Day 3
Lecture 2: Formulating Stochastic models

Short course on modelling infectious disease dynamics in R

Ankara, Türkiye, September 2025

Dr Juan F Vesga

Aims of the session

- Learn how to formulate stochastic models
- Learn how stochasticity is introduced in compartmental and individual base models
- Understand the methodological difference with deterministic models

Formulations of a stochastic model

- Individual based/agent based models
  - Simulate each individual and events which happen to each individual
  - Can include much realistic detail, but can be slow to simulate

- Population or compartmental based models
  - Only keep track of total number of individuals in each compartment
  - Simulate the number of events that will happen to that group of individuals
  - Faster to simulate

Hazard

- The hazard is the probability per unit time of an event happening.
- It is equivalent to the instantaneous per-capita rate used in a deterministic model.

E.g. If an infected person has a hazard or rate of recovery of σ:

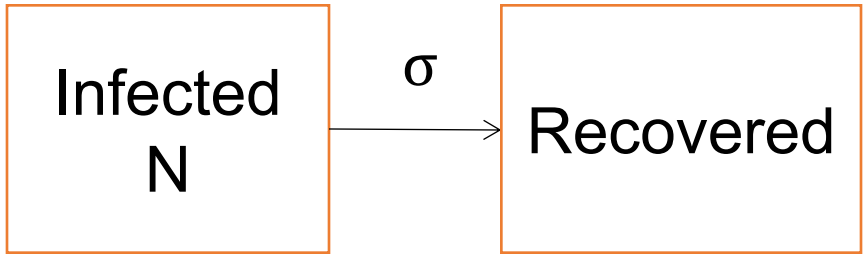
Infected Recovered

Probability of recovering in time dt is **σdt**, for small dt.

Hazard

- If we have N individuals who are all identical (with respect to the process we're talking about), the **binomial distribution** describes the distribution of the number of events in a given time.
- N trials of something with probability p, number of successes is a binomially-distributed random number:

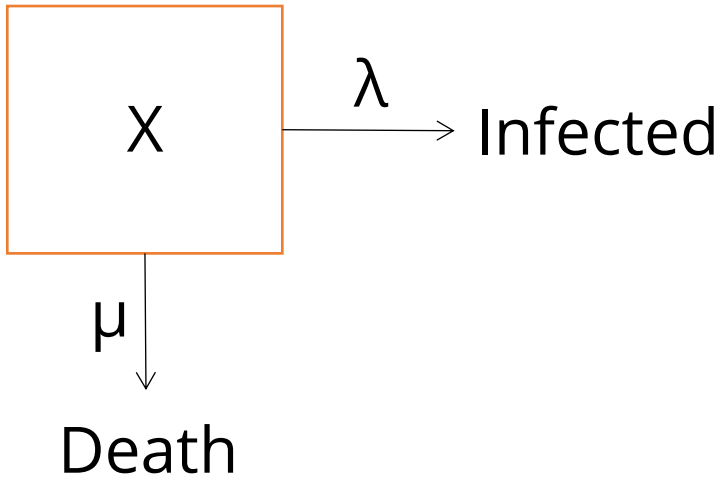
$$\text{Binomial}(p,N)$$



Number of recoveries in dt is from Binomial(σdt, N)

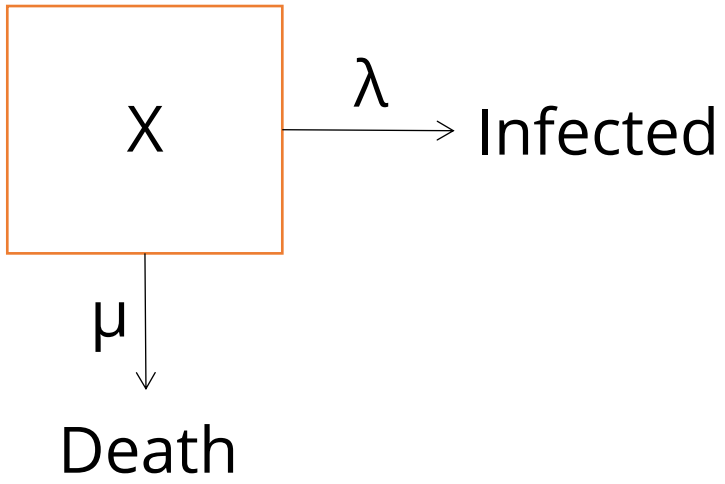
Competing Hazards

- Consider a population of susceptible which can become infected (rate: λ) but can also die (rate: μ).
- But they can't do both at the same time!
- In a time period, *dt*, how many die and how many become infected?



Competing Hazards

- It turns out that we can add the rates of the different events to
- get the total number to which anything happened.
- Then we can decide which of the two things happened to each.



Total number of events: $H_1 + H_2 = \text{Binomial}\left(\left(\mu + \lambda\right)dt, X\right)$

Split into H_1 and H_2: $H_1 = \text{Binomial}\left(\frac{\lambda}{\lambda + \mu}, H_1 + H_2\right)$

Simple infection process

A stochastic model is completely described by two things:

- The state of the population. E.g. the number of individuals in susceptible, infected, recovered states, etc.
- All possible events and their rates. E.g. infection rate, death rate, recovery rate, etc.

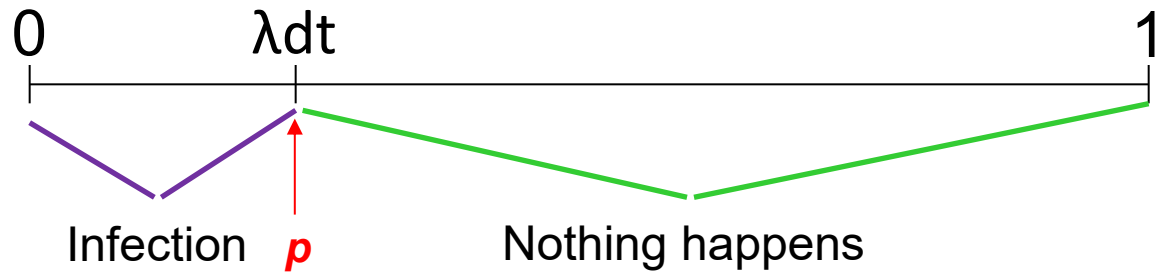
For the present case:

Population: S identical susceptible individuals,
I identical infected individuals.

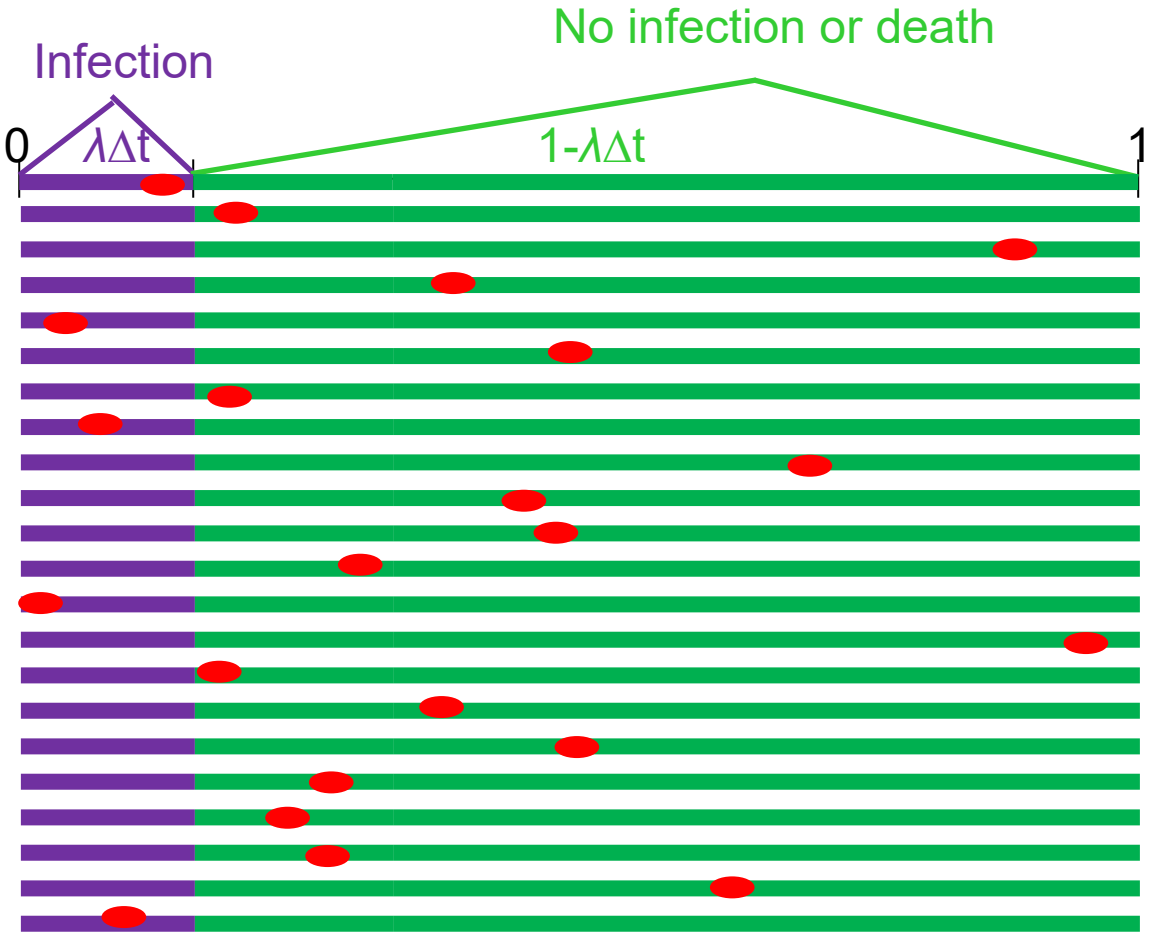
Name	What happens to population	Rate/individual
Infection	$S \rightarrow S - 1, I \rightarrow I + 1$	λ
recovery	$I \rightarrow I - 1$	σ

Individual base models

- Treat every individual separately.
- For *synchronous* update case, choose small update time-step, dt.
- Only one event could happen to the individual – get infected (rate λ)
- Every time-step, for each individual pick random number p (uniformly distributed between 0 and 1). If
  - $p<\lambda dt$ - the individual has become infected, so S$\rightarrow$S-1, I $\rightarrow$ I+1.
  - $p>\lambda dt$ - nothing happens.

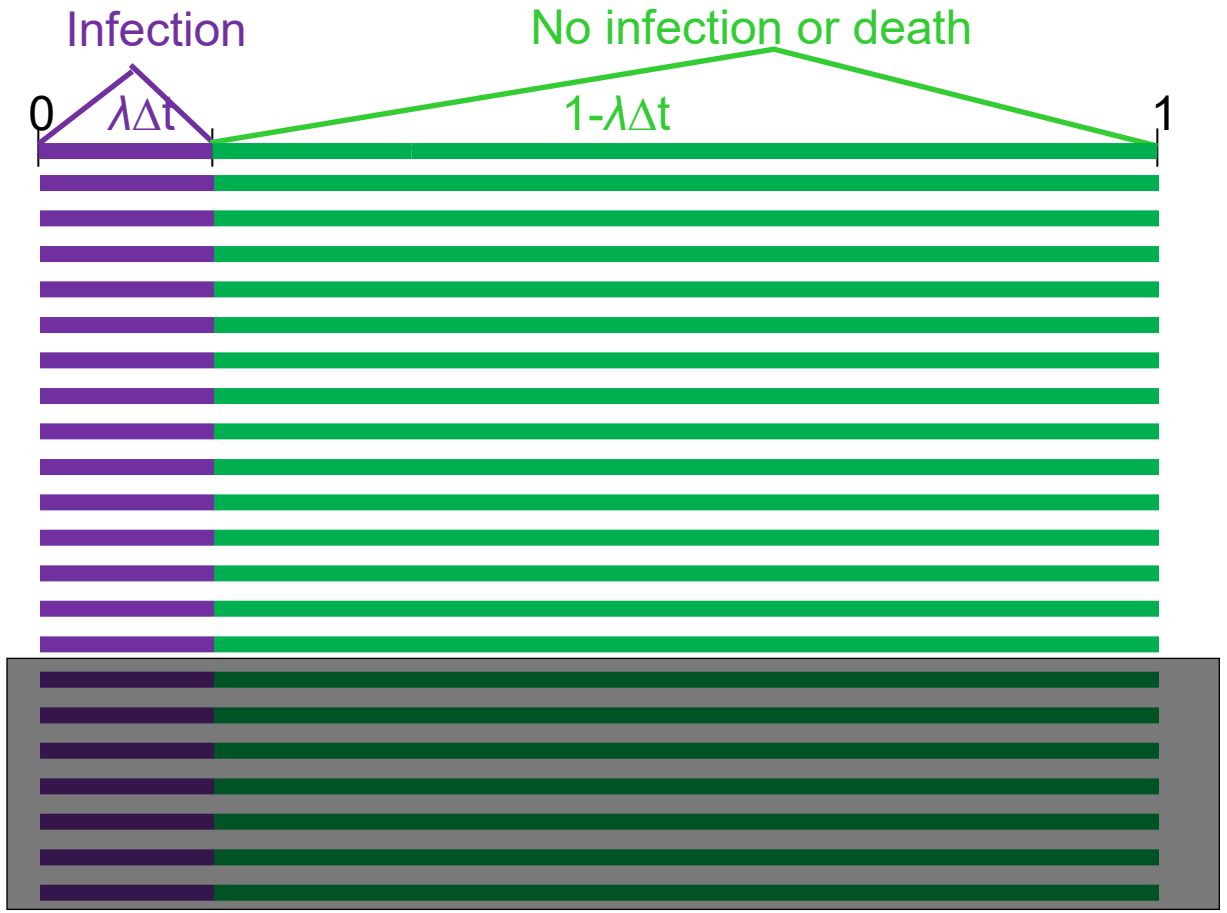


Multiple individuals-IBM



Each susceptible individual is ‘identical’
Could generate a random number for each (22 random numbers) and discover if infection or death
e.g. 5 new infections, 17 no change

Multiple individuals-Compartment



- 22 individuals, probability of infection per individual λDt
- Use Binomial to generate number of ‘successes’,
- Infections = Binomial(λDt,S) = 7 infections (for example)

SEIR

Susceptibles $$\frac{dX}{dt} = -R_0 \nu \frac{X}{N} Y$$

Exposed $$\frac{dW}{dt} = R_0 \nu \frac{X}{N} Y - \sigma W$$

Infectious $$\frac{dY}{dt} = \sigma W - \nu Y$$

Hollingsworth *et al* (2006) Nature Medicine

SEIR

Susceptibles $$\frac{dX}{dt} = -R_0 \nu \frac{X}{N} Y$$

Exposed $$\frac{dW}{dt} = R_0 \nu \frac{X}{N} Y - \sigma W$$

Infectious $$\frac{dY}{dt} = \sigma W - \nu Y$$

Susceptibles $$X(t+\delta t) = X(t) - \iota(t)$$

Exposed $$W(t+\delta t) = W(t) + \iota(t) - \kappa(t)$$

Infectious $$Y(t+\delta t) = Y(t) + \kappa(t) - \rho(t)$$

incidence $$\iota(t) = \text{Binomial}\left(R_0 \nu \frac{1}{N} Y(t) \delta t, X(t) \right)$$

infectious $$\kappa(t) = \text{Binomial}\left(\sigma \delta t, W(t) \right)$$

recovery $$\rho(t) = \text{Binomial}\left(\nu \delta t, Y(t) \right)$$

Hollingsworth *et al* (2006) Nature Medicine

SEIR

Susceptibles $\frac{dX}{dt} = -R_0 \nu \frac{X}{N} Y$

Exposed $\frac{dW}{dt} = R_0 \nu \frac{X}{N} Y - \sigma W$

Infectious $\frac{dY}{dt} = \sigma W - \nu Y$

Number moving from susceptible to exposed

Susceptibles $X(t+\delta t) = X(t) - \iota(t)$

Exposed $W(t+\delta t) = W(t) + \iota(t) - \kappa(t)$

Infectious $Y(t+\delta t) = Y(t) + \kappa(t) - \rho(t)$

incidence $\iota(t) = \text{Binomial}\left(R_0 \nu \frac{1}{N} Y(t)\delta t, X(t)\right)$

infectious $\kappa(t) = \text{Binomial}\left(\sigma\delta t, W(t)\right)$

recovery $\rho(t) = \text{Binomial}\left(\nu\delta t, Y(t)\right)$

Hollingsworth *et al* (2006) Nature Medicine

SEIR

$$\text{Susceptibles} \quad \frac{dX}{dt} = -R_0 \nu \frac{X}{N} Y$$

$$\text{Exposed} \quad \frac{dW}{dt} = R_0 \nu \frac{X}{N} Y - \sigma W$$

$$\text{Infectious} \quad \frac{dY}{dt} = \sigma W - \nu Y$$

Number moving from susceptible to exposed

rate

$$\text{Susceptibles} \quad X(t+\delta t) = X(t) - \iota(t)$$

$$\text{Exposed} \quad W(t+\delta t) = W(t) + \iota(t) - \kappa(t)$$

$$\text{Infectious} \quad Y(t+\delta t) = Y(t) + \kappa(t) - \rho(t)$$

$$\text{incidence} \quad \iota(t) = \text{Binomial}\left(R_0 \nu \frac{1}{N} Y(t) \delta t, X(t)\right)$$

$$\text{infectious} \quad \kappa(t) = \text{Binomial}\left(\sigma \delta t, W(t)\right)$$

$$\text{recovery} \quad \rho(t) = \text{Binomial}\left(\nu \delta t, Y(t)\right)$$

Hollingsworth *et al* (2006) Nature Medicine

Summary

- Stochastic effects important when considering
  - persistence
  - dealing with small populations (start and tail of epidemics)
  - or spatial spread.

- By chance can have outbreaks, even when $R_0<1$
- By chance can have no outbreak when $R_0>1$

- Stochastic models can be relatively simple to program, difficult to analyse

Extra materials

Further reading and something on distribution of stochastic events

Further reading

Books:

Renshaw, E., *Modelling biological populations in space and time*, Cambridge Univ. press, 1991

Bailey, N.T.J., *The mathematical theory of infectious diseases and its applications*, 2nd edition, Griffin, 1975

Papers:

Bartlett, M. S. (1957). "Measles periodicity and community size." J. Roy. Stat. Soc. A **120**: 48-70.

Bolker, B. M. and B. T. Grenfell (1995). "Space, persistence and dynamics of measles epidemics." Proc. Roy. Soc. Lond. B **348**: 308-320.

Jansen, V. A. A., N. Stollenwerk, et al. (2003). “Measles outbreaks in a population with declining vaccine uptake." Science 301(5634): 804-804.

Farrington, C. P., M. N. Kanaan, et al. (2003). "Branching process models for surveillance of infectious diseases controlled by mass vaccination." Biostatistics **4**(2): 279-295.

Advanced: Binomial versus Poisson

- Poisson distribution simulates the number of events that will happen in a small timestep = Poisson($\omega \Delta t N(t)$)

- Binomial formulation Binomial($\omega \Delta t$,N(t)) says there are N(t) tries, each with probability of success approximated by $\omega \Delta t$, how many are successful?
  - Can't have more events than there are people
  - For small timesteps won't matter, since there won't be many events
  - Have to calculate multinomials when there are multiple competing hazards

Population simulations

Individual based models can be very slow, and computing time increases as *N*.

If events are independent, the number of events of type *i* in a fixed time, D*t*, can be simulated by sampling from a binomial distribution (it can also be simulated by sampling from a Poisson distribution). Hence in a population of size S, the number of events of type *i*, is given by

$$k \sim Binomial\left(\lambda \Delta t, S\right)$$

When there are multiple events competing with each other, binomial or multinomial approximations perform better than using sampling from several Poisson distributions, and allow more consistent model formulation.

Advanced: Synchronous versus asynchronous

- *Synchronous:* at each timestep simulate all the events that could have occurred.

- *Asynchronous*: simulate when the next event is, and move forward to that time point,

- If there are many possible events each with a low probability of occurring, particularly in individual based simulations, its may be more efficient to have an asynchronous model