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
- … and recent open source re-implementation *Openm++*
- *simmer* DES library for specific process-oriented simulations, with model specification in R and a C++ core
OPEN SOURCE INFRASTRUCTURE

RngStreams
a random number streams library for C++

SSIM
a C++ discrete event simulation library. The library defines the basic interface for a process and provides the main simulation scheduler

Boost libraries
for R library to be compliant with C++98

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

- 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

![Diagram showing the relationship between Rcpp, OpenMP, and Rmpi in a hybrid system.]
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 $10^7$
PERFORMANCE

The graph shows the performance of different implementations as the number of cores increases. The x-axis represents the number of cores, while the y-axis represents the time in seconds. The implementations include R parallel, OpenMP, MPI, and OpenMP/MPI. The graph indicates that as the number of cores increases, the time decreases, suggesting improved performance with parallel processing.
OUTLINE

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

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

- 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 (RCTs)
  - predict effects for different screening protocols
  - calculate cost-effectiveness using predicted lifetime costs and quality of life.
## DATA SOURCES

<table>
<thead>
<tr>
<th>Study name</th>
<th>Description</th>
<th>Study size</th>
</tr>
</thead>
<tbody>
<tr>
<td>STHLM0</td>
<td>Population of men with a PSA test in Stockholm from 2003. Linked with registrations, deaths, migration, prescribed drugs etc.</td>
<td>400,000 men</td>
</tr>
<tr>
<td>STHLM3</td>
<td>Diagnostic trial for biomarker development of a prostate cancer screening test in 2013-2014.</td>
<td>60,000 men</td>
</tr>
<tr>
<td>PCBaSe</td>
<td>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.</td>
<td>80,000 cases</td>
</tr>
</tbody>
</table>
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

Healthy → Local-regional Gleason 2-6 → Preclinical

Local-regional Gleason 7 → Distant stage Gleason 2-6

Local-regional Gleason 8-10 → Distant stage Gleason 8-10

Clinical

Local-regional Gleason 2-6 → Distant stage Gleason 2-6

Local-regional Gleason 7 → Distant stage Gleason 7

Local-regional Gleason 8-10 → Distant stage Gleason 8-10

Other-cause mortality
MICROSIMULATION TRACES AND SCREENING

- Screening detected pre-clinical disease
- Sojourn time
- Delay time
- Lead time
- Screening benefit
- Death from other causes
- Death from cancer
- Cancer onset
- Log(PSA)
- Age (years)
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

Screening pattern
- No screening
- 2-yearly
- 4-yearly
- Current

Prevalence of prostate cancer

Age
COST-EFFECTIVENESS INTRO

- The cost-effectiveness from the microsimulation can be described using utilities and costs for screening intervention $k$, such that:

$$\text{Effectiveness}_k = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\infty} \frac{dU_{ik}(t)}{(1 + \delta)^t}$$

$$\text{Costs}_k = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\infty} \frac{dC_{ik}(t)}{(1 + \delta)^t}$$

where we simulate for $n$ individuals with index $i$, with individual-based cumulative utilities $U_{ik}(t)$ and costs $C_{ik}(t)$ at time $t$, with discounting $\delta$ (e.g. $\delta = 0.03$).
CONCLUSIONS

- Microsimulation is increasingly being used to plan cancer screening (e.g. by CISNET, NICE, USPSTF).
- Four-yearly testing would reduce costs and have similar effectiveness as current PSA testing.
- Coupling R and C++ eases software dissemination and allows for high-level R methods.
- The performance of the hybrid OpenMP/MPI came at the cost of significant re-factoring.
THANK YOU ALL FOR LISTENING!

We acknowledge funding support from the Swedish eScience Research Centre, the Nordic Information for Action eScience Center and the Swedish Cancerfonden (CAN 2012/765).
BIBLIOGRAPHY