Catalyst: agents of change – Comparing compartment and agent-based models

Shannon K. Gallagher and William F. Eddy

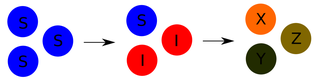
Goal: understand the spread of disease

We want to know, e.g.

- The severity of the recent **Ebola** outbreak
- The effect of cohort on spread of **measles**
- How to deal with the **Next One**

Stochastic epidemic modelling has answers*

SI-framework
Susceptible and
Infectious states +



Estimate number of individuals in each state
Two common types of models are

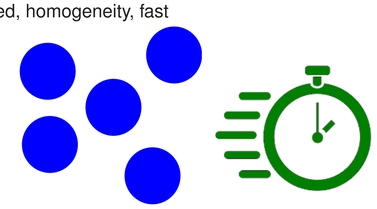
- Compartment (**CM**)
- Agent-based (**AM**)

* some of the time

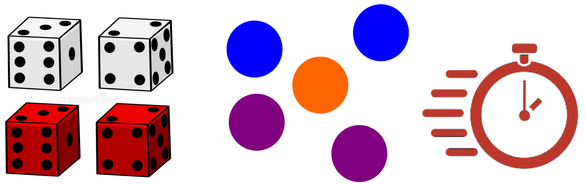
Compartment and agent-based models

CM: equation-based, homogeneity, fast

$$\begin{aligned} \frac{dS}{dt} &= -\beta\frac{SI}{N} \\ \frac{dI}{dt} &= \beta\frac{SI}{N} - \gamma I \\ \frac{dR}{dt} &= \gamma I \\ N &= S + I + R \end{aligned}$$



AM: simulation-based, heterogeneity, less fast



Idea: We can leverage the advantages

Some background theory

$\mathbf{X}(t) = (X_1(t), \ldots, X_K(t))$ for $t = 0, 1, \ldots, T$
$X_i(t)$: Number of individuals in state i at time t

Theorem

For a given CM, we can create an equivalent AM with **independent** individuals such that

$$X(t)^{CM} \stackrel{d}{=} X(t)^{AM}$$

Extension to some **dependent** individuals

Q: When should we use one over the other?

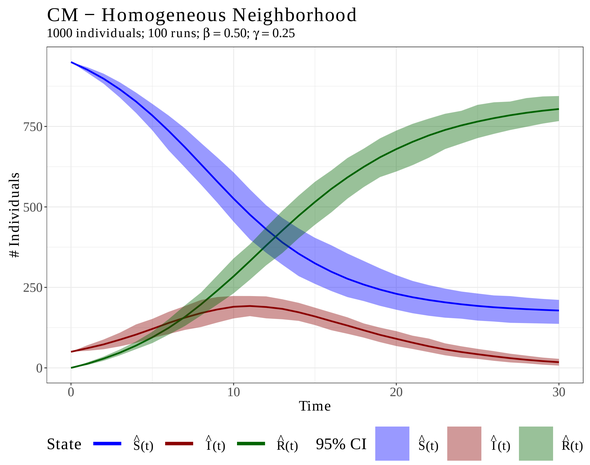
Models, statistics, distances, and distributions (for an SIR system)

Base model

For $t = 0, 1, \ldots, T$

$$\begin{aligned} \hat{S}(t) &= \hat{S}(t-1) - \hat{\epsilon}_{\hat{S}}(t-1) \\ \hat{I}(t) &= N - \hat{S}(t) - \hat{R}(t-1) \\ \hat{R}(t) &= \hat{R}(t-1) + \hat{\epsilon}_{\hat{R}}(t-1) \\ \hat{\epsilon}_{\hat{S}}(t-1) &\sim \textbf{Binomial}\left(\hat{S}(t-1), \frac{\beta\hat{I}(t-1)}{N}\right) \\ \hat{\epsilon}_{\hat{R}}(t-1) &\sim \textbf{Binomial}\left(\hat{I}(t-1), \gamma\right) \end{aligned}$$

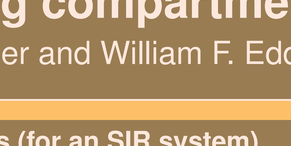
- $(S(0), I(0), R(0)) = (\hat{S}(0), \hat{I}(0), \hat{R}(0))$
- N is the fixed # of individuals

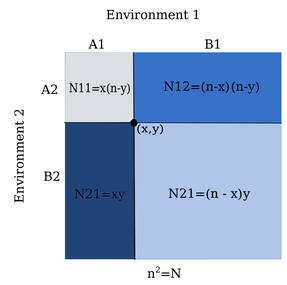


Process

(1) Simulate sample paths from a fixed AM
(2) Estimate $\tilde{f} = (\tilde{S}(t), \tilde{I}(t), \tilde{R}(t))$
(3) Simulate sample paths from CM using $\hat{f}$
(4) Compare distributions of sample paths
(5) Repeat with varying heterogeneity of AM

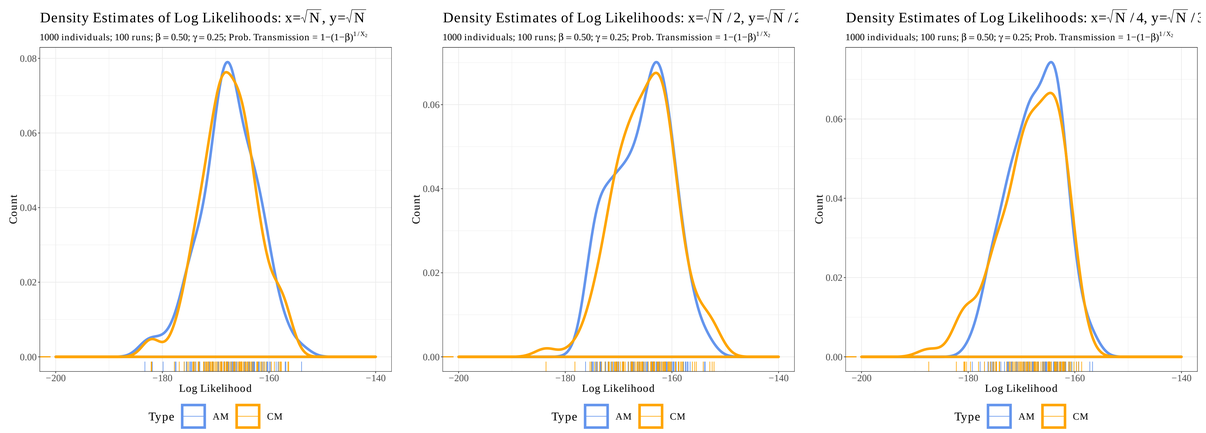
- e.g. two agent interaction locations
- e.g. vary x and y





Catalyst: compartment and agent-based temporal analysis technique

Results from estimating and comparing log likelihood



Conclusions

- The CM can estimate the mean AM well
- Distance ⊥ heterogeneity in this case
- CM is $\sim$ **30*x*** faster than the AM
- Plug-in estimate of log like. performs well

Remaining Questions:

- How should we compare distributions?
- Exact relationship between dep. AM and CM?

Future Work: Testing on real data

(1) **Measles in Hagelloch, Germany 1861**
Available from `surveillance` package in `R`
188 unique cases over the course of a year
SIR system
Age, sex, and household location covariates

(2) **Ebola in Freetown, Sierra Leone 2014-2017**
Over 8,000 cases
Population of $\sim$ 1.4 million people
Covariates provided through SPEW and OpenStreetMaps
SEIR system

Future of the software Catalyst

- All code is currently available at `https://github.com/shannong19/catalyst`
- Synchronization with Synthetic Populations and Ecosystems of the World (SPEW)
  Link: `http://stat.cmu.edu/~spew`
- A Shiny interface for easy use
- Faster simulations with `Rcpp`

Acknowledgments

We would like to thank the MIDAS Informatics Systems Group whom with NIH/NIGMS Grant 1 U24 GM110707-01 made this poster possible.

We would also like to thank Joel Greenhouse, Howard Seltman, and Samuel L. Ventura for their input and advice.

Dept. of Statistics & Data Science, Carnegie Mellon University

Shannon Gallagher: `sgallagh@stat.cmu.edu`