically very convenient (Exercise 2)

5.2 Cox shared log-normal frailty

Let $Z_i = \exp(b_i)$ where $b_i \sim N(0, \sigma^2)$. The conditional Cox shared log-normal frailty model is given by

$$\lambda_{ij}^{\mathcal{H}}(t) = Y_{ij}(t)Z_i\alpha_0(t)e^{X_{ij}\beta} = Y_{ij}(t)\alpha_0(t)e^{X_{ij}\beta + b_i}.$$

- Possible interpretation: the correlation is induced by an unmeasured mean zero normal covariate b_i shared by all cluster members and independent of the observed covariates.
- $\sigma^2 = 0$ corresponds to independence
- A higher σ^2 value indicates stronger within-cluster correlation.
- Parameterized by β, $\alpha_0(\cdot)$, σ^2
- The model can be fitted by the packages `coxme` and `frailtySurv`

6 Marginal vs frailty models

- Marginal models
  - Models the marginal intensity $\lambda_{ij}^{\mathcal{F}^{ij}}$
  - Models population averaged covariate effects.
  - Standard errors corrected for clustering.
- Frailty models
  - Models the basic intensity λ_{ij}^*
  - Covariate effects interpreted conditionally on the frailty (within clusters). Cluster-specific effects.

- For the conditional Cox model with binary X **within the same cluster** i, the hazard ratio is

$$\frac{Z_i\alpha_0(t)e^{\beta}}{Z_i\alpha_0(t)e^{0}} = e^{\beta}$$

- Models within-cluster correlation

6.1 Marginal and observed intensities

The frailty density describes the population at time zero. The frailty density conditional on $T > t$ will change over time as more frail subjects die earlier.

Marginal intensity

We derive the marginal intensity based on the conditional intensity $\lambda_{ij}^{\mathcal{H}}(t)) = Z_1\lambda_{ij}^*(t)$. By the innovation theorem, the marginal intensity is

$$\lambda_{ij}^{\mathcal{F}_{ij}} = E[\lambda_{ij}^{\mathcal{H}}(t)|\mathcal{F}_{t-}^{ij}] = E[Z_i|\mathcal{F}_{t-}^{ij}]\lambda_{ij}^*(t). \tag{1}$$

Let $\phi(u) = E\left[\exp(-uZ)\right]$ denote the Laplace transform of (the density $p = p_\theta$ of) Z. Note that (assuming that we can move the derivative into the integral),

$$\begin{aligned} D^h\phi(u) &= \int \frac{\partial^h}{\partial u^h}\exp(-uz)p(z)dz \\ &= \int (-z)^h \exp(-uz)p(z)dz \\ &= (-1)^h E\left[Z^h \exp(-uZ)\right]. \end{aligned}$$

The expectation of the frailty conditional on the marginal history is

$$E\left[\left. Z_i \right| \mathcal{F}_{t-}^{ij}\right] = \int zp(z|\mathcal{F}_{t-}^{ij})dz$$

The likelihood of $\mathcal{F}_t^{ij}$ conditional on $Z_i = z$ is proportional to

$$\left(z\lambda_{ij}^*(T_{ij})\right)^{N_{ij}(t)} \exp\left(-z\Lambda_{ij}^*(t)\right)$$

Bayes' rule yields

$$\begin{aligned} p(z|\mathcal{F}_t^{ij}) &= \frac{\left(z\lambda_{ij}^*(T_{ij})\right)^{N_{ij}(t)} \exp\left(-z\Lambda_{ij}^*(t)\right) p(z)}{\int (z\lambda_{ij}^*(T_{ij}))^{N_{ij}(t)} \exp\left(-z\Lambda_{ij}^*(t)\right) p(z)dz} \\ &= \frac{z^{N_{ij}(t)} \exp\left(-z\Lambda_{ij}^*(t)\right) p(z)}{\int z^{N_{ij}(t)} \exp\left(-z\Lambda_{ij}^*(t)\right) p(z)dz} \end{aligned}$$

and thus

$$\begin{aligned} E\left[\left. Z_i \right| \mathcal{F}_{t-}^{ij}\right] &= \frac{\int \overbrace{zz^{N_{ij}(t)}}^{z^{1+N_{ij}(t)}} \exp\left(-z\Lambda_{ij}^*(t)\right) p(z)dz}{\int z^{N_{ij}(t)} \exp\left(-z\Lambda_{ij}^*(t)\right) p(z)dz} \\ &= \frac{(-1)^{1+N_{ij}(t)} D^{1+N_{ij}(t)}\phi\left(\Lambda_{ij}^*(t)\right)}{(-1)^{N_{ij}(t)} D^{N_{ij}(t)}\phi\left(\Lambda_{ij}^*(t)\right)} \\ &= -\frac{D^{1+N_{ij}(t)}\phi\left(\Lambda_{ij}^*(t)\right)}{D^{N_{ij}(t)}\phi\left(\Lambda_{ij}^*(t)\right)}. \end{aligned}$$

We are typically only interested in the conditional expectation when the individual is still at risk. When $Y_{ij}(t) = 1$ such that $N_{ij}(t) = 0$,

$$
\begin{aligned}
Y_{ij}(t)E\left[\, Z_i \middle| \mathcal{F}_{t-}^{ij} \right] &= -Y_{ij}(t)\frac{D\phi\left(\Lambda_{ij}^*(t)\right)}{\phi\left(\Lambda_{ij}^*(t)\right)} \\
&= -Y_{ij}(t)\left(D\log\phi\right)\left(\Lambda_{ij}^*(t)\right). \qquad (2)
\end{aligned}
$$

Observed intensity

Now turn to the observed intensity based on same conditional unobserved intensity as above, $\lambda_{ij}^{\mathcal{H}}(t)) = Z_1\lambda_{ij}^*(t)$.. Noting that the likelihood of the whole cluster i in $[0,t]$, $\mathcal{F}_t^i$, conditional on $Z_i = z$ is proportional to

$$
\prod_{j=1}^{k}\left(z\lambda_{ij}^*(T_{ij})\right)^{N_{ij}(t)}\exp\left(-z\Lambda_{ij}^*(t)\right).
$$

Similar calculations as above show that the observed intensity is

$$
\lambda_{ij}^{\mathcal{F}}(t) = E[\lambda_{ij}^{\mathcal{H}}(t)|\mathcal{F}_{t-}] = E[Z_i|\mathcal{F}_{t-}^i]\lambda_{ij}^*(t).
$$

where

$$
\begin{aligned}
E\left[\, Z_i \middle| \mathcal{F}_{t-}^i \right] &= \frac{\int z\prod_{j=1}^k\left(z\lambda_{ij}^*(T_{ij})\right)^{N_{ij}(t)}\exp\left(-z\Lambda_{ij}^*(t)\right)p(z)dz}{\int\prod_{j=1}^k\left(z\lambda_{ij}^*(T_{ij})\right)^{N_{ij}(t)}\exp\left(-z\Lambda_{ij}^*(t)\right)p(z)dz} \\
&= \frac{\int z^{1+\sum_j N_{ij}(t)}\exp\left(-z\sum_j\Lambda_{ij}^*(t)\right)p(z)dz}{\int z^{\sum_j N_{ij}(t)}\exp\left(-z\sum_j\Lambda_{ij}^*(t)\right)p(z)dz} \\
&= \frac{(-1)^{1+\sum_j N_{ij}(t)}D^{1+\sum_j N_{ij}(t)}\phi\left(\sum_j\Lambda_{ij}^*(t)\right)}{(-1)^{\sum_j N_{ij}(t)}D^{\sum_j N_{ij}(t)}\phi\left(\sum_j\Lambda_{ij}^*(t)\right)} \\
&= -\frac{D^{1+\sum_j N_{ij}(t)}\phi\left(\sum_j\Lambda_{ij}^*(t)\right)}{D^{\sum_j N_{ij}(t)}\phi\left(\sum_j\Lambda_{ij}^*(t)\right)}.
\end{aligned}
$$

6.2 The Cox shared gamma frailty

Assume $Z_i \sim \text{Gamma}(1/\theta,\theta)$, i.e. the density is

$$
p(z) = \frac{1}{\Gamma(1/\theta)\theta^{1/\theta}}z^{1/\theta-1}\exp(-z/\theta),\ \text{where } \Gamma(u) = \int_0^\infty x^{u-1}e^{-x}dx,
$$

and conditional proportional hazards,

$$
\lambda_{ij}^{\mathcal{H}}(t) = Y_{ij}(t)Z_i\exp(X_{ij}^T\beta)\alpha_0(t).
$$

The Laplace transform of p is

$$\begin{aligned}
\phi(u) &= \int_0^\infty \exp(-uz)p(z)dz \\
&= \int_0^\infty \exp(-uz)\frac{1}{\Gamma(1/\theta)\theta^{1/\theta}} z^{1/\theta-1}\exp(-z/\theta)dz \\
&= \frac{1}{\Gamma(1/\theta)\theta^{1/\theta}} \int_0^\infty z^{1/\theta-1}\exp(-z(u+1/\theta))dz \\
&= \frac{1}{\Gamma(1/\theta)\theta^{1/\theta}} \int_0^\infty \left(\frac{x}{u+1/\theta}\right)^{1/\theta-1}\exp(-x)\frac{1}{u+1/\theta}dx \\
&= \frac{1}{\Gamma(1/\theta)\theta^{1/\theta}}\frac{1}{(u+1/\theta)^{1/\theta}} \int_0^\infty x^{1/\theta-1}\exp(-x)dx \\
&= \frac{1}{\Gamma(1/\theta)}\frac{1}{(\theta(u+1/\theta))^{1/\theta}}\Gamma(1/\theta) \\
&= (\theta u+1)^{-1/\theta}.
\end{aligned}$$

The fourth equality followed from the variable substitution $x = z(u + 1/\theta)$.

Shorter argument:

$$\begin{aligned}
\phi(u) &= E\left(e^{-uZ}\right) \\
&= \int_0^\infty e^{-uz}\frac{1}{\Gamma(1/\theta)\theta^{1/\theta}} z^{1/\theta-1}\exp\left(-z/\theta\right)dz \\
&= \frac{1}{\Gamma(a)\theta^{1/\theta}}\int_0^\infty z^{1/\theta-1}\exp\left(-z(u+1/\theta)\right)dz \\
&= \frac{1}{\Gamma(1/\theta)\theta^{1/\theta}}\Gamma(1/\theta)(u+1/\theta)^{-1/\theta} \\
&= (1+\theta u)^{-1/\theta}
\end{aligned}$$

The third equality follows from the fact that $\int_0^\infty \frac{1}{\Gamma(a)b^a} z^{a-1}e^{-z/b}dz = 1$ yields $\int_0^\infty z^{a-1}e^{-z/b}dz = \Gamma(a)b^a$.

Marginal intensity in the Cox gamma frailty model

Thus, for the Cox shared gamma frailty model, by (2), the conditional expectation of Z_i for patients in the risk set is

$$\begin{aligned}
Y_{ij}(t)E[Z_i|\mathcal{F}_{t-}^{ij}] &= -Y_{ij}(t)\left(D\log\phi\right)\left(\Lambda_{ij}^*(t)\right) \\
&= -Y_{ij}(t)(-(1+\theta\Lambda_{ij}^*(t))^{-1}) \\
&= Y_{ij}(t)\frac{1}{1+\theta e^{X_{ij}^T\beta}A_0(t)}
\end{aligned} \tag{3}$$

- Since the more frail subjects die sooner, the population at risk changes over time and this has an important impact on the population (marginal) hazard function.
- The expected value decreases with time as the subjects remaining at risk tend to have lower Z.
- The population intensity contains the frailty parameter θ.

Now, by (1),

$$\lambda_{ij}^{\mathcal{F}^{ij}}(t) = Y_{ij}(t)\frac{1}{1+\theta e^{X_{ij}^T\beta}A_0(t)}\exp(X_{ij}^T\beta)\alpha_0(t)$$

Observed intensity in the Cox gamma frailty model

Note that the hth derivative of the Laplace transform of the gamma distribution with mean 1 is

$$D^h\phi(t) = (-1)^h(1/\theta+1)\dots(1/\theta+h-1)(1+\theta t)^{-1/\theta-h}\theta^{h-1}.$$

Using this, the observed intensity is

$$\begin{aligned}
&\lambda_{ij}^{\mathcal{F}}(t)\\
&= E[\lambda_{ij}^{\mathcal{H}}(t)|\mathcal{F}_{t-}^i]\\
&= Y_{ij}(t)E[Z_i|\mathcal{F}_{t-}^i]\alpha_0(t)\exp(X_{ij}^T\beta)\\
&= -Y_{ij}(t)\frac{D^{1+\sum_j N_{ij}(t)}\phi\left(\sum_j \Lambda_{ij}^*(t)\right)}{D^{\sum_j N_{ij}(t)}\phi\left(\sum_j \Lambda_{ij}^*(t)\right)}\alpha_0(t)e^{X_{ij}^T\beta}\\
&= -Y_{ij}(t)\frac{(-1)^{1+\sum_{j'}N_{ij'}(t)}\prod_{h=1}^{\sum_{j'}N_{ij'}(t)}(1/\theta+h)(1+\theta\sum_j\Lambda_{ij}^*(t))^{-1/\theta-\sum_{j'}N_{ij'}(t)-1}\theta^{\sum_{j'}N_{ij'}(t)}}{(-1)^{\sum_{j'}N_{ij'}(t)}\prod_{h=1}^{\sum_{j'}N_{ij'}(t)-1}(1/\theta+h)(1+\theta\sum_j\Lambda_{ij}^*(t))^{-1/\theta-\sum_{j'}N_{ij'}(t)}\theta^{\sum_{j'}N_{ij'}(t)-1}}\alpha_0(t)e^{X_{ij}^T\beta}\\
&= Y_{ij}(t)\frac{1/\theta+\sum_{j'}N_{ij'}(t)}{1/\theta+\sum_{j'}\Lambda_{ij'}^*(t)}\alpha_0(t)e^{X_{ij}^T\beta}
\end{aligned}$$

7 Combining marginal and conditonal (frailty) models

The marginal models describe regression effects on the population level, but don't provide estimates of the within-clusted correlation. Marginal models can be combined with random effects models for the within-cluster dependence.

In Excercise 2, we assumed a model for the marginal intensity $\lambda_{ij}^{\mathcal{F}_{ij}}$ and that dependency within a cluster was induced by an unobserved shared random effect Z acting multiplicatively on the conditional intensity

$$\lambda_{ij}^{\mathcal{H}}(t) = Z_i\lambda_{ij}^*(t).$$

We derived the following expressions for the basic and observed intensities,

$$\begin{aligned}
\lambda_{ij}^*(t) &= -\lambda_{ij}^{\mathcal{F}^{ij}}(t)\exp\left(-\Lambda_{ij}^{\mathcal{F}^{ij}}(t)\right)(D\phi^{-1})\left(\exp\left(-\Lambda_{ij}^{\mathcal{F}^{ij}}(t)\right)\right)\\
\lambda_{ij}^{\mathcal{F}}(t) &= \lambda_{ij}^{\mathcal{F}^{ij}}(t)\exp\left(-\Lambda_{ij}^{\mathcal{F}^{ij}}(t)\right)(D\phi^{-1})\left(\exp\left(-\Lambda_{ij}^{\mathcal{F}^{ij}}(t)\right)\right)\\
&\quad\times\frac{\left(D^{1+\sum_{j=1}^k N_{ij}(t)}\right)\phi\left(\sum_{j=1}^k\phi^{-1}\left(\exp\left(-\Lambda_{ij}^{\mathcal{F}^{ij}}(t)\right)\right)\right)}{\left(D^{\sum_{j=1}^k N_{ij}(t)}\right)\phi\left(\sum_{j=1}^k\phi^{-1}\left(\exp\left(-\Lambda_{ij}^{\mathcal{F}^{ij}}(t)\right)\right)\right)}.
\end{aligned}$$

Using the observed intensities $\lambda_{ij}^{\mathcal{F}}$, one can construct write the likelihood in terms of the observed data.

7.1 The two-stage method

Inference for marginal models with dependency induced by an unobserve frailty can be achieved in two steps (Shih and Louis, 1995; Glidden, 2000):

Step 1 Estimate the marginal parameters (ignoring the random effect).

Step 2 Plug the marginal estimates into the likelihood for θ and maximize.

With Cox marginals and a gamma distributed frailty, the observed likelihood for cluster i is proportional to

$$\sum_j \int_0^\tau \log\left(1+\theta\sum_j N_{ij}(t-)\right) dN_{ij}(t) + \theta\sum_j N_{ij}(\tau)H_{ij}$$
$$-\left(\frac{1}{\theta}+\sum_j N_i(\tau)\right)\log\left(1+\sum_j\left(\exp\left(\theta H_{ij}\right)-1\right)\right)$$

where $H_{ij} = e^{\beta^T X_{ij}} A_0(T_{ij})$.

Replacing β and A_0 by the marginal estimates we obtain a pseudolikelihood. Glidden (2000) showed consistency and asymptotic normality of the estimator. A more efficient, but also more complicated, estmator results from maximizing over both the marginal parameters and the frailty parameter simultaneousle in one step.

7.2 Copula models

Multiplicative shared random effects models, with conditonal hazards $\alpha_{ij}^{\mathcal{H}} = Z_i\alpha_{ik}^*$, are a subclass of copula models. A copula C_θ is a K-dimensional function with uniform marginals such that the joint marginal function is

$$\mathrm{pr}(T_{i1}^* > t_1, \ldots, T_{iK}^* > t_K) = C_\theta(S_{i1}(t_1), \ldots, S_{iK}(t_K)),$$

where S_{ij} is the marginal survival function for T_{ij}^*. Different C_θ gives different joint distribution, but the marginals are unaltered.

For a frailty model with conditonal hazard $Z_i\alpha_{ik}^*$, the marginal survival function is

$$S_{ik}(t_k) = E(\exp(-Z_iA_{ik}^*(t_k))) = \phi(A_{ik}^*(t_k)),$$

such that

$$A_{ik}^*(t_k) = \phi^{-1}(S_{ik}(t_k)).$$

The joint marginal survival function is

$$\begin{aligned}\mathrm{pr}(T_{i1} > t_1, \ldots, T_{iK} > t_K) &= E(\exp(-Z_i(\overbrace{A_{i1}^*(t_1)}^{=\phi^{-1}(S_{i1}(t_1))} + \ldots + A_{iK}^*(t_K)))) \\ &= \phi(\phi^{-1}(S_{i1}(t_1)) + \ldots + \phi^{-1}(S_{iK}(t_K))) \\ &= C_\theta(S_{i1}(t_1), \ldots, S_{iK}(t_K))\end{aligned}$$

Consider the gamma distribution with mean one. The Laplace transform for the gamma distribution was shown to be $\phi(u) = (1+\theta u)^{-1/\theta}$ and thus has inverse

$$\phi^{-1}(u) = \frac{u^{-\theta}-1}{\theta}.$$

The copula is

$$\begin{aligned} C_\theta(u_1, \ldots, u_K) &= (1 + \theta((u_1^{-\theta} - 1)/\theta + \ldots + u_K^{-\theta} - 1)/\theta))^{-1/\theta} \\ &= (1 + u_1^{-\theta} + \ldots + u_K^{-\theta} - K)^{-1/\theta}. \end{aligned}$$

Inserting marginal instensities of Cox form, yields a marginal survival distribution corresponding to the setting of Excercise 2. From this, the observed intensity and likelihood can be derived.

8 Attenuation

For the Cox proportional hazards model, it is known that excluding an important factor which is independent of the observed covariates from the analysis leads to estimates of the covariates which are too small in absolute value. If a frailty is present, but ignored then the covariate effects will be smaller than those in the underlying frailty model.

In Exercise 1 we noted that the frailty model estimates were "larger" than the marginal estimates. The phenomenon is called attenuation and is a consequence of the models. Note that smaller estimates (in absolute value) gives hazard ratios closer to one.

Assume a Cox shared gamma frailty model with binary X. The marginal intensity is

$$Y(t)\alpha_0(t)\exp(\beta X)\frac{1}{1 + \theta A_0(t)e^{\beta X}}$$

giving the marginal hazard ratio

$$\exp(\beta)\frac{1 + \theta A_0(t)}{1 + \theta A_0(t)e^{\beta}}. \tag{4}$$

- The HR (4) is equal to $\exp(\beta)$ at time $t = 0$ and tends to 1 as $t \to \infty$.
- The hazard ratio based on the marginal intensity depends on time, and is not on proportional hazards form.
- Only in the positive stable frailty model that the marginal and conditional formulations both yield proportional hazards.

Assume that the frailty Z has Laplace transform ϕ and that $\beta > 0$. The hazard ratio in the marginal model is

$$\exp(\beta)\underbrace{\frac{(D\log\phi)(e^\beta A_0(t))}{(D\log\phi)(A_0(t))}}_{k(t)} = \exp(\beta)k(t),$$

and we see that $k(t) \leq 1$ if and only if

$$(D\log\phi)(e^\beta A_0(t)) \leq (D\log\phi)(A_0(t)).$$

The inequality holds because $\log\phi$ is convex, as

$$\left(D^2\log\phi\right)(t) = \underbrace{E(Z^2h(t,Z)) - (E(Zh(t,Z)))^2}_{\text{variance of a r.v. with density } h(t,z)p(z)} \geq 0,$$

with $h(t,Z) = \exp(-tZ)/E(\exp(-tZ))$.

9 Characterizing dependence

Different measures of the dependence within clusters have been suggested. The usual corrlelation can typically not be used, as the tails of the distributions are not often not possible to estimate due to censoring.

9.1 Kendall's τ

Kendall's τ is a global rank-based measure of dependence. Kendall's τ is defined given two independent pairs (T_1, T_2) and $(\tilde{T}_1, \tilde{T}_2)$ and measures the the concordance probability minus the discordance probability between the pairs

$$E(I((T_1 - \tilde{T}_1)(T_2 - \tilde{T}_2) > 0)) - E(I((T_1 - \tilde{T}_1)(T_2 - \tilde{T}_2) < 0)) = 2p - 1$$

where p is the concordance probability; the probability that Mr. Smith (T_1) dies before (after) Mr. Peterson ($\tilde{T}_1$) and that Mrs. Smith (T_2) dies before (after) Mrs. Peterson($\tilde{T}_2$).

- Kendall's τ is a global measure of dependence
- $p = 0.5$ or $\tau = 0$ corresponds to independence.
- For the Cox shared gamma frailty model Kendall's τ is $\tau = \frac{\theta}{\theta+2}$.
- For a log-normal frailty model it is not possible to give a closed form expression for Kendall's τ.

9.2 Cross ratio

The cross ratio is a local measure of dependence. Consider (T_1, T_2) from the same pair. Let $\alpha_1(t|T_2 > t_2)$ denote the hazard of T_1 given that $T_2 > t_2$,

$$\begin{aligned}
\alpha_1(t|T_2 > t_2) &= \lim_{h \to 0} \frac{1}{h} \mathrm{pr}(T_1 < t + h | T_1 > t_1, T_2 > t_2) \\
&= -\frac{\partial/(\partial s_1) S(s_1, t_2)|_{s_1 = t_1}}{S(t_1, t_2)} \\
&= -\frac{D\phi(A_1^*(t_1) + A_2^*(t_2))}{\phi(A_1^*(t_1) + A_2^*(t_2))} \alpha_1^*(t)
\end{aligned}$$

and $\alpha_1(t|T_2 = t_2)$ denote the hazard of T_1 given that $T_2 = t_2$,

$$\begin{aligned}
\alpha_1(t|T_2 = t_2) &= \lim_{h \to 0} \frac{1}{h} \mathrm{pr}(T_1 < t + h | T_1 > t_1, T_2 = t_2) \\
&= -\frac{\partial^2/(\partial s_1 \partial s_2) S(s_1, s_2)|_{(s_1, s_2) = (t_1, t_2)}}{\partial/(\partial s_2) S(t_1, s_2)|_{s_2 = t_2}} \\
&= -\frac{D^2\phi(A_1^*(t_1) + A_2^*(t_2))}{D\phi(A_1^*(t_1) + A_2^*(t_2))} \alpha_1^*(t)
\end{aligned}$$

The two hazards concern different hypothetical event histories of the other individual in the cluster. They are equal if there is no dependence between individuals.

The cross ratio is defined the ratio

$$CR(t_1,t_2) = \alpha_1(t|T_2 = t_2)\alpha_1(t|T_2 > t_2) = \frac{\frac{D^2\phi(A_1^*(t_1)+A_2^*(t_2))}{D\phi(A_1^*(t_1)+A_2^*(t_2))}}{\frac{D\phi(A_1^*(t_1)+A_2^*(t_2))}{\phi(A_1^*(t_1)+A_2^*(t_2))}}$$

Note that $CR(t_1,t_2) = 1$, if the two event times are independent, and $CR(t_1,t_2) > 1$, if there is positive dependence between the event times.

For the shared gamma frailty model is $CR = 1 + \theta$, a time-independent quantity, but for other frailty distributions the cross ratio depends on time.

10 Estimation and inference for frailty models

The likelihood of the frailty model is

$$\prod_{i=1}^{n}\left(\left(\prod_{j=1}^{K}\lambda_{ij}^*(T_{ij})^{N_{ij}(\tau)}\right) E\left(Z_i^{\sum_{j=1}^K N_{ij}(\tau)}\exp\left(-Z_i\sum_{j=1}^{K}\Lambda_{ij}^*(\tau)\right)\right)\right)$$
$$= \prod_{i=1}^{n}\left(\prod_{j=1}^{K}\lambda_{ij}^*(T_{ij})^{N_{ij}(\tau)}\right)(-1)^{\sum_{j=1}^K N_{ij}(\tau)} D^{\sum_{j=1}^K N_{ij}(\tau)}\phi\left(\sum_{j=1}^{K}\Lambda_{ij}^*(\tau)\right)$$

When the basic hazard is modelled nonparametrically, the dimension of λ_{ij} equals the number of disting uncensored event times.

- Semi-parametric maximum likelihood theory says that the MLE exists, is consistent and converges weakly to a Gaussian process under regularity conditions.
- The theory is complex, but all details have been worked out (Murphy, 1994, 1995; Parner, 1998).
- The MLE can be found by the EM-algorithm.
- For the Cox shared gamma frailty model `coxph` uses a faster penalized partial likelihood approach to find the MLE.
- Inference for the frailty variance is more challenging. The likelihood ratio test for $\theta = 0$ is on the boundary of the parameter space and follows a 50:50 mixture of χ^2 distributions with 0 and 1 degrees of freedom.

11 Challenges

- In frailty models, the dependence between correlated observations changes over time and the frailty distribution dictates how the dependence changes. For example
  - The gamma frailty model induces high late dependency (high probability of events close by at later time points)
  - The positive stable model induces early dependence (high probability of events close by early in the follow-up)
  - The log-normal is in the middle

However, it is difficult to make a well-informed choice for the frailty distribution. Local depencency measures, such as cross-ratios, can give some guidance. In principle models can be compared via their log-likelihood, e.g., by likelihood ratio comparison within the large family of power variance distributions. Cross-ratios can give some insight.

- Some goodness-of-fit procedures exist for the shared gamma frailty model (Glidden, 1999; Shih and Louis, 1995; Geerdens et al., 2013).

- The frailty distribution (with finite mean) may be estimated on univariate data when there are covariates and a Cox model is assumed. This is because the marginal hazards depend on the frailty variance. In practice the frailty variance is primarily identified from correlation, but for small cluster sizes violation of the propotional hazards assumption and presence of a frailty are confounded.

12 Summary

There are multiple approaches for clustered survival data

- Marginal Cox approach. Given covariates, the marginal intensity is

$$Y_{ij}(t)\alpha_0(t)\exp(X_{ij}^T\beta).$$

  This gives the population regression parameters, without modelling the dependency.

- Shared frailty Cox models. Given random effects Z the survival times are independent with intensity

$$Y_{ij}(t)Z_i\alpha_0(t)\exp(X_{ij}^T\beta).$$

- Copula model. A frailty model where the marginal hazard given the covariates is the Cox model

$$Y_{ij}(t)\alpha_0(t)\exp(X_{ij}^T\beta).$$

  The two approaches above are combined by assuming that the dependence of marginal hazards is characterized by an underlying frailty.

References

Geerdens, C., Claeskens, G., and Janssen, P. (2013). Goodness-of-fit tests for the frailty distribution in proportional hazards models with shared frailty. *Biostatistics*, 14(3).

Glidden, D. (1999). Checking the adequacy of the gamma frailty model for multivariate failure times. *Biometrika*, 86(2).

Glidden, D. (2000). A two-stage estimator of the dependence parameter for the clayton-oakes model. *LIDA*, 6(2).

Murphy, S. (1994). Consistency in a proportional hazards model incorporating a random effect. *Annals of Statistics*, 22(1).

Murphy, S. (1995). Asymptotic theory for the frailty model. *Annals of Statistics*, 23(1).

Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *Annals of Statistics*, 26(1).

Shih, J. and Louis, T. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51(4).

Spiekerman, C. and Lin, D. (1998). Marginal regression models for multivariate failure time data. *JASA*, 93(443).