Empirical Transition Matrix of Multistate Models: The etm Package

Arthur Allignol[1,2,*] Martin Schumacher[2] Jan Beyersmann[1,2]

[1]Freiburg Center for Data Analysis and Modeling, University of Freiburg
[2]Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg
*arthur.allignol@fdm.uni-freiburg.de

DFG Forschergruppe FOR 534

Introduction

- Multistate models provide a relevant modelling framework for complex event history data
- MSM: Stochastic process that at any time occupies one of a set of discrete states
  - Health conditions
  - Disease stages

Introduction

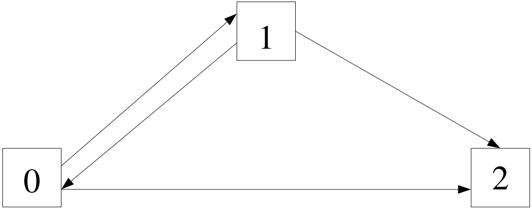
- Multistate models provide a relevant modelling framework for complex event history data
- MSM: Stochastic process that at any time occupies one of a set of discrete states
    - Health conditions
    - Disease stages
- Data consist of:
    - Transition times
    - Type of transition
- Possible right-censoring and/or left-truncation

Introduction

- Survival data

Introduction

- Illness-death model with recovery



Introduction

- Time-inhomogeneous Markov process $X_{t \in [0, +\infty)}$
  - Finite state space $\mathcal{S} = \{0, \ldots, K\}$
  - Right-continuous sample paths $X_{t+} = X_t$

Introduction

- Time-inhomogeneous Markov process $X_{t \in [0,+\infty)}$
  - Finite state space $\mathcal{S} = \{0, \dots, K\}$
  - Right-continuous sample paths $X_{t+} = X_t$
- Transition hazards

$$\alpha_{ij}(t)dt = P(X_{t+dt} = j \mid X_t = i), \; i,j \in \mathcal{S}, \; i \neq j$$

  - Completely describe the multistate process

Introduction

- Time-inhomogeneous Markov process $X_{t \in [0, +\infty)}$
  - Finite state space $\mathcal{S} = \{0, \dots, K\}$
  - Right-continuous sample paths $X_{t+} = X_t$
- Transition hazards

$$\alpha_{ij}(t)dt = P(X_{t+dt} = j \,|\, X_t = i), \; i, j \in \mathcal{S}, \; i \neq j$$

  - Completely describe the multistate process
- Cumulative transition hazards

$$A_{ij}(t) = \int_0^t \alpha_{ij}(u)du$$

Introduction

- Transition probabilities

$$P_{ij}(s,t) = P(X_t = j \mid X_s = i),\ i,j \in \mathcal{S},\ s \leq t$$

Introduction

- Transition probabilities

$$P_{ij}(s, t) = P(X_t = j \mid X_s = i),\ i, j \in \mathcal{S},\ s \leq t$$

- Matrix of transition probabilities

$$\mathbf{P}(s, t) = \prod_{(s,t]} (\mathbf{I} + d\mathbf{A}(u))$$

  - a $(K + 1) \times (K + 1)$ matrix

Introduction

- The covariance matrix is computed using the following recursion formula:

$$
\begin{aligned}
&\widehat{\mathrm{cov}}(\hat{\mathbf{P}}(s,t)) = \\
&\{(\mathbf{I} + \Delta\hat{\mathbf{A}}(t))^{\mathsf{T}} \otimes \mathbf{I}\}\widehat{\mathrm{cov}}(\hat{\mathbf{P}}(s,t-))\{(\mathbf{I} + \Delta\hat{\mathbf{A}}(t)) \otimes \mathbf{I}\} \\
&+ \{\mathbf{I} \otimes \hat{\mathbf{P}}(s,t-)\}\widehat{\mathrm{cov}}(\Delta\hat{\mathbf{A}}(t))\{\mathbf{I} \otimes \hat{\mathbf{P}}(s,t-)^{\mathsf{T}}\}
\end{aligned}
$$

Introduction

- The covariance matrix is computed using the following recursion formula:

$$\widehat{\text{cov}}(\hat{\mathbf{P}}(s,t)) =$$
$$\{(\mathbf{I} + \Delta\hat{\mathbf{A}}(t))^{\mathsf{T}} \otimes \mathbf{I}\}\widehat{\text{cov}}(\hat{\mathbf{P}}(s,t-))\{(\mathbf{I} + \Delta\hat{\mathbf{A}}(t)) \otimes \mathbf{I}\}$$
$$+ \{\mathbf{I} \otimes \hat{\mathbf{P}}(s,t-)\}\widehat{\text{cov}}(\Delta\hat{\mathbf{A}}(t))\{\mathbf{I} \otimes \hat{\mathbf{P}}(s,t-)^{\mathsf{T}}\}$$

  - Estimator of the Greenwood type
    - Enables integrated cumulative hazards of not being necessarily continuous
    - Reduces to usual Greenwood estimator in the univariate setting

Introduction

- The covariance matrix is computed using the following recursion formula:

$$\begin{aligned}&\widehat{\text{cov}}(\hat{\mathbf{P}}(s,t)) = \\ &\{(\mathbf{I} + \Delta\hat{\mathbf{A}}(t))^{\mathsf{T}} \otimes \mathbf{I}\}\widehat{\text{cov}}(\hat{\mathbf{P}}(s,t-))\{(\mathbf{I} + \Delta\hat{\mathbf{A}}(t)) \otimes \mathbf{I}\} \\ &+ \{\mathbf{I} \otimes \hat{\mathbf{P}}(s,t-)\}\widehat{\text{cov}}(\Delta\hat{\mathbf{A}}(t))\{\mathbf{I} \otimes \hat{\mathbf{P}}(s,t-)^{\mathsf{T}}\}\end{aligned}$$

  - Estimator of the Greenwood type
    - Enables integrated cumulative hazards of not being necessarily continuous
    - Reduces to usual Greenwood estimator in the univariate setting
  - Found to be the preferred estimator in simulation studies for survival and competing risks data

In R

- **survival** and **cmprsk** estimate survival and cumulative incidence functions, respectively
  - Outputs can be used to compute transition probabilities in more complex models when transition probabilities take an explicit form
  - **cmprsk** does not handle left-truncation
  - Variance computation "by hand"

In R

- **survival** and **cmprsk** estimate survival and cumulative incidence functions, respectively
  - Outputs can be used to compute transition probabilities in more complex models when transition probabilities take an explicit form
  - **cmprsk** does not handle left-truncation
  - Variance computation “by hand”
- **mvna** estimates cumulative transition hazards
- **changeLOS** computes transition probabilities
  - Lacks variance estimation
  - Cannot handle left-truncation

In R

- **survival** and **cmprsk** estimate survival and cumulative incidence functions, respectively
  - Outputs can be used to compute transition probabilities in more complex models when transition probabilities take an explicit form
  - **cmprsk** does not handle left-truncation
  - Variance computation "by hand"
- **mvna** estimates cumulative transition hazards
- **changeLOS** computes transition probabilities
  - Lacks variance estimation
  - Cannot handle left-truncation

$\Longrightarrow$ **etm**

Package Description

- The main function `etm`

```
etm(data, state.names, tra, cens.name, s, t = "last",
    covariance = TRUE, delta.na = TRUE)
```

Package Description

- The main function `etm`

```
etm(data, state.names, tra, cens.name, s, t = "last",
    covariance = TRUE, delta.na = TRUE)
```

- 4 methods
  - `print`
  - `summary`
  - `plot`
  - `xyplot`
- 2 data sets

Illustration: DLI Data

- 614 patients who received allogeneic stem cell transplantation for chronic myeloid leukaemia between 1981 and 2002
  - All patients achieved complete remission

Illustration: DLI Data

- 614 patients who received allogeneic stem cell transplantation for chronic myeloid leukaemia between 1981 and 2002
  - All patients achieved complete remission
- Patients in first relapse were offered a donor lymphocyte infusion (DLI)
  - Infusion of lymphocytes harvested from the original stem cell donor
  - DLI produces durable remissions in a substantial number of patients

Illustration: DLI Data

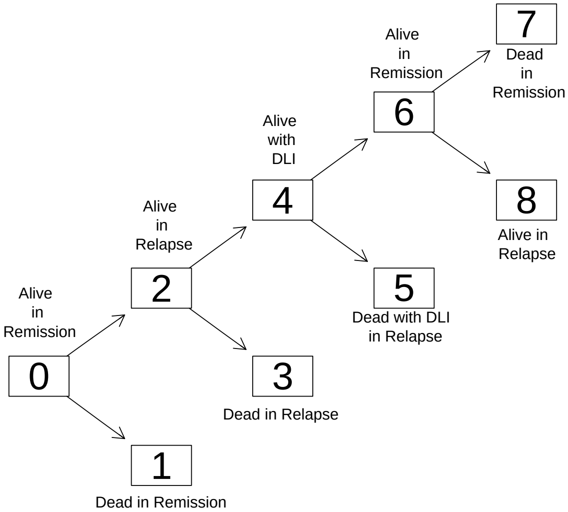


Illustration: DLI Data

- Current leukaemia free survival (CLFS): Probability that a patient is alive and leukaemia-free at a given point in time after the transplant
  - Probability of being in state 0 or 6 at time t

Illustration: DLI Data

- Current leukaemia free survival (CLFS): Probability that a patient is alive and leukaemia-free at a given point in time after the transplant
  - Probability of being in state 0 or 6 at time t

$$\widehat{\text{CLFS}}(t) = \hat{P}_{00}(0, t) + \hat{P}_{06}(0, t)$$

Illustration: DLI Data

- Current leukaemia free survival (CLFS): Probability that a patient is alive and leukaemia-free at a given point in time after the transplant
  - Probability of being in state 0 or 6 at time t

$$\widehat{\text{CLFS}}(t) = \hat{P}_{00}(0, t) + \hat{P}_{06}(0, t)$$

$$\hat{P}_{06}(s, t) = \sum_{s<u\leq v\leq r\leq t} \hat{P}(s, u-)\frac{dN_{02}(u)}{Y_0(u)}\hat{P}_{22}(u, v-) \times \frac{dN_{24}(v)}{Y_2(v)}$$
$$\times \hat{P}_{44}(v, r-)\frac{dN_{46}(r)}{Y_4(r)}\hat{P}_{66}(r, t)$$

Illustration: DLI Data

- Current leukaemia free survival (CLFS): Probability that a patient is alive and leukaemia-free at a given point in time after the transplant
  - Probability of being in state 0 or 6 at time t

$$\widehat{\mathrm{CLFS}}(t) = \hat{P}_{00}(0, t) + \hat{P}_{06}(0, t)$$

$$\widehat{\mathrm{var}}(\widehat{\mathrm{CLFS}}(t)) =$$

$$\widehat{\mathrm{var}}(\hat{P}_{00}(0, t)) + \widehat{\mathrm{var}}(\hat{P}_{06}(0, t)) + 2\widehat{\mathrm{cov}}(\hat{P}_{00}(0, t), \hat{P}_{06}(0, t))$$

Illustration: DLI Data

```
> tra <- matrix(FALSE, 9, 9)
> tra[1, 2:3] <- TRUE
> tra[3, 4:5] <- TRUE
> tra[5, 6:7] <- TRUE
> tra[7, 8:9] <- TRUE
> tra
      0     1     2     3     4     5     6     7     8
0 FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
1 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
2 FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE
3 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
4 FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE
5 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
6 FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE
7 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
8 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

> dli.etm <- etm(dli.data, as.character(0:8), tra, "cens", s = 0)
```

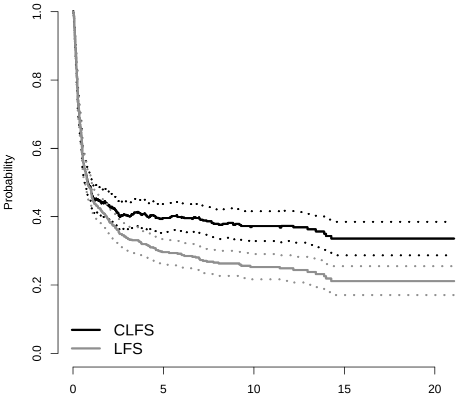
Illustration: DLI Data

```
> clfs <- dli.etm$est["0", "0", ] + dli.etm$est["0", "6", ]
> var.clfs <- dli.etm$cov["0 0", "0 0", ] +
+     dli.etm$cov["0 6", "0 6", ] + 2 * dli.etm$cov["0 0", "0 6", ]
```

Illustration: DLI Data

```
> clfs <- dli.etm$est["0", "0", ] + dli.etm$est["0", "6", ]

> var.clfs <- dli.etm$cov["0 0", "0 0", ] +
+     dli.etm$cov["0 6", "0 6", ] + 2 * dli.etm$cov["0 0", "0 6", ]
```



Summary

- **etm** provides a way to easily estimate and display the matrix of transition probabilities from multistate models
- Permits to compute interesting quantities that depend on the matrix of transition probabilities
- Empirical transition matrix valid under the Markov assumption
  - Stage occupation probability estimates still valid for more general models

Bibliography

Allignol, A., Schumacher, J. and Beyersmann, J. (2009). A note on variance estimation of the Aalen-Johansen Estimator of the cumulative incidence function in competing risks, with a view towards left-truncated data.
Under revision.

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993).
Statistical Models Based on Counting Processes.
Springer-Verlag, New-York.

Datta, S. and Satten, G. A. (2001).
Validity of the Aalen-Johansen Estimators of Stage Occupation Probabilities and Nelson-Aalen Estimators of Integrated Transition Hazards for non-Markov Models.
Statistics & Probability Letters, 55:403–411.

Klein, J. P., Szydlo, R. M., Craddock, C. and Goldman, J. M. (2000)
Estimation of Current Leukaemia-Free Survival Following Donor Lymphocyte Infusion Therapy for Patients with Leukaemia who Relapse after Allografting: Application of a Multistate Model
Statistics in Medicine, 19:3005–3016.

- Thanks to Mei-Jie Zhang (Medical College of Wisconsin) for providing us with the DLI data

Empirical Transition Matrix

- $\mathbf{A}(t)$ the matrix of cumulative transition hazards

Empirical Transition Matrix

- $\mathbf{A}(t)$ the matrix of cumulative transition hazards
  - Non-diagonal entries estimated by the Nelson-Aalen estimator

$$\hat{A}_{ij}(t) = \sum_{t_k \leq t} \frac{\Delta N_{ij}(t_k)}{Y_i(t_k)}, \; i \neq j$$

Empirical Transition Matrix

- $\mathbf{A}(t)$ the matrix of cumulative transition hazards
  - Non-diagonal entries estimated by the Nelson-Aalen estimator

$$\hat{A}_{ij}(t) = \sum_{t_k \leq t} \frac{\Delta N_{ij}(t_k)}{Y_i(t_k)},\ i \neq j$$

  - Diagonal entries

$$\hat{A}_{ii}(t) = -\sum_{j \neq i} \hat{A}_{ij}(t)$$

Empirical Transition Matrix

- Empirical transition matrix

$$\hat{\mathbf{P}}(s, t) = \prod_{(s,t]} (\mathbf{I} + d\hat{\mathbf{A}}(u))$$

Empirical Transition Matrix

- Empirical transition matrix

$$\hat{\mathbf{P}}(s,t) = \prod_{(s,t]} (\mathbf{I} + d\hat{\mathbf{A}}(u))$$

- $\hat{\mathbf{A}}(t)$ is a matrix of step-functions with a finite number of jumps on $(s, t]$

$$\hat{\mathbf{P}}(s,t) = \prod_{s<t_k\leq t} \left(\mathbf{I} + \Delta\hat{\mathbf{A}}(t_k)\right)$$

  - $\Delta\hat{\mathbf{A}}(t) = \hat{A}(t) - \hat{A}(t-)$

DLI Example

Current Leukaemia Free Survival