Swiss TPH

Introduction to model fitting

August 2025

Emilie Pothin – emilie.pothin@swisstph.ch

Course objective

At the end of this lecture, you should be able to:

1. Describe the principle of model fitting
2. Understand the different methods that exist for model fitting
3. Fit a simple model to data

Quick recap

Why modelling?

- **Understanding disease dynamics :** Provide insight on how disease spead and what factor influences outbreaks
- **Forecasting:** Predict future cases, epidemic peaks and potential resurgence.
- **Evaluating impact of interventions :** Asses the impact of control and elimination measures
- **Informing policy:** Help identify the most effective ways to allocate resources and prepare for outbreak responses.

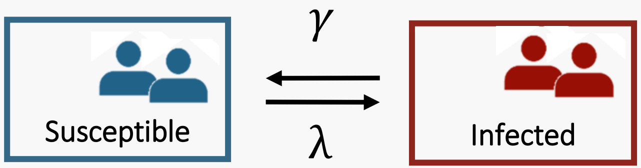
BUT, only useful if models can accurately represent real-world situations

What is an infectious disease model?

Infectious disease models: Sets of mathematical equations that describe the transmission dynamics of the disease within a population.

Already studied examples:

SIS model



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

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

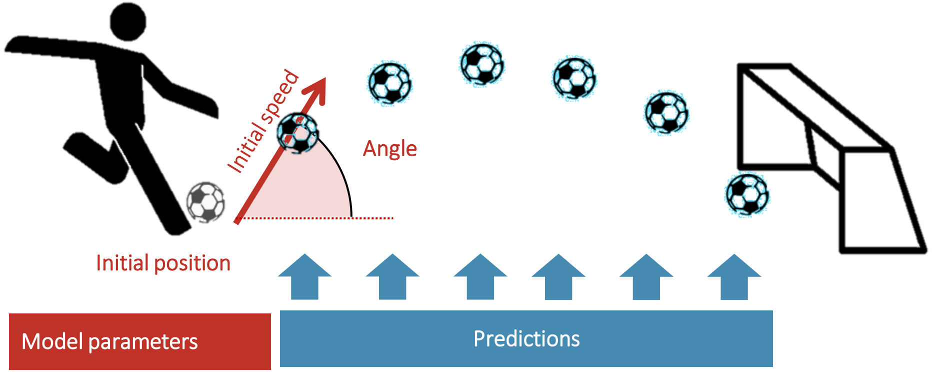
SIR model

$$\frac{dS}{dt} = -\beta \frac{SI}{N}$$

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

$$\frac{dR}{dt} = \gamma I$$

Mathematical Models

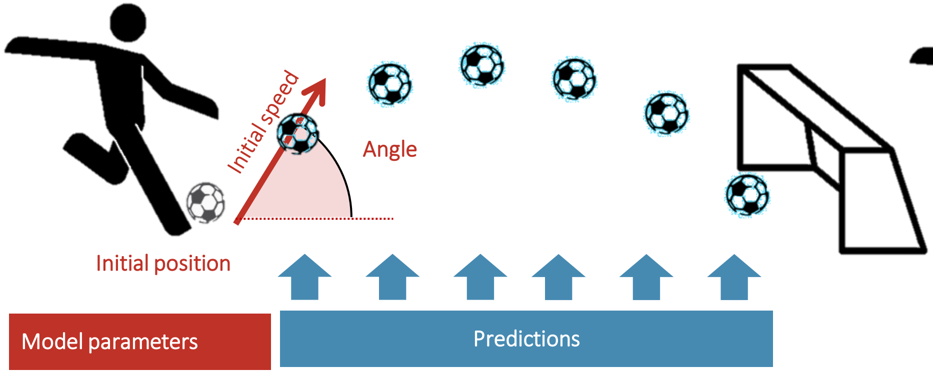


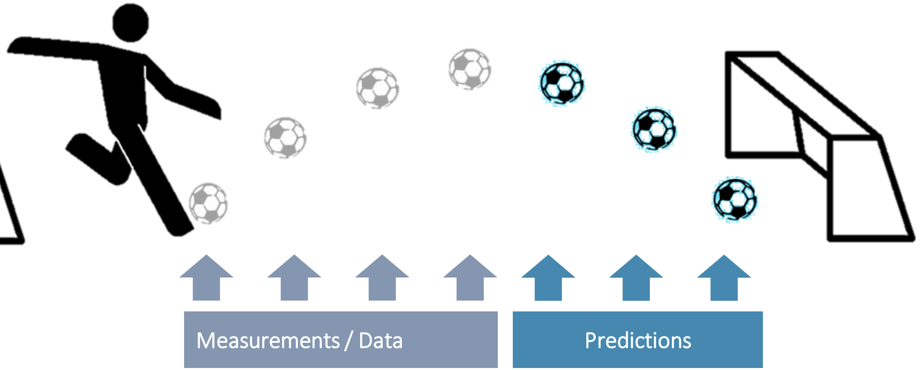
A mathematical model is build on our knowledge of the mechanisms that govern the dynamics of a system.

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

Mathematical Models vs Statistical Models

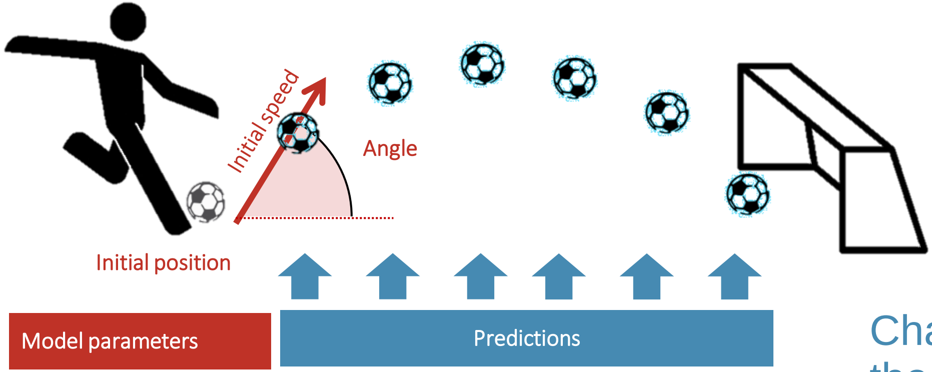
Statistical models



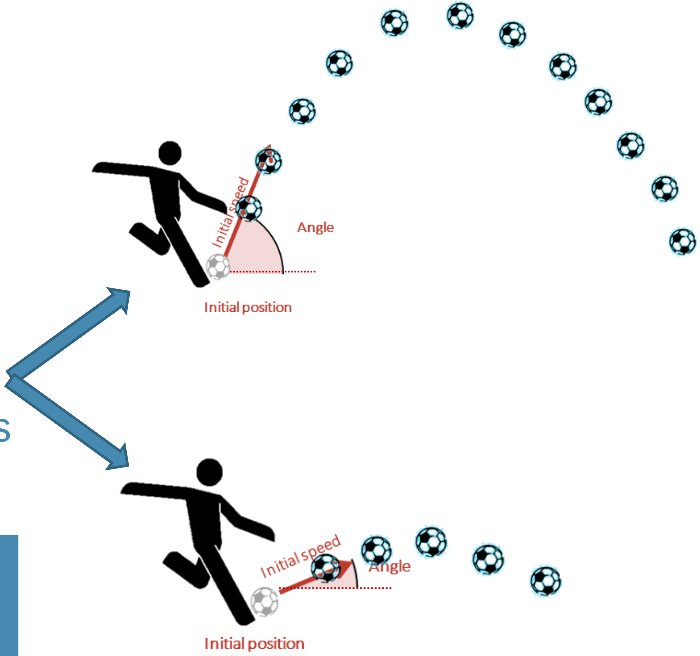


Statistical models rely on observed data to identify and analyze patterns within the available dataset, without explicitly incorporating the underlying mechanistic processes or dynamics governing the system.

The added value of Mathematical Models



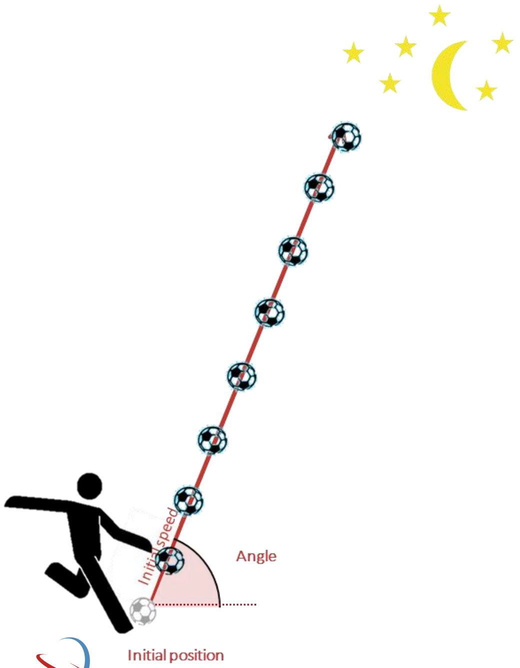
Change in the model parameters



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

How to choose model parameters?

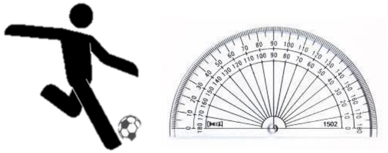
Parameters should be selected to reflect realistic scenarios



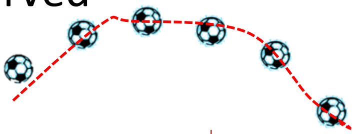
Data is key to inform model parameters

Two possible approaches:

1. Try to measure the parameters directly



2. Fit the parameters, i.e. finding the input parameter values that best reproduce observed model outputs



How to choose model parameters?

Parameters should be selected to reflect realistic scenarios

Data is key to inform model parameters

SIR model

$$\frac{dS}{dt} = -\beta \frac{SI}{N}$$

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

$$\frac{dR}{dt} = \gamma I$$

Two possible approaches:

1. Try to measure the parameters directly **or find known values for these**

Extracted from the literature (e.g. influenza)

- *R0: between 1.3 to 1.8 (Transmission rate 0.65-1.18)*
- *Recovery time: 1 to 2 weeks (Recovery rate 0.5-1)*

How to choose model parameters?

Parameters should be selected to reflect realistic scenarios

Data is key to inform model parameters

SIR model

$$\frac{dS}{dt} = -\beta \frac{SI}{N}$$

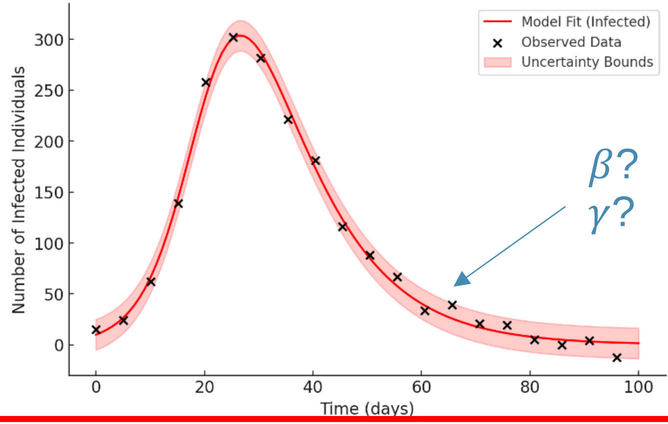
$$\frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

Two possible approaches:

Model fitting

2. Fit the parameters, i.e. finding the input parameter values that best reproduce observed model outputs

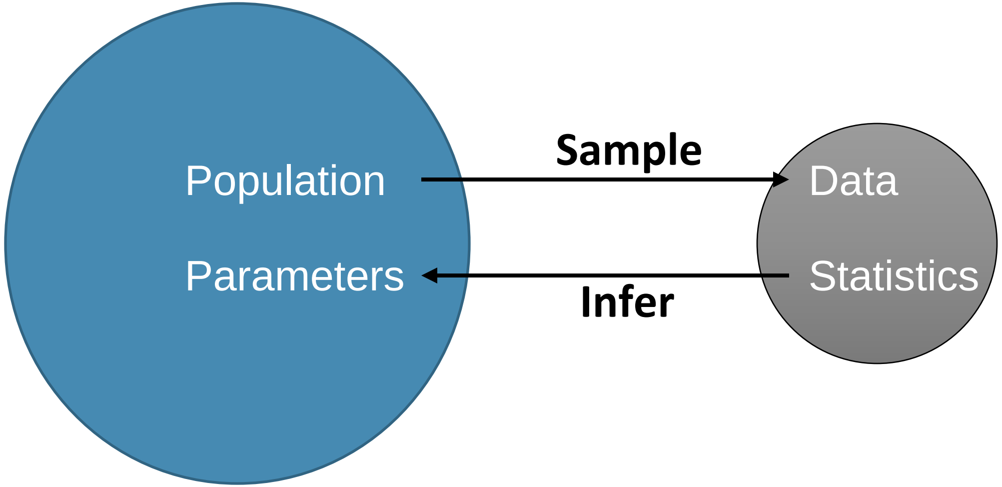


Principles of model fitting

Key concepts – Inference

Statistical Inference: the process of using data analysis techniques to infer properties of an underlying probability distribution based on sample data.

We want to learn about population parameters…



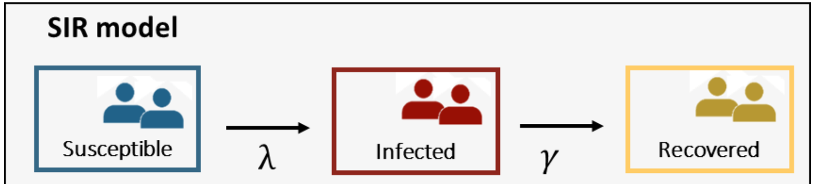
… using statistics calculated in the sample

Adapted from Gerstman, B. Burt. Basic Biostatistics : Statistics for Public Health Practice. Second edition., Jones & Bartlett Learning, 2015.

Key concepts – Model fitting

Goal of model fitting: Using statistical inference to connect mathematical models to real-world data

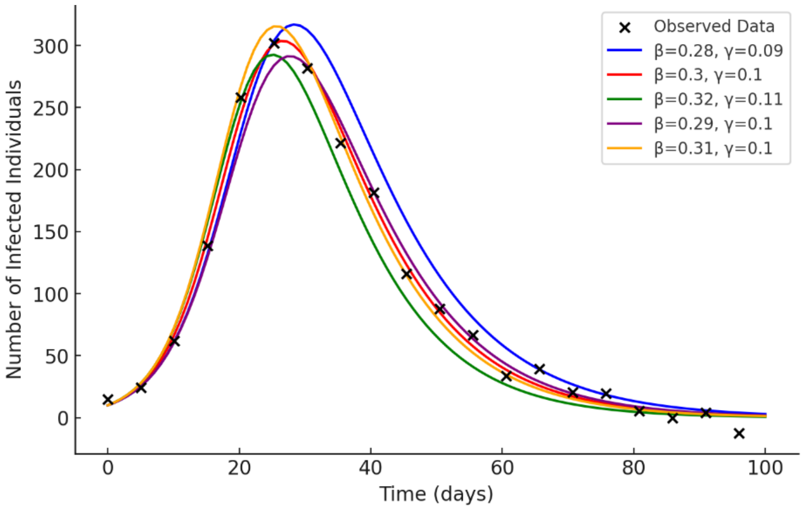
Model fitting consists in estimating model parameters of a mathematical model so to minimize the discrepancy between model predictions and observed data.



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

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

$$\frac{dR}{dt} = \gamma I$$



Key concepts – Parameter Estimation

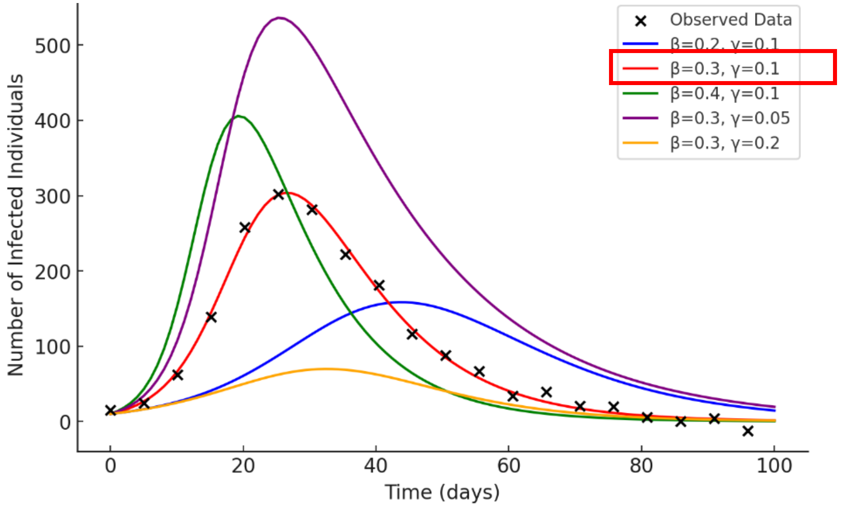
Estimating the parameters of a model means finding the parameters that make that model fit the data the best.

SIR model

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

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

$$\frac{dR}{dt} = \gamma I$$



Model fitting process

1 **Define problem statement & select model**

- Identify the question of interest
- Use understanding of disease mechanisms to define the model

2 **Collect & pre-process data**

- Gather and clean real-world malaria data
- Ensure the post-processed data and model outputs are in the same format

3 **Choose a fitting method & define objective function**

- Select a fitting method, e.g. Least Square Minimization, Maximum Likelihood Estimation or Bayesian approach
- The objective function quantifies the error between the model predictions and observed data (e.g. Sum of Squared Error or Likelihood)

4 **Fit the model & assess goodness of fit**

- Use an optimization algorithm to find parameter values that minimize the objective function.
- Plot observed vs. predicted curves to ensure good fit of the model to the data. Indicators can also be used to quantify goodness of fit.

5 **Validate, interpret & use the model**

- Validate model prediction with independent datasets or perform cross-validation using part of the data for fitting and part for validation.
- Use the fitted model for predictions and scenario analysis. Infer key epidemiological insights

In case of poor fit

If the model does not fit well, the process might need refinements

These can consist in:

- adjusting the model structure (e.g. adding external factors like seasonality)
- collecting more data
- testing alternative fitting techniques

Fitting methods

Fitting methods covered in this course

Least Square Minimization

Minimizes the difference between observed and model-predicted values

Maximum Likelihood Estimation

Finds parameter values that maximize the likelihood of observing the data

Bayesian Estimation

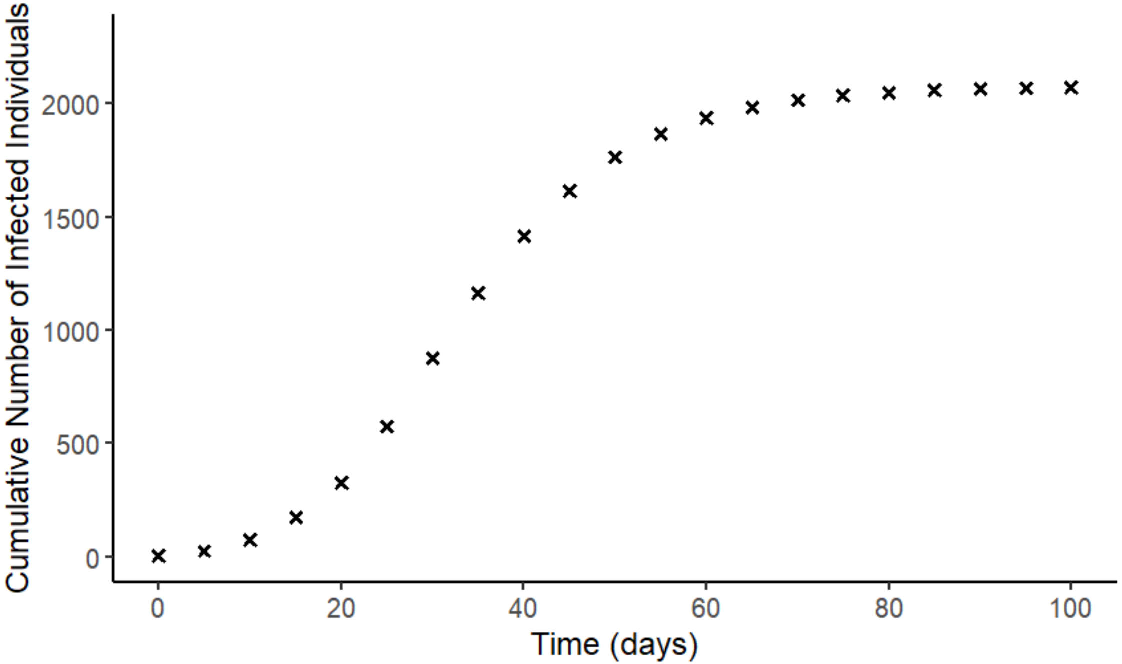
Uses prior knowledge and observed data to estimate parameters with uncertainty

Least Square Minimization

Let's assume this is the observed data we aim to reproduce with a model

Notations

$\boldsymbol{X_{obs}}$ the observed data
(here, the cumulative number of infected individuals)



Least Square Minimization

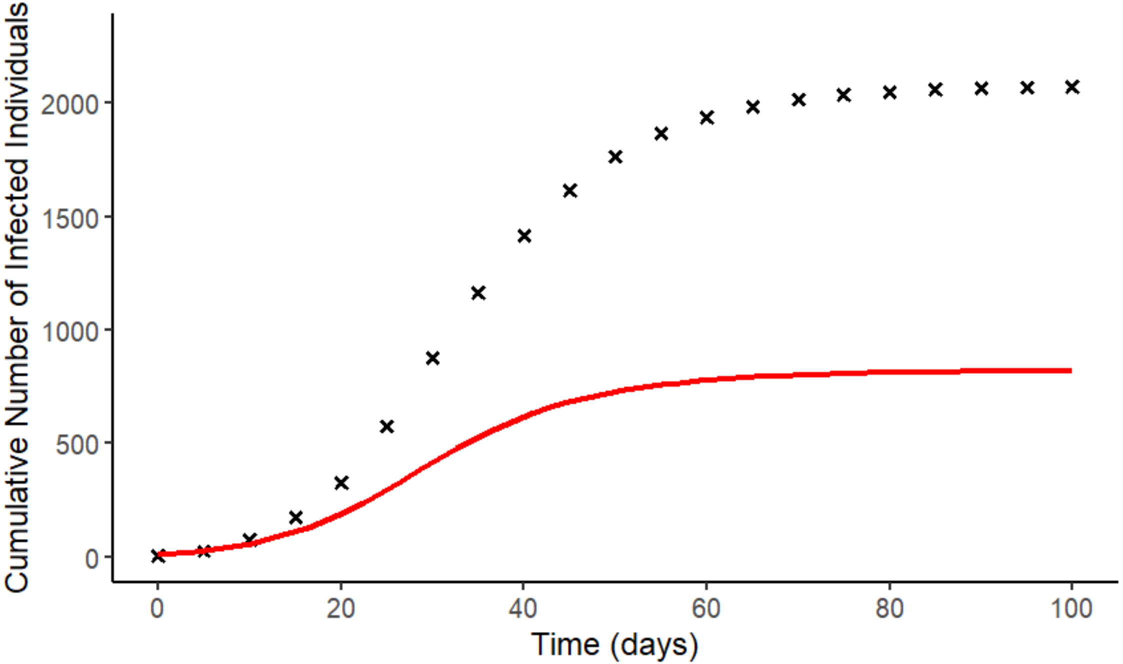
Models can simulate number of infected individuals based on parameters $\boldsymbol{\theta}$

Notations

$\boldsymbol{X_{obs}}$ the observed data

$\boldsymbol{\theta}$ the set of model parameters

$\boldsymbol{X_{sim_\theta}}$ the model prediction for parameters $\boldsymbol{\theta}$



Least Square Minimization

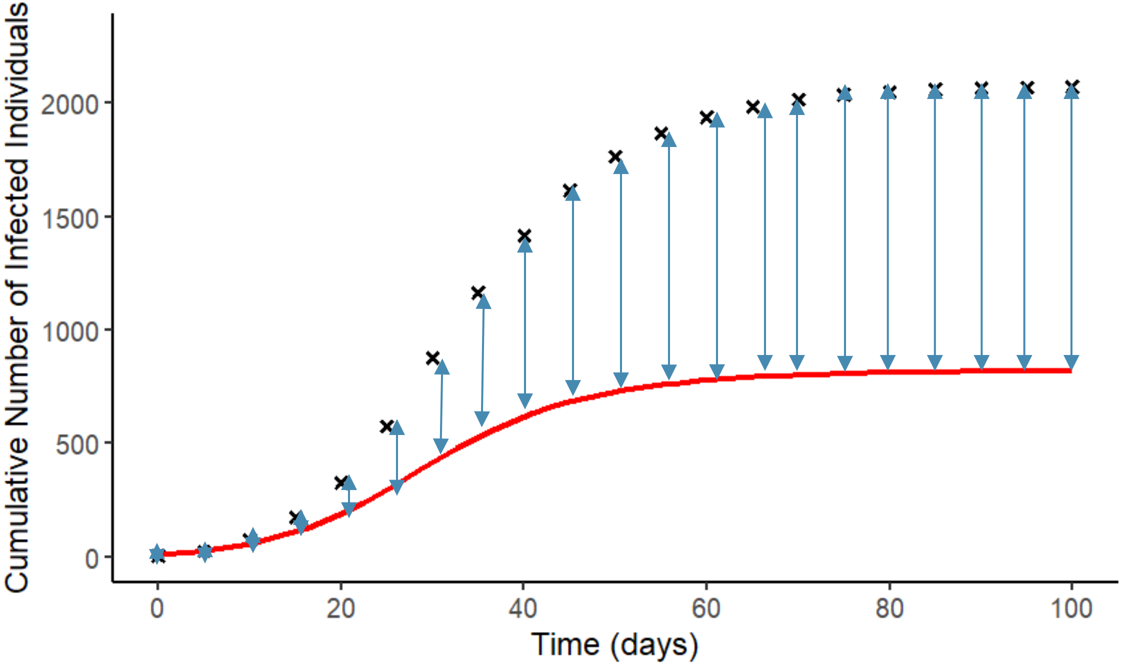
The goal is to find $\boldsymbol{\theta}$ that minimizes the distance between the model predictions and the observed data

Notations

$\boldsymbol{X_{obs}}$ the observed data

$\boldsymbol{\theta}$ the set of model parameters

$\boldsymbol{X_{sim_{\theta}}}$ the model prediction for parameters $\boldsymbol{\theta}$



Least Square Minimization

We define a loss function, **the least square distance,** to compare the model predictions and the observed data

Notations

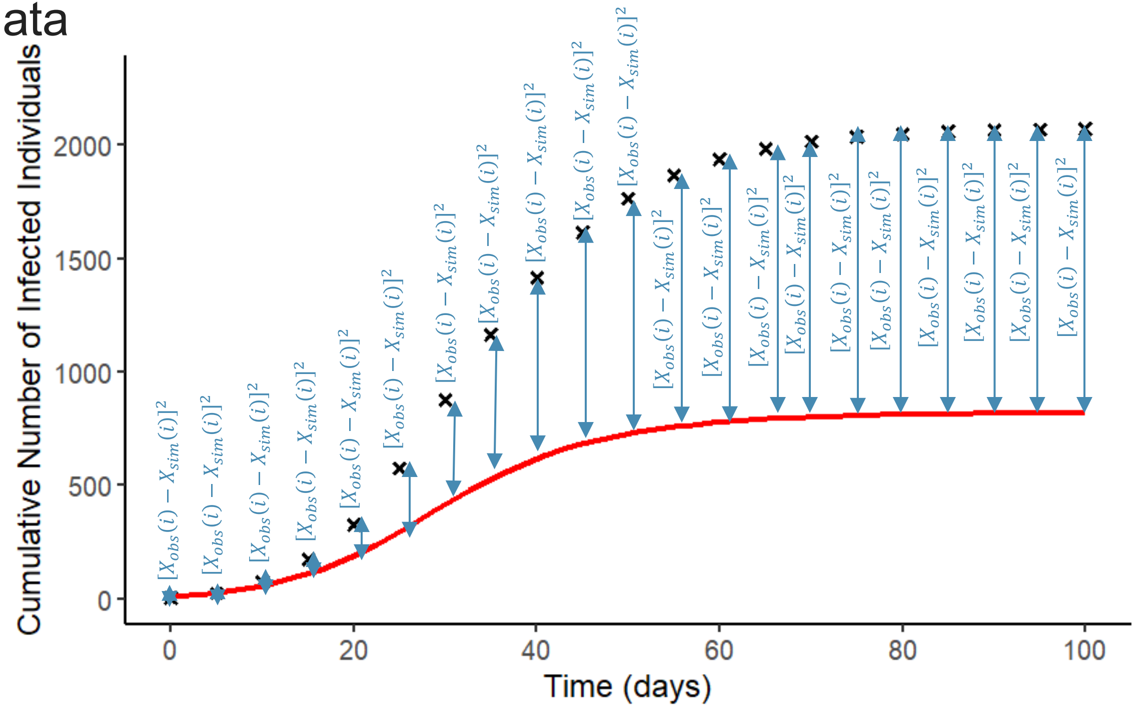
$\boldsymbol{X_{obs}}$ the observed data

$\boldsymbol{\theta}$ the set of model parameters

$\boldsymbol{X_{sim_\theta}}$ the model prediction for parameters $\boldsymbol{\theta}$

Least square distance

$$\sum_i \left[X_{obs}(i) - X_{sim_\theta}(i)\right]^2$$



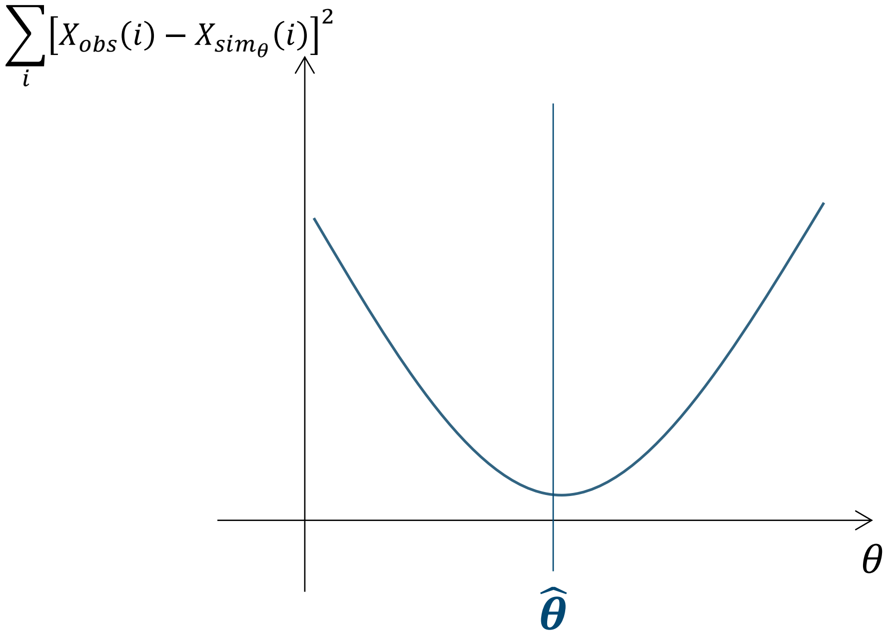
Note: Least square distance = Sum of squared residuals

Least Square Minimization

The objective is to find the parameters $\hat{\boldsymbol{\theta}}$ that **minimizes** the **least square distance**

$$\min_{\boldsymbol{\theta}} \sum_i \left[X_{obs}(i) - X_{sim_\theta}(i)\right]^2$$

Methods for minimization include gradient descent, newton's method,…

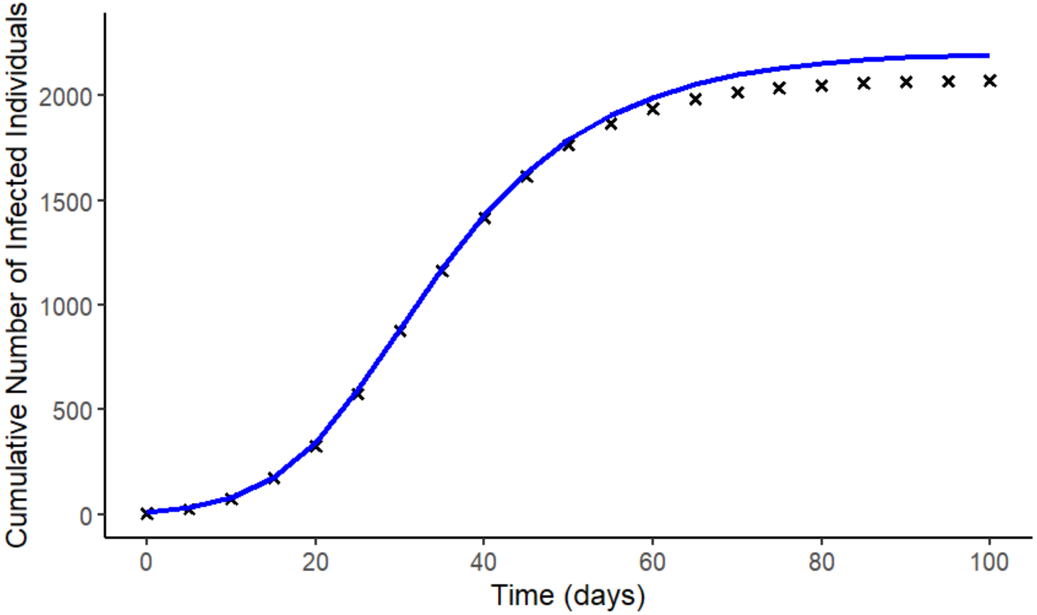


Least Square Minimization

The objective is to find the parameters $\hat{\boldsymbol{\theta}}$ that **minimizes** the **least square distance**

$$\min_{\boldsymbol{\theta}} \sum_i \left[X_{obs}(i) - X_{sim_\theta}(i)\right]^2$$

$$\hat{\theta} = (\hat{\beta} = 0.3 \;; \hat{\gamma} = 0.1)$$



Least Square Minimization

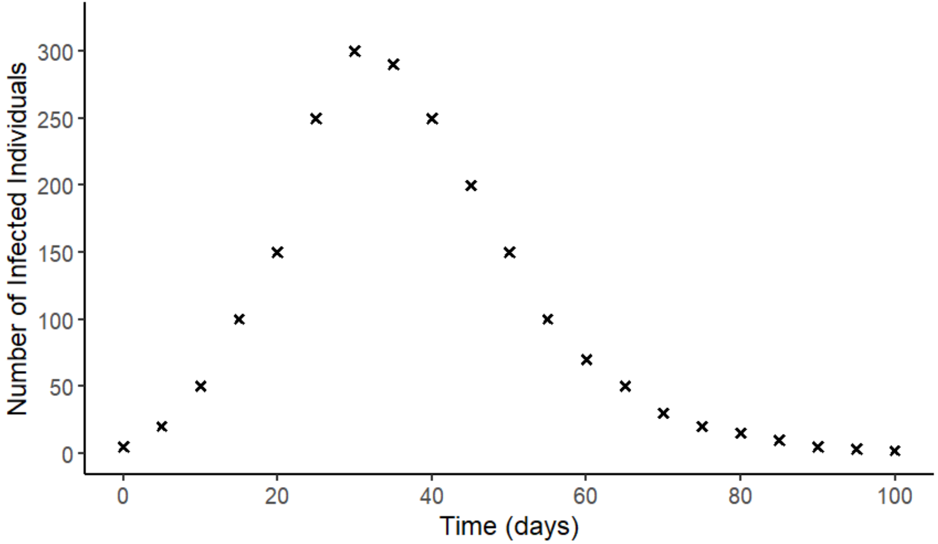
Limitations

- Not robust to non-normal errors (assumes normally distributed residuals)
- Does not model uncertainty explicitly
- Ideal for continuous data with normally distributed errors but not appropriate for count data, prevalence or categorical outcomes

Maximum Likelihood Estimation

Likelihood methods allows for more flexibility for the data to be fitted to the model, including **case counts and proportions**

Notations
X_{obs} the observed data (here, the number of infected individuals)



Maximum Likelihood Estimation

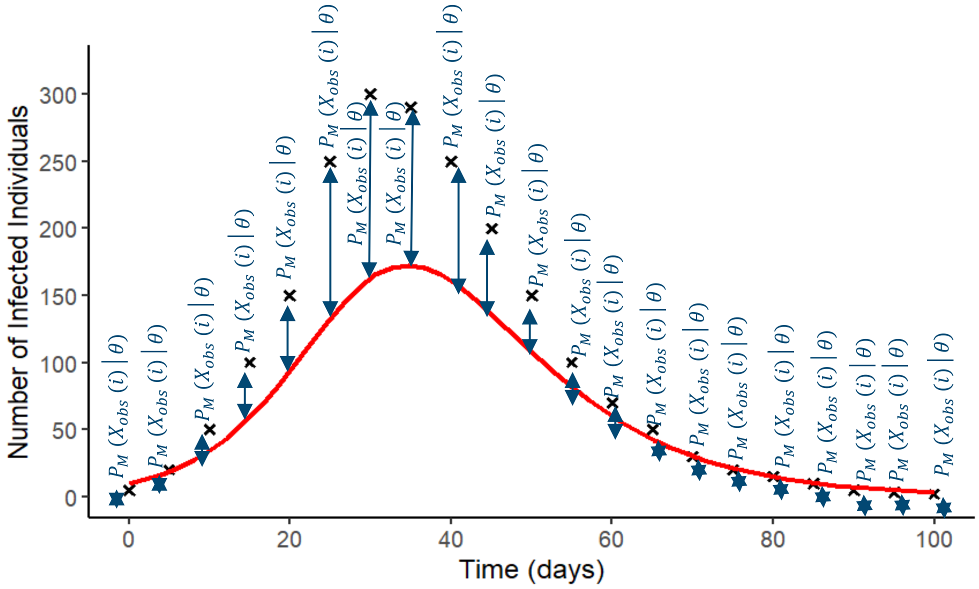
The **likelihood function** measures how well a given set of parameters $\boldsymbol{\theta}$ explains the observed data $\boldsymbol{X_{obs}}$

Likelihood function:
Probability of observing the data given θ

$$L(\theta) = P_M(X_{obs}|\theta) = \prod_i P_M(X_{obs}(i)|\theta)$$

Common likelihood functions:

- **Poisson** for count data
- **Negative Binomial** for over dispersed case counts
- **Binomial** for proportions
- **Gaussian** for continuous outputs like prevalence



Swiss TPH

Maximum Likelihood Estimation

The objective is to find the parameters $\hat{\boldsymbol{\theta}}$ that **maximize** the **likelihood function**

Notations

$\boldsymbol{X_{obs}}$ the observed data

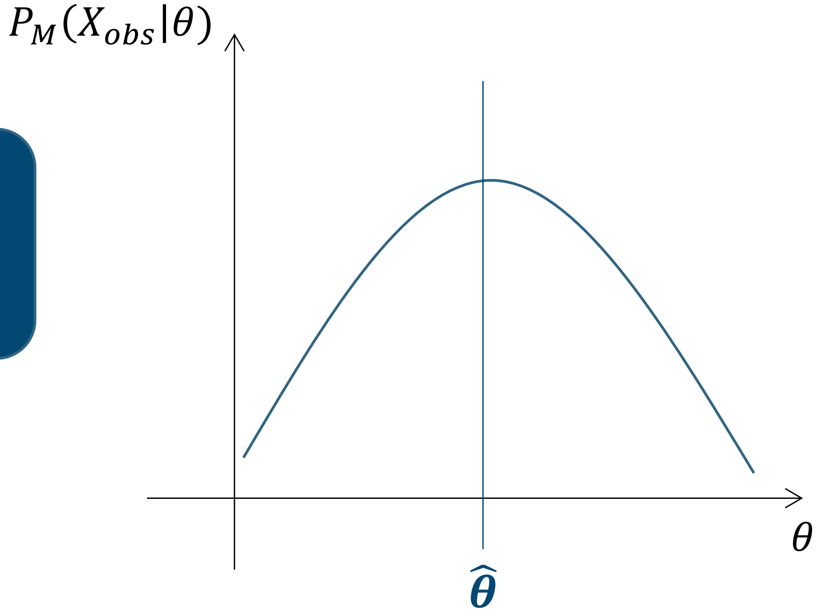
$\boldsymbol{\theta}$ the set of model parameters

Likelihood function:

Probability of observing the data given θ

$$L(\theta) = P_M(X_{obs}|\theta) = \prod_i P_M(X_{obs}(i)|\theta)$$

$$\hat{\theta} = \underset{\theta}{\arg\max}\{L(\theta)\}$$



Note: Maximizing $\boldsymbol{L(\theta)}$ is equivalent to minimizing $-\log\{\boldsymbol{L(\theta)}\}$

Maximum Likelihood Estimation

Simple example

Example of the SIR model
(we observe the number of infected individuals)

- Transition equation:

$$\frac{dS}{dt} = -\beta \frac{SI}{N}$$

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

$$\frac{dR}{dt} = \gamma I$$

$$I_{1,,,n}$$
$$\theta = (\beta, \gamma)$$

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

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

i	1	2	3	4	5	6	7	8	9	10	11
Time	0	10	20	30	40	50	60	70	80	90	100
X_i^{obs}	5	50	150	300	250	150	70	30	15	5	2
I_i	10	65	240	290	178	88	40	18	8	3	1

$\theta = (\beta = 0.3, \gamma = 0.1)$

L_i	$3.8E^{-2}$	$3.3E^{-11}$	$1.2E^{-10}$	$1.9E^{-2}$	$7.4E^{-8}$	$5.4E^{-10}$	$7.1E^{-6}$	$2.8E^{-3}$	$9.1E^{-3}$	$1.3E^{-1}$	$2.5E^{-1}$

$\theta = (\beta = 0.3, \gamma = 0.1)$ ➡ $L = \prod_i L_i$ = 7.11 x 10^{-52}
Log_likelihood: -117.77

Note: Maximizing $L(\theta)$ is equivalent to minimizing $-\log\{L(\theta)\}$

Maximum Likelihood Estimation

Simple example

Example of the SIR model
(we observe the number of infected individuals)

- Transition equation: $\frac{dS}{dt} = -\beta\frac{SI}{N}$

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

$$\frac{dR}{dt} = \gamma I$$

$I_{1,,,n}$
$\theta = (\beta, \gamma)$

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

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

$\theta = (\beta = 0.28, \gamma = 0.09)$ ➡ *-log_Likelihood* = 71.65

$\theta = (\beta = 0.30, \gamma = 0.10)$ ➡ *-log_Likelihood* = 117.77

$\theta = (\beta = 0.32, \gamma = 0.11)$ ➡ *-log_Likelihood* = 195.47

$\theta = (\beta = 0.29, \gamma = 0.10)$ ➡ *-log_Likelihood* = 92.65

$\theta = (\beta = 0.31, \gamma = 0.10)$ ➡ *-log_Likelihood* = 146.94

Note: Maximizing $L(\theta)$ is equivalent to minimizing $-\log\{L(\theta)\}$

Maximum Likelihood Estimation

The objective is to find the parameters $\hat{\boldsymbol{\theta}}$ that **maximize** the **likelihood function**

Notations

$\boldsymbol{X_{obs}}$ the observed data

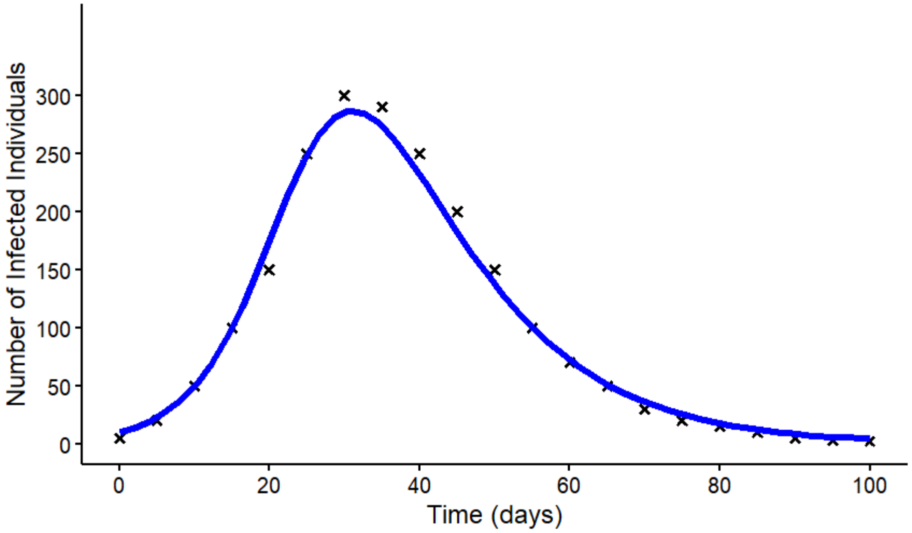
$\boldsymbol{\theta}$ the set of model parameters

Likelihood function:

Probability of observing the data given θ

$$L(\theta) = P_M(X_{obs}|\theta) = \prod_i P_M(X_{obs}(i)|\theta)$$

$$\hat{\theta} = (\hat{\beta} = 0.26 \; ; \hat{\gamma} = 0.09)$$

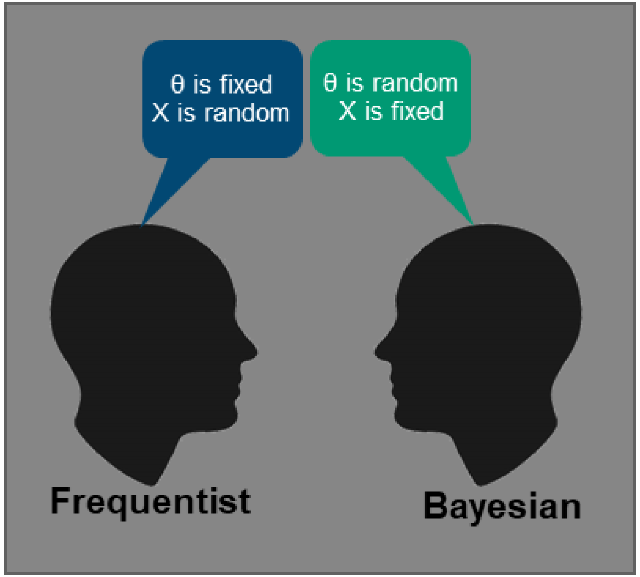


Maximum Likelihood Estimation

Limitations

- Computationally more intensive than Least Square method
- Depends on correct likelihood assumptions
- Sensitive to outliers
- Uncertainty estimation is not straight forward (need likelihood ratio test or bootstrapping)

Frequentist vs Bayesian approaches



Frequentist framework:

- The true value for θ is considered fixed but unknown.
- Point estimate $\hat{\boldsymbol{\theta}}$ represents a single set of values believed to be close to the true parameter value, with uncertainty quantified through confidence intervals.

Bayesian framework:

- The true value for θ is considered random with a probability distribution representing the uncertainty.
- We are looking for the regions of the parameter space where it is most likely to be, given the data we have, i.e. $P_M(\theta|X_{obs})$

Bayesian Estimation

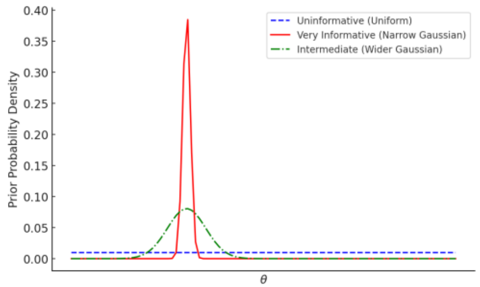
Bayesian approach incorporates **prior beliefs** and updates them with **observed data** to obtain **posterior distributions**.

Marginal Likelihood
Same as Likelihood function for MLE

Prior distribution

Posterior distribution:

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



Expected value of the likelihood function
Unknown most of the time and equation simplified as:

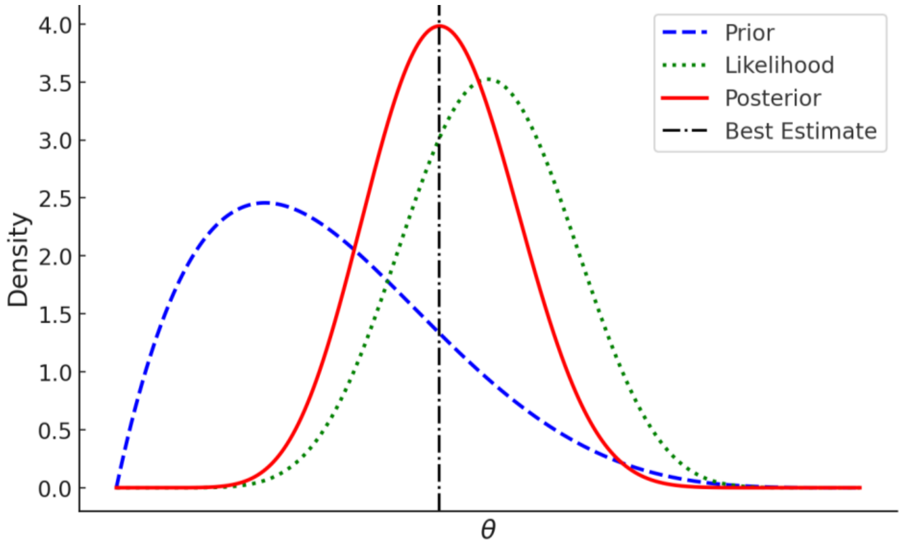
$$\boldsymbol{Posterior\ distribution \propto Likelihood\ function\ * Prior\ distribution}$$

Bayesian Estimation

Posterior distribution:

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

- The **best estimate** of the parameters is determined from the **mean of the posterior distribution**.
- **Credible intervals** are derived from the posterior distribution to provide a probability statement about the parameters, e.g. there is a 95% probability that θ lies within this credible interval



Bayesian Estimation

A few notes

- When paired with a specific likelihood, conjugate prior yields a posterior of the same family, allowing for analytical inference
- The data are much more influential over the posterior distribution than the prior except when the sample size and/or the prior is very informative (small variance)
- If the posterior distribution is complex (no closed form), use MCMC sampling methods like (Metropolis-Hastings, Gibbs Sampling,…)

Data type	Prior	Likelihood	Posterior
Binary	Beta	Bernouilli	Beta
Count	Gamma	Poisson	Gamma
Overdispersed counts	Gamma	Negative Binomial	No closed form
Proportions	Beta	Binomial or Beta	Beta
Continuous	Normal	Normal	Normal

Will often be the case for infectious disease models due to non-linearities and complex likelihood

Bayesian Estimation

Limitations

- Computationally extensive
- Choice of prior can be subjective
- Harder to implement and interpret
- Prior can dominate results

In conclusion

Key considerations

- ✓ **Assess goodness of fit**: Ensure the model captures key data patterns using appropriate metrics when relevant

- ✓ **Account for uncertainty**: Incorporate confidence/credible intervals and stochasticity to reflect real-world variability.

- ✓ **Consider computational efficiency**: Balance model complexity and runtime to avoid unnecessary overfitting or impractical simulations.

- ✓ **Validate with independent data**: Test as much as possible model predictions on new datasets to confirm reliability and generalizability.

- ✓ **Check identifiability**: Ensure parameters can be uniquely estimated—avoid overparameterization or poorly constrained models.

Key take away messages

Model fitting aims to **align model predictions with observed data** as closely as possible

Parameters are estimated by optimizing the model to match data using appropriate methods.

The choice of **fitting method** depends on data type, model complexity, and analysis constraints.

The Bayesian approach provides greater flexibility in handling uncertainty but requires more computational effort

Swiss TPH

Practical on model fitting

August 2025

Emilie Pothin – emilie.pothin@swisstph.ch

The data

- Weekly reported data of malaria cases in an region which historically had very few cases of malaria
- Over a year, the annual incidence has increased from 10 to 110 cases per 1’000.
- Data recorded between January 1st and December 31st 2024
- Data includes the week number, the number of uncomplicated cases, the incidence and the population size.

Part 1:

Manual fitting

Part 1:

Data Visualisation & Model

1. Upload in R the file entitled “sis_si_weekly_cases.csv”
2. Visualize the data in R with ggplot
3. Code the ODE system (SIR-SI without interventions) in R with deSolve
4. Simulate and plot the model solution for $\alpha_{scale} = 0.8$ and then $\alpha_{scale} = 1.2$ overlaying the data
5. How well does the model correspond to the data?
6. Repeat for a few values of parameters searching for the best fit

Part 2:

Fitting using Least Square Error

Part 2:

Fit the model using the Least Square method (RMSE)

1. Define the model and the initial parameters and states
2. Write the objective function using the cumulative number of cases
3. Write the Least Square Error = $\sum_{t=1}^{52}(X_{obs}(t) - X_{sim_\theta}(t))^2$

   ✓ With $X_{obs}\ (t)$ the observed number of malaria cases

   ✓ $X_{sim_\theta}(t)$ the cumulated number of simulated cases of malaria at time t with parameters $\theta = \{\alpha_{scale}, \rho_{scale}\}$

4. Find the parameters that minimize the Least Square Error using the optim function
5. Extract the parameters and visualize the fit

Part 3:

Fitting using MLE

Practical 2: Maximum Likelihood Estimation

Model fitting

1. Write the likelihood linking the model to the data
   - Assuming a negative binomial function as the observational model
   - Total number of cases scaled with reporting rate ρ_scale and multiplying factor α_scale so that: $\mu_t = \rho\_scale \cdot I\_t(\alpha\_scale)$

2. Use the R function **nlminb** to minimize the Negative Log Likelihood function and estimate α_scale and ρ_scale
   - Assume α_scale and γ are bounded on a log-scale to keep the solver in stable regions (e.g., alpha_scale ∈ [0.05, 20], rho_scale ∈ [1e−3, 100], size ∈ [0.5, 1e4]).
   - Pick your start values using the method of moments for the minimization algorithm

3. Visualize the fit. What are the estimated parameters?

Practical 2: Maximum Likelihood Estimation

Model fitting

1. Refit the model with a Poisson distribution.
   - Replace the NB term with $\Sigma\_t - \log \mathrm{dpois}(Y\_t; \mu\_t)$
   - Drop size from θ.
2. How do the extracted values compare?

Swiss TPH

Emilie.Pothin@SwissTPH.ch

Acknowledgement: Clara Champagne and Cameline Orlendo

Gates Foundation