STAT7630: Bayesian Statistics
Lecture Slides # 1

Preliminaries & Chapter 1 - The Big (Bayesian) Picture

Elvan Ceyhan
Department of Mathematics & Statistics
Auburn University

Fall 2024,
Updated: August, 2024

Outline

Needed Background for this course

What is needed or expected:

- The expected level of statistics is equivalent to that obtained by a graduate student in their first year of study of the theory of statistics and probability.
- An understanding of maximum likelihood (ML) methods, simple inference (HT) & estimation, and linear models is most important. Many of these topics are reviewed in the Refresher Notes on Prob & Stat (posted under Canvas).
- Familiarity with multivariate methods and some non-linear models.
- The level of mathematics needed in this course does not extend much beyond Taylor series and linear algebra (with basic understanding of matrix algebra).
- A working knowledge of software package like `R`, Python, or SAS.

Goals of this course

Goals: At the end of this course, students will be able to

1. understand the theory and implementation of key methods in Bayesian statistics
2. apply Bayesian methods to solve problems
3. develop new computational methods in Bayesian statistics

Textbook(s)

The slides will be (mostly) developed from the book

Bayes Rules! An Introduction to Applied Bayesian Modeling,
by Alicia A. Johnson, Miles Q. Ott, & Mine Dogucu (CRC Press).
Book website: https://www.bayesrulesbook.com

- Suggested but not required:
  - Bayesian Computation with R, 2nd ed. (Springer), by Albert, J.
  - Introducing Monte Carlo Methods with R (Springer), by Robert, C.P. and Casella, G.
  - Bayesian Data Analysis, Second Edition, by A. Gelman, J. B. Carlin, H. S. Stern and D. B. Rubin (Chapman & Hall) [Can be bought on Amazon.com]
  - Introduction to Statistical Thought, by Michael Lavine, available free as a pdf download at http://www.math.umass.edu/ lavine/Book/book.html

Topics

- Emphasis will be on practical use of Bayesian inference in a variety of problems, esp. on physical sciences but should be useful generally
- Will not emphasize strictly mathematical results; the mathematics is not difficult or advanced, but a different way of thinking is required from classical statistical thinking
- Topics:
  - Review of probability calculus, interpretations, coherence, Bayes's theorem. Joint, conditional and marginal distributions. Independence. Prior distribution, likelihood, posterior distribution. Bayesian estimation and inference on discrete state spaces. Likelihoods, odds and Bayes factors. Simple and composite alternatives
  - Markov chain Monte Carlo (MCMC). Gibbs and Metropolis-Hastings samplers. Metropolis-within- Gibbs. Computer tools, e.g., R, Stan, JAGS, BUGS.

Topics

- Topics (Cont'd):
  - Bayesian point and interval parameter estimation. Bayesian credible intervals (e.g. HDP intervals). Bayesian inference on Gaussian, Poisson, Cauchy, and arbitrary distributions. MLE as an approximation to Bayesian inference. Laplace approximation. Linear and nonlinear models. Selection models. Hierarchical models.
  - Prior selection. Subjective and objective priors. Priors as encoding knowledge. Sensitivity to prior and robustness. Priors for hierarchical models.
  - Bayesian hypothesis testing. Comparison with frequentist hypothesis testing. Model selection and model averaging. Reversible jump MCMC. Approximations, e.g., AIC, BIC. Likelihood principle. Bayesian Ockham's Razor. Bayesian point and interval parameter estimation.

Software

- R
- We will be using the statistical computer language R for most of the examples that will be computed in class. You are encouraged to go to the R website:
- http://www.R-project.org/
- Download a free copy of R, and the R documentation (RStudio is strongly encouraged too), to your computer. Play around with it to get familiar with it. Consult the "tutorial" in the back of the "Introduction to R" at http://cran.r-project.org/doc/manuals/R-intro.pdf for some ideas to get started.
- Download the bayesrules R package that goes with the book from the CRAN website and add it to your R program.

Outline

Introduction

- In real life, we continuously update our knowledge as we gather more data.
- Bayesian thinking reflects this natural process of updating beliefs with new information.
- Bayesian inference utilizes **Bayes' Rule** to combine prior knowledge with sample data, allowing us to draw conclusions about parameters of interest. That is, it uses **Bayes' Law** (**Bayes' Theorem**) to combine prior information and sample data to make conclusions about a parameter of interest.
- Unlike (classical) frequentist methods, Bayesian inference also models the probability distribution of the parameters.

Bayesian Knowledge-Building Process

- Acknowledge/elicit prior information.
- Collect data.
- Update your knowledge based on new data.
- Repeat the process as more data is gathered.

Bayesian Knowledge-Building Diagram

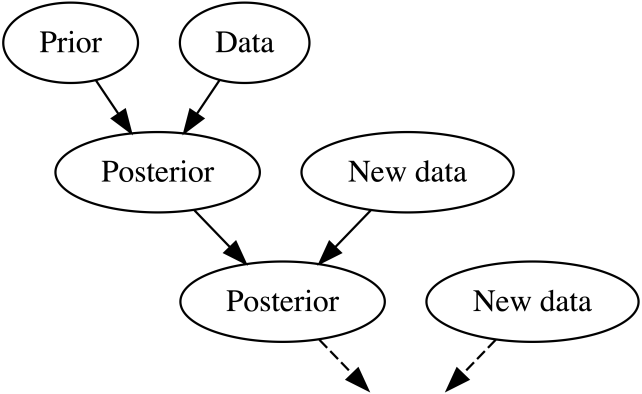

Figure 1: A Bayesian knowledge-building diagram.

Thinking Like a Bayesian

- Bayesian and frequentist analyses share the common goal of **learning from data** (or **making sense of data**).
- The distinction between Bayesian and frequentist approaches lies in their interpretation of probability and the role of prior knowledge.

Quiz Yourself

- How do you interpret the probability of flipping Heads as 0.5?
- How do you interpret a 0.9 probability that a candidate will win an election?
- Assess your confidence in two different claims based on the evidence presented.
- Which question would you ask your doctor if you tested positive for a rare disease?

Interpreting Probability

- **Bayesian approach:** Probability represents the relative plausibility of an event.
- **Frequentist approach:** Probability is the long-run relative frequency of an event that can be repeated.
- **Example:** Interpreting the probability of flipping a head in a coin toss.

Challenges in Defining Probability for Certain Events

- Some events, such as election outcomes or weather predictions, are difficult to conceive as repeatable.
- Bayesian analysis provides a flexible framework for these one-time events.

The Bayesian Balancing Act

- Balance between prior knowledge and new data determines the posterior.
- The strength of the data versus the strength of the prior influences the final conclusion.
- As more data is collected, the influence of prior knowledge diminishes.

Asking Questions

- **Bayesian:** What's the chance that the hypothesis is correct given the data?
- **Frequentist:** What's the chance of observing the data given that the hypothesis is incorrect?

Example from Bayes Rules! Textbook

- Consider a scenario where you tested positive for a rare disease. Which question is more informative?
    - **Bayesian question:** What is the chance that I actually have the disease given the positive test result?
    - **Frequentist question:** What is the chance of getting this positive result if I do not have the disease?
- The frequentist "*p*-value" is often misinterpreted and less intuitive.

A Quick History Lesson

- Bayesian statistics originated with *Thomas Bayes* in the 1740s.
- The philosophy gained momentum in the late 20th century due to advances in computing.
- The Bayesian framework is now used globally in various fields.

Resurgence of Bayesian Methods:

- Advances in computing have made complex Bayesian models feasible (and practical).
- Growing recognition of the value of incorporating subjectivity in scientific analysis.

Why Use Bayesian Methods?

- Incorporate prior knowledge about parameters.
- Update our understanding logically after observing new data.
- Make formal probability statements regarding the parameters.
- Assess model assumptions and sensitivity in a straightforward manner.

Why Use Classical Methods?

- Useful when parameters are considered fixed, as in controlled experiments.
- Applicable when there is no prior information available.
- Preferred by those who favor standardized, "cookbook"-type formulaic approaches with minimal input from the researcher.

Historical Preference for Classical Methods:

- Many classical methods were developed for controlled experiments.
- Bayesian methods historically required more complex mathematical formalism.
- Realistic Bayesian analyses were once impractical due to limited computing power.

Bayesian Perspective on Data and Parameters

- Bayesian inference treats unobserved data and unknown parameters similarly by assigning a probability distribution to each.
- Bayesian models specify:
    - A joint density function describing the distribution of the full dataset given the parameter values.
    - A prior distribution reflecting either uncertainty about a fixed parameter or the possible values of a stochastic parameter.
- The prior could reflect:
    - Uncertainty about a parameter that is actually fixed, OR
    - the variety of values that a truly stochastic parameter could take.

Chapter Summary & Outline of the Course

Chapter Summary:

- Bayesian thinking involves updating knowledge by balancing prior information with new data.
- As more data is gathered, different analysts will converge on the same conclusions.

Course Outline:

- Foundations of Bayesian Models
- Posterior Simulation and Analysis, including MCMC methods
- Bayesian Regression and Classification
- Hierarchical Bayesian Models

Different Interpretations of Probability

- **Frequentist** definition of the probability of an event: If we repeat an experiment a very large number of times, what is the proportion of times the event occurs?
  - **Problem:** For some situations, it is impossible to repeat (or even conceive of repeating) the experiment many times.
  - **Example:** The probability that Governor Ivey is re-elected in 2022.
- **Subjective probability:** Based on an individual's degree of belief that an event will occur.
  - **Example:** A bettor is willing to risk up to \$200 betting that Ivey will be re-elected, in order to win \$100. The bettor's subjective probability, $P(\text{Ivey wins})$, is $2/3$.
  - The Bayesian approach can naturally incorporate subjective probabilities about the parameter, where appropriate.

Some Probability Notation - Events

- We denote events by letters such as A, B, C, $\ldots$
- The idea of **conditional probability** is crucial in Bayesian statistics:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- We denote random variables by letters such as X, Y, Z, etc., taking on values denoted by x, y, z, etc.
- The space of all possible values of the rv is called its **support** (or support of its distribution).

Some Probability Notation - RVs

- We will deal with both **discrete** and **continuous** rv's.
- In general, let $f(\cdot)$ denote the probability distribution (p.m.f. or p.d.f.) of a rv.
- Thus, $f(x)$ is the **marginal** distribution of X and $f(x, y)$ is the **joint** distribution of X and Y.
- In general, $f(x, y) = f(x|y)f(y)$ and $f(x) = \int f(x, y)dy = \int f(x|y)f(y)dy$.
- If X, Y independent, then $f(x, y) = f(x)f(y)$.

Some Probability Notation

- The expected value of any function $h(x)$ of X is:

$$\mathbf{E}[h(X)] = \begin{cases} \sum_{x \in \mathcal{X}} h(x)f(x), & \text{if } X \text{ is discrete} \\ \int_{\mathcal{X}} h(x)f(x)dx, & \text{if } X \text{ is continuous.} \end{cases}$$

- Typically, the distribution of X depends on some parameter(s), say θ, so in fact $f(x) = f(x|\theta)$.

Exchangeability

- Bayesians usually assume the data values in the sample are **exchangeable**: that is, reordering the data values does not change the model.
- **Example:** In a social survey, respondents are asked whether they are generally happy. Let

$$Y_i = \begin{cases} 1, & \text{if respondent } i \text{ is happy} \\ 0, & \text{otherwise.} \end{cases}$$

Exchangeability

- Consider the first 5 respondents. What are the probabilities of these 3 outcomes?

$$p(1,0,0,1,1) = ?$$

$$p(0,1,1,0,1) = ?$$

$$p(1,1,0,1,0) = ?$$

- If the data values are exchangeable, these three outcomes will have the same probability.

Exchangeability and iid Property

- **Theorem:** If the data are **independent** and **identically distributed** (iid), i.e., a random sample, from a distribution with parameter θ which itself follows the distribution $p(\theta)$, then the data are exchangeable.
- **Proof:**

Exchangeability and iid

- A famous theorem (de Finetti's Theorem) shows the converse is* **usually** true as well:
- $Y_1, \ldots, Y_n$ are exchangeable for all n
  $\Rightarrow Y_1, \ldots, Y_n$ are iid given θ, $\theta \sim p(\theta)$.
- * It is only approximate when sampling from a finite population without replacement.

Bayes' Law

- In its simplest form, with two events A and B, Bayes' Law relates the conditional probabilities $P(A|B)$ and $P(B|A)$.
- Recall

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- and

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)} =$$

- Hence, $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$

$$\Rightarrow P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

- Similarly,

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}.$$