Joint modelling of multivariate longitudinal and time-to-event data

Graeme L. Hickey\textsuperscript{1}, Pete Philipson\textsuperscript{2}, Andrea Jorgensen\textsuperscript{1}, Ruwanthi-Kolamunnage-Dona\textsuperscript{1}

\textsuperscript{1}Department of Biostatistics, University of Liverpool, UK
\textsuperscript{2}Mathematics, Physics and Electrical Engineering, University of Northumbria, UK

graeme.hickey@liverpool.ac.uk

3rd July 2017
Motivation for *multivariate* joint models

- Clinical studies often repeatedly measure *multiple* biomarkers or other measurements and an event time
- As of 2015: $> 200$ methodological papers and $> 60$ clinical applications of joint models (Sudell et al. 2016)
- Many software packages available for univariate models, but not for multivariate models (Hickey et al. 2016) $\Rightarrow$ limits many applied statisticians to univariate models only
Motivation for *multivariate* joint models

- Clinical studies often repeatedly measure *multiple* biomarkers or other measurements and an event time.
- As of 2015: > 200 methodological papers and > 60 clinical applications of joint models (Sudell et al. 2016).
- Many software packages available for univariate models, but not for multivariate models (Hickey et al. 2016) ⇒ limits many applied statisticians to univariate models only.

**Objective**

Develop an [R](https://www.r-project.org/) package that extends the existing capabilities of `joineR`. 
For each subject $i = 1, \ldots, n$, we observe

- $y_i = (y_{i1}, \ldots, y_{iK})^\top$ is a $K$-variate continuous outcome vector, where each $y_{ik}$ denotes an $(n_{ik} \times 1)$-vector of observed longitudinal measurements for the $k$-th outcome type:
  
  \[ y_{ik} = (y_{i1k}, \ldots, y_{in_{ik}k})^\top \]

- Observation times $t_{ijk}$ for $j = 1, \ldots, n_{ik}$, which can differ between subjects and outcomes

- $(T_i, \delta_i)$, where $T_i = \min(T_i^*, C_i)$, where $T_i^*$ is the true event time, $C_i$ corresponds to a potential right-censoring time, and $\delta_i$ is the failure indicator equal to 1 if the failure is observed ($T_i^* \leq C_i$) and 0 otherwise
Following Henderson et al. (2000) for the **univariate case**

\[ y_i(t) = \mu_i(t) + W_{1i}(t) + \varepsilon_i(t), \]

where

- \( \varepsilon_i(t) \) is the model error term, which is i.i.d. \( N(0, \sigma^2) \) and independent of \( W_{1i}(t) \)
- \( \mu_i(t) = x_i^\top(t)\beta \) is the mean response
- \( x_i(t) \) is a \( p \)-vector of (possibly) time-varying covariates with corresponding fixed effect terms \( \beta \)
- \( W_{1i}(t) \) is a zero-mean **latent** Gaussian process
Longitudinal sub-model

We can extend it to $K$-separate sub-models (with $k = 1, \ldots, K$)

$$y_{ik}(t) = \mu_{ik}(t) + W_{1i}^{(k)}(t) + \epsilon_{ik}(t),$$

where

- $\epsilon_{ik}(t)$ is the model error term, which is i.i.d. $N(0, \sigma_k^2)$ and independent of $W_{1i}^{(k)}(t)$
- $\mu_{ik}(t) = x_{ik}^\top(t)\beta_k$ is the mean response
- $x_{ik}(t)$ is a $p_k$-vector of (possibly) time-varying covariates with corresponding fixed effect terms $\beta_k$
- $W_{1i}^{(k)}(t)$ is a zero-mean latent Gaussian process
Time-to-event sub-model

\[ \lambda_i(t) = \lambda_0(t) \exp \left\{ v_i^\top(t) \gamma_v + W_{2i}(t) \right\}, \]

where

- \( \lambda_0(\cdot) \) is an unspecified baseline hazard function
- \( v_i(t) \) is a \( q \)-vector of (possibly) time-varying covariates with corresponding fixed effect terms \( \gamma_v \)
- \( W_{2i}(t) \) is a zero-mean latent Gaussian process, independent of the censoring process
Following Laird and Ware (1982):

\[ W_{1i}^{(k)}(t) = z_{ik}(t)b_{ik} \text{ for } k = 1, \ldots, K \]
Following Laird and Ware (1982):

\[ W^{(k)}_{1i}(t) = z_{ik}^\top(t)b_{ik} \text{ for } k = 1, \ldots, K \]

Three sources of correlation:

1. Within-subject correlation between longitudinal measurements:
   \[ b_{ik} \sim N(0, D_{kk}) \]

2. Between longitudinal outcomes correlation: \( \text{cov}(b_{ik}, b_{il}) = D_{kl} \) for \( k \neq l \)

3. Correlation between sub-models\(^1\):
   \[ W_{2i}(t) = \sum_{k=1}^{K} \gamma_{yk} W^{(k)}_{1i}(t) \]

\(^1\)Extends model proposed Henderson et al. 2000, although many other \( W_{2i}(t) \)
   specifications have been proposed in literature
Likelihood

We can re-write the longitudinal sub-model as

\[ y_i \mid b_i, \beta, \Sigma_i \sim N(X_i\beta + Z_ib_i, \Sigma_i), \text{ with } b_i \mid D \sim N(0, D), \]

where \( \beta = (\beta_1^T, \ldots, \beta_K^T) \), \( b_i = (b_{i1}^T, \ldots, b_{iK}^T)^T \), and

\[
X_i = \begin{pmatrix}
X_{i1} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & X_{iK}
\end{pmatrix}, \quad
D = \begin{pmatrix}
D_{11} & \cdots & D_{1K} \\
\vdots & \ddots & \vdots \\
D_{1K}^\top & \cdots & D_{KK}
\end{pmatrix}
\]

\[
Z_i = \begin{pmatrix}
Z_{i1} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & Z_{iK}
\end{pmatrix}, \quad
\Sigma_i = \begin{pmatrix}
\sigma_1^2 I_{n_{i1}} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \sigma_K^2 I_{n_{iK}}
\end{pmatrix}
\]
MCEM algorithm

- A standard estimation approach for
\[ \theta = (\beta^\top, \text{vech}(D), \sigma_1^2, \ldots, \sigma_K^2, \lambda_0(t), \gamma_v^\top, \gamma_y^\top) \] is the EM algorithm treating \( b_i \) as missing data

- Quadrature can be slow and difficult if \( \text{dim}(b_i) \) is large \( \Rightarrow \) might not scale well as \( K \) increases

- Instead, we use the Monte Carlo Expectation-Maximization (MCEM; Wei and Tanner 1990)

- M-step updates similar to the univariate case (Wulfsohn and Tsiatis 1997; Lin et al. 2002)
Conventional EM algorithm: use quadrature to compute

\[ \mathbb{E} \left[ h(b_i) \mid T_i, \delta_i, y_i; \hat{\theta} \right] = \frac{\int_{-\infty}^{\infty} h(b_i)f(b_i \mid y_i; \hat{\theta})f(T_i, \delta_i \mid b_i; \hat{\theta})db_i}{\int_{-\infty}^{\infty} f(b_i \mid y_i; \hat{\theta})f(T_i, \delta_i \mid b_i; \hat{\theta})db_i}, \]

where

\[ h(\cdot) = \text{any known function}, \]

\[ b_i \mid y_i, \theta \sim N \left( A_i \left\{ Z_i^\top \Sigma_i^{-1} (y_i - X_i \beta) \right\}, A_i \right), \]

and

\[ A_i = \left( Z_i^\top \Sigma_i^{-1} Z_i + D^{-1} \right)^{-1}. \]
Monte Carlo E-step

**MCEM algorithm E-step:** use Monte Carlo integration to compute

\[
\mathbb{E} \left[ h(b_i) \mid T_i, \delta_i, y_i; \hat{\theta} \right] \approx \frac{1}{N} \sum_{d=1}^{N} h \left( b_i^{(d)} \right) \frac{f \left( T_i, \delta_i \mid b_i^{(d)}; \hat{\theta} \right)}{\frac{1}{N} \sum_{d=1}^{N} f \left( T_i, \delta_i \mid b_i^{(d)}; \hat{\theta} \right)}
\]

where

\[
h(\cdot) = \text{any known function},
\]

\[
b_i \mid y_i, \theta \sim N \left( A_i \left\{ Z_i^\top \Sigma_i^{-1} (y_i - X_i \beta) \right\}, A_i \right), \text{ and}
\]

\[
A_i = \left( Z_i^\top \Sigma_i^{-1} Z_i + D^{-1} \right)^{-1}
\]

\[
b_i^{(1)}, b_i^{(2)}, \ldots, b_i^{(N)} \sim b_i \mid y_i, \theta \text{ a Monte Carlo draw}
\]
Monte Carlo E-step

As proposed by Henderson et al. (2000), we use antithetic simulation
for variance reduction instead of directly sampling from the MVN
distribution for $b_i \mid y_i; \theta$:

Sample $\Omega \sim N(0, I_r)$ and obtain the pairs

$$A_i \left\{ Z_i^\top \Sigma_i^{-1} (y_i - X_i \beta) \right\} \pm C_i \Omega,$$

where $C_i$ is the Cholesky decomposition of $A_i$ such that $C_i C_i^\top = A_i$

Negative correlation between the pairs $\Rightarrow$ smaller variance in the
sample means than would be obtained from $N$ independent
simulations
The monotonicity property of the EM algorithm is lost in the MCEM algorithm, so we need to increase $N$ at each iteration.

**Problem**

- Using large $N$ when far from maximizer = computationally inefficient
- Using small $N$ when close to maximizer = unlikely to detect convergence
Convergence

The monotonicity property of the EM algorithm is lost in the MCEM algorithm, so we need to increase $N$ at each iteration.

Solution (Ripatti et al. 2002)

After a ‘burn-in’ phase, calculate the coefficient of variation statistic

$$cv(\Delta_{\text{rel}}^{(m+1)}) = \frac{sd(\Delta_{\text{rel}}^{(m-1)}, \Delta_{\text{rel}}^{(m)}, \Delta_{\text{rel}}^{(m+1)})}{mean(\Delta_{\text{rel}}^{(m-1)}, \Delta_{\text{rel}}^{(m)}, \Delta_{\text{rel}}^{(m+1)})},$$

and increase $N$ to $N + \lfloor N/\delta \rfloor$ if $cv(\Delta_{\text{rel}}^{(m+1)}) > cv(\Delta_{\text{rel}}^{(m)})$ for some small positive integer $\delta$, where

$$\Delta_{\text{rel}}^{(m+1)} = \max \left\{ \frac{|\hat{\theta}(m+1) - \hat{\theta}(m)|}{|\hat{\theta}(m)| + \epsilon} \right\}.$$
Standard error estimation

Method 1: Bootstrap

Hsieh et al. (2006) demonstrated that the profile likelihood approach in the EM algorithm leads to underestimation in the SEs, so recommended bootstrapping.
Standard error estimation

**Method 1: Bootstrap**

Hsieh et al. (2006) demonstrated that the profile likelihood approach in the EM algorithm leads to underestimation in the SEs, so recommended bootstrapping.

**Method 2: Empirical information matrix approximation**

Following McLachlan and Krishnan (2008), $\text{SE}(\theta) \approx I_e^{-1/2}(\hat{\theta})$, where

$$I_e(\theta) = \sum_{i=1}^{n} s_i(\theta)s_i^\top(\theta) - \frac{1}{n} S(\theta)S^\top(\theta),$$

$S(\theta) = \sum_{i=1}^{n} s_i(\theta)$ is the score vector.
**joineRML**

Rich collection of associated methods

*associated with additional plot methods*

- `getVarCov()`
- `vcov()`
- `fixef()`
- `ranef()`*
- `AIC()`
- `BIC()`
- `confint()`
- `formula()`
- `sampleData()`
- `dynSurv()`*
- `dynLong()`*
- `print()`
- `summary()`
- `plot()`
- `sigma()`
- `coef()`
- `update()`
- `baseHaz()`
- `residuals()`
- `fitted()`
- `logLik()`
- `bootSE()`

**Version 0.2.2 available on CRAN**

[https://cran.r-project.org/web/packages/joineRML/](https://cran.r-project.org/web/packages/joineRML/)

**Developmental version available on GitHub**

[https://github.com/graemeleehickey/joineRML](https://github.com/graemeleehickey/joineRML)
Primary biliary cirrhosis (PBC) is a chronic liver disease characterized by inflammatory destruction of the small bile ducts, which eventually leads to cirrhosis of the liver (Murtaugh et al. 1994)

We analyse a subset of 154 patients randomized to placebo

Multiple biomarkers repeatedly measured at intermittent times:

1. serum bilirunbin (mg/dl)
2. serum albumin (mg/dl)
3. prothrombin time (seconds)

Patients drop out if they die
Joint modelling of multivariate data
$n_{\text{events}} = 69, 44.8\%$

Time from registration (years)

Survival probability

$\text{GL. Hickey}$

Joint modelling of multivariate data
Proposed model for PBC data

Longitudinal sub-model

\[
\begin{align*}
\log(\text{serBilir}) &= (\beta_{0,1} + b_{0i,1}) + (\beta_{1,1} + b_{1i,1})\text{year} + \varepsilon_{ij1}, \\
\text{albumin} &= (\beta_{0,2} + b_{0i,2}) + (\beta_{1,2} + b_{1i,2})\text{year} + \varepsilon_{ij2}, \\
(0.1 \times \text{prothrombin})^{-4} &= (\beta_{0,3} + b_{0i,3}) + (\beta_{1,3} + b_{1i,3})\text{year} + \varepsilon_{ij3}, \\
\end{align*}
\]

\[b_i \sim N_6(0, D), \quad \text{and } \varepsilon_{ijk} \sim N(0, \sigma_k^2) \text{ for } k = 1, 2, 3\]

Time-to-event sub-model

\[
\begin{align*}
\lambda_i(t) &= \lambda_0(t) \exp \{\gamma_{\text{age}} + W_{2i}(t)\}, \\
W_{2i}(t) &= \gamma_{\text{bil}} (b_{0i,1} + b_{1i,1}t) + \gamma_{\text{alb}} (b_{0i,2} + b_{1i,2}t) + \gamma_{\text{pro}} (b_{0i,3} + b_{1i,3}t)
\end{align*}
\]
data(pbc2)
placebo <- subset(pbc2, drug == "placebo")
fit.pbc <- mjoint(
  formLongFixed = list(
    "bil" = log(serBilir) ~ year,
    "alb" = albumin ~ year,
    "pro" = (0.1 * prothrombin)^-4 ~ year),
  formLongRandom = list(
    "bil" = ~ year | id,
    "alb" = ~ year | id,
    "pro" = ~ year | id),
  formSurv = Surv(years, status2) ~ age,
  data = placebo,
  timeVar = "year",
  control = list(tol0 = 0.001, burin = 400))
## Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{0,1}$</td>
<td>0.5541</td>
<td>0.0858</td>
<td>(0.3859, 0.7223)</td>
</tr>
<tr>
<td>$\beta_{1,1}$</td>
<td>0.2009</td>
<td>0.0201</td>
<td>(0.1616, 0.2402)</td>
</tr>
<tr>
<td>$\beta_{0,2}$</td>
<td>3.5549</td>
<td>0.0356</td>
<td>(3.4850, 3.6248)</td>
</tr>
<tr>
<td>$\beta_{1,2}$</td>
<td>-0.1245</td>
<td>0.0101</td>
<td>(-0.1444, -0.1047)</td>
</tr>
<tr>
<td>$\beta_{0,3}$</td>
<td>0.8304</td>
<td>0.0212</td>
<td>(0.7888, 0.8719)</td>
</tr>
<tr>
<td>$\beta_{1,3}$</td>
<td>-0.0577</td>
<td>0.0062</td>
<td>(-0.0699, -0.0456)</td>
</tr>
<tr>
<td>$\gamma_v$</td>
<td>0.0462</td>
<td>0.0151</td>
<td>(0.0166, 0.0759)</td>
</tr>
<tr>
<td>$\gamma_{bil}$</td>
<td>0.8181</td>
<td>0.2046</td>
<td>(0.4171, 1.2191)</td>
</tr>
<tr>
<td>$\gamma_{alb}$</td>
<td>-1.7060</td>
<td>0.6181</td>
<td>(-2.9173, -0.4946)</td>
</tr>
<tr>
<td>$\gamma_{pro}$</td>
<td>-2.2085</td>
<td>1.6070</td>
<td>(-5.3582, 0.9412)</td>
</tr>
</tbody>
</table>
References


