Simulation of life histories with MicSim

Sabine Zinn (DIW / SOEP)

Women in Statistics: Microsimulation Workshop

Statistische Woche, 9. September 2019

Material available under

https://github.com/bieneSchwarze/MicSimCourse

MicSim Package:

https://cran.r-project.org/web/packages/MicSim/index.html

Outline

- Continuous-Time Microsimulation
- Illustration 1: Transition to First Child
- Illustration 2: Changing Marital Status, Educational Attainment, Fertility Status
- Some Additional Advices
- Extensions

Outline

- Continuous-Time Microsimulation
- Illustration 1: Transition to First Child
- Illustration 2: Changing Marital Status, Educational Attainment, Fertility Status
- Some Additional Advices
- Extensions

Continuous-Time Microsimulation: Basics

Key idea: Individuals `live' their lives according to some (stochastic model)

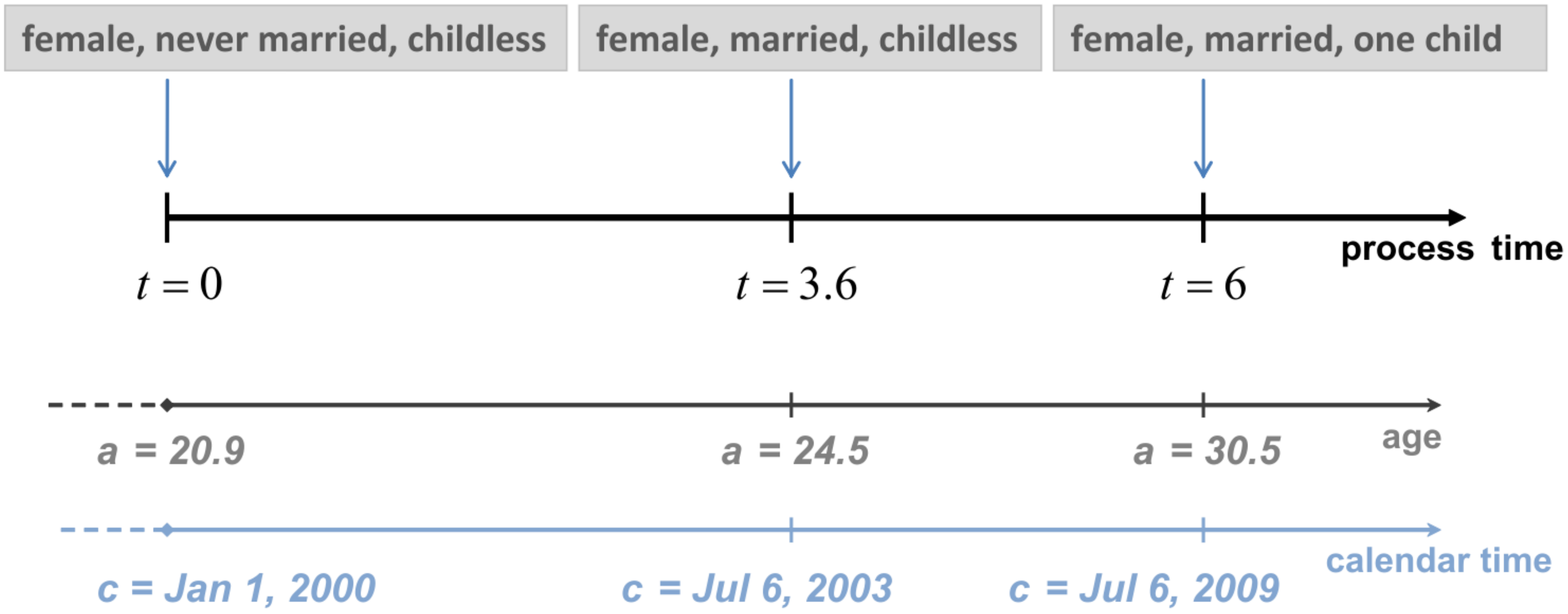
- Continuous-time multi-state model: standard approach to describe individual life-courses
- Multi-state model: stochastic process that at any point in time occupies one out of a set of discrete states
- Each life-course: trajectory of a stochastic process
- Discrete states: summarize relevant categories an individual can belong to (→ state space)
- State space: set of all possible states
- Virtual population: all individuals that are considered during simulation
  (=longitudinal sample from a synthetically constructed population that resembles a real one)

Output: Any conceivable implication of the processes that drive individual lives/population development

Simulation Model

Non-homogeneous continuous-time Markov chain:
life-course considered along age and calendar time

Non-homogeneous semi-Markov processes:
additionally along `time that has elapsed since the last event'



Data Requirements

Base population

- Number of individuals in the different states at starting time of simulation, according to birth date

Transition rates

- only for feasible transitions
- depend on: individual age, calendar time, current state and next state, & optional also on time elapsed
- have to be provided for complete calendar time and age range considered in simulation

Optional

- immigrants, according to birth date, immigration date, & state at immigration
- for newborns: states-at-birth & occurrence probabilities

Simulation

Transition rates → distribution function of waiting times in distinct states of state space

$$F\left(w_{s_j,s_k}, c, a\right) = 1 \; - \exp\left\{-\Lambda_{s_j,s_k}\left(w_{s_j,s_k}, c, a\right)\right\},$$

where

$$\Lambda_{s_j,s_k}\left(w_{s_j,s_k}, c, a\right) = \int_0^{w_{s_j,s_k}} \lambda_{s_j,s_k}(c+v, a+v)dv \qquad \text{(Markov)}$$

$$\Lambda_{s_j,s_k}\left(w_{s_j,s_k}, c, a\right) = \int_0^{w_{s_j,s_k}} \lambda_{s_j,s_k}(v, c+v, a+v)dv \qquad \text{(semi-Markov)}$$

Simulation (cont‘d)

Construction of life-courses: generation of sequences of random waiting times

- Waiting time variates to all possible destination states s_k:

$$w_{s_j,s_k} = \Lambda^{-1}_{s_j,s_k}\left(w_{s_j,s_k}, c, a\right)[-\ln(1-u)], u \sim U_{[0,1]}$$

- Of all generated random waiting times: realize the shortest one
- Computation of ‘shortest’ waiting times repeated for each individual

Event Queue: Along Calendar time …

Individual: 35 Current state: *fem, married, one child* Transition time: 12.6.2019 Next state: *fem, divorced, one child*	Individual: 147 Current state: *male, single, childless* Transition time: 14.6.2019 Next state: *dead*	Individual: 965 Current state: *fem, single, one child* Transition time: 2.1.2020 Next state: *fem, single, two children*	…

1. De-queue of *head* of event-queue
2. Determine next transition time and next new state
3. If the new state is an absorbing state do not enqueue again; otherwise enqueue at the right position in the event queue
4. Pick new head of event-queue and continue until
   - ending time is reached
   - event-queue is empty (all individuals died/ emigrated)

Outline

- Continuous-Time Microsimulation
- Illustration 1: Transition to First Child
- Illustration 2: Changing Marital Status, Educational Attainment, Fertility Status
- Some Additional Advices
- Extensions

Illustration 1: Set up

Study topic

- Pathways of women to transition to first child with respect to living arrangement

Model of individual behavior

- Non-homogeneous continuous-time Markov chains

Simulation horizon

- 01.01.2010 – 31.12.2030

Data

- Empirical rates: SHARELife (retrospective survey carried out in 2010, part of the 3rd wave of the SHARE project)
- Synthetic cohorts: females aged from 20 to 25, in each age group N = 100

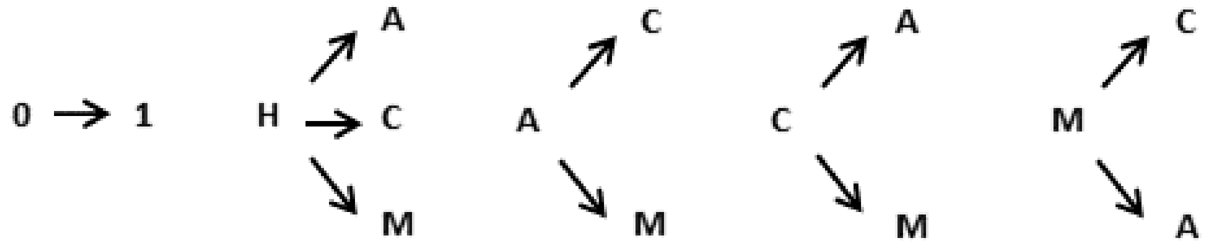
State Space & Transitions

State variables:

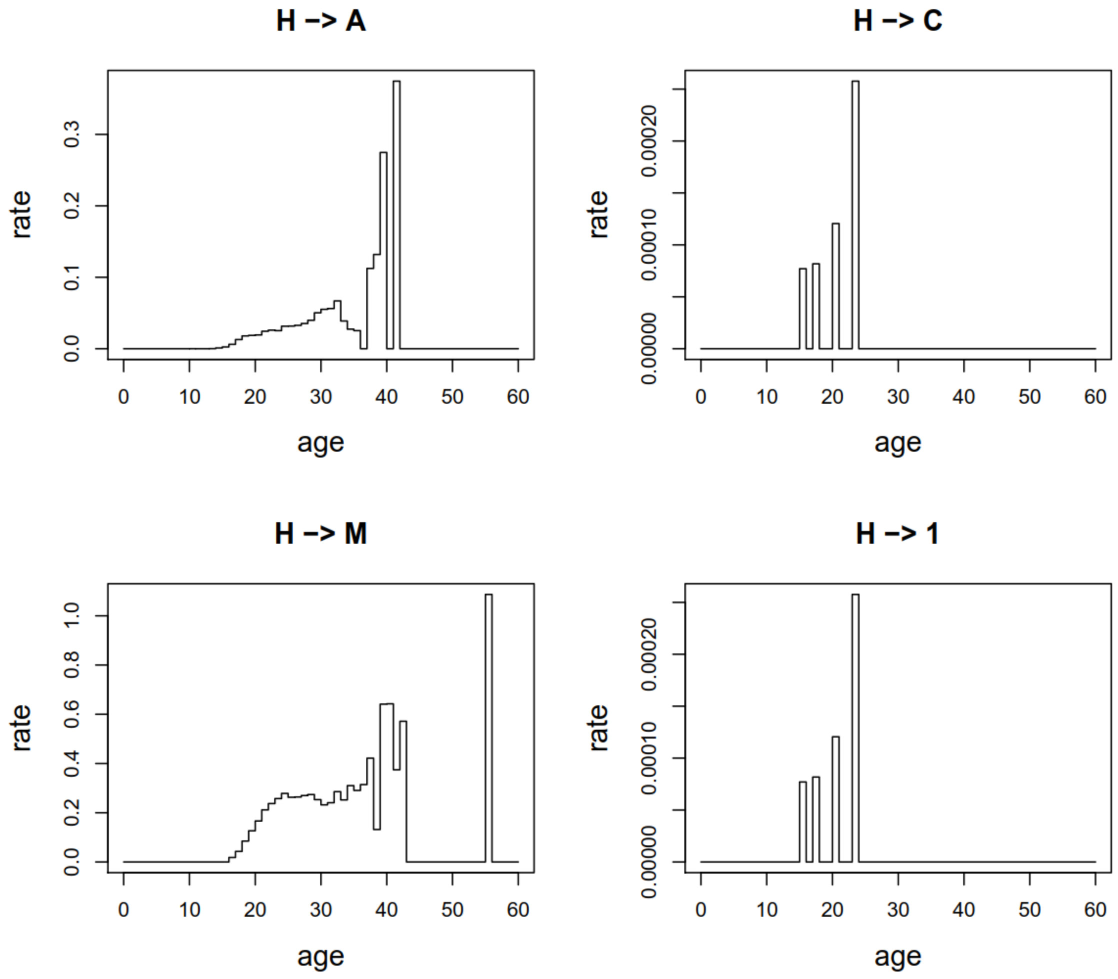
- *Gender*: female (f), male (m)
- *Living Arrangement*: living at parental home (H), living alone (A), living in cohabitation (C), being married (M)
- *Fertility*: childless (0), being parent (1)

State space: $\{f,m\} \times \{H,A,C,M\} \times \{0,1\}$

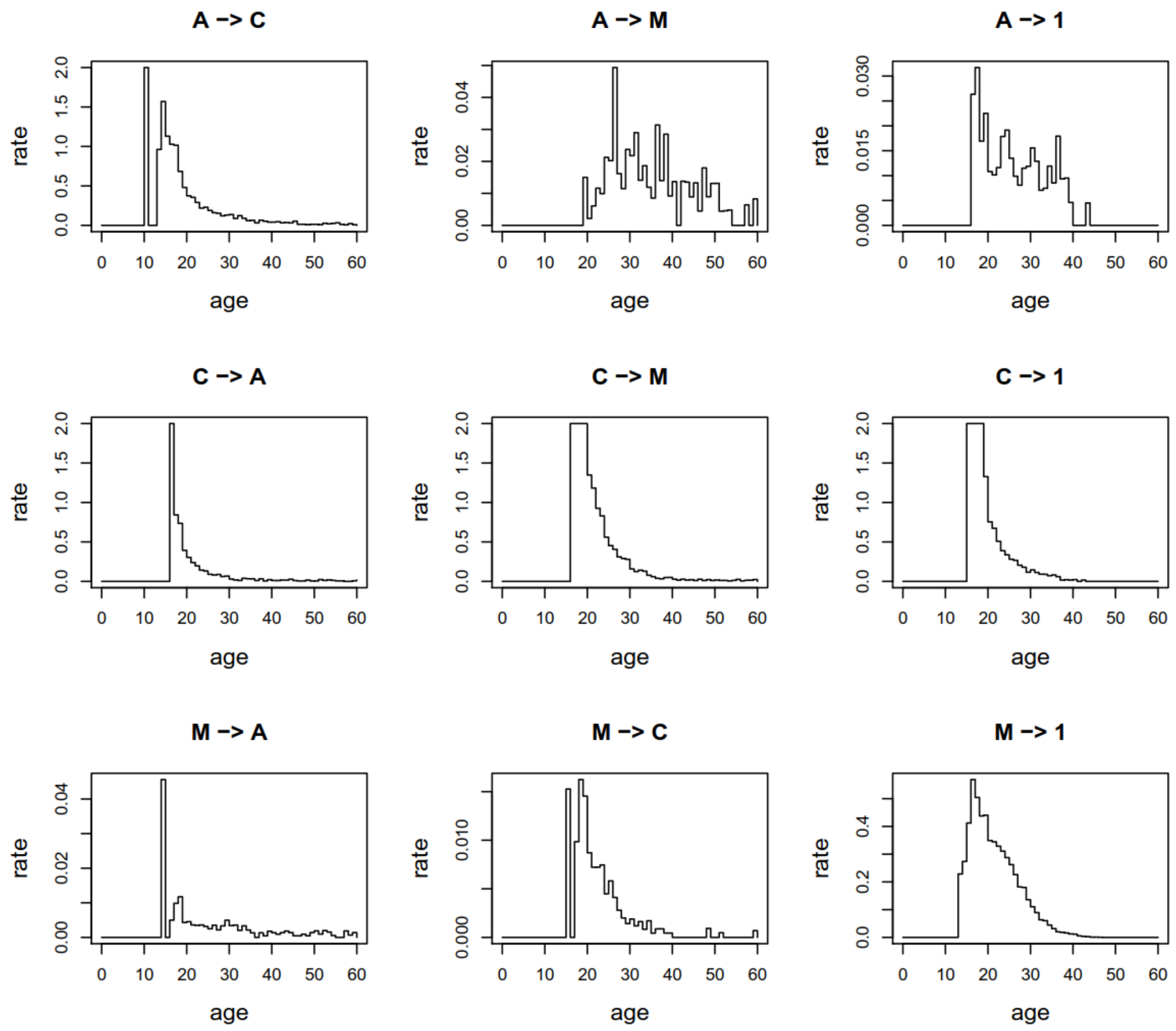
Possible transitions:



Transition Rates



Transition Rates (cont‘d)



Using MicSim …

Outline

- Continuous-Time Microsimulation
- Illustration 1: Transition to First Child
- Illustration 2: Changing Marital Status, Educational Attainment, Fertility Status
- Some Additional Advices
- Extensions

Illustration 2: Set up

Study topic

- Changes in marital status, educational attainment, fertility status over the life-course

Model of individual behavior

- Non-homogeneous continuous-time Markov chains & semi-Markov processes

Simulation horizon

- 01.01.2014 – 31.12.2024

Data

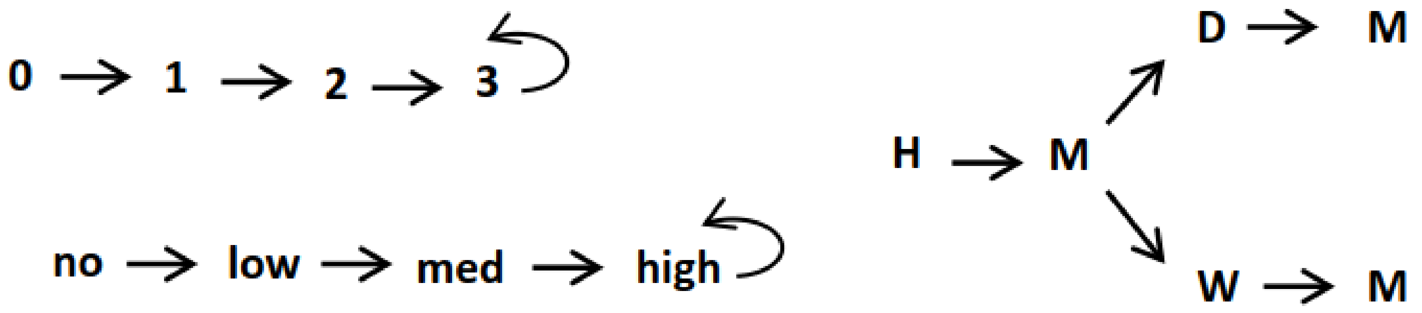
- Hypothetical rates (e.g., Hadwiger model, Gompertz distribution)
- Synthetic base population (N=1000)
- Migrants (M=200)

State Space & Transitions

State variables:

- *Gender*: female (f), male (m)
- *Marital Status*: never married (NM), married (M), divorced (D), widowed (W)
- *Fertility:* 0, 1, 2, 3+
- *Educational Attainment:* no, primary (low), lower secondary (med), upper secondary & tertiary (high)

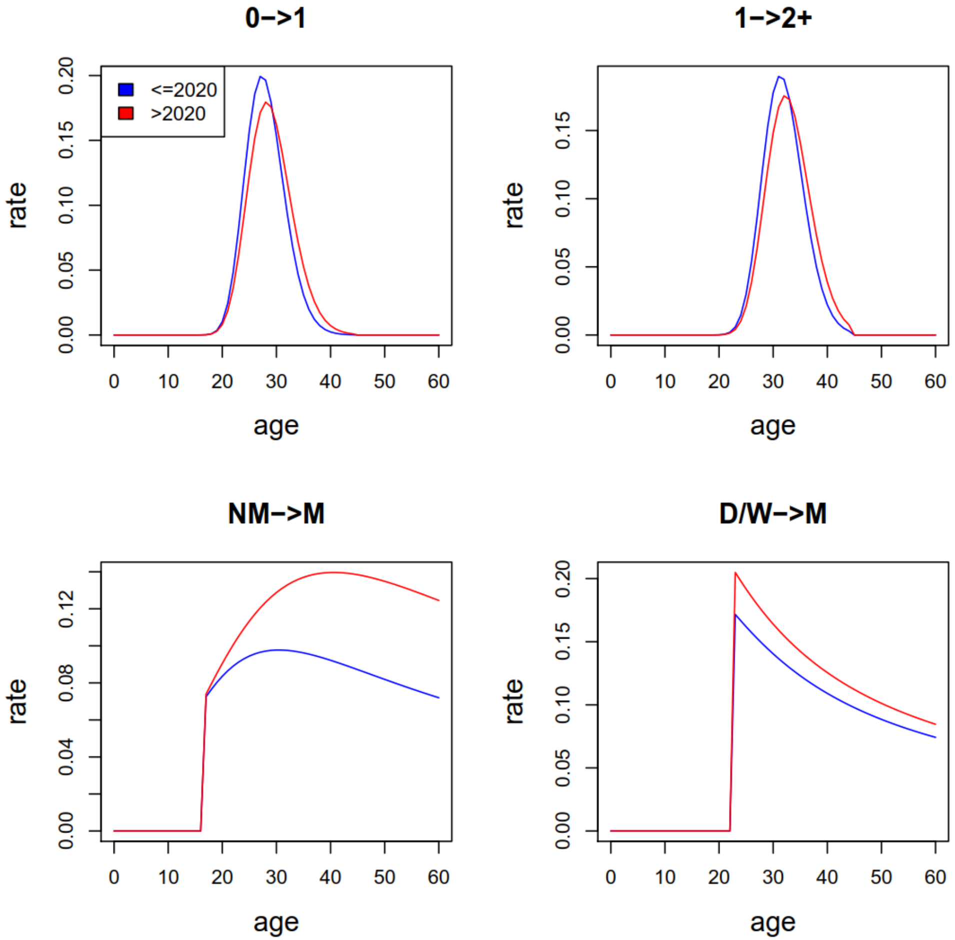
Possible transitions:



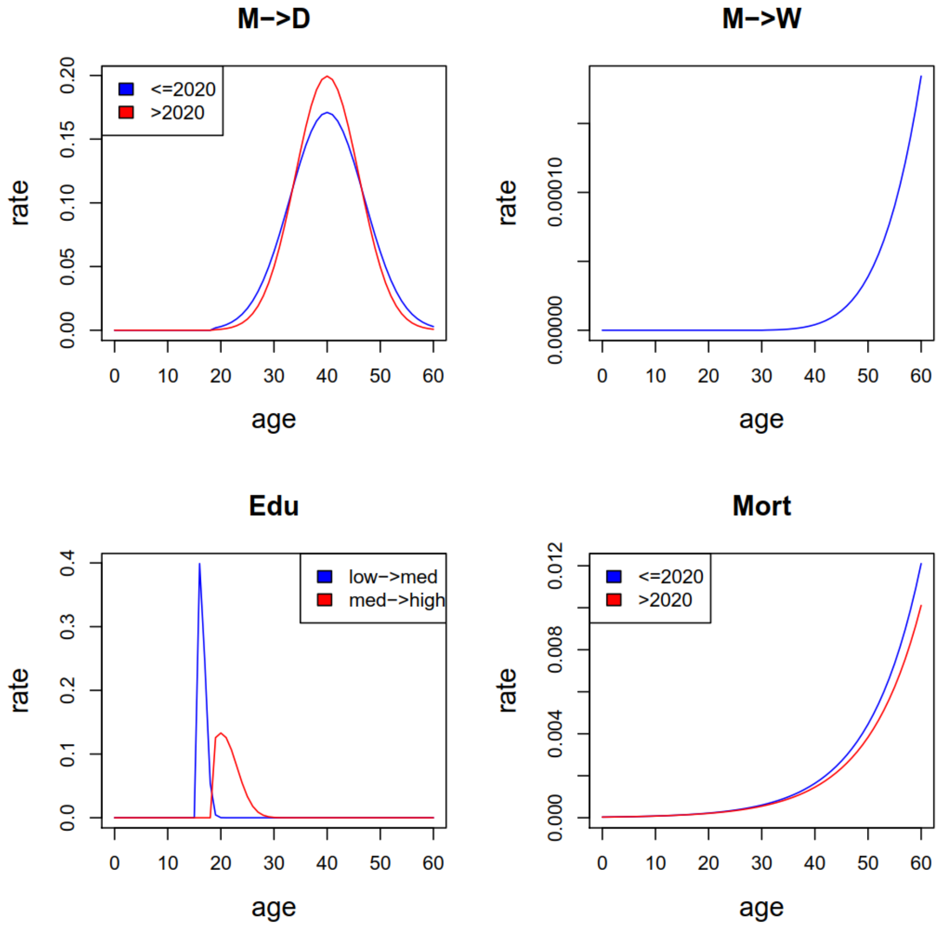
Parametric Models For Transition Rates

- Fertility rates: Hadwiger mixture model
- (Re)Marriage rates: log-logistic model
- Divorce rates: normal density
- Widowhood rates: gamma cdf
- Change in educational attainment: deterministic age & time or normal density
- Mortality rates: Gompertz model

Transition Rates



Transition Rates (cont‘d)



Using MicSim …

Outline

- Continuous-Time Microsimulation
- Illustration 1: Transition to First Child
- Illustration 2: Changing Marital Status, Educational Attainment, Fertility Status
- Some Additional Advices
- Extensions

Pseudo Random Number Generator (PRNG) & Seed

PRNG

- Starting from a pre-specified seed same sequence of pseudo random
- The period is the maximum length PRNG begins to repeat numbers
- MicSim uses R default: Mersenne Twister (based on a matrix linear recurrence over a finite binary field) → one of the best PRNG currently available

Seed

- The repeatability of simulation results is inevitable for their reliability
- Identical seeds result in identical simulation results
- Others should be able to redo what you have done
- However: Interpretation of results should never depend on seed
- Therefore: Run more than one experiment, with different seeds (Zero is never a good choice!)

Parallel Computing & Fast Version of MicSim

Parallel Computing

… look at an example

(Code available in help function of the MicSim package.)

Fast Version of MicSim (requires installing JVM 64bit)

… look at an example

(Code & documentation available under https://github.com/bieneSchwarze/fastMicSim)

Outline

- Continuous-Time Microsimulation
- Illustration 1: Transition to First Child
- Illustration 2: Changing Marital Status, Educational Attainment, Fertility Status
- Some Additional Advices
- Extensions

Extensions

… it can get pretty complicted, once you add

- Linked lives
- Networks
- Behavioral components
- and etc.

… for simple applications use NetLogo or R

… for complex applications program it by yourself (it is worth learning a general purpose programming language such as Python, C#, Java)