Mutlivariate and Joint Models for Longitudinal Data

Longitudinal studies are commonly designed in many research fields in order to see changes over a time interval shared by all participants. Joint modeling consists of two interlinked sub-models with any type of outcome (continuous, binomial, etc). One of the most commonly used is longitudinal sub-model is the linear mixed effect model and a cox proportional hazard with time-to-event sub-model.

Joint modeling reduces the bias of parameter estimates by accounting for the association between the longitudinal and time-to-event data. In clinical trials this often leads to more efficient estimation of the treatment effect on both sub-model outcomes.

Let's assume Y_{i1} and Y_{i2} are two outcomes measured on subject i. We can attempt to specify a joint density $f(y_{i1}, y_{i2})$, but this is only feasible if we assume certain things about the marginal association among the longitudinally measured elements within each of the outcome vectors. This is easier when both outcomes and conditional models are of the same type, but this is not a requirement. When Y_{i1} and Y_{i2} are of different types, or in the case of unbalanced data, this becomes cumbersome. Thus, extensions to higher dimensional data involves considerable challenges as this would require assumption on larger covariance structure and higher order associations.

Conditional Models

We can avoid direct specification of the joint density $f(y_{i1}, y_{i2})$ are and reduce the modeling tasks by specifying a model for each outcome separately:

$$\begin{aligned} f(y_{i1}, y_{i2}) &= f(y_{i1} \mid y_{i2})\, f(y_{i2}) \\ &= f(y_{i2} \mid y_{i1})\, f(y_{i1}) \end{aligned}$$

The main drawback is that this requires reflection about the plausible association between Y_{i1} and Y_{i2} that may be inappropriate in some settings. For example, in a clinical trial with two main post-randomization outcomes conditioning on one of the outcomes may attenuate (make smaller) the treatment effect on the other. Also, it is difficult to implement in high-dimension data.

Shared-Parameter Models

In previous chapters we've seen how random effects can be used to generate an association structure between repeated measurements of a specific outcome. We can use a similar approach to generate association between two outcomes.

\(f(y_{i1}, y_{i2}) = \int f(y_{i1}, y_{i2} | b) f(b) db = \int f(y_{i1} | b) f(y_{i2} | b) f(b) db \)

Where b is a vector of ranom effects common to the outputs Y_{i1} and Y_{i2}, and assume independence of both outcomes conditional on b.

Again the outcomes do not have to be of the same type, but we assume b represents a common set of unobserved characteristics of the subject that governs both models.

Examples:

Let Y_{i1} represent a vector of longitudinal continuous outcomes and $Y_{i2} = (U_i, \square_i)$ with $U = \min(T_i, C_i)$ with T_i as the potential time to the event of interest, C_i as the potential censoring time, and $\square_i = I(T_i <= C_i)$. Let Y_{i1} follow a random intercept model:
$Y = \beta_0 + b_{0i} + \beta_1 t + \varepsilon_{i1t}$
and let the event time of interest follow a proportional hazard model so that:

$$\lim_{h \to 0} \Pr\{t_i \leq T_i \leq t_i + h | T_i \geq t_i, b_{0i}\} = \lambda_0(t) \exp(\alpha b_{0i})$$

where $\lambda_0(t)$ is an unspecified baseline hazard function

In another case, the shared-parameter random effect models may imply very strong assumptions for the association structure between outcomes. Consider this example where Y_{i1} and Y_{i2} are two longitudinal continuous outcomes that share a common random intercept:

$$\begin{aligned} y_{i1t} &= \beta_1 + b_{1i} + \beta_2 t + e_{i1t} \qquad e_{i1t} \sim N(0, \sigma_1^2) \\ y_{i2t} &= \beta_3 + \gamma b_{1i} + \beta_4 t + e_{i2t} \qquad e_{i2t} \sim N(0, \sigma_2^2) \end{aligned}$$

with $\mathrm{Var}(b_{0i}) = \sigma_b^2$. Then

$$\begin{aligned} Corr(Y_{1s}, Y_{1t}) &= \frac{\sigma_b^2}{\sigma_b^2 + \sigma_1^2} \\ Corr(Y_{2s}, Y_{2t}) &= \frac{\gamma^2 \sigma_b^2}{\gamma^2 \sigma_b^2 + \sigma_2^2} \\ Corr(Y_{1s}, Y_{2t}) &= \frac{\gamma \sigma_b^2}{\sqrt{\sigma_b^2 + \sigma_1^2}\sqrt{\gamma^2 \sigma_b^2 + \sigma_2^2}} \\ &= \sqrt{Corr(Y_{1s}, Y_{1t})}\sqrt{Corr(Y_{2s}, Y_{2t})} \end{aligned}$$

The joint random-effects imply more flexible correlation patterns. Consider this example where Y_{i1} and Y_{i2} are dependent on separate random effects b1 and b2 which are themselves correlated:

$$\begin{aligned} y_{i1t} &= \beta_1 + b_{1i} + \beta_2 t + e_{i1t} \qquad e_{i1t} \sim N(0, \sigma_1^2) \\ y_{i2t} &= \beta_3 + b_{2i} + \beta_4 t + e_{i2t} \qquad e_{i2t} \sim N(0, \sigma_2^2) \end{aligned}$$

with $\begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \sim \mathrm{N}(0, \mathrm{G})$. Then

$$\begin{aligned} Corr(Y_{1s}, Y_{1t}) &= \frac{Var(b_1)}{Var(b_1) + \sigma_1^2} \\ Corr(Y_{2s}, Y_{2t}) &= \frac{Var(b_2)}{Var(b_2) + \sigma_2^2} \\ Corr(Y_{1s}, Y_{2t}) &= \frac{Cov(b_1, b_2)}{\sqrt{Var(b_1) + \sigma_1^2}\sqrt{Var(b_2) + \sigma_2^2}} \\ &= \frac{Cov(b_1, b_2)}{\sqrt{Var(b_1)}\sqrt{Var(b_2)}} \frac{\sqrt{Var(b_1)}\sqrt{Var(b_2)}}{\sqrt{Var(b_1) + \sigma_1^2}\sqrt{Var(b_2) + \sigma_2^2}} \\ &= Corr(b_1, b_2)\sqrt{Corr(Y_{1s}, Y_{1t})}\sqrt{Corr(Y_{2s}, Y_{2t})} \\ &\leq \sqrt{Corr(Y_{1s}, Y_{1t})}\sqrt{Corr(Y_{2s}, Y_{2t})} \end{aligned}$$

This demonstrates the role of the correlation between the outcome-specific random effects in the between-process outcome correlation at any two time points. The restriction is imposed by the shared-parameter model is relaxed in the same sense that the model no longer assumes the product of within process correlations equals the between-process correlation and thus allows a more general dependence structure.

In many situations (especially when we have multiple outcomes/high-dimensional data) the previous methods will not be feasible to estimate. Methods that reduce the dimensions are more appropriate. The idea is to use a factor-analytic or principal component type of analysis to reduce the dimensions. After, the estimated factors can be analyzed using any of the classical longitudinal models. The disadvantage is that the inferences are restricted to the factors and cannot be extended to the original outcome variables, so certain assumptions need to be imposed in the case of unequally spaced observations or unbalanced designs.

Defining the Generalized Linear Models

For a bi-variate response $Y_i = (Y_{i1}, Y_{i2})'$, we can assume a general model in the form of:

$$\begin{aligned} Y_i &= \boldsymbol{\mu}_i(\boldsymbol{\eta}_i) + \boldsymbol{\epsilon}_i \\ \boldsymbol{\mu}_i(\boldsymbol{\eta}_i) &= \boldsymbol{g}^{-1}(X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i) \end{aligned}$$

with g^{-1} as the inverse link function which is allowed to change for each univariate outcome and $b_i \sim N(0, G)$.

A general first-order approximate expression for the variance-covariance matrix of Y_i is:

$$Var(Y_i) \sim= \Delta_i Z_i GZ'_i\Delta'_i + V_i$$

Where

$$\Delta_i = \left(\frac{\partial \boldsymbol{\mu}(\boldsymbol{\eta}_i)}{\partial \boldsymbol{\eta}_i}\right)\Bigg|_{\boldsymbol{b}_i=0} \text{ and } \quad V_i \simeq \Xi_i^{1/2} A_i^{\frac{1}{2}} R(\alpha) A_i^{\frac{1}{2}} \Xi_i^{1/2}$$

and $R(\alpha)$ is the correlation matrix and the Ξ_i is a diagonal matrix with over-dispersion parameters along the diagonal:

$$A_i = \begin{bmatrix} v(\mu_{i1}) & 0 & \cdots & 0 \\ 0 & v(\mu_{i2}) & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & v(\mu_{ip}) \end{bmatrix}$$

If there are no random effects then the model is referred to as a **marginal generalized linear model (MGLM).** For the special case of a continuous and binary endpoint the MGLM is in the form of:

$$\begin{pmatrix} Y_{1i} \\ Y_{2i} \end{pmatrix} = \begin{pmatrix} \alpha_0 + \alpha_1 X_i \\ \frac{\exp(\beta_0+\beta_1 X)}{1+\exp(\beta_0+\beta_1 X)} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix}$$

with

$$\Xi_i = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 1 \end{pmatrix}, A_i = \begin{pmatrix} 1 & 0 \\ 0 & v_{i2} \end{pmatrix}, R(\alpha) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

$v_{i2} = \mu_{i2}(1-\mu_{i2})$

In a shared parameter the variance of Y_i is approximately equal to:

$$\begin{aligned} Var(Y_i) &= \begin{pmatrix} \lambda^2 & v_{i2}\lambda \\ v_{i2}\lambda & v_{i2}^2 \end{pmatrix}\tau^2 + \begin{pmatrix} \sigma^2 & \rho\sigma\sqrt{v_{i2}} \\ \rho\sigma\sqrt{v_{i2}} & v_{i2} \end{pmatrix} \\ &= \begin{pmatrix} \lambda^2\tau^2 + \sigma^2 & v_{i2}\lambda\tau^2 + \rho\sigma\sqrt{v_{i2}} \\ v_{i2}\lambda\tau^2 + \rho\sigma\sqrt{v_{i2}} & v_{i2}^2\tau^2 + v_{i2} \end{pmatrix} \end{aligned}$$

As a result the approximate correlation among the two outcomes is:

$$\rho_{12} = \frac{v_{i2}\lambda\tau^2 + \rho\sigma\sqrt{v_{i2}}}{\sqrt{\lambda^2\tau^2 + \sigma^2}\sqrt{v_{i2}^2\tau^2 + v_{i2}}}$$

The correlation depends on the fixed effects through v_{i2}

SAS Code

```
libname s857 "/Datasets";
options fmtsearch=(s857.formats);
options nofmterr;

proc mixed data=s857.Tbi;
class sex naccid  status;
model Mem1=time  naccageb status sex time*educ time*naccageb  /solution;
random Int time/sub=naccid;
run;

proc mixed data=s857.Tbi;
class sex naccid  status;
model Att1=time  naccageb status sex time*educ time*naccageb  /solution;
random Int time/sub=naccid;
run;

/* Multivariate models  */
data s857.tbi_long;
length scale $10;
set s857.tbi;
Outcome=Mem1;scale='Memory';output;
Outcome=Att1;scale='Attention';output;
keep Outcome Scale time trauma_dich dep2yrs status sex educ naccageb naccid;
```

```
run;
proc mixed data=s857.tbi_long covtest method=ML;
class Scale sex naccid  status;
model Outcome=Scale*time Scale*naccageb Scale*status Scale*time*status
 Scale*sex Scale*time*educ Scale*time*naccageb  /solution;
random Int time/sub=naccid;
*random int/sub=Scale type=un;
run;

proc nlmixed data=s857.tbi_long;
parms b0=-0.17 b1=0.43 b2=0.39 b3=-0.83 b4=0.08 b5=-0.01
s2e_mem=0.361 g11=0.60
 c0=0.05 c1=-0.05 c2=0.5 c3=0.1
s2e_att=0.19;
if Scale='Memory' then do;
 mean=b0+b1*time + b2*status + b3*sex +b4*time*status +b5*time*naccageb  +u1 ;
 ll = -0.5*log(3.14159265358) - log(sqrt(s2e_mem)) -0.5*(outcome-mean)**2/(s2e_mem);
end;
if Scale='Attention' then do;
 mean=c0+c1*time + c2*status + c3*time*status  +u1;
 ll = -0.5*log(3.14159265358) - log(sqrt(s2e_att)) -0.5*(outcome-mean)**2/(s2e_att);
end;
model outcome~general(ll);
random  u1~normal([0],[g11]) subject=naccid;
run;

/*Shared parameter model*/
title 'Memory';
proc mixed data=s857.tbi covtest method=ML;
class trauma_dich sex naccid  status;
model Mem1=time status  sex time*status time*naccageb  /solution;
random intercept/sub=naccid;
run;
title 'Attention';
proc mixed data=s857.tbi covtest method=ML;
class trauma_dich sex naccid  status;
model Att1=time /solution;
random intercept/sub=naccid;
run;

proc nlmixed data=s857.tbi_long;
```

```
parms b0=-0.13 b1=0.43 b2=0.39 b3=-0.83 b4=0.08 b5=-0.01
s2e_mem=0.361 g11=0.60
 c0=0.05 c1=-0.05 c2=0.5 c3=0.1 gamma=0.20
s2e_att=0.19;
if Scale='Memory' then do;
 mean=b0+b1*time + b2*status + b3*sex +b4*time*status +b5*time*naccageb  +u1 ;
 ll = -0.5*log(3.14159265358) - log(sqrt(s2e_mem)) -0.5*(outcome-mean)**2/(s2e_mem);
end;
if Scale='Attention' then do;
 mean=c0+c1*time + c2*status + c3*time*status  +gamma*u1;
 ll = -0.5*log(3.14159265358) - log(sqrt(s2e_att)) -0.5*(outcome-mean)**2/(s2e_att);
end;
model outcome~general(ll);
random  u1~normal([0],[g11]) subject=naccid;
run;
proc nlmixed data=s857.tbi_long;
parms b0=-0.13 b1=0.43 b2=0.39 b3=-0.83 b4=0.08 b5=-0.01
s2e_mem=0.361 g11=0.60 g22=.01
 c0=0.05 c1=-0.05  c2=-0.3 c3=-0.1 gamma=1 g33=0.01
s2e_att=0.19;
if Scale='Memory' then do;
 mean=b0+b1*time + b2*status + b3*sex +b4*time*status +b5*time*naccageb  +u1 +u2*time;
 ll = -0.5*log(3.14159265358) - log(sqrt(s2e_mem)) -0.5*(outcome-mean)**2/(s2e_mem);
end;
if Scale='Attention' then do;
 mean=c0+c1*time + c2*status + c3*time*status  +gamma*u1 +u3*time;
 ll = -0.5*log(3.14159265358) - log(sqrt(s2e_att)) -0.5*(outcome-mean)**2/(s2e_att);
end;
model outcome~general(ll);
random  u1 u2 u3~normal([0,0,0],[g11,0,g22,0,0,g33]) subject=naccid;
run;
proc nlmixed data=s857.tbi_long;
parms b0=-0.13 b1=0.43 b2=0.39 b3=-0.83 b4=0.08 b5=-0.01
s2e_mem=0.361 g11=0.60 g22=.01 g12=0.01
 c0=0.05 c1=-0.05  c2=-0.3 c3=-0.1
s2e_att=0.19;
if Scale='Memory' then do;
 mean=b0+b1*time + b2*status + b3*sex +b4*time*status +b5*time*naccageb  +u1 ;
 dens = -0.5*log(3.14159265358) - log(sqrt(s2e_mem)) -0.5*(outcome-mean)**2/(s2e_mem);
 ll=dens;
end;
```

```
if Scale='Attention' then do;
 mean=c0+c1*time + c2*status + c3*time*status  +u2;
 dens = -0.5*log(3.14159265358) - log(sqrt(s2e_att)) -0.5*(outcome-mean)**2/(s2e_att);
 ll=dens;
end;
model outcome~general(ll);
random  u1 u2~normal([0,0],[g11,g12,g22]) subject=naccid;
run;

/*
if dist='Binary' then do;
  eta=v0 + v1*u1   ;
  p=exp(eta)/(1+exp(eta));
  ll = (outcome)*log(p) + (1-outcome)*log(1-p);
end;
*/
```

Revision #10
Created 30 March 2023 19:31:02 by Elkip
Updated 31 March 2023 20:17:40 by Elkip