Practical: Introduction to stochastic simulation

James Truscott and Peter Winskill

Introduction

This practical introduces the principles of stochastic simulation: that is, simulations which capture the randomness inherent in real disease and demographic processes. Most of the models encountered so far involve the construction of rate equations (ordinary differential equations or ODEs) to describe the mean numbers of events (e.g. new infections) per unit time. When numbers are small, this method can provide nonsensical results, as subjects typically come in integer numbers!

Both ODE models (known as deterministic) and stochastic models emerge from a consideration of the random events of the epidemic. An event is an occurrence such as a susceptible becoming infected, or an individual dying or being born. In a deterministic approach, we consider the mean number of a particular event occurring in unit time. This is a rate and we construct ODE rate equations from them. In a stochastic simulation, we draw a random number with the same mean to represent the number of events that occurred. A key difference between deterministic and stochastic models is that a deterministic model will always produce the same output, whereas a stochastic model will produce different output for every simulation.

Example 1: Growth model

We begin with a deterministic growth model. This is a simple simulation of a birth process. Each individual has a birth rate (births/unit time), r. In a short time, dt, we would expect $r \times dt$ new births for each individual. Thus, in a population of size N there will be $rNdt$ births. The ODE for change in population size is

$$\frac{dN}{dt} = rN$$

where r is the growth rate per individual per unit time. The solution of this equation is exponential growth:

$$N(t) = N(0)e^{rt},$$

where $N(0)$ is the initial population at time $t = 0$.

In this model, how long does it take for the population to double?

How long does it take for the population to reach four times its original size?

Exercise

1. Open a web browser, and go to https://shiny.dide.imperial.ac.uk/infectiousdiseasemodels-2021/ and click on the link for the deterministic and stochastic growth model in the stochasticity practical.

2. Check the `odin` code in to see how the deterministic model is implemented. Note that as the ODE equation deals with mean values, N doesn't have to be a whole number.

3. Go to the “Visualise” tab and click “Run model” (near the bottom, on the left). You will see the model solution plotted on the graph. For now, click to hide N_stoch and N_Stoch (mean).

4. Try changing the birth rate parameter, r, to 0.01 and 0.5.

How does changing the birth rate influence population growth and the final population size?

5. Use the graph settings to choose a log scale on the y axis.

Looking at the graph on a log scale, what do you notice?

In this example we can extend the code to perform a stochastic simulation of the growth model. For a stochastic simulation, the probability that an individual gives birth in a short time, dt, is given by $r\ dt$. The same probability applies to each of the N individuals. Hence the number of new births will be binomially distributed:

$$\text{births in time dt} \sim Bin(N, rdt)$$

For the stochastic simulation, N is always a whole number of individuals.

Exercise

1. Go to the “Visualise” tab and set “Replicates” = 20 and click “Run model”. Make sure that the stochastic outputs are being plotted.

What do you notice?

What is the difference between the solutions' behaviour at early, low populations and at later times with larger populations? Explain this behaviour (It may help to view the output on the log scale).

In the long term, are the stochastic runs getting closer to the exact solution or further away, on the log scale?

What about on the linear scale?

2. Try some different values of r using the "Model parameters" control panel on the left.

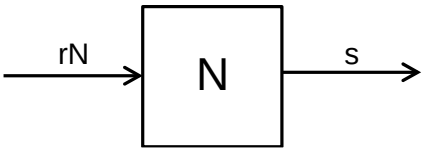
How does the value of r affect your comparisons between the exact and stochastic simulations?

3. Increase the number of replicates to 100. Due to the large number of simulations, individual runs will not be plotted, but the mean at each time point will be estimated. This is plotted instead of the individual simulations.

How does the mean compare with the exact solution?

For linear differential equation models, the exact solution is the mean of the stochastic simulations. The more stochastic simulations are run, the closer the mean gets to the exact solution. However, this conclusion breaks down when models are non-linear or there is the possibility of extinction. We will look at these possibilities in later examples.

Example 2: Stochastic birth and death model



We next extend the growth model by introducing deaths at a per-capita rate s per unit time. The ODE for N is made up of the difference between the mean growth rate and the mean death rate:

$$\frac{dN}{dt} = rN - sN$$

The difference $(r - s)$ is the net growth rate and the solution is either exponential growth or decline i.e., $N(t) = N(0)\, exp((r - s)t)$. If births exceed deaths $(r > s)$ the population grows. If deaths exceed births $(s > r)$ the population shrinks.

In the stochastic simulation, we have to be more careful as we are dealing with individuals. Each individual has a probability of reproducing and of dying in some short interval, dt. However, it can't die and then give birth! These two dependent events that can happen are known as *competing hazards*. To calculate the number of births and deaths in a short time, we first calculate the number of births or deaths:

$$\text{Births or deaths in dt, } n \sim Bin(N, (r + s)dt)$$

We then decide which were births and which deaths:

$$\text{Deaths in dt} \sim Bin(s / (r + s), n)$$

Another consequence of dealing with individuals is the possibility of extinction. If the population contains a single individual and they die, no more can be generated and the population will be zero from then on. In this system, individuals are created by individuals. Extinction can also occur in an infectious disease model, since infected individuals are only generated by other infected individuals. In an ODE model, populations are not whole numbers and extinction cannot occur. This is an important shortcoming of deterministic models.

Exercise

1. Open a web browser, and go to https://shiny.dide.imperial.ac.uk/infectiousdiseasemodels-2021/ and click on the

link for the stochastic growth and death model in the stochasticity practical. Have a look at the model code in the “Code & documentation” tab.

In this model, r = 0.1 and s = 0.06. What do you predict will happen to the population?

2. Set number of replicates to 20 and run the simulation a few times. Select the variable “N” to be plotted.

Roughly how often does the population die out?

For models which exhibit extinction behaviour, extinction is possible whenever the relevant population is small. As the population grows larger, the size of random fluctuations relative to the population size reduces and behaviour becomes closer to the equivalent deterministic model.

3. Run the simulation again, varying the value of “N0”

Is there a safe threshold initial population above which growth (rather than extinction) looks assured?

How might we estimate the probability of extinction? Look at the model equations. The output extinct takes the value 1 if N is 0, and 0 otherwise. The average value of extinct at a given time is therefore the probability that a population gone extinct by that time.

4. Set N0 back to 1 and re-run the model and examine the plot of extinct (mean).

What is the extinction fraction at the end of the run?

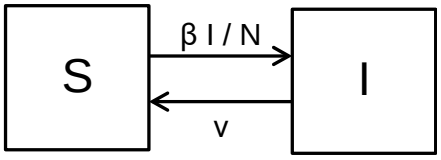
5. Try reversing the values of r and s (so s >r), to give a negative net growth rate.

What is the new extinction fraction?

In this model, the extinction fraction tends to s/r if r > s and 100% if s > r.

What happens when births equal deaths?

Example 3: Stochastic SIS model



The models discussed so far have looked at population size: growth, extinction, or simply "wandering about". In the next example we consider a model of infection transmission within a closed population (no births or deaths): the susceptible-infected-susceptible (SIS) model.

In this model we consider two types of event:

Infection events, which occur in each time step with probability $\frac{\beta\, I\, dt}{N}$ per susceptible individual, and

Recovery events, which occur in each time step with a probability $\nu\, dt$ per infected individual.

The ODE representation of the model is:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} + \nu I$$
$$\frac{dI}{dt} = \frac{\beta SI}{N} - \nu I$$

The analytical solution to this ODE can be shown to be:

$$I(t) = I^* \frac{1}{1 + \left({}^{I^*}/_{I_0} - 1\right) \exp\left(-(\beta - \nu)t\right)}$$

where I_0 is the initial number of infected individuals and $I^* = N\left(1 - {}^{\nu}/_{\beta}\right)$. After sufficient time, the infection and recovery processes balance each other out. This occurs when the number of infectious individuals reaches its equilibrium value I^*. We can write the equilibrium value as

$$I^* = N\left(1 - {}^{1}/_{R_0}\right)$$

since R_0 is defined as ${}^{\beta}/_{\nu}$. At this point, we say that the disease is *endemic* in the population.

Using Equation for the equilibrium number of infectious individuals, how do you expect the contact rate (β), recovery rate (ν) and total population (N) to affect the equilibrium number of infected individuals?

Action	Predicted effect on equilibrium I	Correct?
Increase contact rate (β)		
Increase recovery rate (ν)		
Increase total population (N)		

The stochastic simulation of this system is constructed in a similar way to the previous examples (see the code). Its behaviour, however, differs significantly from the deterministic simulation for some parameter values. We'll investigate these in the rest of this section.

Exercise

1. Open a web browser, and go to https://shiny.dide.imperial.ac.uk/infectiousdiseasemodels-2021/ and click on the link for stochastic SIS model in the stochasticity practical. Have a look at the model code in the "Code & documentation" tab.
2. First, on the "visualise" tab, we look at the endemic situation where there is a continuous population of infected individuals. Check that the variable I0_at_steady_state is set to 1 (TRUE). Set the number of replicates to 20.

Try varying the parameters from the table above in the model. Does the model behave as you expect?

3. Now look at the accuracy of the deterministic solution. With default parameter values (click "reset"), increase the number of replicates to 100 and set N to 30

How does the deterministic solution compare to the mean value of I from the stochastic simulation for small N?

4. Extinction can occur when the endemic infected population is small. Using 20 replicates, set $\beta = 0.5$, $\nu = 0.3$, and reduce N from the default value of 1000.

From what value do extinctions start happening?

Can you explain what is happening to the difference between mean stochastic behaviour and the deterministic?

5. Go to the "Sensitivity" tab. Do a parameter plot of the proportion of simulations that have gone extinct by the end of the simulation, against N, from 10 to 150.

What do you notice? Estimate a critical community size.

6. Now, we can look at a growing epidemic starting from a single individual. Go back to the "Visualise" tab and change I0_at_steady_state to 0 (FALSE). Set N high (300+) and the number of replicates to 1000.

What proportion runs are going extinct?

You should get the same extinction rate as in the birth and death model. Can you explain why?

7. Finally, let's re-examine the accuracy of the deterministic solution with a larger population size. With default parameter values (click "reset"), increase the number of replicates to 100 and set N to 100

How does the deterministic solution compare to the mean value of I from the stochastic simulation for large N?

When N is large, the differences between the simulations are due to the non-linear infection term: $\beta SI/_N$. The number of susceptibles (S) and infecteds (I) are negatively correlated, since when an individual leaves one class, (s)he enters the other. This correlation term is absent from the ODE model which then overestimates the force of infection, giving a higher population of infecteds.

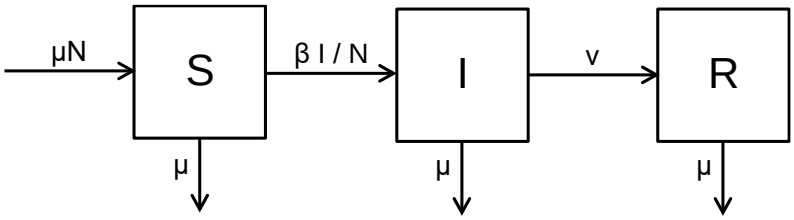
Example 4: Stochastic SIR model

The final part of the practical is to investigate the behaviour of a Susceptible-Infected-Recovered (SIR) model. The code to do this in `odin` can be easily adapted from the SIS code in previous section. Look again at the diagram of SIS model structure in Example 4. The SIS model has two compartments, S and I, although since the total population, N, is constant, S = N - I and we only need to keep track of the size of either S or I. There are only two possible events:

Event	Rate per head
Infection (S → I)	$\beta I/N$
Recovery (I → S)	ν

As described in the stochasticity lecture, at each time step, the number of each kind of event is evaluated and the populations updated. Check the code for Example 4 to see how this model is constructed in `odin`.

The structure of the SIR model is shown below. It has one extra compartment, R, representing recovered, immune individuals and four extra events: death of susceptible, infected and recovered individuals at rate μ, and birth of new susceptibles at rate B. To simplify, in any time period, we can make the number of births equal to the sum of the number of deaths. This keeps the total population constant.



The table of events for the SIR system is therefore:

Event	Rate per head
Infection (S → I)	$\beta I/N$
Recovery (I → R)	ν
Death of susceptible (S →)	μ
Death of infected (I →)	μ
Death of recovered (R →)	μ
Births (→ S)	μN

Why is the birth rate equal to μN?

Note that there are *two* competing hazards for S (death and infection) and for I (death and recovery). For this system, $R_0 = \beta/(\mu + \nu)$ and the new equilibria are:

$$S^* = 1/R_0$$

$$I^* = \frac{N\mu}{\beta}(R_0 - 1)$$

Exercise

1. Open a web browser and go to https://shiny.dide.imperial.ac.uk/infectiousdiseasemodels-2021/ and click on the link for the stochastic SIR model in the stochasticity practical. Have a look at the model code in the "Code & documentation" tab. The time units in this example are weeks.

From the default parameters, what is the mean infectious lifetime and the mean lifespan for individuals in this model?

2. The behaviour of the SIR model is in some ways similar and in some ways different from the simpler SIS model. Ensure that you are looking at the epidemic situation (i.e. set I0_at_steady_state to 0). Run the model and look at the curves for the mean number of infected people over time.

How does this differ from the prediction under the SIS model? Why?

3. Look at the curves for the number of infected people in each of the independent simulations (un-check S, R, S mean and R mean). The simulations take off at different times.

What could cause the epidemic simulations to take off at different times?

Why does the number of susceptibles slowly increase over time after the main epidemic (you may need to increase the time period to see this)?

4. We can look at probability of extinction for the SIR system in the same way as for the SIS. Increase the number of replicates to 200, set the initial infecteds to 1, and produce a graph of the variable extinction and sketch it below. What do the two phases of extinction in this graph represent?

The probability of extinction (in the first of the two phases) from a single infected individual is the same for SIR as SIS and is equal to 1/R0.

Given the different structures of SIS and SIR, why should this behaviour be the same?

5. Try varying the fraction of the population that's initially immune (by setting prop_immune).

What happens to the proportion that go extinct? Why?

6. We can look at critical community size for the SIR system in the same way as for the SIS.
    a. Now look at the endemic solution, by setting I0_at_steady_state to 1.
    b. As with the SIS model, using the "Sensitivity" tab, investigate the extinction fraction as the total population is varied. Set runtime to 1000 and R0 to 3 by changing β. When you've identified roughly the size of the critical population, run a parameter plot of extinct against N to illustrate it. Scan total population size from around 1000 up to about 1.5 million (this may be quite slow due to the large populations). Using a logarithmic scale for scanning N and setting replicates to 20 is also recommended.

How does your answer compare to that from the SIS model? Comment on the difference.

How does this result compare to the result derived in the Stochasticity lecture?

7. The open SIR system also exhibits some interesting and complex behaviours, such as stochastically-driven resonance. This phenomenon occurs when the system is close to equilibrium, so we initiate the model at the steady state.
    a. In the "Visualise" tab reset the parameters, set I0_at_steady_state = 1, N to 2 million, a single replicate, and the end time to 2000.
    b. Run the model with a single replicate. Examine the dynamics of the number of infected individuals over time.

What do you notice about the shape of the disease prevalence time series over the time period? Can you explain this behaviour?