Introduction to stochastic modelling and its applications

Nicky McCreesh

nicky.mccreesh@lshtm.ac.uk

Learning outcomes

- Understand the differences between deterministic and stochastic models
- Be familiar with two methods types of stochastic models
- Appreciate the advantages and disadvantages of stochastic models

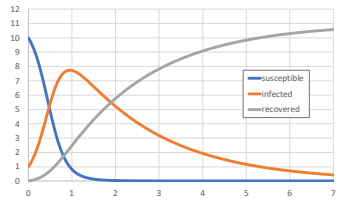
What is a stochastic model?

- A stochastic model is a model that includes some element of randomness or chance
- With a deterministic model, if you use the same initial conditions and parameter values, you always get the same results
- With a stochastic model, you get different results each time the model is run

Why do we need stochastic models?

- Suppose that one case of an SIR infection (eg measles) is introduced into a closed population of 10 susceptibles.
- How many of those 10 people will have been infected by the end of the outbreak?

Deterministic model



R0 and the probability of an epidemic

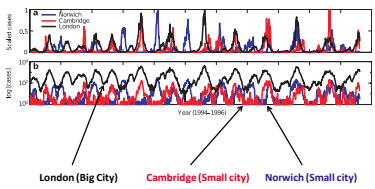
	R0 < 1	R0 > 1
Deterministic		
Stochastic	There is no chance of a major epidemic, but some small outbreaks may still occur	There is some chance of a major epidemic, but it will not always occur

R0 and the probability of an epidemic

- In reality, the probability of an epidemic depends on:
  - R0
  - The variance in R0
    - With R0>1, the probability decreases with increasing variance in the number of secondary cases
    - With R0<1, the probability increases with increasing variance in the number of secondary cases
  - The number of initial cases/introductions

Effect of population size – deterministic model

- Measles
  - Persists in cities
  - Dies out in town
- Deterministic model cannot capture die out

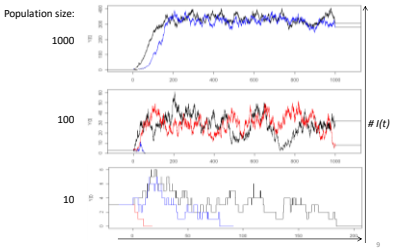


London (Big City) Cambridge (Small city) Norwich (Small city)

(Grenfell, Nature 2001)

Effect of population size – stochastic model

- Stochastic SIS model
- R0=1.5



When are stochastic models used?

- Describing the transmission dynamics of infections in small populations
  - e.g. hospital epidemics
- Describing transmission dynamics of infections where small numbers are infected in large populations
  - e.g. emerging or largely eliminated diseases
  - e.g. diseases with strong seasonality
- Stochastic models are also useful in exploring the critical population size or other criteria for the persistence of infection
  - e.g. measles
- Stochastic models are sometimes used for other reasons, with stochasticity being a (sometimes unwanted) side effect of the choice of model type

Part 2: How to build a stochastic model

Two approaches covered in lecture and practicals, and a third in notes

1. Set up a model which tracks the disease dynamics for *each individual* in the population and allow chance to help determine whether or not he/she becomes infected and develops disease at each time step.
2. Set up a model which allows chance to help determine the *number of secondary cases* which result from the cases in each generation.
3. Consider a continuous-time stochastic model using the Gillespie algorithm (in notes)

Possible approaches

	Individual	Population-based
Discrete time	Method 1 (Practical 17)	Method 2 (Practical 20)
Continuous time	Not covered, but possible	Method 3 (in notes)

- Choice 1: individual vs. population based
  - Individual-based model (IBM), method 1
  - Population-based model, method 2
- Choice 2: discrete or continuous time
  - Equivalent to difference and differential equations (earlier part of course)
- Methods 1 and 2 are both discrete time models
  - (stochastic equivalent to difference equations)
- Method 3 (in notes) is a continuous time model
  - (stochastic equivalent to differential equation model)

Method 1 (Read-Frost model)

- Back to our simple model…

- SIR
- One infected introduced into population of 10 susceptibles
- Assume:
  - All infected individuals become infected in one time step
  - All infectious individuals recover in one time step
- Take time steps of one serial interval (eg one week)

Method 1 (Read-Frost model)

- Model the infection status of each individual separately (S or I or R?)
- Use chance to determine whether each individual is infected in each time interval
- Let's say the risk of getting infected each week is 0.2
- Draw a number between 0-1 for each susceptible individual in the model
  - If the number is <0.2, they become infected
  - If the number is >0.2, they remain susceptible
- Usually interested in dynamic models: the risk of infection each week is related to the number of infected people

Making the model dynamic

- If the probability of being infected when there is one infected person is p=0.2, what is the probability when there are two infected people?
- NOT p=2 x 0.2
- Instead, need to calculate the probability that an individual comes into contact with *at least one* infectious case
- The easiest way to calculate this is to first calculate the probability that a susceptible person avoids effective contact with *all* of the infectious cases

The Reed-Frost formula

- Probability that 2 specific individuals come into effective contact between time t and t+1 is p. Therefore:
  - (1-p) is probability that individual avoids contact with 1 case.
  - (1-p)x(1-p) is probability that individual avoids contact with 2 cases.
  - (1-p)x(1-p)x(1-p) is probability that individual avoids contact with 3 cases.
- $(1-p)^{I_t}$ is probability that an individual avoids contact with all of the I_t cases.
- So the probability of coming into effective contact with at least one infectious case is

$$1-(1-p)^{I_t}$$

The algorithm

- Step 1: Calculate the expected risk λ_t that a susceptible individual becomes infected in the next time interval (using, for instance, the Reed-Frost formula)
- Step 2: Draw a random number between 0 and 1 for each of the susceptible individuals.
- Step 3: If the random number drawn for any individual is less than λ_t, then that individual becomes infected and hence a case by time t+1; otherwise that individual remains susceptible.
- Step 4: Count up number of cases at time t+1 (I_{t+1}), assuming that all those who were cases at time t are now immune.
- Step 5: If I_{t+1}=0, transmission ceases - continue to step 6, otherwise return to step 1.
- Step 6: The size of the outbreak is given by the sum of the number of cases at time t=1, 2, 3, 4,..,t.

A worked example

- You will also have a chance to try it for yourself in the next practical

An illustration of the IBM approach: time step 1

- Assume that:
  - $p = 0.15$ is the risk of 2 specific individuals coming into effective contact per unit time
  - $\lambda_0 = 1-(1-0.15)^1 = 0.15$ is therefore the risk that a susceptible individual becomes infected between time t=0 and t=1 when one initially infected

Individual number	Random number	Status by t=1
1	0.76	susceptible
2	0.14	case
3	0.30	susceptible
4	0.46	susceptible
5	0.76	susceptible
6	0.33	susceptible
7	0.61	susceptible
8	0.97	susceptible
9	0.31	susceptible
10	0.85	susceptible

An illustration of the IBM approach: time step 2

- Return to step 1, substitute I_1=1 into the Reed-Frost formula
  $\lambda_1 = 1-(1-0.15)^1 = 0.15$ (again) is the risk that a susceptible individual becomes infected between time t=1 and t=2

Individual number	Random number	Status by t=2
1	0.23	susceptible
2	-	immune
3	0.86	susceptible
4	0.41	susceptible
5	0.73	susceptible
6	0.03	case
7	0.09	case
8	0.02	case
9	0.53	susceptible
10	0.34	susceptible

An illustration of the IBM approach: time step 3

- Return to step 1, substitute I_2=3 into the Reed-Frost formula
  $\lambda_2 = 1-(1-0.15)^3 = 0.386$ is the risk that a susceptible individual becomes infected between time t=2 and t=3

An illustration of the IBM approach: time step 3

- Return to step 1, substitute I_2=3 into the Reed-Frost formula
  $\lambda_2 = 1-(1-0.15)^3 = 0.386$ is the risk that a susceptible individual becomes infected between time t=2 and t=3

Individual number	Random number	Status by t=3
1	0.215361	case
2	-	immune
3	0.270405	case
4	0.862182	susceptible
5	0.696761	susceptible
6	-	immune
7	-	immune
8	-	immune
9	0.098544	case
10	0.012308	case

An illustration of the IBM approach: time step 4

- In this instance, I_3=4 and $\lambda_3 = 1-(1-0.15)^4 = 0.478$.

Individual number	Random number	Status by t=4
1	-	immune
2	-	immune
3	-	immune
4	0.751125	susceptible
5	0.602339	susceptible
6	-	immune
7	-	immune
8	-	immune
9	-	immune
10	-	immune

Method 2: population-based discrete time stochastic model

- Individual-based model can be inefficient and computationally intensive, particularly with large populations:
  - Have to keep track of each individual
  - Need to draw a random number for each individual each time step
- Our second (population-based) method:
  - Keeps track of the *total number* of susceptibles, cases, and immune at each time step
  - Random numbers are used to determine *numbers of events* from the appropriate statistical distribution

Bernoulli trials and the binomial distribution

- For each susceptible,
  - λ: probability of being infected
  - (1- λ): probability of not being infected
- An experiment with a random binary outcome is called a Bernoulli trial.
  - Trial = infected? yes or no
- The binomial distribution gives the probability of k successes of n independent Bernoulli trials
  - Therefore, binomial distribution gives us the number of susceptibles who would be infected in each generation

Deriving the bionomial distribution

- The probability that k susceptibles will be infected is calculated by combining
  1) the probability that *a particular* sequence of outcomes is observed (eg SSSF)
  2) With the number of different sequences that result in k successes overall (eg FSSS, SFSS, SSFS, SSSF)

1) the probability that *a particular* sequence of outcomes is observed

- What if the probability of obtaining the sequence SSSF?
- Let:
  - S: no. of independent Bernoulli trials (= no. of susceptibles)
  - λ: probability of the event occurring (the same in each trial)
  - The probability of the first trial being a success is λ.
  - The probability of the first three trials being successes is λ. x λ. x λ.
  - The probability of SSSF is λ. x λ. x λ. x (1−λ.)
- This probability is the same for FSSS, SFSS, SSFS, and SSSF
- More generally, the probability of observing exactly *k* infections (successes) is

$$\lambda^k(1-\lambda)^{S-k}$$

2) The number of different sequences that result in k sucesses overall

- There are $\binom{S_t}{k}$ ways of choosing k out of S_t individuals to become infected.

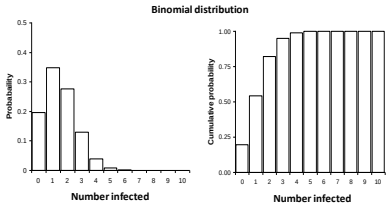
$$\binom{S_t}{k} = \frac{S_t!}{k!(S_t - k)!}$$

Bionomial distribution

- Combining the two parts, this gives us the probability that exactly k out of S_t susceptibles at time t are infected and develop disease by time t+1

$$P(I_{t+1} = k) = \binom{S_t}{k} \lambda_t^k (1-\lambda_t)^{S_t - k}$$

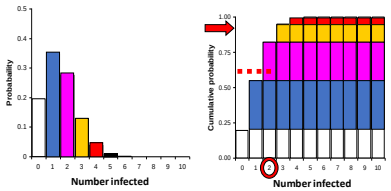
Probability and cumulative probability distribution of the number of newly infected individuals in the next time interval



P=0.15

$$P(i(S_t) = k) = \binom{S_t}{k} I_t^k (1 - I_t)^{S_t - k}$$

Probability and cumulative probability distribution of the no. of newly infected individuals in the next time interval



$$P(i(S_t)=k)=\binom{S_t}{k}\lambda_t^{k}(1-\lambda_t)^{S_t-k}$$

Implementing the model in Berkley Madonna

- Many statistical and modelling packages will draw a number from an appropriate binomial distribution
- In BM, code is binomial(*p*,*n*)
- Each time this code is run (i.e. at each time step) will draw a number from the binomial distribution with probability of success given by *p*, and number of trials by *n*
- NB: *p* here is "λ_t"; *n* is "S_t" from last slide notation

An illustration of Method 2

p = 0.15 risk of 2 specific individuals coming into effective contact per unit time

$\lambda_0 = 1-(1-0.15)^1 = 0.15$ risk that a susceptible individual becomes infected between time t=0 and t=1 when one initially infected

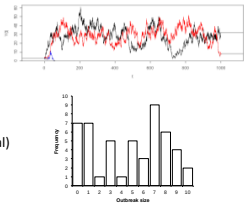
Number of individuals get infected (x)	Probability that x get infected	Cumulative probability
0	0.20	0.20
1	0.34	0.54
2	0.28	0.82
3	0.13	0.95
4	0.04	0.99
5	0.01	1.00
6	0.00	1.00
7	0.00	1.00
8	0.00	1.00
9	0.00	1.00
10	0.00	1.00

Adding additional stochasticity

- So far, the only part of the model that we have made stochastic is the number of people infected each time step
- Using similar methods (eg binomial distribution), we can also make other rates in the model stochastic
- For instance, if an infection has a variable infectious duration, we can make the risk of an infected individual becoming immune each time step stochastic
- Or the rate of becoming infectious following exposure

How to we display stochastic output?

- Wide range of different methods, that may be appropriate in different scenarios
- For example
  - Display all runs or 'sample' runs
  - Proportion of runs that 'take-off'
  - Mean and confidence interval
  - Median and IQR
  - Summary graphs
  - Mean only (if stochasticity incidental)



Advantages and disadvantages of stochastic models (compared to deterministic)

- They can be more realistic (incorporating chance)
  - e.g. transmission doesn't always occur
  - e.g. for small populations such as at the beginning and end of an epidemic when small number of infecteds
- They provide estimates of the variation in an outcome
  - i.e. not only mean (expected number of cases) but also variance in number of cases
- The main disadvantage is computational
  - It may be necessary to run many simulations to get reasonable estimate of average results and variance.

Summary and next steps

- So far today you have learnt:
  - What stochastic models are and when to use them
  - Two methods for setting up stochastic models
  - Some of the advantages and disadvantages of stochastic models
- Next you will explore a Read-Frost model (method 1) in Excel
- Tomorrow morning you will then set up your own discrete-time population-based stochastic model in Berkley Madonna