Bayesian Data Analysis

Lecture 1

Basics of Bayesian Inference

邓婉璐

WANLUDENG@TSINGHUA.EDU.CN

清华大学统计学研究中心

ChatGPT’s Reply

Q: What is the difference between the Bayesian and Frequentist frameworks of Probability Theory?

The Bayesian and frequentist frameworks are two different approaches to probability theory that are used to describe and analyze statistical data. The main difference between the two frameworks is their interpretation of probability.

In the Bayesian framework, probability is interpreted as a measure of belief or degree of certainty about an event or a hypothesis. In this framework, probability is subjective and depends on the individual’s beliefs and values. The Bayesian approach involves updating these probabilities based on new information, using Bayes’ theorem.

In contrast, in the frequentist framework, probability is interpreted as the long-term relative frequency of an event occurring in a large number of trials. In this framework, probability is objective and does not depend on an individual’s beliefs or values. The frequentist approach involves using statistical techniques to estimate the probability of an event based on the observed data.

Both the Bayesian and frequentist frameworks have their own strengths and limitations, and they are used in different contexts depending on the needs of the analysis and the type of data being analyzed.

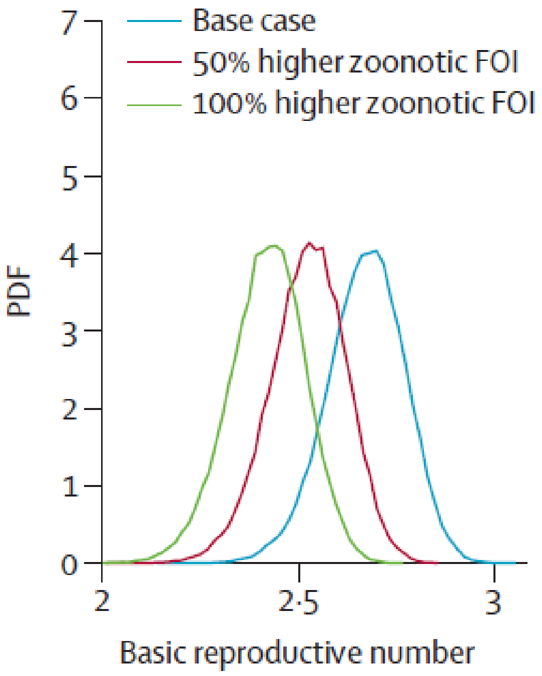
Reference: https://bitdribble.github.io/chatgpt/statistics/2022/12/31/chatgpt/

COVID-19

Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study

Joseph T Wu, Kathy Leung*, Gabriel M Leung*

- Lancet (Wu et al. 2020)
- Bayesian model: Used MCMC method to estimate R_0 (基本繁殖数/传染速率) with a noninformative prior and a SEIR model. 用境外城市的确诊人数和武汉到这些地区的航班数据估计R_0.



Susceptible (S) → Exposure (E) → Infectious (I) → Recovered (R)

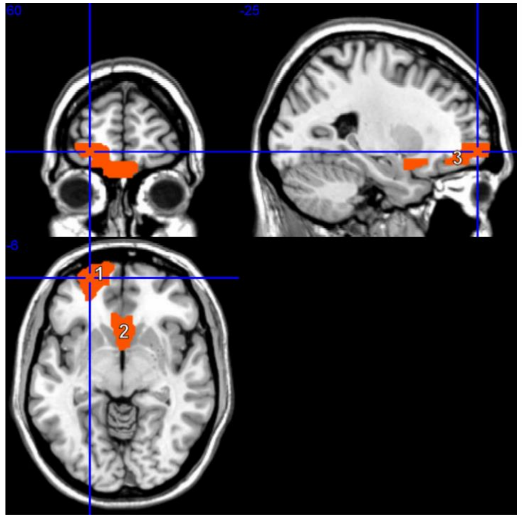
Brain Disease

Published in final edited form as:

J Am Stat Assoc. 2013 January 1; 108(503): . doi:10.1080/01621459.2013.804409.

An Integrative Bayesian Modeling Approach to Imaging Genetics

- Goal: identify brain regions of interest (ROIs) with discriminating activation patterns between schizophrenic patients and healthy controls
- Approach: a Bayesian hierarchical modeling approach



Outline

- What is it? (& Why is it important?)
  - What is ‘Statistics’?
  - What does ‘Bayesian’ mean?
  - Bayesian vs Frequentist
  - Examples to illustrate Bayesian inference
- Basic course information
  - Outline of the course --- How to implement?
- Summary

What is Statistics?

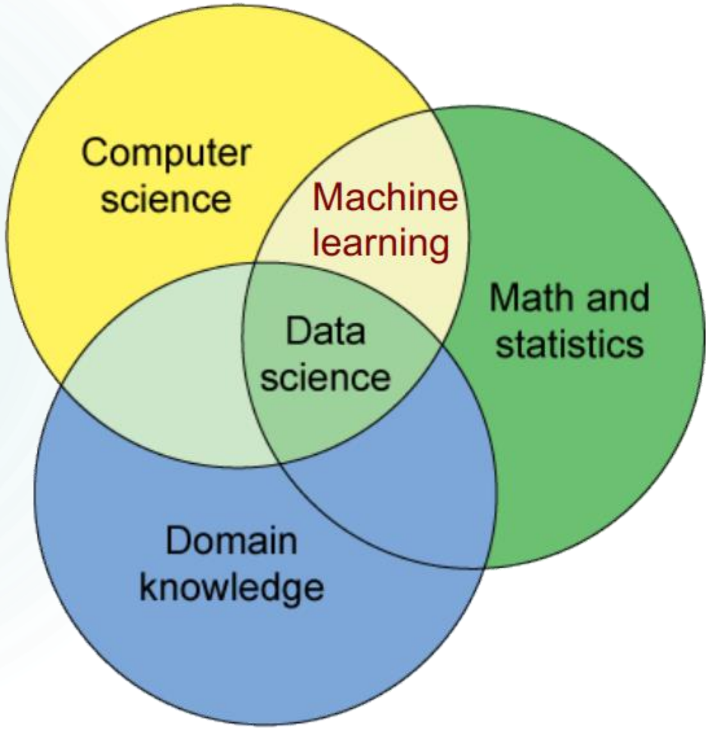
Role of Statistics in Sciences

Statistical Problems:

Any problem that involves uncertainties,

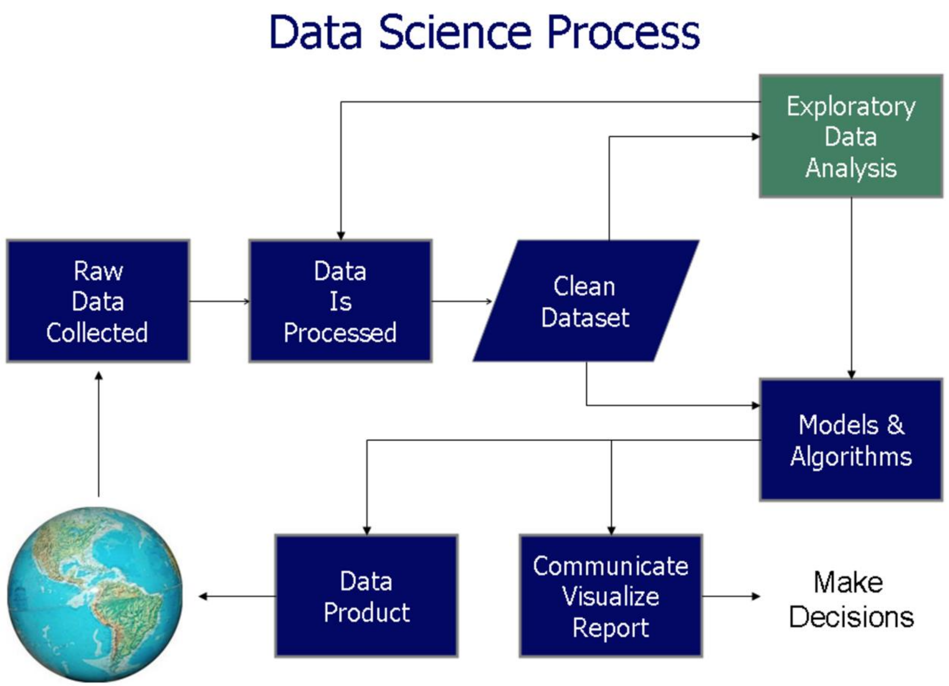
basically all the real problems.

Role of Statistics in Data Science



The Process of Data Analysis

- *Data requirements*
- *Data collection*
- *Data processing*
- *Data cleaning*
- *Exploratory data analysis*
- *Modeling and algorithms*
- *Data product*
- *Communication*



John Turkey

The Future of Data Analysis (1962)

Statistical Modeling

A statistical model is "a device through which we make sense of actual world experiences"

- Express knowledge quantitatively
- Describe the data collection process
- Define primary quantities of interest and connections with data
- Enable formal and principled solutions to questions of interest
- Iterative model improvements

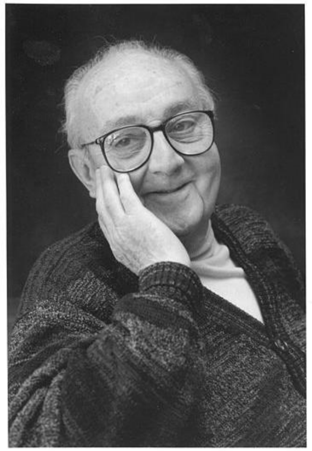
Should be (1) reasonably behaved (consistent), (2) as simple as possible, and (3) faithful to relevant scientific knowledge.

Statistical Modeling: an Example

- The binary model
  - ✓ $P(head\,|\,\theta) = \theta$
  - ✓ $P(tail\,|\,\theta) = 1 - \theta$
- Model parameter: θ
- parameter space: $[0,1]$

A Simplified Model can be very Useful

- *"Essentially, all models are wrong, but some are useful."* -- **George E.P. Box**



- "矛盾是无处不在的, 我们要抓住主要矛盾。"

  -- 毛泽东

What are behind a Model

Assumptions & our understandings of the problem (science)
either explicitly or implicitly

- Mathematical Models: mathematical descriptions of the above;
- Statistical Models: using probabilistic models to connect observations and "theory".

Probability vs Statistics

Distribution $f(\theta)$ *governing the data generating process*

Parameter θ →  → Observation X

- **Probability** $f(\theta) \Rightarrow X\,?$
  - For a given distribution $f(\theta)$, study the properties of $f(\theta)$ itself and samples from $f(\theta)$.
- **Statistics** $?\,f(\theta) \Leftarrow X$
  - Given samples (data X) generated from some unknown distribution $f(\theta)$, study what $f(\theta)$ or θ is.
- **Relation**
  - Two sides of one coin

Statistical Inference

- ***Inference***
  - To infer: "To conclude based on fact and/or premise"
  - Everyday: Make inferences about things unseen based on the observed
- ***Statistical Inference***
  - Facts are the data
  - Premise carried by a probability model
  - Conclusions on unknowns
  - Examples:
    - ✓ Sample mean to estimate (what?)
    - ✓ Linear regression – "estimating" the slope
    - ✓ Whether a certain drug/treatment is effective.
    - ✓ What is the true signal?
    - ✓ Who will win the election?

Key Challenges of Statistical Inferences

- *How to quantify uncertainty of the interested unobserved quantities?*

Is this coin a fair one?

Given outcomes of 4 tosses

Ways to Understand Probability

➢ Long run relative frequencies

- If an experiment of independent trials is repeated over and over, the relative frequency of an event will converge to the probability of the event.
- Let A be the event of interest and let $p = P(A)$. Then for a sequence of independent trials, let

$$Y_i = I(A \text{ occurs in trial } i), \qquad X_n = \sum_{i=1}^{n} Y_i$$

- Then by the law of large numbers, the sample proportion of successes

$$\frac{X_n}{n} \to p, \qquad \text{as } n \to \infty$$

Ways to Understand Probability

- For example, three different experiments looked at the probability of getting a head when flipping a coin.
  - The French naturalist Count Buffon: 4040 tosses, 2048 heads ($\hat{p} =$ 0.5069).
  - While imprisoned during WWII, the South African statistician John Kerrich: 10000 tosses, 5067 heads ($\hat{p} =$ 0.5067)
  - Statistician Karl Pearson: 24000 tosses, 12012 heads ($\hat{p} =$ 0.5005)

> **Frequency argument:** probability = relative frequency obtained in a long sequence of tosses, assumed to be performed in an identical manner, physically independently of each other.

Prob(head)=0.5

Ways to Understand Probability

- Models, physical understanding, etc.
- The structure of the problem will often suggest a probability model.
- For example, the physics of rolling a coin suggest that no one side should be favored (equally likely outcomes), giving the uniform model used in the earlier example. However this could be verified by looking at the long run frequencies.

Symmetry or exchangeability argument:

$$\text{probability} = \frac{\text{number of favorable cases}}{\text{number of possibilities}},$$

assuming equally likely possibilities. For a coin toss this is really a physical argument, based on assumptions about the forces at work in determining the manner in which the coin will fall, as well as the initial physical conditions of the toss.

Prob(head)=0.5

Ways to Understand Probability

- Subjective beliefs
- Can be used for experiments that can’ t be repeated exactly, such as a sporting event.
- For example, what is the probability that the A will win B in the football game next year?
  - Can be done through comparison (i.e. is getting a head on a single flip of a coin more or less likely, getting a 6 when rolling a fair 6 sided die, etc).
  - Can also be done by comparing different possible outcomes (1.5 times more likely than the C, 10 more likely than than D, 1,000,000 times more likely than E, etc).
- Often expressed in terms of odds

$$Odds = \frac{Pr}{1 - Pr}; \qquad Pr = \frac{Odds}{1 + Odds}.$$

Ways to Understand Probability

- The idea of subjective probability fits into the idea of a fair bet.
- Let $p \in [0,1]$ be the amount that you are willing to bet for a return of ¥1 if the event A occurs, i.e. gain ¥$(1-p)$ if A occurs, lose ¥p if A^c occurs.
- If you want this to be a fair bet ($E[gain] = 0$), then p is your subjective probability of the event E occuring.
- Let q be the probability of success. Then
$$E[gain] = (1-p)q - p(1-q)$$
For this to be 0, $q = p$.

What does ‘Bayesian’ mean?

Bayes' Rule

X is a continuous variable

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}$$

$$= \frac{f_{Y|X}(y|x)f_X(x)}{\int_{\Omega_X} f_{Y|X}(y|x)f_X(x)dx}$$

X is a discrete variable

$$P(x|y) = \frac{P(x,y)}{P(y)} = \frac{P(y|x)P(x)}{P(y)}$$

$$= \frac{P(y|x)P(x)}{\sum_x P(y|x)P(x)dx}$$

Generally I will just use the continuous version, as the discrete version will be analogous.

Bayes' Rule

X is a continuous variable

$$P(x|y) = \frac{P(x,y)}{P(y)} = \frac{P(y|x)P(x)}{P(y)}$$

$$= \frac{P(y|x)P(x)}{\int_{\Omega_X} P(y|x)P(x)dx}$$

X is a discrete variable

$$P(x|y) = \frac{P(x,y)}{P(y)} = \frac{P(y|x)P(x)}{P(y)}$$

$$= \frac{P(y|x)P(x)}{\sum_x P(y|x)P(x)dx}$$

Who is Bayes?

What does the rule imply?

Generally I will just use the continuous version, as the discrete version will be analogous.

/beɪz/

Who is Bayes?

- This is reportedly the only known picture of the Reverend Thomas Bayes, F.R.S. — 1701?- 1761.

- As no source is given, the authenticity of this portrait is open to question.

- So what is the probability that this is actually Reverend Thomas Bayes?

Who is Bayes?

- However there is some additional information. How does this change our belief about who this is?

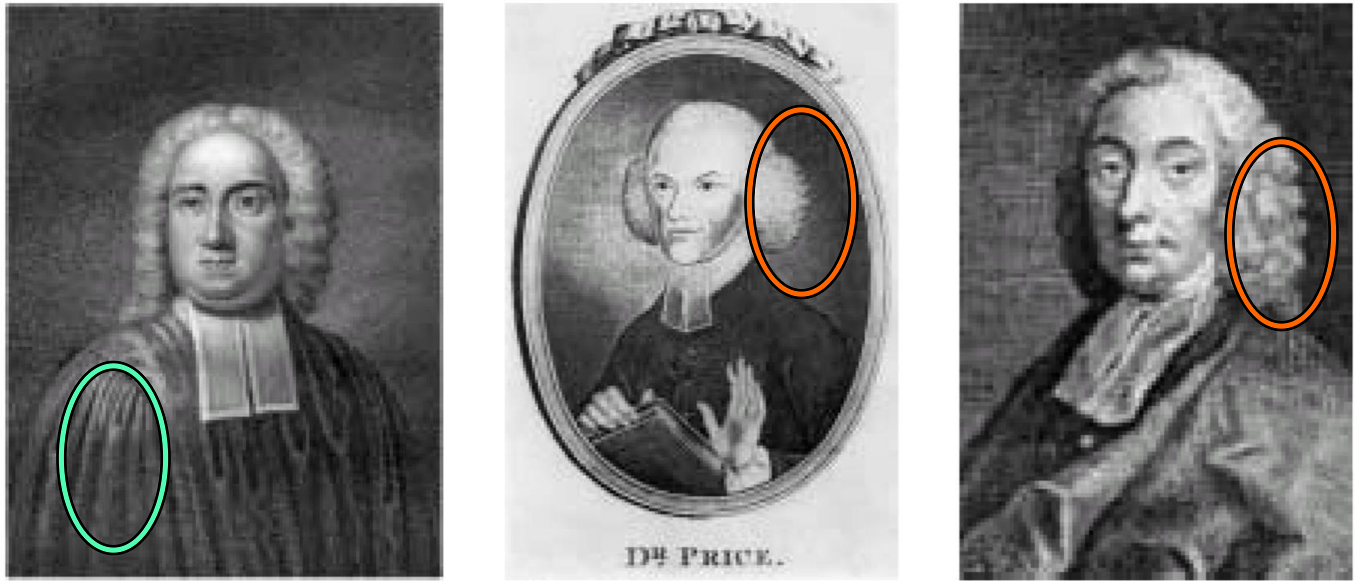
1. The caption under the photo in O'Donnell's book was "Rev. T. Bayes: Improver of the Columnar Method developed by Barrett".

   There are some problems with this claim. First Barrett was born in 1752 and would have been about 9 years old when Bayes died. In addition, the method that Bayes allegedly improved was apparently developed between 1788 and 1811 and read to the Royal Society in 1812, long after Bayes' death.

   So there is a problem here, but whether it is a problem with just the caption or the picture as well isn't clear.

Who is Bayes?

2. Bayes was a Nonconformist Minister. Does the clothing in the picture match that of a Nonconformist Minister in the 1740's and 1750's? The picture has been compared to three other Ministers, Joshua Bayes, Bayes' father, Richard Price (portrait dated 1776), the person who read Bayes' paper to the Royal Society, and Philip Doddridge, a friend of Bayes' brother-in-law.



Joshua Bayes (1671-1746) | Richard Price (1723-1791) | Philip Doddridge (1702-1751)

Who is Bayes?

- Question: How to we incorporate this information to adjust our probability that this is actually a picture of Bayes?
- Answer: $P[This\ is\ Bayes|Data]$ which can be determined by Bayes' Theorem.

$$P[B|\text{Data}] = \frac{P[B]P[\text{Data}|B]}{P[B]P[\text{Data}|B] + P[B^c]P[\text{Data}|B^c]}$$

- Many details about Bayes are sketchy.
- "The date of his birth is not known: Bayes's posterior is better known than his prior. "

Stigler SM (1983). Who Discovered Bayes's Theorem. American Statistician 37: 290-296.

Way to Understand Bayes’ Rule

1. Inverse Probability.
   (Laplace 1749-1827.)

$$P(y|x) \Rightarrow P(x|y)$$

2. Belief Update.
   (e.g. the last example)

$$P(x) \Rightarrow P(x|y)$$

What is "Bayesian Statistics"?

Ways to understand probabilities

1. Long run relative frequencies
2. Physical understanding, etc.
3. Subjective beliefs

+

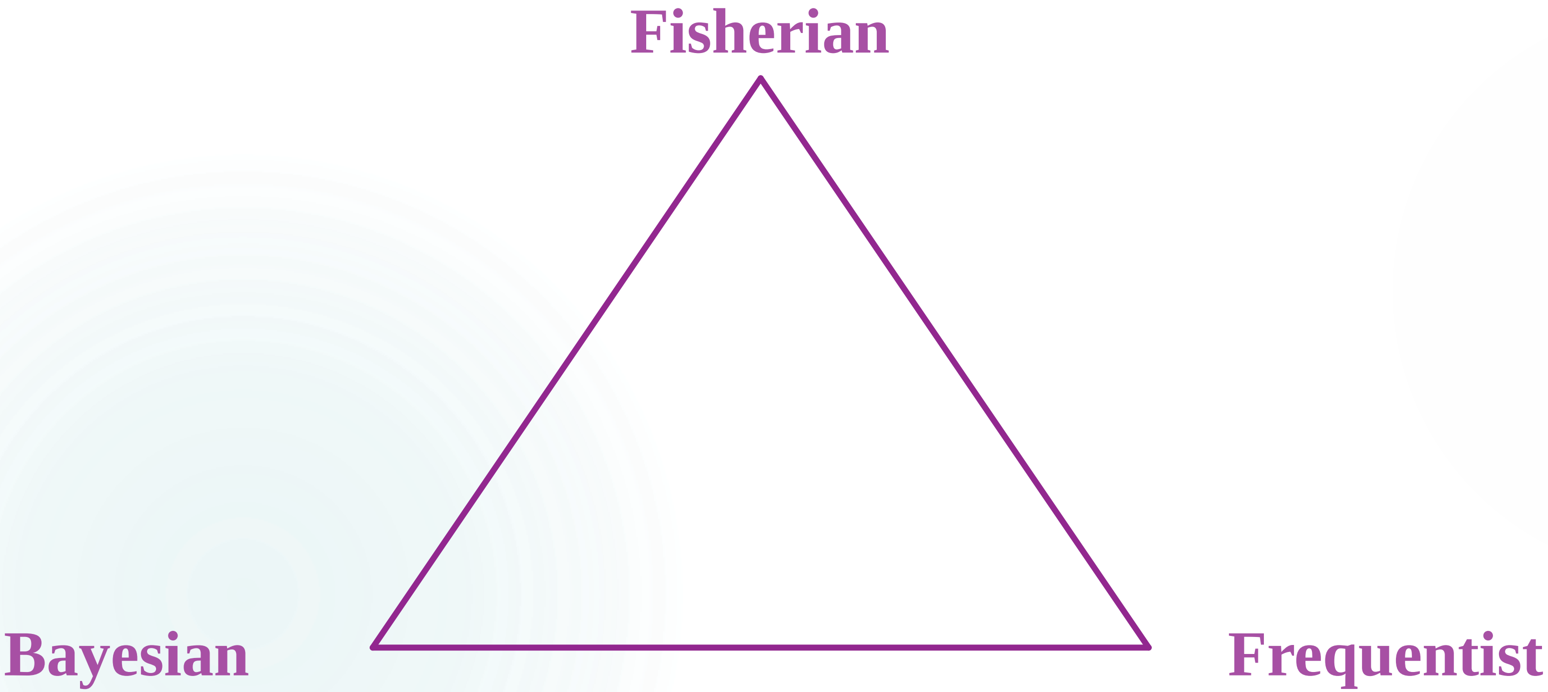
Way to understand Bayes' rule

1. Inverse Probability
2. Belief Update.

In whom do we care the belief, during our traditional statistical inference?

Bayesian vs Frequentist

Different Paradigms of Statistical Inference



The Frequentist Paradigm

- ***Key opinion***
  - treat model parameters as unknown but fixed constants
- ***Main approaches***
  - Point estimation: $S = f(data)$
  - Interval estimation: $[a, b]$ where $a = f_1(data) < b = f_2(data)$
  - Hypothesis testing:

$$H_0\text{: } \theta = \theta_0 \quad v.s. H_1\text{: } \theta = \theta_1$$

- ***Frequentist interpretation***
  - A hypothetical sequence of repeated experiments

Limitations of Frequentist Paradigm

- Need a hypothetical sequence of repeated experiments
- Uncertainty is quantified in a population base (i.e., patients like me)

Patient: what the successful rate of my surgery?
Doctor: 95% based on my history?

Limitations of Frequentist Paradigm

- Logically cannot quantify uncertainty for a specific data set
- But, in practice we always make decision for a particular data set, which is unique (i.e., nobody is really identical to me)

Patient: what the successful rate of my surgery?
Doctor: 95% based on my history?

Giving a correct answer to a wrong problem!

The Modified Coin Example

Toss a double-head coin
or
a double-tail coin

Double-head coin Double-tail coin

Prob(head) = 0.5 ?

- The symmetry argument still works well
- The frequency argument encounters difficulty

The Bayesian Paradigm

- ***Key opinion***
  - Treat unknowns (model parameters) as random variables
- ***Main elements***
  - Prior distribution: $p(\theta)$
  - Likelihood function: $p(data\,|\,\theta)$
  - Joint distribution: $p(data, \theta)$
  - Posterior distribution: $p(\theta\,|\,data)$
- ***Bayesian inference***
  - Describe quantities of scientific interest with probability
  - Let the ***law of probability*** work its way out
  - Summarize all uncertainty into posterior distribution

Bayesian Framework of Inference

- ***Generic Notations***

  $p(y \mid \theta)$ --- Sampling distribution or data likelihood

  $p(\theta)$ --- Prior distribution

  Why do we incorporate prior?

- ***The Bayes Theorem***

$$\underbrace{p(\theta|y)}_{\text{Posterior}} = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)} \propto \overbrace{p(\theta)}^{\text{Prior}}\underbrace{p(y|\theta)}_{\text{Data likelihood}}$$

- ***Forecast***

$$\begin{aligned} p(\tilde{y}|y) &= \int p(\tilde{y}, \theta|y)d\theta \\ &= \int p(\tilde{y}|\theta, y)p(\theta|y)d\theta \\ &= \int p(\tilde{y}|\theta)p(\theta|y)d\theta. \end{aligned}$$

Post predictive distribution

Example: Spell Correction

Spell Correction

radom → random ?
radom → radon ?
radom → radom ?

Prior probability:

θ	$p(\theta)$
random	7.60×10^{-5}
radon	6.05×10^{-6}
radom	3.12×10^{-7}

Error rate:

θ	$p(\text{‘radom’}\mid\theta)$
random	0.00193
radon	0.000143
radom	0.975

Bayes theorem:

$$\Pr(\theta \mid y = \text{‘radom’}) \propto p(\theta)\Pr(y = \text{‘radom’} \mid \theta)$$

$$p(\text{random}\mid\text{‘radom’}) = \frac{p(\theta_1)p(\text{‘radom’}\mid\theta_1)}{\sum_{j=1}^{3} p(\theta_j)p(\text{‘radom’}\mid\theta_j)}$$

Posterior probability:

θ	$p(\theta)p(\text{‘radom’}\mid\theta)$	$p(\theta\mid\text{‘radom’})$
random	1.47×10^{-7}	0.325
radon	8.65×10^{-10}	0.002
radom	3.04×10^{-7}	0.673

θ_1=random, θ_2=radon, and θ_3=radom.

Example: Binomial Model

- Laplace' s work on boy birth rate in Paris. (1786)
  - Want to test if the rate is over 0.5.
  - 1745-1770, 251527 boys, 241945 girls.

- Assume that each baby is independent and the outcome each time is a Bernoulli trial with success probability θ.

- This implies that $y|\theta \sim Bin(n, \theta)$

- The sample size n is fixed and the success probability θ can be any number in $[0, 1]$.

Example: Binomial Model

- The measurement model is

$$p(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}, \qquad \theta \in [0,1]$$

Since n is fixed in this analysis, we' ll drop it as a conditioning argument for ease of notation.

- Goal: make inference on θ given y and n.
- Need: a prior distribution $p(\theta)$ for θ.
- Bayes' choice: $\theta \sim U(0,1)$, i.e.

$$p(\theta) = \begin{cases} 1 & 0 \le \theta \le 1 \\ 0 & \text{Otherwise} \end{cases}$$

Example: Binomial Model

- So the Bayes' rule gives

$$p(y,\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y},$$

$$p(y) = \int_0^1 \binom{n}{y}\theta^{y}(1-\theta)^{n-y}d\theta = \frac{1}{n+1},$$

$$p(\theta|y) = (n+1)\binom{n}{y}\theta^{y}(1-\theta)^{n-y}.$$

- Given data, Laplace calculated that
  $Prob(boy\ birth\ rate\ \le 0.5) < 10^{-42}$.
  So he concluded that the rate is over 0.5.

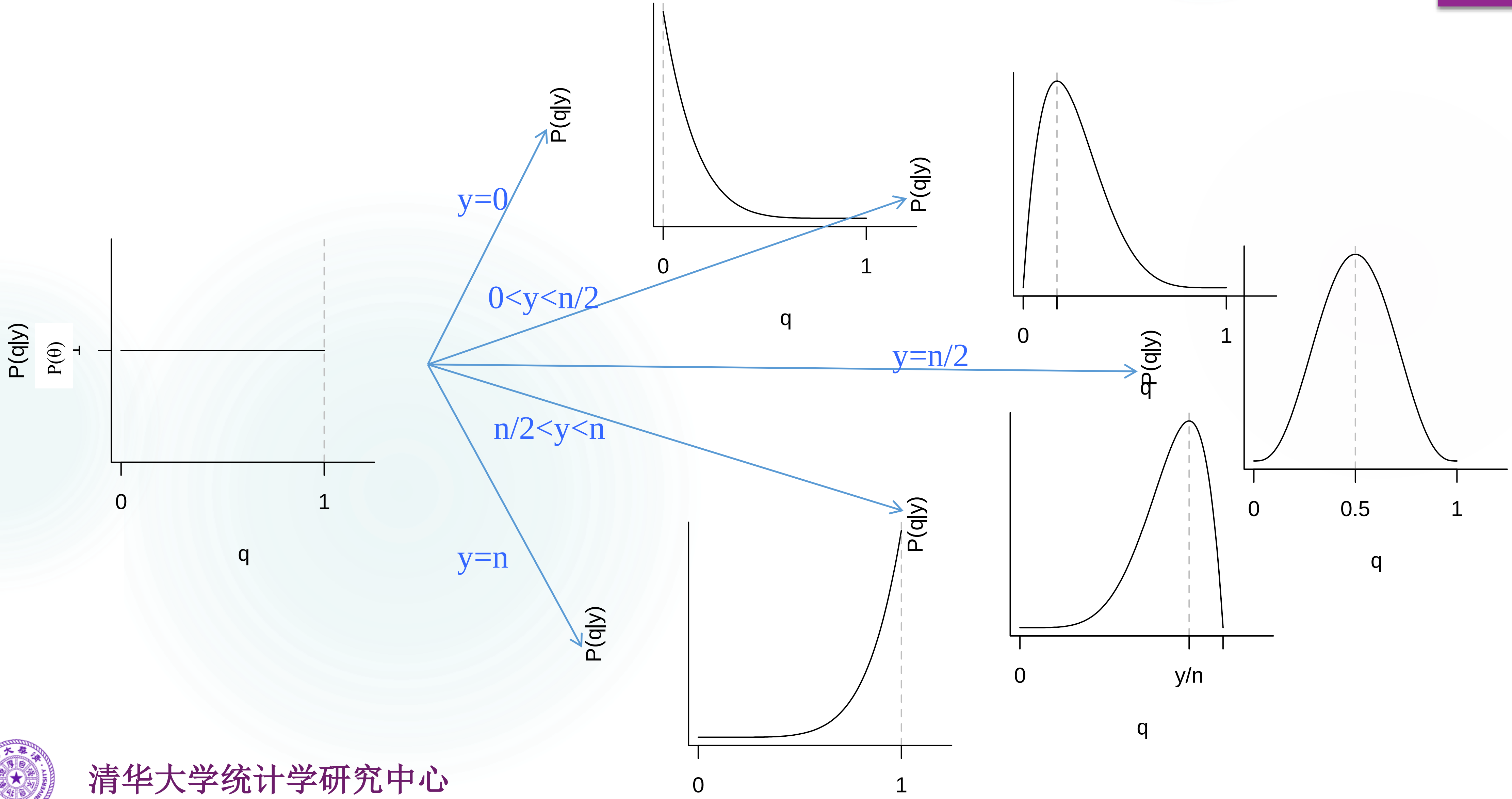
Example: Binomial Model

- $p(\theta|y) = (n+1)\binom{n}{y}\theta^y(1-\theta)^{n-y}$. $\theta|y \sim ?$
- The PDF for the for the beta distribution ($Beta(\alpha,\beta)$) is $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$.
- It is often easier to deal with the proportional form of Bayes' Rule

$$p(\theta|y) \propto p(\theta)p(y|\theta)$$

- So in this example,
  - $p(\theta|y) \propto 1 \times \theta^y(1-\theta)^{n-y}$.
  - This is proportional to the density of a $Beta(y+1, n-y+1)$ distribution.

Example: Binomial Model



Example: Binomial Model

- Assume that we have two different samples:
  - A: we checked 8 children one by one, got the ordered sample (1,1,1,0,1,0,1,0), with boy denoted as 1 and girl denoted as 0.
  - B: we checked 8 children one by one, got the ordered sample (0,0,1,0,1,1,1,1), with boy denoted as 1 and girl denoted as 0.
- Will the posteriors be different?

- Bayesian analysis satisfies the Likelihood Principle: the two data sets, with the same likelihood function, should lead to the same inferences.

Thinking

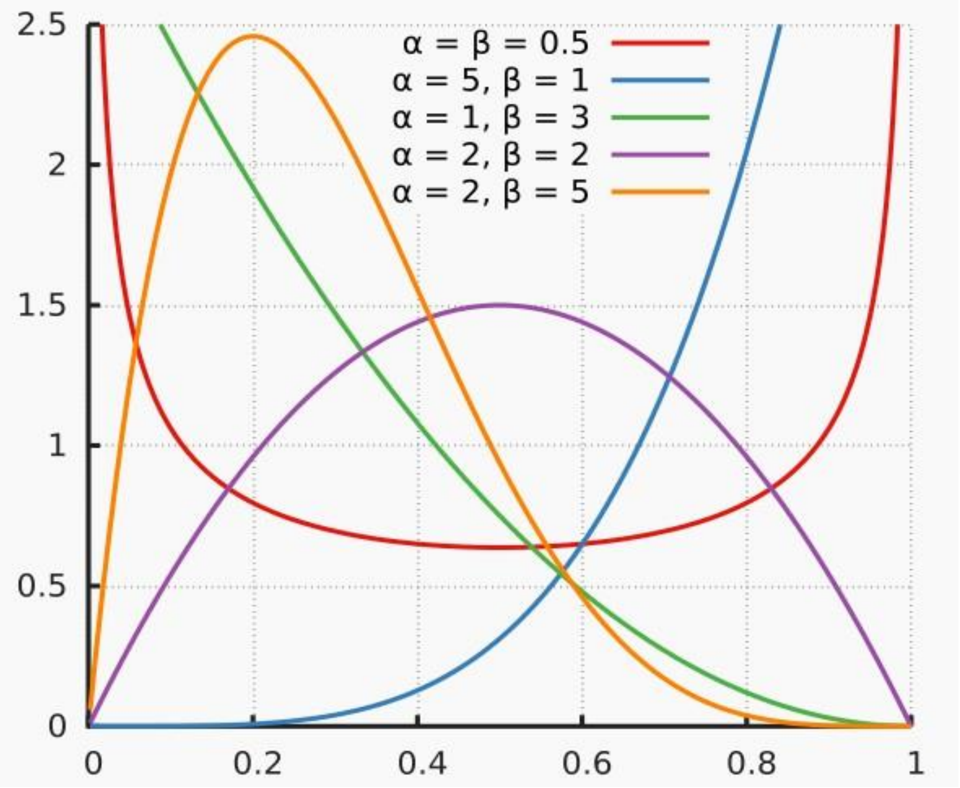
- If we got evidence sequentially, will the order influence our results?
  - There is an example on a genetic status (Section 1.4 in Chapter 1). You may use this example to check you conclusion.

- Under Frequentist paradigm, what inference about the parameters do we care? How can we achieve these goals under Bayesian paradigm?
  - Point estimation
  - Confidence Interval
  - Hypothesis testing

Example: $Beta(y+1, n-y+1)$

Notation	Beta(α, β)
Parameters	α > 0 shape (real) β > 0 shape (real)
Support	$x \in (0,1)$
PDF	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}$
CDF	$I_x(\alpha,\beta)$
Mean	$\mathrm{E}[X] = \frac{\alpha}{\alpha+\beta}$ $\mathrm{E}[\ln X] = \psi(\alpha) - \psi(\alpha+\beta)$ (see digamma function and see section: Geometric mean)
Median	$I_{\frac{1}{2}}^{[-1]}(\alpha,\beta)$ (in general) $\approx \frac{\alpha-\frac{1}{3}}{\alpha+\beta-\frac{2}{3}}$ for $\alpha,\beta>1$
Mode	$\frac{\alpha-1}{\alpha+\beta-2}$ for α, β >1
Variance	$\mathrm{var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ $\mathrm{var}[\ln X] = \psi_1(\alpha) - \psi_1(\alpha+\beta)$ (see trigamma function and see section: Geometric variance)

Beta density functions



Summarizing Posterior Inference

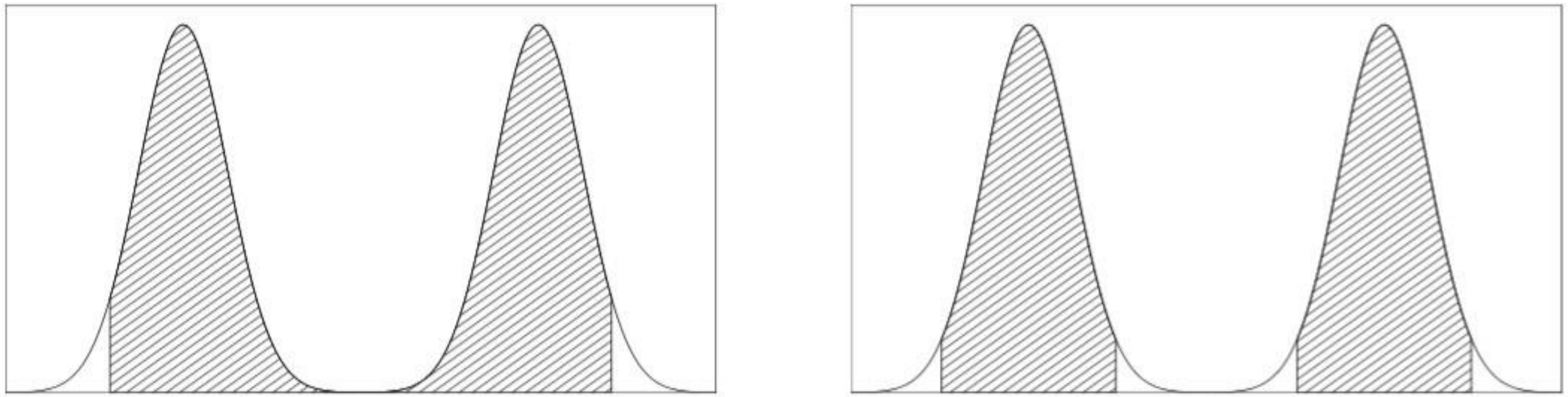
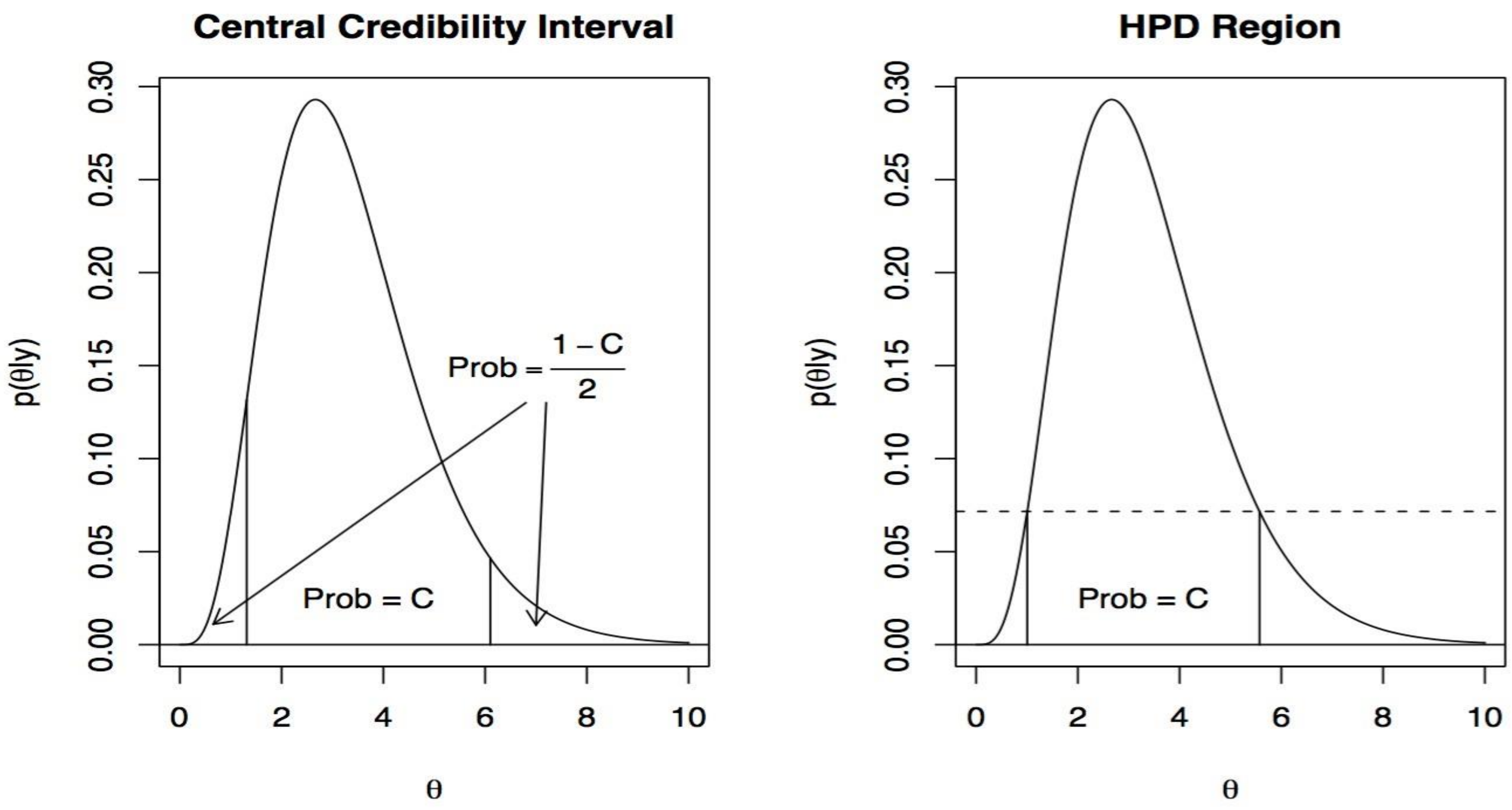


Figure 2.2 *Hypothetical density for which the 95% central interval and 95% highest posterior density region dramatically differ: (a) central posterior interval, (b) highest posterior density region.*

- summaries of location: mean, median, and mode(s) / Max A Posteriori (MAP)
- summaries of variation: standard deviation, interquartile range, other quantiles
- Bayesian version of confidence interval: **credible interval**

  Central posterior interval & highest posterior density region

Summarizing Posterior Inference



Central interval: $(1.313, 6.102)$; length $= 4.789$

HPD interval: $(1.006, 5.571)$; length $= 4.565$

Example: Binomial Model

- How did you calculated?
  - Routine methods: calculate $p(\theta_i|y), i = 1,2$, and compare.
  - Odds ratios:

    Bayes' rule has a nice form in terms of odds ratios

Prior Odds

$$\frac{p(\theta_1|y)}{p(\theta_2|y)} = \frac{p(\theta_1)p(y|\theta_1)}{p(\theta_2)p(y|\theta_2)} = \frac{p(\theta_1)}{p(\theta_2)} \cdot \frac{p(y|\theta_1)}{p(y|\theta_2)} = \frac{p(\theta_1)}{p(\theta_2)} \cdot \frac{L(\theta_1|y)}{L(\theta_2|y)}$$

Posterior Odds

Likelihood Ratio

Hypothesis Testing

- Suppose we observe n=5, y=1, and we know θ=0.1 or 0.2.

Frequentist

$$H_0: \theta = 0.1, H_1: \theta = 0.2$$

P-value

$$P(y \geq 1|H_0) = 1 - P(y = 0|H_0)$$
$$= 1 - 0.9^5 \approx 0.41$$

$$H_0: \theta = 0.2, H_1: \theta = 0.1$$

P-value

$$P(y < 1|H_0) = P(y = 0|H_0)$$
$$= 0.8^5 \approx 0.33$$

Bayesian

$$H_1: \theta = 0.1, H_2: \theta = 0.2$$

Prior $P(H_1) = 0.5, P(H_2) = 0.5$

Likelihood

$$P(y = 1|H_1) = \binom{n}{y} 0.1 \times 0.9^4 \approx 0.33$$

$$P(y = 1|H_2) \approx 0.41$$

Posterior

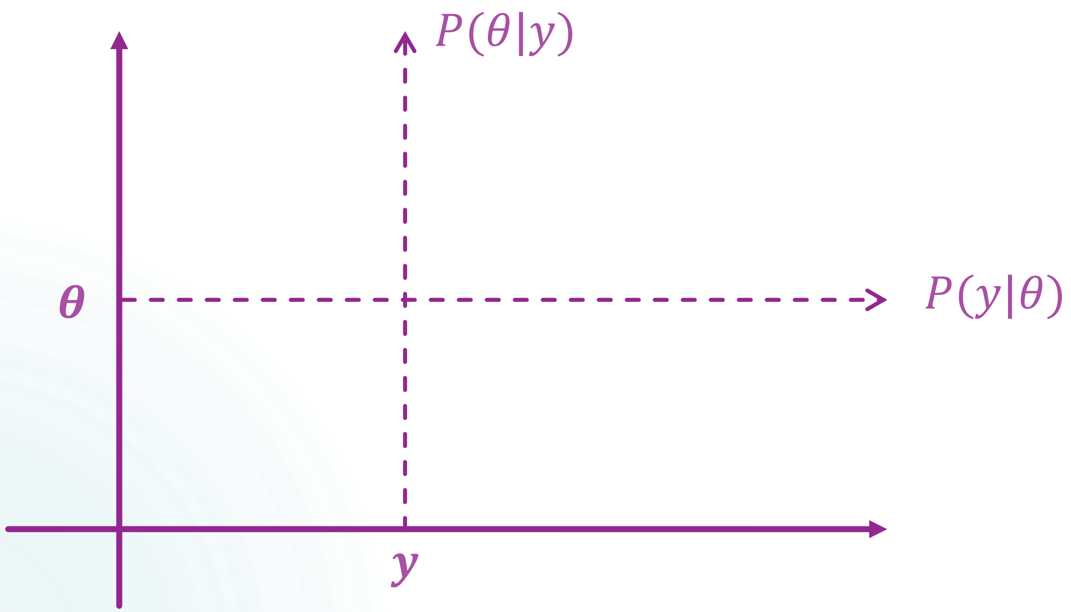
$$P(H_1|y = 1) = \frac{P(H_1)P(y = 1|H_1)}{P(y = 1)} \approx 0.45$$

$$P(H_2|y = 1) \approx 0.55$$

Posterior Odds

$$\frac{P(H_1|y = 1)}{P(H_2|y = 1)}$$

Bayesian versus Frequentist



- Bayesian inference proceeds vertically, given data y; requires a prior distribution $p(\theta)$.
- Frequentist inference proceeds horizontally, given parameter θ; requires a method or algorithm $t(x)$ designed to answer the specific question at hand.

Bayesian versus Frequentist

- ***Advantages***
  - Answers the questions that researchers are usually interested in, "What is the probability that …"
  - Almost all frequentist methods have Bayesian counterparts
  - Easy and consistent to combine information: Formal method for combining prior beliefs with observed (quantitative) information. Natural way of combining information from multiple studies or dynamic contexts.
  - Easy to deal with real problems with complicate structures
  - Many powerful sampling techniques can be used to support the analysis
- ***Critical issue***
  - Selection of the prior (subjectivity of Bayesian approach)

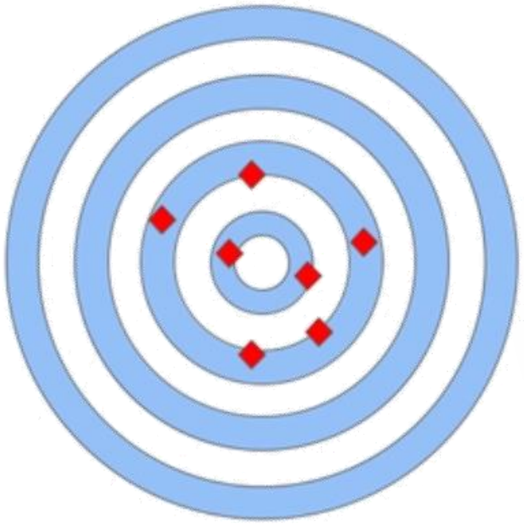
Subjectivity & Objectivity

- All statistical methods that use probability are subjective in the sense of relying on mathematical idealizations of the world.
- Bayesian methods are sometimes said to be especially subjective because of their reliance on a prior distribution, but in most problems, scientific judgment is necessary to specify both the ‘likelihood’ and the ‘prior’ parts of the model.
- For example, linear regression models are generally at least as suspect as any prior distribution that might be assumed about the regression parameters.

Some people think that:

Bayesian paradigm tries to answer the correct problem!

Accuracy & Precision

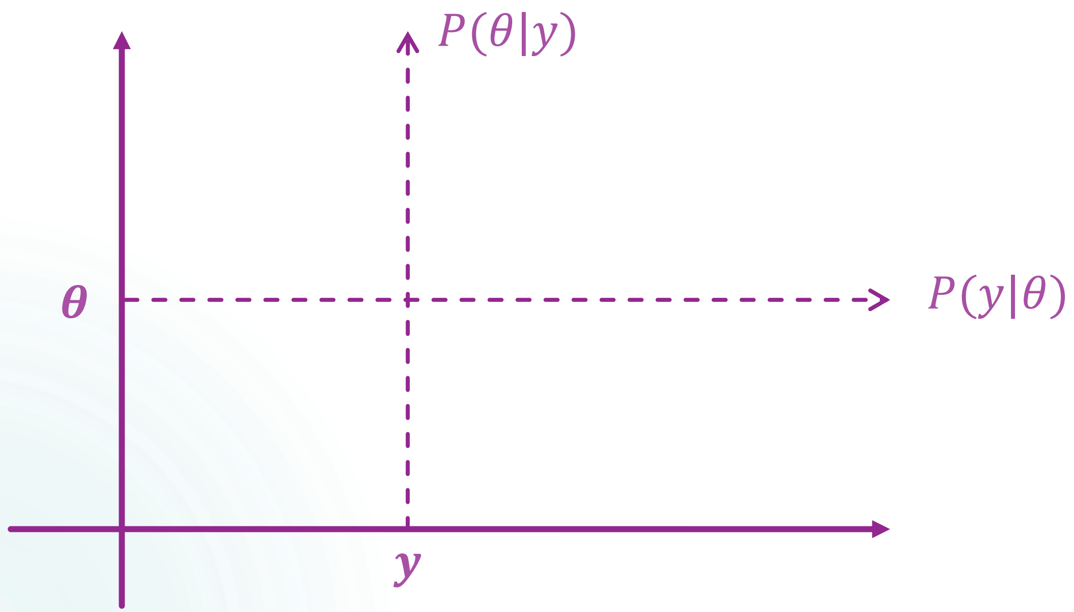


High precision (capability)
Low accuracy (alignment)

Low precision (capability)
High accuracy (alignment)

Reference: https://www.assemblyai.com/blog/how-chatgpt-actually-works/

Bayesian versus Frequentist



- A Bayesian analysis answers all possible questions at once.
- Frequentism focuses on the problem at hand, requiring different estimators for different questions. This is more work, but allows for more intense inspection of particular problems.

Basic Course Information

Remarks

- Forms of data representation
  - *i.i.d.*
  - Time series
  - the semantic type of data (statements, texts, pictures, etc.).
- Principal aims of statistical data analysis
  - compact representation of data,
  - estimation of model parameters explaining and/or revealing data structure,
  - prediction.

Not covered:

- Nonparametric Bayesian
- Variational Bayesian
- Bayesian Networks
- Missing data

Outline of the course

- Basics of Bayesian Inference
- Single-Parameter Models, Multi-Parameter Models, Hierarchical Models
- Asymptotic Normality
- Model Checking, Evaluating and Comparing

Models

Evaluation

- Bayesian Computation: non-iterative methods, iterative sampling methods (Markov Chain Monte Carlo, MCMC); including Metroplis-Hastings algorithm, Gibbs sampling, etc.
- Linear Regression
- Hierarchical Linear Models

Computation

Goal for the course

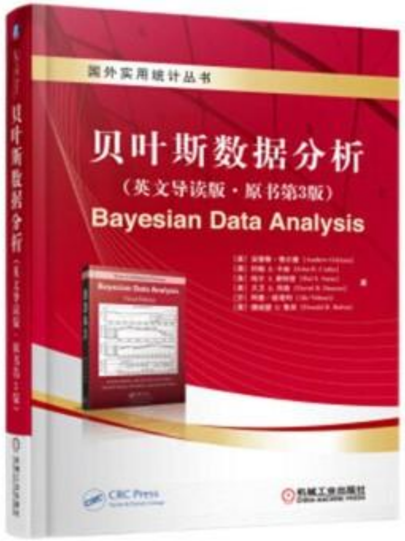
- Understand the basic idea of Bayesian statistics
- Be Capable of building reasonable Bayesian models in real data analysis.
- Be Capable of deriving elements of interest in Bayesian analysis
- Be Capable of performing computation for Bayesian analysis
- Have the potential to combine Bayesian analysis with new models / your domain knowledge to deal with real data analysis

Prerequisite & Requirements

- Prerequisite:
  - Undergraduate level probability, statistical inference, linear regression analysis and multivariate statistical analysis.
  - Computation: R
- Grading:
  - Homework 35% (电子版，附相应代码；可以讨论、拒绝抄袭，如发现雷同，当次作业记零分；迟交扣分，详见助教公告)
  - Midterm Exam (basic concept, model, procedure) 30%
  - Final Exam (model, procedure, computation) 35%
- 请关注网络学堂：课件、数据、作业、通知等。

Reference books

✧ Gelman A, Carlin JB, Stern HS, etc. (2013) Bayesian Data Analysis, 3rd edition. Chapman & Hall/CRC.



- ✓ 电子版 – 校内教参服务: http://reserves.lib.Tsinghua.edu.cn
- ✓ 纸板 – 《贝叶斯数据分析》(英文导读版) 机械工业出版社，Available from 京东

✧ References:

- Liu J (2001). Monte Carlo Strategies in Scientific Computing. Springer-Verlag.
- Faming Liang, Chuanhai Liu, Raymond J. Carroll (2010). Advanced Markov Chain Monte Carlo Methods, Wiley.

Software

- R. The tool that we will use in this class. Familiarity with R is assumed.

- Python. (PYMC3)

- Other choice:
  - Stan (Used in our main reference book)
  - JAGS
  - BUGS (Bayesian inference Using Gibbs Sampling). Windows and Linux (Intel based) versions are available.

Personnel

- Instructor: 邓婉璐
  - E-mail: wanludeng@mail.tsinghua.edu.cn
  - Office: 伟清楼203C
  - Office Hours: 周三上午 10-11am

- Teaching Assistant: 于浩洋、潘庆一
  - E-mail: yhy22@mails.tsinghua.edu.cn , panqingyi19@gmail.com
  - Location: 伟清楼二层
  - Office Hours: 周四下午 4-5pm、周五上午 10-11am

Summary

Key Points for Today

- Statistical inference deals with uncertainty, and involves assumptions and models.
- 3 ways to understand probability.
- Basic concepts in Bayesian statistics: prior, posterior, likelihood
- Basic rule in Bayesian statistics: Bayes' rule
- Basic inference procedure in Bayesian statistics
- Comparison between Bayesian and Frequentist

Reference

- Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2003). Bayesian Data Analysis (3rd ed), Chapman & Hall: London. (Textbook) – Chapter 1

- *"Since all models are wrong the scientist cannot obtain a 'correct' one by excessive elaboration. ... Just as the ability to devise simple but evocative models is the signature of the great scientist so overelaboration and overparameterization is often the mark of mediocrity. Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad. "* -- **George E.P. Box**