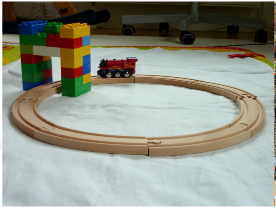
Model fitting and inference for infectious disease dynamics

Sebastian Funk, Gwen Knight, Anton Camacho
with contributions from Helen Johnson

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

1. Introduction

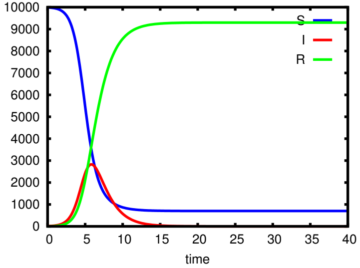
Model fitting and inference for infectious disease dynamics



Model fitting and inference for infectious disease dynamics

SIR-type models

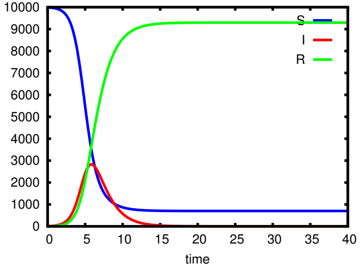
$$\frac{dS}{dt} = -\beta I \frac{S}{N}$$
$$\frac{dI}{dt} = \beta I \frac{S}{N} - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$



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

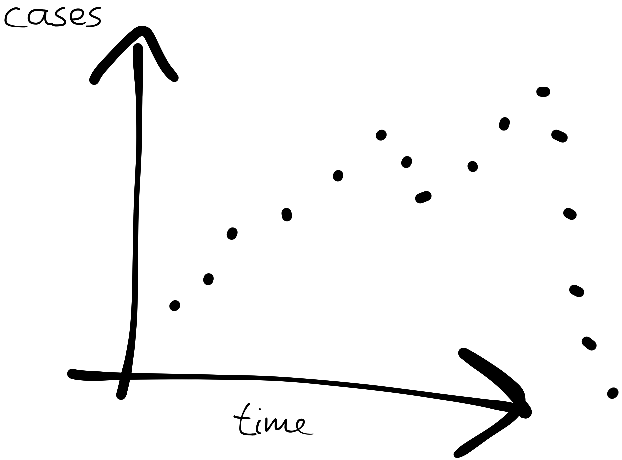
Given a model, what are the parameter combinations that best fit the data (in whichever way)

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

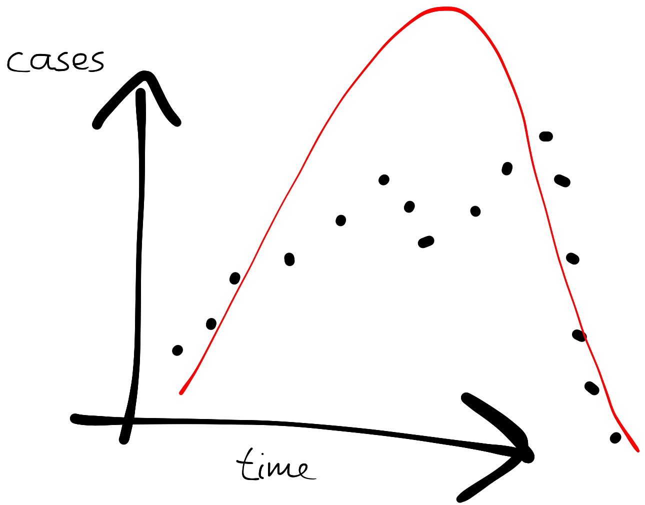
Model fitting and inference for infectious disease dynamics

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



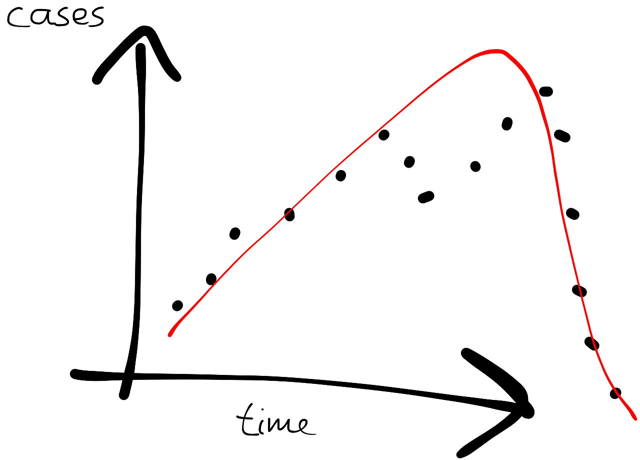
Model fitting and inference for infectious disease dynamics

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



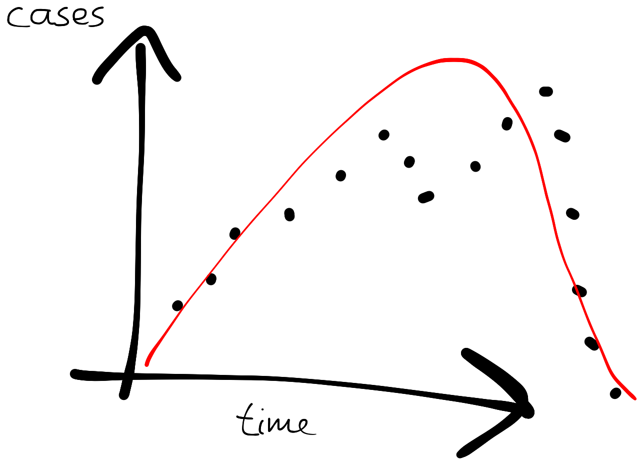
Model fitting and inference for infectious disease dynamics

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



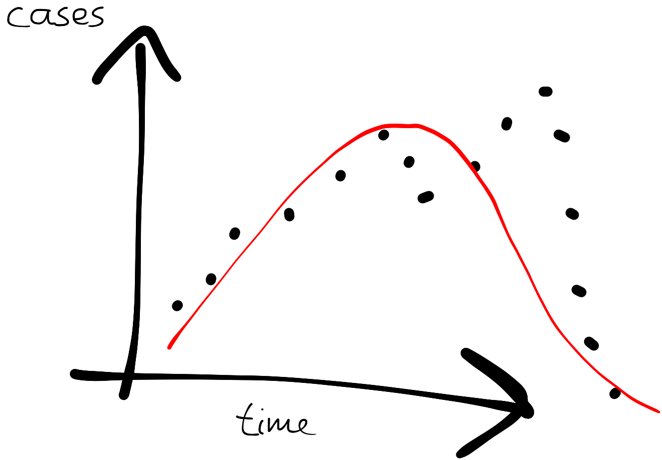
Model fitting and inference for infectious disease dynamics

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



Model fitting and inference for infectious disease dynamics

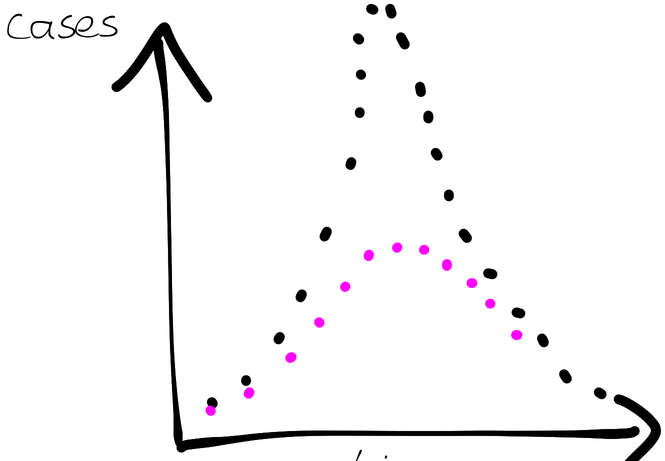
What do we mean by “best fit the data”?



Model fitting and inference for infectious disease dynamics

State estimation

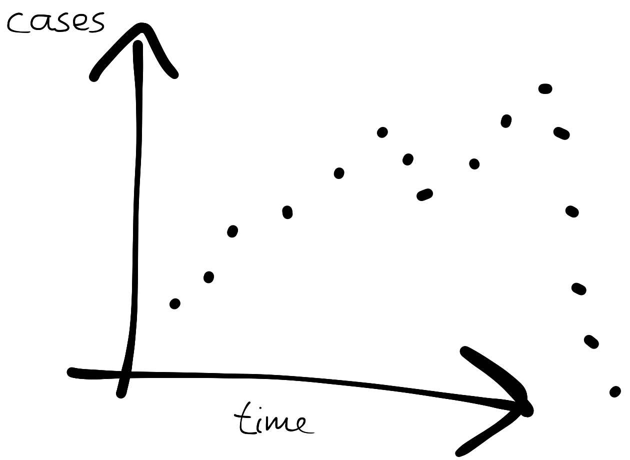
Given what we observe, what is the state of the sytem?



Model fitting and inference for infectious disease dynamics

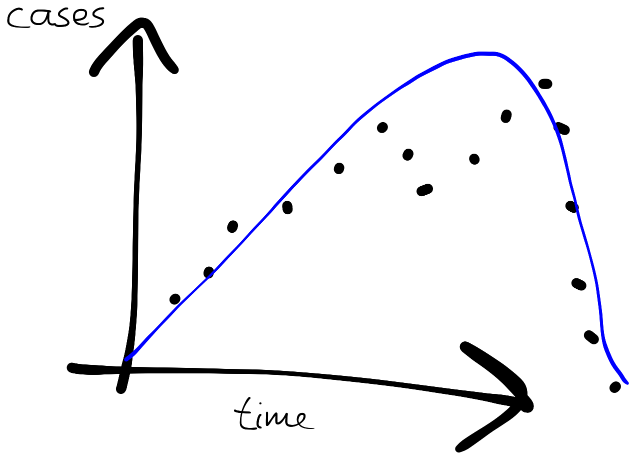
Model selection

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



Model selection

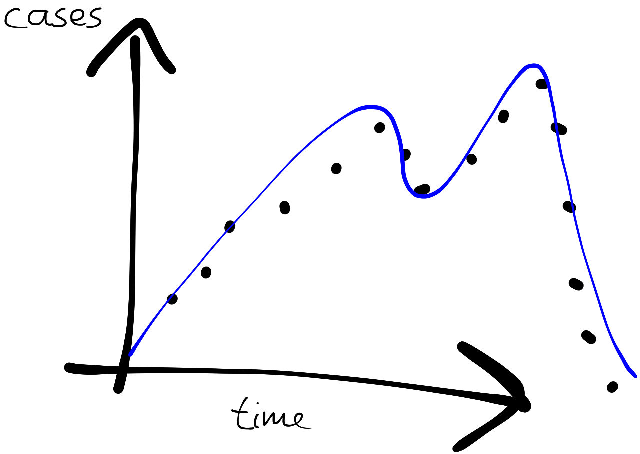
Given a set of potential models, how do we decide which is the right one?



Model fitting and inference for infectious disease dynamics

Model selection

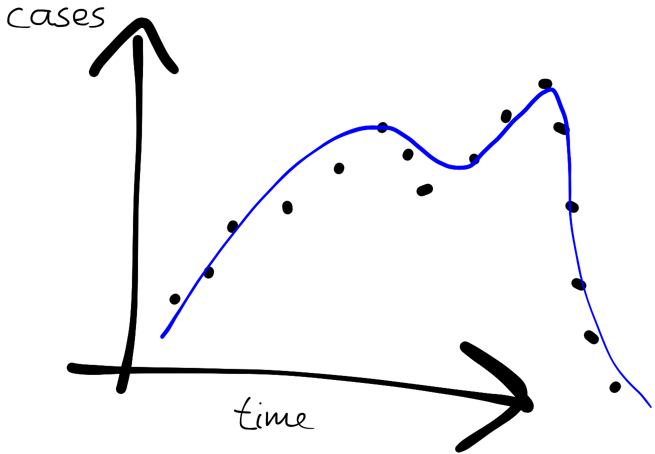
Given a set of potential models, how do we decide which is the right one?



Model fitting and inference for infectious disease dynamics

Model selection

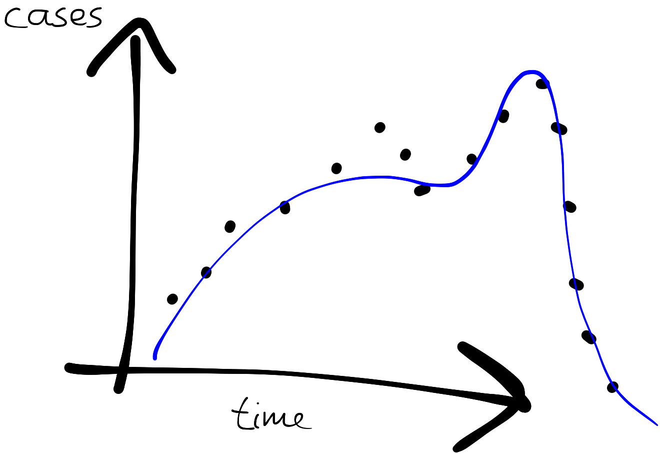
Given a set of potential models, how do we decide which is the right one?



Model fitting and inference for infectious disease dynamics

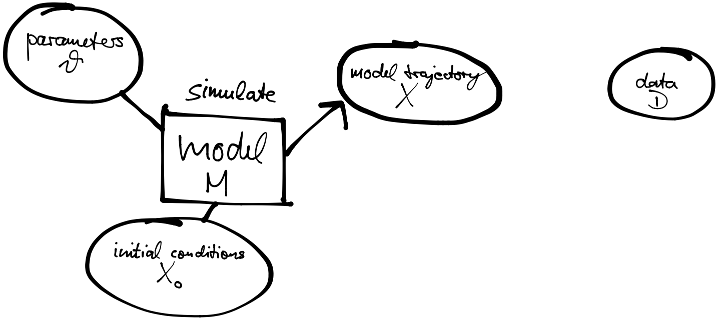
Model selection

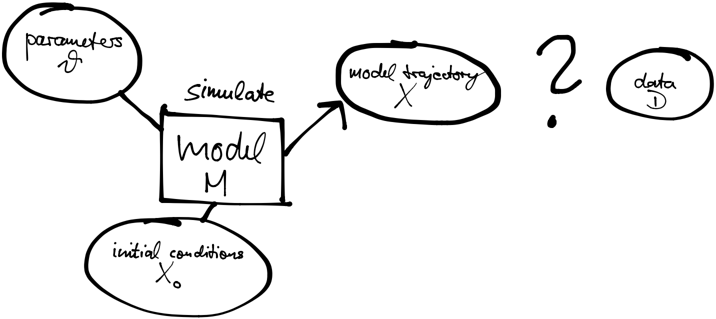
Given a set of potential models, how do we decide which is the right one?



2. Linking models to data


Model
M
data
D


parameters
ϑ
simulate
model
M
model trajectory
X
data
D
initial conditions
X0


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

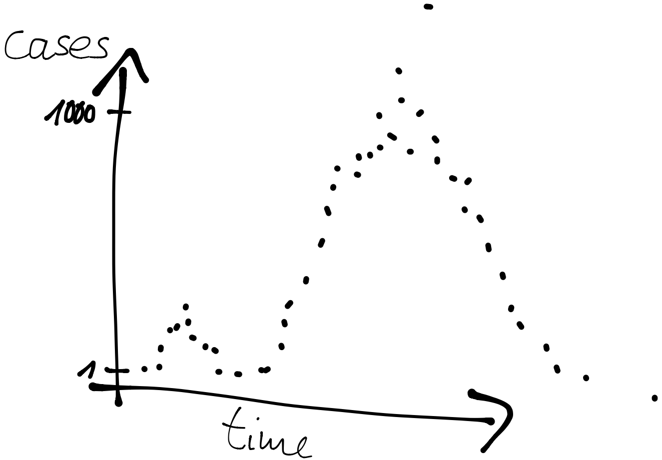
Assessing the “closeness” of model output and data

- eyeballing
- absolute distance
- squared distance

Assessing the “closeness” of model output and data

- eyeballing
- absolute distance
- squared distance

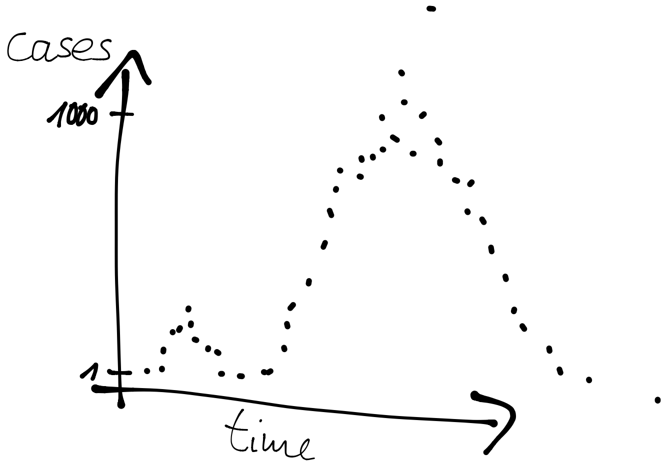
Do these work?



Assessing the “closeness” of model output and data

- eyeballing
- absolute distance
- squared distance

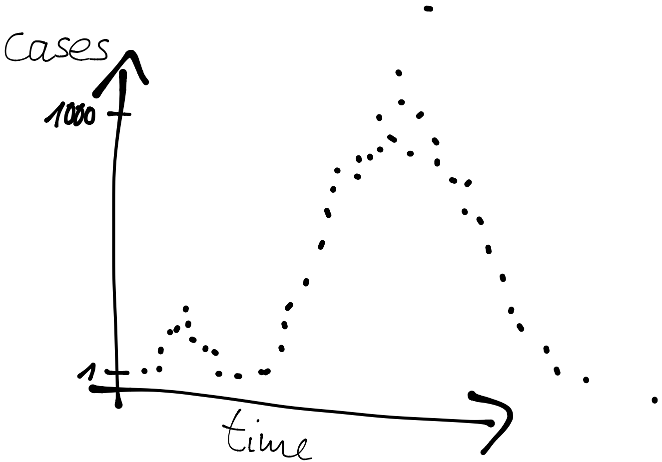
Do these work?



Assessing the "closeness" of model output and data

- eyeballing
- absolute distance
- squared distance

Do these work?



Probabilistic formulation

- We can think of the relationship between the model and data as probabilistic
- For example, 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 $p(\mathrm{data}|\mathrm{underlying\ process})$
- By including this in our model, we get $p(\mathrm{data}|\mathrm{model\ output})$

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 it rains tomorrow $p(\mathrm{W} = \mathrm{rain}) = p(\mathrm{rain})$
- Normalisation $\sum_a p(a) = 1$

Interlude: probabilities II

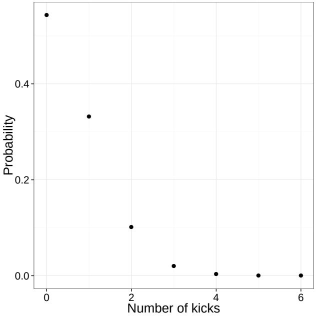
- If A and B are random variables, we write $p(A = a, B = b) = p(a, b)$ for the joint probability that A takes value a and B takes value b
- Example: The probability that it rains tomorrow and India wins at the cricket $p(\mathrm{W} = \mathrm{rain}, \mathrm{C} = \mathrm{India}) = p(\mathrm{rain}, \mathrm{India})$
- 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 India wins at the cricket, given that it rains $p(C = \mathrm{India}|W = \mathrm{rain}) = p(\mathrm{India}|\mathrm{rain})$
- 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

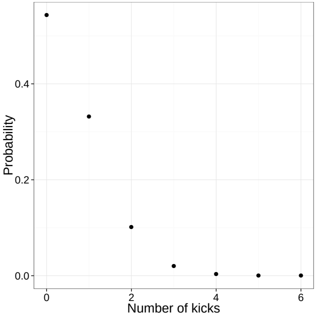


Two directions

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)
```

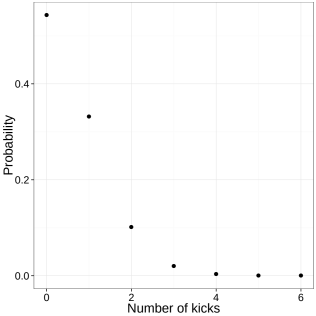
```
[1] 0.1010904
```

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)
```

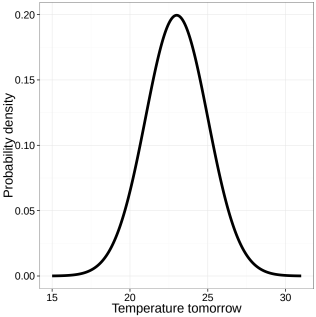
```
[1] 2
```

Two directions

1. Evaluate the probability
2. Randomly sample

Probability distributions (continuous)

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



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

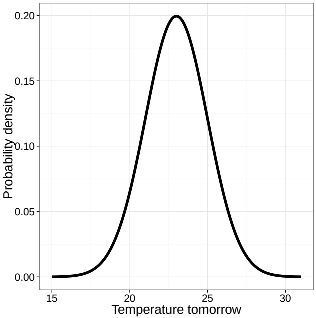
Marginal probabilities:

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

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?

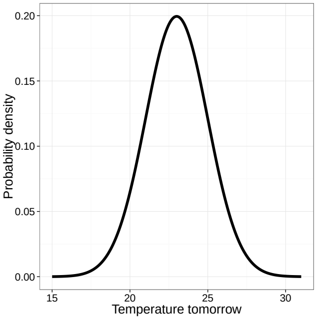
```
dnorm(x = 30,
      mean = 23,
      sd = 2)
```

```
[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)
```

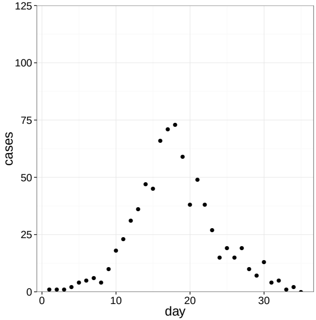
```
[1] 25.61007
```

Two directions

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

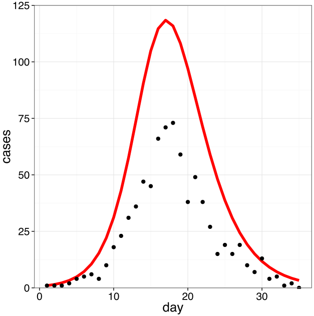
Example: observation uncertainty

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



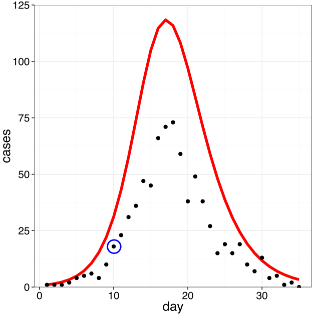
Example: observation uncertainty

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



Example: observation uncertainty

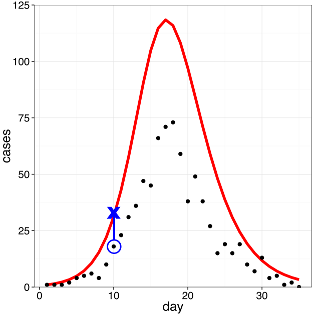
SIR model, assume that cases are detected with independent reporting probability $\rho = 0.5$ per day.



At day 10, 18 cases observed.

Example: observation uncertainty

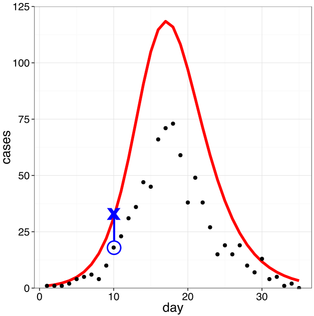
SIR model, assume that cases are detected with independent reporting probability $\rho = 0.5$ per day.

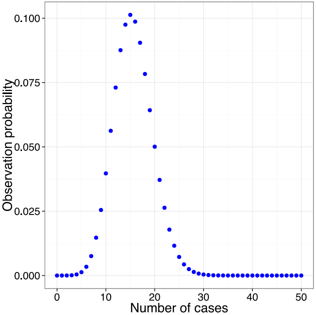


At day 10, 18 cases observed, 31.1 cases in the model.

Example: observation uncertainty

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

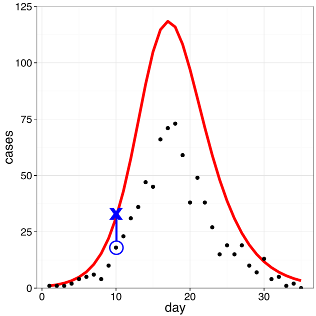


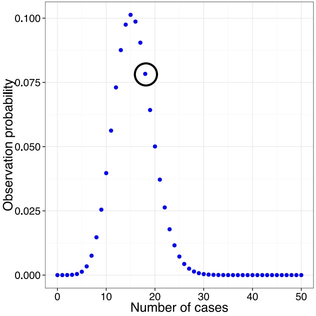


At day 10, 18 cases observed, 31.1 cases in the model.

Example: observation uncertainty

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

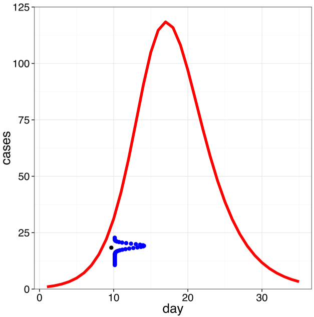


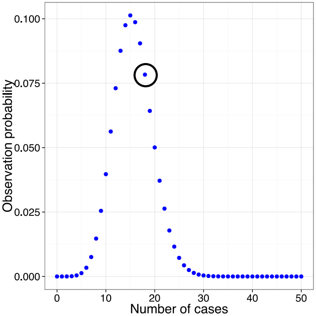


At day 10, 18 cases observed, 31.1 cases in the model.

Example: observation uncertainty

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



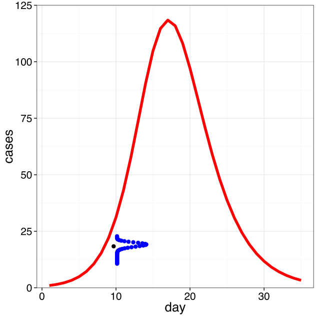


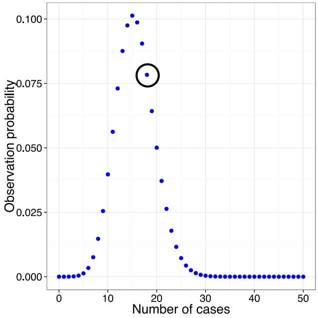
At day 10, 18 cases observed, 31.1 cases in the model.

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

Example: observation uncertainty

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





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

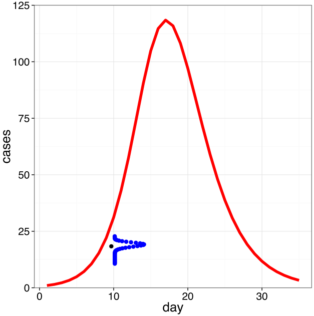
$$p(\text{data point } 10|\theta) = 0.078$$

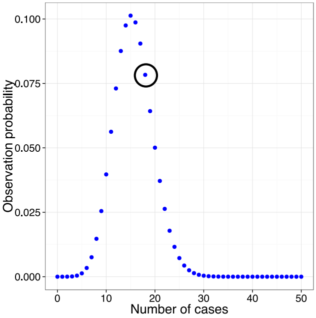
Multiply across the data to get the full trajectory likelihood.

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

Example: observation uncertainty

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



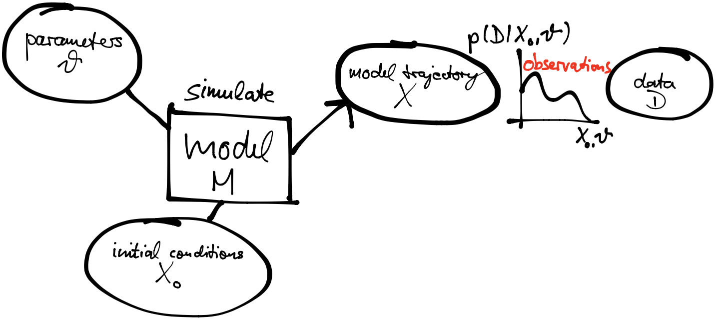


At day 10, 18 cases observed, 31.1 cases in the model.

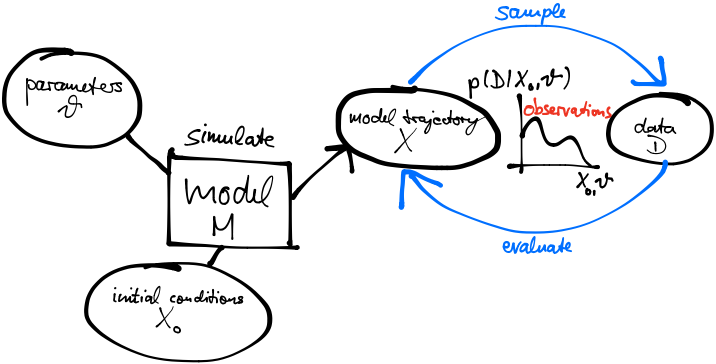
$$p(\text{data point } 10|\theta) = 0.078$$

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
$p(D|X_0,\vartheta)$
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

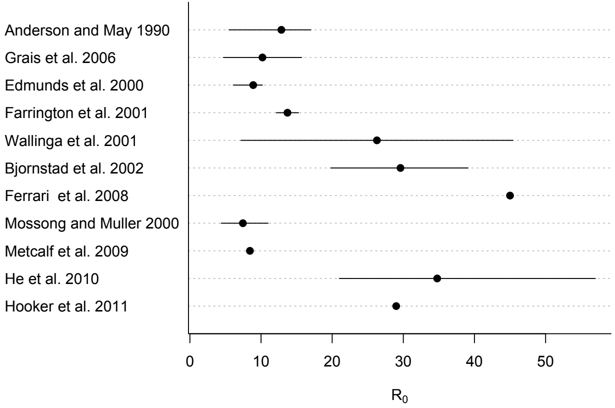
- We said that in Bayesian inference, we need to calculate $p(\theta|\mathrm{data})$. Applying the rule of conditional probabilities, we can write this as

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

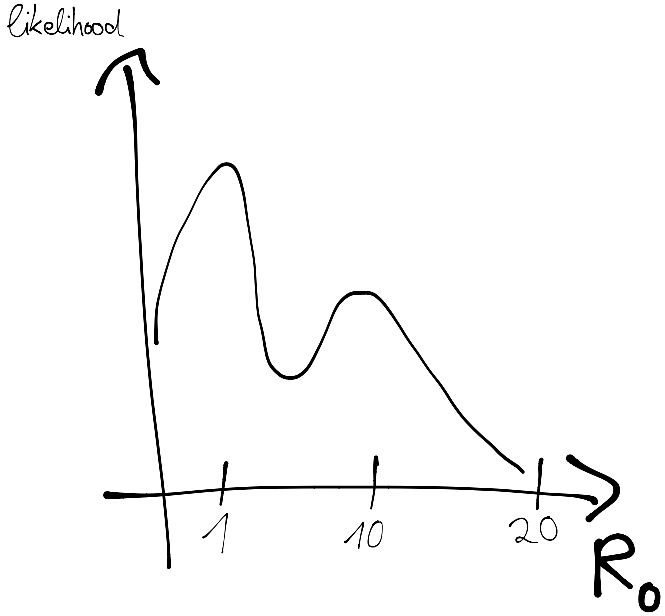
- $p(\theta|\mathrm{data})$ is the *posterior*
- $p(\mathrm{data}|\theta)$ is the *likelihood*
- $p(\theta)$ is the *prior*
- $p(\mathrm{data})$ is a *normalisation constant*
- In words, $(\mathrm{posterior}) \propto (\mathrm{normalised\ likelihood}) \times (\mathrm{prior})$

Prior probabilities

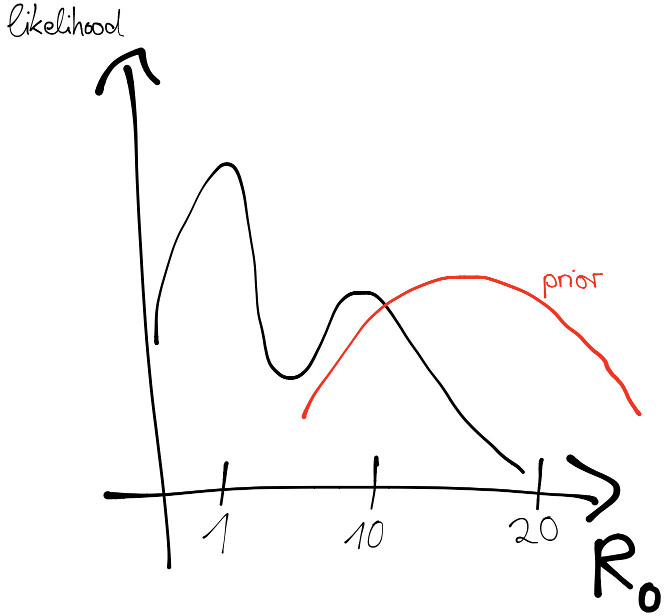
- $p(\theta)$ quantifies our degree of belief via a probability distribution before confronting the model with data: $p(\theta)$ E.g., from previous measurements, literature, experts etc.
- Example: R_0 of measles



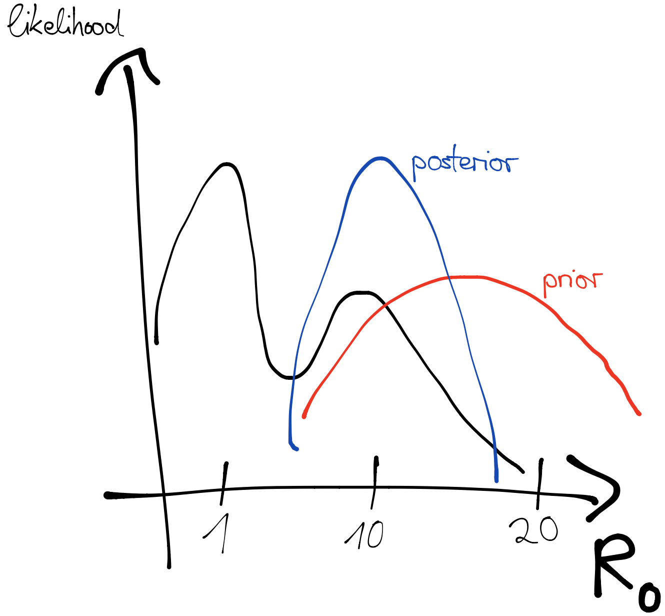
Example: estimating R_0 of measles



Example: prior for estimating R_0 of measles



Example: posterior for estimating R_0 of measles



Expectation values

Bayesian statistics

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

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

- To calculate the expected value of any quantity given the data, we integrate over $p(\theta)$

$$E[A] = \int p(\theta|\mathrm{data})X(\theta)\,d\theta$$

- For example, in an SIR, if we know $p(\beta, \gamma)$, we can calculate the expected value of R_0

$$E[R_0] = \int p(\beta, \gamma|\mathrm{data})\frac{\beta}{\gamma}\,d\beta\,d\gamma$$

Sample approximation

- How do we find an expression for $p(\beta, \gamma | \mathrm{data})$? Generally, this is impossible.
- Instead, we can use a Monte-Carlo approximation:

$$\int f(x) p(x) \, dx \approx \sum_{\mathrm{x}} p(x) f(x)$$

Sample approximation

- How do we find an expression for $p(\beta, \gamma|\mathrm{data})$? Generally, this is impossible.
- Instead, we can use a Monte-Carlo approximation:

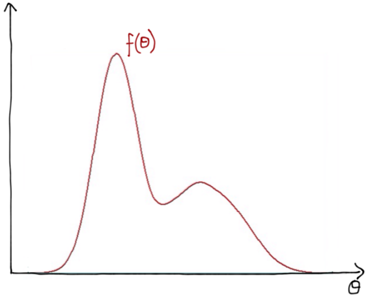
$$\int f(x)p(x)\,dx \approx \sum_{\mathrm{x}} p(x)f(x)$$

- Or: draw N samples from $p(x)$ and calculate

$$\int f(x)p(x)\,dx \approx \frac{1}{N}\sum_{\mathrm{x} \sim \mathrm{p(x)}} f(x)$$

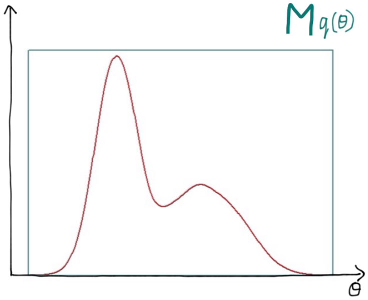
4. Monte-Carlo sampling

Rejection sampling



- Consider a distribution $f(\theta)$,which we can evaluate for any θ
- How do we draw samples?

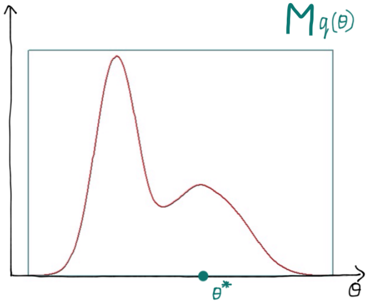
Rejection sampling



Rejection sampling uses a proposal distribution $q(\theta)$ which:

- is simple to evaluate
- is easy to sample from
- one can find $M > 1$ such that $f(\theta) < Mq(\theta)$ for all θ

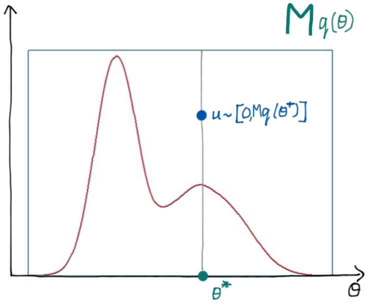
Rejection sampling



The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$

Rejection sampling

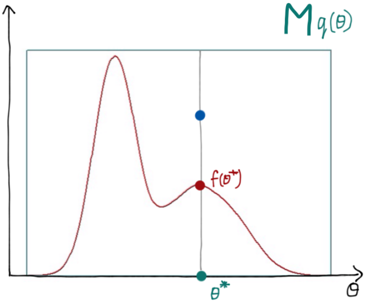


The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$

2. Draw $u \sim \mathit{Uniform}[0, Mq(\theta^*)]$

Rejection sampling



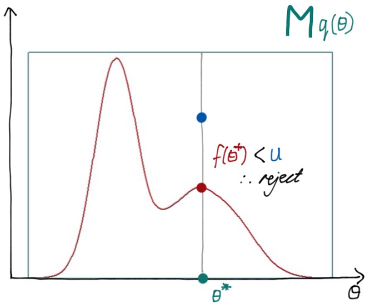
The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$

2. Draw $u \sim \mathit{Uniform}[0, Mq(\theta^*)]$

3. Evaluate $f(\theta^*)$

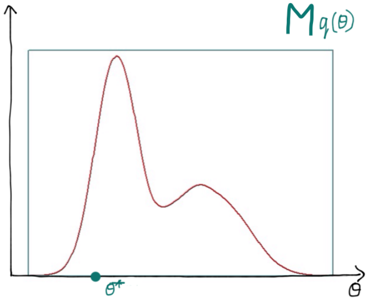
Rejection sampling



The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$
2. Draw $u \sim Uniform[0, Mq(\theta^*)]$
3. Evaluate $f(\theta^*)$
4. If $f(\theta^*) > u$ accept, else reject

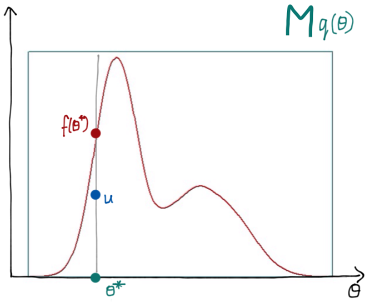
Rejection sampling



The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$
2. Draw $u \sim Uniform[0, Mq(\theta^*)]$
3. Evaluate $f(\theta^*)$
4. If $f(\theta^*) > u$ accept, else reject
5. Repeat steps 1-4

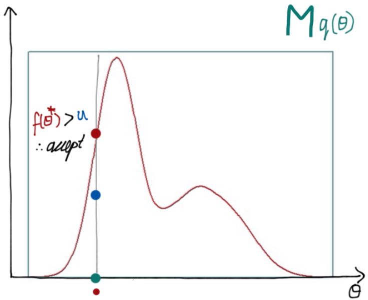
Rejection sampling



The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$
2. Draw $u \sim \mathit{Uniform}[0, Mq(\theta^*)]$
3. Evaluate $f(\theta^*)$
4. If $f(\theta^*) > u$ accept, else reject
5. Repeat steps 1-4

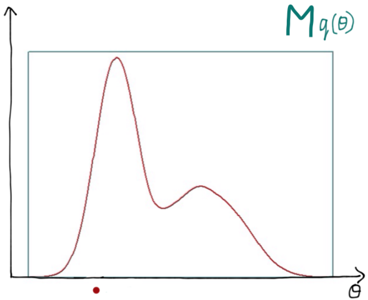
Rejection sampling



The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$
2. Draw $u \sim \mathit{Uniform}[0, Mq(\theta^*)]$
3. Evaluate $f(\theta^*)$
4. If $f(\theta^*) > u$ accept, else reject
5. Repeat steps 1-4

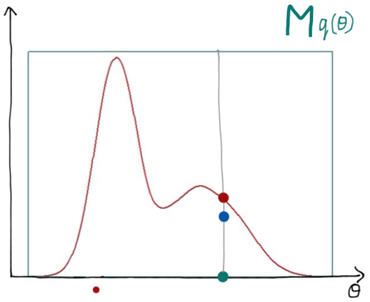
Rejection sampling



The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$
2. Draw $u \sim \mathit{Uniform}[0, Mq(\theta^*)]$
3. Evaluate $f(\theta^*)$
4. If $f(\theta^*) > u$ accept, else reject
5. Repeat steps 1-4

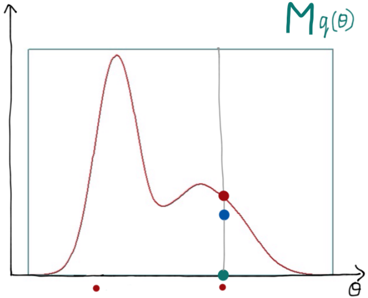
Rejection sampling



The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$
2. Draw $u \sim \mathit{Uniform}[0, Mq(\theta^*)]$
3. Evaluate $f(\theta^*)$
4. If $f(\theta^*) > u$ accept, else reject
5. Repeat steps 1-4

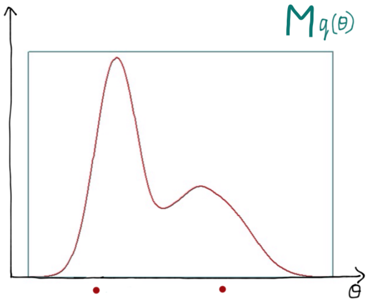
Rejection sampling



The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$
2. Draw $u \sim \mathit{Uniform}[0, Mq(\theta^*)]$
3. Evaluate $f(\theta^*)$
4. If $f(\theta^*) > u$ accept, else reject
5. Repeat steps 1-4

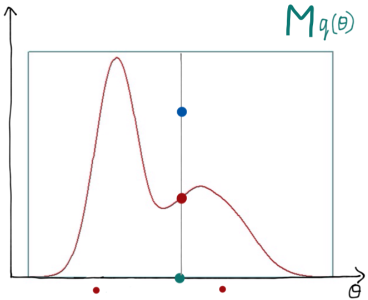
Rejection sampling



The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$
2. Draw $u \sim \mathit{Uniform}[0, Mq(\theta^*)]$
3. Evaluate $f(\theta^*)$
4. If $f(\theta^*) > u$ accept, else reject
5. Repeat steps 1-4

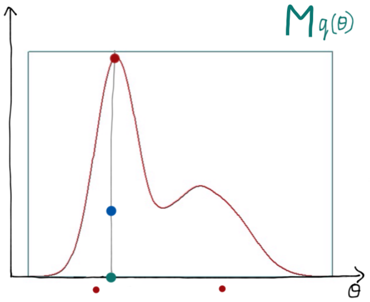
Rejection sampling



The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$
2. Draw $u \sim \mathit{Uniform}[0, Mq(\theta^*)]$
3. Evaluate $f(\theta^*)$
4. If $f(\theta^*) > u$ accept, else reject
5. Repeat steps 1-4

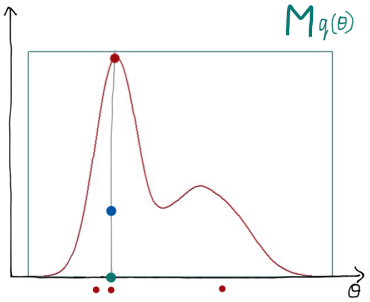
Rejection sampling



The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$
2. Draw $u \sim \textit{Uniform}[0, Mq(\theta^*)]$
3. Evaluate $f(\theta^*)$
4. If $f(\theta^*) > u$ accept, else reject
5. Repeat steps 1-4

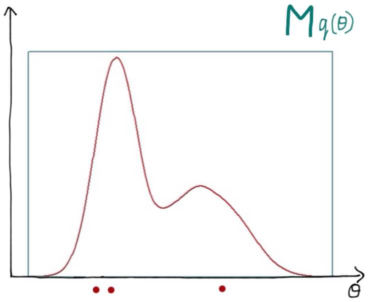
Rejection sampling



The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$
2. Draw $u \sim \mathit{Uniform}[0, Mq(\theta^*)]$
3. Evaluate $f(\theta^*)$
4. If $f(\theta^*) > u$ accept, else reject
5. Repeat steps 1-4

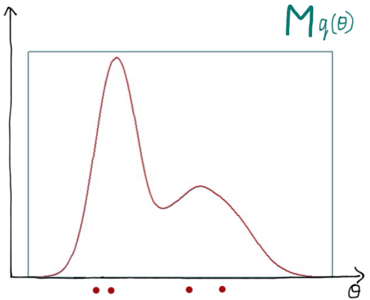
Rejection sampling



The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$
2. Draw $u \sim \mathit{Uniform}[0, Mq(\theta^*)]$
3. Evaluate $f(\theta^*)$
4. If $f(\theta^*) > u$ accept, else reject
5. Repeat steps 1-4

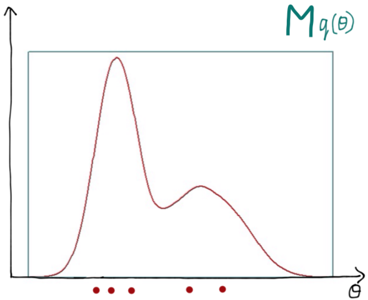
Rejection sampling



The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$
2. Draw $u \sim Uniform[0, Mq(\theta^*)]$
3. Evaluate $f(\theta^*)$
4. If $f(\theta^*) > u$ accept, else reject
5. Repeat steps 1-4

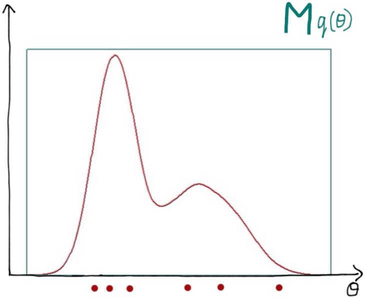
Rejection sampling



The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$
2. Draw $u \sim Uniform[0, Mq(\theta^*)]$
3. Evaluate $f(\theta^*)$
4. If $f(\theta^*) > u$ accept, else reject
5. Repeat steps 1-4

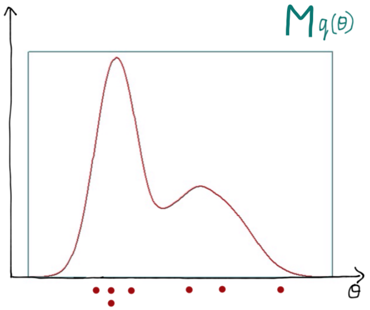
Rejection sampling



The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$
2. Draw $u \sim \mathit{Uniform}[0, Mq(\theta^*)]$
3. Evaluate $f(\theta^*)$
4. If $f(\theta^*) > u$ accept, else reject
5. Repeat steps 1-4

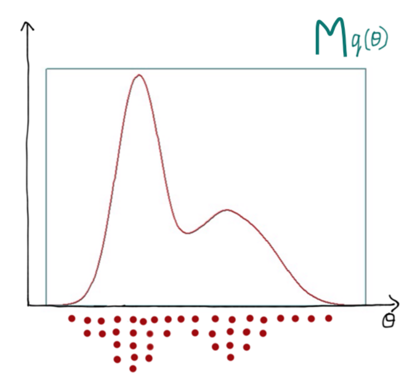
Rejection sampling



The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$
2. Draw $u \sim Uniform[0, Mq(\theta^*)]$
3. Evaluate $f(\theta^*)$
4. If $f(\theta^*) > u$ accept, else reject
5. Repeat steps 1-4

Rejection sampling

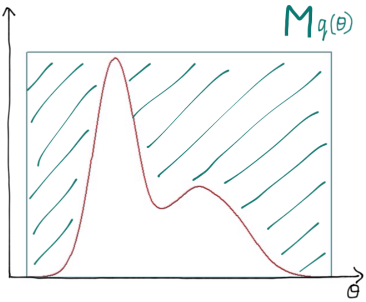


The algorithm proceeds as follows:

1. Sample θ^* from $q(\theta)$
2. Draw $u \sim \mathit{Uniform}[0, Mq(\theta^*)]$
3. Evaluate $f(\theta^*)$
4. If $f(\theta^*) > u$ accept, else reject
5. Repeat steps 1-4

Rejection sampling

- Rejection sampling works best if $q(\theta) \approx f(\theta)$ ($M \gtrapprox 1$)
- Acceptance rate of rejection sampler is $\frac{1}{M}$
- Requiring $f(\theta) < Mq(\theta)$ for all θ can make rejection rate v. high
- Even more limited in high dimensions



Markov Chain Monte Carlo

- In Markov Chain Monte Carlo (MCMC) we do not define one proposal density $q(\theta)$ such that $f(\theta) < Mq(\theta)$.

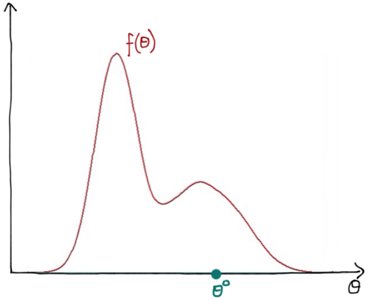
- Rather we build up a chain of samples where each proposed θ^* depends on the previous one

  i.e the proposal density takes the form $q(\theta^*|\theta)$

- A commonly used MCMC algorithm is Metropolis-Hastings (M-H).

- The acceptance rate of M-H is carefully derived to ensure unbiased samples.

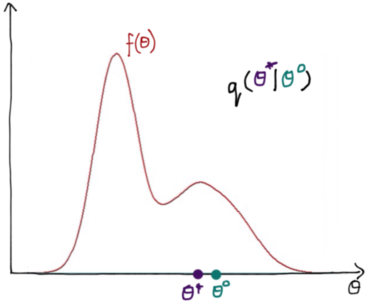
Metropolis-Hastings



The algorithm proceeds as follows:

1. Initialise θ^0, set $\theta = \theta^0$

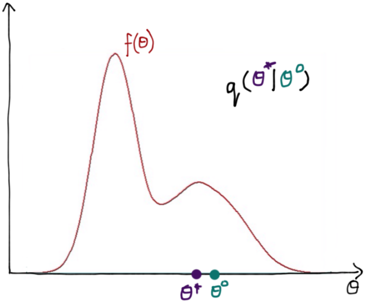
Metropolis-Hastings



The algorithm proceeds as follows:

1. Initialise θ^0, set $\theta = \theta^0$
2. Sample $\theta^* \sim q(\theta^*|\theta)$

Metropolis-Hastings



The algorithm proceeds as follows:

1. Initialise θ^0, set $\theta = \theta^0$
2. Sample $\theta^* \sim q(\theta^*|\theta)$
3. Compute acceptance probability, r

Metropolis-Hastings

Acceptance

- If $q(\theta^*|\theta)$ symmetric, then

$$r = min\left(1, \frac{f(\theta^*)}{f(\theta)}\right)$$

The algorithm proceeds as follows:

1. Initialise θ^0, set $\theta = \theta^0$
2. Sample $\theta^* \sim q(\theta^*|\theta)$
3. Compute acceptance probability, r

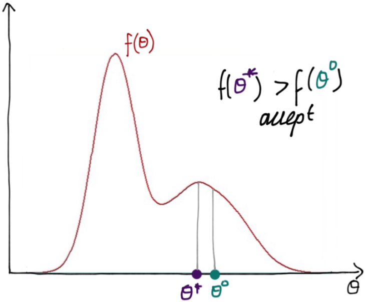
Metropolis-Hastings

Acceptance

- If $q(\theta^*|\theta)$ symmetric, then

$$r = min\left(1, \frac{f(\theta^*)}{f(\theta)}\right)$$

- Definitely move to θ^* if more probable than θ



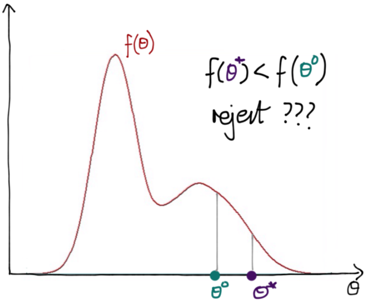
Metropolis-Hastings

Acceptance

- If $q(\theta^*|\theta)$ symmetric, then

$$r = min\left(1, \frac{f(\theta^*)}{f(\theta)}\right)$$

- Definitely move to θ^* if more probable than θ
- May move if θ^* less probable



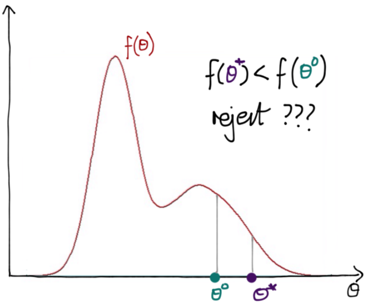
Metropolis-Hastings

Acceptance

- If $q(\theta^*|\theta)$ symmetric, then

$$r = min\left(1, \frac{f(\theta^*)}{f(\theta)}\right)$$

- Definitely move to θ^* if more probable than θ
- May move if θ^* less probable



- If $q(\theta^*|\theta)$ asymmetric, then

$$r = min\left(1, \frac{f(\theta^*)q(\theta|\theta^*)}{f(\theta)q(\theta^*|\theta)}\right)$$

Metropolis-Hastings

Acceptance

- If $q(\theta^*|\theta)$ symmetric, then

$$r = min\left(1, \frac{f(\theta^*)}{f(\theta)}\right)$$

- Definitely move to θ^* if more probable than θ
- May move if θ^* less probable

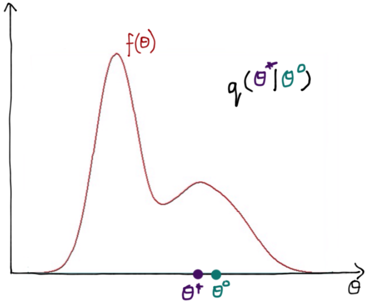
- If $q(\theta^*|\theta)$ asymmetric, then

$$r = min\left(1, \frac{f(\theta^*)q(\theta|\theta^*)}{f(\theta)q(\theta^*|\theta)}\right)$$

The algorithm proceeds as follows:

1. Initialise θ^0, set $\theta = \theta^0$
2. Sample $\theta^* \sim q(\theta^*|\theta)$
3. Compute acceptance probability, r

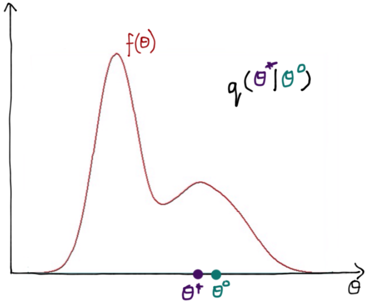
Metropolis-Hastings



The algorithm proceeds as follows:

1. Initialise θ^0, set $\theta = \theta^0$
2. Sample $\theta^* \sim q(\theta^*|\theta)$
3. Compute acceptance probability, r

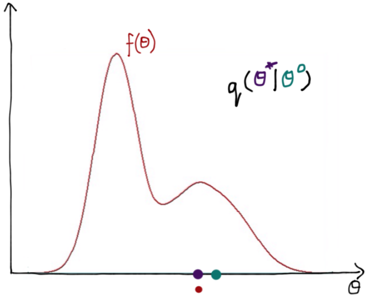
Metropolis-Hastings



The algorithm proceeds as follows:

1. Initialise θ^0, set $\theta = \theta^0$
2. Sample $\theta^* \sim q(\theta^*|\theta)$
3. Compute acceptance probability, r
4. Draw $u \sim \mathit{Uniform}[0, 1]$

Metropolis-Hastings

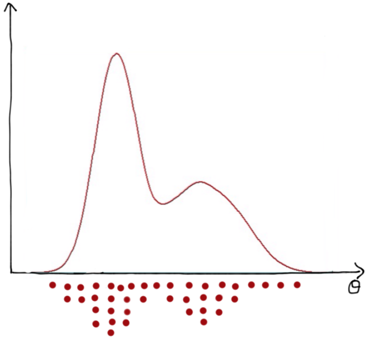


The algorithm proceeds as follows:

1. Initialise θ^0, set $\theta = \theta^0$
2. Sample $\theta^* \sim q(\theta^*|\theta)$
3. Compute acceptance probability, r
4. Draw $u \sim \mathit{Uniform}[0, 1]$
5. Set new sample to

$$\theta^{(s+1)} = \begin{cases} \theta^*, & \text{if } u < r \\ \theta^{(s)}, & \text{if } u \geqslant r \end{cases}$$

Metropolis-Hastings



The algorithm proceeds as follows:

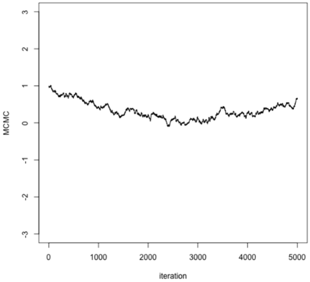
1. Initialise θ^0, set $\theta = \theta^0$
2. Sample $\theta^* \sim q(\theta^*|\theta)$
3. Compute acceptance probability, r
4. Draw $u \sim Uniform[0,1]$
5. Set new sample to

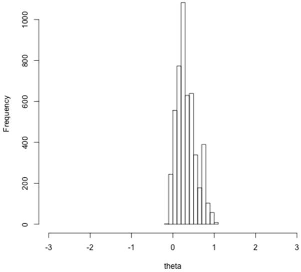
$$\theta^{(s+1)} = \begin{cases} \theta^*, & \text{if } u < r \\ \theta^{(s)}, & \text{if } u \geqslant r \end{cases}$$

6. Repeat steps 2-5

Choosing a proposal distribution

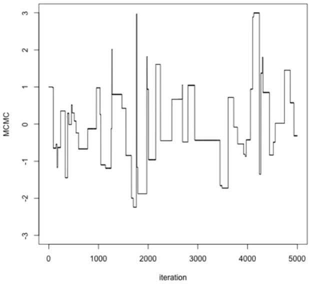
If variance is too small, the chain will be slow to reach the target distribution.

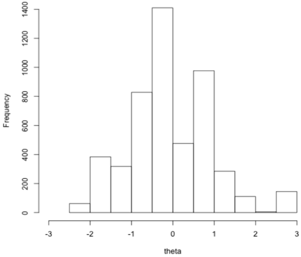




Choosing a proposal distribution

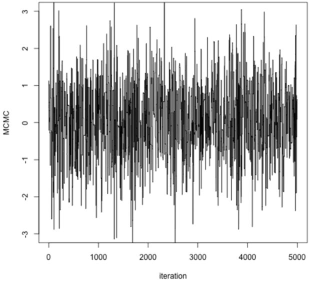
If variance is too high, many proposed values will be rejected and the chain will *stick* in one place for many steps.

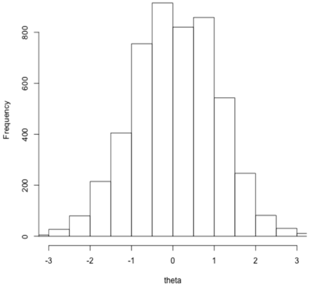




Choosing a proposal distribution

If variance is just right, the chain will efficiently explore the full shape of the target distribution.



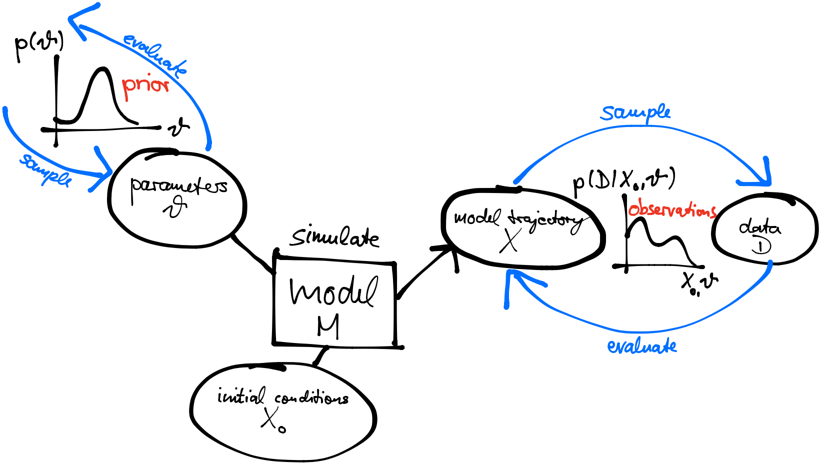


Try several different proposal distributions (pilot runs), aiming for an acceptance rate between 24% and 40%.

5. Summary

Summary

- Likelihood as $p(\text{data}|\theta)$ to express closeness of model to data
- Bayesian inference:
  $(\text{posterior}) \propto (\text{normalised likelihood}) \times (\text{prior})$
- To estimate quantities or project into the future, need to calculate $E[A] = \int p(\theta|\text{data}) X(\theta) d\theta$
- Monte Carlo sampling as a method to calculate this
- Metropolis-Hastings Markov-Chain Monte Carlo method



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

Tomorrow: Try it yourself in the lab