Biostatistics 682: Applied Bayesian Inference
Lecture 1: Introduction

Jian Kang

Department of Biostatistics
University of Michigan, Ann Arbor

Introduction to the course

Instructor

- Jian Kang, PhD, Associate Professor of Biostatistics
- Research Interests: Bayesian modeling and theory; Imaging statistics; Large-scale complex data analysis.
- Email: jiankang@umich.edu
- Office: M4055, SPH II (1415 Washington Heights)
- Office Hour:
  - Tuesday, 3:30pm – 4:30pm
  - Thursday (**expect September 7th**), 1:00pm – 2:00pm
- Class Time: Tuesday & Thursday, 2:00PM - 3:30PM
- Class Room: G322 DENT

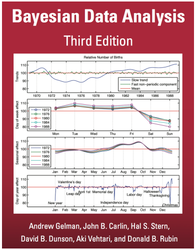
Objectives

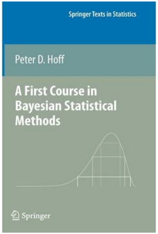
- Basic theory and methods for Bayesian data analysis.
- Statistical programming language for Bayesian computation
  - R, `https://cran.r-project.org`
  - JAGS (Just Another Gibbs Sampler), `http://mcmc-jags.sourceforge.net`
  - Stan, `http://mc-stan.org`

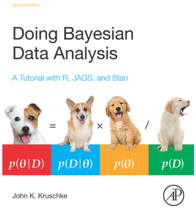
Prerequisites

- Linear algebra (matrix theory)
- Multivariate calculus
- Basic statistics including probability distribution
- Linear model and hypothesis testing.
- Biostatistics 601, Biostatistics 650, Biostatistics 651, Biostatistics 653 or equivalent is required prior to or in parallel to taking Biostatistics 682.
- Previous experience in programming in R

Reference Books (Not Required)







- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. *Bayesian data analysis (Third Edition)*, CRC press, 2014
- Hoff, Peter D. *A first course in Bayesian statistical methods.* Springer Science & Business Media, 2009.
- Kruschke, John. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.* Academic Press, 2014.

Grading

- Homework: 50%
  - Homework assignments will be given out at every other week
  - Due in two weeks.
  - Plagiarism will NOT be tolerated and late homework will NOT be graded
- Midterm Exam (October 31st): 25%
  - In-class, closed-book
  - Three questions for Bayesian theory and methods
  - Two questions for Bayesian data analysis
- Course project: 25%
  - Two or three students form a team
  - To address scientific questions, develop Bayesian methods for real datasets
  - Write a report to describe the methods and summarize the analysis results.
  - Submit the software package or source code for the analysis
- Letter grades: A+ = [95,100], A = [90, 94], A- = [85, 90], B+ = [80, 84], B = [70, 79], B- = [60, 69]

Important Dates

- September 5th: Class begins
- September 12th: Homework 1 distributed
- September 25th: Drop / Add deadline for full term classes
- September 26th: Homework 1 due & Homework 2 distributed
- October 10th: Homework 2 due & Homework 3 distributed
- October 17th: No class for fall break
- October 24th: Homework 3 due & Homework 4 distributed
- October 31st: Midterm exam
- November 2nd: Project team member and topic due
- November 7th: Homework 4 due & Homework 5 distributed
- November 21th: Homework 5 due
- November 23th: No class for Thanksgiving
- December 12th: Class ends (last class)
- December 15th: Course project due

Topics

- Part I: Fundamentals of Bayesian Inference and data analysis
    - Basic concepts of probability and Bayesian statistics
    - Single-parameter models
    - The multivariate normal model
    - The multinomial model
    - Group comparisons and hierarchical modeling
    - Linear regression model
    - Generalized linear model
    - Optional: Generalized linear mixed-effects model, Gaussian process models and Dirichlet process models
    - Model checking
    - Model comparisons
- Part II: Bayesian Computations
    - Approximating posterior distributions
    - Introduction to Monte Carlo methods
    - Importance sampling
    - Gibbs sampler, Metropolis-Hastings algorithms and Introduction to JACS
    - Hamiltonian Monte Carlo and Introduction to Stan

Core Competencies

After taking this class, students are expected to be able to

- Understand basic theory for Bayesian statistics
- Use existing methods or develop new methods for Bayesian data analysis
- Interpret the Bayesian data analysis results to address the scientific questions
- Implement Bayesian computational algorithms using a statistical package
- Perform model checking and model diagnostics

Standards of Academic Act

- From the School of Public Health's Student Code of Conduct,

  "Student academic misconduct includes behavior involving plagiarism, cheating, fabrication, falsification of records or official documents, intentional misuse of equipment or materials, and aiding and abetting the perpetration of such acts. The preparation of reports, papers, and examinations, assigned on an individual basis, must represent each student's own effort. Reference sources should be indicated clearly. The use of assistance from other students or aids of any kind during a written examination, except when the use of books or notes has been approved by an instructor, is a violation of the standard of academic conduct."

- In the context of this course, any work the student hand-in should be his/her own and any material that is a transcript (or interpreted transcript) of work by others must be clearly labeled as such.

Basic Concepts in Bayesian Inference

How to interpret probability?

Long run frequencies

OR

Subjective degree of belief

An Example – Why the sidewalk is wet?

- Suppose we notice that the sidewalk is wet and wonder why.
- What are the possibilities?
  - recent rain
  - recent garden irrigation
  - a newly erupted underground spring,
  - a broken sewage pipe
  - a passerby who spilled a drink
  - ......
- What is the probability of each event?
- It depends on what information being given
  - No additional information: only know the sidewalk is wet.
  - Information A: Trees, cars and everything else outside are also wet.
  - Information B: The wetness is only localized to a small area and there is an empty drink cup a few feet away?

What is the Bayesian Inference?

- Reallocation of credibility across possibilities

$$\text{Prior Knowledge} \xrightarrow[]{\substack{\text{Data}\\ \text{Information}}} \text{Posterior Inference}$$

- Bayesian versus frequentist
  - Difference in the goal:
    - Bayesian: Quantify and analyze the subjective degree of belief
    - Frequentist: Create procedure that have frequency guarantees
  - Difference in the assumption of parameters
    - Frequentist: Data are random while parameters are fixed but unknown
    - Bayesian: Both data and parameters are random
  - Which school is correct?
    - Can Bayesian inference enjoy the good frequentist properties?
    - Can frequentist approach achieve Bayesian goal?

Thomas Bayes (1701–1761)

- Bayes's Theorem: Let A and B be two random events with $\Pr(B) > 0$. Then

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)}.$$

  How to prove it?
- Why Bayes Theorem is useful for Bayesian?
- Does a frequentist use Bayes Theorem or not?

An Example – Screening Tests

- Accuracy of a diagnostic tests is usually measured by
  - Sensitivity: probability that a diseased patient tests positive.

$$\Pr(\text{test}+ \mid \text{disease}+)$$

  - Specificity: probability that a healthy person tests negative.

$$\Pr(\text{test}- \mid \text{disease}-)$$

  - Positive predictive value (PPV): probability of a patient has disease given the test is positive.

$$\Pr(\text{disease}+ \mid \text{test}+)$$

  - Negative predictive value (NPV): probability of a person is healthy given the test is negative.

$$\Pr(\text{disease}- \mid \text{test}-)$$

An Example – Screening Tests (Continued)

- The partial accuracy information of a screening test for breast cancer is given

	Diagnostic Test	
Breast Cancer	Positive	Negative
Yes	0.82	0.18
No	0.01	0.99

- We know that the prevalence of breast cancer is 0.1, i.e. $\Pr(\text{disease}+) = 0.1$.
- Can we compute the PPV and NPV based on the Bayes Theorem?

$$
\begin{aligned}
PPV &= \Pr(\text{disease}+ \mid \text{test}+) \\
&= \frac{\Pr(\text{test}+ \mid \text{disease}+)\Pr(\text{disease}+)}{\Pr(\text{test}+)} \\
&= \frac{\Pr(\text{test}+ \mid \text{disease}+)\Pr(\text{disease}+)}{\Pr(\text{test}+ \mid \text{disease}+)\Pr(\text{disease}+) + \Pr(\text{test}+ \mid \text{disease}-)\Pr(\text{disease}-)} \\
&= \frac{0.82(0.1)}{0.82(0.1) + 0.18(0.9)} \\
&= 0.63
\end{aligned}
$$

Similar results hold for the NPV.

Bayesian data analysis

- Making Bayesian inferences from data
- Probability models are used for all quantities – both observed and unobserved (those we wish to learn something about).
- Three basic steps to an applied Bayesian analysis
  - Set up a full probability model
    - A joint probability distribution for all observable and unobservable quantities in a problem.
    - The model should be consistent with knowledge about the underlying scientific problem and the data collection process
  - Condition on the observed data
    - Involves calculating and interpreting the appropriate posterior distribution
    - The conditional probability distribution of the unobserved quantities of interest given the observed data and perhaps nuisance parameters
  - Evaluate the fit of the model
    - Does the model fit the data?
    - Are substantive conclusions reasonable?
    - How sensitive are results to model assumptions?

Notations

- θ–unobservable (vector) quantities or population parameters
- y–observed (vector) data
- $\tilde{y}$– unknown, potentially observable, data (e.g. the outcome for the next patient enrolled in a trial)
- Vectors are assumed to be column vectors
- Greek letters will be used for parameters
- Lower case Roman letters for observed data (scalar or vector)
- Upper case Roman letters for observed matrices
- $\pi(y \mid \theta)$ denotes the sampling distribution (as a function of θ) as opposed to the likelihood notation used in the frequentist literature ($f(\theta \mid y)$)
- $\pi(\cdot \mid \cdot)$ in general represents a conditional density or probability mass function
- $\pi(\cdot)$ is a marginal density or probability mass function. (Sometimes we use $\Pr(\cdot)$ for the probability mass function)
- $[\theta \mid \phi] \sim F(\phi)$ means that θ has distribution of F with parameter set ϕ. For convenience, the conditional notation is often suppressed: $\theta \sim F(\phi)$.

Bayes Mechanics

- To make probabilistic inference on a parameter (vector) θ given data y,
- Specify the likelihood or sampling distribution $\pi(y \mid \theta)$.
- Specify the prior distribution of θ: $\pi(\theta)$.
- Obtain the joint distribution of data and parameters

$$\pi(y, \theta) = \pi(y \mid \theta)\pi(\theta).$$

- By Bayes' theorem, the posterior distribution of θ:

$$\pi(\theta \mid y) = \frac{\pi(y \mid \theta)\pi(\theta)}{\pi(y)},$$

  where $\pi(y) = \int \pi(y, \theta) d\theta = \int \pi(y \mid \theta)\pi(\theta) d\theta$.
- When θ is discrete and takes values $\{\theta_i\}$, $\pi(y) = \sum_i \pi(y \mid \theta_i)\pi(\theta_i)$.
- When $\pi(y)$ is analytically intractable, we usually works on unnormalized densities:

$$\pi(\theta \mid y) \propto \pi(y \mid \theta)\pi(\theta).$$

Statistical inference under Bayesian framework

- Point estimator: summary statistics from the posterior distribution, e.g. the posterior mean
$$\bar{\theta} = E(\theta \mid y) = \int \theta \pi(\theta \mid y) d\theta.$$
- Interval/region estimation: the posterior $1-\alpha$ credible interval/region C:
$$\Pr\{\theta \in C \mid y\} = \int_C \pi(\theta \mid y) d\theta = 1 - \alpha.$$
When θ is a one-dimensional parameter, $C = (l, u)$.
- Remark:
    - Although the Bayesian credible interval is also a function of data, the interpretation is quite different from that of the frequentist confidence interval
    - Typically, the $1-\alpha$ Bayesian interval has coverage lower than $1-\alpha$. What does this mean? Who cares?

Bayesian Hypothesis Testing

- Suppose that $\theta \in \mathbb{R}$ and we want to test

$$H_0 : \theta = \theta_0, \quad \text{and} \quad H_1 : \theta \neq \theta_0.$$

- If we really believes that there is a positive prior probability that H_0 is true when we can use a prior of the form

$$\pi_0 \delta_{\theta_0} + (1 - \pi_0) g(\theta),$$

where $\pi_0 (0 < \pi_0 < 1)$ is the prior probability that H_0 is true and g is a smooth prior density over θ which represents our prior beliefs about θ when H_0 is false. It follows from Bayes' theorem that

$$\text{Pr}(\theta = \theta_0 \mid y) = \frac{\pi_0 \pi(y \mid \theta_0)}{\pi_0 \pi(y \mid \theta_0) + (1 - \pi_0) \int \pi(y \mid \theta) g(\theta) d\theta}.$$

- Note: The interpretation is quite different from the p-value. Bayes factor is a more commonly used criterion.

Prediction

- Predictions are more straightforward in the Bayesian framework.
- Suppose we would like to make a prediction on a new patient's outcome, $\tilde{y}$.
- Before any data, y, has been collected, the prior predictive distribution of $\tilde{y}$ is

$$\pi(\tilde{y}) = \int \pi(\tilde{y}, \theta) d\theta = \int \pi(\tilde{y} \mid \theta)\pi(\theta) d\theta.$$

- After data, y, has been collected, we can predict $\tilde{y}$ given the data y using the posterior predictive distribution

$$\pi(\tilde{y} \mid y) = \int \pi(\tilde{y} \mid \theta)\pi(\theta \mid y) d\theta.$$