A parallel microsimulation package for modelling cancer screening policies

eScience, 2016-10-26, Baltimore

Andreas Karlsson
andreas.a.karlsson@ki.se

Co-authors: Niten Olofsson, Erwin Laure & Mark Clements

Outline

- Microsimulation R Package
- Parallelisation - four approaches
- Case Study - prostate cancer screening

Microsimulation R Package

- Our package: https://github.com/mclements/microsimulation
  - Open source tool for planning cancer screening policies
  - Fast and flexible discrete-event simulation (DES)
  - Common random number support
  - In-simulation report reductions
  - Support for distributed/shared memory parallelism

Notable related work

- Closed source *MODGEN* microsimulation language by Statistics Canada [1]
- … and recent open source re-implementation *Openm++* [2]
- *simmer* DES library for specific process-oriented simulations, with model specification in R and a C++ core [3]

Open source infrastructure

RngStreams
: a random number streams library for C++ [4]

SSIM
: a C++ discrete event simulation library. The library defines the basic interface for a process and provides the m simulation scheduler [5]

Boost libraries
: for `R` library to be compliant with `C++98` [6]

Rcpp
: R/C++ interface library, extended to wrap vectors of tuples, and maps using tuples as keys [7]

Pseudocode

```
for j in 1...J do                  // iterate over people
  schedule events                  // initialise events
  while queue is not empty do
    event = pop(queue)
    handle event: begin            // new state?
      schedule new events
      write to report
    end
  end
end
```

Outline

- Microsimulation R Package
- Parallelisation - four approaches
- Case Study - prostate cancer screening

Parallelisation

- computationally intensive to run the simulation with high precision
- it is also computationally intensive to calibrate the model

- Microsimulations are computationally intensive, particularly for model calibration.
- Four different methods of parallelisation:
    1. Shared memory: R-parallel
    2. Shared memory: OpenMP
    3. Distributed memory: MPI
    4. Hybrid: OpenMP/MPI

1. Shared memory: R-parallel

2. Shared memory: OpenMP

3. Distributed memory: MPI

4. Hybrid: OpenMP/MPI

Benchmark

- Eight core nodes on 16 node cluster
- Software:
    - *OpenMP* with *gcc* version 4.8.1
    - *R* version 3.0.2
    - *Open MPI* version 1.4.1
- Model further described in case study
- Simulation size \(\text{10}^\text{7}\)

Performance

Efficiency

Outline

- Microsimulation R Package
- Parallelisation - four approaches
- Case Study - prostate cancer screening

Case Study

Adapting the FHCRC model to Sweden

- Prostate cancer is the most common cancer diagnosis for men in both Sweden and the
- Opportunistic prostate-specific antigen (PSA) testing causes over-diagnosis and over-treatment
- Cost-effectiveness analysis (CEA) can be used to plan cancer screening policies

- Based on a prostate cancer screening model from the Fred Hutchinson Cancer Research Center (FHCRC) [8] [9]
- *Incorporate* Swedish input data
- *Calibrate* (or fit) parameters using Swedish calibration targets
- *Validate* the model predictions using Swedish data

Natural history modelling

- A natural history (NH) model describes the course from being healthy, to disease onset, progression and death. The model is motivated by biological mechanisms.
- **Natural history modelling can be used to:**
  - to generalise results from randomised controlled trials (RCT)
  - predict effects for different screening protocols
  - calculate cost-effectiveness using predicted lifetime costs and quality of life.

Data Sources

Study name	Description	Study size
STHLM0	Population of men with a PSA test in Stockholm from 2003. Linked with registrations, deaths, migration, prescribed drugs etc.	400,000 men
STHLM3	Diagnostic trial for biomarker development of a prostate cancer screening test in 2013-2014.	60,000 men
PCBaSe	Survival at 10 and 15 years by PSA, grade and stage. PCBaSe links the national quality register on prostate cancer with the cause of death register	80,000 cases

Research question: How should we plan for better prostate cancer testing?

Parameters affecting screening include:

- Test characteristics for different tests (e.g. PSA, S3M, 4K and PHI)
- Screening ages
- Re-screening intervals

- Screening history
- Screening test compliance (if invitations are organised)
- Biopsy compliance
- Treatment effectiveness
- etc…

Modelled states

Microsimulation traces and screening

- only in a simulated data, not be observable.
- Stochastic life lines generated by the microsimulation

- man in a scenario without screening
- man in a scenario with screening
- classic timings in a screening context
- how do we use this?

Predictions

- PSA testing scenarios:
  - No screening
  - 2-yearly ages 50-70
  - 4-yearly ages 50-70
  - Current
- Outcomes:
  - Prevalence
  - Mortality rate ratios
  - Cost-effectiveness

Prevalence

Mortality rate ratios