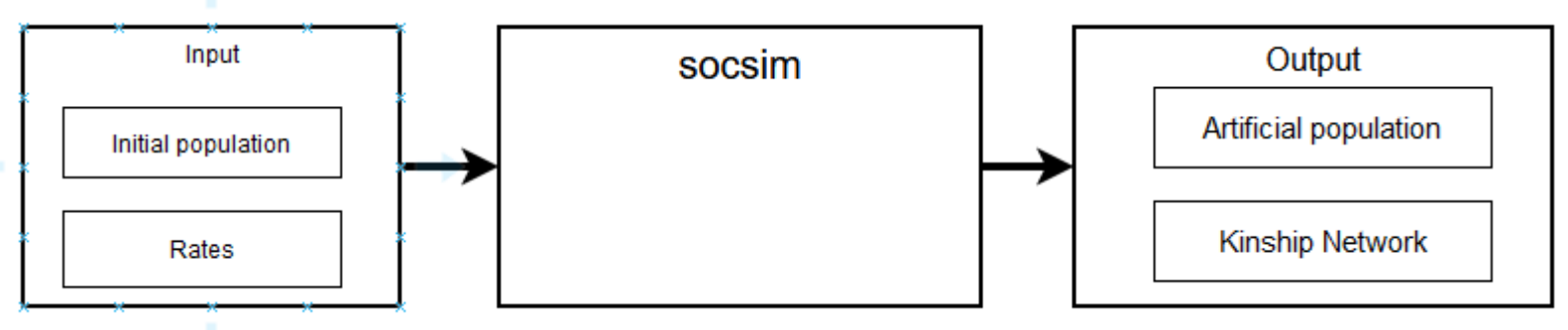
RSOCSIM

Tom Theile – Software Developer

Max-Planck Institute for demographic research

Are we living in a simulation?

Are we living in a simulation?



AGENDA

- History
- How does socsim work internally?
- Why is socsim now an R-package?

HISTORY

* first developed in the 70ies by Kenneth Wachter (?) and colleagues at the UC Berkeley
* written in FORTRAN
* rewritten in C in the 80ies
* Used throughout the years

Similar and derived software emerged:

WHY SOCSIM

- It is established and well tested
- It is complex enough to support all the things that are interesting to us,
- While still simple enough to not introduce too much moving parts and unforeseeable inner interactions
- The basic simulation mechanic is still the same as it was 50 years ago.

HOW DOES SOCSIM WORK?

OVERVIEW

- Every person is an individual object
- The population is a long list of persons
- When a birth happens, a new person is created and appended to that list
- At the start of the simulation (or after an event), every person gets a “next event”
- Events can be marriage, divorce, childbirth, death, etc.
- Time increments in discrete timesteps
- In every time step, all events scheduled for this time “happen”
- A simulation can consist of 1 or more “segments”, every segment can have different rates
- At the end of the simulation, socsim writes the population into output files

HOW DOES SOCSIM WORK?

PERSON

Every person has some parameters:

- dead/alive
- Sex
- *female or male, nothing in between*
- marital status
- *single, married, divorced, cohabiting, widowed*
- parity
- *total number of children born to a woman*
- Group Number
- *between 0 and 63. Can be used to simulate different groups/countries/towns.... with transition rates between groups*
- Age
- *In months, changes with time steps, starts at 0. Max age is 200 years*
- Next event

HOW DOES SOCSIM WORK?

EVENTS

- birth
- *creation of a new person with age 0 and parameters that are random (sex) or derived (marital status=single at birth, …)*
- death
- *according to mortality rates. These are specific to parameters of the individual persons (age, gender, groups, parity, marital status)*
- marriage
- *Just like in real live, Marriage is the most complicated event, because it involves 2 persons.*
- *Socsim comes with 3 Marriage market systems*
  - **1 - Marriage rates for both males and females**
  - **2 - Marriage rates only for females - males just get picked**
  - **3 - No marriage rates; Marriage just before a childbirth to an unmarried mother**
- Divorce
- Childbirth

HOW DOES SOCSIM WORK?

EVENT SCHEDULING

- Every person has 1 next event scheduled for the future
- Every timestep, all events scheduled for this timestep are executed in random order
- After an event, new events are scheduled for affected persons

- For every possible event, a waiting time is calculated based on rates and random numbers
- The event with the shortest waiting time is picked as the next event

HOW DOES SOCSIM WORK?

EVENT SCHEDULING

The waiting time algorithm is conceptually equivalent to drawing a random number u, from a uniform (0,1) distribution, calling u the probability that the event will not yet have occurred, then finding the first month by which the probability of non-occurrence is less than or equal to u. The probability that an event will not have occurred by a particular month T is given by the expression

$$\prod_{t=0..T} (1 - p_t) \tag{3}$$

Where p_t is the probability of the event's occurrence in period t conditioned on it not having occurred at any time before t. Since $(1 - p_t)$ is always between 0 and 1, the expression given above is nonincreasing in T. Consequently, beginning with $t = 0$ we can successively multiply the $(1 - p_t)$ terms together until the value of the product falls below u. What Socsim does is mathematically equivalent to this procedure, however, the implementation in function `datev` takes advantage of fact that the probabilities can be the same over months or years and works with powers of $(1 - p_t)$.

HOW DOES SOCSIM WORK?

SEGMENTS

- One Simulation can consist of many segments
- A new segment can introduce new rates

HOW DOES SOCSIM WORK?

AT THE END OF A SIMULATION

- The virtual population will be written to output files
  - *opop*
    - **contains every person that has lived during the whole simulation.**
      - **Month of birth, month of dead, etc.**
  - *omar*
    - **Contains information of every marriage**

- These output files can then be read in with R to analyze the virtual kinship network, retrieve rates etc.

19711	0	2	3	3197	13043	12555	18526	18526	37574	15933	4	4096	0.000000
19712	0	1	3	3197	14767	14397	19106	19106	33949	18076	4	3988	0.000000
19713	0	2	3	3197	14473	14068	0	0	32148	17978	4	4125	0.000000
19714	0	2	3	3197	13039	12587	18302	18302	35137	11100	4	4064	0.000000
19715	0	2	3	3197	13052	12552	0	0	37177	16596	4	3872	0.000000
19716	0	1	3	3197	12832	12557	17985	17985	37329	11103	4	4047	0.000000
19717	0	3	3	3197	15153	14778	0	0	0	0	1	3208	0.000000
19718	1	4	3	3197	13153	12680	19007	19007	31797	10616	4	3632	0.640667
19719	1	4	3	3197	12531	12453	19114	19114	34126	10623	4	3975	0.837893
19720	1	4	3	3197	12464	11373	19141	19141	30587	10620	4	3600	1.889329
19721	1	1	3	3197	14480								
19722	1	4	3	3197	12317								
19723	0	3	3	3197	13433								
19724	1												
19725	0												
19726	1												
19727	1												
19728	1	2	3	3197	14854								
19729	0	2	3	3197	14696								
19730	0	4	3	3198	11861								
19731	0	1	3	3198	12940								
19732	0	4	3	3198	11876								
19733	0	2	3	3198	13229								
19734	0	3	3	3198	15207								
19735	1	2	3	3198	12876								
19736	0	4	3	3198	15622								
19737	0	3	3	3198	12419								
19738	1	1	3	3198	11735								
19739	0	1	3	3198	14927								
19740	0	4	3	3198	12426								
19741	0	4	3	3198	11400								
19742	1	4	3	3198	14395								
19743	1	1	3	3198	14076								
19744	0	1	3	3198	15248								
19745	0	3	3	3198	15213								

.opop-file

1 row = 1 person

position	name	description
1	pid	Person id unique identifier assigned as integer in birth order
2	fem	1 if female 0 if male
3	group	Group identifier 1..60 current group membership of individual
4	nev	Next scheduled event
5	dob	Date of birth integer month number
6	mom	Person id of mother
7	pop	Person id of father
8	nesibm	Person id of next eldest sibling through mother
9	nesibp	Person id of next eldest sibling through father
10	lborn	Person id of last born child
11	marid	Id of marriage in .omar file
12	mstat	Marital status at end of simulation integer 1=single;2=divorced; 3=widowed; 4=married
13	dod	Date of death or 0 if alive at end of simulation
14	fmult	Fertility multiplier

17443	22641	23563	3789	4039	3	12313	13246
17444	31585	30837	3789	3998	3	0	0
17445	13207	12701	3789	3841	3	14940	15718
17446	31573	30845	3789	0	16	0	0
17447	13999	13320	3789	3807	3	15558	16566
17448	31568	30840	3789	0	16	0	0
17449	31588	30828	3789	4167	3	0	0
17450	27568	30875	3789	4205	3	15203	0
17451	31577						
17452	31576						
17453	17214						
17454	31560						
17455	29796						
17459	31595						
17460	25431						
17461	31607						
17462	31626						

Marriage file

1 row = 1 marriage

position	name	description
1	mid	Marriage id number (unique sequential integer)
2	wpid	Wife's person id
3	hpid	Husband's person id
4	dstart	Date marriage began
5	dend	Date marriage ended or zero if still in force at end of simulation
6	rend	Reason marriage ended 2 = divorce; 3 = death of one partner
7	wprior	Marriage id of wife's next most recent prior marriage
8	hprior	Marriage id of husband's next most recent prior marriage

WHAT IS DIFFERENT NOW? WHY R?

- Most of the code is still the old C code (mostly from Carl Mason)
- Windows-support
- R-package instead of command-line program
- R-package includes useful functions:
- create_simulation_folder(),Retrieve_rates(), retrieve_kin(),

FUTURE WORK

- Making it more robust on all platforms, better tested
- Improving the documentation
- More R-functions to analyze the simulation output
- Parallelization
- Integrating user feedback

THANK YOU FOR YOUR ATTENTION

Tom Theile
Research Software Developer
DCoDe Departement MPIDR

theile@demogr.mpg.de