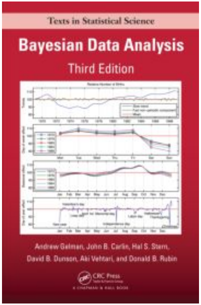
Bayesian data analysis (Aalto fall 2022)

- Book: Gelman, Carlin, Stern, Dunson, Vehtari & Rubin: Bayesian Data Analysis, Third Edition. (online pdf available)
- The course website has more detailed information than these slides
  https://avehtari.github.io/BDA_course_Aalto/Aalto2022.html
- Timetable: see the course website
- TAs: Anna Riha, Elena Shaw, Kunal Ghosh, Andrew Johnson, Noa Kallioinen, David Kohns, Leevi Lindgren, Yann McLatchie, Teemu Sailynoja, Niko Siccha



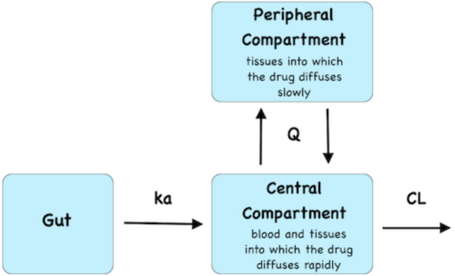
Uncertainty and decision making

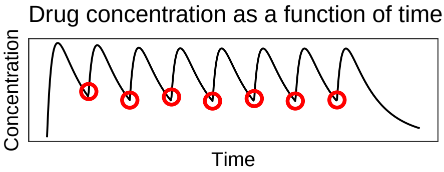
- Predicting concrete quality

$\rightarrow$

Uncertainty and decision making[1]

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

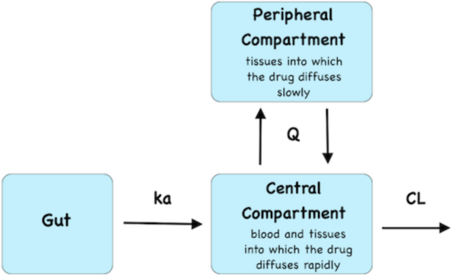




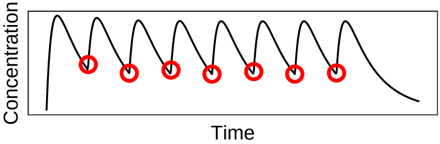
[1] with E. Siivola, Aalto and S. Weber, Novartis Pharma

Uncertainty and decision making[1]

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



Drug concentration as a function of time

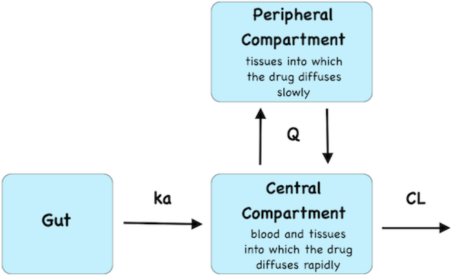


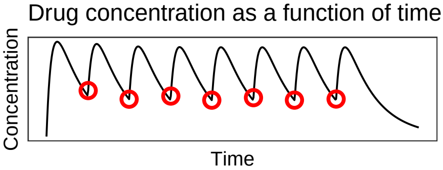
- Model fitted with 500 adults, extrapolation to children?

[1] with E. Siivola, Aalto and S. Weber, Novartis Pharma

Uncertainty and decision making[1]

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

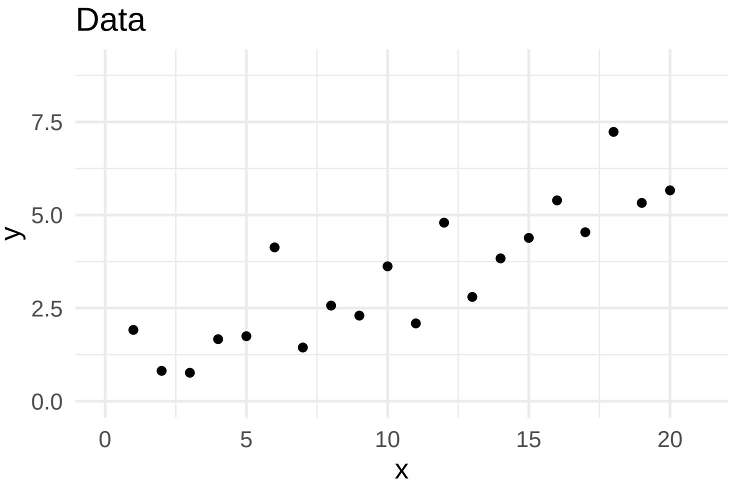




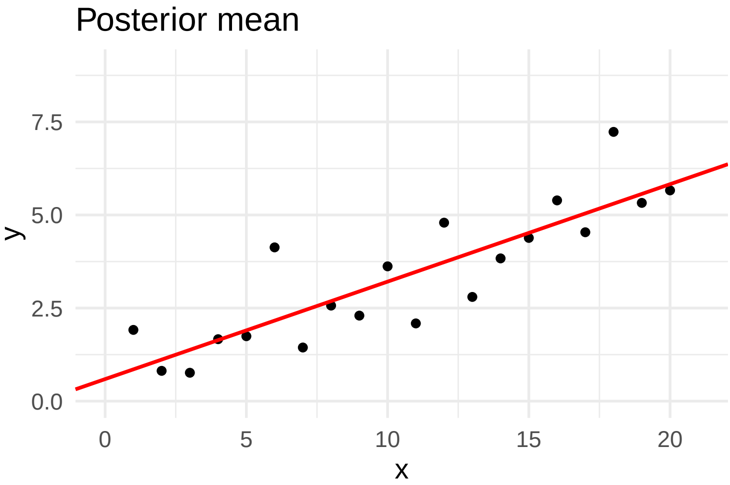
- Model fitted with 500 adults, extrapolation to children?
- Maturation effect, 17 observations from children

[1] with E. Siivola, Aalto and S. Weber, Novartis Pharma

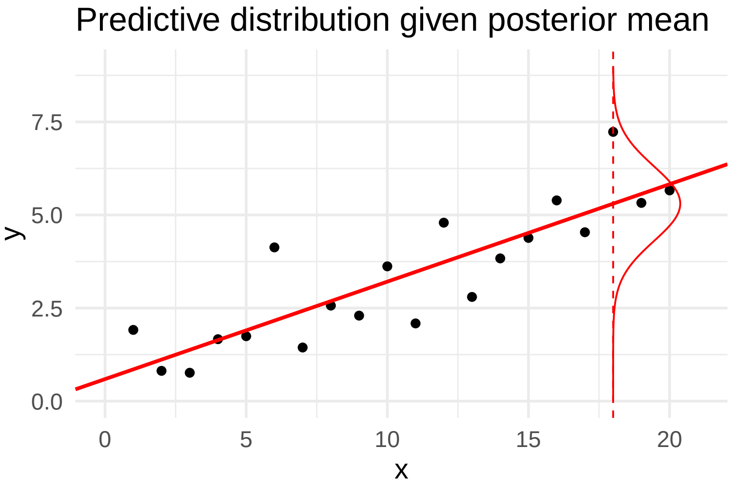
Uncertainty in modeling



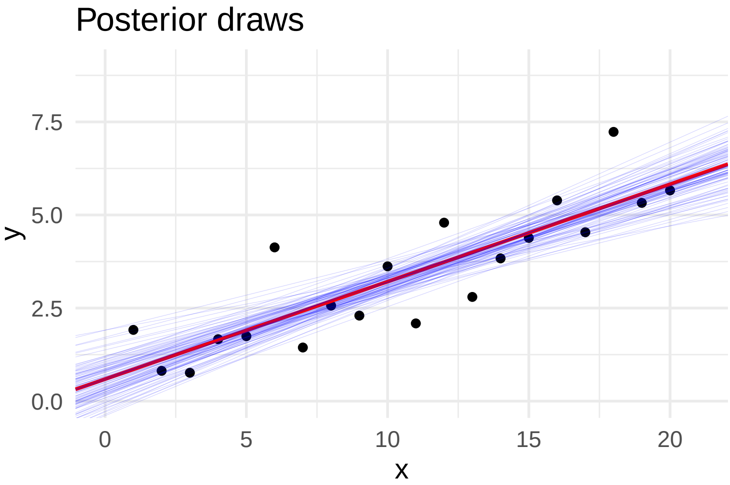
Uncertainty in modeling



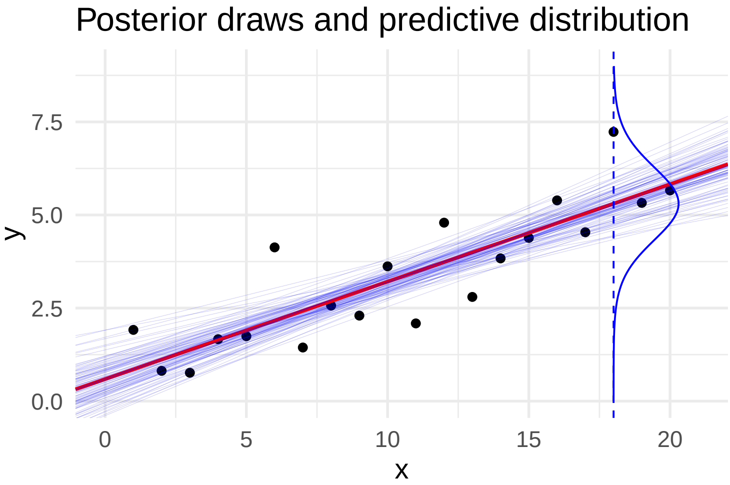
Uncertainty in modeling



Uncertainty in modeling



Uncertainty in modeling



Bayesian probability theory

expert information

Bayesian probability theory

expert information

$\searrow$

mathematical model

+

uncertainty with probabilities

Bayesian probability theory

expert information $\searrow$ mathematical model + uncertainty with probabilities

data

Bayesian probability theory

expert information $\searrow$ mathematical model + uncertainty with probabilities

+

data $\swarrow$ 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

expert information $\searrow$ mathematical model + uncertainty with probabilities

+

data $\swarrow$ Bayesian probability theory

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

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

$\downarrow$

updated uncertainty

Bayesian probability theory

expert information $\searrow$ mathematical model + uncertainty with probabilities

+

data $\swarrow$ Bayesian probability theory

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

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

$\downarrow$

updated uncertainty

+

understandable trusted models

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

Probabilistic programming and Stan

Stan is a probabilistic programming framework and ecosystem
40+ developers, 100+ contributors, 100K+ users

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

$$\mathrm{E}_{\theta|y}\left[g(\theta)\right] = \int p(\theta|y) g(\theta) d\theta$$

  - 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

$$\mathrm{E}_{\theta|y}\left[g(\theta)\right] = \int p(\theta|y) g(\theta) d\theta$$

  - 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

$\rightarrow$ better science

$\rightarrow$ 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

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

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 $\frac{\#\text{red}}{\#\text{red}+\#\text{yellow}} = \theta$

Updating uncertainty

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

- $p(y = \#\text{red}|\theta) = \theta$ aleatoric uncertainty

Updating uncertainty

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

Updating uncertainty

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

Updating uncertainty

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

- Likelihood: $p(y|\boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$ 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 θ describes the aleatoric uncertainty

- Likelihood: $p(y|\boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$ 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

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?
- Is the quantum uncertainty aleatoric or epistemic?

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?
- Is the quantum uncertainty aleatoric or epistemic?
- What is your own example with both aleatoric and epistemic uncertainty?

Pre-requisites

- 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

Ambiguous notation in statistics

$\ln p(y|\theta)$

Ambiguous notation in statistics

In $p(y|\theta)$

- y can be variable or value

  we could clarify by using $p(Y|\theta)$ or $p(y|\theta)$

Ambiguous notation in statistics

In $p(y|\theta)$

- y can be variable or value

  we could clarify by using $p(Y|\theta)$ or $p(y|\theta)$

- θ can be variable or value

  we could clarify by using $p(y|\Theta)$ or $p(y|\theta)$

Ambiguous notation in statistics

In $p(y|\theta)$

- y can be variable or value

  we could clarify by using $p(Y|\theta)$ or $p(y|\theta)$
- θ 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 θ

  we could clarify by using P_Y, P_Θ, p_Y or p_Θ

Ambiguous notation in statistics

In $p(y|\theta)$

- y can be variable or value

  we could clarify by using $p(Y|\theta)$ or $p(y|\theta)$
- θ 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 θ

  we could clarify by using P_Y, P_Θ, p_Y or p_Θ
- $P_Y(Y|\Theta = \theta)$ is a probability mass function, sampling distribution, observation model

Ambiguous notation in statistics

In $p(y|\theta)$

- y can be variable or value

  we could clarify by using $p(Y|\theta)$ or $p(y|\theta)$
- θ 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 θ

  we could clarify by using P_Y, P_Θ, p_Y or p_Θ
- $P_Y(Y|\Theta = \theta)$ is a probability mass function, sampling distribution, observation model
- $P(Y = y|\Theta = \theta)$ is a probability

Ambiguous notation in statistics

In $p(y|\theta)$

- y can be variable or value

  we could clarify by using $p(Y|\theta)$ or $p(y|\theta)$
- θ 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 θ

  we could clarify by using P_Y, P_Θ, p_Y or p_Θ
- $P_Y(Y|\Theta = \theta)$ 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)

Ambiguous notation in statistics

In $p(y|\theta)$

- y can be variable or value

  we could clarify by using $p(Y|\theta)$ or $p(y|\theta)$
- θ 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 θ

  we could clarify by using P_Y, P_Θ, p_Y or p_Θ
- $P_Y(Y|\Theta=\theta)$ 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 distrbution, observation model

Ambiguous notation in statistics

In $p(y|\theta)$

- y can be variable or value

  we could clarify by using $p(Y|\theta)$ or $p(y|\theta)$
- θ 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 θ

  we could clarify by using P_Y, P_Θ, p_Y or p_Θ
- $P_Y(Y|\Theta = \theta)$ 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 distrbution, observation model
- $p(Y = y|\Theta = \theta)$ is a density

Ambiguous notation in statistics

In $p(y|\theta)$

- y can be variable or value
  - we could clarify by using $p(Y|\theta)$ or $p(y|\theta)$
- θ 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 θ
  - we could clarify by using P_Y, P_Θ, p_Y or p_Θ
- $P_Y(Y|\Theta = \theta)$ 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 distrbution, observation model
- $p(Y = y|\Theta = \theta)$ is a density
- $p_\Theta(Y = y|\Theta)$ is a likelihood function (can be discrete or continuous)

Ambiguous notation in statistics

In $p(y|\theta)$

- y can be variable or value

  we could clarify by using $p(Y|\theta)$ or $p(y|\theta)$
- θ 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 θ

  we could clarify by using P_Y, P_Θ, p_Y or p_Θ
- $P_Y(Y|\Theta = \theta)$ 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 distrbution, 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 θ can also be mix of continuous and discrete

Ambiguous notation in statistics

In $p(y|\theta)$

- y can be variable or value

  we could clarify by using $p(Y|\theta)$ or $p(y|\theta)$
- θ 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 θ

  we could clarify by using P_Y, P_Θ, p_Y or p_Θ
- $P_Y(Y|\Theta = \theta)$ 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 distrbution, 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 θ can also be mix of continuous and discrete
- Due to the sloppines sometimes likelihood is used to refer $P_{Y,\theta}(Y|\Theta)$, $p_{Y,\theta}(Y|\Theta)$

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