Model fitting and inference for infectious disease dynamics

Centre for the Mathematical Modelling of Infectious Diseases
London School of Hygiene & Tropical Medicine

Outline

Introduction

Linking models to data

Bayesian inference

Practical session in R

1. Introduction

Model fitting and inference for infectious disease dynamics

Model

A simplified description, especially a mathematical one, of a system or process, to assist calculations and predictions

Oxford English Dictionary

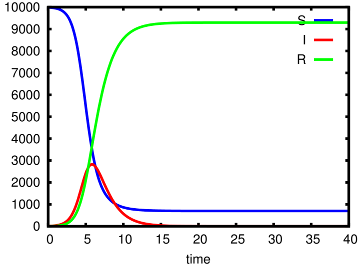
Mathematical model

Takes *parameters* and produces *output*
(using some set of rules / equations)

Model fitting and inference for infectious disease dynamics

SIR-type models

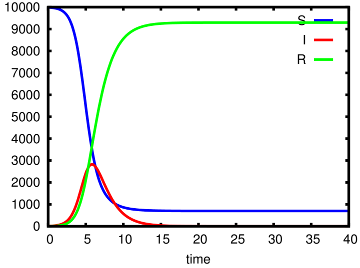
$$\frac{dS}{dt} = -\beta I \frac{S}{N}$$
$$\frac{dI}{dt} = \beta I \frac{S}{N} - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$



Model fitting and inference for infectious disease dynamics

SIR-type models

$$\frac{dS}{dt} = -\beta I \frac{S}{N}$$
$$\frac{dI}{dt} = \beta I \frac{S}{N} - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$



Mechanistic models

description vs mechanism

Model fitting and inference for infectious disease dynamics

Parameter estimation

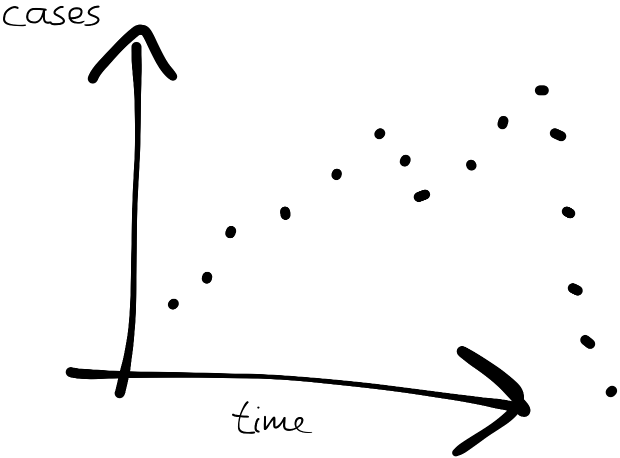
Given a model, what are the parameter combinations that best fit the data (in whichever way)

Why are we doing this?

- Learn something about the system
  - test a scientific hypothesis
    - e.g., why did the UK H1N1 epidemic wane in summer 2009? (Dureau et al., 2013)
  - estimate parameters
    - e.g. which fraction of infections with cholera in Bangladesh are asymptomatic? (King et al., 2008)
  - sometimes in real time
- Validate the model
  - especially: for prediction

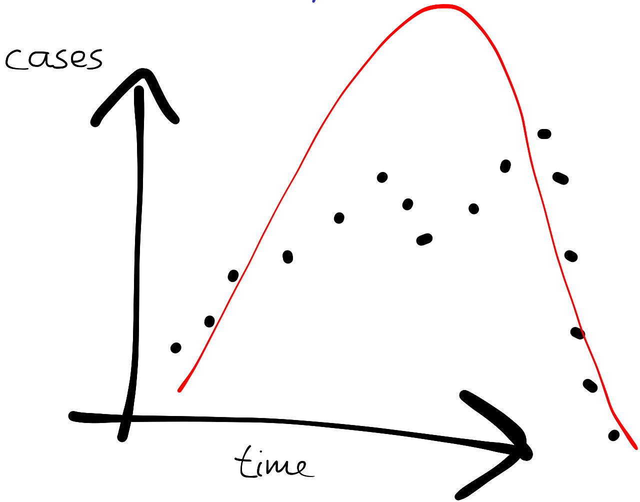
Model fitting and inference for infectious disease dynamics

What do we mean by “best fit the data”?



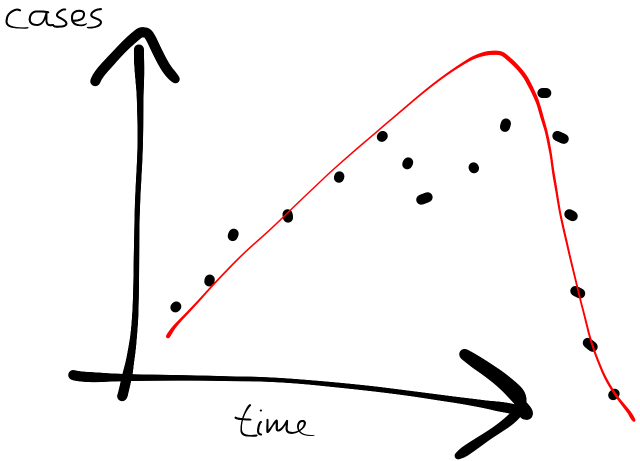
Model fitting and inference for infectious disease dynamics

What do we mean by “best fit the data”?



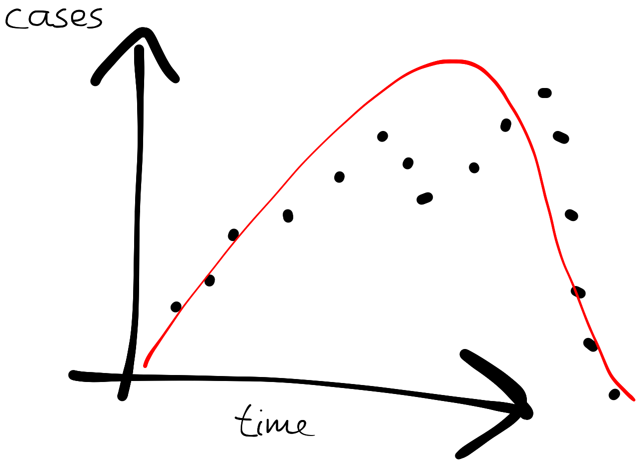
Model fitting and inference for infectious disease dynamics

What do we mean by “best fit the data”?



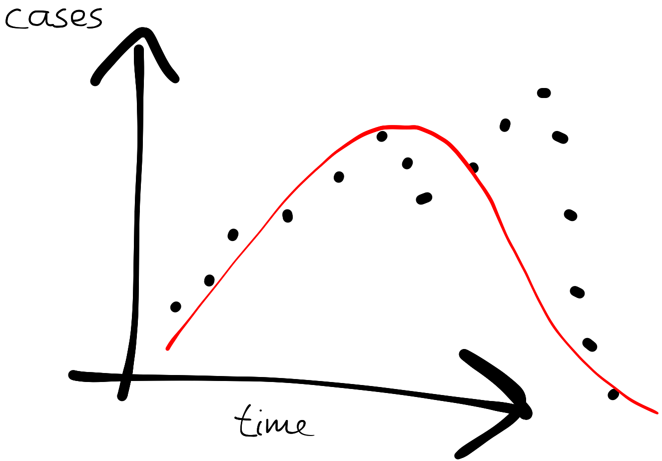
Model fitting and inference for infectious disease dynamics

What do we mean by “best fit the data”?



Model fitting and inference for infectious disease dynamics

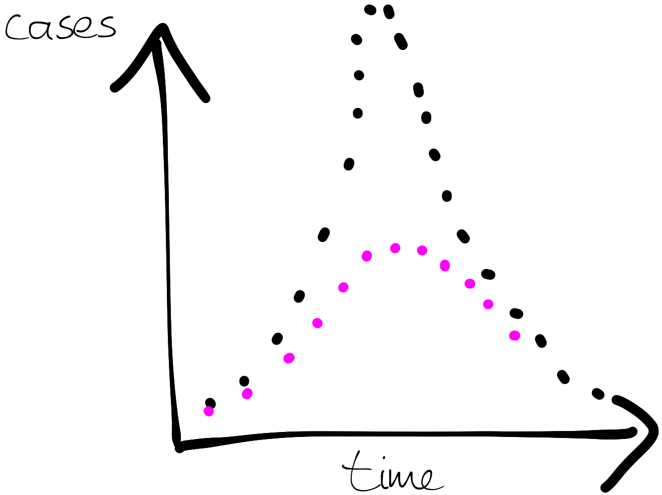
What do we mean by “best fit the data”?



Model fitting and inference for infectious disease dynamics

State estimation

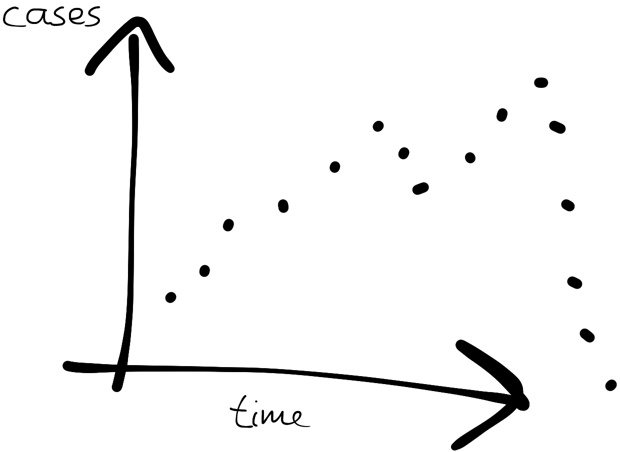
Given what we observe, what is the state of the sytem?



Model fitting and inference for infectious disease dynamics

Model selection

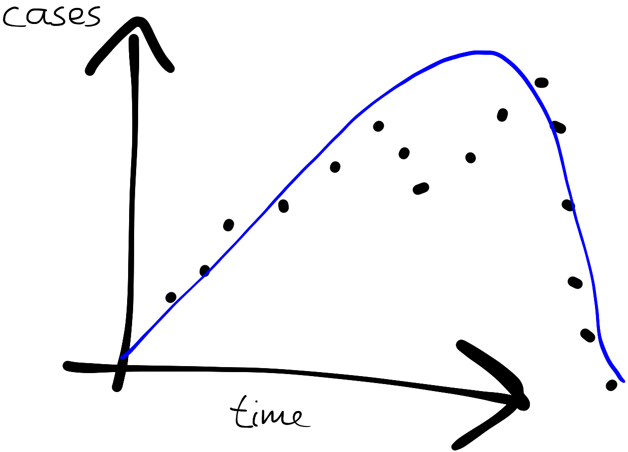
Given a set of potential models, how do we decide which is the right one?



Model fitting and inference for infectious disease dynamics

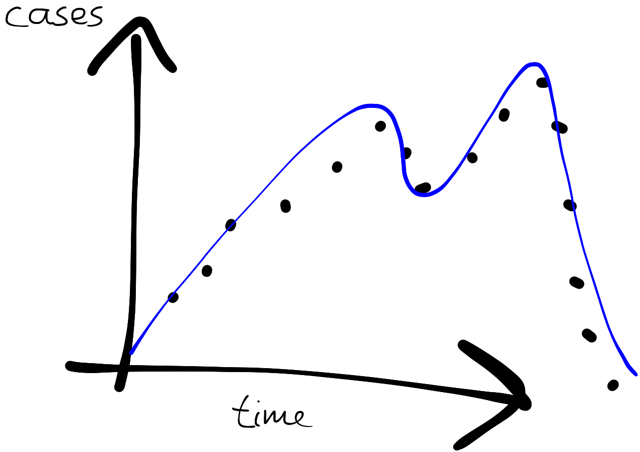
Model selection

Given a set of potential models, how do we decide which is the right one?



Model selection

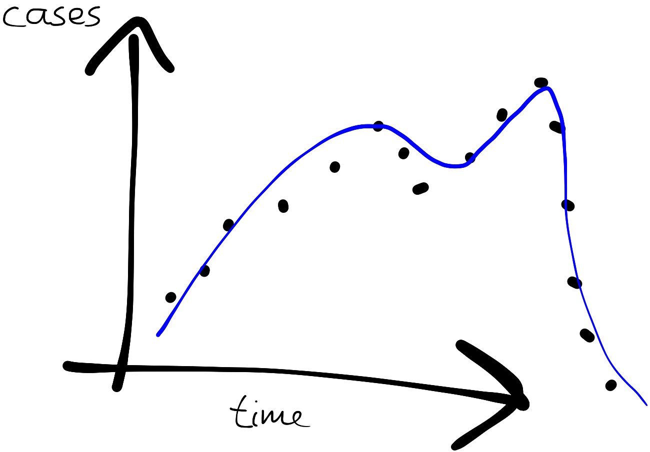
Given a set of potential models, how do we decide which is the right one?



Model fitting and inference for infectious disease dynamics

Model selection

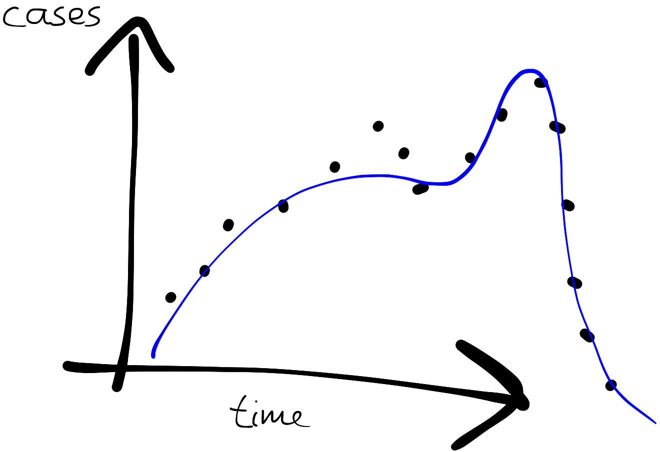
Given a set of potential models, how do we decide which is the right one?



Model fitting and inference for infectious disease dynamics

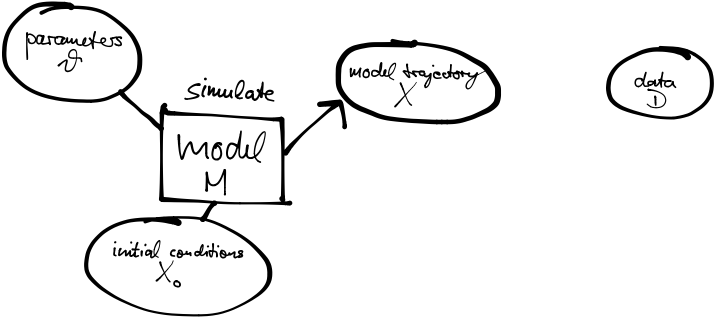
Model selection

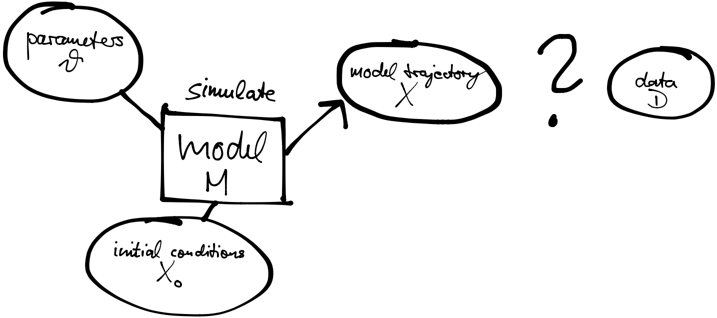
Given a set of potential models, how do we decide which is the right one?



2. Linking models to data


Model
M
data
D


parameters
ϑ
simulate
Model
M
model trajectory
X
data
D
initial conditions
X_0


parameters
ϑ
simulate
Model
M
model trajectory
X
?
data
D
initial conditions
X_0

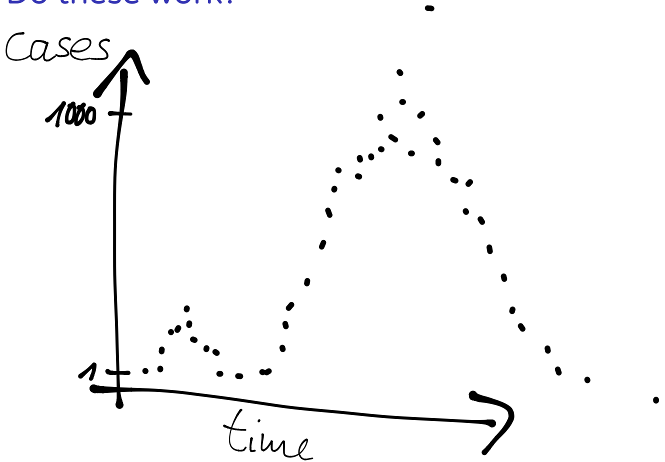
Assessing the "closeness" of model output and data

- eyeballing
- absolute distance
- squared distance

Assessing the “closeness” of model output and data

- eyeballing
- absolute distance
- squared distance

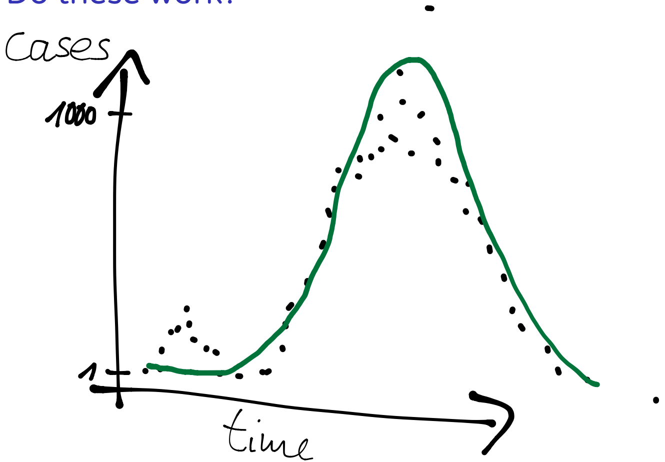
Do these work?



Assessing the “closeness” of model output and data

- eyeballing
- absolute distance
- squared distance

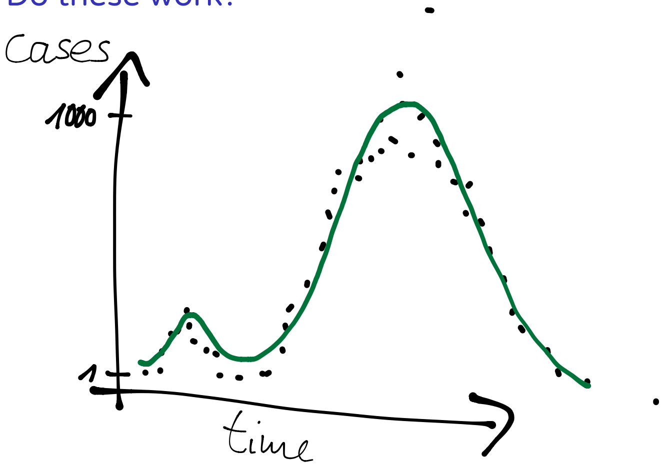
Do these work?



Assessing the "closeness" of model output and data

- eyeballing
- absolute distance
- squared distance

Do these work?



Probabilistic formulation

- Often we know something about how the data were taken $\rightarrow$ observations introduce uncertainty
- We can express the uncertainty in observing the process as a probability

$$p(\text{data}|\text{underlying process})$$

- By including this in our model, we get

$$p(\text{data}|\text{model output})$$

Interlude: probabilities I

Probability theory is nothing but common sense reduced to calculation.

Laplace, 1812

- If A is a random variable, we write

$$p(A = a)$$

for the probability that A takes value a.

- We often write

$$p(A = a) = p(a)$$

- Example: The probability that Andy Murray wins Wimbledon this year

$$p(\mathrm{T} = \mathrm{Andy}) = p(\mathrm{Andy})$$

- Normalisation

$$\sum_a p(a) = 1$$

Interlude: probabilities II

- If A and B are random variables, we write

$$p(A = a, B = b) = p(a, b)$$

  for the joint probability that A takes value a and B takes value b

- Example: The probability that Andy Murray wins Wimbledon and it is sunny final day

$$p(\mathrm{T} = \mathrm{Andy}, \mathrm{W} = \mathrm{sunny}) = p(\mathrm{Andy}, \mathrm{sunny})$$

- We can obtain a marginal probability from joint probabilities by summing

$$p(a) = \sum_b p(a, b)$$

Interlude: probabilities III

- The conditional probability of getting outcome a from random variable A, given that the outcome of random variable B was b, is written as

$$p(A = a|B = b) = p(a|b)$$

- Example: the probability that Andy Murray wins Wimbledon, given that it is sunny on final day

$$p(T = \text{Andy}|W = \text{sunny}) = p(\text{Andy}|\text{sunny})$$

- Conditional probabilities are related to joint probabilities as

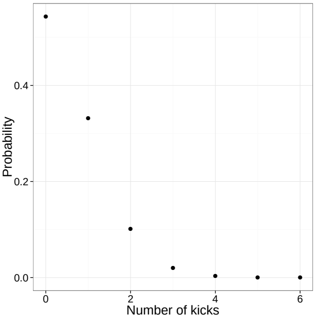
$$p(a|b) = \frac{p(a, b)}{p(b)}$$

- We can combine conditional probabilities in the chain rule

$$p(a, b, c) = p(a|b, c)p(b|c)p(c)$$

Probability distributions (discrete)

- E.g., how many people die of horse kicks if there are 0.61 kicks per year
- Described by the Poisson distribution

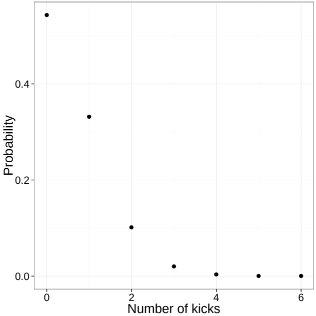


Two directions

1. Evaluate the probability
2. Randomly sample

Evaluating under the (Poisson) probability distribution

- E.g., how many people die of horse kicks if there are 0.61 kicks per year
- Described by the Poisson distribution



Evaluate

What is the probability of 2 deaths in a year?

```
dpois(x = 2,
      lambda = 0.61)
```

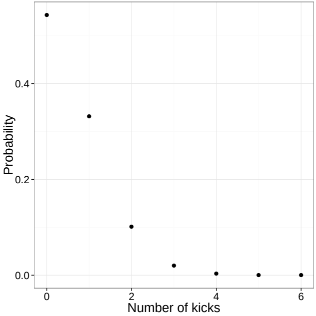
```
[1] 0.1010904
```

Two directions

1. Evaluate the probability
2. Randomly sample

Generating a random sample (Poisson distribution)

- E.g., how many people die of horse kicks if there are 0.61 kicks per year
- Described by the Poisson distribution



Sample

Give me a random sample from the probability distribution

```
rpois(n = 1,
      lambda = 0.61)
```

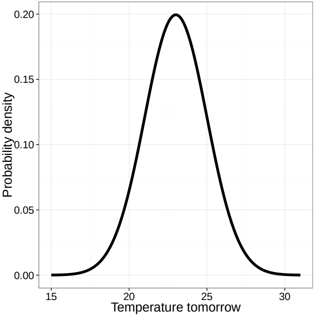
```
[1] 0
```

Two directions

1. Evaluate the probability
2. Randomly sample

Probability distributions (continuous)

- Extension of probabilities to continuous variables
- E.g., the temperature in London tomorrow



Normalisation:

$$\int p(a)\,da = 1$$

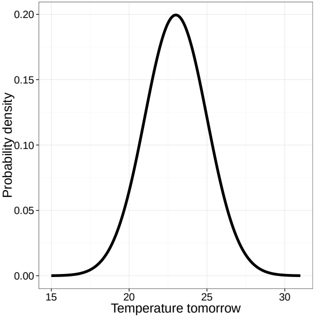
Marginal probabilities:

$$p(a) = \int p(a, b)\,db$$

Two directions

1. Evaluate the probability (density)
2. Randomly sample

Evaluating under the (normal) probability distribution



Evaluate

What is the probability density of $30°C$ tomorrow?

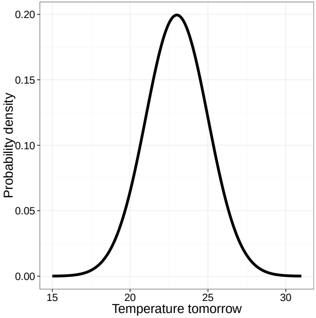
```
dnorm(x = 30,
      mean = 23,
      sd = 2)
```

```
[1] 0.0004363413
```

Two directions

1. Evaluate the probability (density)
2. Randomly sample

Generating a random sample (normal distribution)



Sample

Give me a random sample from the probability distribution

```
rnorm(n = 1,
      mean = 23,
      sd = 2)
```

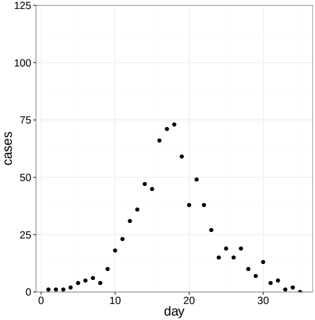
```
[1] 23.05824
```

Two directions

1. Evaluate the probability (density)
2. Randomly sample

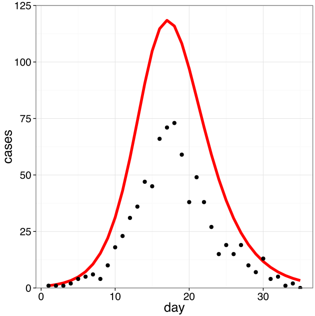
Example: observation uncertainty

SIR model, assume that cases are detected with independent reporting probability $\rho = 0.5$.



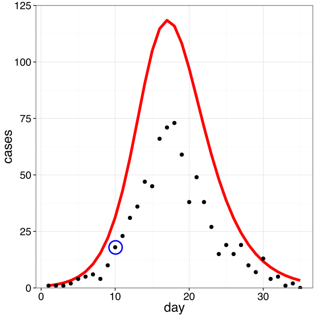
Example: observation uncertainty

SIR model, assume that cases are detected with independent reporting probability $\rho = 0.5$.



Example: observation uncertainty

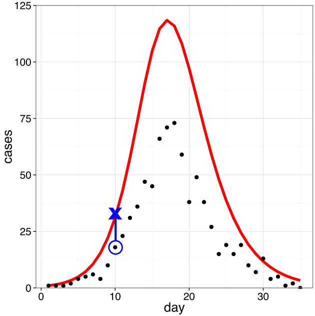
SIR model, assume that cases are detected with independent reporting probability $\rho = 0.5$.



At time 10, 18 cases observed.

Example: observation uncertainty

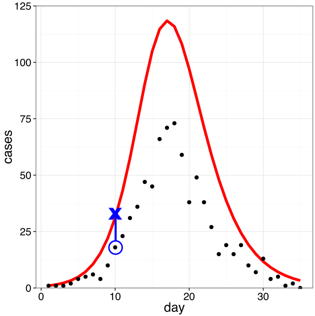
SIR model, assume that cases are detected with independent reporting probability $\rho = 0.5$.

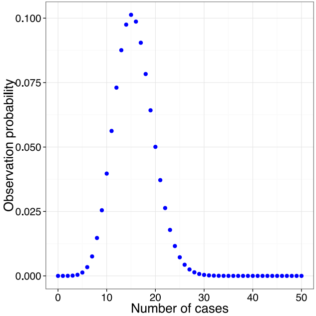


At time 10, 18 cases observed, 31.1 cases in the model.

Example: observation uncertainty

SIR model, assume that cases are detected with independent reporting probability $\rho = 0.5$.

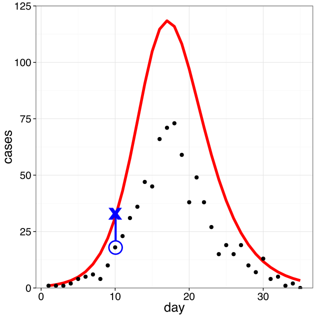


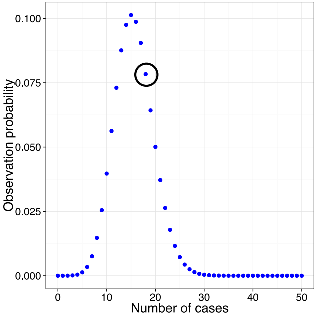


At time 10, 18 cases observed, 31.1 cases in the model.

Example: observation uncertainty

SIR model, assume that cases are detected with independent reporting probability $\rho = 0.5$.

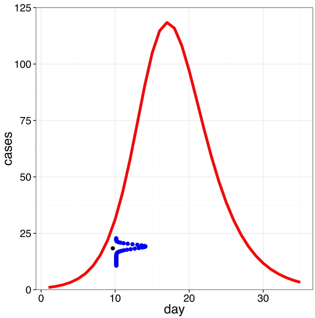


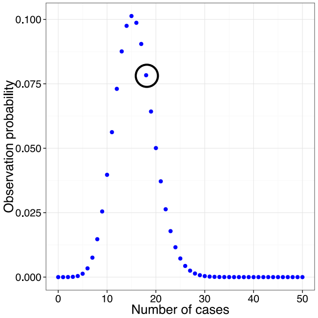


At time 10, 18 cases observed, 31.1 cases in the model.

Example: observation uncertainty

SIR model, assume that cases are detected with independent reporting probability $\rho = 0.5$.



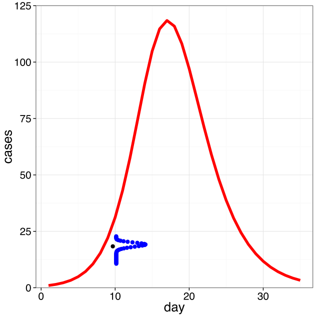


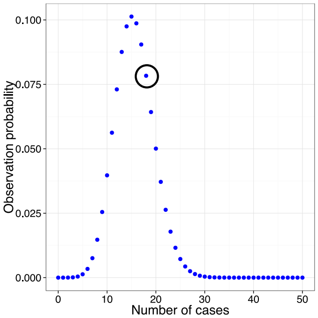
At time 10, 18 cases observed, 31.1 cases in the model.

$$p(\text{data point } 10|\theta) = 0.078$$

Example: observation uncertainty

SIR model, assume that cases are detected with independent reporting probability $\rho = 0.5$.





At time 10, 18 cases observed, 31.1 cases in the model.

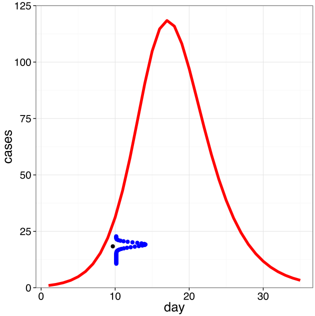
$$p(\text{data point } 10|\theta) = 0.078$$

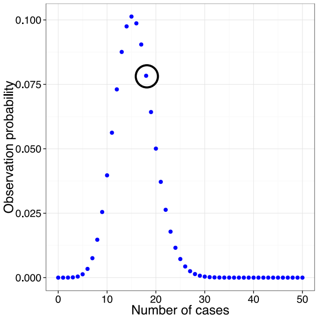
Multiply across the data to get the full trajectory likelihood.

$$p(\text{data}|\theta) = \prod_i p(\text{data point } i|\theta)$$

Example: observation uncertainty

SIR model, assume that cases are detected with independent reporting probability $\rho = 0.5$.



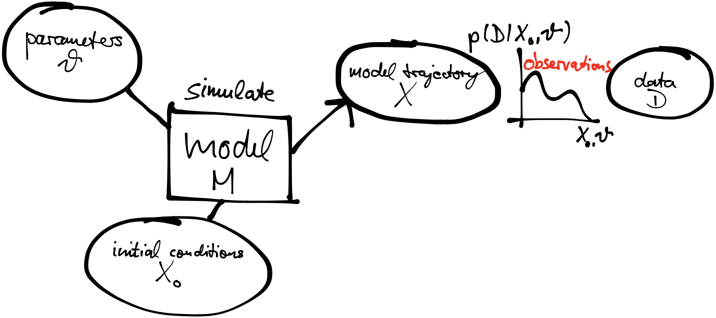


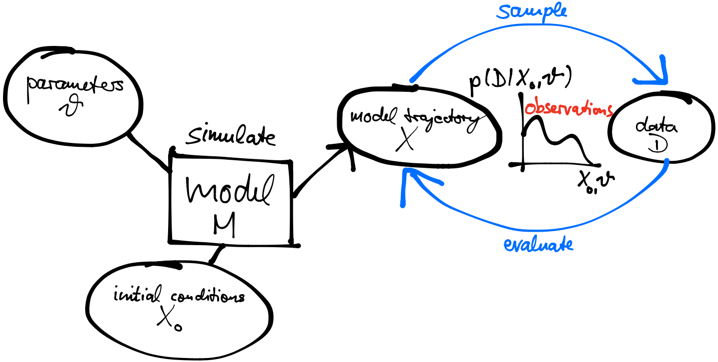
At time 10, 18 cases observed, 31.1 cases in the model.

$$p(\text{data point } 10|\theta) = 0.078$$

Sum across the data to get the full trajectory log-likelihood.

$$\log(p(\text{data}|\theta)) = \sum_i \log(p(\text{data point } i|\theta))$$


parameters
ϑ
simulate
Model
M
initial conditions
X_0
model trajectory
X
p(D|X_0,ϑ)
observations
X_0,ϑ
data
D


parameters
ϑ
simulate
Model
M
initial conditions
X_0
model trajectory
X
sample
$p(D|X_0,\vartheta)$
observations
X_0, ϑ
data
D
evaluate

The likelihood

- We have argued that it makes sense to write

$$p(\text{data}|\text{model output})$$

- For a given model the output depends on the parameters θ. So we can write

$$p(\text{data}|\theta)$$

(note: θ encompasses all parameters; e.g., $\theta = \{\beta, \gamma\}$)

- This is called the likelihood of parameters θ
- likelihoods can span a wide range of orders of magnitude, which can lead to numerical problems

Solution: take the logarithm to get the log-likelihood

$$\log p(\text{data}|\theta) = \sum_i \log p(\text{data point } i|\theta)$$

Frequentist vs Bayesian inference

Frequentist inference:

- there are ***true*** parameters in the world, the uncertainty comes from the data
- this is encoded in the likelihood: $p(\mathrm{data}|\theta)$
- in inference, I try to estimate these parameters
- probabilities express outcomes of repeated experiments

Frequentist vs Bayesian inference

Frequentist inference:

- there are ***true*** parameters in the world, the uncertainty comes from the data
- this is encoded in the likelihood: $p(\mathrm{data}|\theta)$
- in inference, I try to estimate these parameters
- probabilities express outcomes of repeated experiments

Bayesian inference

- there are no true parameters, the ***data*** are true; uncertainty is in parameters / hypotheses
- this is encoded in the posterior: $p(\theta|\mathrm{data})$
- probabilities express my belief in a given parameter
- the posterior is interpreted as the ***probability distribution*** of a ***random variable*** θ

3. Bayesian inference

Bayes' rule

- We said that in Bayesian inference, we need to calculate $p(\theta|\mathrm{data})$. Applying the rule of conditional probabilities, we can write this as

$$p(\theta|\mathrm{data}) = \frac{p(\mathrm{data}|\theta)p(\theta)}{p(\mathrm{data})}$$

- $p(\theta|\mathrm{data})$ is the ***posterior***
- $p(\mathrm{data}|\theta)$ is the ***likelihood***
- $p(\theta)$ is the ***prior***
- $p(\mathrm{data})$ is a ***normalisation constant***
- In words,

$$(\mathrm{posterior}) \propto (\mathrm{normalised\ likelihood}) \times (\mathrm{prior})$$

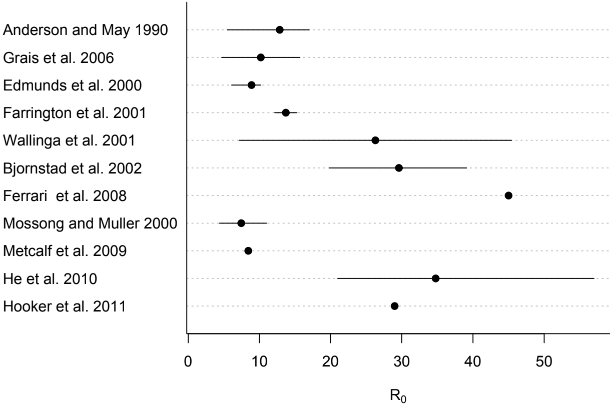
Prior probabilities

- $p(\theta)$ quantifies our degree of belief via a probability distribution before confronting the model with data:

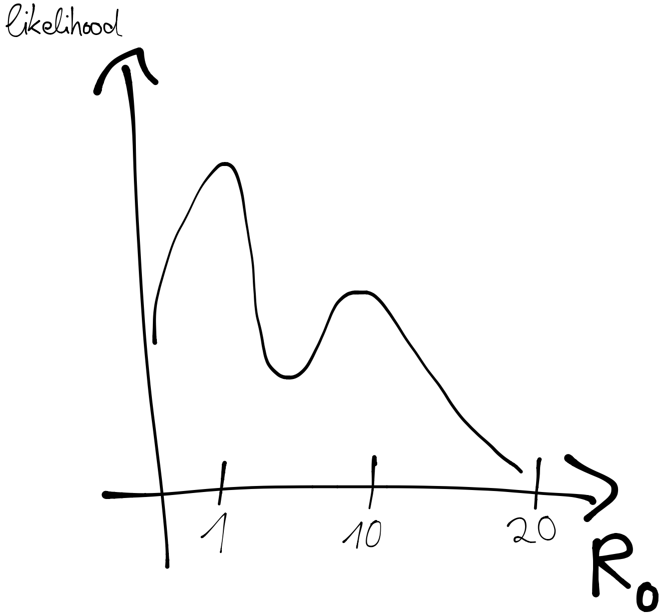
$$p(\theta)$$

E.g., from previous measurements, literature, experts etc.

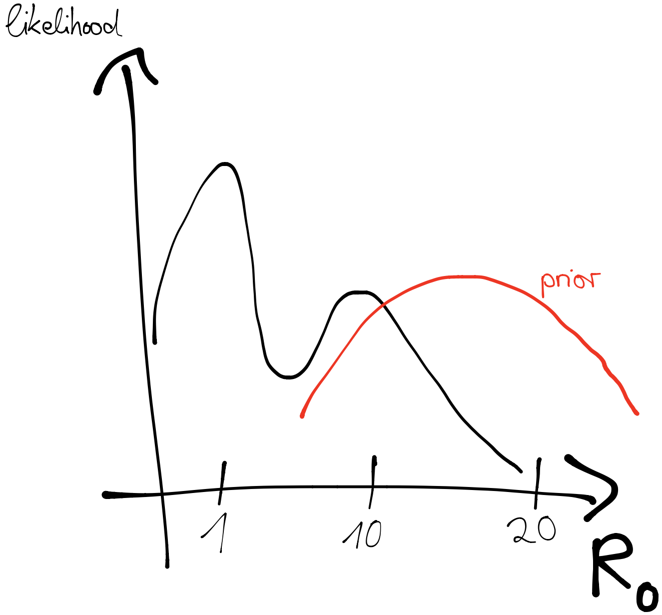
- Example: R_0 of measles



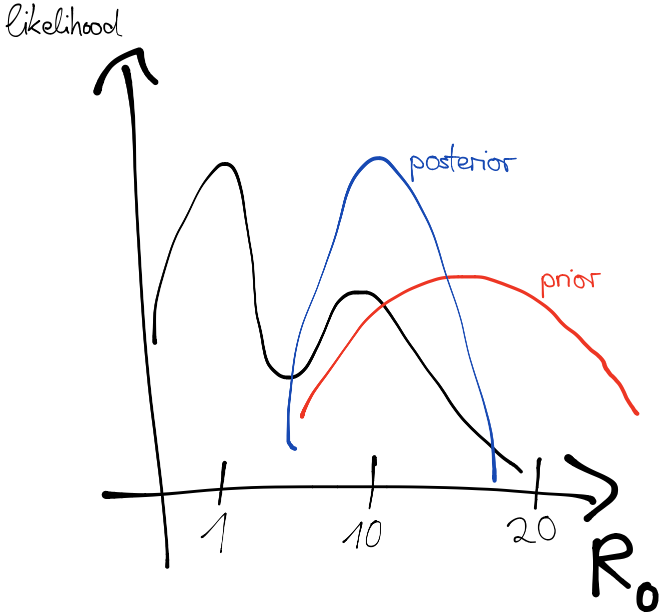
Example: estimating R_0 of measles



Example: prior for estimating R_0 of measles



Example: posterior for estimating R_0 of measles



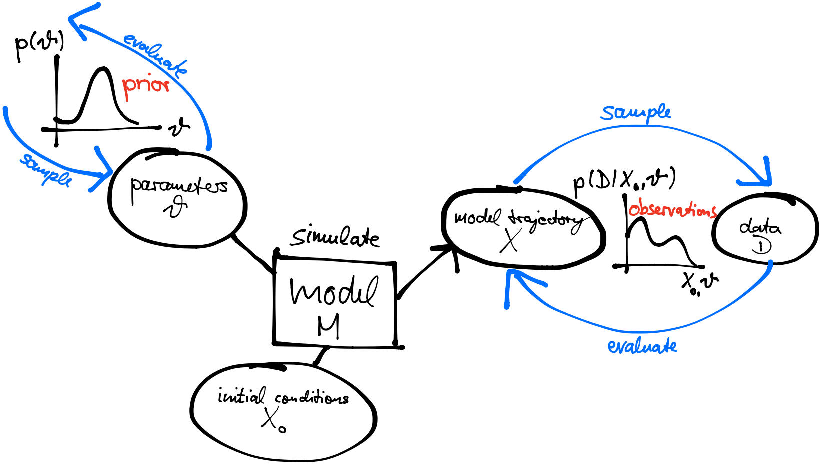
Sampling from the posterior

Bayesian statistics

Parameter(s) θ are interpreted as a ***random*** variable, distributed according to the posterior.

$$p(\theta|\text{data}) \propto p(\text{data}|\theta)p(\theta)$$

We want to generate samples of θ from this distribution.



p(ϑ)
prior
evaluate
sample
parameters ϑ
simulate
Model M
initial conditions X₀
model trajectory X
p(D|X₀,ϑ)
observations
X₀,ϑ
data D
sample
evaluate

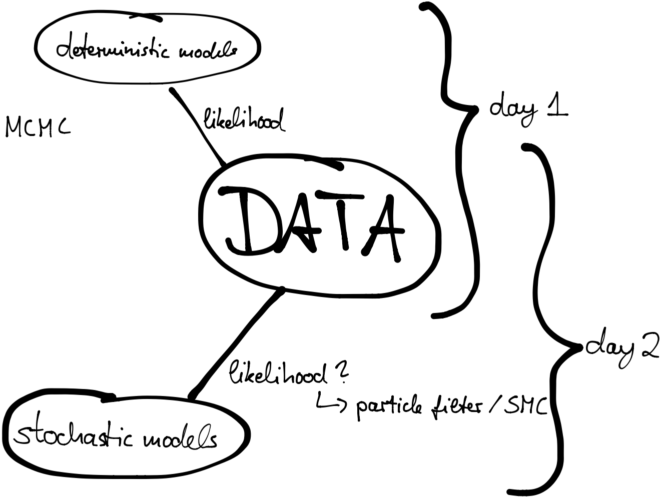
deterministic models

MCMC

likelihood

DATA

day 1


deterministic models
MCMC
likelihood
DATA
likelihood ?
↳ particle filter / SMC
stochastic models
day 1
day 2

deterministic models

MCMC

likelihood

DATA

day 1

day 2/3

MCMC

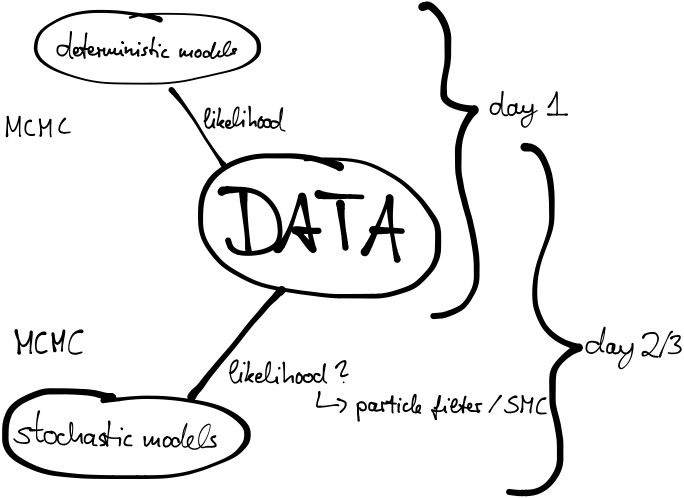
likelihood ?

↳ particle filter / SMC

stochastic models

day 3 (afternoon):

- discussion / open session



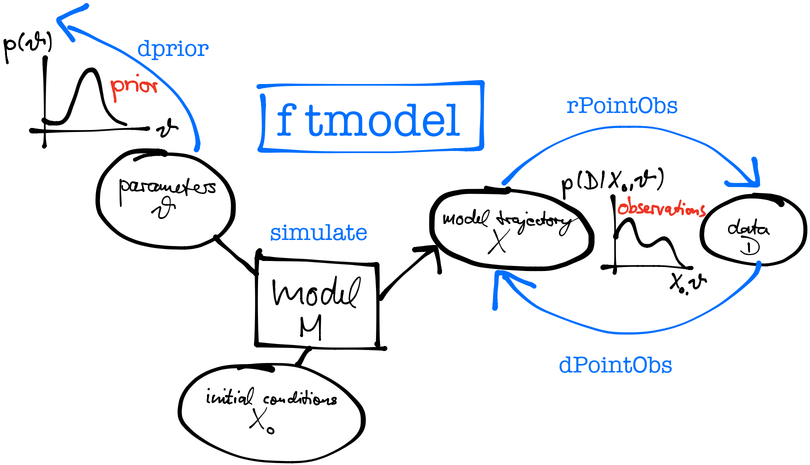
day 3 (afternoon):

- discussion / open session

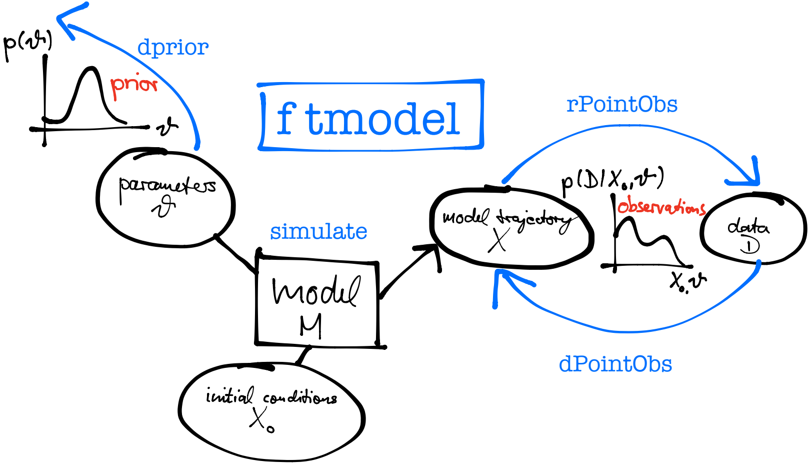
day 4:

- other methods
- available software
- final discussion

4. Practical session in R



fitmodel
dprior
$p(\vartheta)$
prior
ϑ
parameters ϑ
simulate
model M
initial conditions X_0
model trajectory X
rPointObs
$p(D|X_0,\vartheta)$
observations
X_0,ϑ
data D
dPointObs



http://sbfnk.github.io/mfiidd