

Background

This is based on an exercise that our research staff run in a short course introducing researchers to infectious disease modelling. It captures some elements of the work that you would be doing here, though is simpler (in this job the work would also be much more collaborative and involve research staff directly - you would not be expected to know anything about infectious diseases or how they are modelled).

One can divide models into "deterministic" and "stochastic" - often it is possible to create a stochastic model and a deterministic model where the deterministic model represents the average of a very large number of runs of the stochastic model.

The exercise

We would like you to implement a simple epidemiological model so that we can see your approach to:

- * structuring a project that might grow in future
- * interface design
- * testing

The model is described below. We would like you to implement this with the idea that a researcher would be able to run simulations with this model changing the starting population and parameters ('beta' and 'sigma'). Because this model is stochastic, the user will want to be able to generate a number of simulations at once.

You can implement this model in any language or framework you like, so long as you provide sufficient instructions for us to be able to run the model. For the purposes of this exercise, we are not interested in speed of execution (we are also not looking for a GUI or any fancy user interaction - just code that could be used or extended).

Please share your completed model via GitHub or equivalent platform, or as a zip file.

The spec

A population of 'N' individuals is structured into "compartments" which contain Susceptible 'S', Infected ('I') and Recovered ('R') individuals. Susceptible individuals can be infected by infected individuals at rate 'beta' per interaction and infected individuals eventually recover and become immune to infection at rate 'sigma'. This model can be expressed as a set of deterministic differential equations:

...

$$dS/dt = \text{beta} * S * I / N$$

$$dI/dt = \text{beta} * S * I / N - \text{sigma} * I$$

$$dR/dt = \text{sigma} * I$$

...

(in this model $N = S + I + R$ and does not change over time).

To convert this to a discrete stochastic model we imagine that in a small period of time dt the chance of an event happening follows a Bernoulli trial.

In a discrete time step of size dt we model the number of infections (i.e., the number of individuals who move from S to I) as a binomial draw with $n = S$ and $p = \beta * I / N * dt$ and the number of recoveries (movements from I to R) as a binomial draw with $n = I$ and $p = \sigma * dt$ (see below for details).

Sensible parameters for the above are $\beta = 10$, $\sigma = 1$ with population size of ~ 1000 and a single initial infected individual. In this case the epidemic will complete over ~ 5 time units. You will qualitatively see S decrease, R increase and I peak around time unit 1.

Why the binomial distribution?

If an event happens at a constant rate r , then over a period of time dt the probability that an event has occurred is p , is $1 - \exp(-r * dt)$ and the chance that it does not happen is $1 - p$. The $1 - \exp(-r * dt)$ part comes from the cumulative probability function of the exponential distribution - if $r * dt$ is very small then this is approximately $r * dt$.

So if we assume events are independent, then for a single time period dt and fixed number of individuals n , the probability of seeing k events occur follows the binomial distribution.

In this particular model, if there are S individuals who can be infected, and per-capita infection happens at rate $\beta * I / N$ then over a time dt , $p = 1 - \exp(-\beta * I / N * dt)$ and the probability of k infections is described as $\text{Binomial}(S, 1 - \exp(-\beta * I / N * dt))$.