Competing risks

Here we consider K duration variables $X_1, \ldots, X_K$ for the same individual (e.g. times to death for either of K diseases).

Let $S(t_1, \ldots, t_K)$ denote joint survivor function of $(X_1, \ldots, X_K)$.

Marginal survivor function for X_i is

$$S_{i,M}(t_i) = S(0, \ldots, 0, t_i, 0, \ldots, 0)$$

In practice we only get to observe

$$T = \min(X_1, \ldots, X_k) \quad \text{and} \quad \Delta = \text{ cause of death}$$

With this kind of data we can not infer S (we would need observations of $(X_1, \ldots, X_K)$ to infer S).

Overall and cause-specific hazard

The hazard function of T (using $T > t \Leftrightarrow X_1 > t, \ldots, X_k > t$)

$$h_T(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \log S(t, \ldots, t) = \sum_{i=1}^{K} h_i(t)$$

where

$$h_i(t) = -\frac{\partial}{\partial t_i} \log S(t_1, \ldots, t_K)|_{t_1 = t_2 = \ldots, t_K = t}$$

is the *cause-specific hazard* (here we used the multivariate chain rule on $S_T(t) = S(g(t))$ with $g(t) = (t, \ldots, t)$).

We can show (later) that

$$\begin{aligned} h_i(t) &= P(T \in [t, t + \mathrm{d}t[, \Delta = i | T \geq t) \\ &= P(X_i \in [t, t + \mathrm{d}t[, X_j > X_i, j \neq i | T \geq t) \end{aligned} \tag{1}$$

Thus $h_i(t)\mathrm{d}t$ is the probability of dying of cause i in the infinitesimal interval $[t; t + \mathrm{d}t[$ given alive at time t.

Estimation of cause-specific hazard

Suppose $[u_{l-1}, u_l[$ is a small interval with d_{li} number of cause i events in the interval then (assuming at most one type of cause in interval)

$$\hat{h}_i(u_l) = \frac{d_{li}}{r(u_{l-1})}$$

where $r(u_{l-1})$ is the number of individuals alive at time u_{l-1}.

This gives Nelson-Aalen estimator of ith cumulative cause specific hazard

$$\hat{H}_i(t) = \sum_{t^* \in D_i : t^* \leq t} \frac{d_i(t^*)}{r(t^*)}$$

where D_i set of cause i death times and $d_i(t^*)$ number of type i deaths at time t^*.

We can also estimate $S_T(t)$ simply as proportion of individuals who were at risk at time t.

See note by Rodriguez regarding likelihood for competing risk data: again only cause-specific hazards can be inferred.

Rodriguez also proposes Kaplan-Meier type estimate of $S_i(t) = \exp(-H_i(t))$ but does not give details about this.

Not clear to me what is the argument behind this estimator (S_i is not a survivor function in general !)

Independent competing risks

Suppose $X_1, \ldots, X_k$ are independent. Then

$$S(t_1, \ldots, t_K) = \prod_{i=1}^{K} S_{i,M}(t_i)$$

and we immediately get

$$h_{M,i} = -\frac{\mathrm{d}}{\mathrm{d}t_i} \log S_{i,M}(t_i) = h_i(t_i)$$

- thus marginal and cause-specific hazards coincide.

Hence in case of independent competing risks we are able to infer marginal hazards (and distributions) of $X_1, \ldots, X_k$.

However, given data (T, Δ) we can not infer S and hence not assess whether independence is fulfilled.

Note: so far we have been treating censoring as a competing risk and assumed censoring times independent of death times.

Example 2.7 in KM

X_1, X_2 correlated survival times from shared Gamma frailty model:

$$S(t_1, t_2) = [1 + \theta(\lambda_1 t_1 + \lambda_2 t_2)]^{-1/\theta}$$

Cause-specific hazard for $i = 1$:

$$h_1(t) = -\frac{\partial}{\partial t_1} \log S(t_1, t_2)|_{t_1 = t_2 = t} = \frac{\lambda_1}{1 + \theta t(\lambda_1 + \lambda_2)}$$

is smaller than marginal hazard

$$h_{1,M}(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \log S_{1,M}(t) = \frac{\mathrm{d}}{\mathrm{d}t} - \log S(t, 0) = \frac{\lambda_1}{1 + \theta \lambda_1 t}$$

Lack of identifiability

Suppose we have cause specific hazard

$$h_i(t) = \frac{\lambda_i}{1 + \theta t \lambda}$$

By previous slide, this is consistent with model for correlated X_1 and X_2 (letting $\lambda = \lambda_1 + \lambda_2$).

However, it is also consistent with model where X_1 and X_2 are independent with marginal hazards

$$h_{i,M}(t) = h_i(t)$$

We can not tell which underlying joint model is true.

Summary: competing risks are tricky. We may be interested in marginal hazards for different risk types but given only data (T, Δ) these can in general not be inferred.

Only possible if independent competing risks.

Assumption of independence can not be tested given only data of the form (T, Δ).

What is probability of dying of cause i ?

This is

$$P(X_i < X_j, j \neq i) = \int_0^\infty h_i(t) S_T(t) \mathrm{d}t \qquad (2)$$

Makes intuitive sense:

$$\begin{aligned} P(X_i < X_j, j \neq i) &\approx \sum_{l=0}^{\infty} P(X_i \in [u_l, u_{l+1}[, X_j \geq u_{l+1}, j \neq i) \\ &\approx \sum_{l=0}^{\infty} h_i(u_l) S(u_l)(u_{l+1} - u_l) \end{aligned}$$

Let $u_{l+1} - u_l$ tend to zero.

Proof of (2)

Let $i = 1$ wlog.

Consider fixed t - only u varying:

$$S(\infty, t, \dots, t) - S(u, t, \dots, t) = \int_u^\infty s(z)\mathrm{d}z$$

where $s(u) = \frac{\partial}{\partial u} S(u, t, \dots, t)$.

On the other hand,

$$\begin{aligned} &S(\infty, t, \dots, t) - S(u, t, \dots, t) \\ =&0 - \int_u^\infty \int_t^\infty \dots \int_t^\infty f(u_1, \dots, u_k)\mathrm{d}u_1\mathrm{d}u_2 \dots \mathrm{d}u_k. \end{aligned}$$

Hence

$$\frac{\partial}{\partial u} S(u, t, \dots, t) = -\int_t^\infty \dots \int_t^\infty f(u, u_2 \dots, u_k)\mathrm{d}u_2 \dots \mathrm{d}u_k$$

Thus we obtain

$$\begin{aligned} P(X_1 < X_j, j > 1) &= \int_0^\infty \int_t^\infty \dots \int_t^\infty f(t, u_2 \dots, u_k) \mathrm{d}t \mathrm{d}u_2 \dots \mathrm{d}u_k \\ &= \int_0^\infty \frac{-\frac{\partial}{\partial u} S(u, t, \dots, t)|_{u=t}}{S_T(t)} S_T(t) \mathrm{d}t \\ &= \int_0^\infty h_1(t) S_T(t) \mathrm{d}t \end{aligned}$$

In a similar way we can compute probability of surviving to time t and eventually die from cause i:

$$
\begin{aligned}
P(T > t, X_i < X_j, j \neq i) =& P(X_i > t, X_i < X_j, j \neq i) \\
=& \int_t^\infty h_i(u) S_T(u) \mathrm{d}u
\end{aligned}
$$

The cumulative incidence function is the probability of dying before t of cause i:

$$
\begin{aligned}
F_i(t) :=& P(T \leq t, X_i < X_j, j \neq i) \\
=& P(X_i < X_j, j \neq i) - P(X_i > t, X_i < X_j, j \neq i) \\
=& \int_0^t h_i(u) S_T(u) \mathrm{d}u
\end{aligned}
$$

Cause-specific hazard

We show that (1) is true.

$$
\begin{aligned}
&\frac{P(t \le T < t+\Delta, X_1 < X_j, j \ne i | T \ge t)}{\Delta} \\
=&\frac{P(t \le X_1 < t+\Delta, X_1 < X_j, j \ne i)}{S(t)\Delta} \\
=&\frac{\frac{1}{\Delta}\int_t^{t+\Delta} h_1(u)S(u)\mathrm{d}u}{S(t)}
\end{aligned}
$$

Letting $\Delta \to 0$ we obtain

$$\frac{h_1(t)S(t)}{S(t)} = h_1(t)$$

Back to Spar Nord

Suppose a customer defaults. We then follow customer up to loss (L), quit (Q), stops being default ($\neg D$) or until present date.

Focus on L. Competing risks Q, $\neg D$. Let X_L, X_Q, $X_{\neg D}$ be (discrete-valued) times to competing events and let T be minimum of these.

We are interested in probability of getting a loss:

$$P(X_L < X_Q, X_L < X_{\neg D}) = \sum_{l=1}^{\infty} P(X_L = l, l < X_Q, l < X_{\neg D}, l \leq X_L)$$

$$= \sum_{l=1}^{\infty} P(X_L = l, l \leq T) = \sum_{l=1}^{\infty} P(X_L = l | T \geq l) P(T \geq l)$$

(discrete time analogue of (2), note X_L, X_Q and $X_{\neg D}$ can not coincide)

The probabilities $P(X_L = l | T \geq l)$ and $P(T \geq l)$ can be estimated unbiasedly regardless of whether times to competing risks are independent or not.

To estimate $P(X_l \geq l)$ we need $P(X_L = l | X_L \geq l)$ which coincides with $P(X_L = l | T \geq l)$ in case of independent competing risks.

Back to Thiele: marriage or death

X_M, X_D times to marriage or death. $T = \min(X_M, X_G)$ Again we can compute probability of getting married as

$$\sum_{l=1}^{\infty} P(X_M = l | T \geq l) P(T \geq l) = 0.44$$

If we assume X_M and X_G independent we can (and did) estimate $P(X_M = l | X_M \geq l)$ and compute counterfactual probability of getting married in a world where women are immortal:

$$\sum_{l=1}^{\infty} P(X_M = l | X_M \geq l) P(X_M \geq l) = 0.57$$

Note, naturally latter probability (crude) is larger than the former (net) !