Bayesian Data Analysis

Feng Li
feng.li@cufe.edu.cn

School of Statistics and Mathematics
Central University of Finance and Economics

Revised on July 23, 2017[1]

[1] I am grateful to Professor Aki Vehtari from Aalto University (https://users.aalto.fi/~ave/) for sharing his course materials.

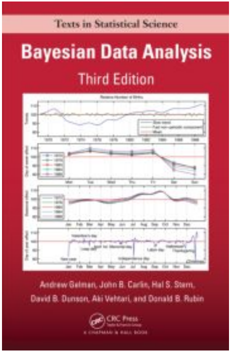
General information

- **Lecturer**
  - Feng Li, `feng.li@cufe.edu.cn`
- **Language**: The course is taught in Chinese. The course materials are in English.
- **Reception hours**: Questions concerned with this course can be asked after each lecture or via email.
- **Lecture notes and course materials**
  - **Bayesian Data Analysis** by Andrew Gelman *et. al* (2013), Third Edition. Chapman and Hall/CRC.
  - **An Introduction to R** `http://cran.r-project.org/doc/manuals/R-intro.pdf`
  - **Course materials and computer code** `http://feng.li/teaching/bda2017summer/`
- **Working load**: Depending on your own situation and you ambition, you decide how much time you want to input.

Don’t worry and it’s fun!

About

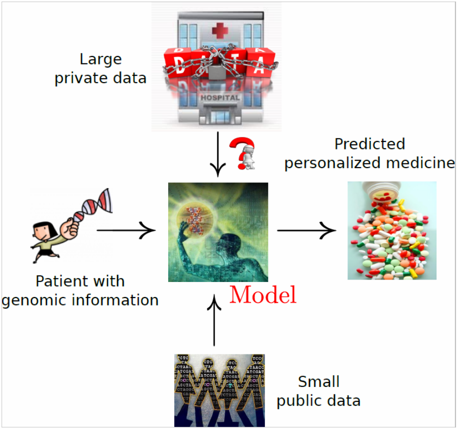
- Bayesian data analysis
  - Bayesian theory, models, and inference
- Probabilistic machine learning
  - Gaussian processes



For updated slides and code, check out:
https://feng.li/teaching/bda2017summer/

Current applications

- Digital health, personalized medicine
  - survival analysis, disease risk prediction
  - biomarkers, genetic data



Graduated students work, e.g., at

- Brain signal analysis, Harvard / Aalto
- Epidemiology, National institue for health and welfare, Finnish Institue of Molecular Medicine
- Fisheries and environmental management analysis, University of Helsinki
- Forecast analyst (consumer goods trade), SOK
- Monitoring and imaging systems for industrial processes, Rocsole
- Portfolio analysis, Investment Research Finland
- Engineer, ZenRobotics
- Research scientist, Virtual Air Guitar Company
- Data scientist, TeliaSonera, ...

Other quasi-random examples

- Facebook, news feed ordering and ad selection
- Google, A/B testing
- Nate Silver, USA election polls
- Reaktor, kannattaakokauppa.fi
- F1 teams, aerodynamics
- ...

Some other application areas

- Archeology
- Astronomy
- Bio-sciences
- Cognitive science
- Data mining
- Decision analysis
- Economy
- Epidemiology
- Genetics
- Image analysis
- Law
- Medicine
- Meteorology
- Physics
- Process modeling
- Reliability analysis
- Signal analysis
- Social sciences

Some other application areas

- Archeology
- Astronomy
- Bio-sciences
- Cognitive science
- Data mining
- Decision analysis
- Economy
- Epidemiology
- Genetics
- Image analysis
- Law
- Medicine
- Meteorology
- Physics
- Process modeling
- Reliability analysis
- Signal analysis
- Social sciences

- *Anything related to real world,*
  where inference is made based on observations

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: $p(y|\theta)$ as a function of y given fixed θ describes the aleatoric uncertainty
- Likelihood: $p(y|\theta)$ as a function of θ given fixed y provides information about epistemic uncertainty

Model vs. likelihood

- Bayes rule $p(\theta|y) \propto p(y|\theta)p(\theta)$
- Model: $p(y|\theta)$ as a function of y given fixed θ describes the aleatoric uncertainty
- Likelihood: $p(y|\theta)$ as a function of θ given fixed y provides information about epistemic uncertainty
- Bayes rule combines the likelihood with prior uncertainty $p(\theta)$ and transforms them to updated posterior uncertainty

GIST cancer example

- Gastrointestinal stromal tumor (GIST)
- 2560 patients followed after surgery (+ 920 validation set)
- Various predictors available
- Probability of recurrence in five years after surgery?

Reminder: Other quasi-random examples

- Facebook, news feed ordering and ad selection
- Google, A/B testing
- Nate Silver, USA election polls
- Reaktor, kannattaakokauppa.fi
- F1 teams, aerodynamics
- ...

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

Rest of the course

- Basic models which can be used as building blocks
- Basic computation
- Typical simple scientific data analysis cases
  - e.g. comparison of treatments
- Presentation of the results

Reminder: Uncertainty and probabilistic modeling

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

Reminder: Uncertainty and probabilistic modeling

- Two types of uncertainty: aleatoric and epistemic
- Representing uncertainty with probabilities
- Updating uncertainty
- Additional reading material: Dicing with unknown by Tony O'Hagan

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?

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

Bayesian Analysis

- Based on Bayesian probability theory
  - uncertainty is presented with probabilities
  - probabilities are updated based on new information
  - *...common sense reduced to calculation*, Laplace 1819

Bayesian Analysis

- Based on Bayesian probability theory
  - uncertainty is presented with probabilities
  - probabilities are updated based on new information
  - *...common sense reduced to calculation*, Laplace 1819
- Thomas Bayes (170?–1761)
  - English nonconformist, Presbyterian minister, mathematician
  - Richard Price published Bayes' paper on conditional probabilities in 1763 after Bayes had died
  - considered the problem of *inverse probability*
    - significant part of the Bayesian theory

Bayesian Analysis

- Based on Bayesian probability theory
  - uncertainty is presented with probabilities
  - probabilities are updated based on new information
  - *...common sense reduced to calculation*, Laplace 1819
- Thomas Bayes (170?–1761)
  - English nonconformist, Presbyterian minister, mathematician
  - Richard Price published Bayes' paper on conditional probabilities in 1763 after Bayes had died
  - 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

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
- Frequentist 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
- Frequentist 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
  - after this Bayesians started to use term "frequentist"

- Background (Ch 1)
- Single-parameter models (Ch 2)
- Multiparameter models (Ch 3)
- Computational methods (Ch 10)
- Markov chain Monte Carlo (Ch 11-12)
- R
- Hierarchical models (Ch 5)
- Model checking (Ch 6)
- Evaluating and comparing models (Ch 7)
- Decision analysis (Ch 9)
- Large sample properties and Laplace approximation (Ch 4)
- R

- 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

Other courses and books

- Gaussian processes
- sparse data
- time series
- deep learning
- more theoretical books
  - Bayesian Theory by Bernardo & Smith
  - Bayesian inference by O'Hagan & Forster