Frailty - mixed models for duration data

Rasmus Waagepetersen

November 9, 2022

Topics:

- Mixed models for survival data: frailty models
- Example with time-dependent covariate and frailty
- Marginal analysis of correlated survival data

Mixed models for survival data

Suppose we have survival data (T, Δ, Z) where Z is a covariate that can be used to model heterogeneity of the population. T is minimum of X and C.

Thus we use a hazard model $h^*(t; Z)$ that depends on Z.

However, survival may be influenced by a further unobserved factor U - e.g. unobserved genetic predisposition for a disease or genetically influenced ability to recover from disease. This is called a frailty in survival analysis.

We may then use the following so-called frailty model for the hazard

$$h(t; Z, U) = Uh^*(t; Z)$$

This is the typical example of a mixed model in survival analysis.

Marginal survival function and hazard

Since U is unobserved we can not base inference directly on $h(t; Z, U)$ - instead we need to find the marginal hazard function or survival function.

We assume U and Z are independent (note they may be dependent conditionally on T !)

Conditionally on $U = u$ and $Z = z$, the survival function is

$$S(t; z, u) = \exp(-uH^*(t; z))$$

so the marginal survival function is

$$S(t; z) = \mathbb{E}S(t; z, U) = L_U(H^*(t; z)) \quad L_U(t) = \mathbb{E}\exp(-tU)$$

where L_U is the Laplace transform for U ($L_U(t) = M_U(-t)$ where M_U is the moment generating function of U).

Marginal hazard function

Hazard function:

$$h(t;z) = -\frac{\mathrm{d}}{\mathrm{d}t}\log S(t;z) = h^*(t;z)\frac{\mathbb{E}[US(t;z,U)}{S(t;z)]}$$
$$= h^*(t;z)\mathbb{E}[U|X \geq t, Z = z]$$

(note (exercise !) conditional density of $U|X \geq t, Z = z$ is $f(u|t,z) = S(t;z,u)f_U(u)/S(t;z)$

Another calculation leading to same result:

$$\begin{aligned} h(t;z)\mathrm{d}t &= P(X \in [t, t+\mathrm{d}t[|X \geq t, Z = z) \\ &= \mathbb{E}[P(X \in [t, t+\mathrm{d}t[|X \geq t, Z = z, U)|X \geq t, Z = z] \\ &= \mathbb{E}[h(t;z,U)\mathrm{d}t|X \geq t, Z = z] = h^*(t;z)\mathbb{E}[U|X \geq t, Z = z]\mathrm{d}t \end{aligned}$$

In practice we need to assume a distribution (on $[0,\infty[$) for U, e.g. log-normal, gamma,....

Example: gamma frailty

A gamma distributed variable $U \sim \Gamma(\alpha, \beta)$ with shape α and scale β has Laplace transform

$$L_U(t) = (1 + \beta t)^{-\alpha}$$

so in that case

$$S(t; z) = (1 + \beta H^*(t; z))^{-\alpha}$$

The hazard function becomes

$$h(t; z) = -\frac{\mathrm{d}}{\mathrm{d}t} \log S(t; z) = \frac{\alpha\beta h^*(t; z)}{1 + \beta H^*(t; z))}$$

A common simplifying choice is to use $\Gamma(1/\theta, \theta)$ in which case

$$h(t; z) = \frac{h^*(t; z)}{1 + \theta H^*(t; z))}$$

($\Gamma(1/\theta, \theta)$ has mean 1 and variance θ)

Marginal hazard versus conditional hazard

On a previous slide we saw

$$h(t; z) = \mathbb{E}[h(t; z, U)|X \geq t, Z = z] = \mathbb{E}[U|X \geq t, Z = z]h^*(t; z)$$

Interesting phenomenon: $\mathbb{E}[U|X \geq t, Z = z]$ is a decreasing function of time: weak individuals die first so the remaining population with $T \geq t$ becomes stronger as t increases.

We compute (exercise: check this) derivative of conditional expectation wrt t:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[U|X \geq t, Z = z] =$$
$$\frac{-\mathbb{E}[U^2 h^*(t; z) \exp(-UH^*(t; z)]S(t; z) + \mathbb{E}[U \exp(-UH^*(t; z))]^2 h^*(t; z)}{S(t; z)^2}$$
$$= -h^*(t; z)\mathbb{V}\mathrm{ar}[U|X \geq t, Z = z] \leq 0$$

Example ($\Gamma(1/\theta, \theta)$ distribution): $\mathbb{E}[U|X \geq t, Z = z] = S(t; z)^\theta$

Interpretations of marginal and conditional hazards

$h(t; z, u)$: hazard specific for an individual with unobserved factor u.

$h(t; z)$: average hazard for the population remaining (still alive) at time t (i.e. subpopulation with $X \geq t$). This population is stronger than the original population which reduces hazard.

In particular, ratio between marginal hazard and conditional hazard is decreasing:

$$\frac{h(t; z)}{h(t; z, u)} = \frac{\mathbb{E}[U|X \geq t, Z = z]h^*(t; z)}{uh^*(t; z)} = \frac{\mathbb{E}[U|X \geq t, Z = z]}{u}$$

It is less than one for an average individual with $u = \mathbb{E}[U] = \mathbb{E}[U|X \geq 0, Z = z]$!

Example

Simple example: U either 1 or 2 each with probability $1/2$.
$h^*(t;z) = 1$.

Then (exercise !)

$$h(t;z) = 1\mathbb{E}[U|X \geq t, Z = z] = 1 + \frac{1}{1+e^t}$$

Thus $h(t;z)$ is $1.5 = \mathbb{E}[U]$ for $t = 0$ and decreasing afterwards !

From marginal to conditional: example with time-varying effect (Torben Martinussen)

Specify

$$h(t;z) = \exp(\beta_1 z 1[t \leq v])$$

I.e. timevarying effect of z. No effect for $t > v$.

Assume $U \sim \Gamma(1,1) = \text{Exp}(1)$. By results for Gamma distribution we can reverse engineer to find $h^*(t;z)$:

$$S(t;z) = (1+H^*(t;z))^{-1} \Leftrightarrow H^*(t;z) = S(t;z)^{-1}-1 = \exp(H(t;z))-1$$

Thus

$$h^*(t;z) = h(t;z)\exp(H(t;z)) = \begin{cases} \exp(\beta_1 z)\exp[\exp(\beta_1 z)t] & t \leq v \\ \exp[\exp(\beta_1 z)v + t - v] & t > v \end{cases}$$

Population versus individual

Suppose e.g. $\beta_1 < 0$. Then conditional (individual specific hazard ratio)

$$\frac{h(t;1,u)}{h(t;0,u)} = \frac{h^*(t;1)}{h^*(t;0)} < 1$$

for all t and u!

However hazard ratio at the population level

$$\frac{h(t;1)}{h(t;0)} = \exp(\beta_1 1[t \leq v]) \quad \text{is} \quad \begin{cases} < 1 & t \leq v \\ 1 & t > v \end{cases}$$

Thus conclusions holding at the population level (e.g. treatment not reducing hazard for $t > v$) may not be valid at the individual level.

This does not mean that population level hazard $h(t; z)$ is 'wrong'.

It means that conclusions based on marginal distributions and conditional distributions may differ.

It is similar in spirit to what is referred to as Simpson's paradox (try to look it up at Wikipedia).

We know human populations are heterogeneous so we need to be careful when interpreting population level hazard (or equivalently population level survival function).

‘Paradox’

Note

$$\frac{h(t;1)}{h(t;0)} = \frac{h^*(t;1)}{h^*(t;0)} \frac{\mathbb{E}[U|X \geq 1, Z = 1]}{\mathbb{E}[U|X \geq t, Z = 0]}$$

LHS is equal to one for $t > v$. First factor on RHS is < 1. So last factor must be > 1.

This means U is on average bigger when $Z = 1$ than when $Z = 0$ and $X \geq t$. I.e. population with $Z = 1$ more frail (bigger U on average) than $Z = 0$ - weaker individuals saved by treatment.

From an individual point of view treatment would always be beneficial.

However, a policy maker might notice that treatment effect vanishes at the population level when $t > v$.

Nevertheless better survival when $z = 1$: $S(t;1) \geq S(t;0)$ due to beneficial effect for $0 \leq t \leq v$.

Correlated data - shared frailty model

Suppose U represents genetic effect or effect of environment. Then group of closely related individuals may share the same U.

Suppose we have G groups each with n_i individuals sharing a frailty U_i. We assume $U_i \sim \Gamma(1/\theta, \theta)$ are independent and that all death times X_{ij} are independent given $U_1, \ldots, U_G$ with hazard function for jth individual in ith group given by

$$U_i h_0(t) \exp(\beta^\mathsf{T} z_{ij})$$

Joint survival function for ith group:

$$S_i(t_{i1}, \ldots, t_{in_i}) = \left(1 + \theta \sum_{j=1}^{n_i} H_0(t_{ij}) \exp(\beta^\mathsf{T} z_{ij})\right)^{-1/\theta}$$

This does not factorize - hence individuals in ith group are not independent ! (marginally) due to dependence on shared frailty. However, groups are independent.

Likelihood

Conditional likelihood assuming u_i known:

$$L(\beta, \theta | u_1, \ldots, u_G) = \prod_{i=1}^{G} L_i(\beta, \theta, u_i)$$

where $L_i(\beta, \theta, u_i)$ is conditional likelihood for ith group:

$$\begin{aligned} L_i(\beta, \theta, u_i) &= \prod_{j=1}^{n_i} (u_i h_0(t_{ij}) \exp(\beta^{\mathsf{T}} z_{ij}))^{\delta_{ij}} \exp(-u_i H_0(t_{ij}) \exp(\beta^{\mathsf{T}} z_{ij})) \\ &= u_i^{d_i} \exp(-u_i \sum_{j=1}^{n_i} H_0(t_{ij}) \exp(\beta^{\mathsf{T}} z_{ij})) \prod_{j=1}^{n_i} (h_0(t_{ij}) \exp(\beta^{\mathsf{T}} z_{ij}))^{\delta_{ij}} \end{aligned}$$

where $d_i = \sum_{j=1}^{n_i} \delta_{ij}$ (assuming right censored sample).

Marginal (observable) likelihood for ith group:

$$L_i(\beta, \theta) = \mathbb{E}[L_i(\beta, \theta | U_i)] = \prod_{j=1}^{n_i} (h_0(t_{ij}) \exp(\beta^\mathsf{T} z_{ij}))^{\delta_{ij}} \cdot$$

$$\int_0^\infty \frac{u_i^{1/\theta - 1 + d_i} \exp(-u_i(\theta^{-1} + \sum_{j=1}^{n_i} H_0(t_{ij}) \exp(\beta^\mathsf{T} z_{ij})))}{\Gamma(1/\theta)\theta^{1/\theta}} \mathrm{d}u_i =$$

$$\prod_{j=1}^{n_i} (h_0(t_{ij}) \exp(\beta^\mathsf{T} z_{ij}))^{\delta_{ij}} \frac{\Gamma(1/\theta + d_i)(1/\theta + \sum_{j=1}^{n_i} H_0(t_{ij}) \exp(\beta^\mathsf{T} z_{ij}))^{-1/\theta - d_i}}{\Gamma(1/\theta)\theta^{1/\theta}}$$

Case $n_i = 1$ ($d_i = \delta_i$ $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$):

$$L_i(\beta,\theta) =$$

$$(h_0(t_i)\exp(\beta^\mathsf{T} z_i))^{\delta_i} \frac{(1/\theta)^{\delta_i}(1/\theta)^{-1/\theta-\delta_i}}{\theta^{1/\theta}}(1+\theta H_0(t_i)\exp(\beta^\mathsf{T} z_i))^{-1/\theta-\delta_i}$$

$$= \left(\frac{h_0(t_i)\exp(\beta^\mathsf{T} z_i)}{1+\theta H_0(t_i)\exp(\beta^\mathsf{T} z_i)}\right)^{\delta_i}(1+\theta H_0(t_i)\exp(\beta^\mathsf{T} z_i))^{-1/\theta}$$

Consistent with previous results for population hazard and survival function for gamma frailty model !

Log likelihood (order as in KM 13.3.2):

$$d_i \log\theta - \log\Gamma(1/\theta) + \log\Gamma(1/\theta + d_i)$$

$$-(1/\theta + d_i)\log(1+\theta\sum_{j=1}^{n_i} H_0(t_{ij})\exp(\beta^\mathsf{T} z_{ij}))\sum_{j=1}^{n_i}\delta_{ij}[\beta^\mathsf{T} z_{ij} + \log h_0(t_{ij})]$$

If we assume a known form of H_0 - e.g. Weibull(α, λ) then we can maximize likelihood wrt. to all unknown parameters $\beta, \theta, \lambda, \alpha$ and proceed as usual for likelihood based inference.

If we want to use semi-parametric model where H_0 is unspecified we may proceed using EM algorithm as detailed in KM.

Frailty models can be fitted using `survreg` or `coxph` by adding `frailty` statement in model formula.

Correlated data - marginal approach

Suppose as before that $X_{i1}, \ldots, X_{in_i}$ are correlated observations, $i = 1, \ldots, G$ but observations from different groups are independent.

Suppose we are just interested in estimating the marginal hazard function of the X_{ij} and we assume the marginal hazard model

$$h_0(t) \exp(\beta^\mathsf{T} z_{ij})$$

(note we could reverse engineer to identify a corresponding frailty model)

The key point is that if we pretend the observations are independent and just apply the usual Cox partial likelihood then the resulting estimate of β is still consistent !

However, the covariance matrix of $\hat{\beta}$ is no longer the inverse Fisher information due to correlation within groups. In general inverse Fisher information underestimates variance !

A general perspective

Recall that if we obtain β from solving an estimating equation

$$u(\beta) = 0$$

then the approximate covariance matrix of β is

$$\mathbb{V}\mathrm{ar}\hat{\beta} \approx S^{-1} C S^{-\mathsf{T}}$$

where $S = -\mathbb{E}[\frac{\mathrm{d}}{\mathrm{d}\beta^{\mathsf{T}}} u(\beta)]$ is the sensitivity and $C = \mathbb{V}\mathrm{ar}\, u(\beta)$.

Suppose now that

$$u(\beta) = \sum_{i=1}^{G} u_i(\beta)$$

where the $u_i(\beta)$ are independent and identically distributed unbiased estimating functions.

Then

$$C = \mathbb{V}\mathrm{ar}\, u(\beta) = G\,\mathbb{V}\mathrm{ar}\, u_1(\beta)$$

where we can estimate

$$\mathbb{V}\mathrm{ar}\, u_1(\beta) \approx \frac{1}{G} \sum_{i=1}^{G} u_i(\hat{\beta}) u_i(\hat{\beta})^{\mathsf{T}} = \hat{C}/G$$

Thus we can approximate

$$\mathbb{V}\mathrm{ar}\, \hat{\beta} \approx S^{-1} \hat{C} S^{-\mathsf{T}}$$

(known as the ‘sandwich estimator’)

A simple example: linear regression with correlated errors

Suppose

$$Y^i = X\beta + \epsilon^i$$

are *iid* $m \times 1$ vectors each with covariance matrix Σ, $i = 1, \dots, G$.

Least squares estimation: minimize

$$\|Y - \tilde{X}\beta\|^2 = \sum_{i=1}^{G} \|Y^i - X\beta\|^2$$

with respect to β. Here $\tilde{X}$ consists of m times X stacked on top of each other.

Corresponding estimating function

$$u(\beta) = \tilde{X}^\mathsf{T}(Y - \tilde{X}\beta) = \sum_{i=1}^{G} X^\mathsf{T}(Y^i - X\beta) = \sum_{i=1}^{G} u_i(\beta)$$

Least squares estimate is

$$\hat{\beta} = (\tilde{X}^\mathsf{T}\tilde{X})^{-1}\tilde{X}^\mathsf{T}Y = (\sum_{i=1}^{G} X^T X)^{-1}\sum_{i=1}^{G} X^\mathsf{T} Y_i$$

Estimate is not optimal since it ignores correlation within groups.

However, easy to see that estimate is unbiased $\mathbb{E}[\hat{\beta}] = \beta$.

The variance of $\hat{\beta}$ is ($\tilde{\Sigma} = \mathbb{C}\mathrm{ov}(Y)$ is blockdiagonal)

$\mathbb{V}\mathrm{ar}[\hat{\beta}] =$

$$(\tilde{X}^\mathsf{T}\tilde{X})^{-1}\tilde{X}^\mathsf{T}\tilde{\Sigma}\tilde{X}(\tilde{X}^T\tilde{X})^{-1} = (\sum_{i=1}^{G} X^T X)^{-1}[\sum_{i=1}^{G} X^\mathsf{T}\Sigma X](\sum_{i=1}^{G} X^T X)^{-1}$$

$$= \frac{1}{G}(X^T X)^{-1}\mathbb{V}\mathrm{ar} u_1(\beta)(X^T X)^{-1}$$

In practice we do not know $\mathbb{V}\mathrm{ar} u_1(\beta) = X^\mathsf{T} \Sigma X$. However, it can be estimated by

$$\frac{1}{G}\sum_{i=1}^{G} u_i(\hat{\beta}) u_i(\hat{\beta})^\mathsf{T} = \frac{1}{G}\sum_{i=1}^{G} X^T (Y^i - X\hat{\beta})(Y_i - X\hat{\beta})^\mathsf{T} X = X^\mathsf{T} \hat{\Sigma} X$$

where $\hat{\Sigma}$ is the empirical estimate of Σ.

Note: the assumption of identical design matrix X for all groups can be relaxed by replacing X by X^i where the X^i are *iid* design matrices and ϵ^i is independent of X^i.

Back to marginal survival model

In this case S coincides with the Fisher information (called V in KM) and $u_i(\beta)$ corresponds to the partial log likelihood score for the ith group.

We thus arrive at the approximate covariance matrix $\tilde{V}$ on page 437 in KM.

Can be implemented using `cluster()` in connection with `coxph`.

Caution - marginal vs conditional

Consider a gamma $\Gamma(1/\theta, \theta)$ frailty model with

$$h(t|U, z) = U h_0(t) \exp(\beta^\mathsf{T} z)$$

Then the marginal hazard is

$$h(t|z) = h_0(t) \exp(\beta^\mathsf{T} z) \frac{1}{1 + \theta \exp(\beta^\mathsf{T} z) H_0(t)} \neq h_0(t) \exp(\beta^\mathsf{T} z)$$

So it does not really make sense to compare results obtained with `coxph+frailty` with results obtained with `coxph+cluster` !

Exercises

1. Show that the conditional density of $U|X \geq t, Z = z$ is $f(u|t,z) = S(t;z,u)f_U(u)/S(t;z)$.

   Hint: show that

   $$P(U \in A, Z \in B, X \geq t) = \int_A \int_B f(u|t,z) P(X \geq t|Z = z) f(z) \mathrm{d}z \mathrm{d}u$$

   See also last two slides on conditional distributions.
2. Check expression for derivative $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[U|X \geq t, Z = z]$.
3. Show that $\mathbb{E}[U|X \geq t, Z = z] = S(t;z)^\theta$ when $U \sim \Gamma(1/\theta, \theta)$.

4. Assume U is either 1 or 2 each with probability $1/2$ and $h^*(t; z) = 1$. Show that

$$h(t; z) = 1 + \frac{1}{1 + e^t}$$

5. Go through derivations leading from conditional likelihood $L_i(\beta, \theta, u_i)$ to marginal likelihood $L_i(\beta, \theta)$ in case $n_i > 1$.

Conditional distributions

Given two random quantities X and Y taking values in sets M and N, $P(\cdot|\cdot)$ is said to be a conditional distribution of X given Y if for $A \subseteq M$ and $B \subseteq N$,

$$P(X \in A, Y \in B) = \int_B P(A|y) f(y) \mathrm{d}y.$$

Likewise, $f(\cdot|\cdot)$ on $M \times N$ is said to be a conditional density of X given Y if

$$P(X \in A, Y \in B) = \int_B \int_A f(x|y) \mathrm{d}x f(y) \mathrm{d}y.$$

Note that the laws of total probability are just consequences of these definitions.

E.g. if X and Y are continuous random variables with joint density $f(x, y)$ and marginal density $f(y)$ for Y, one can check that $f(x, y)/f(y)$ fulfills the requirement for being a conditional density of X given Y.

To handle exercise 1 note that $X \geq t$ is equivalent to $Y = 1$ where $Y = 1[X \geq t]$. To show that

$$f(u|y,z) = \frac{P(Y = y|U = u, Z = z)f(u)}{P(Y = y|Z = z)}$$

is a conditional density of U given Y and Z we need to verify for all appropriate A, B and C,

$$P(U \in A, Z \in B, Y \in C) = \int_B \sum_{y \in C} \int_A f(u|y,z) \mathrm{d}u p(y,z) \mathrm{d}z$$

This is equivalent to

$$\sum_{y \in C} P(U \in A, Z \in B, Y = y) = \sum_{y \in C} \int_B \int_A f(u|y,z) \mathrm{d}u p(y,z) \mathrm{d}z$$

Thus we need to show for each $y \in \{0,1\}$ that

$$P(U \in A, Z \in B, Y = y) = \int_B \int_A f(u|y,z) \mathrm{d}u p(y,z) \mathrm{d}z$$

For $y = 1$, this is the equality in the hint of exercise 1.