Model fitting and inference for infectious disease dynamics

2019

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

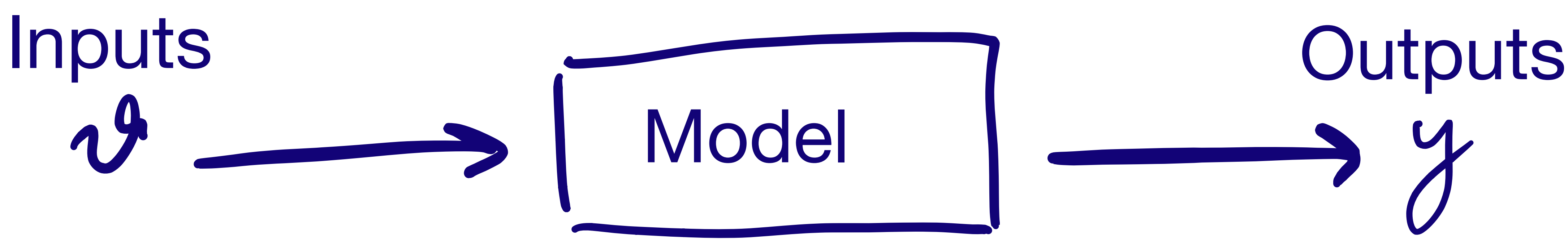
cmmid centre *for the* mathematical modelling *of* infectious diseases

LONDON SCHOOL *of* HYGIENE &TROPICAL MEDICINE

1. Introduction

Model fitting and inference for infectious disease dynamics

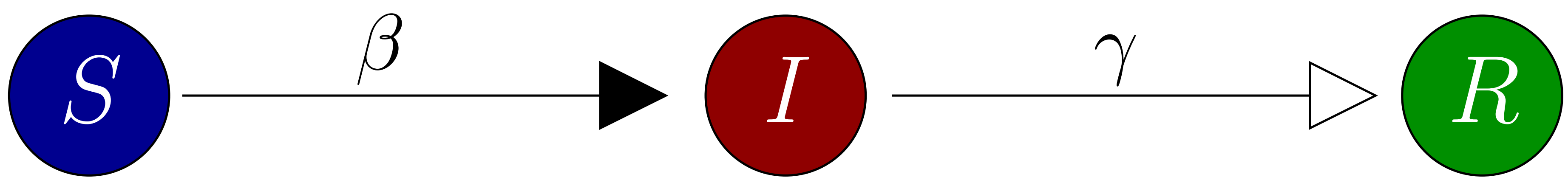
A model takes input values θ (parameters) and returns output values y (observations, data).



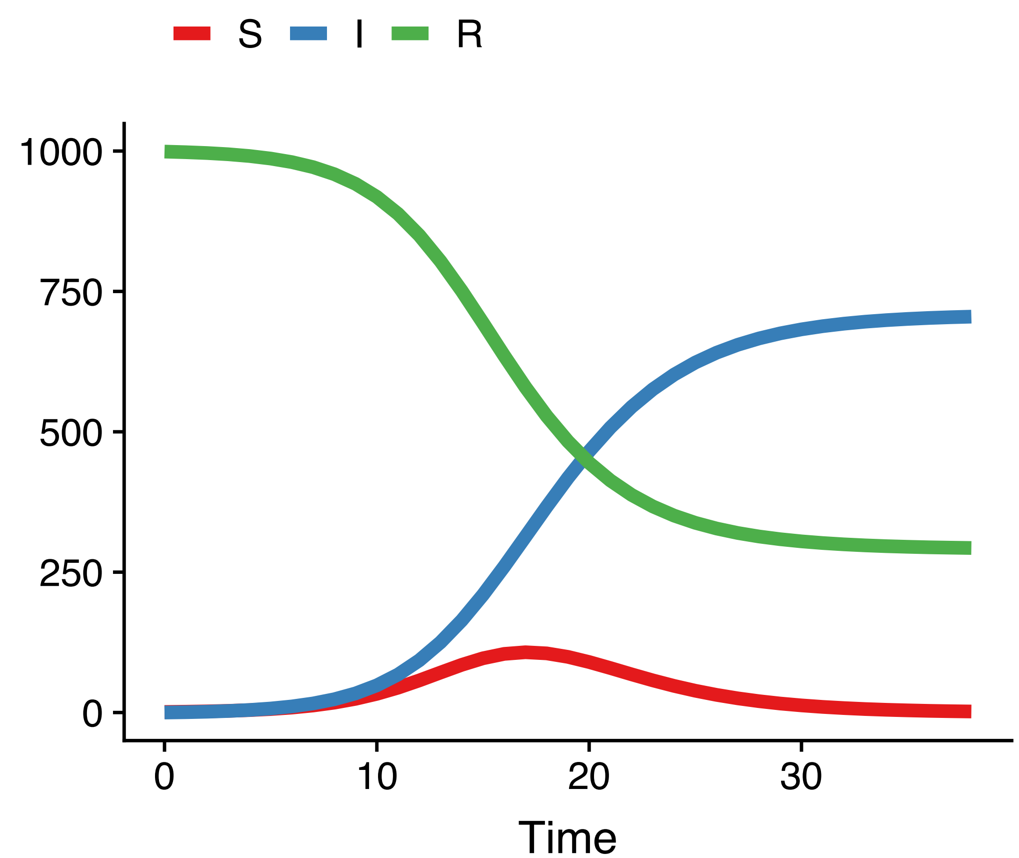
probability $p(y|\theta)$ that y is the output, given inputs θ.

Model fitting and inference for infectious disease dynamics

SIR-type models



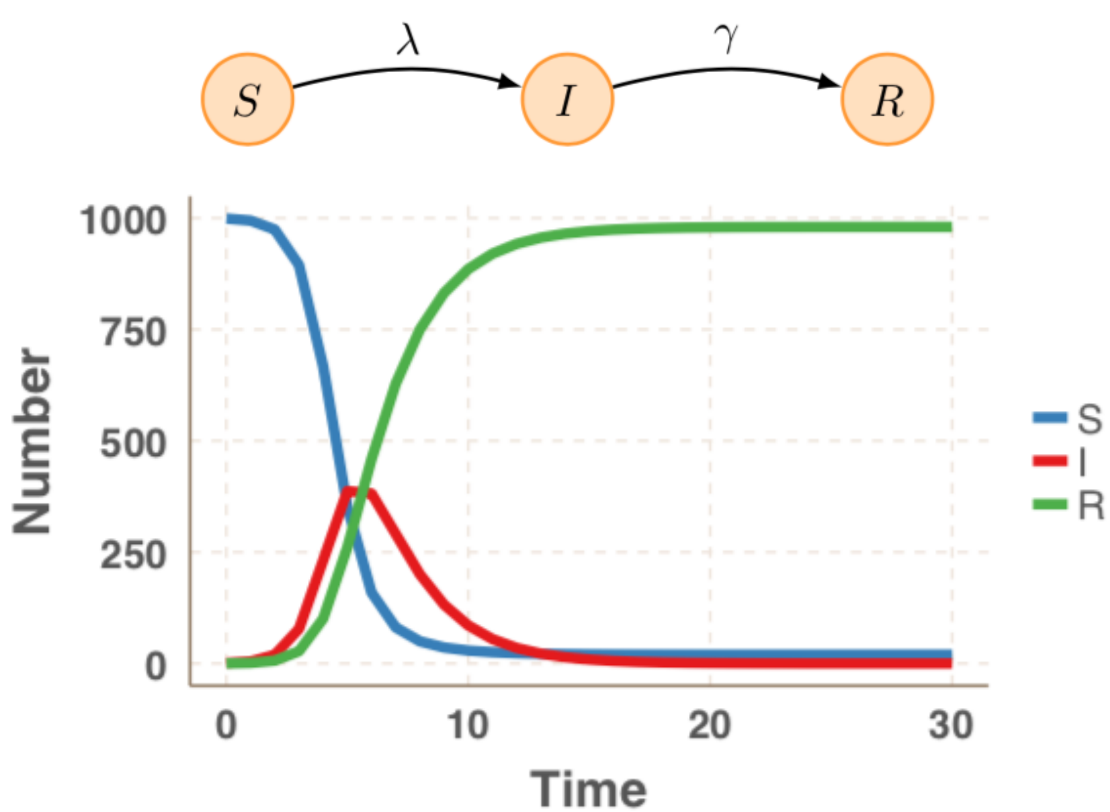
$$\frac{dS}{dt} = -\beta I \frac{S}{N}$$
$$\frac{dI}{dt} = \beta I \frac{S}{N} - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$



Mechanistic approach

- Design model from first principles

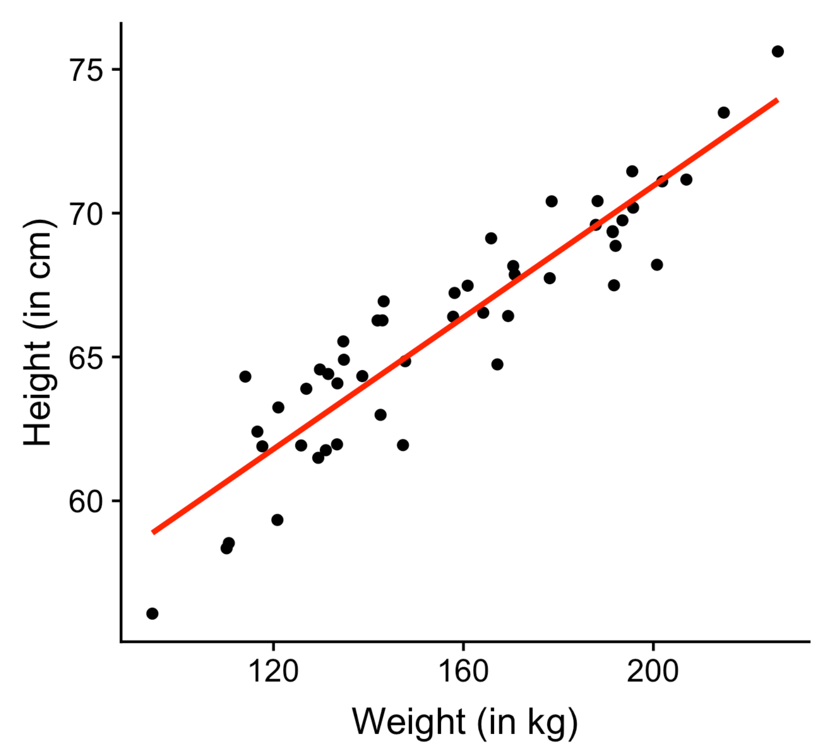
- Focus on model behaviour under different scenarios / parameter ranges



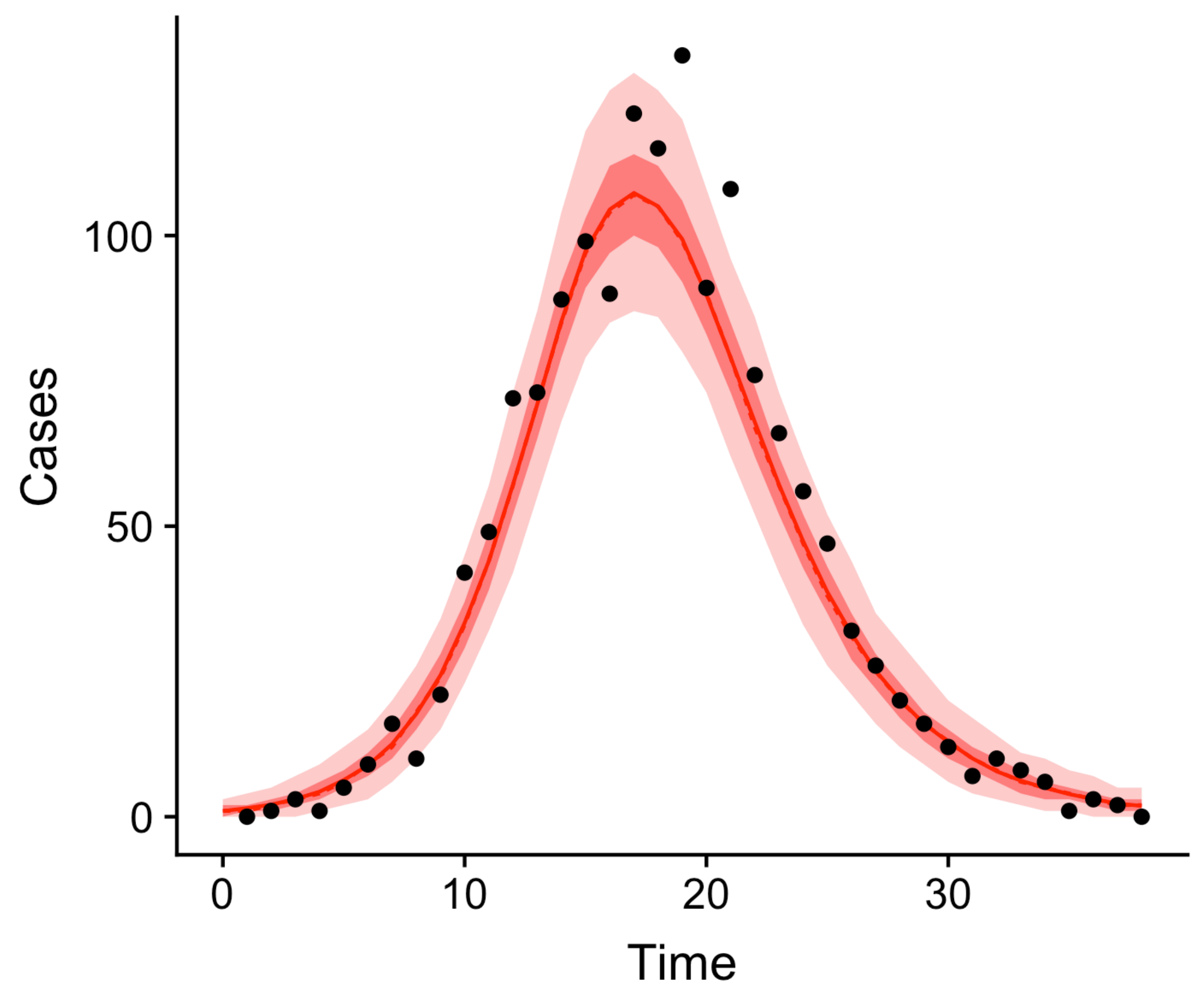
Statistical approach

- Models as a tool to investigate data

- Model choice driven by data and hypotheses about relationships between variables



Simulation-based inference



Combine mechanistic and statistical approach for

- Parameter estimation
- Prediction

Model fitting and inference for infectious disease dynamics

Parameter estimation

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

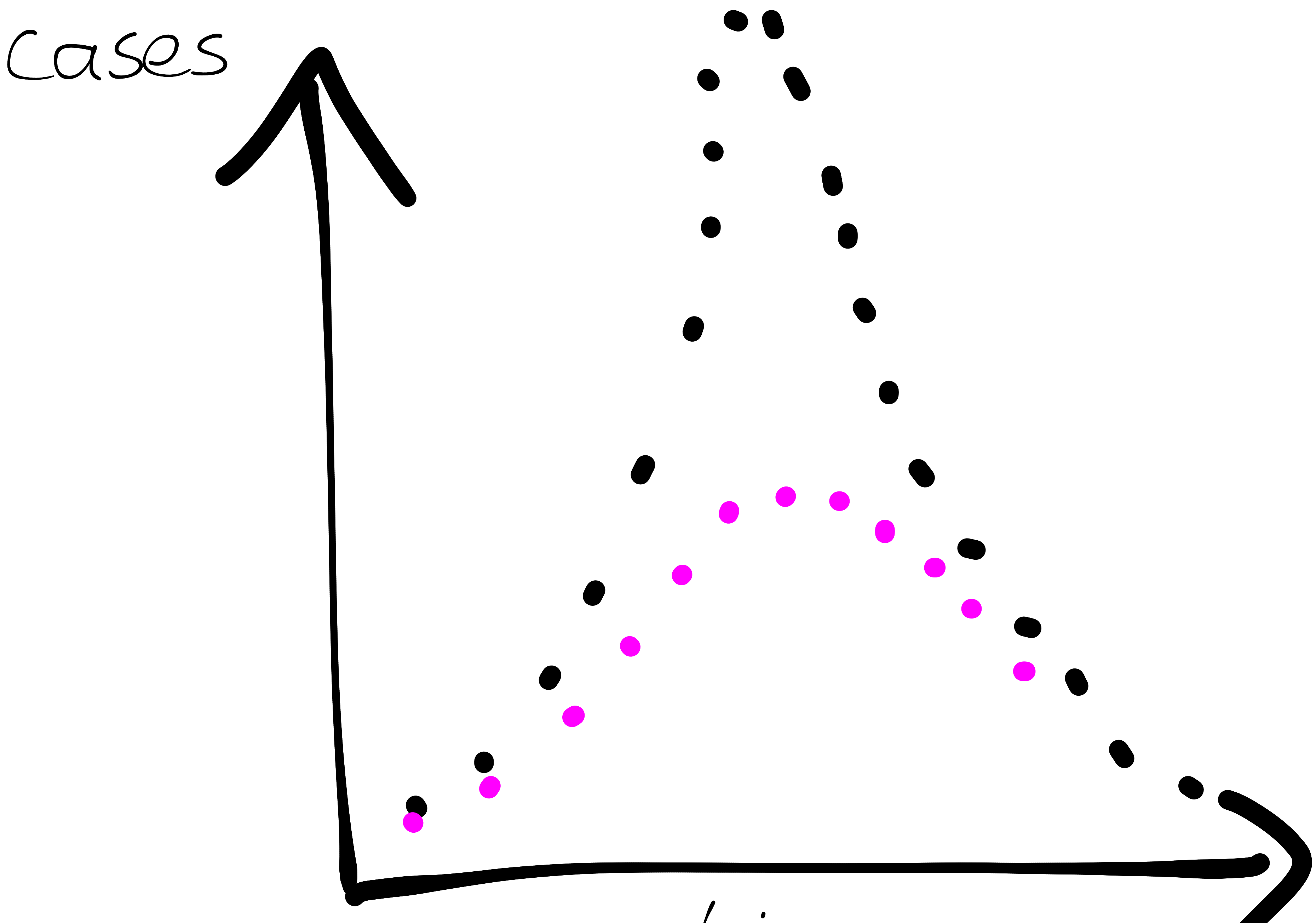
Why do this?

- Inference: 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:)
- Prediction: give probabilities to unobserved data points
  - forecasting: future data points
- Extrapolation:
  - e.g., test the effect of interventions
  - e.g., apply insights to another geographic area

Model fitting and inference for infectious disease dynamics

State estimation

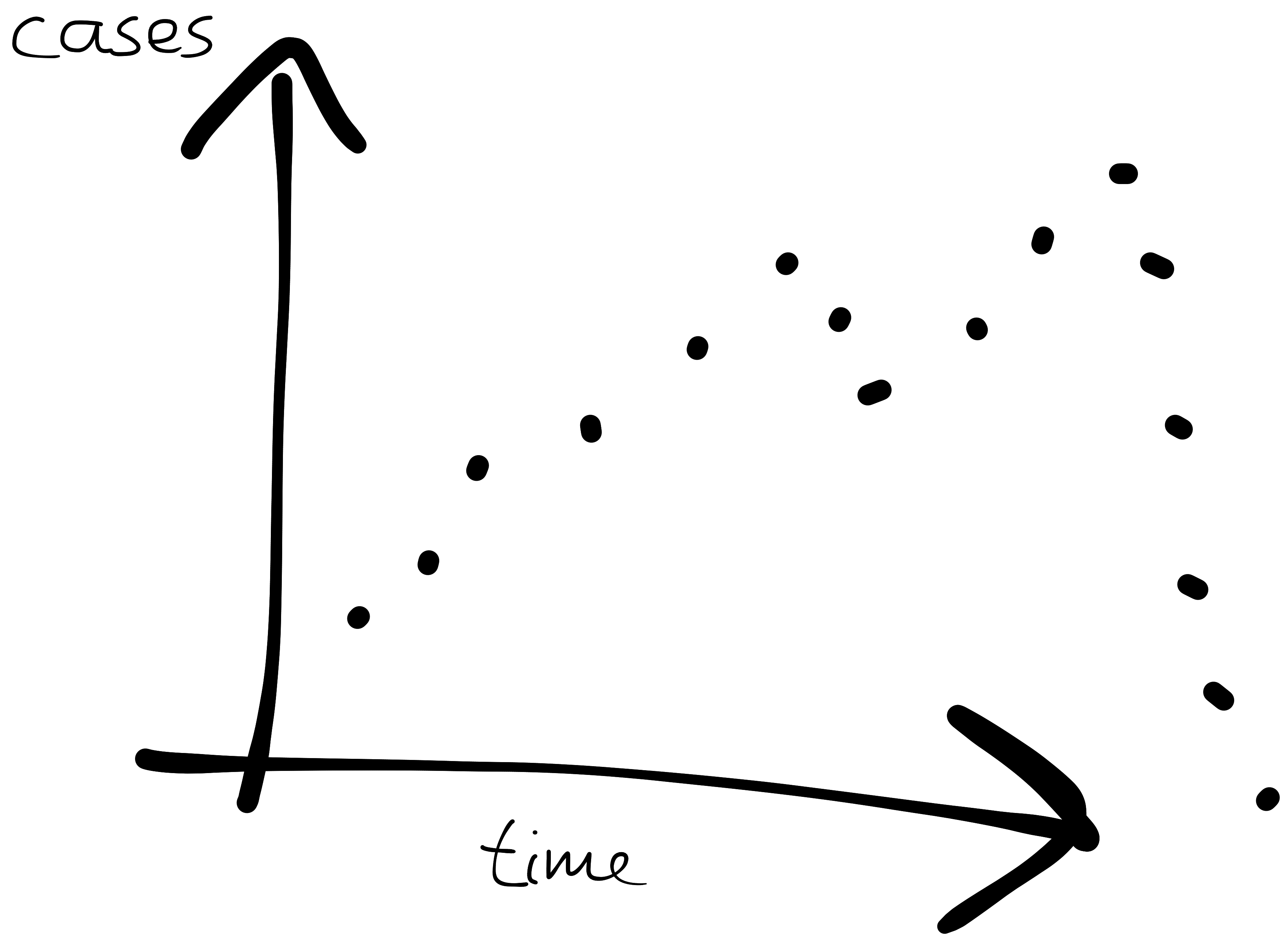
Given what we observe, what is the state of the sytem?



Model fitting and inference for infectious disease dynamics

Model selection

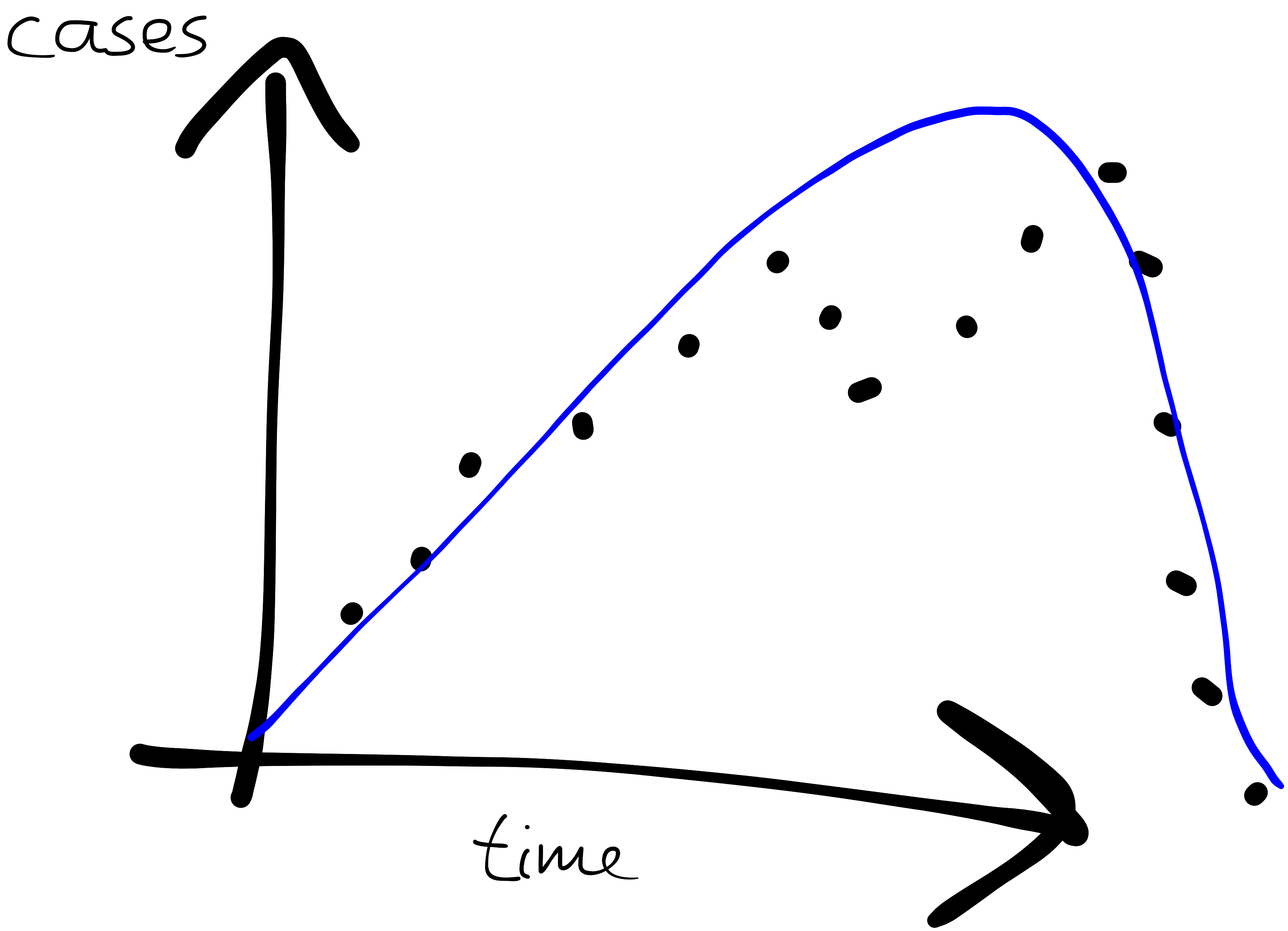
Given a set of potential models, how do we decide which is the right one?



Model fitting and inference for infectious disease dynamics

Model selection

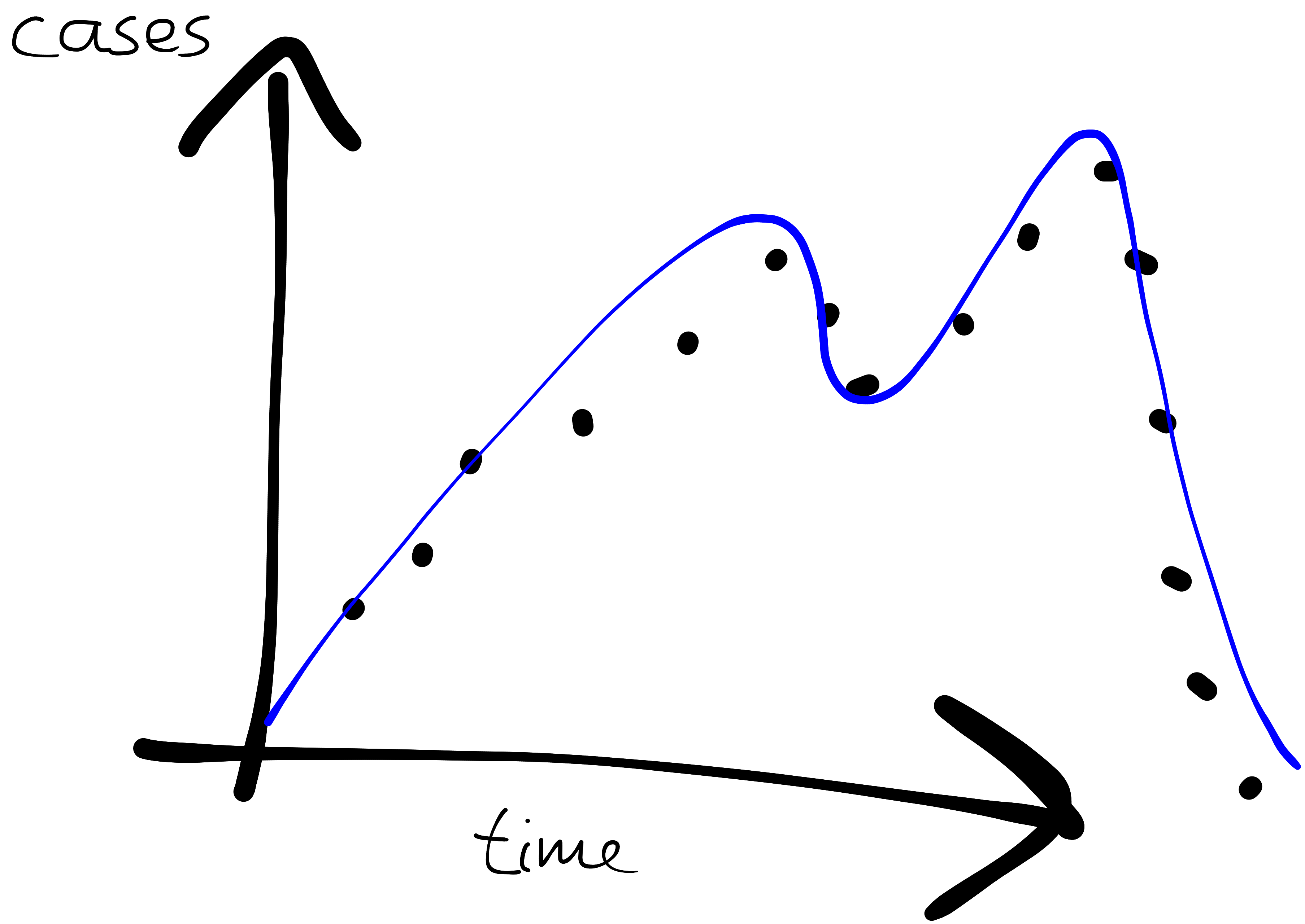
Given a set of potential models, how do we decide which is the right one?



Model fitting and inference for infectious disease dynamics

Model selection

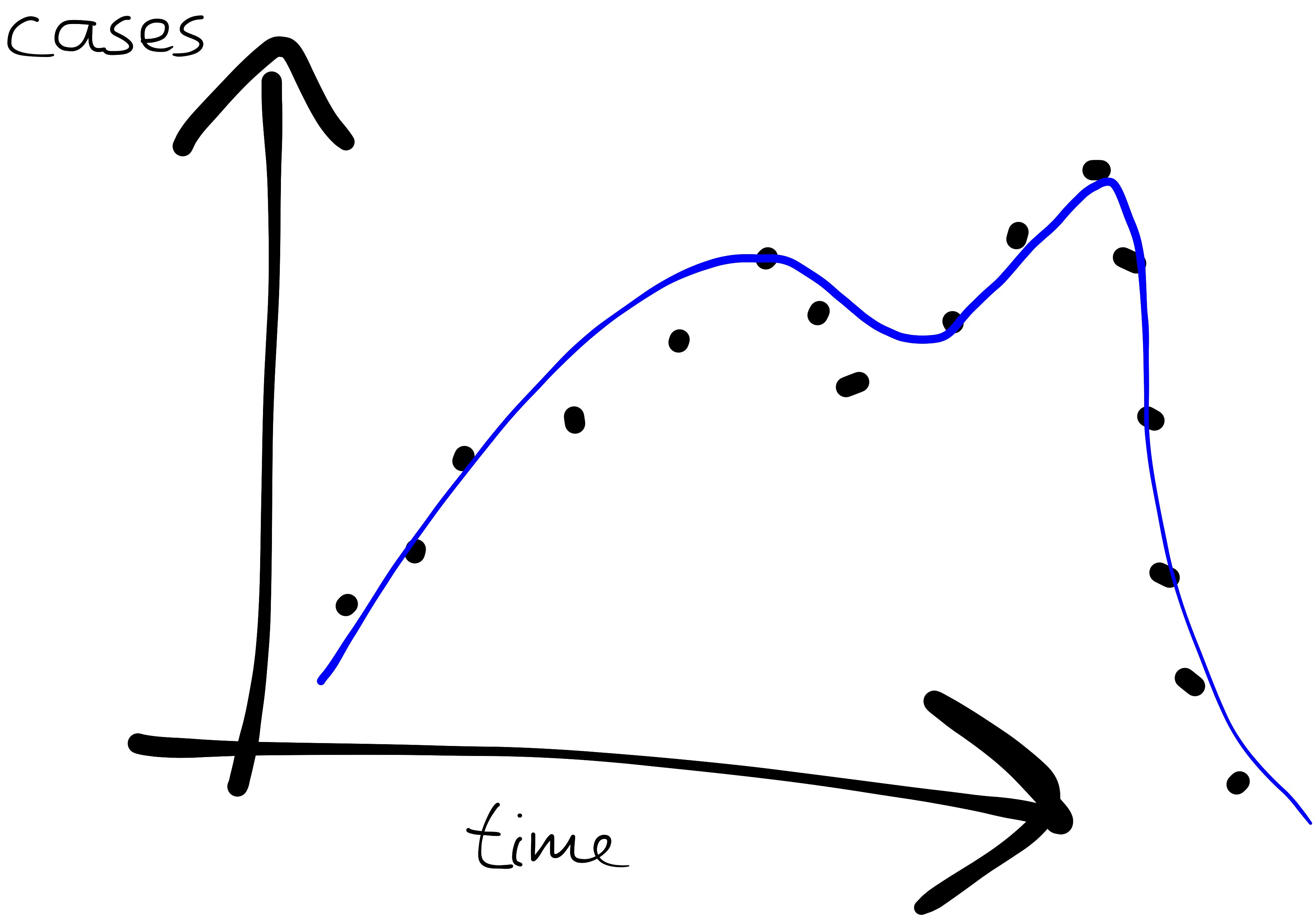
Given a set of potential models, how do we decide which is the right one?



Model fitting and inference for infectious disease dynamics

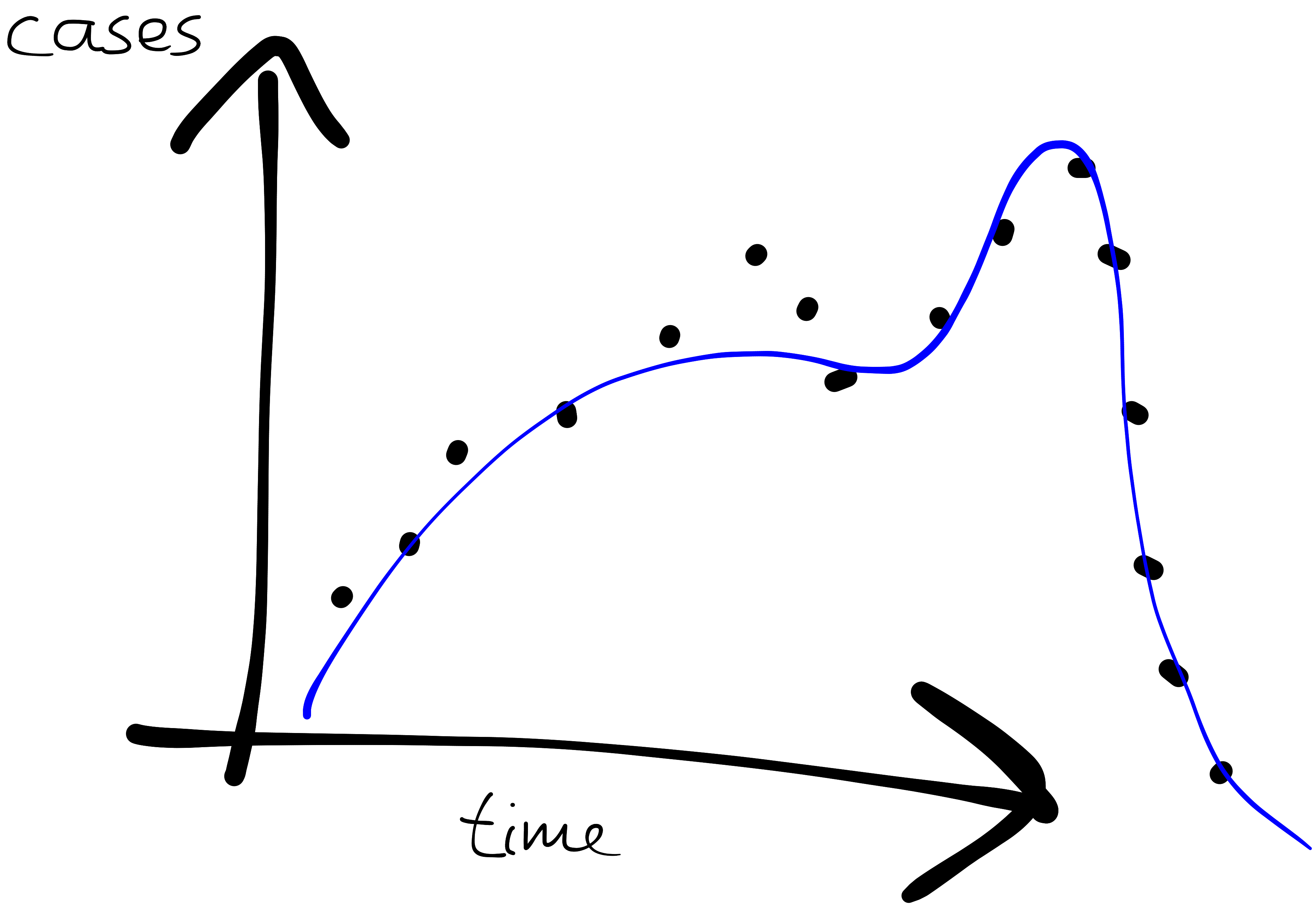
Model selection

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



Model selection

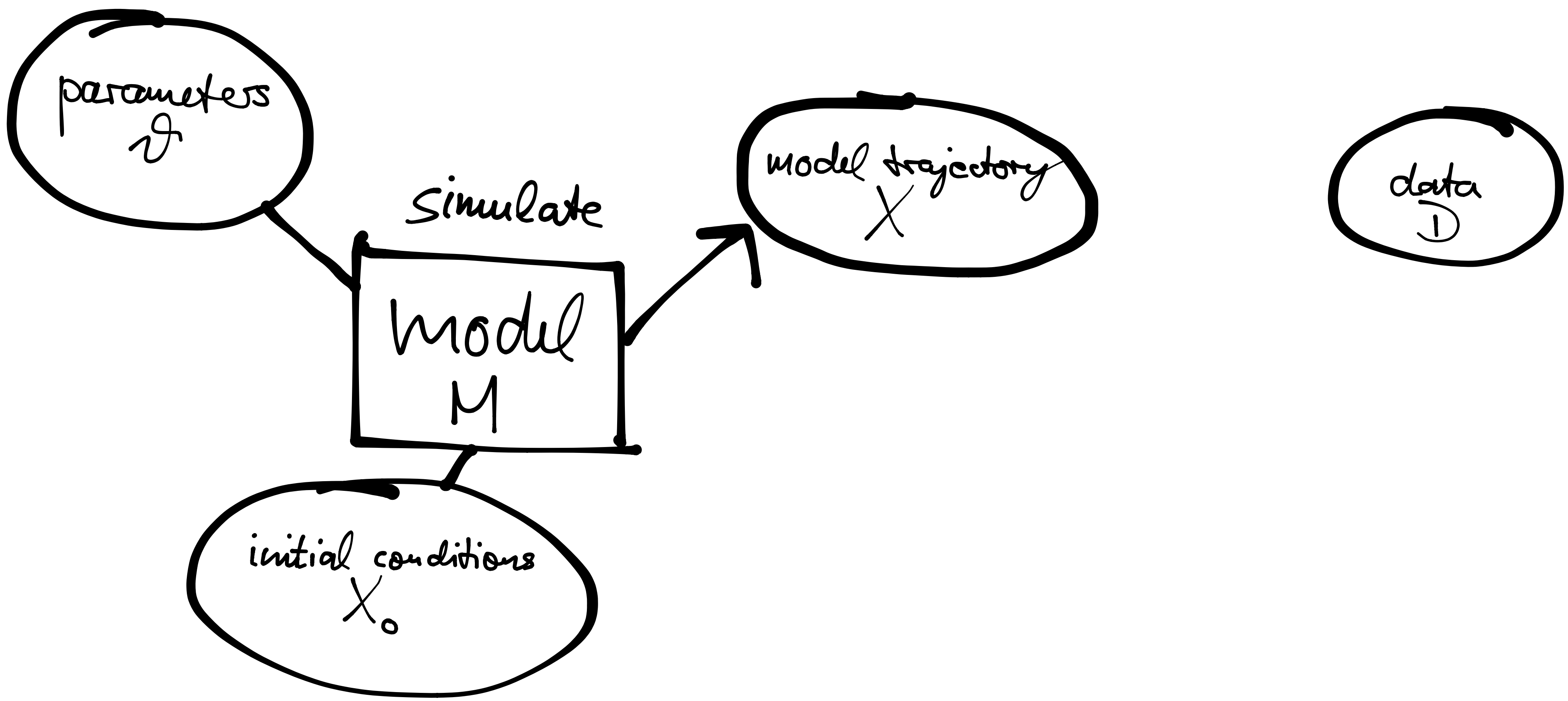
Given a set of potential models, how do we decide which is the right one?

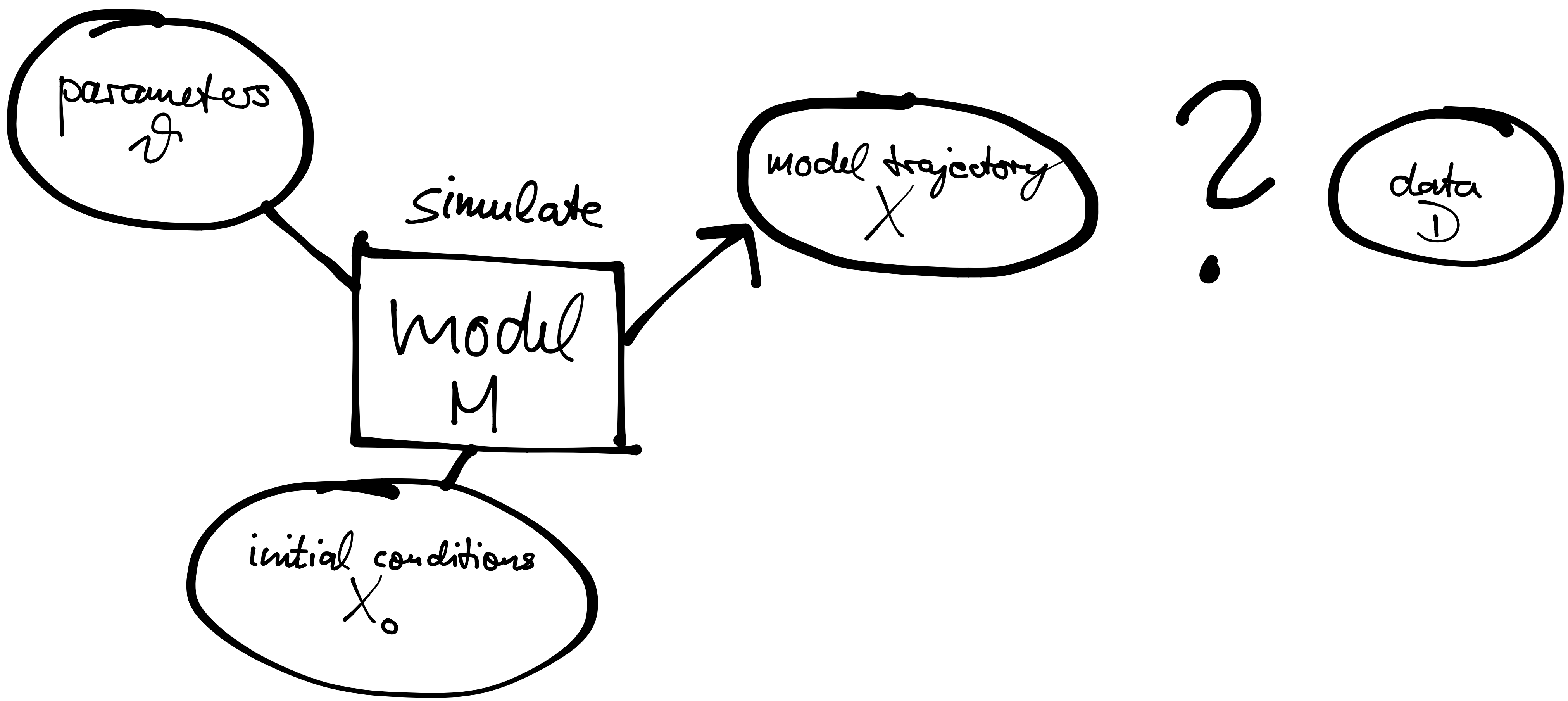


2. Linking models to data

model
M

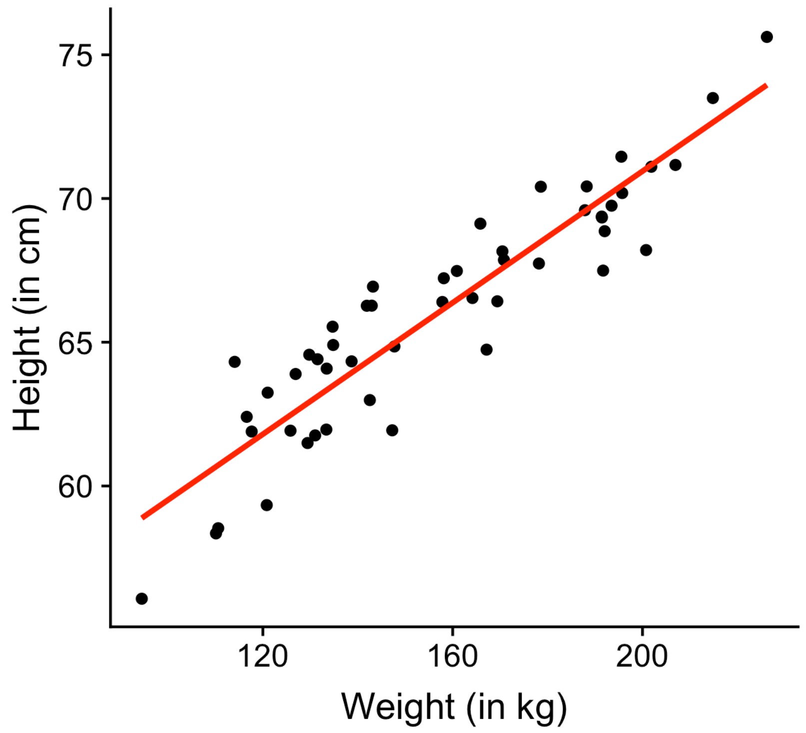
data
D


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


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

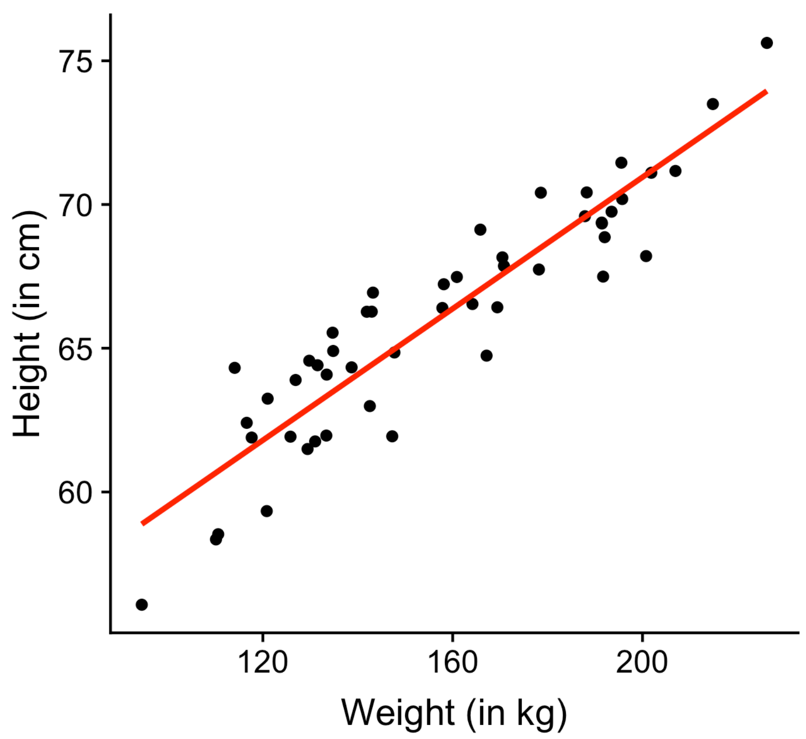
Linear model

$$y = \beta_0 + \beta_1 x + \epsilon$$
$$\epsilon \sim \mathcal{N}\left(0, \sigma^2\right)$$



Linear model

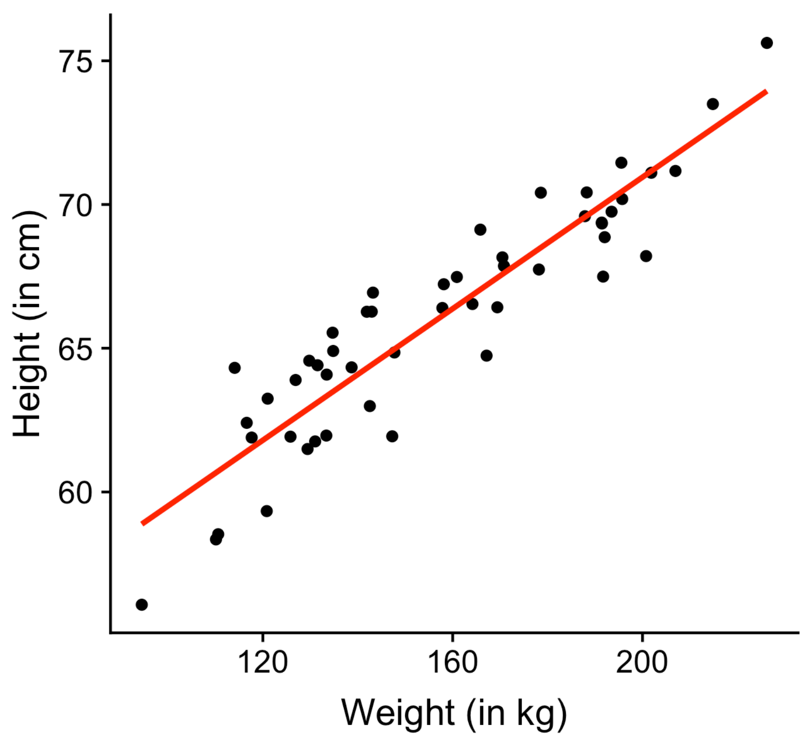
$$y \sim \mathcal{N}\left(\beta_0 + \beta_1 x, \sigma^2\right)$$



Normal
probability density

Linear model

$$p\left(y|\theta=\{\beta_0,\beta_1,\sigma\}\right)=f\left(y|\beta_0+\beta_1 x,\sigma^2\right)$$



Normal
probability density

Linear model

$$p\left(y|\theta = \{\beta_0, \beta_1, \sigma\}\right) = f\left(y|\beta_0 + \beta_1 x, \sigma^2\right)$$

Given *n* data points (x_i, y_i), *i=1..n*

$$p\left(y_i|\theta\right) = \prod_{i=1}^{n} p\left(\left(y_i, x_i\right)|\beta_0, \beta_1, \sigma\right)$$

$$= \left(2\pi\sigma^2\right)^{\left(-\frac{n}{2}\right)} \prod_{i=1}^{n} e^{-\frac{y_i - \left(\beta_0 + \beta_1 x_i\right)}{2\sigma^2}}$$

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

- 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

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

- 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 England win the world cup and the final is decided on penalties

$$p(\mathrm{W} = \mathrm{England}, \mathrm{FD} = \mathrm{penalties}) = p(\mathrm{England}, \mathrm{penalties})$$

- 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 England win the world cup, if the final is decided on penalties

$$p(\mathrm{W} = \mathrm{England}|\mathrm{FD} = \mathrm{penalties}) = p(\mathrm{England}|\mathrm{penalties})$$

- Conditional probabilities are related to joint probabilities as

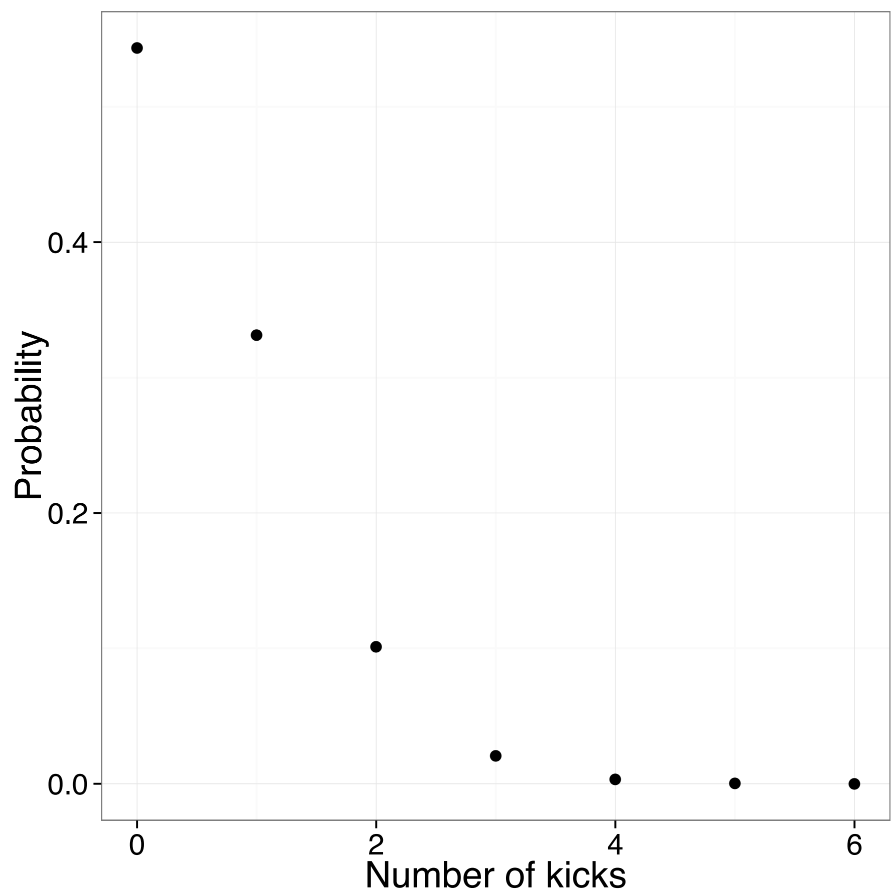
$$p(a|b) = \frac{p(a, b)}{p(b)}$$

- 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

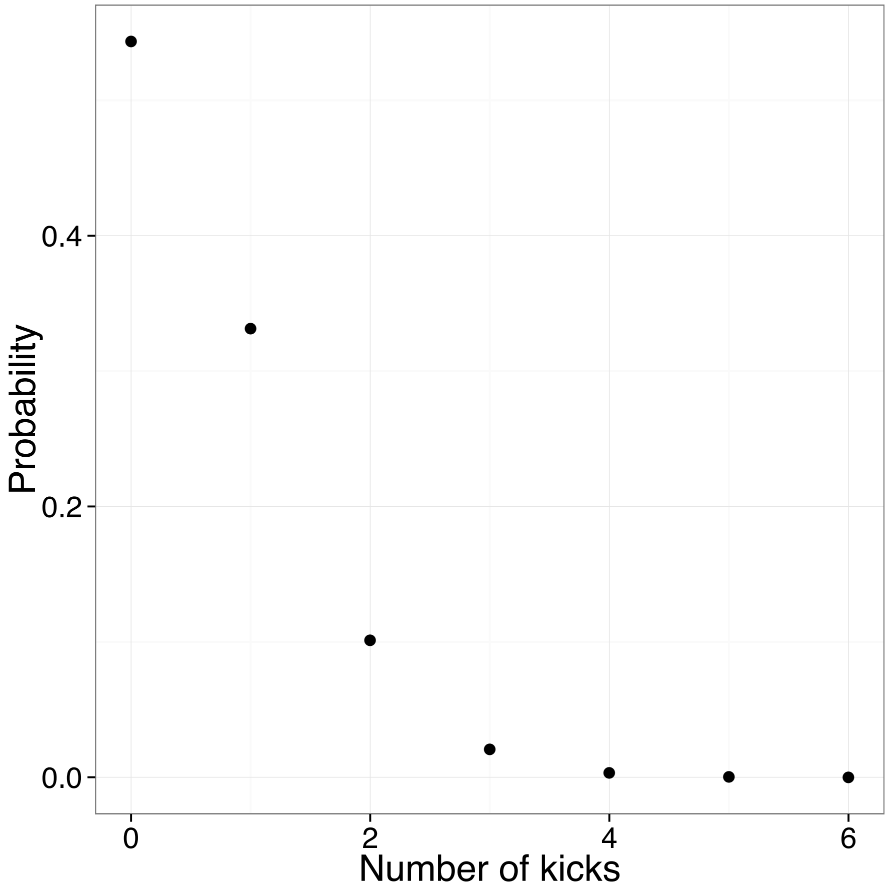


Two directions

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?

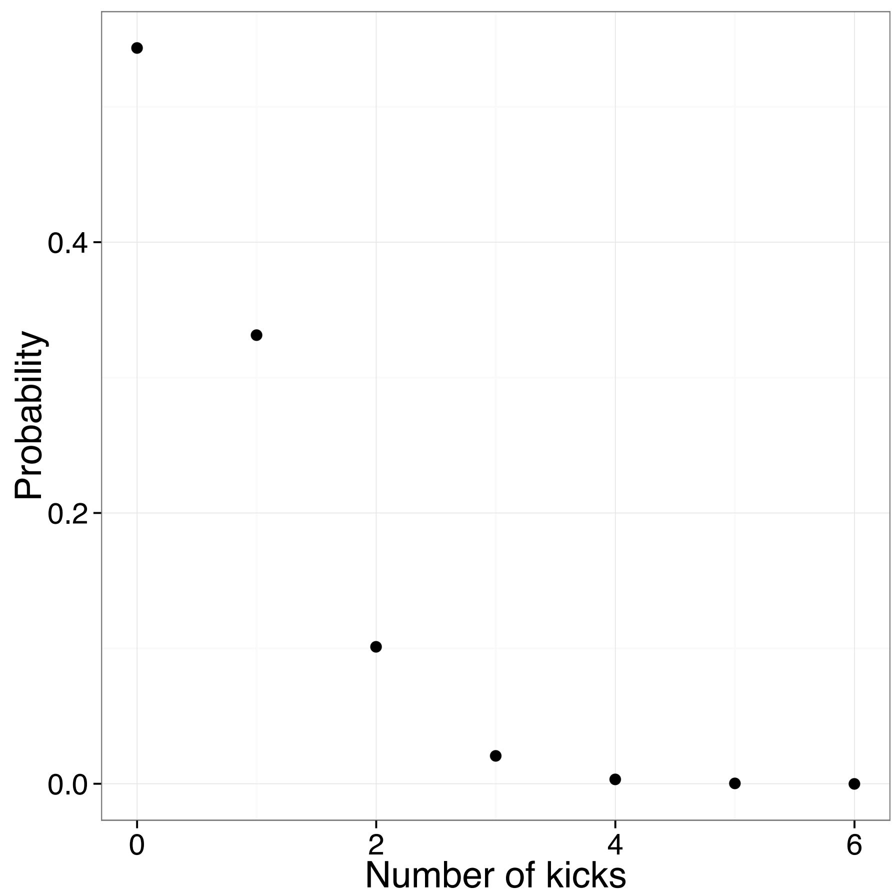
```
dpois(x = 2,
      lambda = 0.61)
```

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

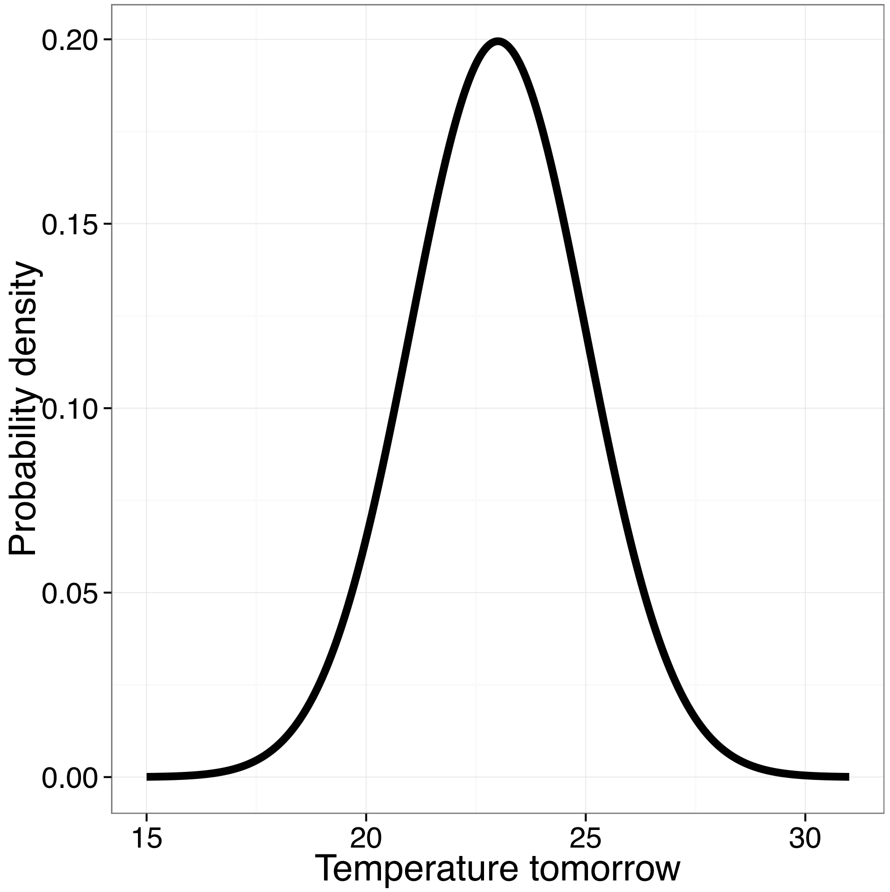
```
rpois(n = 1,
      lambda = 0.61)
```

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:

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

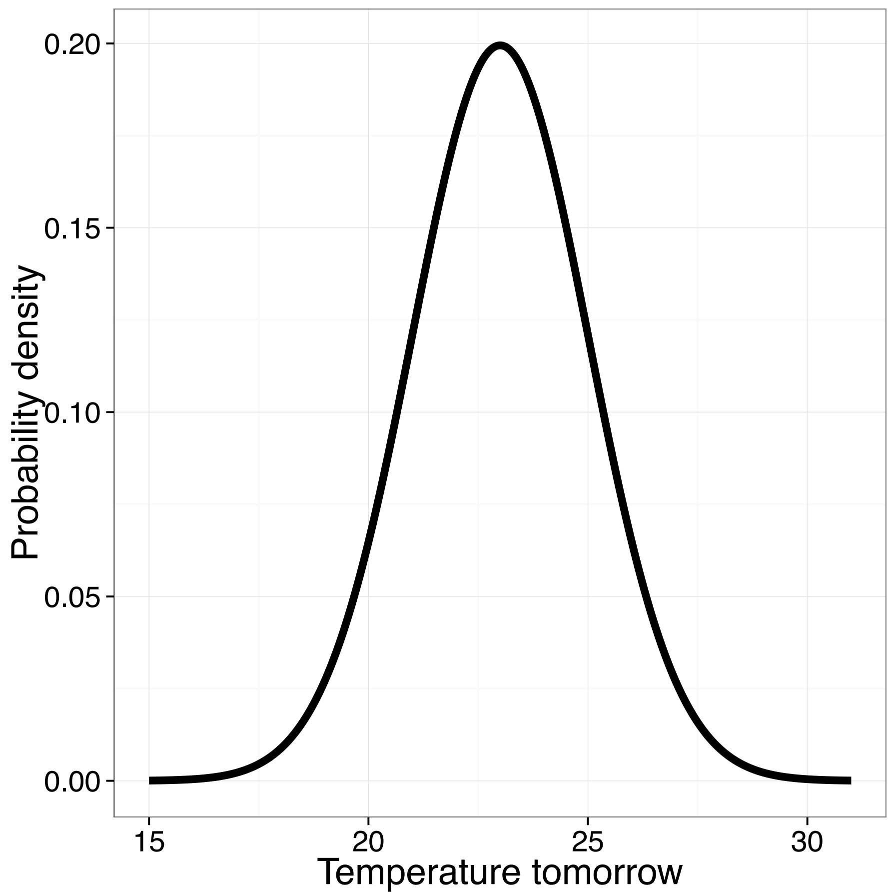
Marginal probabilities:

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

Two directions

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

Evaluating under the (normal) probability distribution



Evaluate

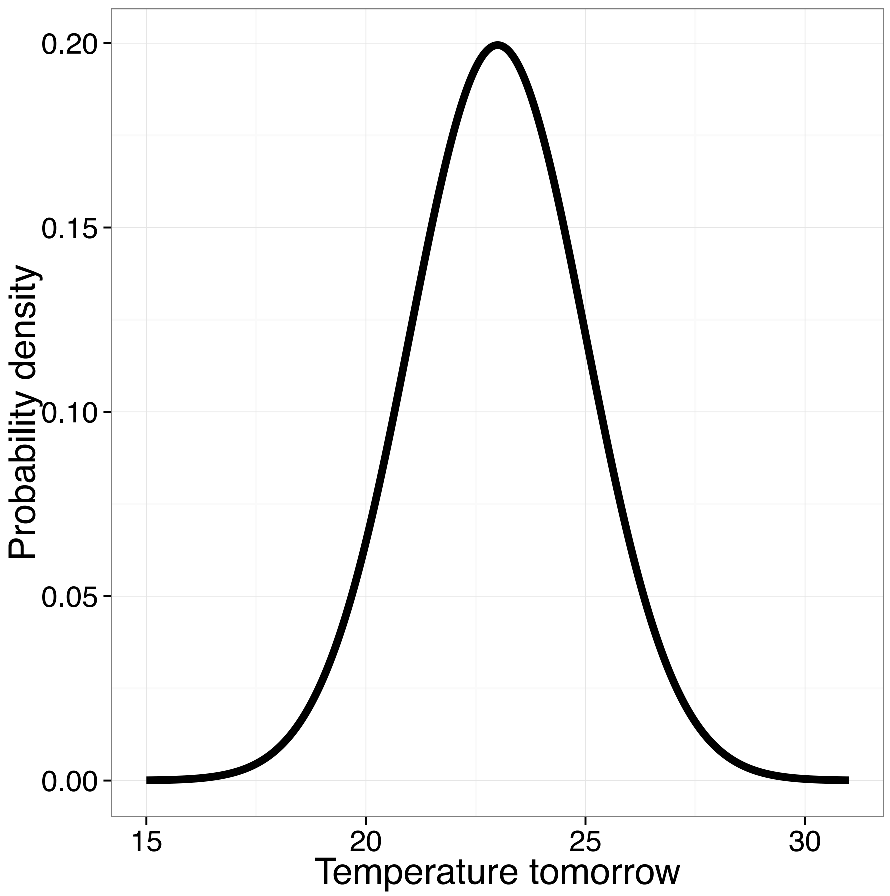
What is the probability density of $30^\circ C$ tomorrow?

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

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)
```

Two directions

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

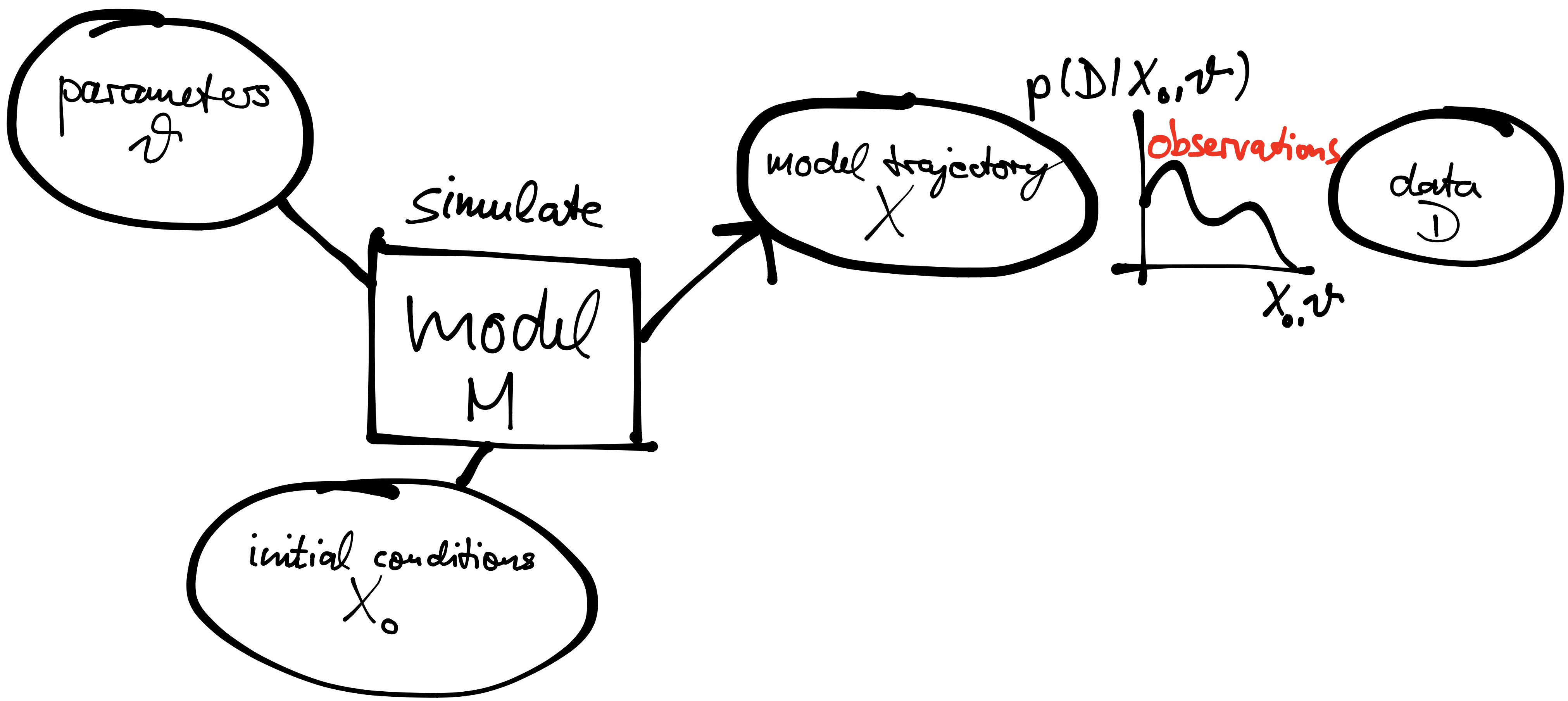
Probabilistic formulation

- 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

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

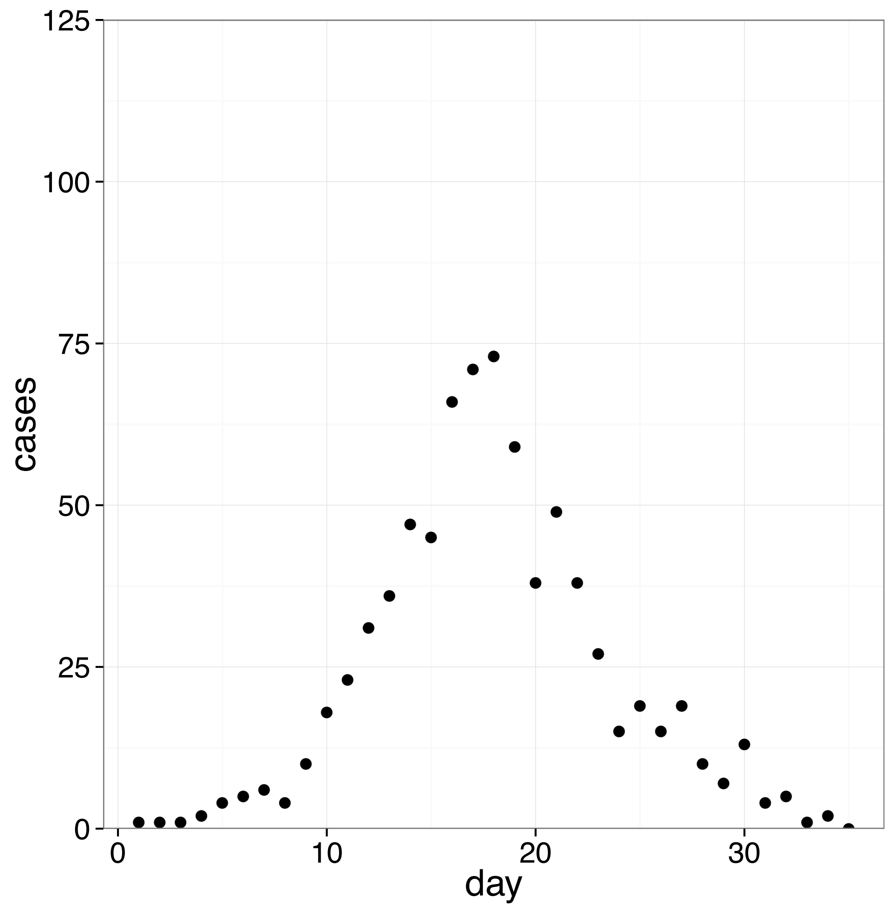
- By including this in our model, we get

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


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

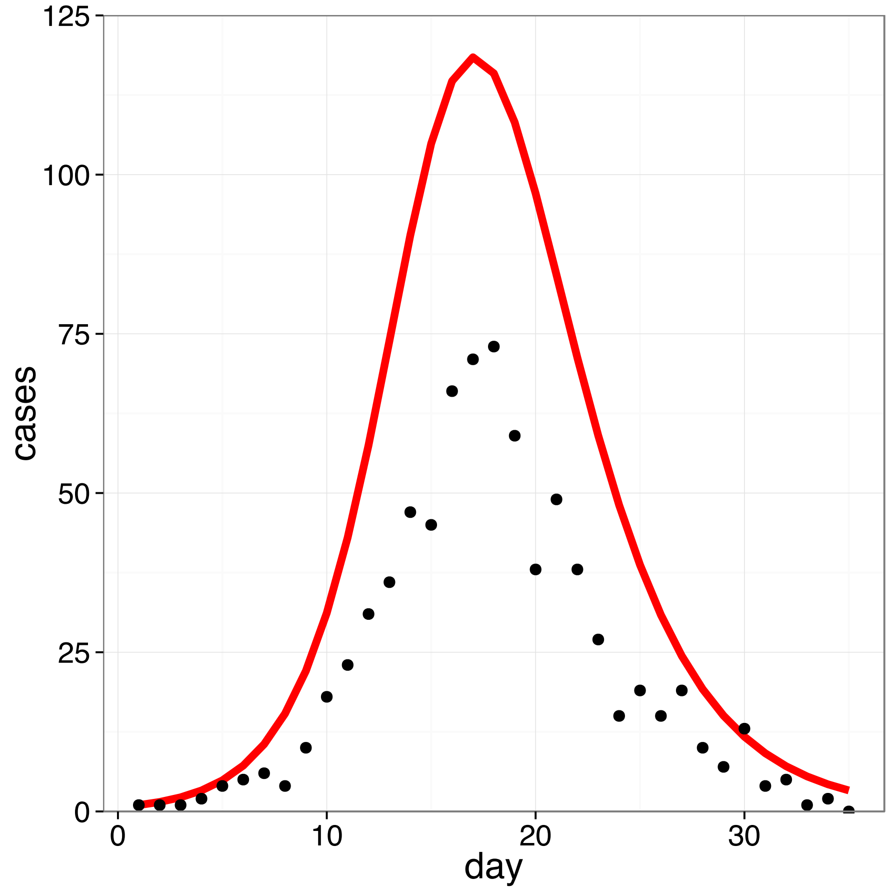
Example: observation uncertainty

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



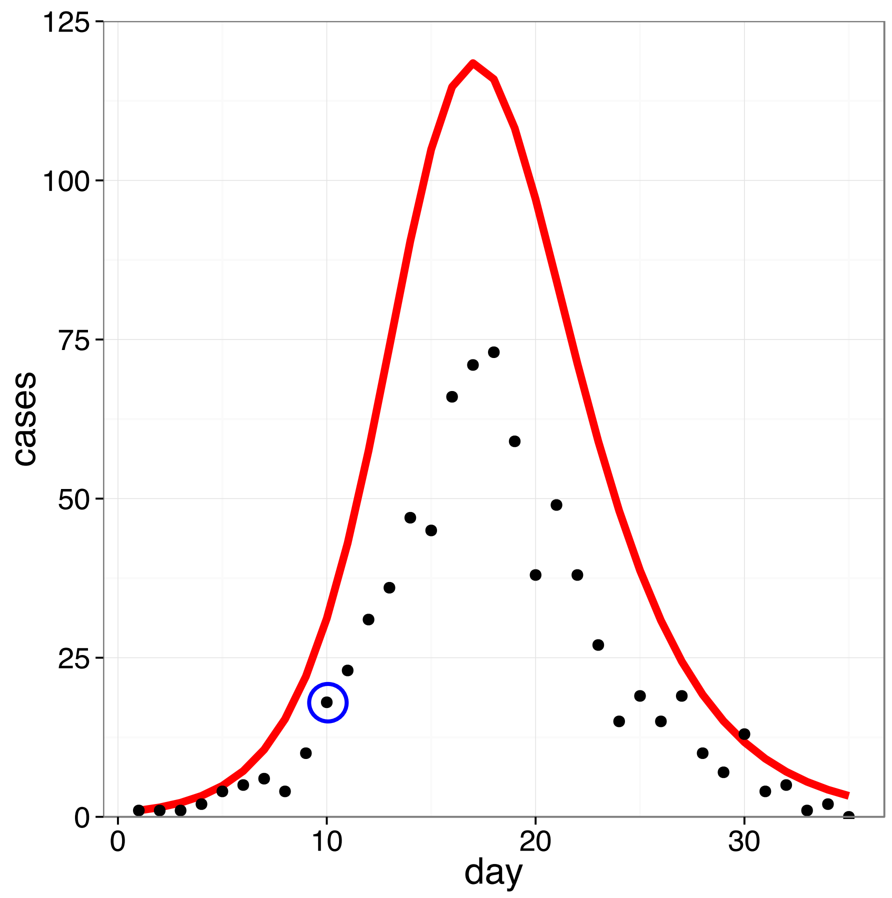
Example: observation uncertainty

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



Example: observation uncertainty

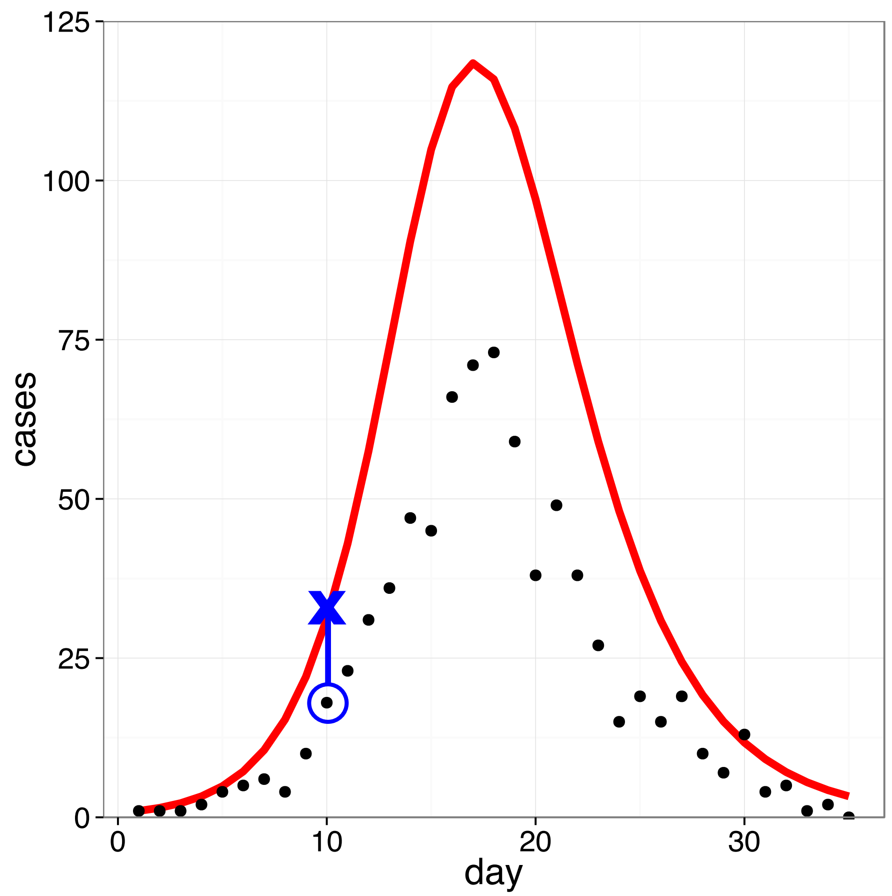
SIR model, assume that cases are detected with independent reporting probability $\rho = 0.5$.



At time 10, 18 cases observed.

Example: observation uncertainty

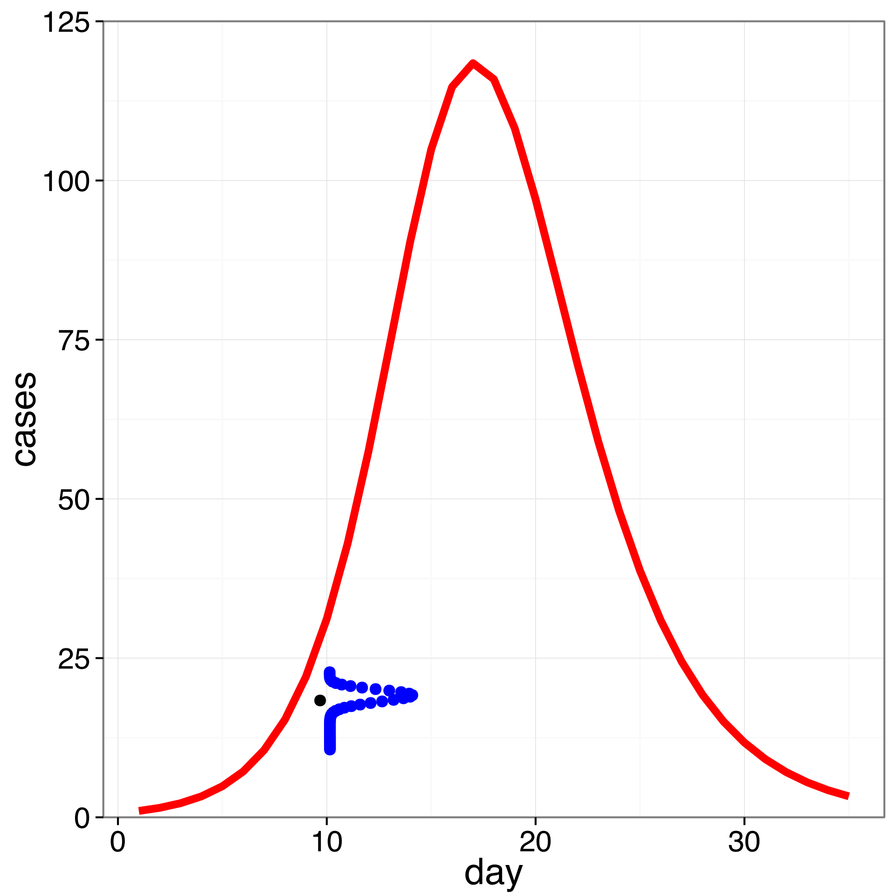
SIR model, assume that cases are detected with independent reporting probability $\rho = 0.5$.

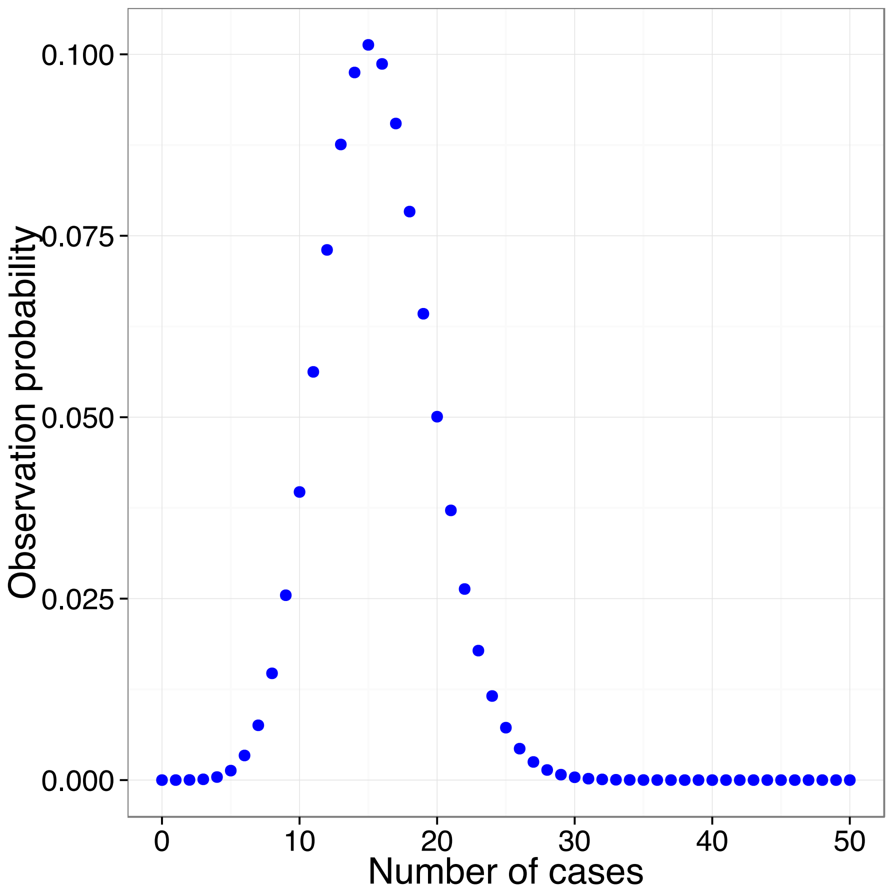


At time 10, 18 cases observed, 31.1 cases in the model.

Example: observation uncertainty

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

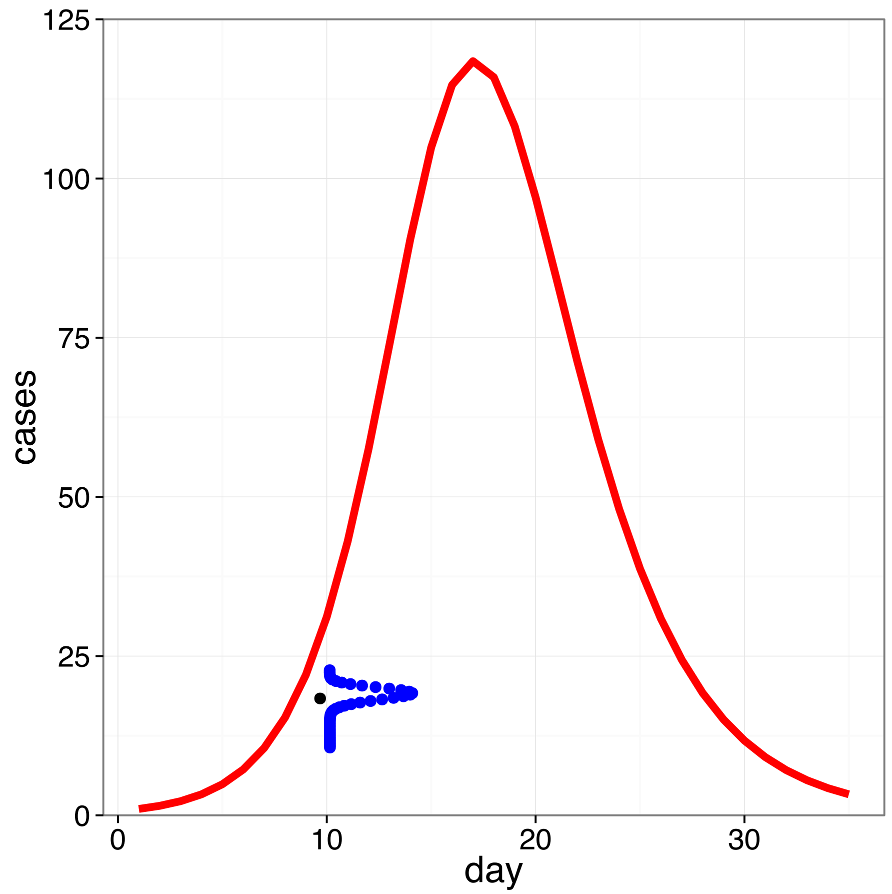


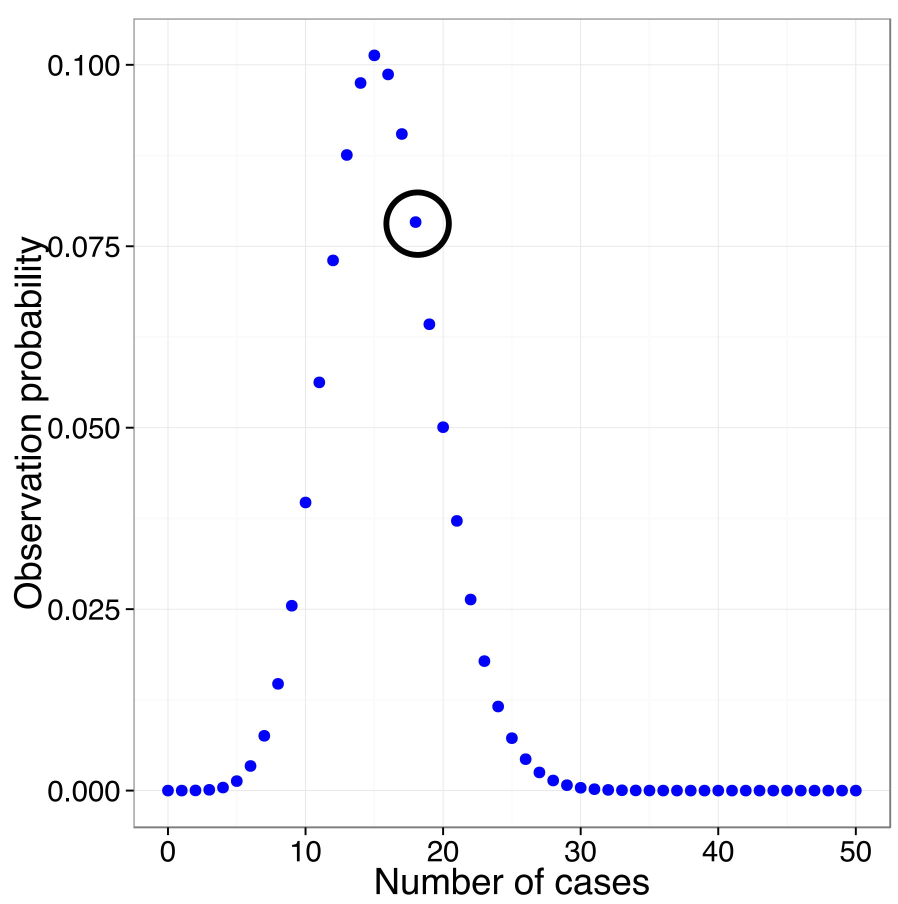


At time 10, 18 cases observed, 31.1 cases in the model.

Example: observation uncertainty

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



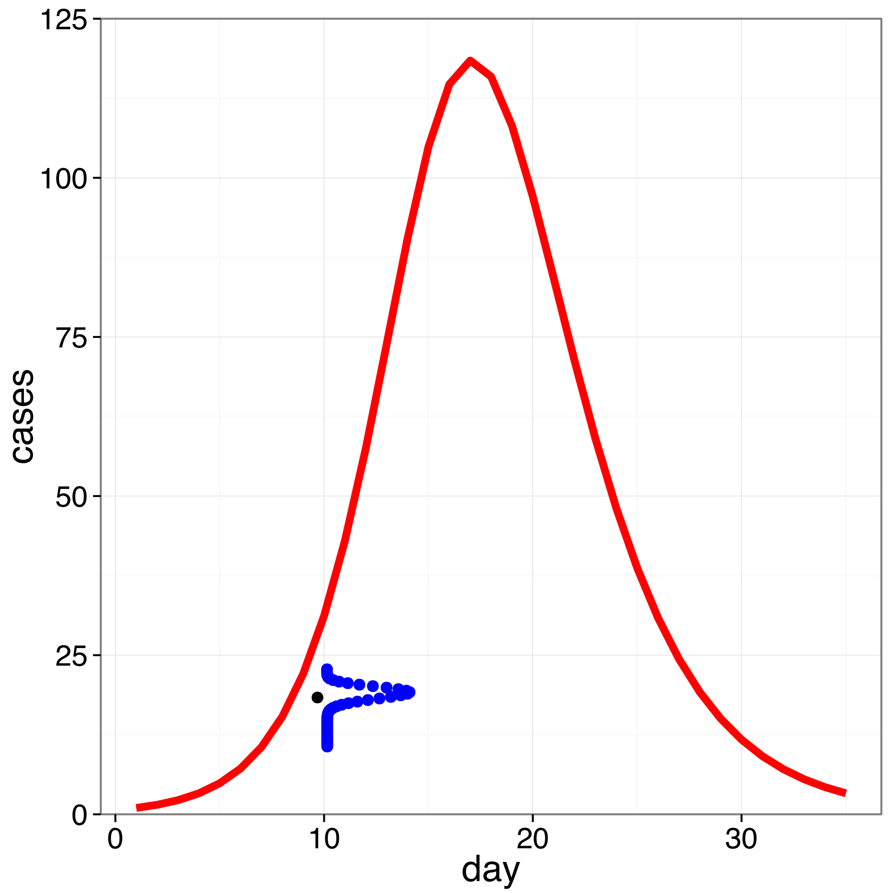


At time 10, 18 cases observed, 31.1 cases in the model.

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

Example: observation uncertainty

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

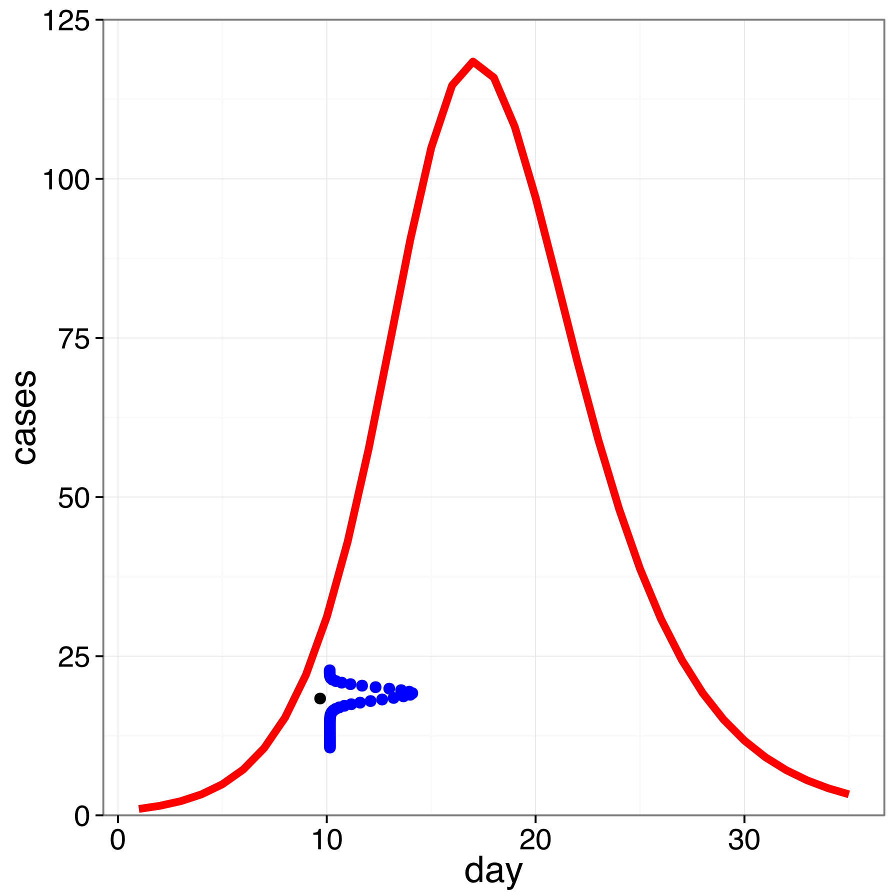


Multiply across the data to get the probability of the whole trajectory.

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

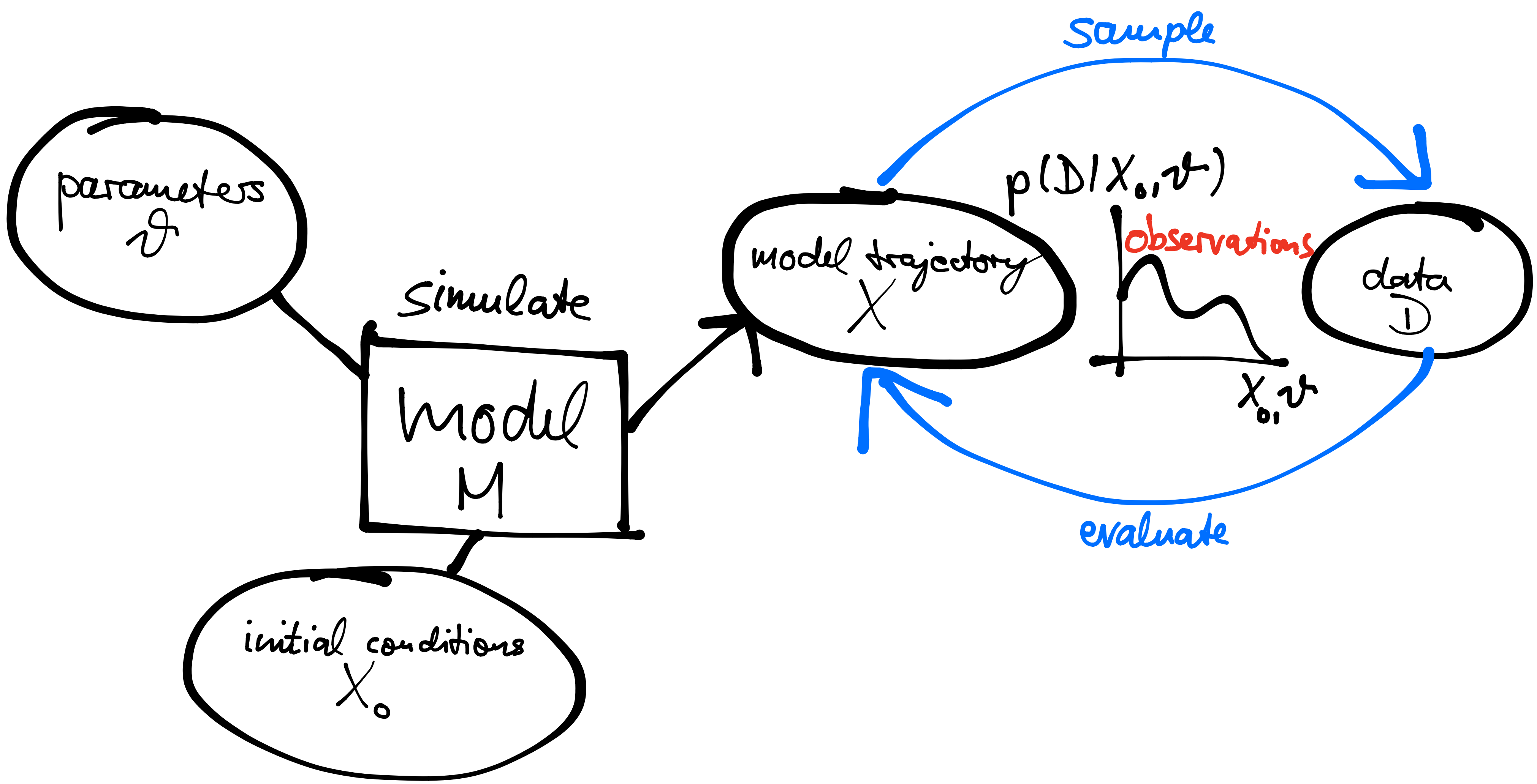
Example: observation uncertainty

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



Sum across the data to get the probability of the whole trajectory.

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



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

The likelihood

- We compare models to data using probabilities

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

- For a given model this depends on the parameters θ.

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

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

- 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

$$\log L(\theta) = \sum_i \log L(\theta)$$

Normal
probability density

Linear model

$$p\left(y|\theta=\{\beta_0,\beta_1,\sigma\}\right)=f\left(y|\beta_0+\beta_1 x,\sigma^2\right)$$

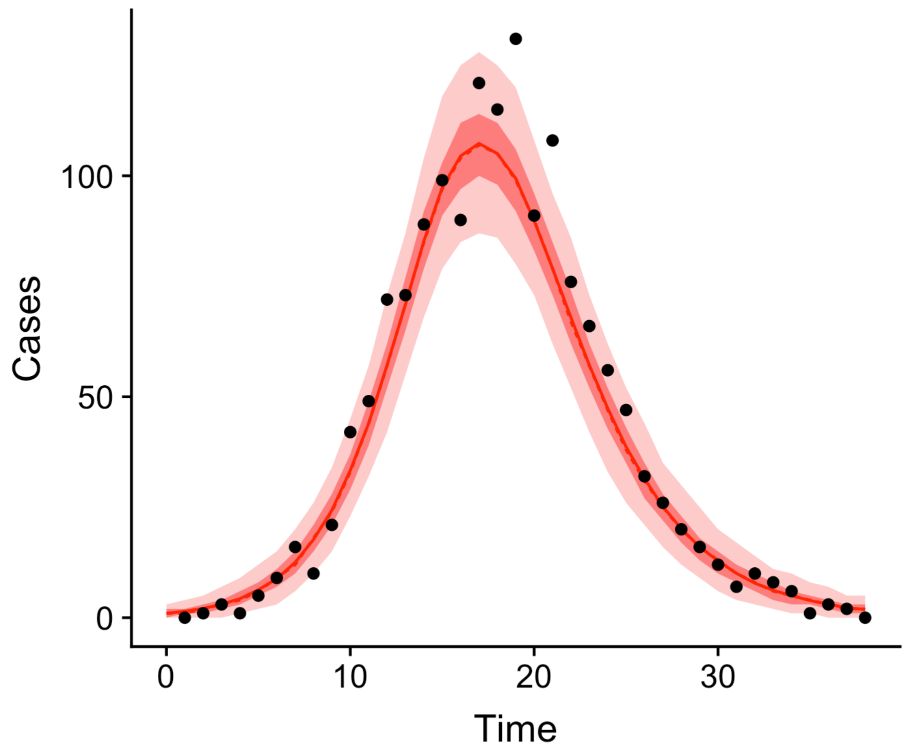
Given *n* data points (x_i, y_i), *i=1..n*

$$p\left(y_i|\theta\right)=\prod_{i=1}^{n}p\left(\left(y_i,x_i\right)|\beta_0,\beta_1,\sigma\right)$$

$$=\left(2\pi\sigma^2\right)^{\left(-\frac{n}{2}\right)}\prod_{i=1}^{n}e^{-\frac{y_i-\left(\beta_0+\beta_1 x_i\right)}{2\sigma^2}}$$

Infectious disease model

$$p\left(y|\theta\right) = ???$$



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: $L(\theta) = p(\mathrm{data}|\theta)$
- in inference, one tries 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: $L(\theta) = p(\mathrm{data}|\theta)$
- in inference, one tries 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,

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

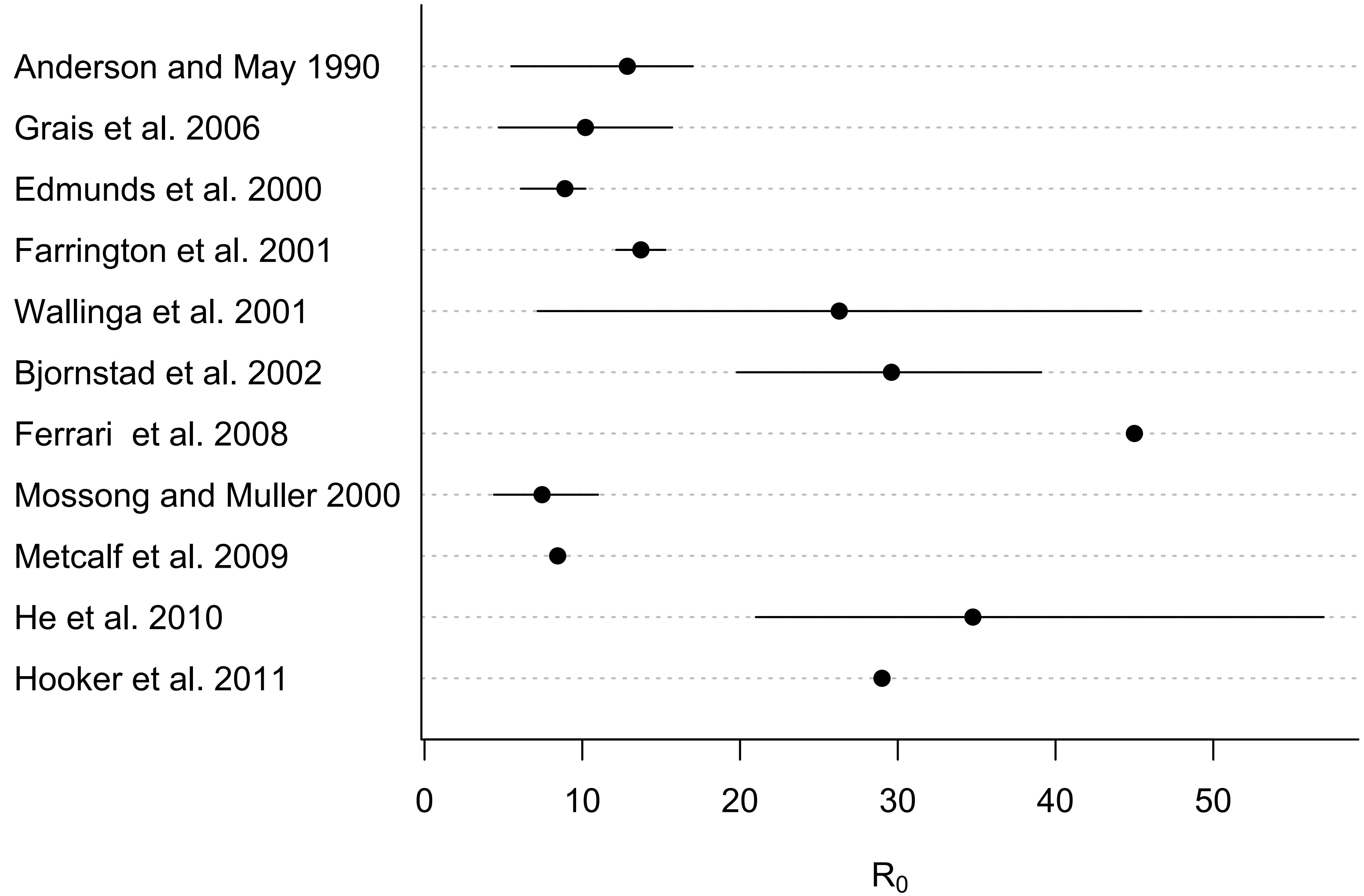
Prior probabilities

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

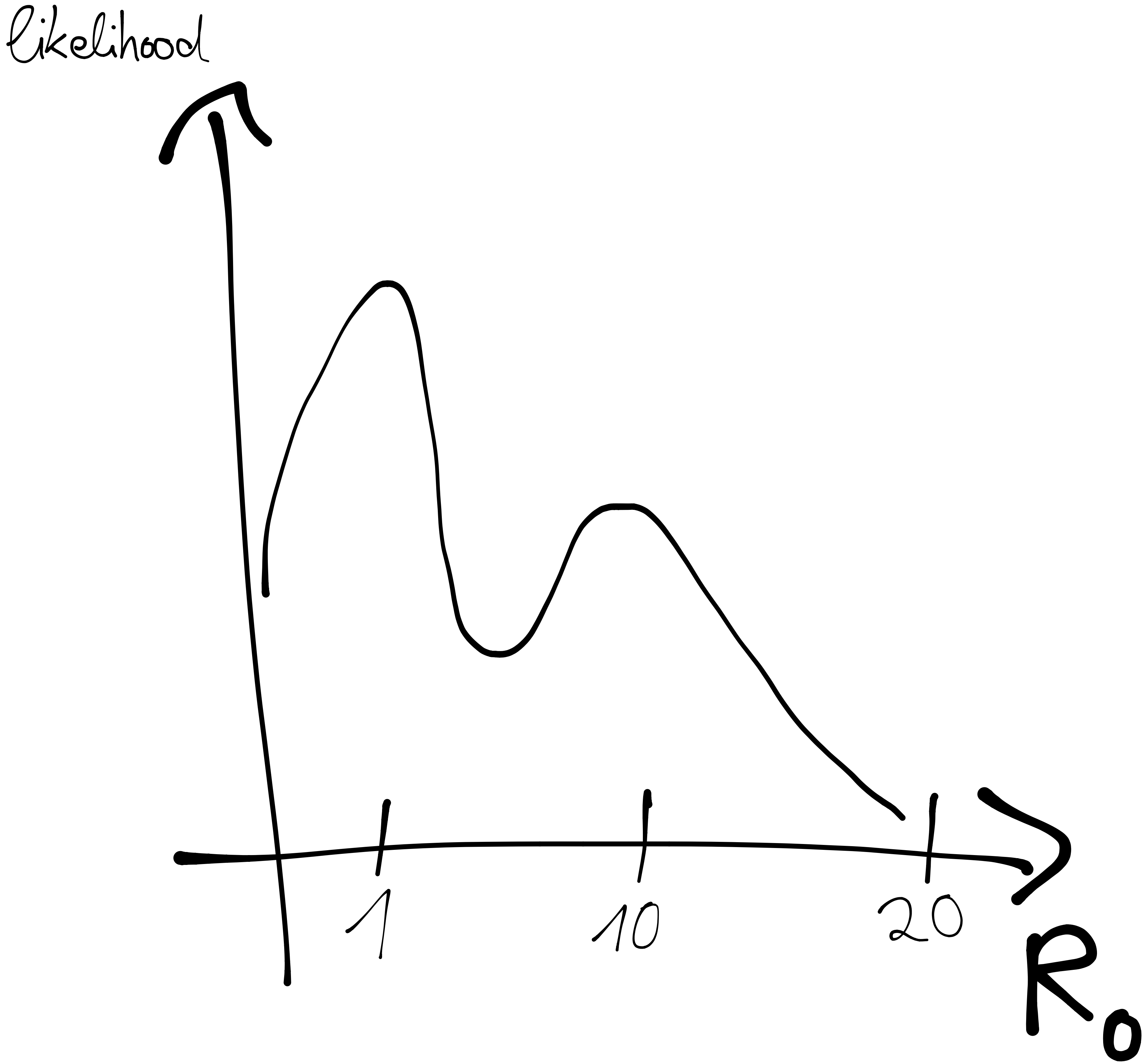
$$p(\theta)$$

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

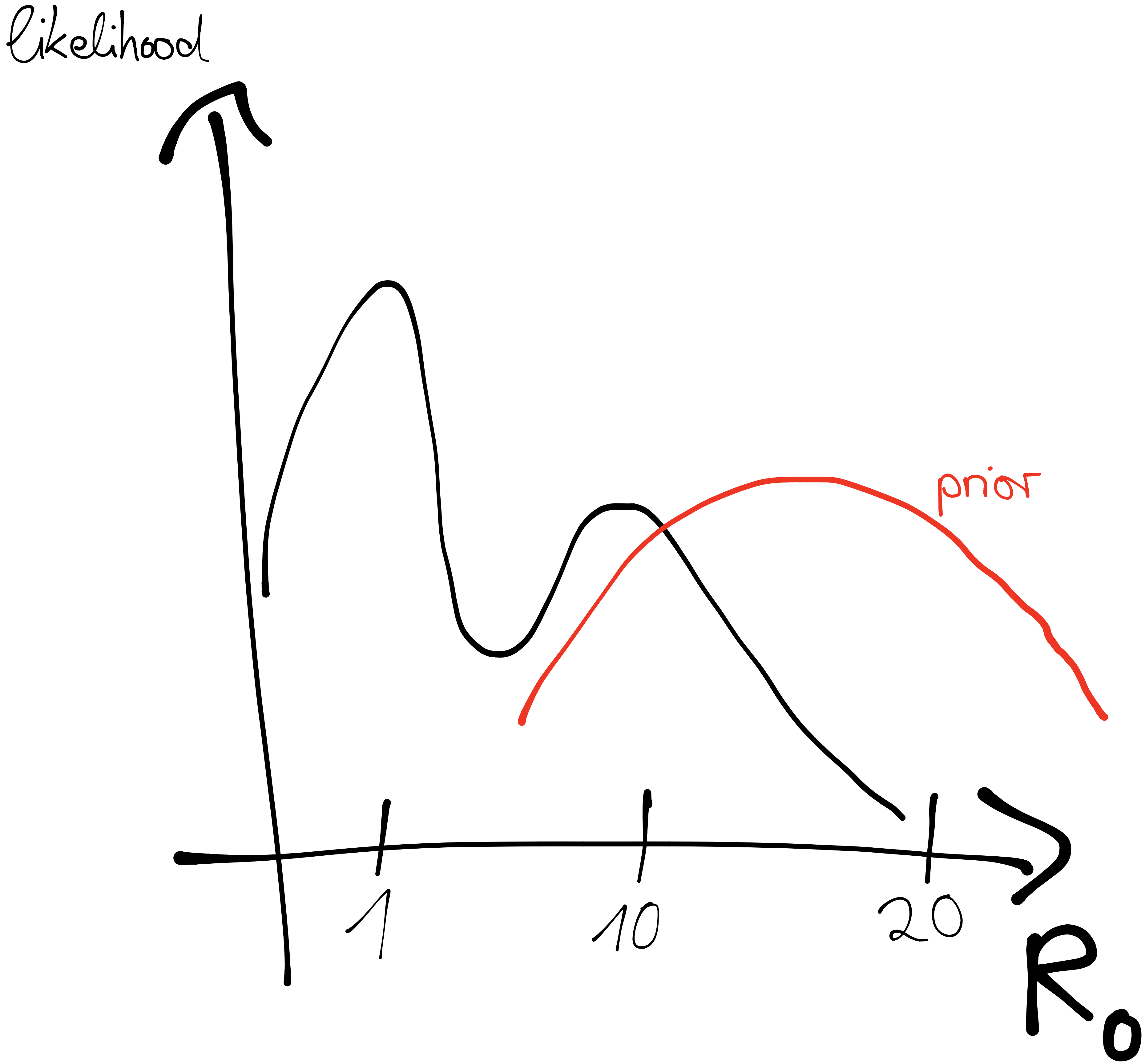
- Example: R_0 of measles



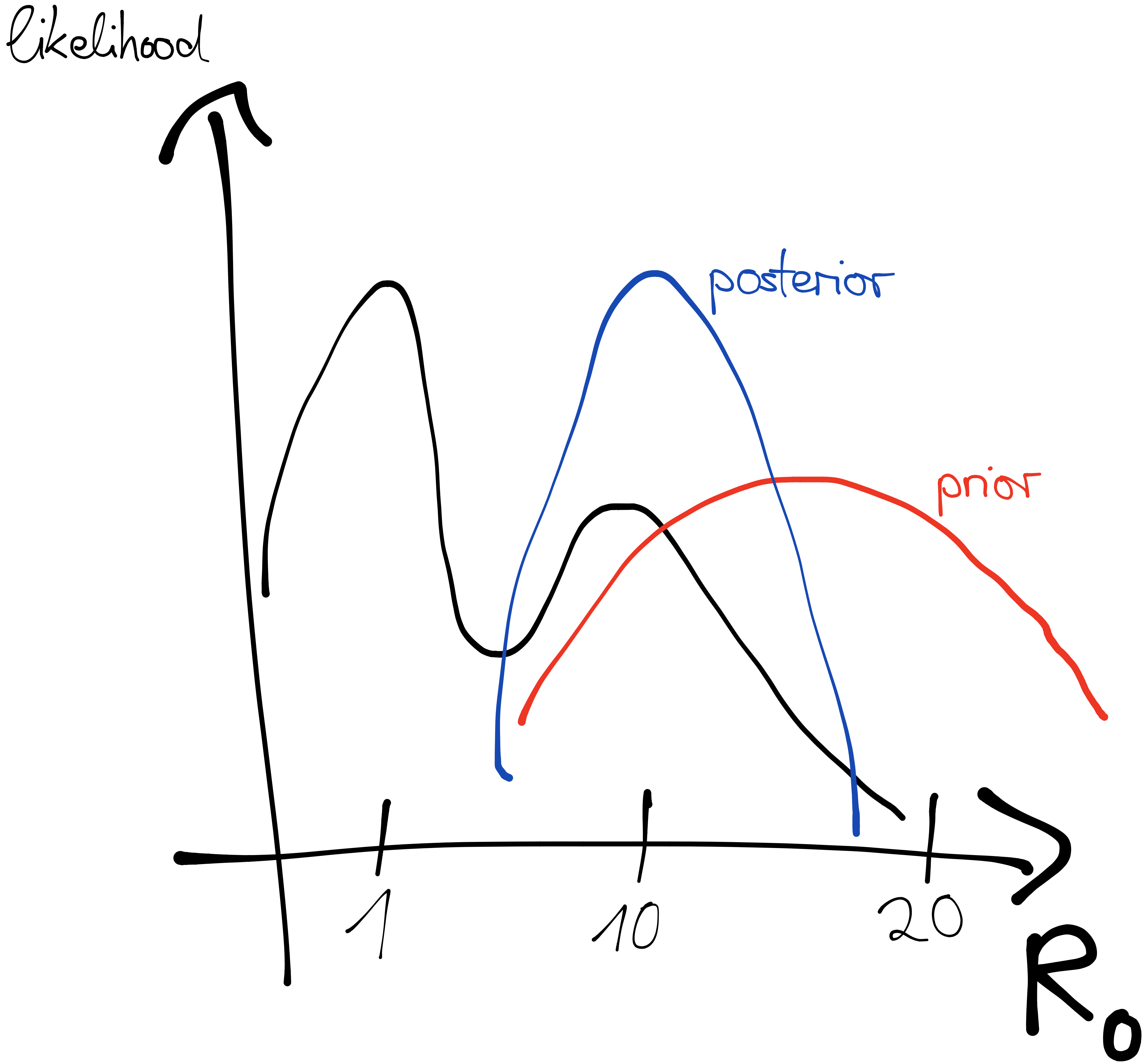
Example: estimating R_0 of measles



Example: prior for estimating R_0 of measles



Example: posterior for estimating R_0 of measles



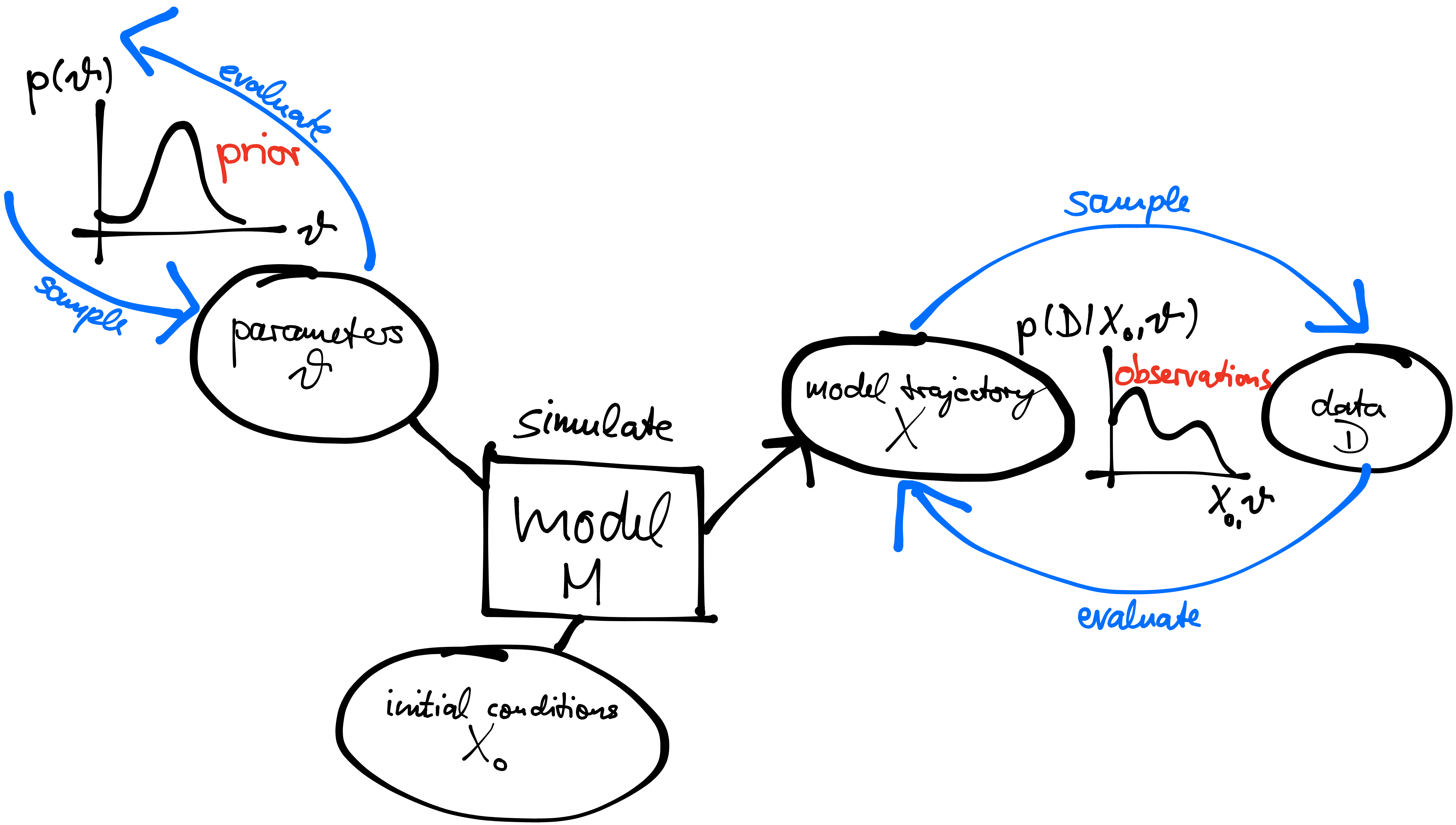
Sampling from the posterior

Bayesian statistics

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

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

We want to generate samples of θ from this distribution.



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

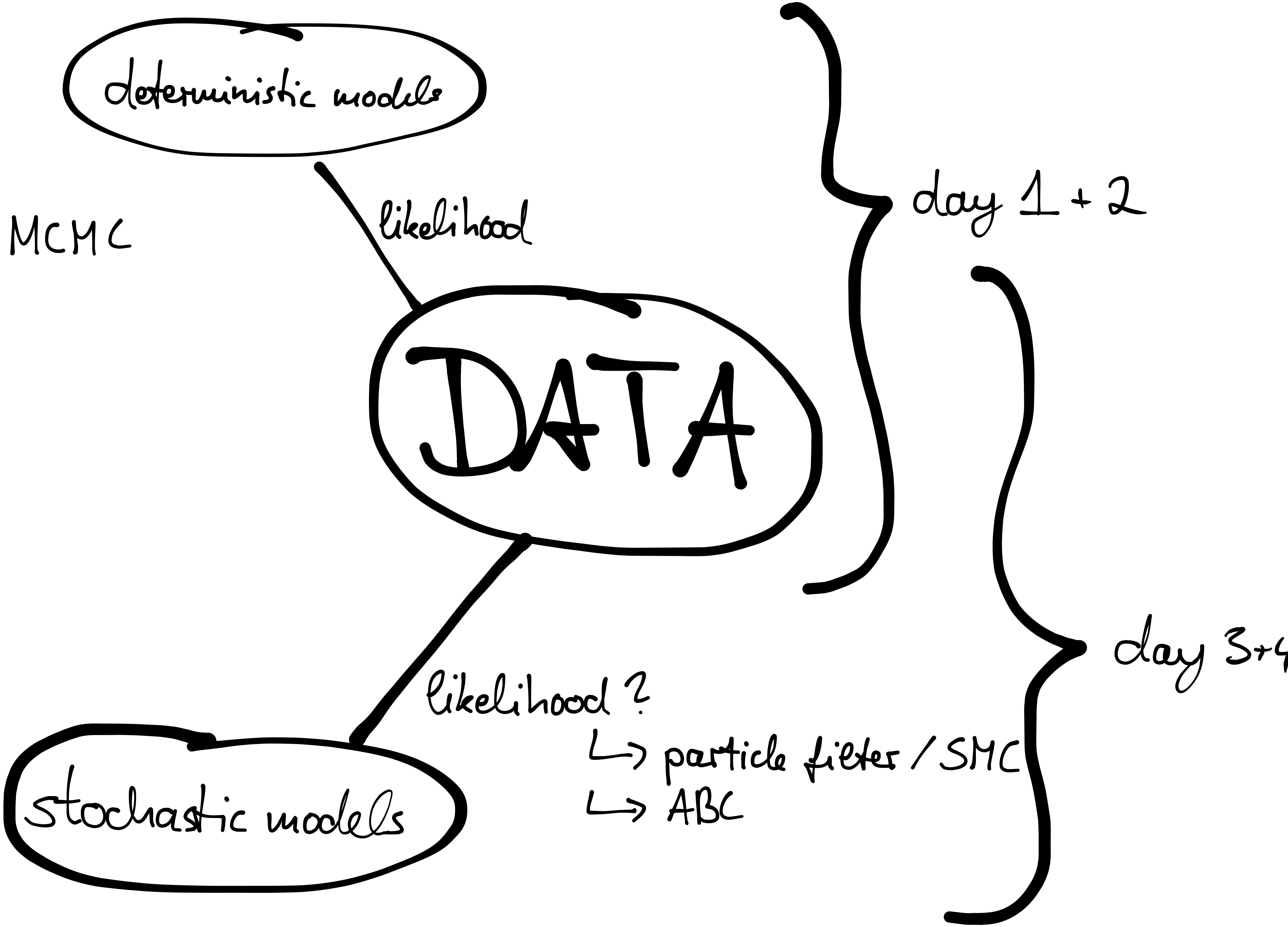
deterministic models

likelihood

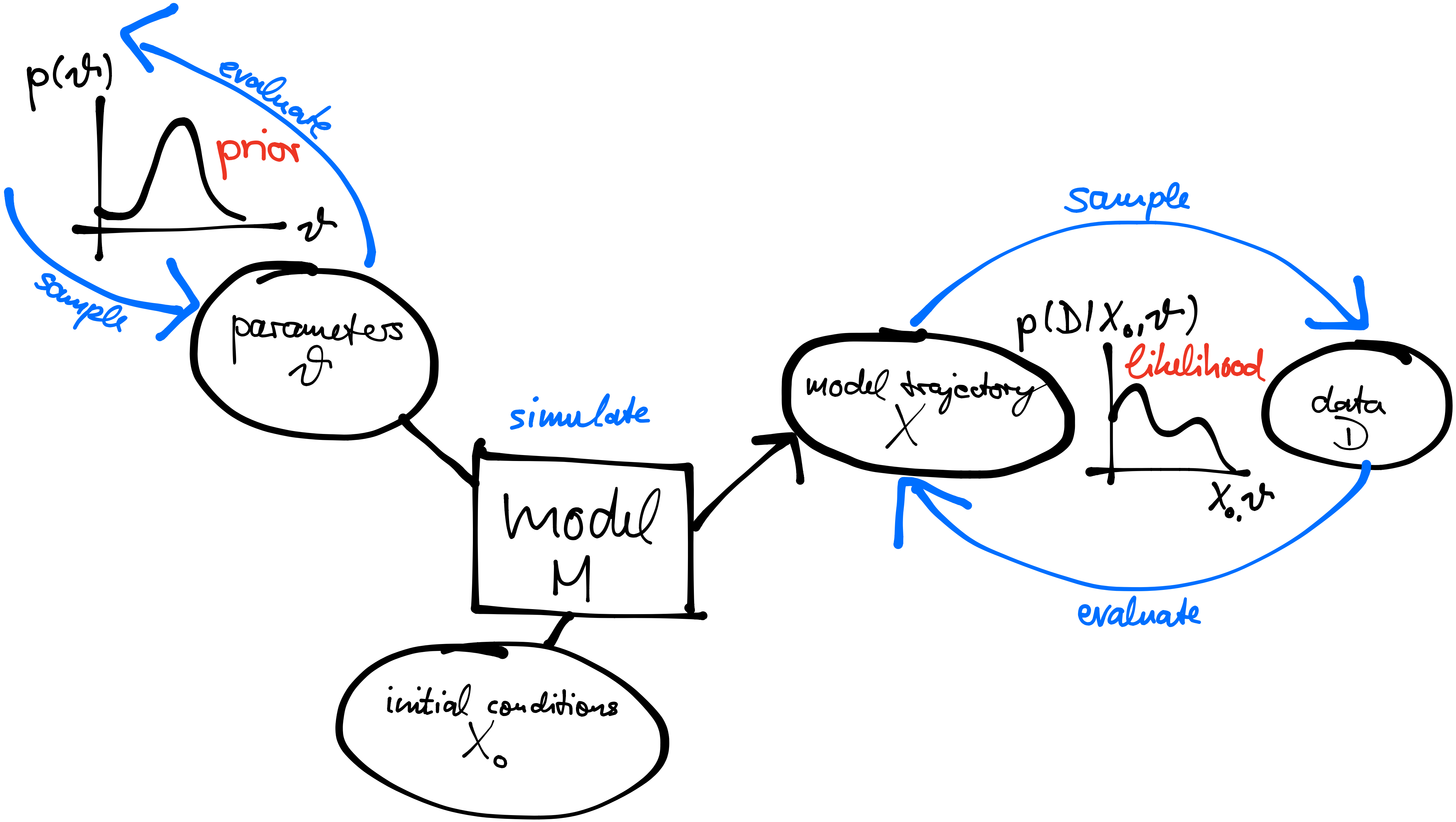
MCMC

DATA

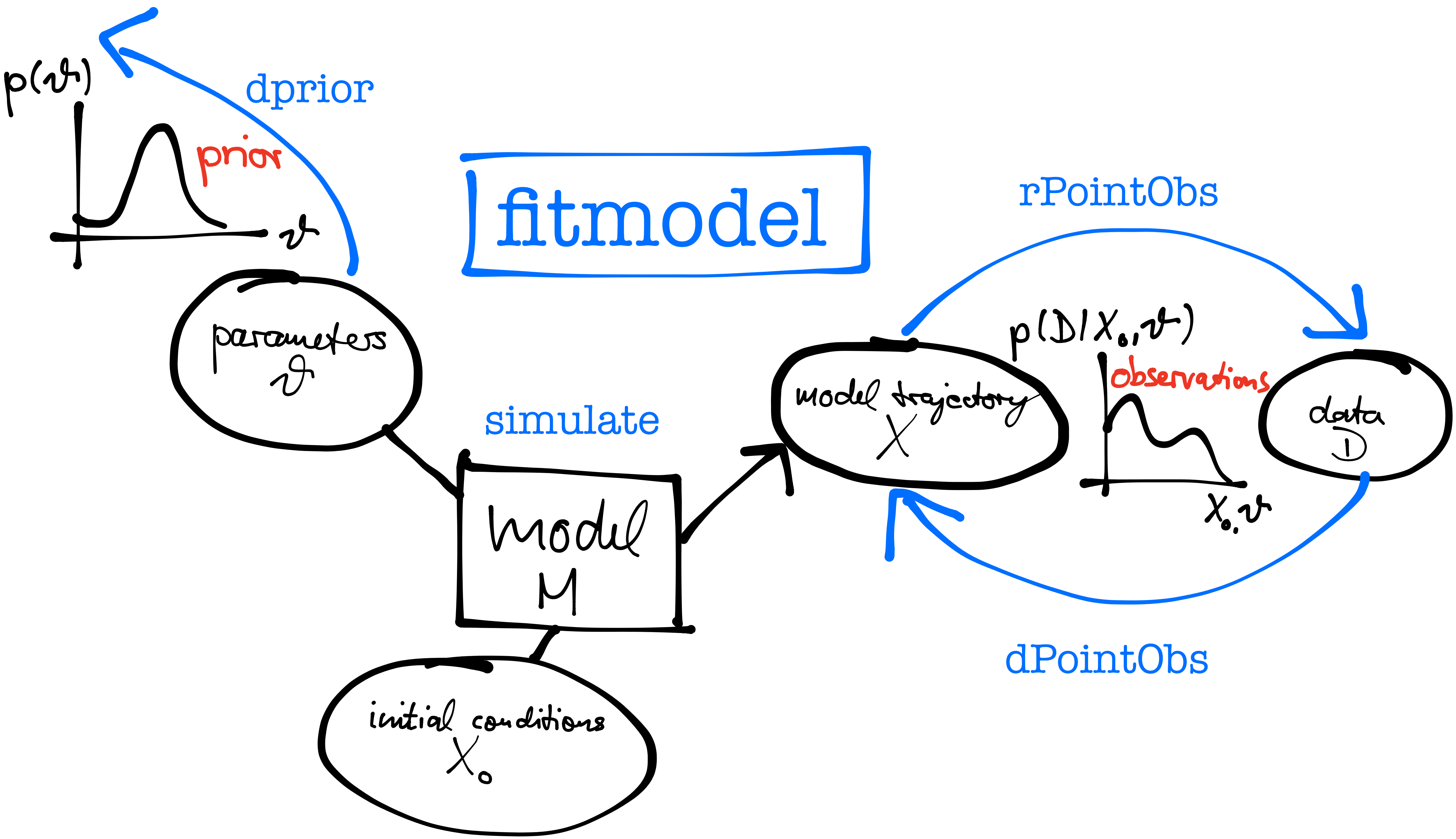
day 1 + 2

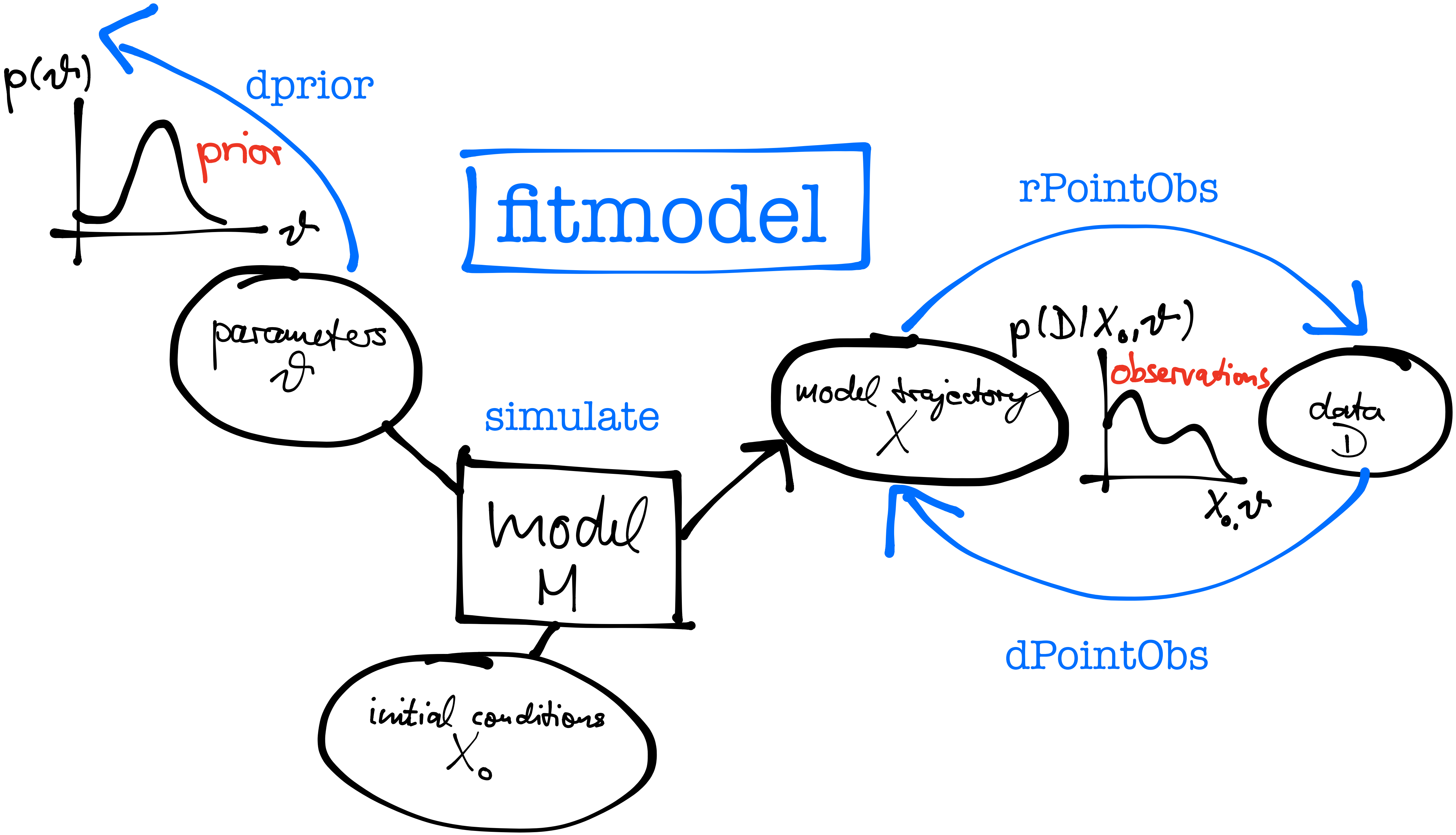

deterministic models
likelihood
MCMC
DATA
day 1 + 2
likelihood ?
↳ particle filter / SMC
↳ ABC
stochastic models
day 3+4

4. Practical session in R



p(ϑ)
prior
ϑ
evaluate
sample
parameters ϑ
simulate
Model M
initial conditions X_0
model trajectory X
sample
p(D | X_0, ϑ)
likelihood
X_0, ϑ
data D
evaluate


fitmodel
dprior
p(ϑ)
prior
ϑ
parameters
ϑ
simulate
Model
M
initial conditions
X₀
rPointObs
p(D|X₀,ϑ)
observations
model trajectory
X
data
D
X₀,ϑ
dPointObs



http://sbfnk.github.io/mfiidd

5. References

Dureau, J., K. Kalogeropoulos, and M. Baguelin (July 2013). “Capturing the time-varying drivers of an epidemic using stochastic dynamical systems.”. *Biostatistics* 3, 541–555.

King, A. A. et al. (Aug. 2008). “Inapparent infections and cholera dynamics.”. *Nature* 7206, 877–880.