Practical: Introduction to stochastic simulation

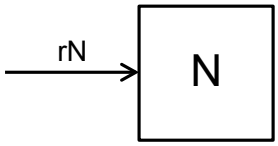
Joanna Lewis, James Truscott & Robert Verity

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 numbers of events (say 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: Deterministic growth model



We begin with EXAMPLE1. 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/stochasticity-practical/example1/` to open the `odin` model Example 1.

2. Check the `odin` code in Example 1 ("Code & documentation" tab) 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 "Run" tab and click "Run model" (near the bottom, on the left). You will see the model solution plotted on the graph. You can use the "Graph settings" (underneath the graph, on the right) to choose what variables to plot.

Which output variable represents the exact solution to the ODE?

Which output variable represents the numerical solution?

Compare the exact and numerical solutions

4. 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?

Example 2: Stochastic growth model

In this example we extend the code developed in Example 1 to perform a stochastic simulation. 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 `https://shiny.dide.imperial.ac.uk/stochasticity-practical/example2/` to open the `odin` model Example 2. Have a look at the model code in the "Code & documentation" tab.

2. Go to the "Run" tab and click "Run model" multiple times.

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.

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?

3. 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?

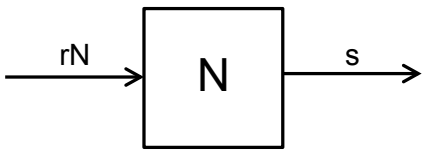
4. The `odin` interface allows the user to perform multiple (replicate) simulations. Try this:
    a. Set the number of replicates to 10 using the "replicates" box under "Run options".
    b. Click "Run".
    c. Now repeat this but enter 100 in the replicates box.

5. Finally repeat once more but check the box to display "mean_N". Odin will now perform 100 simulations and calculate the mean at each time point. This is plotted instead as well as 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 there is the possibility of extinction, as examined in Example 3, and also for non-linear models.

Example 3: Stochastic growth and death model



We next extend Example 2 by introducing deaths at a per-capita rate s per week. 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. Go to `https://shiny.dide.imperial.ac.uk/stochasticity-practical/example3/` to open the `odin` model Example 3. Have a look at the model code in the "Code & documentation" tab.

In this model, r = 0.5 and s =0.3. What do you predict will happen to the population?

2. Set number of replicates to 10, and run the simulation a few times.

Roughly how often does the population die out?

For models which exhibit extinction behaviour, extinction is possible whenever the relevant population is small. Hence, it's often possible to define a critical community size, above which extinction is very unlikely (but not impossible).

3. Set the number of replicates to 100 and run the simulation again.

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

How might we estimate the probability of fade-out? Look at the model equations. The output `fade` takes the value `1` if N is 0, and 0 otherwise. The average value of `fade` at a given time is therefore the probability that a population has faded out by that time.

4. Set $N0$ back to 1 and re-run the model.

What is the fade-out fraction after 14 weeks?

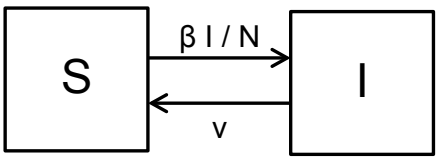
5. Try reversing the values of r and s (so s >r), to give a negative net growth rate.

What is the new fade-out fraction?

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

What happens when births equal deaths?

Example 4: 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 popluation (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 $I^* = N(1 - {}^{\nu}/_{\beta})$. 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

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

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

Using Equation (1), 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 the model EXAMPLE4.

2. First, 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 100.

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"), vary N from around 400 down to about 50.

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

For large N….

For small N….

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.

4. Fade-out can occur when the endemic infected population is small.

   Set $\beta = 0.5$, $\nu = 0.3$, and gradually reduce N from the default value of 100.

From what value do fadeouts start happening?

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

5. Go to the "Explore" tab. Do a parameter plot of the proportion of simulations that have "faded out" by the end of the simulation, against N, from 150 to 10. You can do this in the "report" ad "focal parameter" control panels.

What do you notice? Estimate a critical community size.

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

What proportion of fade-outs are you getting?

You should get the same fade-out rate as in example 3. Can you explain why?

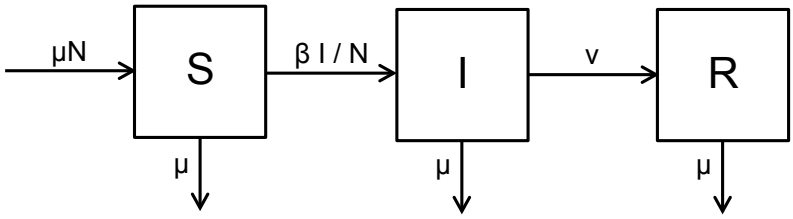
Example 5: Stochastic simulation of an SIR model

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 the `odin` model EXAMPLE 5. This is an adaptation of the EXAMPLE 4, with lines which have been added or changed marked with a comment. Note how the new variable, I, is introduced and initialised and the new events are handled. Ask a demonstrator to explain any changes you don't understand.

2. Dynamics:

   The behaviour of the SIR model is in some ways similar and in some ways different from the simpler SIS model.

   a. 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?

   b. Look at the curves for the number of infected people in each of the independent simulations (un-check S and R). 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?

3. Fadeout:

   We can look at probability of fade-out for the SIR system in the same way as for the SIS.

Produce a graph of the variable percent_fade and sketch it on the axes below. What do the two phases of fadeout in this graph represent?

The probability of fadeout (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?

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

What happens to the proportion that fades out? Why?

4. Critical community size:

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, investigate the fade-out fraction as the total population is varied. Set runtime to 1000 weeks and R0 to 3 by changing β. When you've identified roughly the size of the critical population, run a parameter plot of percent_fade against N to illustrate it. Scan total population size from around 1000 up to about 3 million (this may be quite slow due to the large populations).

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?

5. Stochastically-driven resonance

   The open SIR system also exhibits some interesting and complex behaviours.

   a. Set R0 to 3 and N to 1 million. Look at a single realization.

What do you notice about the shape of the disease prevalence time series over 2000 weeks? Can you explain this behaviour?