Model fitting: Bayesian approach

Clara Champagne

clara.champagne@swisstph.ch

March 2024

Swiss TPH

Recap of previous classes

What is model fitting?

Objective

We will introduce Bayesian methods to fit disease models, with the example of MCMC

This method is very commonly used for fitting disease models

Today: we are explaining the theory (lecture)

Tomorrow: we will apply the method in a practical

At the end of the class, you should know:

1. **What is the Bayesian approach and its advantages (general concept)**
2. **How to apply MCMC to a simple disease model (like SIR or SIS)**

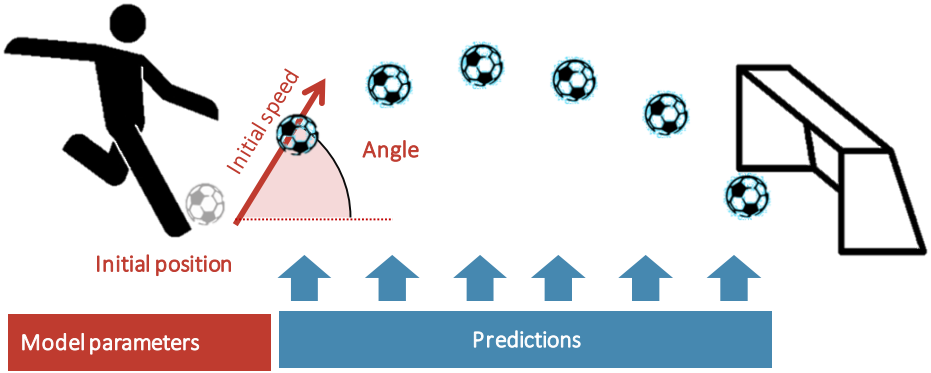
Outline

- Concepts of model fitting and bayesian inference
- Why it is useful?
- How to perform bayesian inference for dynamical models

Outline

- Concepts of model fitting and bayesian inference
- Why is it useful?
- How to perform bayesian inference for dynamical model

Mathematical model



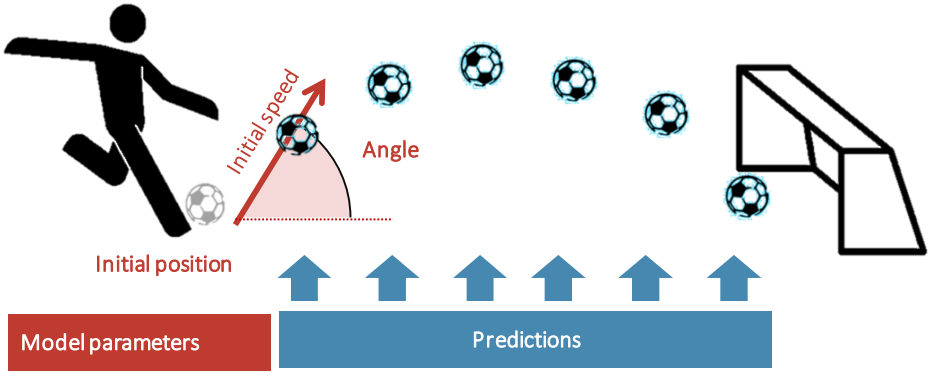
A mathematical model is build on our knowledge of the mechanisms that govern the dynamics of a system.

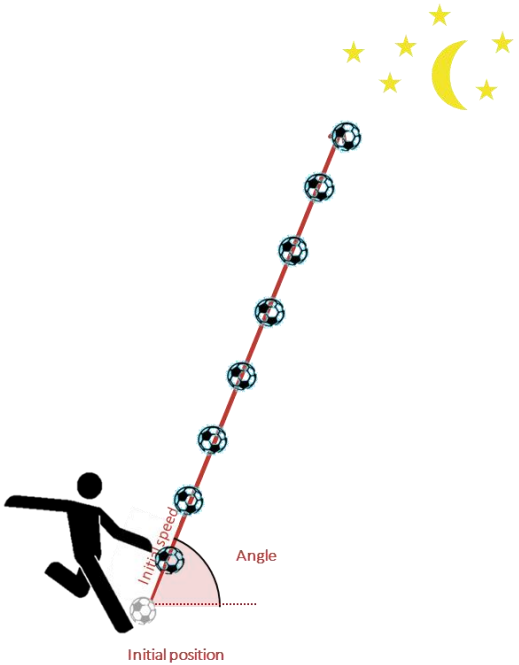
For example, the laws of physics for the trajectory of the football.
Or our knowledge about the biology of malaria in the case of a malaria model

One of the powers of mathematical models is to explore various scenarios (“what if ?”)

12

Mathematical model



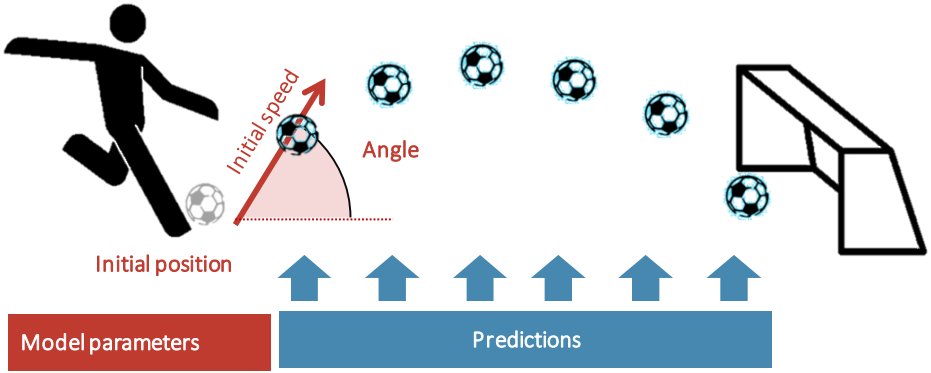


The model relies on various parameters.

But, if we want our model to represent realistic scenarios, how to find the values for the parameters?

Swiss TPH

Mathematical model



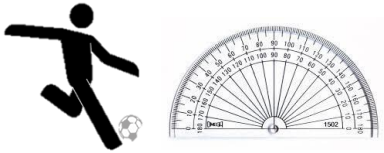
The model relies on various parameters.

But, if we want our model to represent realistic scenarios, how to find the values for the parameters?

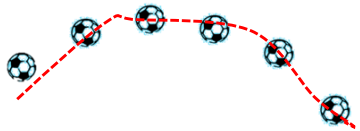
We need **DATA**

2 approaches:

- Try to measure the parameters directly



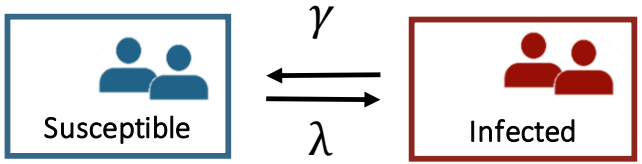
- Fit the parameters, i.e. finding the parameter values that best reproduce known model outputs



Swiss TPH

Infectious disease model

SIS model



$$\frac{dS}{dt} = -\frac{\lambda SI}{N} + \gamma I$$

$$\frac{dI}{dt} = \frac{\lambda SI}{N} - \gamma I$$

The model relies on various parameters.

But, if we want our model to represent realistic scenarios, how to find the values for the parameters?

We need **DATA**

2 approaches:

- Try to measure the parameters directly

Counting the number of contacts

- Fit the parameters, i.e. finding the parameter values that best reproduce known model outputs

Fitting the number of reported cases

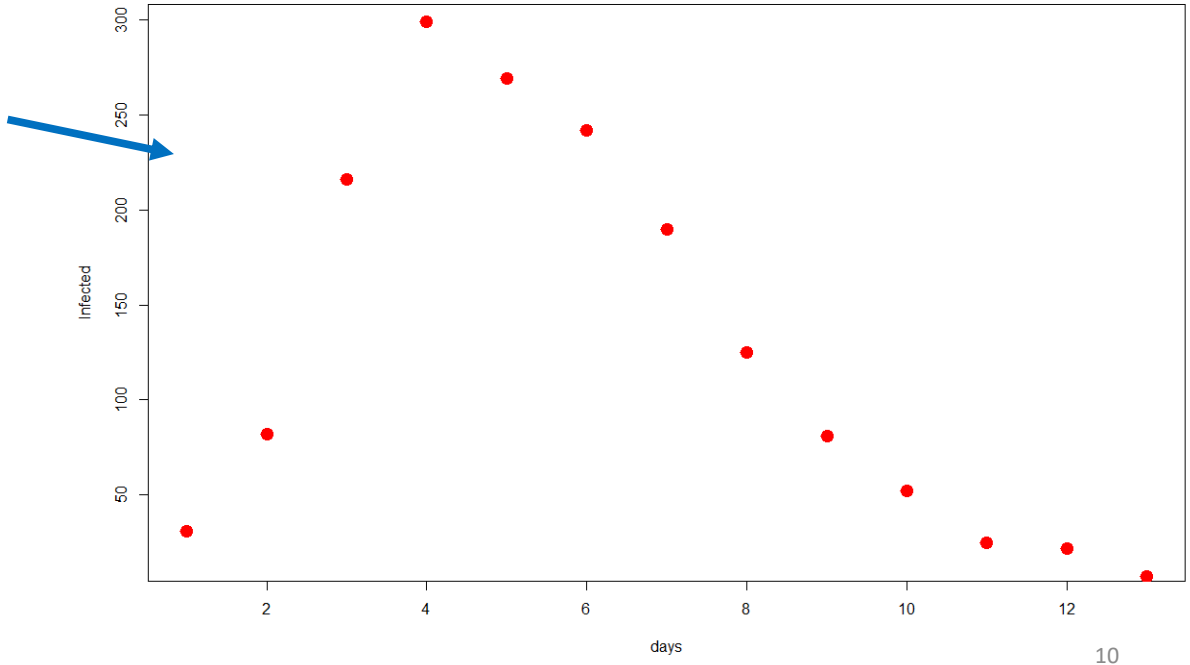
Swiss TPH

Overview and notations

Prevalence: proportion of infectious individuals over time

The data:

X_{obs}

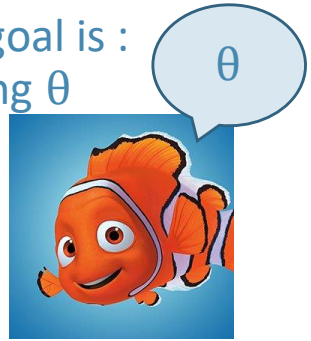


In practice, we very often observe incidence (number of new cases over a certain time, but for simplicity here, we will focus on prevalence)

Overview and notations

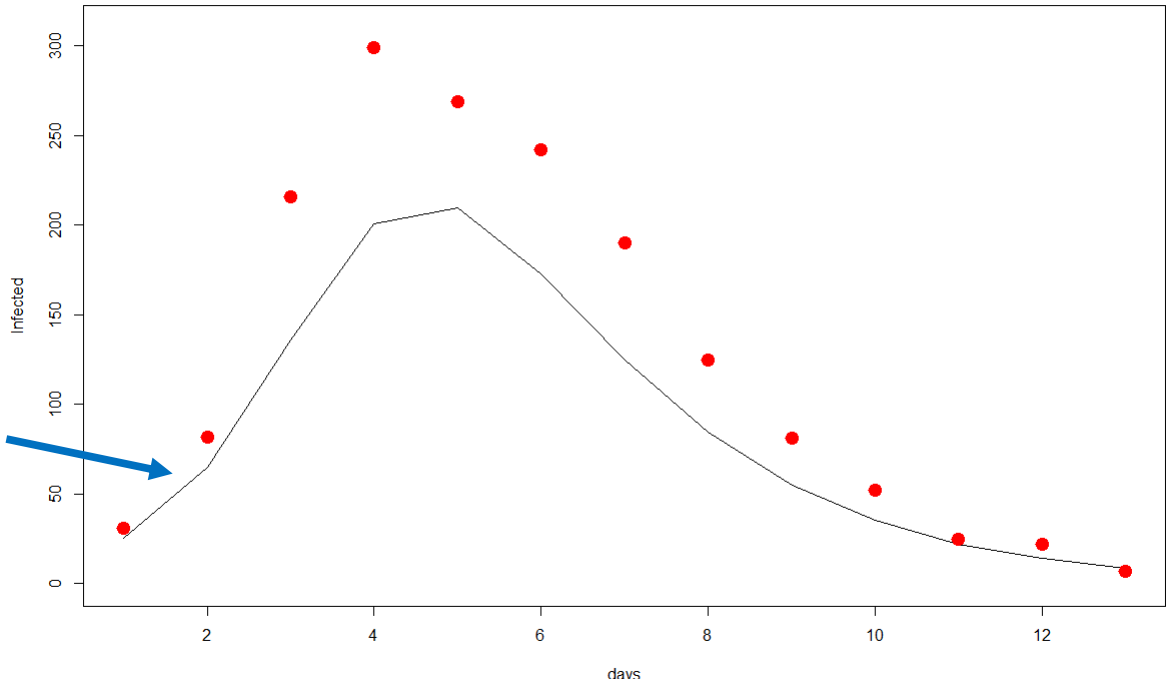
The data: X_{obs}

The goal is : finding θ



A model M that depends on a parameter set θ

(for example, an SIR model, or an SIS model)



Overview and notations

The data: X_{obs}

The parameters: θ

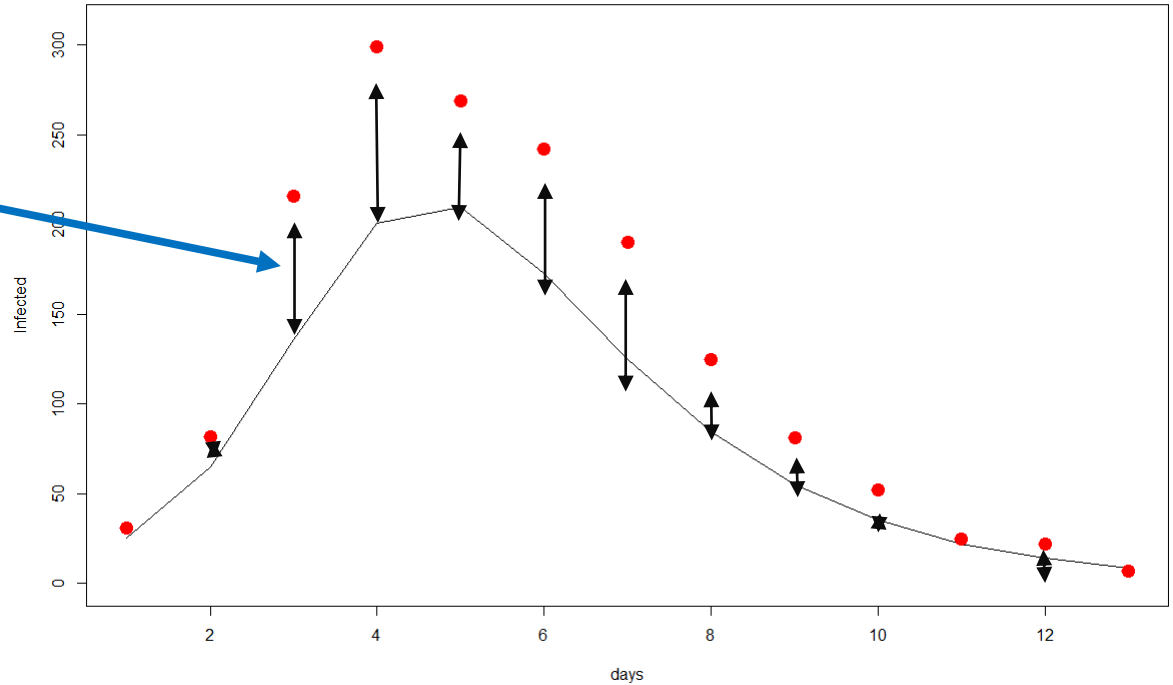
A metric to compare the model output and the data

$$L_M(X_{obs}, \theta) = P_M(X_{obs}|\theta)$$

"How likely is my data with this parameter"

Can be

- Gaussian (→ least squares)
- Poisson
- Negative binomial
- …



The full model: state space model representation

Example of the SIS model, where we observe the number of infected

- Transition equation:

$$\frac{dS}{dt} = -\frac{\lambda SI}{N} + \gamma I$$

$$\frac{dI}{dt} = \frac{\lambda SI}{N} - \gamma I$$

$$\rightarrow \quad I_{1,,,n}$$

$$\theta = (\lambda, \gamma)$$

- Observation equation:

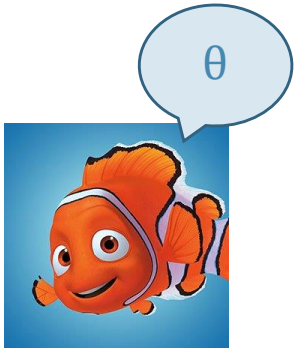
$$X^{obs}_{1,\dots,n} \sim Poisson(I_{1,,,n})$$

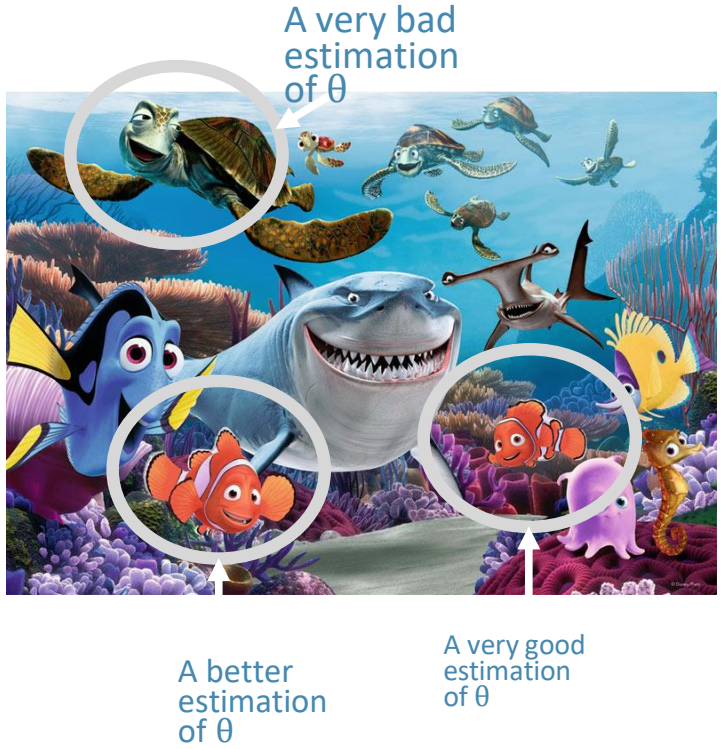
$$L_M\left(X^{obs}_{1,\dots,n}, \theta\right) = \prod_{i=1}^{n} \frac{I_i^{X_i^{obs}} e^{-I_i}}{X_i^{obs}!}$$

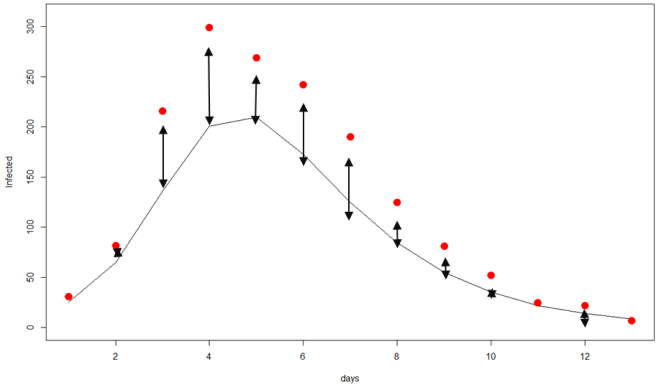
Overview and notations

Frequentist framework:

θ is fixed but unknown. We are looking for a way to find a $\hat{\theta}$ that is most likely close to θ (with some confidence bounds)

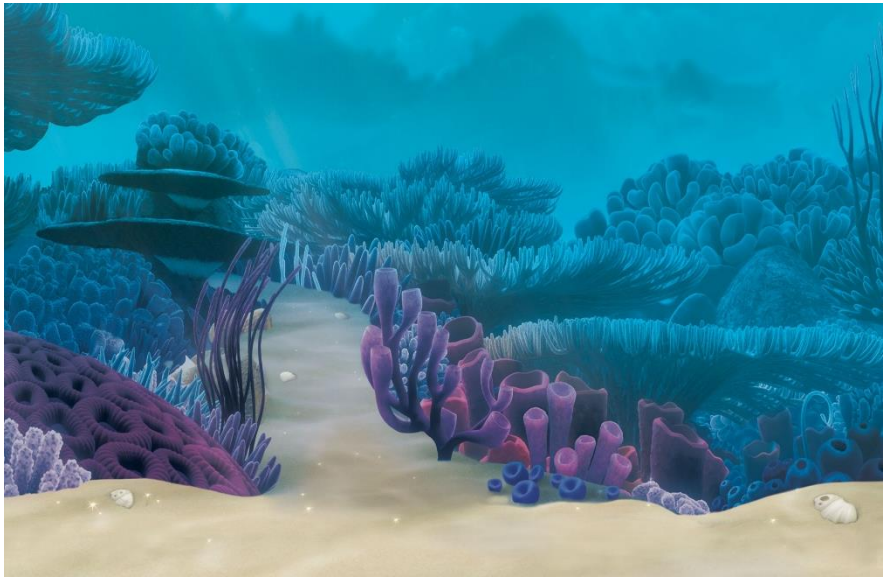






Maximum likelihood: Finding θ such that the distance is minimized (= the likelihood is maximlized)

Overview and notations



Frequentist framework:

θ is fixed but unknown. We are looking for a way to find a $\hat{\theta}$ that is most likely close to θ (with some confidence bounds)

Bayesian framework:

θ random. We are looking for the regions of the parameter space where it is most likely to be, given the data we have, i.e.

$$P_M(\theta|X_{obs})$$

Overview and notations

Frequentist framework:

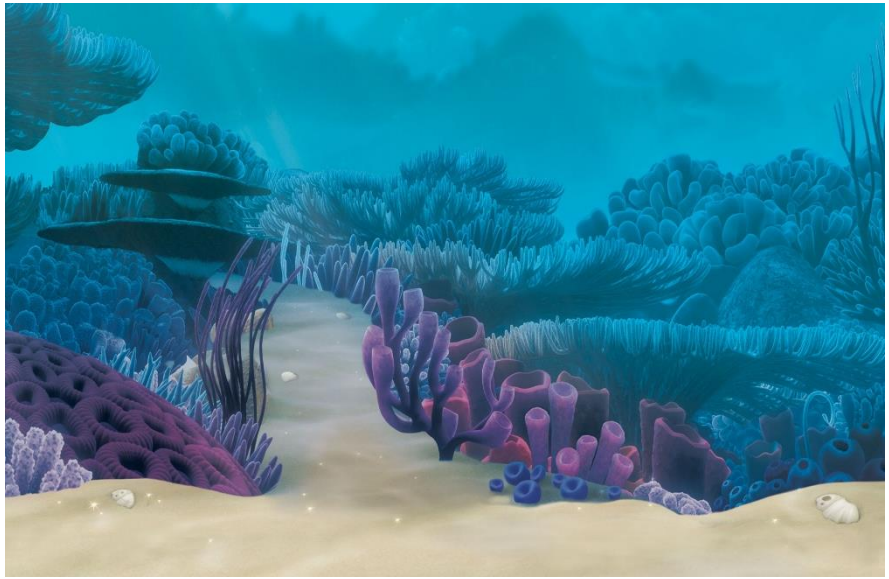
θ is fixed but unknown. We are looking for a way to find a $\hat{\theta}$ that is most likely close to θ (with some confidence bounds)

Bayesian framework:

θ random. We are looking for the regions of the parameter space where it is most likely to be, given the data we have, i.e.

$$P_M(\theta|X_{obs})$$

Overview and notations



Bayesian framework:

Posterior distribution:

$$P_M(\theta|X_{obs}) = \frac{P_M(X_{obs}|\,\theta)P_M(\theta)}{P_M(X)}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Overview and notations

Bayesian framework:

Posterior distribution:

$$P_M(\theta|X_{obs}) = \frac{P_M(X_{obs}|\,\theta)P_M(\theta)}{P_M(X)}$$

Prior distribution

Example of a non-informative prior

Overview and notations

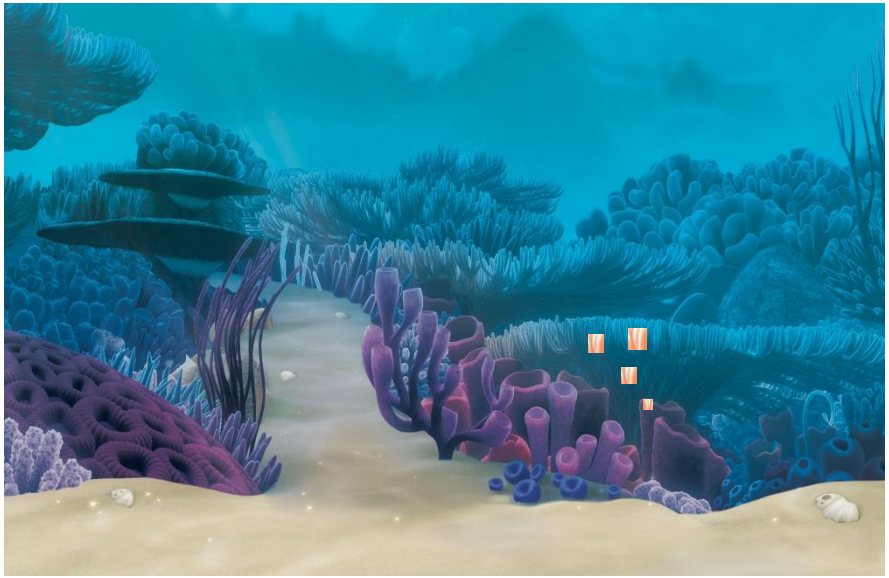
Bayesian framework:

Posterior distribution:

Marginal likelihood (like before)

Prior distribution

$$P_M(\theta|X_{obs}) = \frac{P_M(X_{obs}|\,\theta)P_M(\theta)}{P_M(X)}$$



Overview and notations

Bayesian framework:

Posterior distribution:

Marginal likelihood (like before)

Prior distribution

$$P_M(\theta|X_{obs}) = \frac{P_M(X_{obs}|\,\theta)P_M(\theta)}{P_M(X)}$$

Overview and notations

Bayesian framework:

Posterior distribution:

Marginal likelihood (like before)

Prior distribution

$$P_M(\theta|X_{obs}) = \frac{P_M(X_{obs}|\,\theta)P_M(\theta)}{P_M(X)}$$

Most of the time unknown !

How to calculate the posterior ?

Outline

- Concepts of model fitting and bayesian inference
- **Why is it useful?**
- How to perform bayesian inference for dynamical model

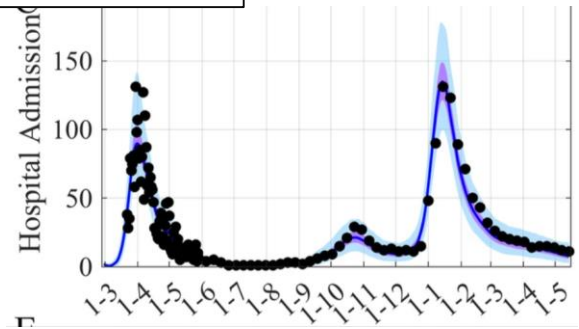
	Frequentist Paradigm:	Bayesian Paradigm:
Focus	Emphasizes the long-run properties of estimators and tests based on repeated sampling from a fixed population	Incorporates prior beliefs and updates them with observed data to obtain posterior distributions.
Parameter estimation	parameters are considered fixed but unknown. Estimation involves finding a single point estimate of the parameter that maximizes the likelihood function.	Parameters are treated as random variables with probability distributions representing uncertainty.
Interpretation	Point estimates obtained from frequentist methods represent a single value believed to be close to the true parameter value, with uncertainty typically quantified through confidence intervals.	Provides full probability distributions for parameters, allowing for a comprehensive assessment of uncertainty.
Assumption	Rely on assumptions such as random sampling, large sample sizes, and the asymptotic behavior of estimators.	Require specification of prior distributions

Swiss TPH

Courtesy of C. N. Orlendo

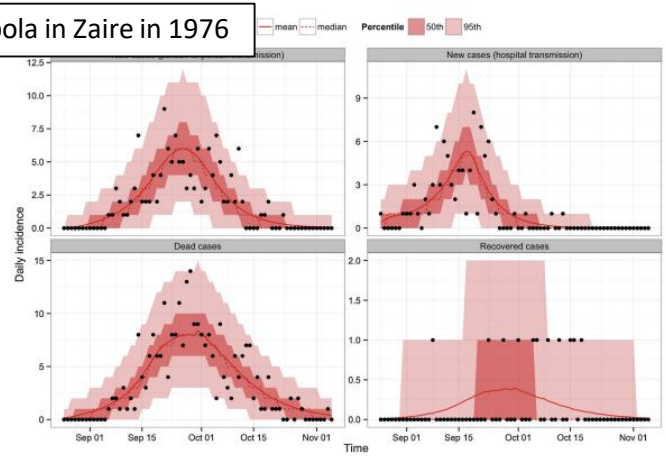
Bayesian methods for fitting are very commonly used for disease models

Covid19 in Ireland



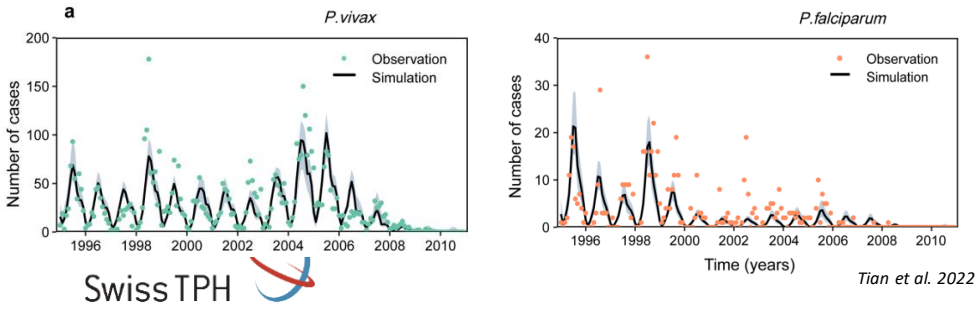
Cazelles et al. 2021

Ebola in Zaire in 1976



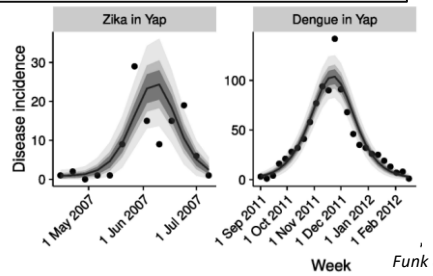
Camacho et al. 2014

Malaria in China



Tian et al. 2022

Zika and Dengue in Micronesia



Funk et al. 2016

What are the advantages of this approach

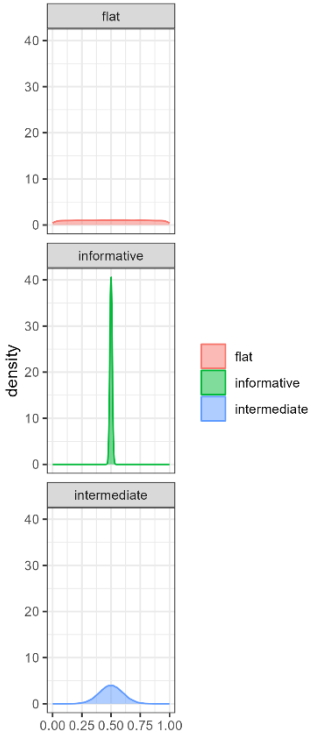
- General methods that can be applied on a very large class of disease models (mainly compartmental models)
- Very good for measuring and propagating uncertainties
  - **Posterior distribution:** we obtain a full distribution for the parameters and we can propagate it in the future when simulating the model
  - **Prior distribution:** we can include what we know before fitting the data.

Some words about priors

- Frequentist approach: all or nothing
  - we consider that we know nothing → we fit the parameter
  - we know the parameter perfectly → we don't fit the parameter
- Bayesian approach:
  - we consider that we know nothing → we take a non-informative prior
  - we know the parameter perfectly → we take a very informative prior
  - we know a little about the parameter → we take a weakly informative prior, based on the literature, expert opinion, previous model fits, etc. (this is not always easy…)

- The more data you have, the smaller the influence of the prior on the posterior

$$P_M(\theta|X_{obs}) = \frac{P_M(X_{obs}|\,\theta)P_M(\theta)}{P_M(X)}$$



Outline

- Concepts of model fitting and bayesian inference
- Why is it useful?
- How to perform bayesian inference for dynamical models? The example of Metropolis-Hastings MCMC

Metropolis-Hastings is one type of MCMC:
there are other types of MCMC (e.g. Gibbs sampling, HMC, etc.)

And there are many other Bayesian inference methods that are not MCMC.
But we will not talk about them in this course.

How to calculate the posterior?

We want to calculate : $P_M(\theta|X_{obs}) = \frac{P_M(X_{obs}|\,\theta)P_M(\theta)}{P_M(X)}$

but we don't know the denominator ….

→ In some simple cases, we can find the posterior distribution through calculations.
E.g. if the prior $P_M(\theta)$ is Gaussian and the likelihood $P_M(X_{obs}|\,\theta)$ is Gaussian, we know that the posterior is a Gaussian, and we can find its parameters through calculations. (not shown here)

Already in this simple case, calculations are a bit lengthy.
As soon as the models become less simple, this approach is not manageable.

The full model: state space model representation

Example of the SIS model, where we observe the number of infected

- Transition equation:

$$\frac{dS}{dt} = -\frac{\lambda SI}{N} + \gamma I$$

$$\frac{dI}{dt} = \frac{\lambda SI}{N} - \gamma I$$

$\rightarrow$ $I_{1,,,n}$

$\theta = (\lambda, \gamma)$

- Observation equation: $X^{obs}_{1,\dots,n} \sim Poisson(I_{1,,,n})$

$$P_M(X_{obs}|\,\theta) = L_M\left(X^{obs}_{1,\dots,n}, \theta\right) = \prod_{i=1}^{n} \frac{I_i^{X_i^{obs}} e^{I_i}}{X_i^{obs}!}$$

We don't have an analytical formula for $I_{1,,,n}$ (we solve the ODE through simulation)
→ We don't have an analytical formula for $P_M(X_{obs}|\,\theta)$

The full r… …epresentation

Even if we can calculate $P_M(\theta|X_{obs})$ numerically, how can we calculate the posterior because we don't know the denominator $P_M(X)$???

…ber of infected

$$\frac{dI}{dt} = \ldots - \gamma I$$

$I_{1,,,n}$

$\theta = (\lambda, \gamma)$

- Observation equation: $X^{obs}_{1,\ldots,n}, \sim N(I_{1,,,n}, \sigma)$

$$P_M(X_{obs}|\,\theta) = L_M\left(X^{obs}_{1,\ldots,n}, \theta\right) = \prod_{i=1}^{n} \frac{I_i^{X_i^{obs}} e^{-I_i}}{X_i^{obs}!}$$

$$P_M(\theta|X_{obs}) = \frac{P_M(X_{obs}|\,\theta) P_M(\theta)}{P_M(X)}$$

We don't have an analytical formula for $I_{1,,,n}$ (we solve the ODE through simulation)
→ We don't have an analytical formula for $P_M(X_{obs}|\,\theta)$

MCMC = Markov Chain Monte Carlo

Monte Carlo:

- Reproduce a numerical experiment to calculate a quantity

Markov Chain:

- Collection of random variables $Y_0, Y_1, Y_2, \ldots$ such that
- $P(Y_{i+1}|Y_0, Y_1, Y_2, \ldots, Y_i) = P(Y_{i+1} \mid Y_i) = Q$ (memoryless)

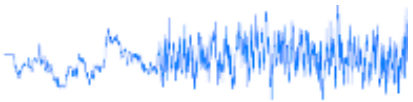
MCMC = Markov Chain Monte Carlo

- **Monte Carlo:**
- Reproduce a numerical experiment to calculate a quantity

- **Markov Chain:**
- Collection of random variables $Y_0, Y_1, Y_2, \ldots$ such that
- $P(Y_{i+1}|Y_0, Y_1, Y_2, \ldots, Y_i) = P(Y_{i+1} \mid Y_i) = Q$ (memoryless)

Under certain conditions:

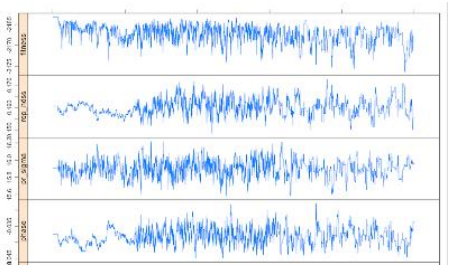
- The chain has stationary distribution Π such that if $Y_i \sim \Pi$ then $Y_{i+1} \sim \Pi$
- If $i \to \infty$ then the limiting distribution for Y_i is Π

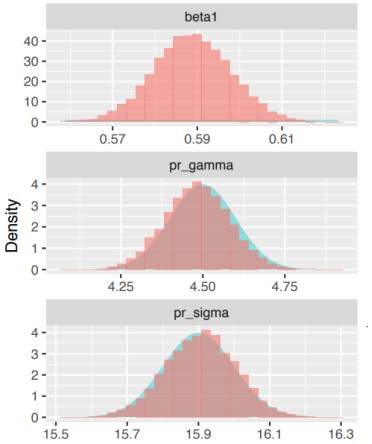


Swiss TPH

Idea of the MCMC

- We want to simulate from a target distribution, here $P_M(\theta|X_{obs})$
- We create a Markov chain $\theta_0, \theta_1, \theta_2, \ldots$ for which Π is $P_M(\theta|X_{obs})$
- When the chain has converged to Π, the simulation of $\theta_{n+1}, \theta_{n+2}, \ldots$ will mimic a sample from $P_M(\theta|X_{obs})$





How do we find such a Markov chain???

Metropolis-Hastings algorithm

The target distribution is our posterior $P_M(\theta|X_{obs})$.
We need an intialisation point θ_0 and a proposal $q(.\mid \theta_i)$

1. Initialise θ_0
2. **for** $\boldsymbol{i = 1 \ldots N}$
3. Sample θ^* from $q(.\mid \theta_i)$
4. Accept θ^* with probability $1 \wedge \frac{P_M(\theta^*|X_{obs})\, q(\theta_i|\theta^*)}{P_M(\theta_i|X_{obs}) q(\theta^*|\theta_i)}$
5. If θ^* is accepted, $\theta_{i+1} = \theta^*$. Otherwise $\theta_{i+1} = \theta_i$
6. **end for**
7. **return** $(\theta_0, \theta_1, \ldots, \theta_N)$

This algorithm generates a Markov chain $(\theta_0, \theta_1, \ldots, \theta_N)$ that will converge to $P_M(\theta^*|X_{obs})$

Swiss TPH

Metropolis-Hastings algorithm

The target distribution is our posterior $P_M(\theta|X_{obs})$.
We need an intialisation point θ_0 and a proposal $q(.\,|\,\theta_i)$

1. Initialise θ_0
2. **for** $\boldsymbol{i = 1 \ldots N}$
3. Sample θ^* from $q(.\,|\,\theta_i)$
4. Accept θ^* with probability $1 \wedge \frac{P_M(\theta^*|X_{obs})\, q(\theta_i|\theta^*)}{P_M(\theta_i|X_{obs}) q(\theta^*|\theta_i)}$
5. If θ^* is accepted, $\theta_{i+1} = \theta^*$. Otherwise $\theta_{i+1} = \theta_i$
6. **end for**
7. **return** $(\theta_0, \theta_1, \ldots, \theta_N)$

Just to ensure the probability remains below 1

This algorithm generates a Markov chain $(\theta_0, \theta_1, \ldots, \theta_N)$ that will converge to $P_M(\theta^*|X_{obs})$

Swiss TPH

Metropolis-Hastings algorithm

The target distribution is our posterior $P_M(\theta|X_{obs})$.
We need an intialisation point θ_0 and a proposal $q(.\,|\,\theta_i)$

1. Initialise θ_0
2. **for** $\boldsymbol{i = 1 \ldots N}$
3. Sample θ^* from $q(.\,|\,\theta_i)$
4. Accept θ^* with probability $1 \wedge \frac{P_M(\theta^*|X_{obs})}{P_M(\theta_i|X_{obs})}\frac{q(\theta_i|\theta^*)}{q(\theta^*|\theta_i)}$
5. If θ^* is accepted, $\theta_{i+1} = \theta^*$. Otherwise $\theta_{i+1} = \theta_i$
6. **end for**
7. **return** $(\theta_0, \theta_1, \ldots, \theta_N)$

Easy to compute as soon as we know q ***(and we can choose q as we want)***

This algorithm generates a Markov chain $(\theta_0, \theta_1, \ldots, \theta_N)$ that will converge to $P_M(\theta^*|X_{obs})$

Swiss TPH

Metropolis-Hastings algorithm

The target distribution is our posterior $P_M(\theta|X_{obs})$.
We need an intialisation point θ_0 and a proposal $q(.\,|\,\theta_i)$

1. Initialise θ_0
2. **for** $\boldsymbol{i = 1 \ldots N}$
3. Sample θ^* from $q(.\,|\,\theta_i)$
4. Accept θ^* with probability $1 \wedge \frac{P_M(\theta^*|X_{obs})}{P_M(\theta_i|X_{obs})}\frac{q(\theta_i|\theta^*)}{q(\theta^*|\theta_i)}$
5. If θ^* is accepted, $\theta_{i+1} = \theta^*$. Otherwise $\theta_{i+1} = \theta_i$
6. **end for**
7. **return** $(\theta_0, \theta_1, \ldots, \theta_N)$

$$\frac{P_M(\theta^*|X_{obs})}{P_M(\theta_i|X_{obs})}$$

Let's plug the formulae

$$= \frac{\frac{P_M(X_{obs}|\theta^*)P_M(\theta^*)}{P_M(X)}}{\frac{P_M(X_{obs}|\theta_i)P_M(\theta_i)}{P_M(X)}}$$

This algorithm generates a Markov chain $(\theta_0, \theta_1, \ldots, \theta_N)$ that will converge to $P_M(\theta^*|X_{obs})$

Swiss TPH

Metropolis-Hastings algorithm

The target distribution is our posterior $P_M(\theta|X_{obs})$.
We need an intialisation point θ_0 and a proposal $q(.\,|\,\theta_i)$

1. Initialise θ_0
2. **for $\boldsymbol{i = 1 \ldots N}$**
3. Sample θ^* from $q(.\,|\,\theta_i)$
4. Accept θ^* with probability $1 \wedge \frac{P_M(\theta^*|X_{obs})}{P_M(\theta_i|X_{obs})}\frac{q(\theta_i|\theta^*)}{q(\theta^*|\theta_i)}$
5. If θ^* is accepted, $\theta_{i+1} = \theta^*$. Otherwise $\theta_{i+1} = \theta_i$
6. **end for**
7. **return** $(\theta_0, \theta_1, \ldots, \theta_N)$

$$\frac{P_M(\theta^*|X_{obs})}{P_M(\theta_i|X_{obs})}$$

Let's plug the formulae

$$= \frac{\frac{P_M(X_{obs}|\theta^*)P_M(\theta^*)}{\cancel{P_M(X)}}}{\frac{P_M(X_{obs}|\theta_i)P_M(\theta_i)}{\cancel{P_M(X)}}}$$

This algorithm generates a Markov chain $(\theta_0, \theta_1, \ldots, \theta_N)$ that will converge to $P_M(\theta^*|X_{obs})$

Swiss TPH

Metropolis-Hastings algorithm

The target distribution is our posterior $P_M(\theta|X_{obs})$.
We need an intialisation point θ_0 and a proposal $q(.\,|\,\theta_i)$

1. Initialise θ_0
2. **for** $\boldsymbol{i = 1 \ldots N}$
3. Sample θ^* from $q(.\,|\,\theta_i)$
4. Accept θ^* with probability $1 \wedge \frac{P_M(\theta^*|X_{obs})}{P_M(\theta_i|X_{obs})}\frac{q(\theta_i|\theta^*)}{q(\theta^*|\theta_i)}$
5. If θ^* is accepted, $\theta_{i+1} = \theta^*$. Otherwise $\theta_{i+1} = \theta_i$
6. **end for**
7. **return** $(\theta_0, \theta_1, \ldots, \theta_N)$

$$\frac{P_M(\theta^*|X_{obs})}{P_M(\theta_i|X_{obs})}$$

Let's plug the formulae

$$= \frac{P_M(X_{obs}|\theta^*)P_M(\theta^*)}{P_M(X_{obs}|\theta_i)P_M(\theta_i)}$$

We have simplified the unknown denominator !!!

This algorithm generates a Markov chain $(\theta_0, \theta_1, \ldots, \theta_N)$ that will converge to $P_M(\theta^*|X_{obs})$

Swiss TPH

Metropolis-Hastings algorithm

The target distribution is our posterior $P_M(\theta|X_{obs})$.
We need an intialisation point θ_0 and a proposal $q(.\mid \theta_i)$

1. Initialise θ_0
2. **for** $\boldsymbol{i = 1 \ldots N}$
3. Sample θ^* from $q(.\mid \theta_i)$
4. Accept θ^* with probability $1 \wedge \frac{P_M(X_{obs}|\theta^*)P_M(\theta^*)q(\theta_i|\theta^*)}{P_M(X_{obs}|\theta_i)P_M(\theta_i)q(\theta^*|\theta_i)}$
5. If θ^* is accepted, $\theta_{i+1} = \theta^*$. Otherwise $\theta_{i+1} = \theta_i$
6. **end for**
7. **return** $(\theta_0, \theta_1, \ldots, \theta_N)$

This algorithm generates a Markov chain $(\theta_0, \theta_1, \ldots, \theta_N)$ that will converge to $P_M(\theta^*|X_{obs})$

Swiss TPH

Metropolis-Hastings algorithm: applied to our disease model

The target distribution is our posterior $P_M(\theta|X_{obs})$.
We need an intialisation point θ_0 and a proposal $q(.\,|\,\theta_i)$

1. Initialise θ_0
2. **for $\boldsymbol{i = 1 \ldots N}$**
3. Sample $\theta^* = (\lambda^*, \gamma^*)$ from $q(.\,|\,\theta_i)$
4. Simulate the SIS with $\theta^*=(\lambda^*, \gamma^*)$. Extract the curve $I^*_{1,,,n}$
5. Compute the value: $P_M(X_{obs}|\,\theta^*) = L_M(X_{obs}, \theta^*) = \prod_{i=1}^{n} \frac{I_i^{X_i^{obs}} e^{-I_i}}{X_i^{obs}!}$

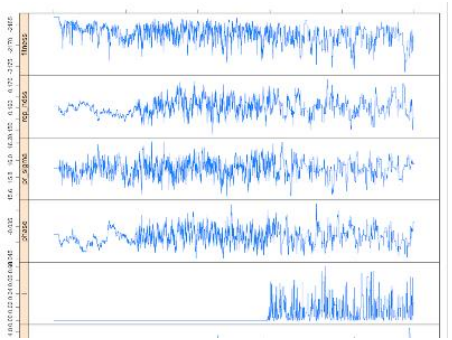
4. Accept θ^* with probability $1 \wedge \frac{P_M(X_{obs}|\theta^*)P_M(\theta^*)q(\theta_i|\theta^*)}{P_M(X_{obs}|\theta_i)P_M(\theta_i)q(\theta^*|\theta_i)}$
5. If θ^* is accepted, $\theta_{i+1} = \theta^*$. Otherwise $\theta_{i+1} = \theta_i$
6. **end for**
7. **return** $(\theta_0, \theta_1, \ldots, \theta_N)$

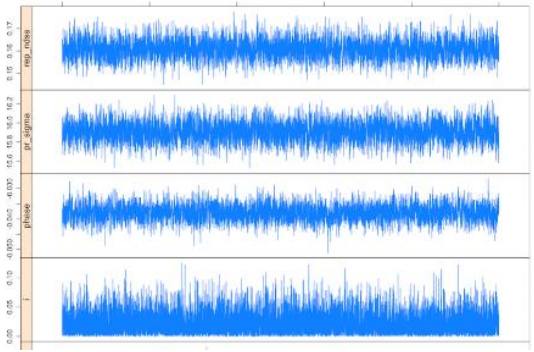
In practice

It is important to check that the chain has converged!

This can be done visually (caterpillar-like) or with specific convergence tests

It is usually recommended to sample several independent chains and ensure they converge to the same distribution.





Metropolis-Hastings is one type of MCMC:
there are other types of MCMC (e.g. Gibbs sampling, HMC, etc.)

And there are many other Bayesian inference methods that are not MCMC.

Swiss TPH

In practice

- Tomorrow, we are going to apply this methodology with an example dataset in a practical.

→You will better understand how it works !!

In practice, you can use libraries where such algorithms are already implemented (e.g. Stan, mc-state, libbi, pomp, etc.). But for this course, we are going to code it ourselves.

What to remember

- What is Bayesian inference for a disease model
- What is MCMC
- How to use MCMC to fit a disease model (e.g. SIS model)

Thank you !

clara.champagne@swisstph.ch

Questions?

Swiss TPH