Multistate analysis of life histories with R

Frans Willekens
Pre-conference workshop
Ottawa, 18 March 2015
The overarching question

“How do events, experiences, conditions, institutions and policies affect the lives of people?”
• Questions multistate models are designed to answer
• Life course data: what are the issues?
• Sequence analysis: a very brief introduction
  – Studies *observed* sequence of states
  – Disregard censoring
• Multistate life history data analysis
  – Studies transitions between states and state occupation times
  – Much attention to censoring and truncation: later entry and early departure
What is the effect of early-life experiences on disability status at old age?

Is increase in longevity associated with more years with disability?

What is the effect of smoking and obesity on life expectancy and years with disability?

Does retirement increase the incidence of cognitive impairment?

What is an optimal lifestyle for a long / healthy life?
Research questions

• What is the probability of being employed at 65?
• What is the probability of a divorce?
• What is the rate of divorce?
• What is the expected waiting time to a tenure?
• What is the expected sojourn time in cohabitation?
Research questions

• What are the most frequent life histories?
• Do life histories differ between cohorts or SES?
• What is the effect of a personal attribute (covariate) on the rate of transition?
• How do transition rates differ between birth cohorts and vary with socio-economic status (SES), gender, level of education, place of residence, etc.
• What is the effect of an additional year of schooling on a transition (rate and age)?
• What is the effect of an additional year of schooling on the subsequent life path?
• What is the effect of age at leaving parental home on age at first birth?
• What is the effect of sequence of living arrangements on age at first birth?
• What is the effect of marriage on employment career?
• What is the effect of postponement of marriage on age at first birth?
• What is the effect of duration of unemployment on likelihood of finding a job?
Two observations on challenges in life course research
“New directions in life course research”
Karl Ulrich Mayer
Annual Review Sociology, 2009

• “There will always be more need for modeling transitions between states.” (p. 425)
• Concept of risk: “Exposure to risk, measured by its incidence and duration, can be a powerful concept in mapping and measuring life courses” (p. 424)
“New Life for Old Ideas: The ‘Second Wave’ of Sequence Analysis Bringing the ‘Course’ Back Into the Life Course”
Aisenbrey and Fasang
Sociological Methods and Research, 2010

“To date, the majority of quantitative life course studies are based on event history applications that model the probability of single transitions and durations under certain conditions.”
Data on subjects **over time** (focus on processes)

- Quantitative: panel, follow-up study, cohort study
  
  => *life histories*

- Qualitative: autobiographies, oral histories
  
  => *life stories*
Lexis diagram

Events
- Events of interest
- Unrelated events (censoring: lost to observation)
- Competing events (competing risks)

Exposures
- Episode
- Exposure
- Exposure ends in event of interest or censoring

Sequences
- Event sequence (event history)
- State sequence
Single event
(single decrement)

\[ \mu_{01}(t) \]

H \quad \text{exit}

\[ \mu_{02}(t) \]

\[ \mu_{03}(t) \]

H \quad A \quad M \quad C

“Survival”

\[ 0 = \text{origin state (H)} \]

- Event of interest and unrelated events
- Exposure

Total exit rate: \[ \mu_{0+}(t) = \mu_{01}(t) + \mu_{02}(t) + \mu_{03}(t) \]
Competing risks

0 = origin state (H)

- Event of interest and unrelated events
- Exposure

Total exit rate: \[ \mu_{0+}(t) = \mu_{01}(t) + \mu_{02}(t) + \mu_{03}(t) \]
Multistate: event sequence or state sequence

Competing risks

H → A → C → M

C, M → M → A
1. Living at parental home (H)
2. Living alone (A)
3. Married (M)
4. Cohabiting (C)

Multistate modeling as extension of competing risks
Multistate life history analysis

• A life domain is characterized by a finite set of possible states (*multistate; multiple states*)
  – Educational status
  – Health status
  – Employment status
  – Living arrangement
  – Marital status

• State space: a finite set of possible states
  – Alphabet (TraMineR): (short) state labels
  – State space depends on research question

• Any time, a person may leave a state and enter one of the other states
  – Multiple possible destinations: *competing risks*
The life course is a sequence of states and transitions between states.

State sequence: ordered list of states (on a time axis).
Event sequence: ordered list of events (on time axis).

At a given time, an individual occupies one state.
- Grade of membership method; mixture models.

State occupied is observable (not latent or hidden).
- Hidden multistate models (e.g. hidden Markov model).
• Sequence representation: use state space or alphabet

• Living arrangement
  – State space:
    • H Living with parents
    • A living alone
    • M married
    • C cohabiting
  – Sequence: $A = \{H, A, C, A, M, A\}$

• Employment history
  – State space (letters in alphabet)
    • N No Job
    • J Job
### Sequence data: data types

<table>
<thead>
<tr>
<th>State sequence</th>
<th>Event sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ID</strong></td>
<td><strong>time</strong></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
### State sequence

<table>
<thead>
<tr>
<th>Year</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marital status</td>
<td>Single</td>
<td>Single</td>
<td>Married</td>
<td>Divorced</td>
</tr>
<tr>
<td>Location</td>
<td>Parents</td>
<td>Alone</td>
<td>Cohabitation</td>
<td>Alone</td>
</tr>
<tr>
<td>Child</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>1</td>
</tr>
</tbody>
</table>

### Episodes

<table>
<thead>
<tr>
<th>ID</th>
<th>Start</th>
<th>End</th>
<th>Marital Status</th>
<th>Location</th>
<th>Child</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1321</td>
<td>1333</td>
<td>Single</td>
<td>Parents</td>
<td>No</td>
</tr>
<tr>
<td>1</td>
<td>1333</td>
<td>1344</td>
<td>Single</td>
<td>Alone</td>
<td>No</td>
</tr>
<tr>
<td>1</td>
<td>1344</td>
<td>1356</td>
<td>Married</td>
<td>Cohabitation</td>
<td>No</td>
</tr>
<tr>
<td>1</td>
<td>1356</td>
<td>1380</td>
<td>Divorced</td>
<td>Alone</td>
<td>No</td>
</tr>
<tr>
<td>1</td>
<td>1380</td>
<td>1384</td>
<td>Divorced</td>
<td>Alone</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Data 1. German Life History Survey (GLHS) (Blossfeld & Rohwer, 2002)

- 201 respondents; three birth cohorts: 1929-31, 1939-41, and 1949-51
- 600 job episodes; 382 out-of-job episodes
Employment history

Time representation
Calendar time
Age
Duration in state
Figure 5.8 Lexis diagram: employment careers of selected GLHS respondents. Display B, using ggplot2 package.
Data 2. Netherlands Family and Fertility Survey (NLOG) 1998

- 5450 respondents
Pathways to first child
Netherland Family and Fertility Survey 1998

- 19425 respondents

Non-disabled \(\rightarrow\) Disabled

Death

Illness-death model with recovery
(extension of hazard model with time varying covariate)
Disability history
US HRSw 1992-2004
Models need to consider data limitations
early exit, late entry

=> Counting process perspective: track transitions and exposures
Sequence analysis

- Sequence (trajectory) is unit of analysis
- Study trajectories (fixed period)
- Compare trajectories
- 1986: Andrew Abbott and John Forrest: “Optimal matching methods for historical sequences”
- 2007: Chesa (Elzinga, Free University Amsterdam)
- 2008: TraMineR (Richards, University of Geneva)
- Research questions
  - Most frequent sequences in stage of life
  - How often do transitions occur in a cohort
  - What are general types of life stage trajectories?
  - How and why do trajectories differ between individuals?

Short course of social sequence analysis: http://www.youtube.com/watch?v=9WJPook9Qsc

library (TraMineR)


ddd <- seqdef(dd)

seqplot(ddd, type="i")
library (Biograph)
data (GLHS)
parameters <- Parameters (GLHS)
occup <- Occup (GLHS)
DTraMineR <- seqconc (occup
$st_age_1,sep="-")namst <- c(parameters
$namstates,"+")
D.seq <- seqdef
(DTraMineR,labels=namst,alphabet=c("N","J","+")
))alphabet (D.seq)
n <- 10
seqfplot (D.seq, tlim=1:n, title="Sequence frequency
plot", xlab=c(0:54), ltext=c("N","J","Censored"), las=1, ylab=paste(n," most frequent sequences (%)", sep=""))
Sequence index plot, 201 respondents
State distribution plot, 201 respondents
Sequence analysis: compare trajectories

• Pairwise comparison of sequences
  – Reference sequence
  – How many sequences need to be inserted or deleted to make two sequences equal
  – Dissimilarity measures
  – Optimal matching (OM)
  – Problem: censoring
The life course is a sequence of states and transitions between states.

Does not focus on description and comparison of trajectories (patterns) but on processes generating the trajectories.

Model transitions between states in a state space.

The process governing transitions is a multistate process in continuous time (e.g. Markov, semi-Markov, non-Markov multistate model).
Life course dynamics (trajectory) is governed by rates of transition that vary with age and depend on history and context.

Rate = events / exposure (duration)

- Hazard rate
- Occurrence-exposure rate (occurrences / exposure)
- Incidence rate
- Transition rate

Ulrich Mayer (2009): “Exposure to risk, measured by its incidence and duration, can be a powerful concept in mapping and measuring life courses” (p. 424)
Events: A counting process counts event occurrences and people at risk (exposed).
- At each occurrence of a transition
- During an interval (observation window): event count and exposure time

Exposure: Numbers of individuals at risk
- Just before event time -> Nelson-Aalen estimator
- During the observation period: number at risk AND duration at risk -> occurrence-exposure rate
Thank you