Model fitting and inference for infectious disease dynamics

Sebastian Funk

June 10, 2014

Contents

1 Introduction

1.1 Model fitting and inference for infectious disease dynamics

1.1.1 Model

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

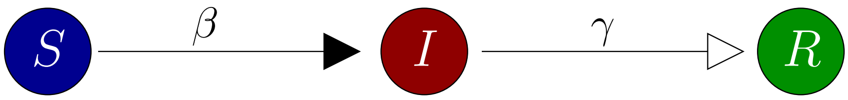
Oxford English Dictionary

1.1.2 Mathematical model

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

1.2 Model fitting and inference for infectious disease dynamics

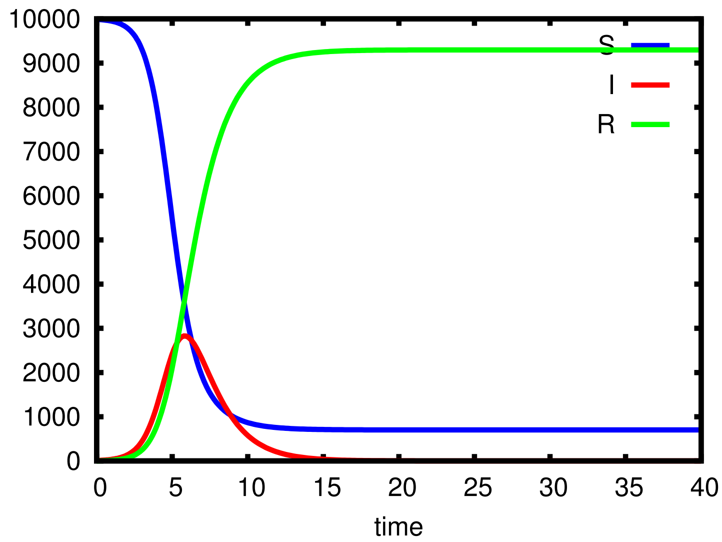
1.2.1 SIR-type models



1. Equations

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

2. Example output



1.2.2 Mechanistic models

description vs **mechanism**

1.3 The difference between statistic and mathematical models

A distinction is often made between so-called *mathematical* (mechanistic, dynamic) and *statistical* (or phenomenological) models. As the labels imply, the former are usually concerned with describing the mechanisms that generate a certain pattern over time, usually based on differential equations or, more generally, simulations, with traditionally no particular concern to uncertainty and noise. The latter, on the other hand, are usually static and descriptive but contain a rigorous treatment of uncertainty.

Quoting Bolker, 2008: "A mathematician is more likely to produce a deterministic, dynamic process model without thinking very much about noise and uncertainty (e.g. the ordinary differential equations that make up the Lotka-Volterra predator prey model). A statistician, on the other hand, is more likely to produce a stochastic but static model, that treats noise and uncertainty carefully but focuses more on static patterns than on the dynamic processes that produce them (e.g. linear regression)."

Recently, there has been growing interest in applying the rigorous methods of statistical models to dynamic models, and this has led to the development of a variety of techniques which we discuss in this course. Through these developments, the traditional and somewhat artificial boundaries between mathematical and statistical models are becoming increasingly blurry.

1.4 Model fitting and inference for infectious disease dynamics

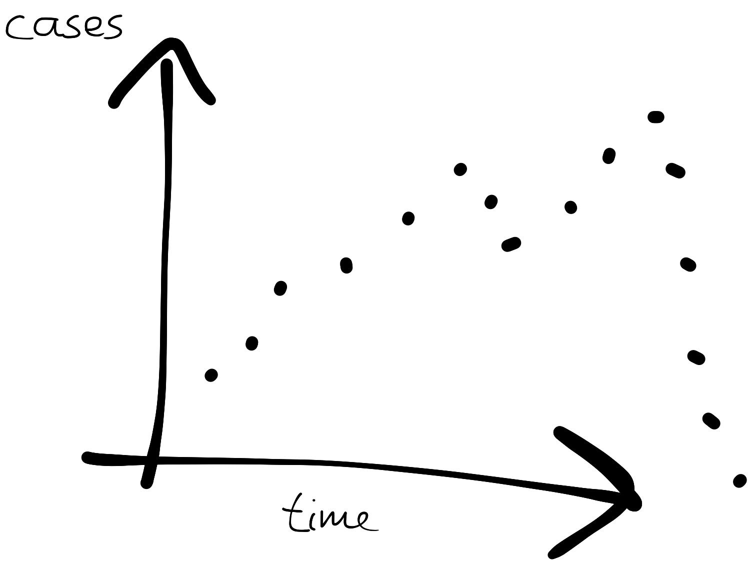
1.4.1 Parameter estimation

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

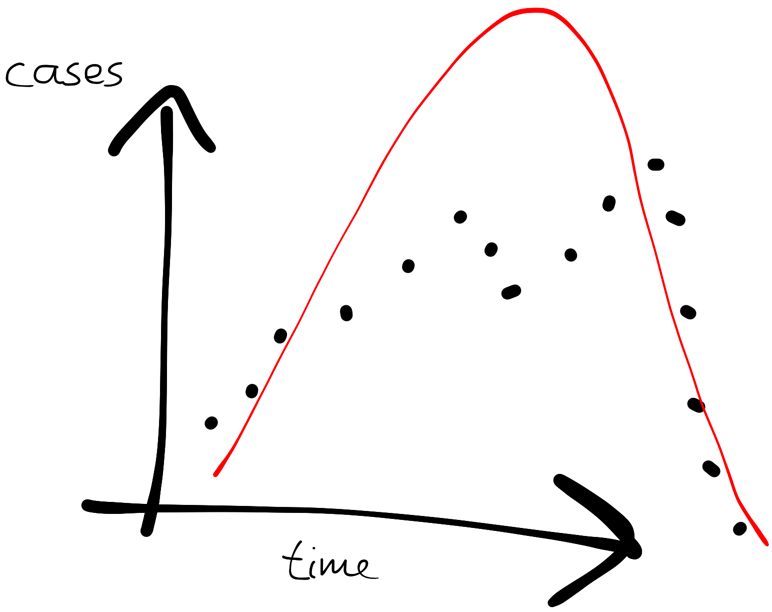
1. Why are we doing this?

    - **Learn** something about the system
        - test a scientific hypothesis
            * e.g., why did the UK H1N1 epidemic wane in summer 2009? (Dureau, Kalogeropoulos, and Baguelin, 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

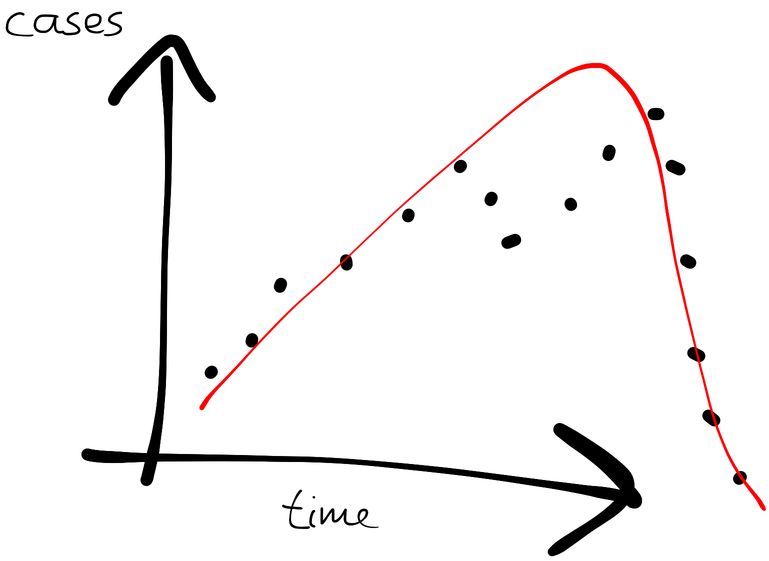
1.4.2 What do we mean by "best fit the data"?



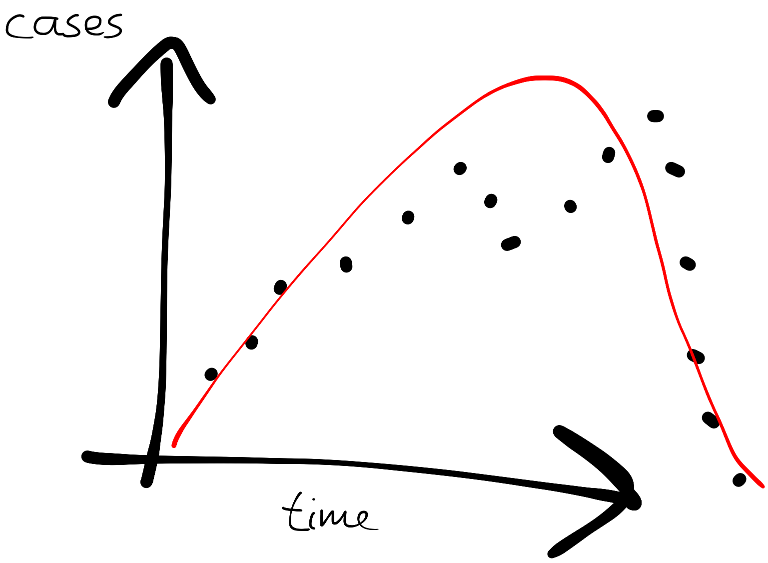
1.



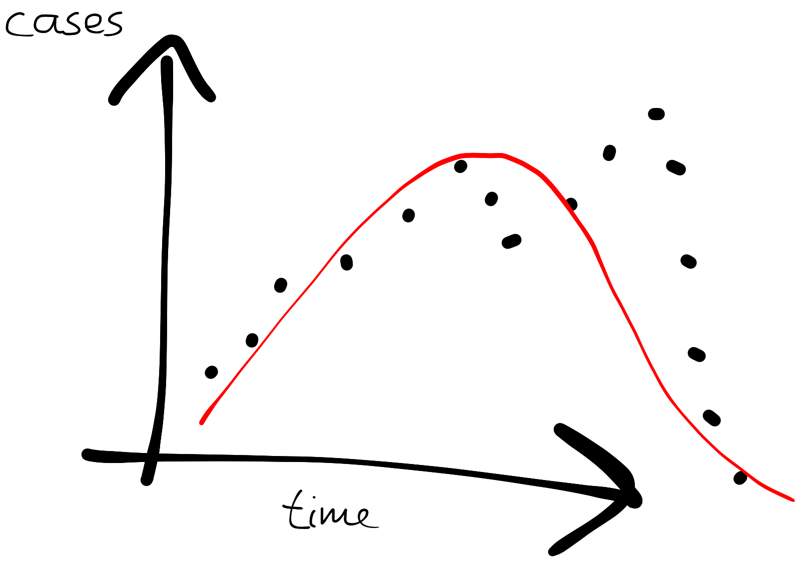
2.



3.

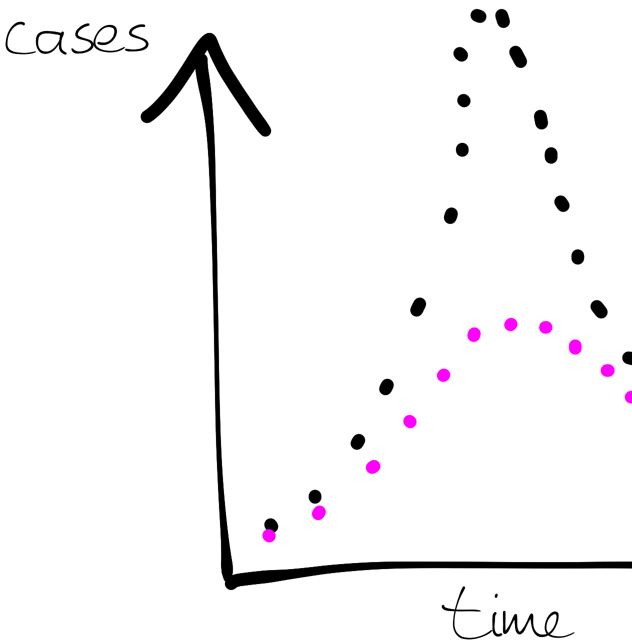


4.

5. 

1.5 Model fitting and inference for infectious disease dynamics

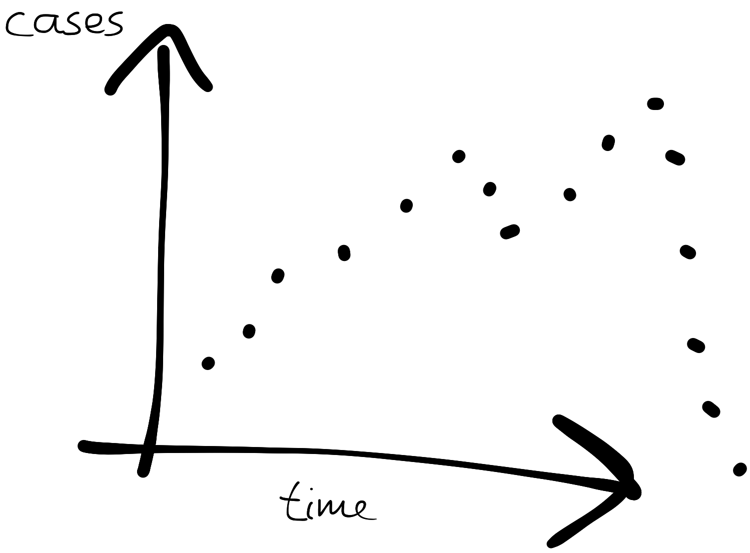
1.5.1 State estimation



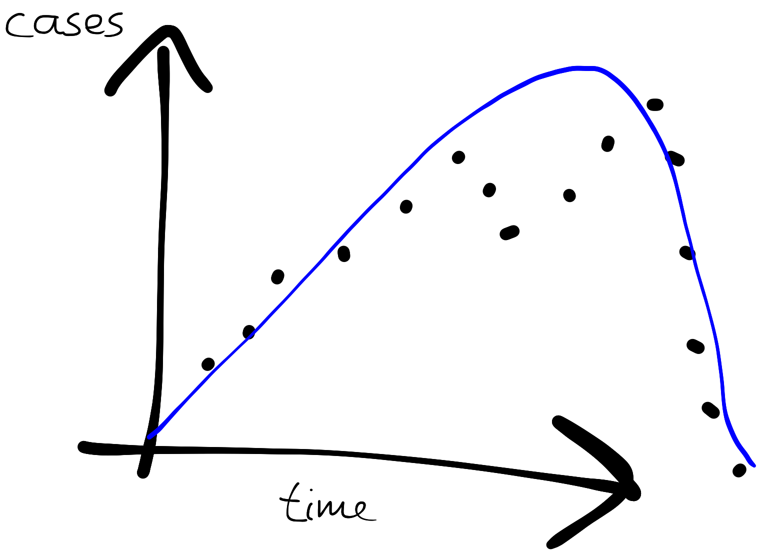
Given what we observe, what is the **state** of the sytem?

1.5.2 Model selection

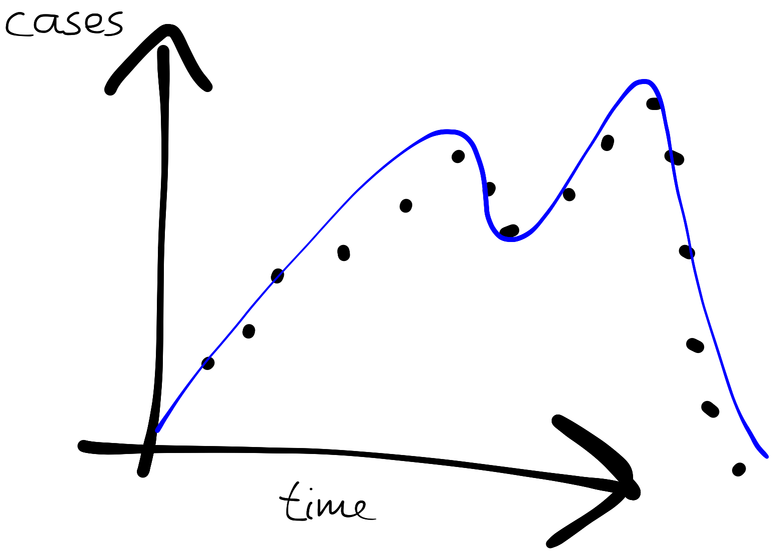
Given a set of potential models, how do we decide which is the right one?



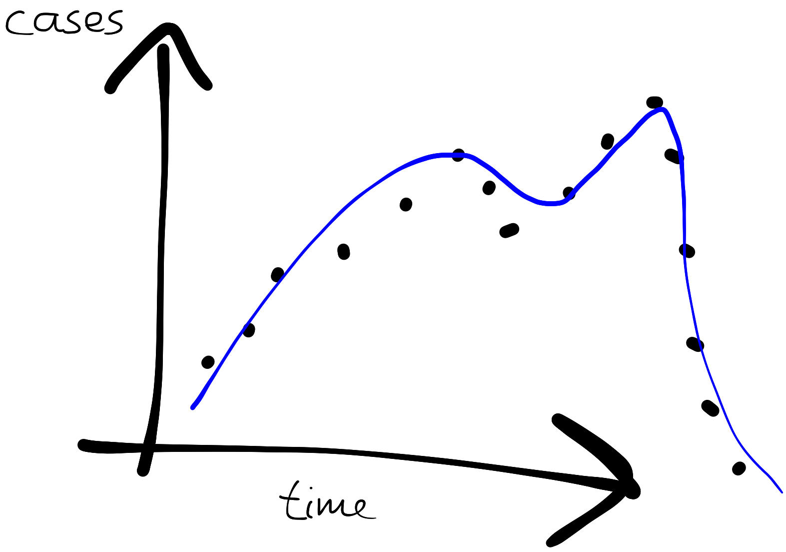
1.



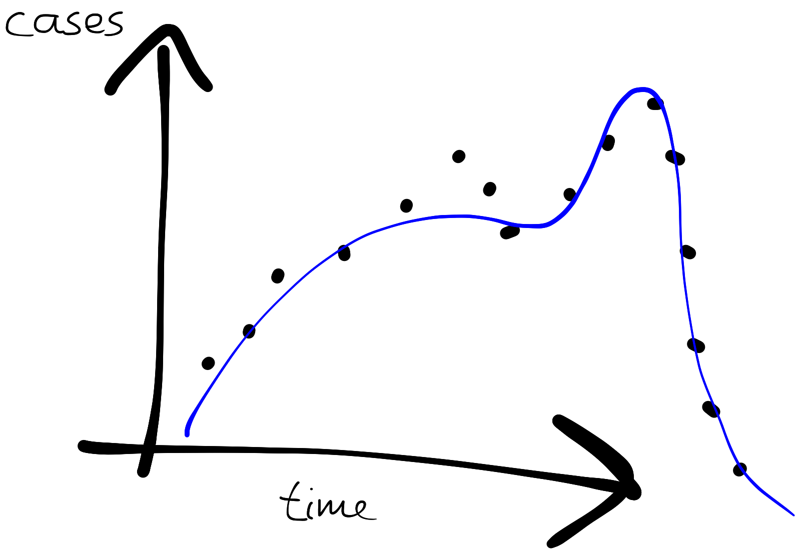
2.



3.



4.

5. 

2 The likelihood

2.1 The need to define a likelihood

Ideally, we would like a model that reproduces (i.e., fits) our data perfectly. However, in reality we are unlikely to be able to do that because

1. we are unlikely to have captured every single aspect that affects the system in question, or to have eliminated every source of disturbance, and
2. measurements/data taking themselves are subject to error.

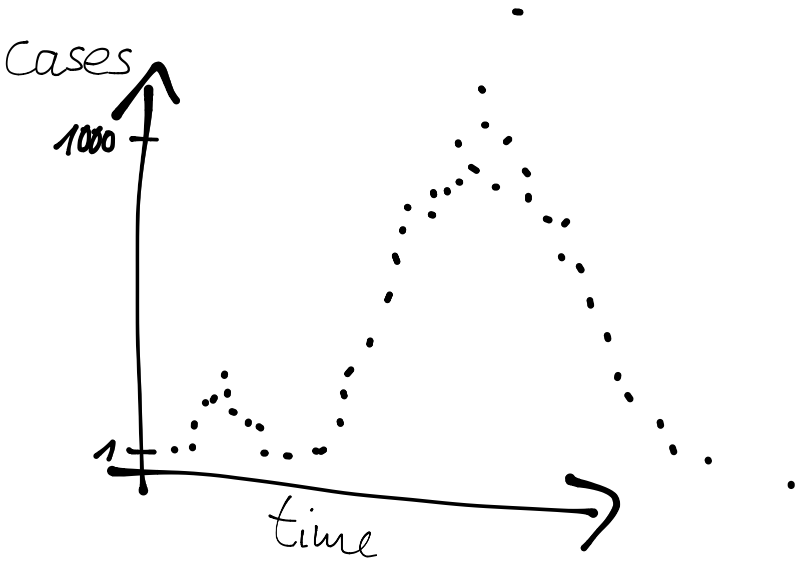
With this in mind, it makes sense to think of relationship between the data and the model to be a *probabilistic* one, i.e. given a core model (which describes the system only as a function of the parameters included, and free of measurement error) we attribute each possible measurement outcome a probability of occurring as a function of the model output under a given set of parameters.

2.2 Assessing the "closeness" of model output and data

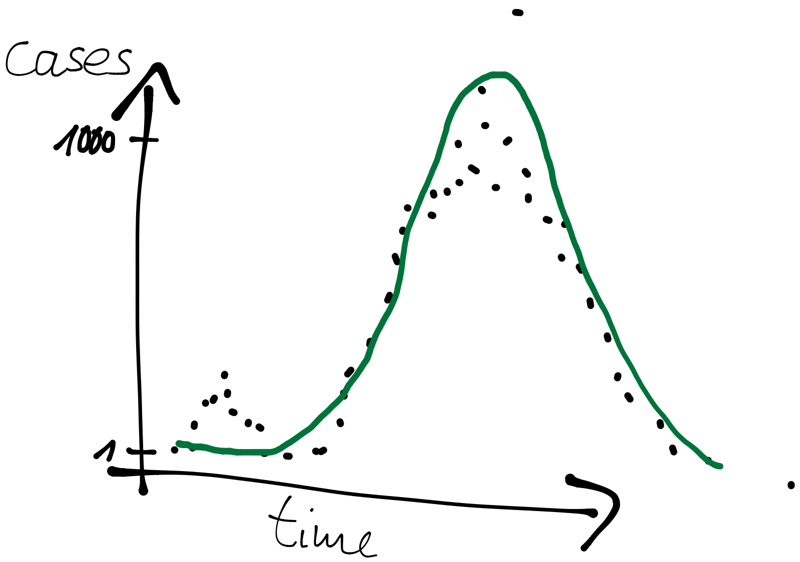
- eyeballing
- absolute distance
- squared distance

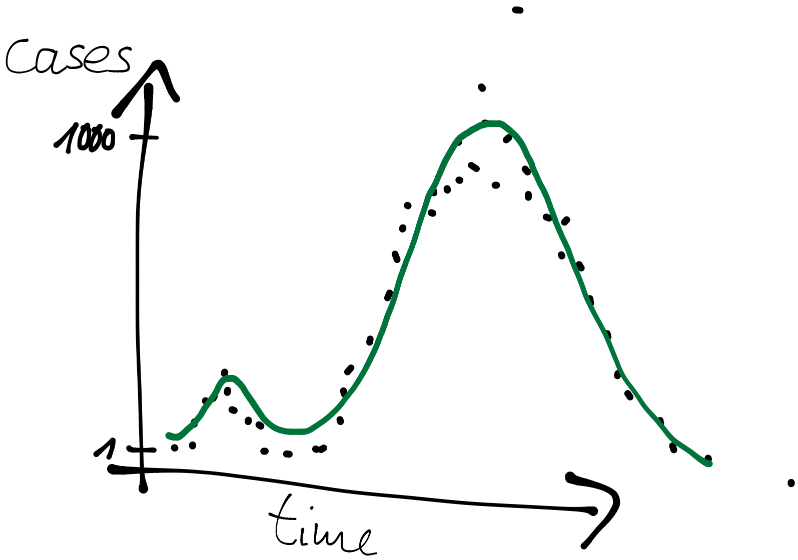
2.2.1 Do these work?

1.



2.



3. 

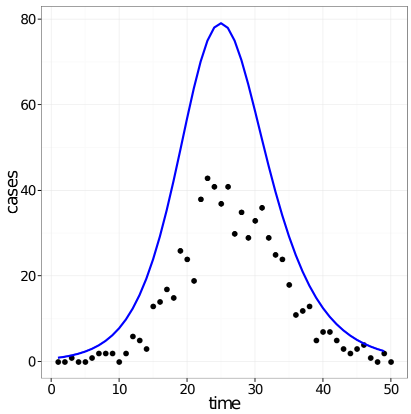
2.3 Probabilistic formulation

- Often we know something about how the data were taken → measurement introduces uncertainty
- We can express the uncertainty in data taking as a probability

$$p(\text{data}|\text{"true" process})$$

- By including this in our model, we get

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



2.4 Interlude: probabilities

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 England win the world cup

$$p(W = \text{England}) = p(\text{England})$$

- Normalisation

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

- 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 England win the world cup and Wayne Rooney is top scorer

$$p(W = \text{England}, TS = \text{Rooney}) = p(\text{England}, \text{Rooney})$$

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

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

- 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 England win the world cup, given that Rooney is top scorer

$$p(W = \text{England}|TS = \text{Rooney}) = p(\text{England}|\text{Rooney})$$

- 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**

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

2.5 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., for the SIR model we could have $\theta = \{\beta, \gamma, N\}$)

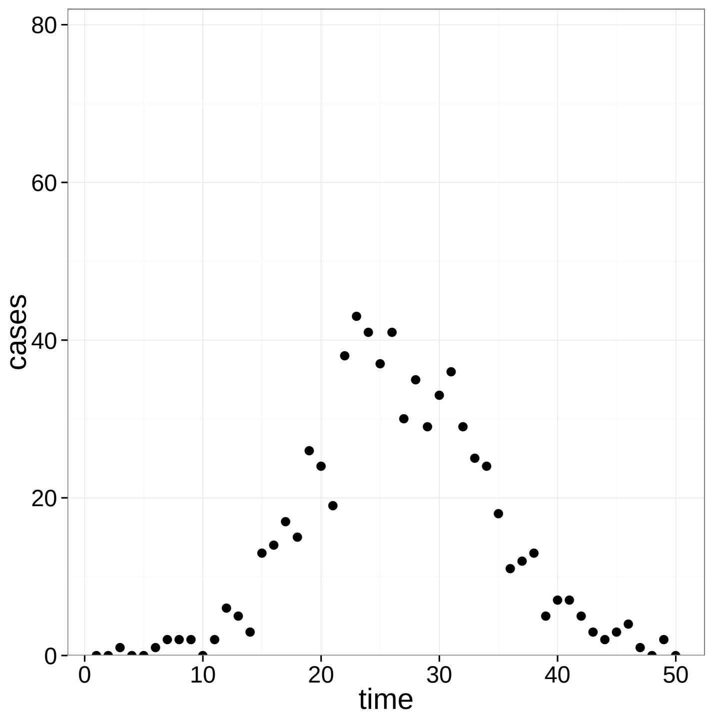
- Inverse problem: We make an observation, what is the probability of θ being a certain value(s).

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

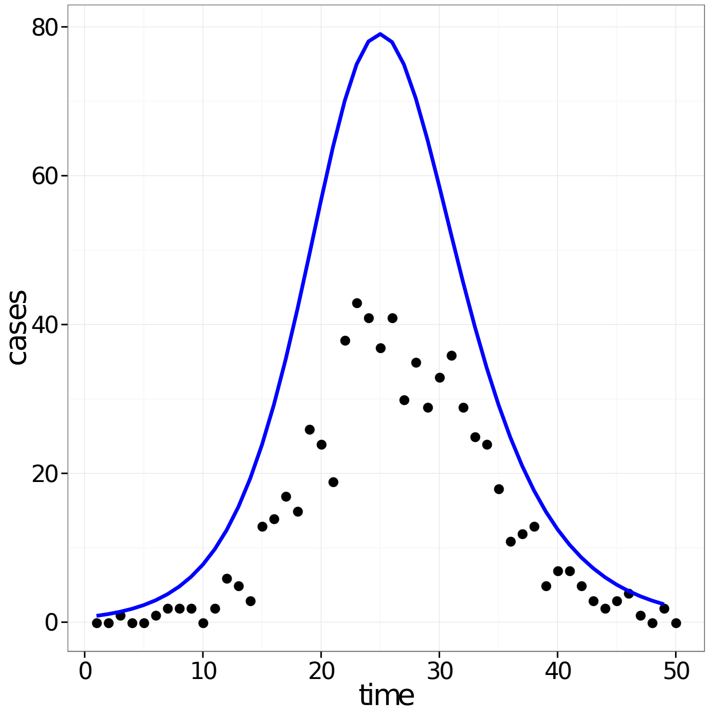
2.6 Example

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

1. Model trajectory

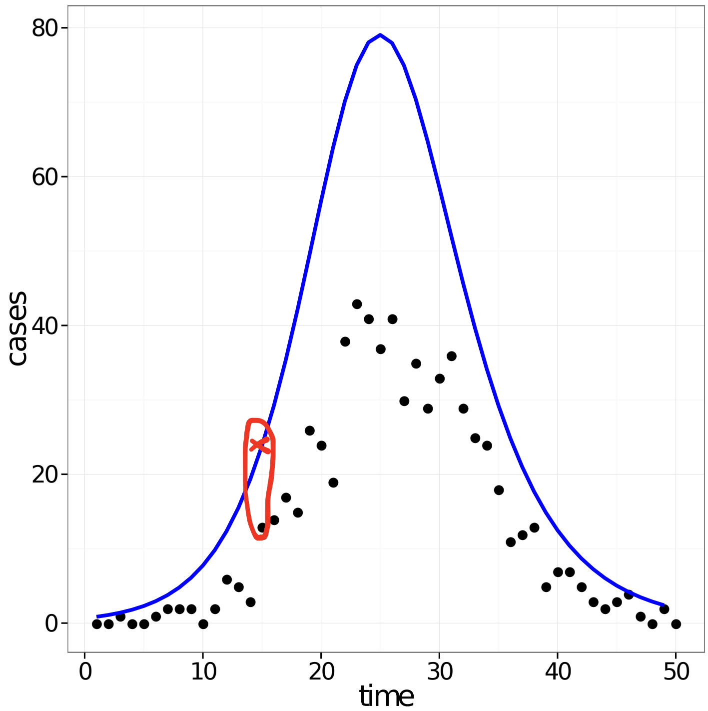


(a)



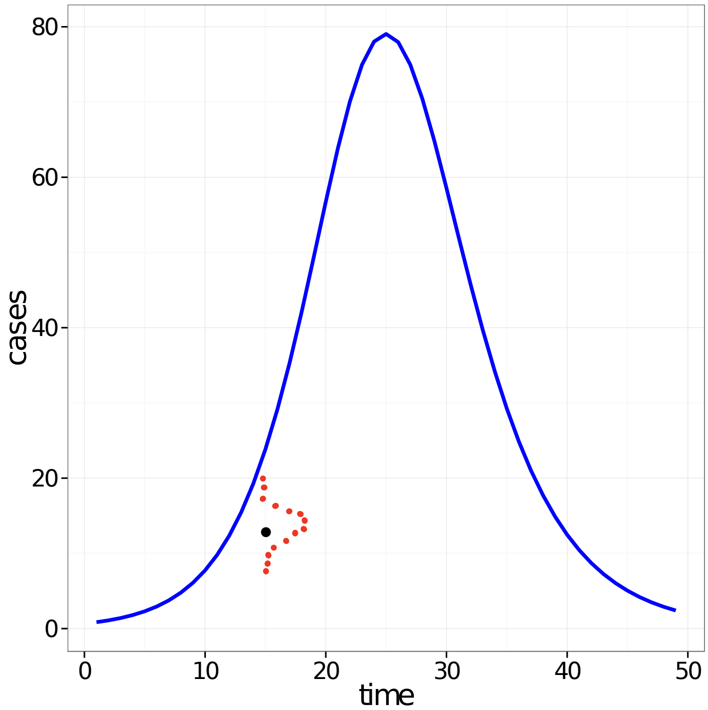
cases
time
0
20
40
60
80
0
10
20
30
40
50

(b)



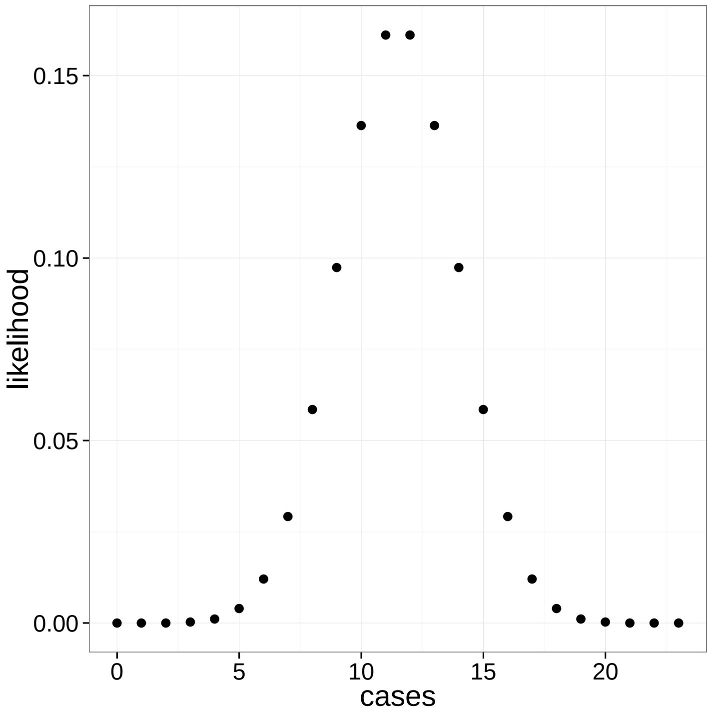
cases
time
0
20
40
60
80
0
10
20
30
40
50

(c)

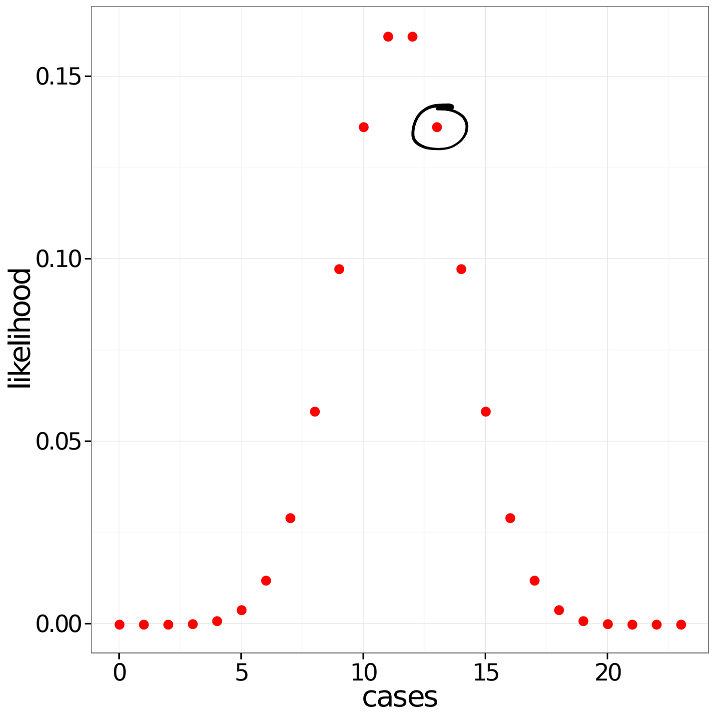


(d)

2. Likelihood



(a)



(b)

2.6.1 At time 15, 24 case in the model, 13 cases observed.

Multiply across the data to get the full observation likelihood.

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

2.7 Properties of the likelihood

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

- Note the **proportionality**
  - we are only ever interested in **comparing** likelihoods
  - likelihoods need not sum to 1
- likelihoods can span a wide range of orders of magnitude, which can lead to numerical problems
  - Solution: take the **logarithm**

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

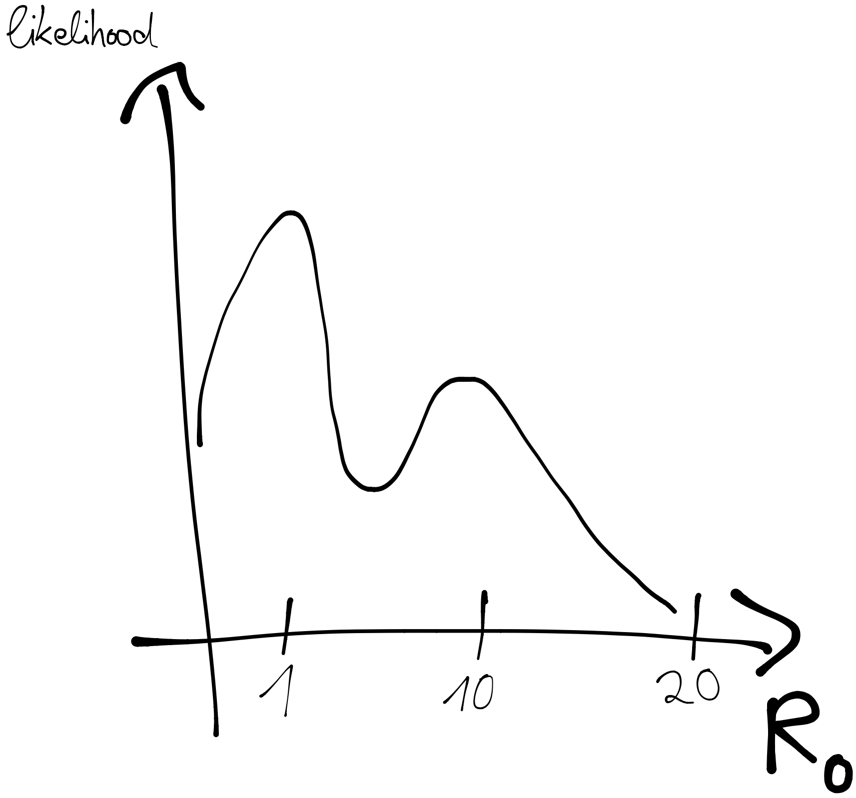
- Maximum likelihood estimation (MLE): find the value of θ that maximises the likelihood.

2.8 Frequentist inference

MLE is based on a *frequentist* interpretation of statistics: a hypothesis is either true or false and, similarly, a parameter model has a *true* value. Any uncertainty in our estimation is due to the fact that the *data* are random (because of stochasticity in the process or measurement) – probabilities represent long-term frequencies of outcomes. In the next section, we will encounter an alternative interpretation, called *Bayesian*.

3 Beyond the likelihood

3.1 Example: estimating R_0 of measles

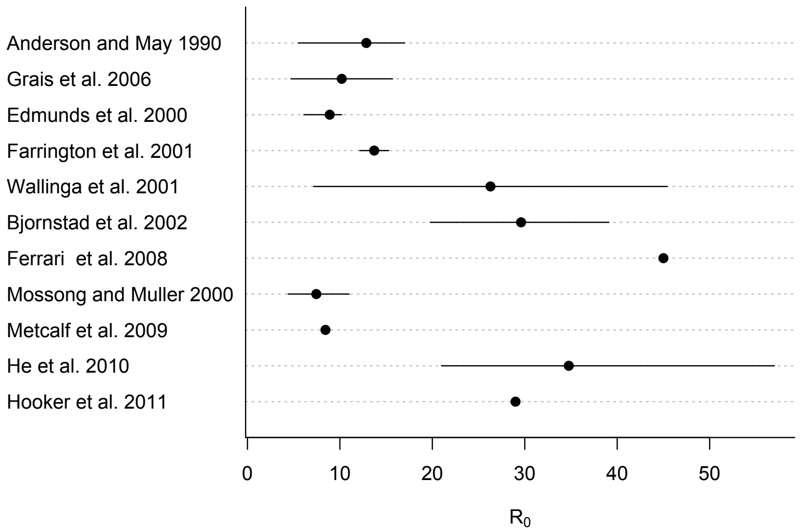


3.2 Prior probabilities

- What if we know something about the parameters *a priori*? E.g., from previous measurements, literature, experts etc.

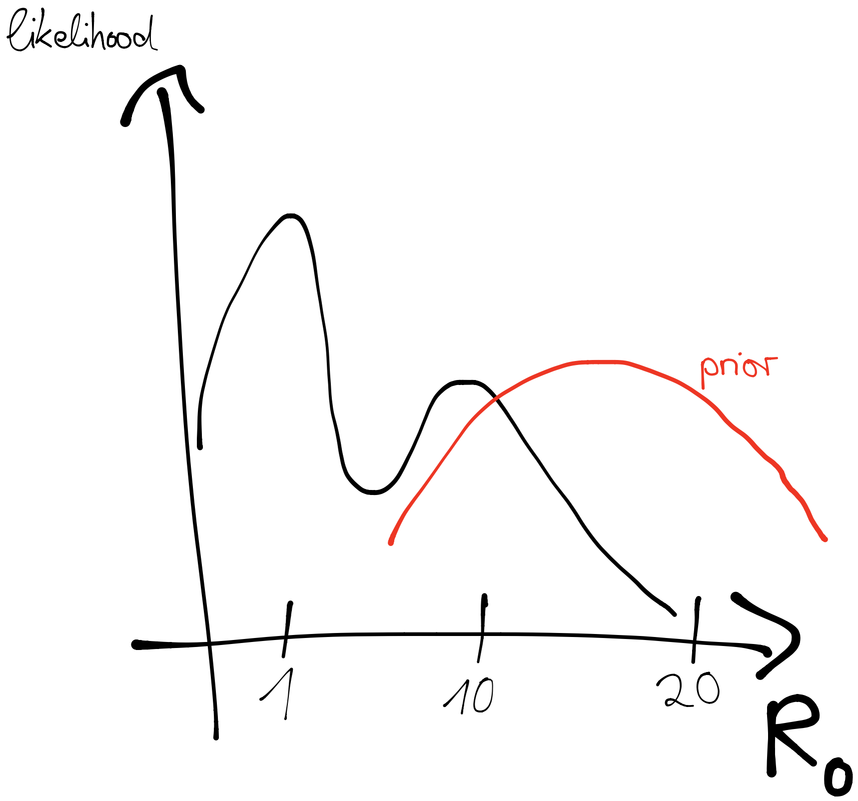
- Quantify our degree of **belief** in a probability distribution before confronting the model with data:

$$p(\theta)$$



- Example: R_0 of measles

3.3 Example: prior for estimating R_0 of measles



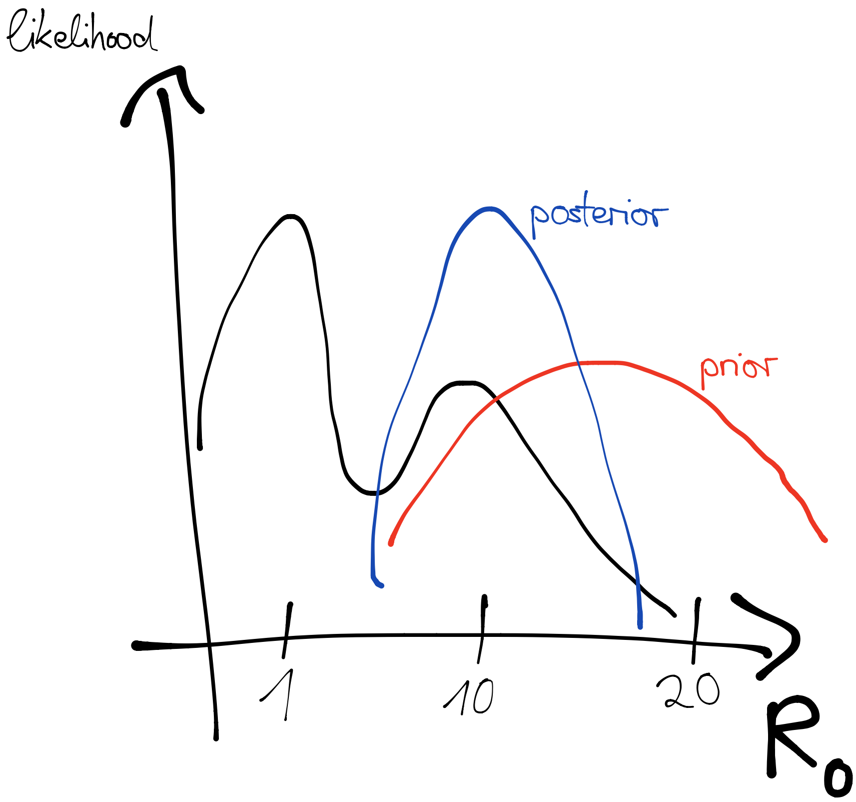
3.4 Bayes' rule

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

- Left hand side: **posterior** probability, or the probability of hypothesis (parameter) θ given the evidence.
- In words,

$$(\text{posterior probability}) \propto (\text{likelihood}) \times (\text{prior probability})$$

3.5 Example: posterior for estimating R_0 of measles



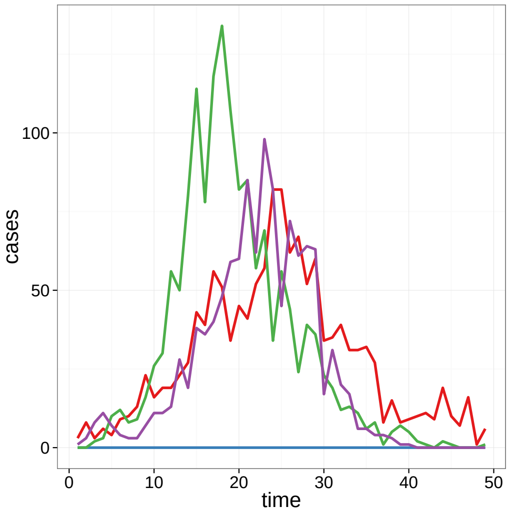
3.6 Bayesian inference

Bayesian inference interprets probabilities to express our degree of *belief* or *certainty* in a given outcome. While in a frequentist interpretation there is a "true" state of the world, Bayesian inference treats the data as the truth instead of a single outcome of a potentially infinite number of experiments. Instead, hypotheses themselves are subject to randomness (i.e., they are true with a certain probability that express our degree of belief in them given the "true" data).

There are many great Statisticians following either a frequentist or Bayesian interpretation of Statistics. A further discussion of these issues is beyond the scope of this course, but can be found elsewhere (e.g., Fisher, 1922; Ellison, 2004; Bolker, 2008; Munroe, 2012). Here, it suffices to realise that the posterior is essentially the likelihood multiplied with the prior.

3.7 Stochastic models

- In a stochastic model, one set of parameters θ can lead to meany different outcomes



- What is $p(\text{data}|\theta)$?
- The model outcome itself occurs with a certain probability:

$$p(\text{data}|\theta) = p(\text{data}|\text{model trajectory})p(\text{model trajectory}|\theta)$$

4 Putting it all together

4.1 What we really want to know

- we want to find out **two** things
  1. the **parameters** of the process

$$\theta$$

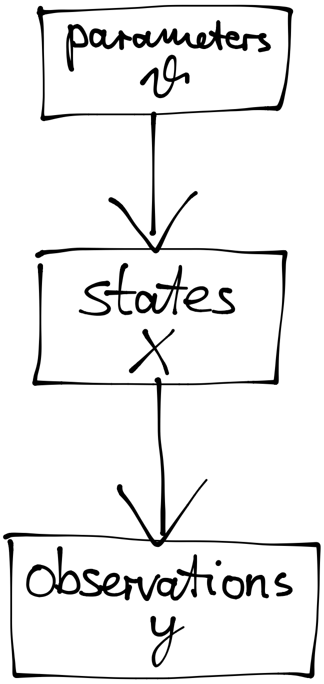
  2. the **state** (as opposed to observed) process

$$x$$

- we want to do this on the basis of **observations** (data)

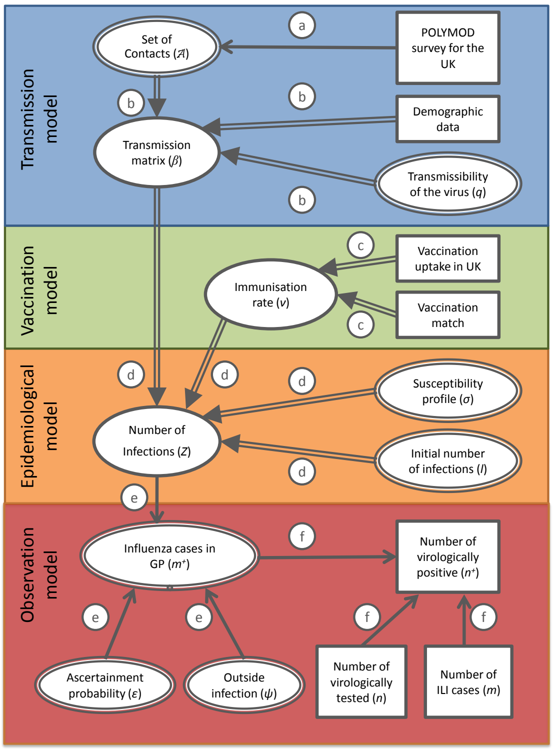
$$y$$

4.2 Hierarchical model



- this kind of model is sometimes called a **Bayesian hierarchical model**

4.3 Example: influenza in the UK



Baguelin et al., 2013

4.4 Combining the probabilities

1. prior probabilities of the parameters

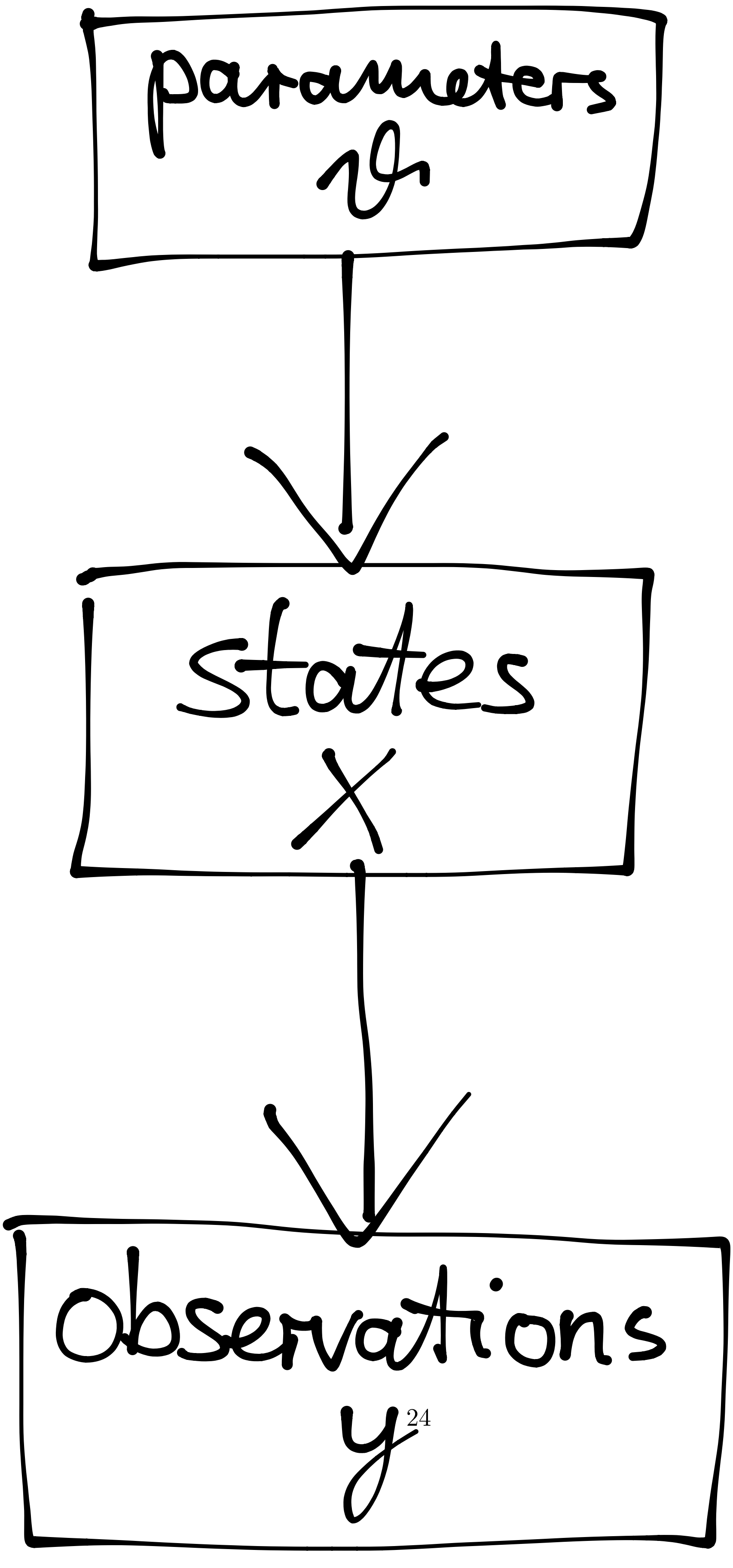
$$p(\theta)$$

2. model trajectories

$$p(x|\theta)$$

3. observations

$$p(y|x,\theta)$$


parameters
ϑ
states
x
observations
y

4.4.1 Joint probability

- applying the chain rule

$$p(y, x, \theta) = p(y|x, \theta)p(x|\theta)p(\theta)$$

4.5 Doing inference

- The joint probability

$$p(y, x, \theta) = p(y|x, \theta)p(x|\theta)p(\theta)$$

  encodes everything about our **model**.

- To do model fitting and inference, set y to the data we observe and calculate the *posterior* probabilities

$$p(x, \theta|y) = \frac{p(y, x, \theta)}{p(y)}$$

- We can rewrite this as

$$p(x, \theta|y) = p(\theta|y)p(x|\theta, y)$$

  - Calculating the first factor is needed for **parameter fitting**.
  - Calculating the second factor is needed for **state estimation**.

- This is what we'll do in this course!

5 Doing inference

5.1 Parameter estimation

$$p(x, \theta|y) = p(\theta|y)p(x|\theta, y)$$

- In a **deterministic** model, every θ leads to one possible x_θ. In that case,

$$p(x_\theta, \theta|y) = p(\theta|y)$$

- In general, we cannot write down a formula for

$$p(x_\theta, \theta, y)$$

  . But we can **sample** from it.

5.2 Sampling

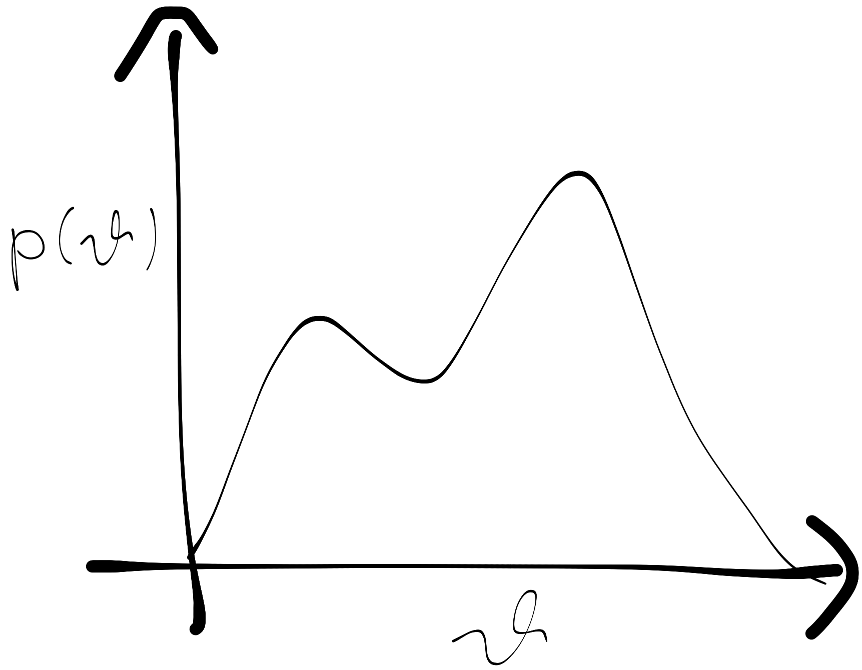
- Imagine you want to know the mean height of people in London
- Cannot measure all people – take a **sample** S
- Mean:

$$\begin{aligned} m &= \frac{1}{N_{\mathrm{Lon}}} \sum_{i \in \mathrm{Lon}} h_i \\ &\approx \frac{1}{N_S} \sum_{i \in S} h_i \end{aligned}$$

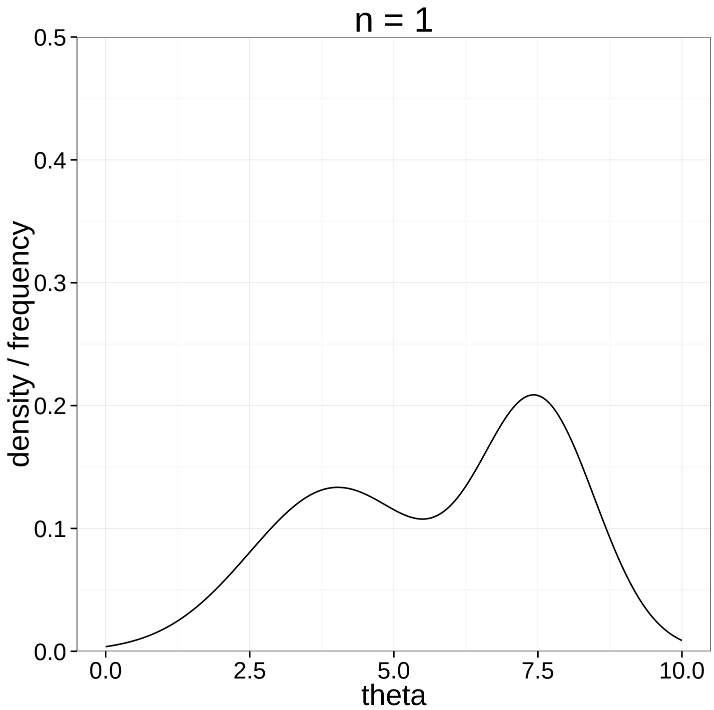
- We can do something similar to study the posterior

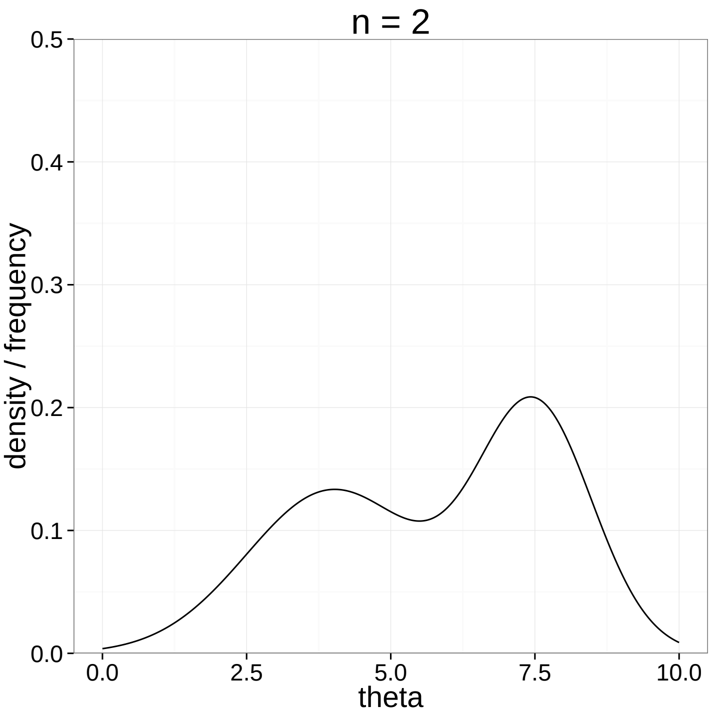
5.3 Monte-Carlo sampling

- Explore a distribution by sampling

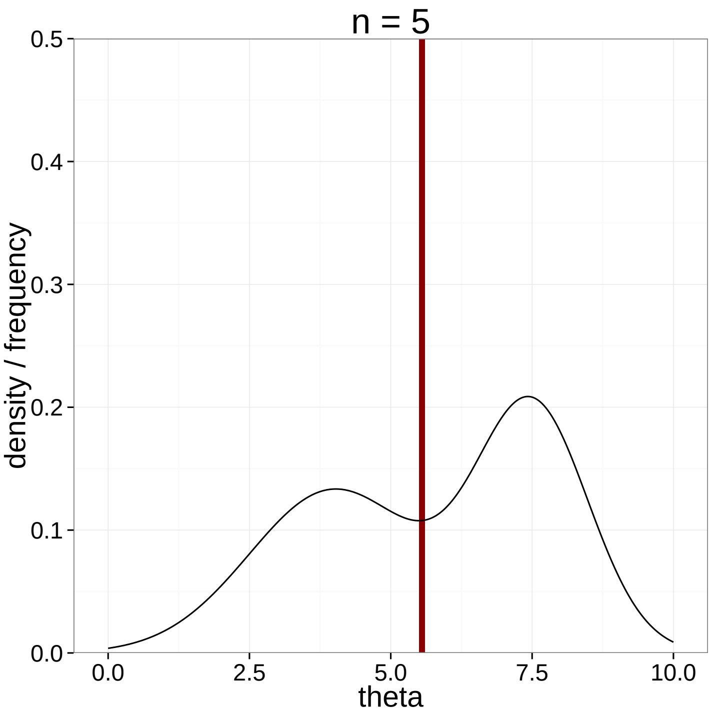


5.3.1

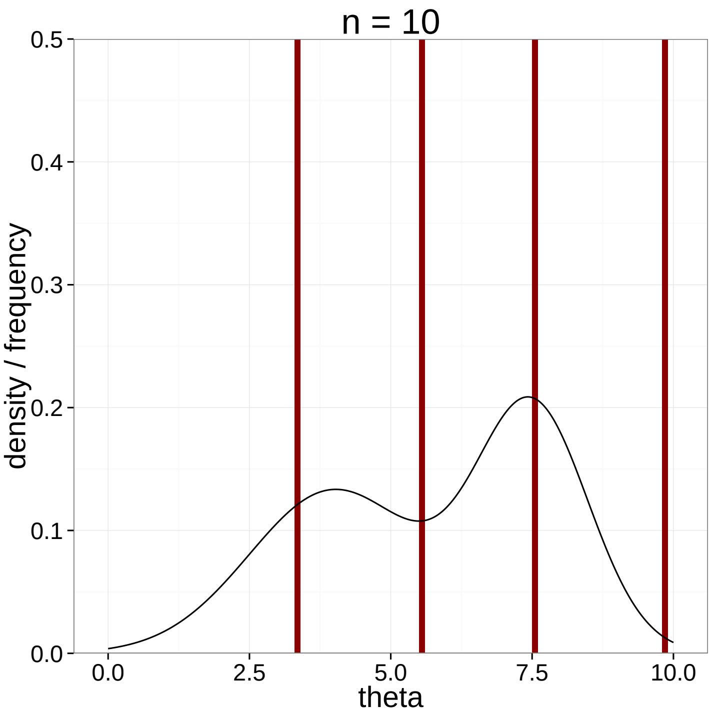


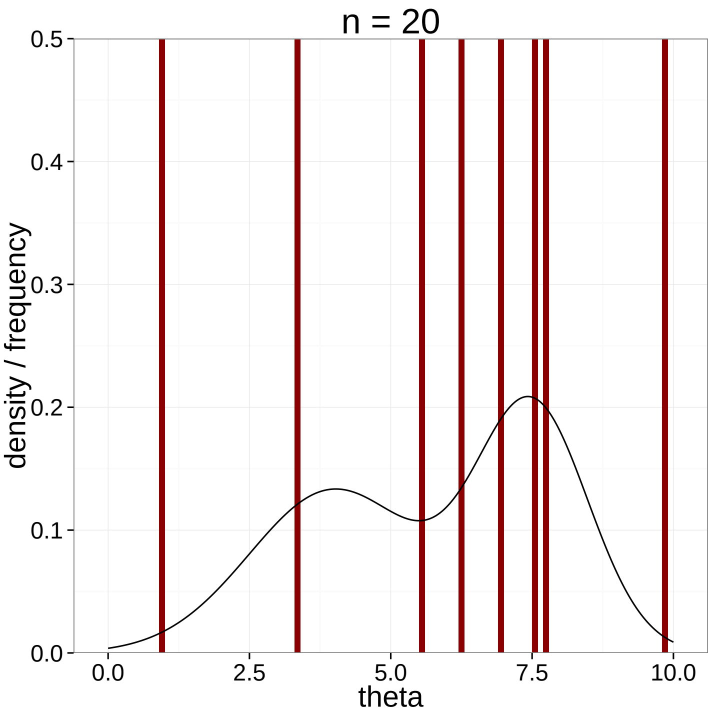


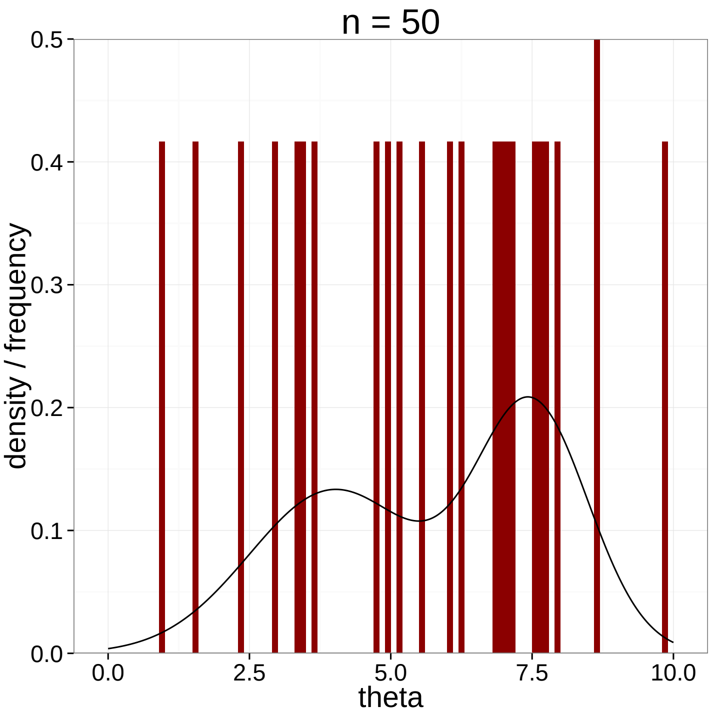
n = 2
0.5
0.4
0.3
0.2
0.1
0.0
density / frequency
0.0
2.5
5.0
7.5
10.0
theta

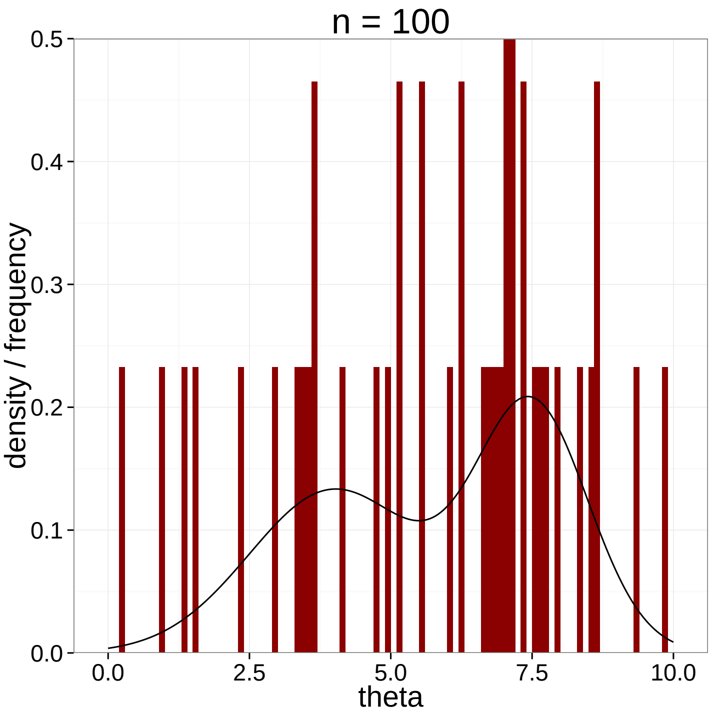


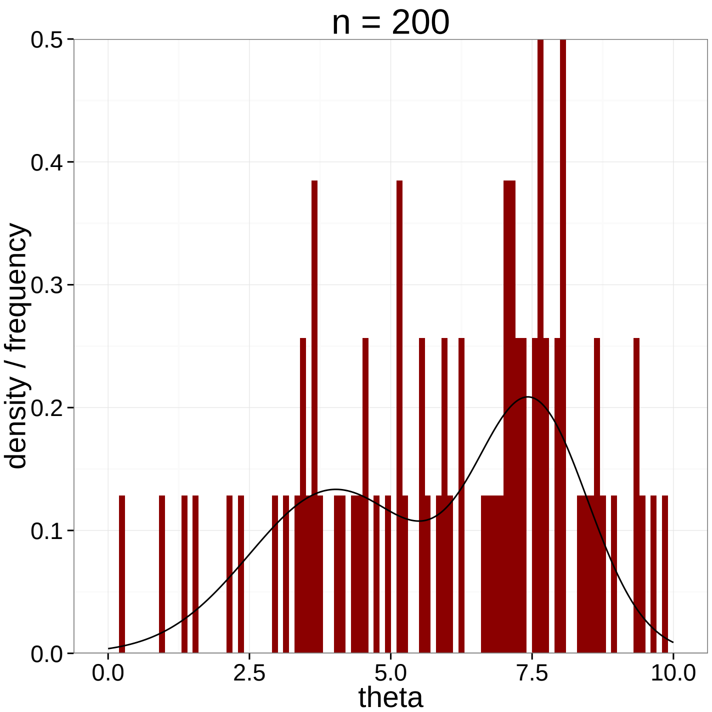
n = 5
0.5
0.4
0.3
0.2
0.1
0.0
density / frequency
0.0
2.5
5.0
7.5
10.0
theta

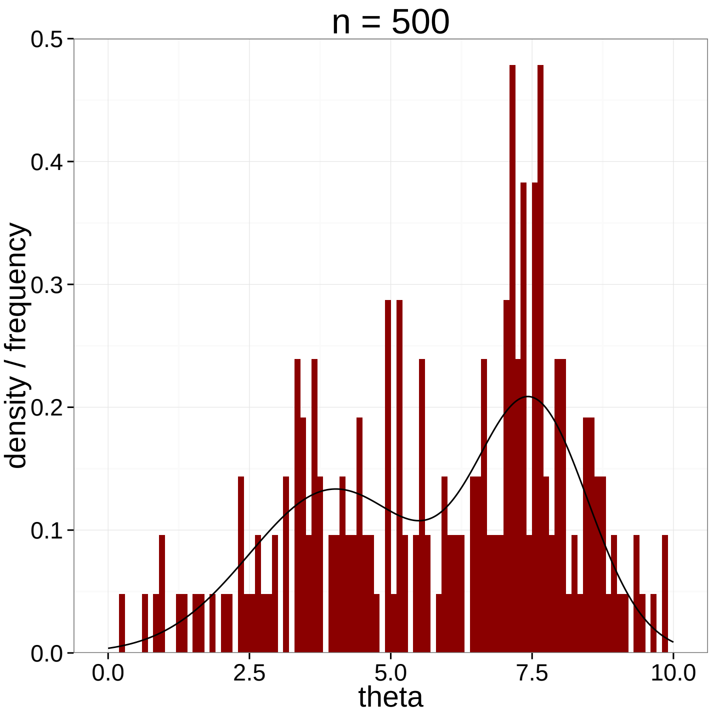

n = 10
density / frequency
0.0
0.1
0.2
0.3
0.4
0.5
0.0
2.5
5.0
7.5
10.0
theta

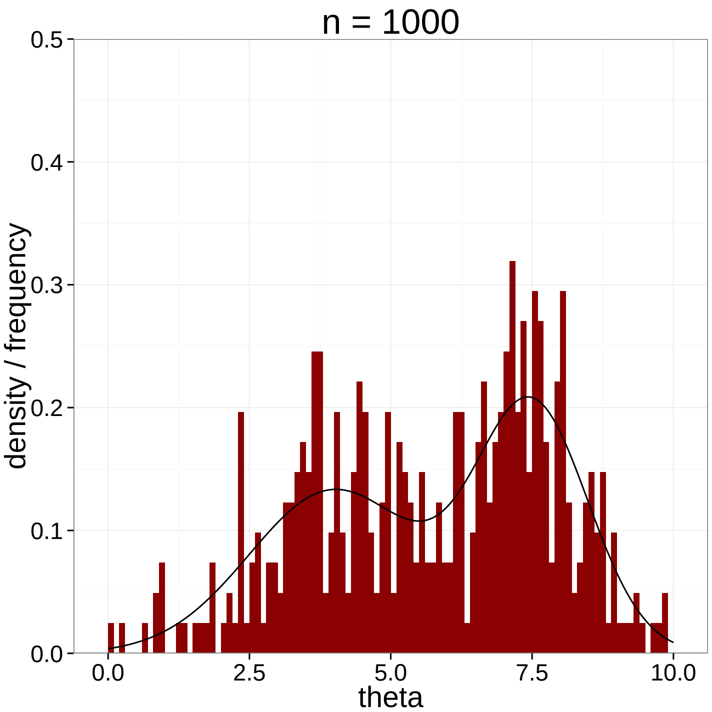

n = 20
density / frequency
0.0
0.1
0.2
0.3
0.4
0.5
0.0
2.5
5.0
7.5
10.0
theta

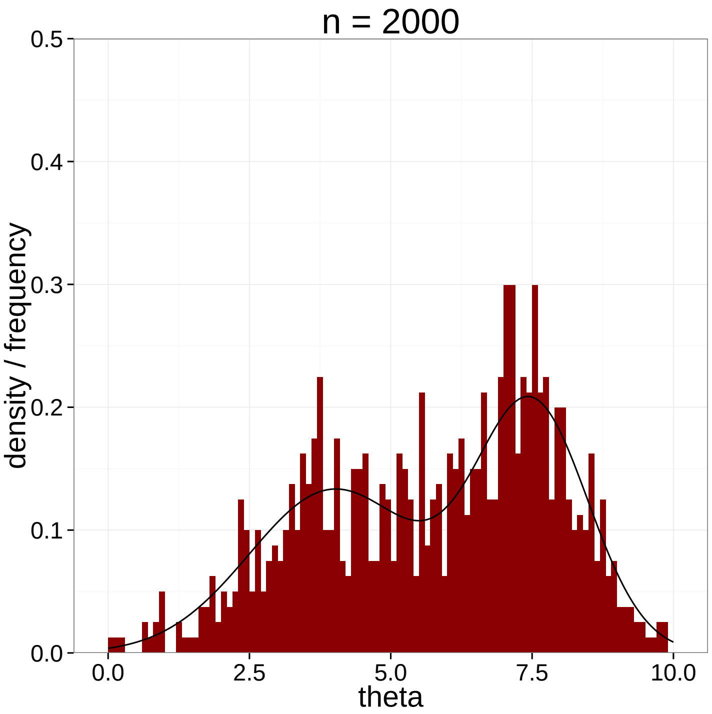

n = 50
density / frequency
0.5
0.4
0.3
0.2
0.1
0.0
0.0
2.5
5.0
7.5
10.0
theta

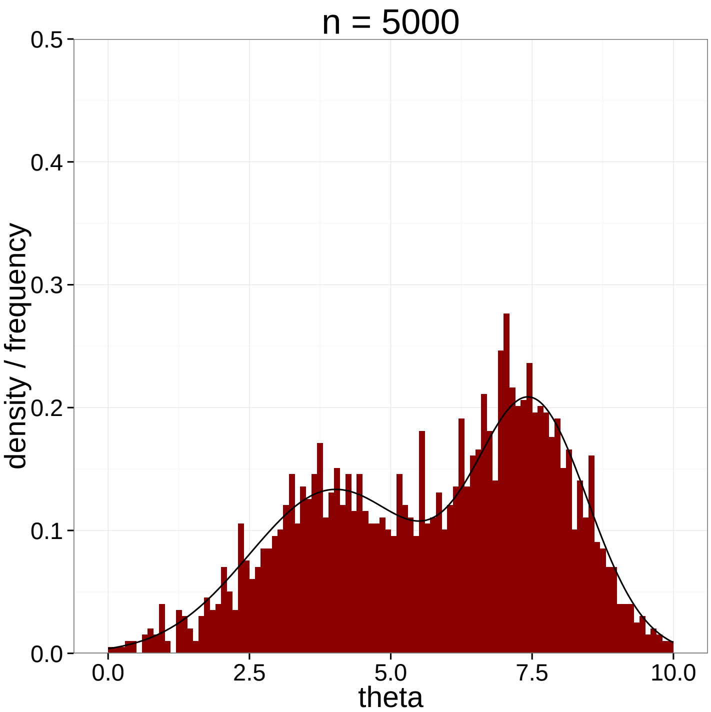

n = 100
density / frequency
0.5
0.4
0.3
0.2
0.1
0.0
0.0
2.5
5.0
7.5
10.0
theta

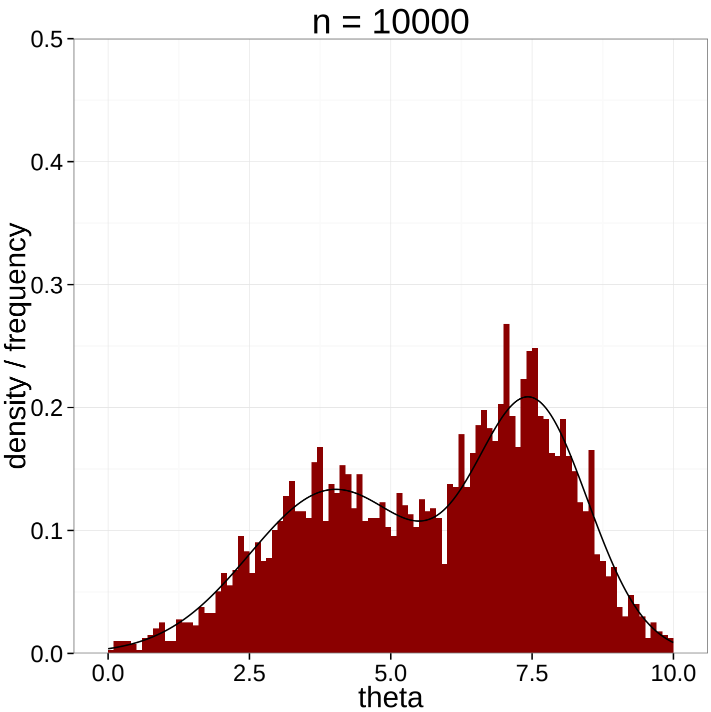

n = 200
density / frequency
0.0
0.1
0.2
0.3
0.4
0.5
0.0
2.5
5.0
7.5
10.0
theta

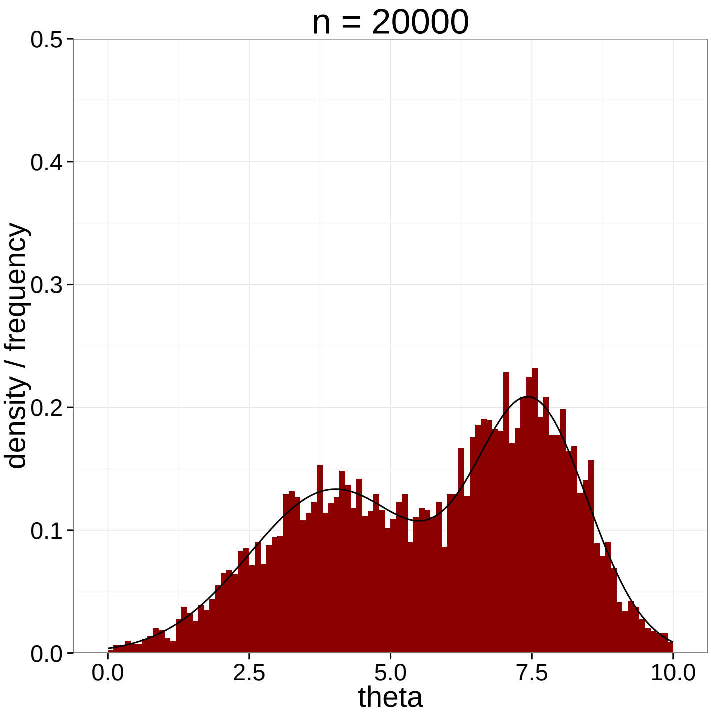

n = 500
density / frequency
0.0
0.1
0.2
0.3
0.4
0.5
0.0
2.5
5.0
7.5
10.0
theta

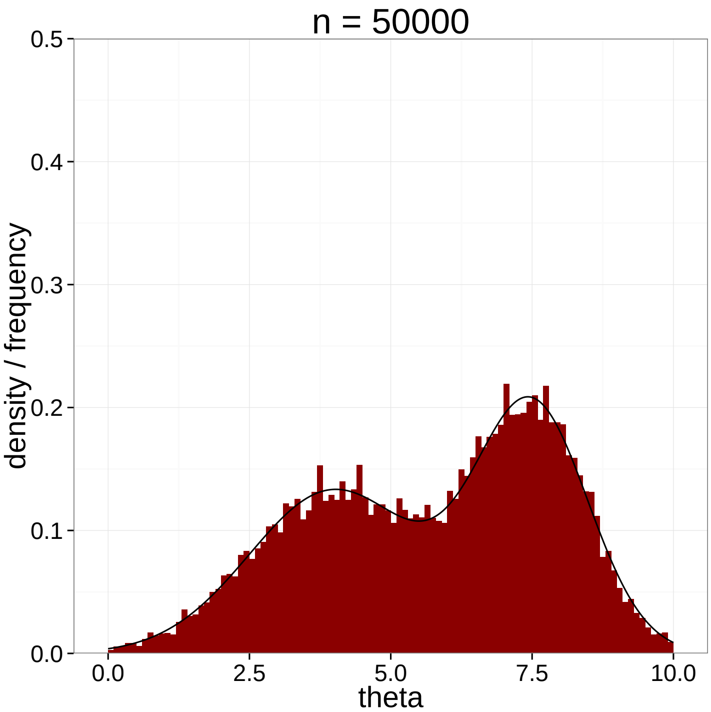

n = 1000
density / frequency
0.5
0.4
0.3
0.2
0.1
0.0
0.0
2.5
5.0
7.5
10.0
theta

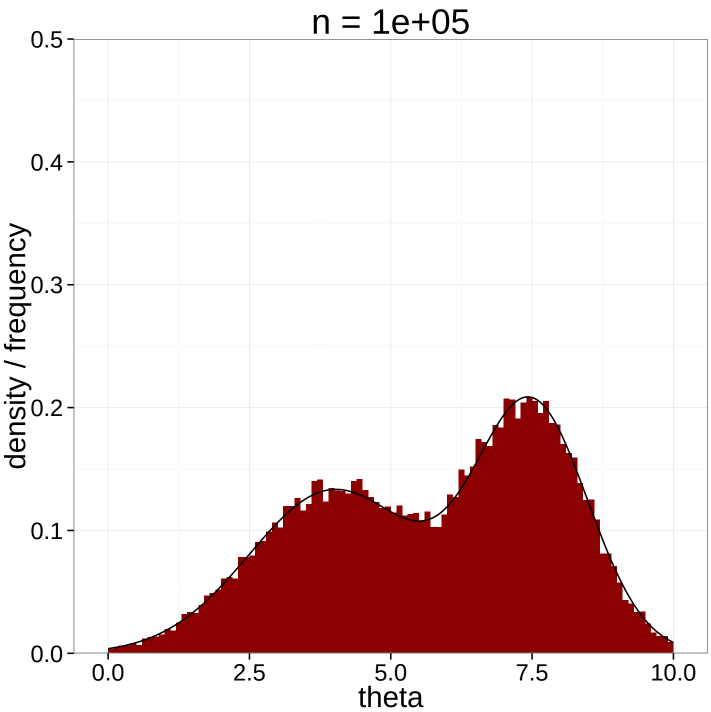

n = 2000
density / frequency
0.5
0.4
0.3
0.2
0.1
0.0
0.0
2.5
5.0
7.5
10.0
theta

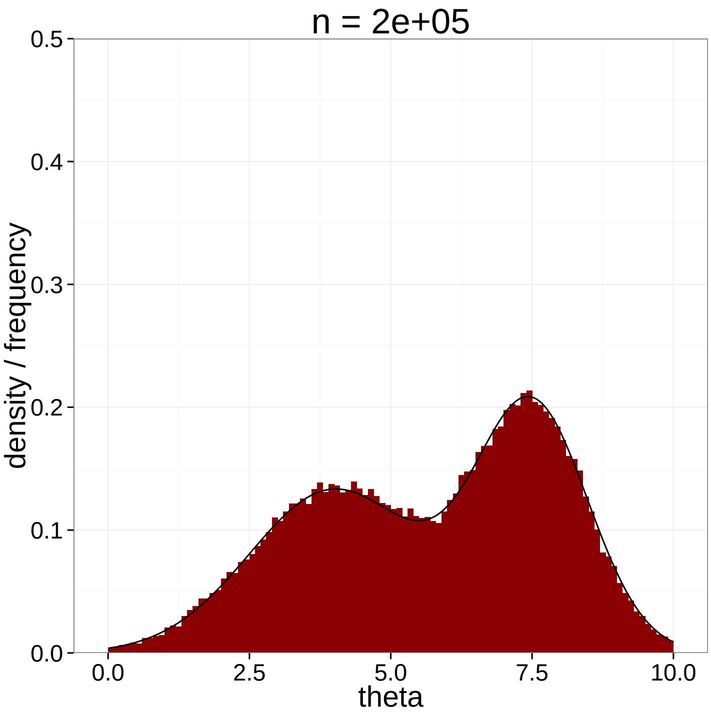

n = 5000
0.5
0.4
0.3
0.2
0.1
0.0
density / frequency
0.0
2.5
5.0
7.5
10.0
theta

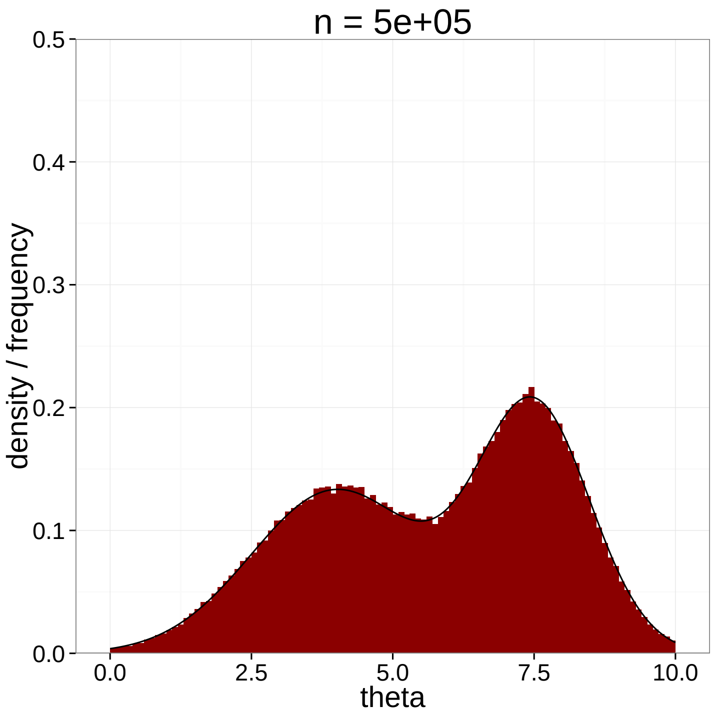

n = 10000
0.5
0.4
0.3
0.2
0.1
0.0
density / frequency
0.0
2.5
5.0
7.5
10.0
theta

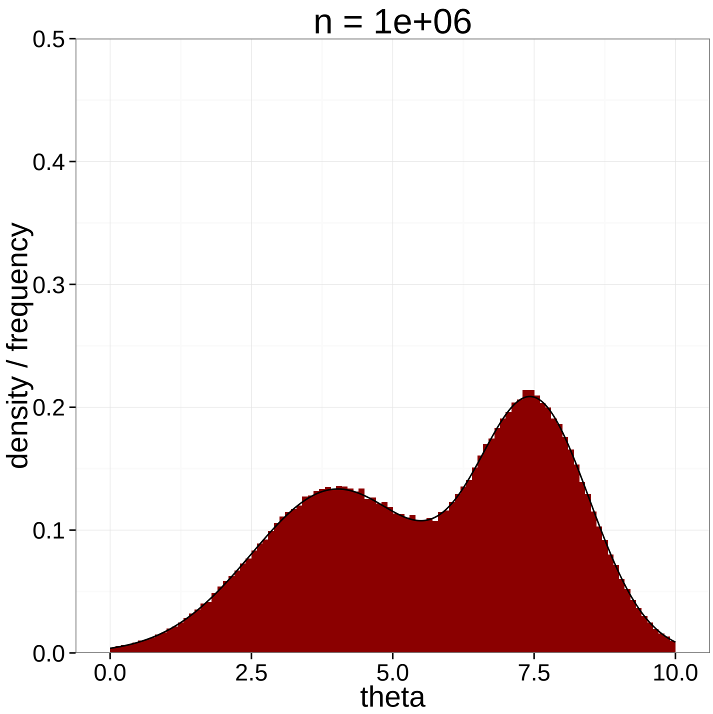

n = 20000
density / frequency
0.0
0.1
0.2
0.3
0.4
0.5
0.0
2.5
5.0
7.5
10.0
theta

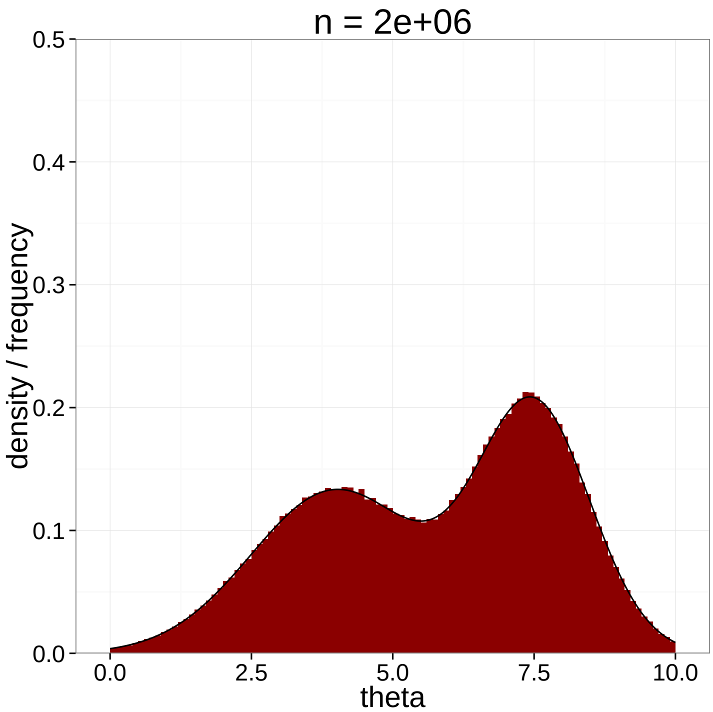

n = 50000
density / frequency
0.0
0.1
0.2
0.3
0.4
0.5
0.0
2.5
5.0
7.5
10.0
theta

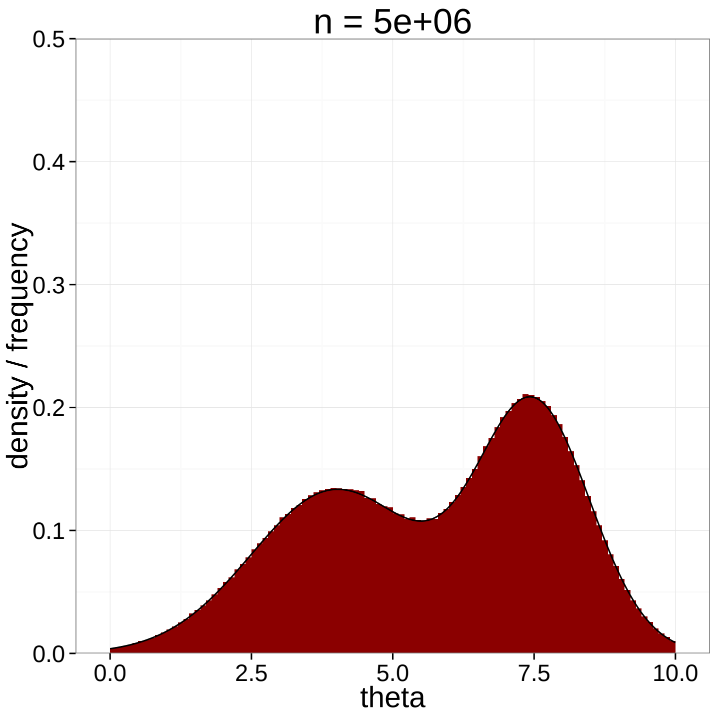

n = 1e+05
density / frequency
0.0
0.1
0.2
0.3
0.4
0.5
0.0
2.5
5.0
7.5
10.0
theta

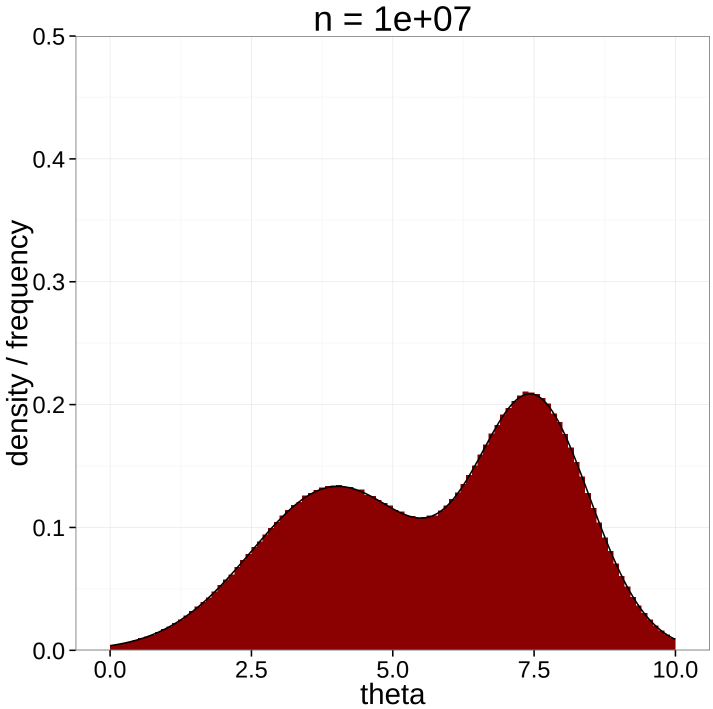

n = 2e+05
density / frequency
0.0
0.1
0.2
0.3
0.4
0.5
0.0
2.5
5.0
7.5
10.0
theta


n = 5e+05
0.5
0.4
0.3
0.2
0.1
0.0
density / frequency
0.0
2.5
5.0
7.5
10.0
theta


n = 1e+06
0.5
0.4
0.3
0.2
0.1
0.0
density / frequency
0.0
2.5
5.0
7.5
10.0
theta

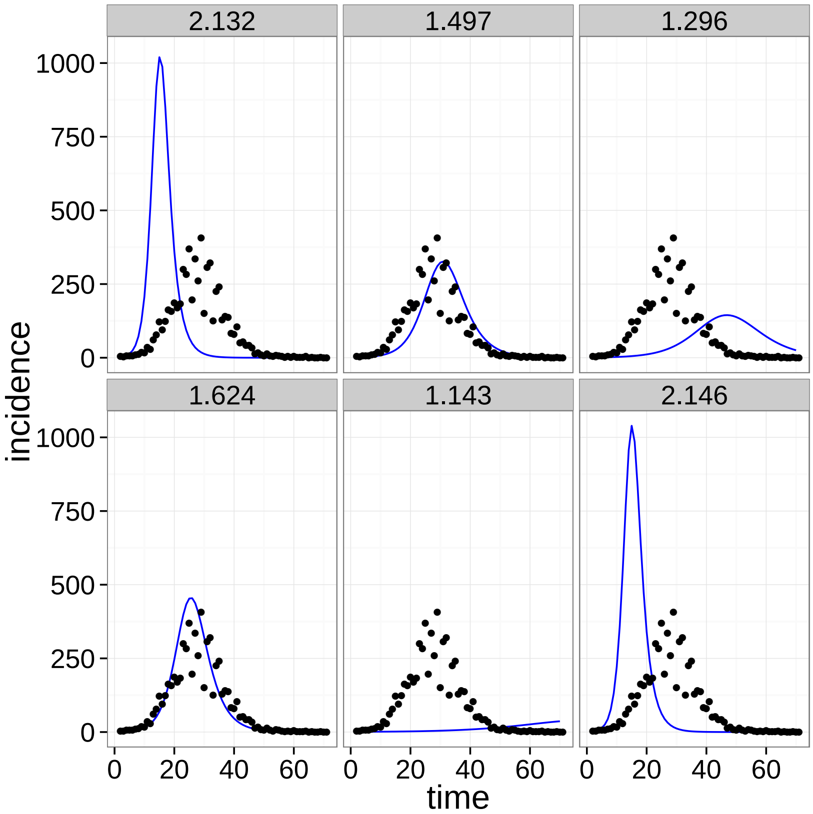

n = 2e+06
density / frequency
0.5
0.4
0.3
0.2
0.1
0.0
0.0
2.5
5.0
7.5
10.0
theta


n = 5e+06
density / frequency
0.5
0.4
0.3
0.2
0.1
0.0
0.0
2.5
5.0
7.5
10.0
theta



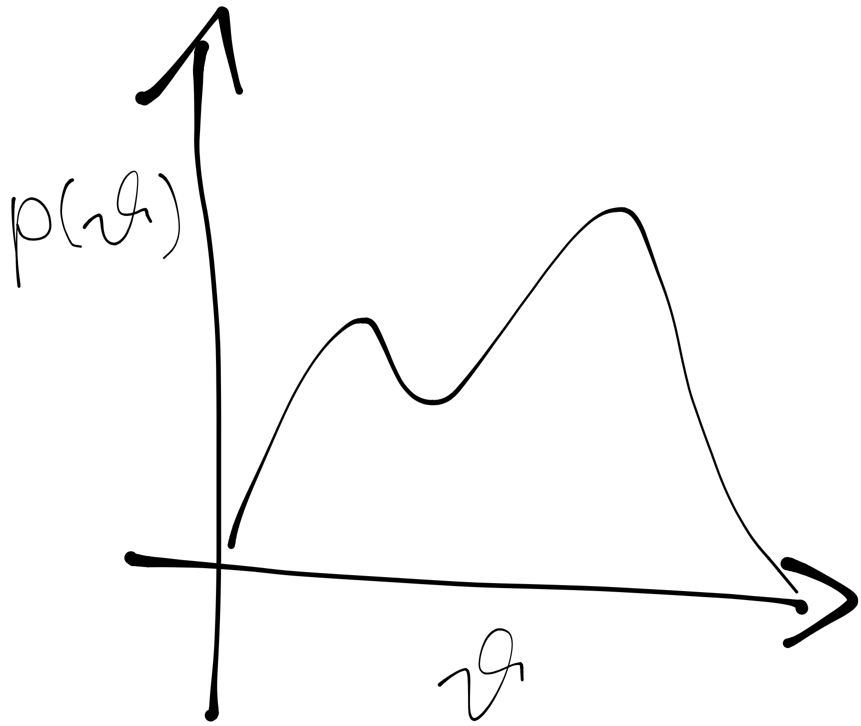
n = 1e+07
0.5
0.4
0.3
0.2
0.1
0.0
density / frequency
0.0
2.5
5.0
7.5
10.0
theta

5.4 Wasting time in sampling

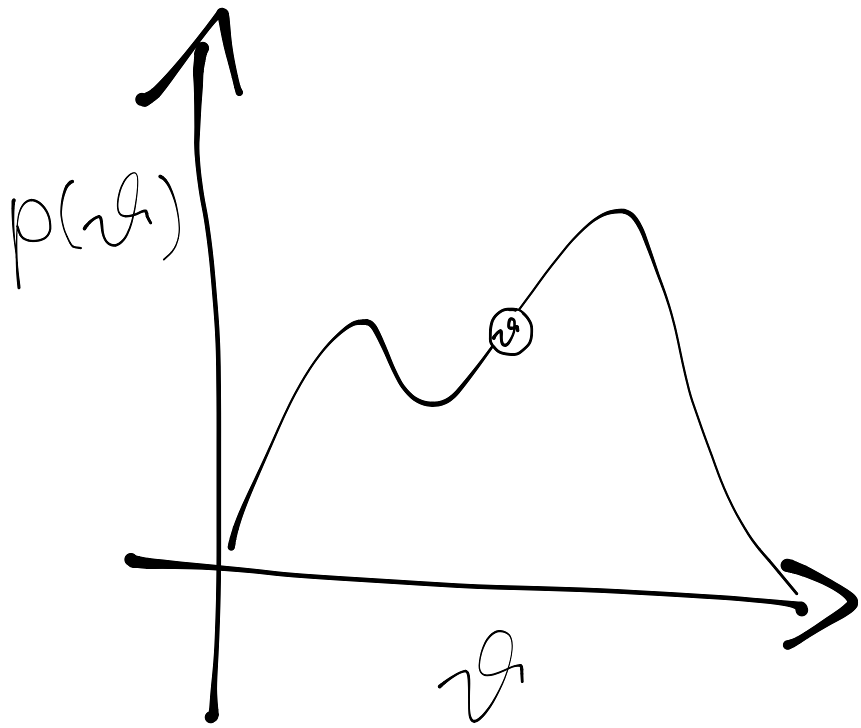


5.5 Metropolis-Hastings sampler

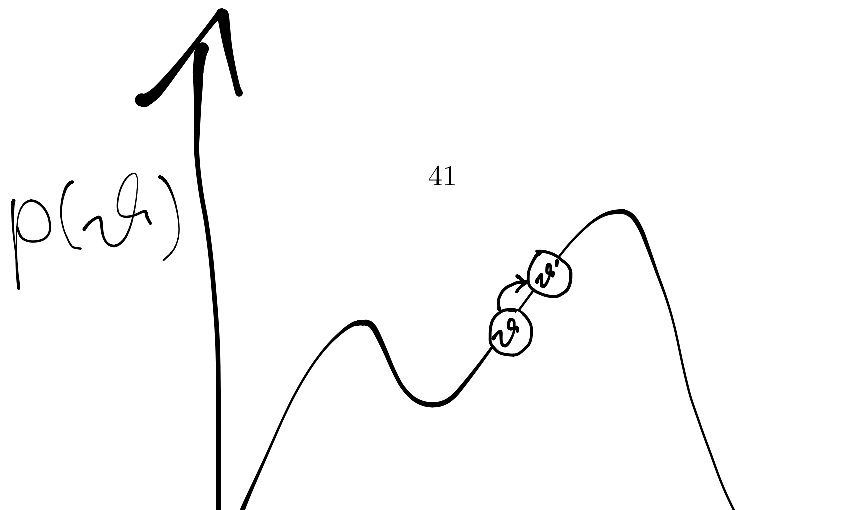
- use the information of the last step to suggest the next one:



5.5.1



5.5.2



[2] J. Dureau, K. Kalogeropoulos, and M. Baguelin. "Capturing the time-varying drivers of an epidemic using stochastic dynamical systems." eng. In: *Biostatistics* 14.3 (July 2013), pp. 541–555.

[3] A. A. King et al. "Inapparent infections and cholera dynamics." In: *Nature* 454.7206 (Aug. 2008), pp. 877–880.

[4] P. S. Laplace. *Théorie analytique des probabilités*. Courcier, 1812.

[5] R. A. Fisher. "On the Mathematical Foundations of Theoretical Statistics". In: *Phil T R Soc A* 222 (1922), pp. 309–368.

[6] A. M. Ellison. "Bayesian inference in ecology". In: *Ecology Letters* 7.6 (2004), pp. 509–520. ISSN: 1461-0248.

[7] R. Munroe. *Frequentists vs. Bayesians*. http://xkcd.com/1132/. 2012.

[8] M. Baguelin et al. "Assessing optimal target populations for influenza vaccination programmes: an evidence synthesis and modelling study." eng. In: *PLoS Med* 10.10 (Oct. 2013), e1001527.