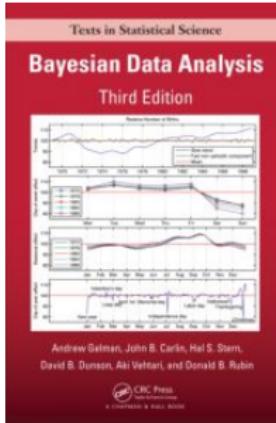


# Bayesian data analysis (Aalto fall 2025)

- Book: Gelman, Carlin, Stern, Dunson, Vehtari & Rubin: Bayesian Data Analysis, Third Edition. (online pdf available)
- The course website has more detailed information  
[https://avehtari.github.io/BDA\\_course\\_Aalto/Aalto2025.html](https://avehtari.github.io/BDA_course_Aalto/Aalto2025.html)
- Timetable: see the course website
- TAs: David Kohns, Osvaldo Martin, Wanxian Chen, Ningyi Peng, Maksim Sinelnikov, Reza Soumi, Herman Tesso, Quan Tran, Aristomenis Tsopelakos, Behram Ulukir



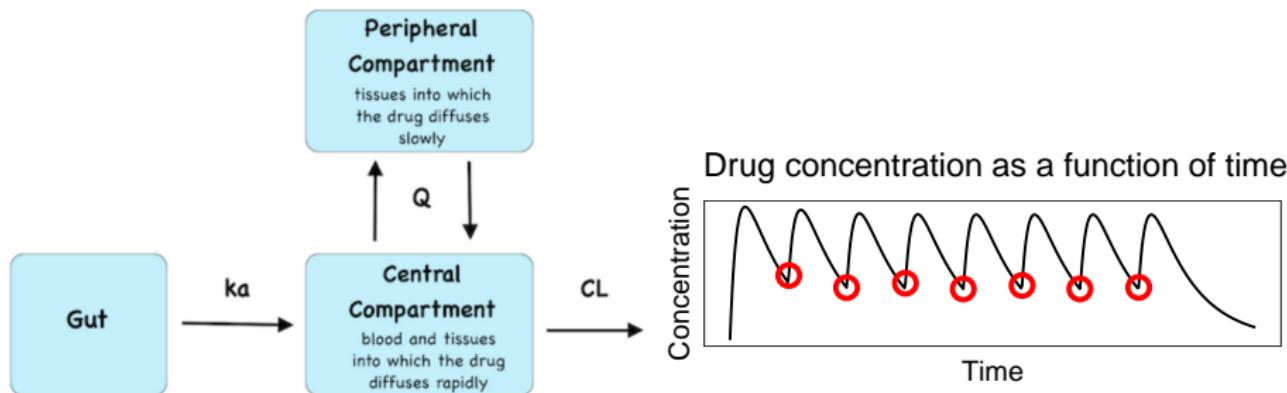
# Uncertainty and decision making

- Predicting concrete quality



# Uncertainty and decision making<sup>1</sup>

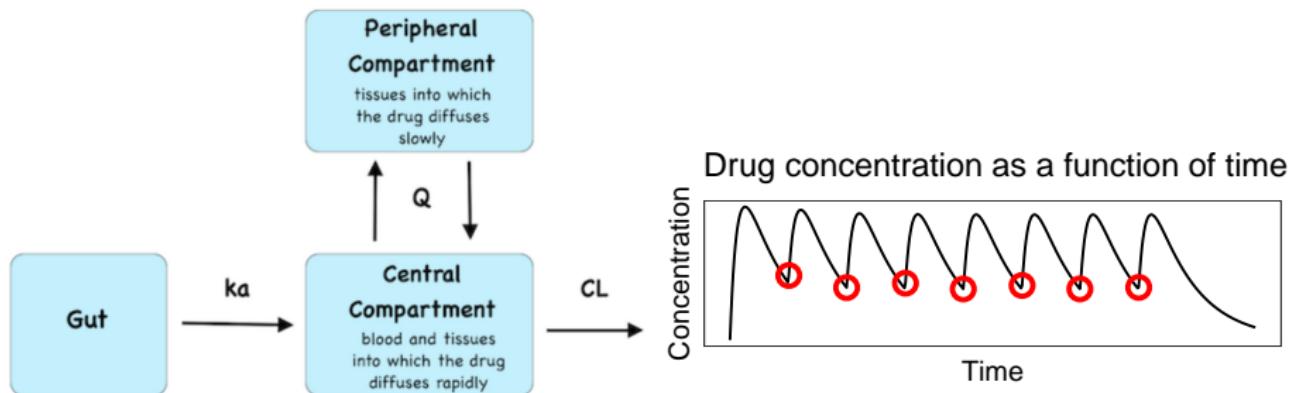
- Everolimus is immunosuppressant to prevent rejection of organ transplants
- Pharmacokinetic model of drug and body, optimal dosage depends on weight



<sup>1</sup>with E. Siivola, Aalto and S. Weber, Novartis Pharma

# Uncertainty and decision making<sup>1</sup>

- Everolimus is immunosuppressant to prevent rejection of organ transplants
- Pharmacokinetic model of drug and body, optimal dosage depends on weight

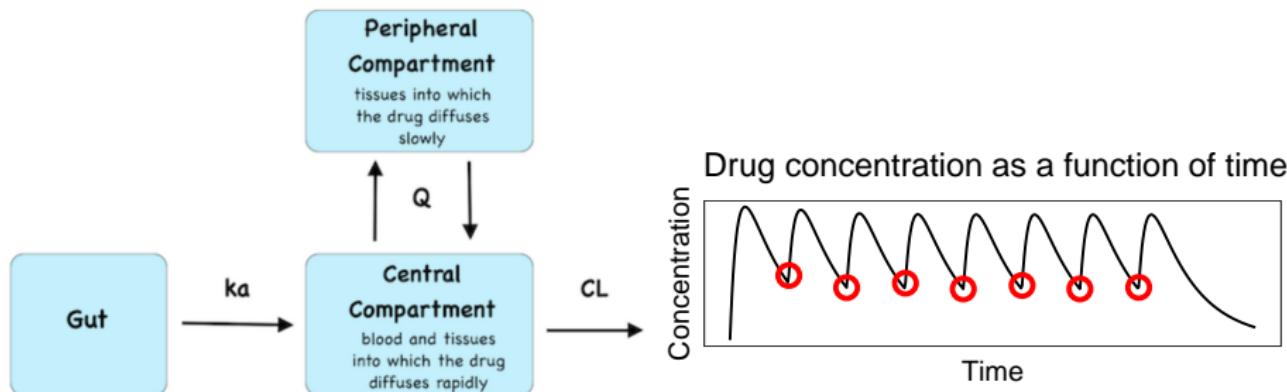


- Model fitted with 500 adults, extrapolation to children?

<sup>1</sup>with E. Siivola, Aalto and S. Weber, Novartis Pharma

# Uncertainty and decision making<sup>1</sup>

- Everolimus is immunosuppressant to prevent rejection of organ transplants
- Pharmacokinetic model of drug and body, optimal dosage depends on weight

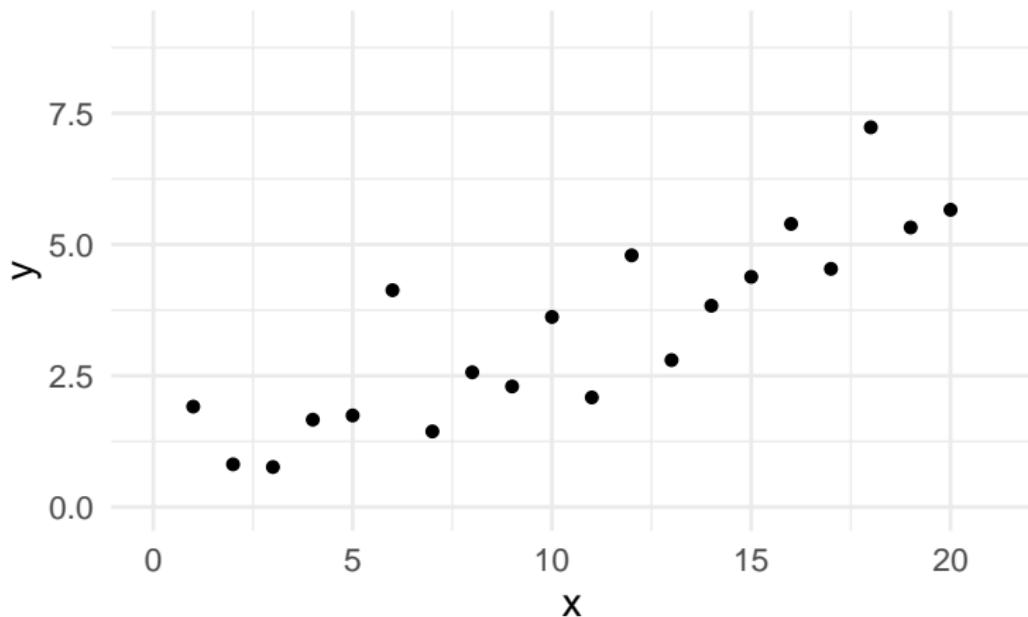


- Model fitted with 500 adults, extrapolation to children?
- Maturation effect, 17 observations from children

<sup>1</sup>with E. Siivola, Aalto and S. Weber, Novartis Pharma

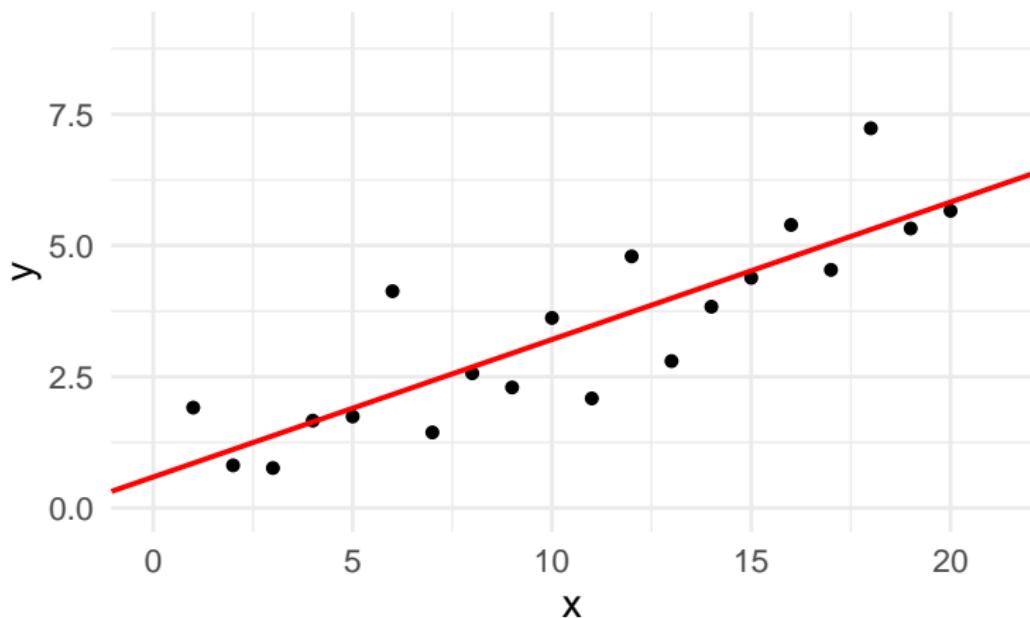
# Uncertainty in modeling

Data



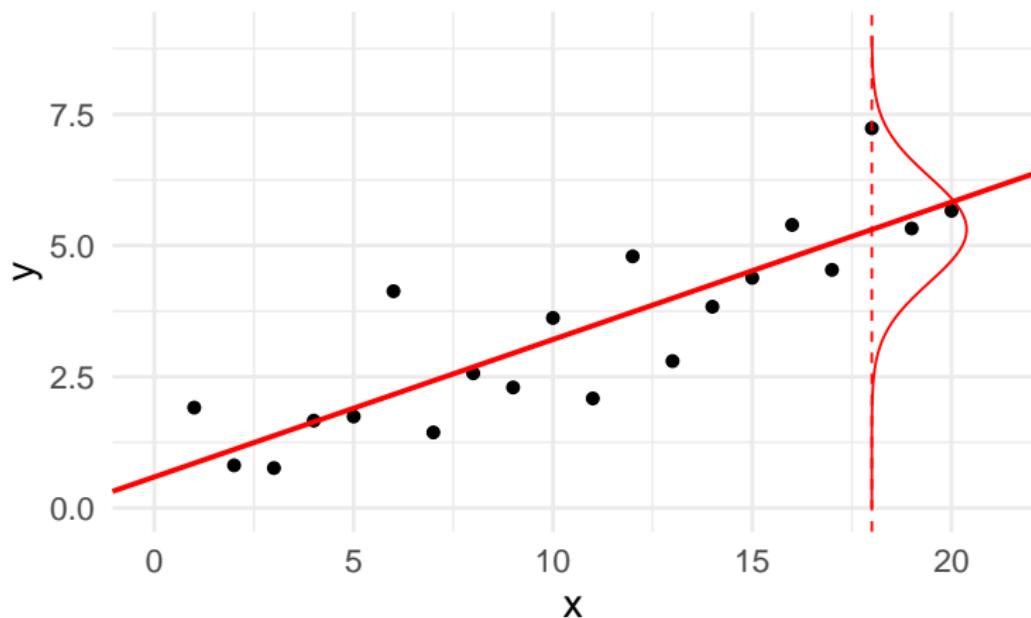
# Uncertainty in modeling

Posterior mean



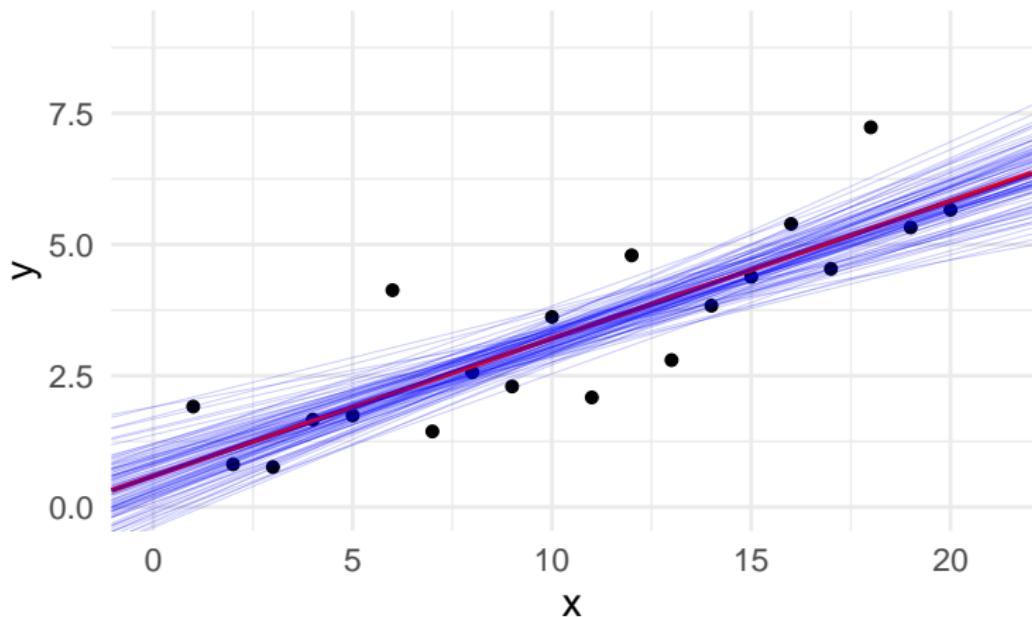
# Uncertainty in modeling

Predictive distribution given posterior mean



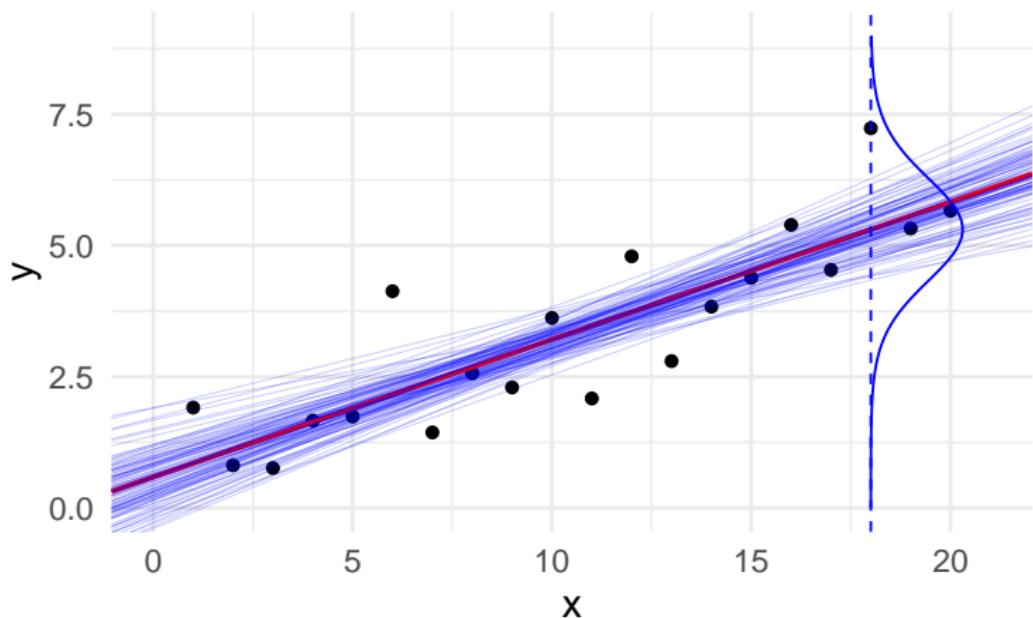
# Uncertainty in modeling

Posterior draws



# Uncertainty in modeling

Posterior draws and predictive distribution



# Bayesian probability theory

expert information

# Bayesian probability theory

expert information



mathematical model

+

uncertainty with probabilities

# Bayesian probability theory

expert information



data

mathematical model

+

uncertainty with probabilities

# Bayesian probability theory

expert information



mathematical model

+

uncertainty with probabilities

+

data

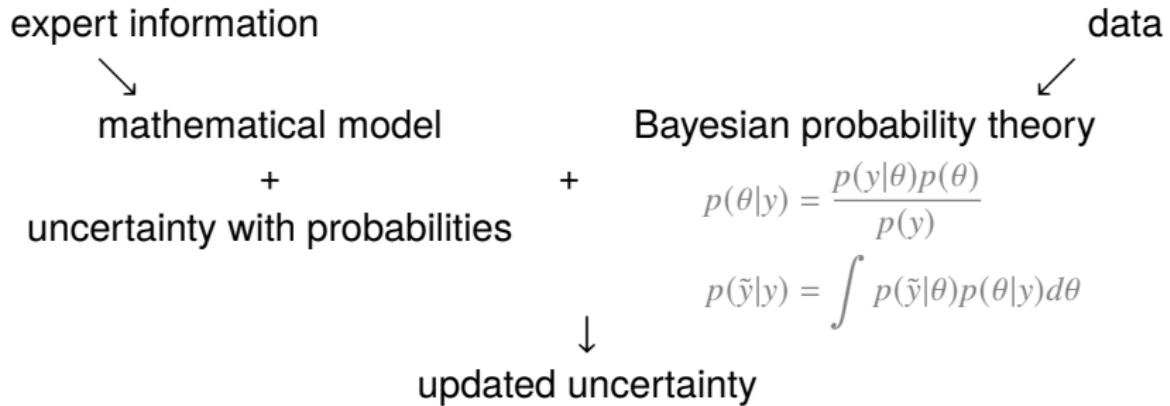


Bayesian probability theory

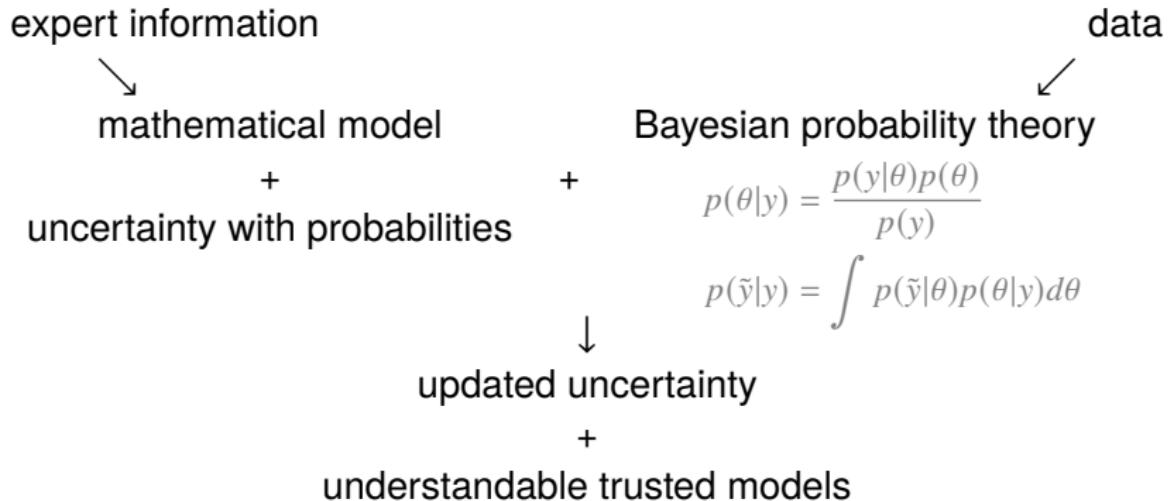
$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

$$p(\tilde{y}|y) = \int p(\tilde{y}|\theta)p(\theta|y)d\theta$$

# Bayesian probability theory



# Bayesian probability theory



# Bayesian inference with computers

mathematical model to computer

probabilistic programming

computation, automatic inference algorithms

limitations of computers

# Yes, but did it work?

computation + inference diagnostics

model diagnostics

limitations of mathematical models

improve and iterate

# Bayesian Workflow

All the pieces put together

# Probabilistic programming and Stan

Stan is a probabilistic programming framework and ecosystem  
40+ developers, 100+ contributors, 200K+ users, 300+ other  
software packages built on top of Stan



[mc-stan.org](http://mc-stan.org)

# Bayesian Data Analysis course

- Probability distributions as model building blocks
  - need to understand the math part (prereq.)
  - continuous vs discrete (prereq.)
  - observation model, likelihood, prior
  - constructing bigger models

# Bayesian Data Analysis course

- Probability distributions as model building blocks
  - need to understand the math part (prereq.)
  - continuous vs discrete (prereq.)
  - observation model, likelihood, prior
  - constructing bigger models
- Computation
  - We need to be able to compute expectations

$$E_{\theta|y} [g(\theta)] = \int p(\theta|y)g(\theta)d\theta$$

- when analytic solutions are not available, computational approximations with finite number of function evaluations
- grid, importance sampling, Monte Carlo, Markov chain Monte Carlo

# Bayesian Data Analysis course

- Probability distributions as model building blocks
  - need to understand the math part (prereq.)
  - continuous vs discrete (prereq.)
  - observation model, likelihood, prior
  - constructing bigger models
- Computation
  - We need to be able to compute expectations

$$E_{\theta|y} [g(\theta)] = \int p(\theta|y)g(\theta)d\theta$$

- when analytic solutions are not available, computational approximations with finite number of function evaluations
- grid, importance sampling, Monte Carlo, Markov chain Monte Carlo
- Workflow
  - steps of model building, inference, and diagnostics

## Impact on society

Better modelling and quantification of uncertainty

- better science
- better informed decision making  
in companies, government, and NGOs

# Bayesian probability theory

- Based on Bayesian probability theory
  - uncertainty is presented with probabilities
  - probabilities are updated based on new information
- Thomas Bayes (170?–1761)
  - English nonconformist, Presbyterian minister, mathematician
  - considered the problem of *inverse probability*
    - significant part of the Bayesian theory

# Bayesian probability theory

- Based on Bayesian probability theory
  - uncertainty is presented with probabilities
  - probabilities are updated based on new information
- Thomas Bayes (170?–1761)
  - English nonconformist, Presbyterian minister, mathematician
  - considered the problem of *inverse probability*
    - significant part of the Bayesian theory
- Bayes did not invent all, but was first to solve problem of inverse probability in special case
- Modern Bayesian theory with rigorous proofs developed in 20th century

# Bayesian probability theory

- Based on Bayesian probability theory
  - uncertainty is presented with probabilities
  - probabilities are updated based on new information
- Thomas Bayes (170?–1761)
  - English nonconformist, Presbyterian minister, mathematician
  - considered the problem of *inverse probability*
    - significant part of the Bayesian theory
- Bayes did not invent all, but was first to solve problem of inverse probability in special case
- Modern Bayesian theory with rigorous proofs developed in 20th century
- A nice book about history: Sharon Bertsch McGrayne, *The Theory That Would Not Die*, 2012.

## Term Bayesian used first time in mid 20th century

- Earlier there was just "probability theory"
  - concept of the probability was not strictly defined, although it was close to modern Bayesian interpretation
  - in the end of 19th century there were increasing demand for more strict definition of probability (mathematical and philosophical problem)

## Term Bayesian used first time in mid 20th century

- Earlier there was just "probability theory"
  - concept of the probability was not strictly defined, although it was close to modern Bayesian interpretation
  - in the end of 19th century there were increasing demand for more strict definition of probability (mathematical and philosophical problem)
- In the beginning of 20th century frequentist view gained popularity
  - accepts definition of probabilities only through frequencies
  - does not accept inverse probability or use of prior
  - gained popularity due to apparent objectivity and "cook book" like reference books

## Term Bayesian used first time in mid 20th century

- Earlier there was just "probability theory"
  - concept of the probability was not strictly defined, although it was close to modern Bayesian interpretation
  - in the end of 19th century there were increasing demand for more strict definition of probability (mathematical and philosophical problem)
- In the beginning of 20th century frequentist view gained popularity
  - accepts definition of probabilities only through frequencies
  - does not accept inverse probability or use of prior
  - gained popularity due to apparent objectivity and "cook book" like reference books
- R. A. Fisher used in 1950 first time term "Bayesian" to emphasize the difference to general term "probability theory"
  - term became quickly popular, because alternative descriptions were longer

## Term Bayesian used first time in mid 20th century

- Earlier there was just "probability theory"
  - concept of the probability was not strictly defined, although it was close to modern Bayesian interpretation
  - in the end of 19th century there were increasing demand for more strict definition of probability (mathematical and philosophical problem)
- In the beginning of 20th century frequentist view gained popularity
  - accepts definition of probabilities only through frequencies
  - does not accept inverse probability or use of prior
  - gained popularity due to apparent objectivity and "cook book" like reference books
- R. A. Fisher used in 1950 first time term "Bayesian" to emphasize the difference to general term "probability theory"
  - term became quickly popular, because alternative descriptions were longer
- The probabilistic programming revolution started in early 1990's

# Uncertainty and probabilistic modeling

- Two types of uncertainty: aleatoric and epistemic
- Representing uncertainty with probabilities
- Updating uncertainty

## Two types of uncertainty

- Aleatoric uncertainty due to randomness
- Epistemic uncertainty due to lack of knowledge

## Two types of uncertainty

- Aleatoric uncertainty due to randomness
  - we are not able to obtain observations which could reduce this uncertainty
- Epistemic uncertainty due to lack of knowledge

## Two types of uncertainty

- Aleatoric uncertainty due to randomness
  - we are not able to obtain observations which could reduce this uncertainty
- Epistemic uncertainty due to lack of knowledge
  - we are able to obtain observations which can reduce this uncertainty
  - two observers may have different epistemic uncertainty

## Updating uncertainty

## Updating uncertainty

- Probability of red  $\frac{\#red}{\#red + \#yellow} = \theta$

## Updating uncertainty

- Probability of red  $\frac{\#red}{\#red + \#yellow} = \theta$
- $p(y = \#red | \theta) = \theta$  aleatoric uncertainty

## Updating uncertainty

- Probability of red  $\frac{\#red}{\#red+\#yellow} = \theta$
- $p(y = \#red | \theta) = \theta$  aleatoric uncertainty
- $p(\theta)$  epistemic uncertainty

## Updating uncertainty

- Probability of red  $\frac{\#red}{\#red + \#yellow} = \theta$
- $p(y = \#red | \theta) = \theta$  aleatoric uncertainty
- $p(\theta)$  epistemic uncertainty
- Picking many chips updates our uncertainty about the proportion
- $p(\theta | y = \#red, \#yellow, \#red, \#red, \dots) = ?$

## Updating uncertainty

- Probability of red  $\frac{\#red}{\#red + \#yellow} = \theta$
- $p(y = \#red | \theta) = \theta$  aleatoric uncertainty
- $p(\theta)$  epistemic uncertainty
- Picking many chips updates our uncertainty about the proportion
- $p(\theta | y = \#red, \#yellow, \#red, \#red, \dots) = ?$
- Bayes rule  $p(\theta | y) = \frac{p(y | \theta)p(\theta)}{\int p(y | \theta)p(\theta)d\theta}$

## Model vs. likelihood

- Bayes rule  $p(\theta|y) \propto p(y|\theta)p(\theta)$
- Model:  $p(\mathbf{y}|\theta)$  as a function of  $\mathbf{y}$  given fixed  $\theta$  describes the aleatoric uncertainty
- Likelihood:  $p(y|\theta)$  as a function of  $\theta$  given fixed  $y$  provides information about epistemic uncertainty, but is not a probability distribution

## Model vs. likelihood

- Bayes rule  $p(\theta|y) \propto p(y|\theta)p(\theta)$
- Model:  $p(\mathbf{y}|\theta)$  as a function of  $\mathbf{y}$  given fixed  $\theta$  describes the aleatoric uncertainty
- Likelihood:  $p(y|\theta)$  as a function of  $\theta$  given fixed  $y$  provides information about epistemic uncertainty, but is not a probability distribution
- Bayes rule combines the likelihood with prior uncertainty  $p(\theta)$  and transforms them to updated posterior uncertainty

## The art of probabilistic modeling

- The art of probabilistic modeling is to describe in a mathematical form (model and prior distributions) what we already know and what we don't know

## The art of probabilistic modeling

- The art of probabilistic modeling is to describe in a mathematical form (model and prior distributions) what we already know and what we don't know
- “Easy” part is to use Bayes rule to update the uncertainties
  - computational challenges

# The art of probabilistic modeling

- The art of probabilistic modeling is to describe in a mathematical form (model and prior distributions) what we already know and what we don't know
- “Easy” part is to use Bayes rule to update the uncertainties
  - computational challenges
- Other parts of the art of probabilistic modeling are, for example,
  - model checking: is data in conflict with our prior knowledge?
  - presentation: presenting the model and the results to the application experts

## Modeling nature

- Drop a ball from different heights and measure time

## Modeling nature

- Drop a ball from different heights and measure time
  - Newton
  - air resistance, air pressure, shape and surface structure of the ball
  - relativity

## Modeling nature

- Drop a ball from different heights and measure time
  - Newton
  - air resistance, air pressure, shape and surface structure of the ball
  - relativity
- Taking into account the accuracy of the measurements, how accurate model is needed?

## Modeling nature

- Drop a ball from different heights and measure time
  - Newton
  - air resistance, air pressure, shape and surface structure of the ball
  - relativity
- Taking into account the accuracy of the measurements, how accurate model is needed?
  - often simple models are adequate and useful
  - *All models are wrong, but some of them are useful*, George P. Box

## Reminder: Uncertainty and probabilistic modeling

- Two types of uncertainty: aleatoric and epistemic
- Representing uncertainty with probabilities
- Updating uncertainty

# Chapter 1

## Reading instructions

- 1.1-1.3 important terms
- 1.4 a useful example
- 1.5 foundations
- 1.6 & 1.7 examples (can be skipped, but may be useful to read)
- 1.8 & 1.9 background material, good to read before doing the exercises
- 1.10 a point of view for using Bayesian inference

## Part of the assignment 1

Refresh your memory on these concepts!

- probability
- probability density
- probability mass
- probability density function (pdf)
- probability mass function (pmf)
- probability distribution
- discrete probability distribution
- continuous probability distribution
- cumulative distribution function (cdf)
- likelihood
- Bayes rule

# Ambiguous notation in statistics

Find this in the Chapter 1 reading instructions on the course web page!

$\ln p(y | \theta)$

- $y$  can be variable or value
  - we could clarify by using  $p(Y | \theta)$  or  $p(y | \theta)$
- $\theta$  can be variable or value
  - we could clarify by using  $p(y | \Theta)$  or  $p(y | \theta)$
- $p$  can be a discrete or continuous function of  $y$  or  $\theta$ 
  - we could clarify by using  $P_Y$ ,  $P_\Theta$ ,  $p_Y$  or  $p_\Theta$
- $P_Y(Y | \Theta = \theta)$  is a probability mass function, sampling distribution, observation model
- $P(Y = y | \Theta = \theta)$  is a probability
- $P_\Theta(Y = y | \Theta)$  is a likelihood function (can be discrete or continuous)
- $p_Y(Y | \Theta = \theta)$  is a probability density function, sampling distribution, observation model
- $p(Y = y | \Theta = \theta)$  is a density
- $p_\Theta(Y = y | \Theta)$  is a likelihood function (can be discrete or continuous)
- $y$  and  $\theta$  can also be mix of continuous and discrete
- due to the sloppiness sometimes likelihood is used to refer  $P_{Y,\theta}(Y | \Theta)$ ,  $p_{Y,\theta}(Y | \Theta)$

# Questions

- Pick a number between 1–5

## Questions

- Pick a number between 1–5
  - raise as many fingers

## Questions

- Pick a number between 1–5
  - raise as many fingers
  - is the number of fingers raised random (by you or by others)?

## Questions

- Pick a number between 1–5
  - raise as many fingers
  - is the number of fingers raised random (by you or by others)?
- If we build a robot with very fast vision which can observe the rotating coin accurately, is the throw random for the robot?

## Questions

- Pick a number between 1–5
  - raise as many fingers
  - is the number of fingers raised random (by you or by others)?
- If we build a robot with very fast vision which can observe the rotating coin accurately, is the throw random for the robot?
- What is your own example with both aleatoric and epistemic uncertainty?