The scientific method

1. Ask a question or pose a problem
2. Collect the existing relevant information
3. Based on information obtained in step 2, design an investigation or experiment to address the question posed in step 1
4. Carry out the investigation or experiment
5. Use the evidence from step 4 to update the previously known information, then draw conclusions.
6. Repeat steps 3-5 as necessary.

Where does statistics fit in?

- Central to steps 2, 3, and 5
  - Experimental design, survey sampling
  - Statistical inference
- Bayesian statistics is particularly well suited for step 2 and step 5
  - Step 2: prior information
  - Step 5: prior information → posterior information

A bit history

- Thomas Bayes
  - Born:1702 in London, England
  - Died: 1761 in Kent, England
  - Elected a fellow of the Royal Society in 1742 despite he had no published work on mathematics at that time.
- *Essay towards solving a problem in the doctrine of chances*
  - Set out Bayes' theory of probability
  - Published in Philosophical Transactions of the Royal Society of London in 1764
  - The paper was sent by a friend of Bayes

Applications of Bayesian statistics

- Economics

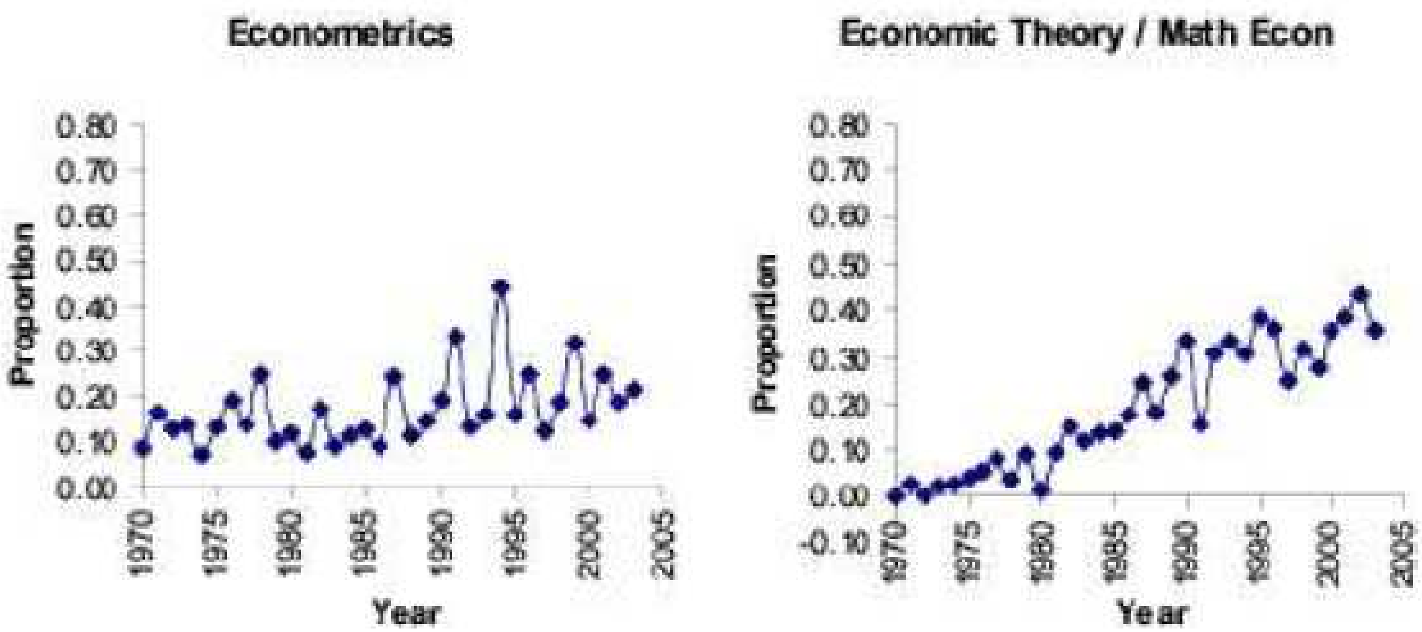


Figure 5: Econometrica Containing "Bayes" or "Bayesian"

Applications of Bayesian Statistics

- Marketing
- Social science
- Education
- Medical research
- Genomics
- Weather

Quantification of uncertainty

- Classical
- Frequentist
- Bayesian

An example of Bayes' theorem

- Enzyme-Linked Immuno Sobert Assay (ELISA)
  - A diagnosis tool for HIV
  - Low cost, about \$20 each
  - Less accurate compared to some more expensive tools
- Let A = {the patient is HIV positive}. Is it proper to use Pr(A)?
- How should we update the uncertainty after a test?

An example of Bayes' theorem (cont)

- Data: test result (**positive**/negative)
- Update the uncertainty using Bayes' theorem

$$P(A|+) = P(+|A)P(A)/[P(+|A)P(A)+P(+|\text{not } A)P(\text{not } A)]$$

- Sensitivity: $P(+|A) = 0.95$
- Specificity: $P(-|\text{not } A) = 1-P(+|\text{not } A) = 0.98$
- Population prevalence: $P(A) = 0.01$

An example of Bayes' theorem (cont)

- $P(A|+) = 0.32$
- False positive rate: $P(\text{not } A|+) = 1-P(A|+) = 0.68$
- False negative rate: $P(A|-) = 0.00052$, about 1 in 1941.

The example in the Bayesian language

- Prior: $P(A) = 0.01$
- Posterior: $P(A|+) = 0.32$
- Another formulation of Bayes' theorem
  - $P(A|+)/P(-|+) = [P(A)/P(\text{not } A)] \times [P(+|A)/P(+|\text{not } A)]$
  - Posterior odds = prior odds x BF

General form of Bayes' theorem

- Model: $y|\theta \sim f(y|\theta)$
  - y: data
  - θ : parameter
- Bayes' theorem
  - $f(\theta|y) = f(\theta,y)/f(y) = f(y|\theta)f(\theta)/f(y) \propto f(y|\theta)f(\theta)$
  - Posterior $\propto$ likelihood x prior

Key issues in Bayesian statistics

- How to specify prior distributions?
- How to compute the posterior $f(\theta \mid y)$?
  - In general, we need to know the marginal distribution $f(y) = \int f(y \mid \theta) f(\theta)\, d\theta$
  - Direct computation of this integral is difficult especially when θ is high-dimensional
  - Simulation is the general solution
- How to summarize the posterior?
  - Posterior mean, median or mode
  - Credible interval

Comparison of Bayesian and frequentist methods

- Interval estimation
  - Frequentist: confidence interval
    - The experiment should be repeatable
    - Not conditional on the data
  - Bayesian: credible interval
- Hypothesis testing
  - Consider testing H0: $\theta=0.5$ versus H1: $\theta>0.5$
  - Frequentist (Neyman-Pearson):
    - pvalue can only be computed when the parameter value is given under the null
    - If we switch the role of two hypotheses, frequentists can not compute the pvalue
  - Bayesian: no trouble

Summary: frequentist v.s. Bayesian

- Disadvantages of Bayesian
  - Subjective choice of prior
  - Computation of posterior can be difficult
- Advantages of Bayesian
  - Can easily incorporate prior information
  - Inference are conditional on the actual data
  - Results are more easily interpretable
  - Inference are based on the posterior, which is conceptually simple
  - In principle, all problems can be solved

Some facts on Bayesian computation

- Posterior is proportional to (likelihood x prior)

- Making inference using sufficient statistics is equivalent to using the original data

- Prediction
  - Predictive distribution

$$f(y_{n+1} \mid y_1, \ldots, y_n) = \int f(y_{n+1} \mid \theta) f(\theta \mid y_1, \ldots, y_n) d\theta$$

Some often used basic facts in probability

- $E(X)=E[E(X|Y)]$
- $Var(X) = E[Var(X|Y)] + Var[E(X|Y)]$
  - On average, posterior variance is smaller than prior variance
- Computing the distribution of a random variable after transformation
  - Change-of-variable formula

Some basic models

- Normal-normal model
  - Model: $y_1, \ldots, y_n \mid \theta \sim N(\theta, \sigma^2)$ with known σ^2, i.i.d.
  - Prior: $\theta \sim N(\mu, \tau^2)$
- Posterior $\theta \mid y_1, \ldots, y_n \sim N(\ ?\ ,\ ?\ )$
- Interpretation of the posterior mean
- Prediction

Some basic models (cont)

- Beta-binomial model
  - $y \mid \theta \sim \text{binomial}(n, \theta)$
  - $\theta \sim \text{beta}(a, b)$
- Posterior: beta(? , ?)
- Interpretation of posterior mean
- prediction

Some other one-parameter models

- Normal with known mean but unknown variance
  - $y_1, \ldots, y_n \mid \sigma^2 \sim N(\theta, \sigma^2)$
  - Prior: Inverse gamma
- Poisson model
  - $y_1, \ldots, y_n \mid \theta \sim \text{Poisson}(\theta)$
  - Prior: gamma
  - Marginal model: negative-binomial
- Exponential model
  - $y_1, \ldots, y_n \mid \theta \sim \text{Exp}(\theta)$
  - Prior: gamma

Conjugate prior

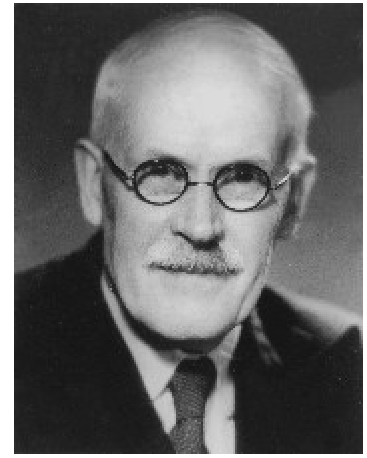
- For exponential family

- Mixture of conjugate prior → mixture of ‘conjugate’ posterior

Prior specification

- Elicited (informative) prior
  - Histogram
  - Determining parameters in priors using moments or quantiles
    - Two different distributions can have the same set of quantiles or moments
  - Interactive computer programs
  - Using a hierarchical structure is often helpful
- Conjugate prior
  - Computationally convenient
- Noninformative prior
- Empirical Bayes

Noninformative prior



- Flat (uniform)
  - Improper when the support is not finite
  - Improper prior can still lead to proper posterior
  - Usually not flat after transformation
- Jeffrey's prior
  - Invariant under one-to-one transformation
- Location invariant prior for location family
- Scale invariant prior for scale family