Overview of Joint Longitudinal-Survival Models: Modeling the Association Between Dependent Outcomes

Isaac Quintanilla Salinas

University of California, Riverside

2021-06-07

Outline

- Overview of Longitudinal and Survival Data
- Joint Longitudinal-Survival Models
- Dynamic Predictions of Joint Models
- Applications in R

Longitudinal and Survival Data

Longitudinal Data

- Longitudinal data is the repeated collection of data from a participant at different time points
- The repeated measurements are correlated with each other
- Standard methods provide biased results due to the correlation
- Two methods can be used to address this correlation:
    - General Least Squares
    - Linear Mixed Effects Models

Survival Data

- Survival data is the length of time, since treatment or diagnosis, until a participant experiences an event, know as time-to-event
- Survival Models are used to model the association between the time-to-event and a set of predictors
    - Cox Proportional Hazard Model (Cox, 1972)

Joint Longitudinal-Survival Models

Overview

- Many biomedical studies collect repeated measurements from a participant until they experience an event
- Given the available data, researchers may be interested in the association between the repeated measurements and time-to-event
- De Gruttola and Tu (1994) and Tsiatis, DeGruttola, and Wulfsohn (1995) proposed different models to characterize the association
- Wulfsohn and Tsiatis (1997) proposed the shared-parameter model to jointly model the dependent outcomes

Shared-Parameter Model

- The shared-parameter model utilizes two separate submodels for each outcome
- Each submodel contain share a set of random effects
- The random effects induce the correlation between the outcomes (Tsiatis and Davidian, 2004)
- The shared-parameter has been extended for
  - Heirarchical Data (Liu, Ma, and O'Quigley, 2008)
  - Generalized Outcomes (Larsen, 2004; Li, Elashoff, and Li, 2009; Rizopoulos, Verbeke, Lesaffre, and Vanrenterghem, 2008)

Data

With n participants, each i^{th} participant has:

Longitudinal Data

- n_i repeated measurements
- $Y_i = (Y_{i1}, Y_{i2}, \cdots, Y_{in_i})^{\mathrm{T}}$
- $t_i = (t_{i1}, t_{i2}, \cdots, t_{in_i})^{\mathrm{T}}$
- $X_{ij} = (X_{ij1}, X_{ij2}, \cdots, X_{ijk})^{\mathrm{T}}$

Survival Data

- T_i: Observed time
- δ_i: Censoring status

Longitudinal Submodel

$$Y_{ij} = m_i(t_{ij}) + \epsilon_i(t_{ii}),$$

where

- $m_i(t_{ij}) = \boldsymbol{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^{\mathrm{T}}b_i$
- $\boldsymbol{X}_{ij}$: design matrix
- $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^{\mathrm{T}}$: regression coefficients
- $b_i \sim N_q(\mathbf{0}, \boldsymbol{G})$
- $\epsilon_i(t_{ij})$: error term at time t_{ij}
- $\boldsymbol{Z}_{ij}$: subset of $\boldsymbol{X}_{ij}$
- $b_i = (b_{i1}, \cdots, b_{iq})^{\mathrm{T}}$: random effects
- $\epsilon_i(t_{ij}) \sim N(0, \sigma^2)$

Survival Submodel - Hazard Function

$$\lambda_i\{t|M_i(t), X_i\} = \lim_{\Delta \to 0} \frac{P\{t \leq T_i < t + \Delta | T_i \geq t, M_i(t), \boldsymbol{X}_{i1}\}}{\Delta}$$

$$\lambda_i\{t|M_i(t), \boldsymbol{X}_{i1}\} = \lambda_0(t) \exp\{\boldsymbol{X}_{i1}^{\mathrm{T}} \boldsymbol{\gamma} + \alpha m_i(t)\}$$

where

- $\lambda_0(t)$: baseline hazard function
- $\boldsymbol{X}_{i1}$: design matrix at first time point
- $\boldsymbol{\gamma}$: regression coefficients
- α: association coefficient
- $M_i(t)$: history of the longitudinal outcome

Estimation

- Due to the random effects, the EM algorithm is used to obtain the maximum likelihood estimates (MLE)(Wulfsohn and Tsiatis, 1997)
- E-Step: Numerical techniques are used to approximate the random effects
  - Gaussian Quadrature (Wulfsohn and Tsiatis, 1997)
  - Laplace Approximation (Rizopoulos, Verbeke, and Lesaffre, 2009)
  - Monte Carlo Techniques (Henderson, Diggle, and Dobson, 2000)
- M-Step: A Newton-Raphson or other numerical techniques are used to maximize the likelihood function.

Joint Density Function

$$P(T_i, \delta_i, \boldsymbol{Y}_i, b_i; \boldsymbol{\theta}) = P(T_i, \delta_i | b_i; \boldsymbol{\theta}) P(\boldsymbol{Y}_i | b_i; \boldsymbol{\theta}) P(b_i; \boldsymbol{\theta})$$

where

$$\begin{aligned} P(\boldsymbol{Y}_i | b_i; \boldsymbol{\theta}) P(b_i; \boldsymbol{\theta}) = & \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2} \left\{ Y_{ij} - (\boldsymbol{X}_{ij}^{\mathrm{T}} \boldsymbol{\beta} + \boldsymbol{Z}_{ij}^{\mathrm{T}} b_i) \right\}^2 \right] \\ & \times (2\pi)^{-q/2} |G|^{-1/2} \exp\left(-\frac{1}{2} b_i^{\mathrm{T}} G^{-1} b_i \right) \end{aligned}$$

$$\begin{aligned} P(T_i, \delta_i | b_i; \boldsymbol{\theta}) &= \lambda_i(T_i)^{\delta_i} S_i(T_i) \\ &= \left[\lambda_0(T_i) \exp\{ \boldsymbol{X}_{i1}^{\mathrm{T}} \boldsymbol{\gamma} + \alpha m_i(t) \} \right]^{\delta_i} \\ & \quad \times \exp\left[- \int_0^{T_i} \lambda_0(s) \exp\{ \boldsymbol{X}_{i1}^{\mathrm{T}} \boldsymbol{\gamma} + \alpha m_i(t) \} ds \right] \end{aligned}$$

Likelihood Formulation

$$
\begin{aligned}
\ell(u, \boldsymbol{\theta}) &= \sum_{i=1}^{n} \ell_i(b_i, \boldsymbol{\theta}) = \sum_{i=1}^{n} \log P(T_i, \delta_i, \boldsymbol{Y}_i, b_i; \boldsymbol{\theta}) \\
\log P(T_i, \delta_i, \boldsymbol{Y}_i, b_i; \boldsymbol{\theta}) &= \delta_i \left[\log\{\lambda_0(T_i)\} + \boldsymbol{X}_{i1}^{\mathrm{T}} \boldsymbol{\gamma} + \alpha m_i(t) \right] \\
&\quad - \int_0^{T_i} \lambda_0(s) \exp\{\boldsymbol{X}_{i1}^{\mathrm{T}} \boldsymbol{\gamma} + \alpha m_i(t)\} ds \\
&\quad + \sum_{j=1}^{n_i} \left[-\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left\{ Y_{ij} - (\boldsymbol{X}_{ij}^{\mathrm{T}} \boldsymbol{\beta} + \boldsymbol{Z}_{ij}^{\mathrm{T}} b_i) \right\}^2 \right] \\
&\quad - \frac{q}{2} \log(2\pi\sigma^2) - \frac{1}{2} \log(|G|) - \frac{1}{2} b_i^{\mathrm{T}} G^{-1} b_i
\end{aligned}
$$

where

- $u = (b_1, \cdots, b_n)^{\mathrm{T}}$

Dynamic Predictions

Overview

- Recent developments in joint models is to expand the predicting capabilities of the models
  - Estimating the probability of that an event will occur in the next moment
  - Develop a mechanism to identify which individuals will experience an event
- Rizopoulos (2011) proposed a method to estimate the probability
- Andrinopoulou, Eilers, Takkenberg, and Rizopoulos (2018) improved the method using time-varying coefficients

Predicting the Survival Function

Rizopoulos (2011) proposed to estimate the probability via a Bayesian Approach with Markov Chain Monte Carlo (MCMC) methods.

- The survival probability for a new participants can be obtained from the survival function and the MLEs
- The standard error is minimized using the proposed Bayesian approach

Predicting the Outcome

To determine whether a new patient will experience an event, Rizopoulos (2011) proposed identifying the threshold using the sensitivity, specificity, and receiver operating characteristics (ROC) curve.

For a specified threshold:

- The sensitivity and specificity are estimated using a Bayesian approach
- The ROC curve and area under curve (AUC) are obtained to determine the efficacy of the threshold

Applications in R

JM in R

The `JM` package by Rizopoulos (2010) fits the joint longitudinal-survival models

- Fit a linear mixed effects model using the `lme` function from the `nlme` by Pinheiro, Bates, DebRoy, Sarkar, and R Core Team (2019)
- Fit a survival model using `coxph` function from the `survival` by Therneau (2015)
- Lastly, use the `jointModel` function from the `JM` package with the following 4 arguments:
  - *lmeObject*: the lme object
  - *survObject*: the survival object
  - *timeVar*: the time variable
  - *method*: the baseline hazard function

JM in R: Data

For R demonstration, we use the `aids` data from the `JM` package. The data set represents a randomized clinical trials to determine the efficacy of two antiretroviral drugs.

- 1408 observations
- 467 patients
- 9 variables

JM in R: Models

Linear Mixed Effects Model

```
me <- lme(CD4 ~ obstime + obstime:drug,
          random = ~ obstime | patient,
          data = aids)
```

Survival Model

```
ph <- coxph(Surv(Time, death) ~ drug,
            data = aids.id,
            x=TRUE)
```

- Note: the *x* argument must be set to TRUE

Joint Models

```
jm.1<-jointModel(me, ph, timeVar = "obstime",
                 method = "piecewise-PH-aGH")
```

JM: Interpretation Longitudinal

```
summary(jm.1)
```

Show 10 entries Search:

	Value	Std.Err	z-value	p-value
(Intercept)	7.2203	0.2218	32.5537	0
obstime	-0.1917	0.0217	-8.8374	0
obstime:drugddI	0.0116	0.0302	0.3834	0.7014

Showing 1 to 3 of 3 entries

Previous 1 Next

JM: Interpretation Survival

```
summary(jm.1)
```

Show 5 entries Search:

	Value	Std.Err	z-value	p-value
drugddI	0.3348	0.1565	2.1397	0.0324
Assoct	-0.2875	0.0359	-8.0141	0
log(xi.1)	-2.5438	0.1913	-13.2953	0
log(xi.2)	-2.2722	0.1784	-12.7328	0
log(xi.3)	-1.9554	0.2403	-8.1357	0

Showing 1 to 5 of 9 entries

Previous 1 2 Next

JMbayes Package

The `JMbayes` package by Rizopoulos (2016) estimates the Joint Models with MCMC techniques.

- Capable of conducting dynamic predictions

The `JMbayes2` package by Rizopoulos, Papageorgiou, and Miranda Afonso (2021) further expands the `JMbayes` package for different data structures.

Thank You

References

Andrinopoulou, E., P. H. C. Eilers, J. J. M. Takkenberg, et al. (2018). "Improved Dynamic Predictions from Joint Models of Longitudinal and Survival Data with Time-Varying Effects Using P-Splines". En. In: *Biometrics* 74.2, pp. 685-693. DOI: 10.1111/biom.12814.

Cox, D. R. (1972). "Regression Models and Life-Tables". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 34.2, pp. 187-220.

De Gruttola, V. and X. M. Tu (1994). "Modelling Progression of CD4-Lymphocyte Count and Its Relationship to Survival Time". En. In: *Biometrics* 50.4, p. 1003. DOI: 10.2307/2533439.

Henderson, R., P. Diggle, and A. Dobson (2000). "Joint Modelling of Longitudinal Measurements and Event Time Data". En. In: *Biostatistics* 1.4, pp. 465-480. DOI: 10.1093/biostatistics/1.4.465.

Larsen, K. (2004). "Joint Analysis of Time-to-Event and Multiple Binary Indicators of Latent Classes". In: *Biometrics* 60.1, pp. 85-92.

Acknowledgments

- Dr. Esra Kürüm
- Debaleena Sain