# Bayesian inference for sample surveys

Roderick Little

Trivellore Raghunathan

Module 1: Course preliminaries, introduction to Bayesian inference

UNIVERSITY OF MICHIGAN

# Instructors

- Rod Little, Professor of Biostatistics and Research Professor, Michigan Program in Survey Methodology, Institute for Social Research
- T. E. Raghunathan, Professor of Biostatistics and Director, Survey Research Institute and Research Professor, Michigan Program in Survey Methodology, Institute for Social Research.

# Locations

- Class Meeting time: Mondays and Wednesdays 10:30am-12pm
- First meeting: Wednesday Jan 4, 2017
- Location: Room 368 (Basement), ISR. This course is offered jointly with the Joint Program in Survey Methodology (JPSM) through an interactive video transmission system.
- ISR 4036 10:30-12:00 on following dates:
- Jan 9th, Feb 13th, March 13th, April 10th

# Course Objectives

- Bayesian methods in statistics are increasingly popular, spurred by advances in computational power and tools.
- Bayesian inference provides solutions to problems that cannot be solved exactly by standard frequentist methods.
- Bayesian approach provides new analysis tools, and a deeper understanding of competing systems of statistical inference, including the frequentist approach.
- Describe the application of the Bayesian approach to survey sampling, where the focus of inference is on finite population quantities. This course will emphasize both theoretical and applied aspects of Bayesian inference, in general, and to sample surveys, in particular.

# Prerequisites

- This is an advanced course in statistics.
  - Students should be familiar with standard statistical methods based on likelihood and other statistical concepts in probabilities and distributions, sufficient statistics, point and interval estimation, testing of hypothesis and basic asymptotic theory. Students should also be well versed in calculus.
- The course will be a mix of theory and computations.
  - Computations will be performed using
  - WinBugs -- free software for Bayesian Analysis that can be downloaded from the website (www.mrc-bsu.cam.ac.uk/bugs),
  - R (free software similar to S-plus (www.r-project.org)
  - IVEware, imputation and analysis software (www.iveware.org).
  - During the first week of classes, download these three software packages and install them on your computers.

# Grading

- Letter grades for the course will be based on
  - homework assignments, given periodically, a mix of theoretical and applied problems. Grade will be based on completing and handing in assignments (20%)
  - a midterm take- home examination (40%)
  - a final project. Either a theoretical project demonstrating an advantage (or disadvantage) of the Bayesian approach to a particular problem or an applied survey data analysis using the Bayesian approach. The final project will be graded based on a written report not exceeding 15 pages including tables, figures and references. (40%)

# Textbook

- Though there are no required textbooks, you might want to buy the third edition of the book "Bayesian Data Analysis" by Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, Donald B. Rubin · CRC Press
- Course notes will be provided from time to time

# Course Topics

1. Modes of inference in surveys; Fundamentals of Bayesian Inference; Application to surveys using simple random sample (srs) design
2. Basic Monte-Carlo methods (Non-iterative)
3. Role of complex sample designs in Bayesian inference; unequal probabilities of selection and stratification
4. Auxiliary variables and nonparametric methods
5. Hierarchical models for cluster sample designs
6. Bayesian models for complex sample survey designs
7. Advanced computational methods; Marko Chain Monte-Carlo Methods
8. Bayesian inference with unit and item nonresponse

# Lecture Schedule (subject to change):

| | | |
|---|---|---|
| 1. Jan 4 | Introduction 1 | |
| 2. Jan 9 | Introduction 2 | ISR 4036 |
| 3. Jan 11 | Maximum Likelihood for SRS | |
| Jan 16 | MLK Day no class | |
| 4. Jan 18 | Bayes for Simple Random Samples | |
| 5. Jan 23 | Bayesian Computation 1: Principles | |
| 6. Jan 25 | Bayesian Computation 2: Software | |
| 7. Jan 30 | Complex Survey Designs | |
| 8. Feb 1 | Models for Complex Survey Designs | |
| 9. Feb 6 | Bayes for Stratified Samples 1 | |
| 10. Feb 8 | Bayes for Stratified Samples 2 | |
| 11. Feb 13 | Bayes for PPS Samples | ISR 4036 |
| 12. Feb 15 | Role of Weights, Classical and Bayes | |
| 13. Feb 20 | Bayes for Clustered Data 1 | |
| 14. Feb 22 | Bayes for Clustered Data 2 | |
| | | |
| Feb 27, Mar 1 | Winter Break, no class | |

# Lecture Schedule (continued)

| | | |
|---|---|---|
| 15. Mar 6 | Midterm review | |
| 16. Mar 8 | Midterm exam | |
| 17. Mar 13 | Multistage Sampling | ISR 4036 |
| 18. Mar 15 | Missing Data 1 | |
| 19. Mar 20 | Missing Data 2 | |
| 20. Mar 22 | Missing Data 3 | |
| 21. Mar 27 | Measurement error as missing data | |
| 22. Mar 29 | Applications 1 | |
| 23. Apr 3 | Applications 2 | |
| 24. Apr 5 | Applications 3 | |
| 25. Apr 10 | Summary of Course material | ISR 4036 |
| 26. Apr 12 | Project presentations 1 | |
| 27. Apr 17 | Project presentations 2 | |
| 28. Apr 21 | Extra Class for Project presentations 3 (if needed) | |

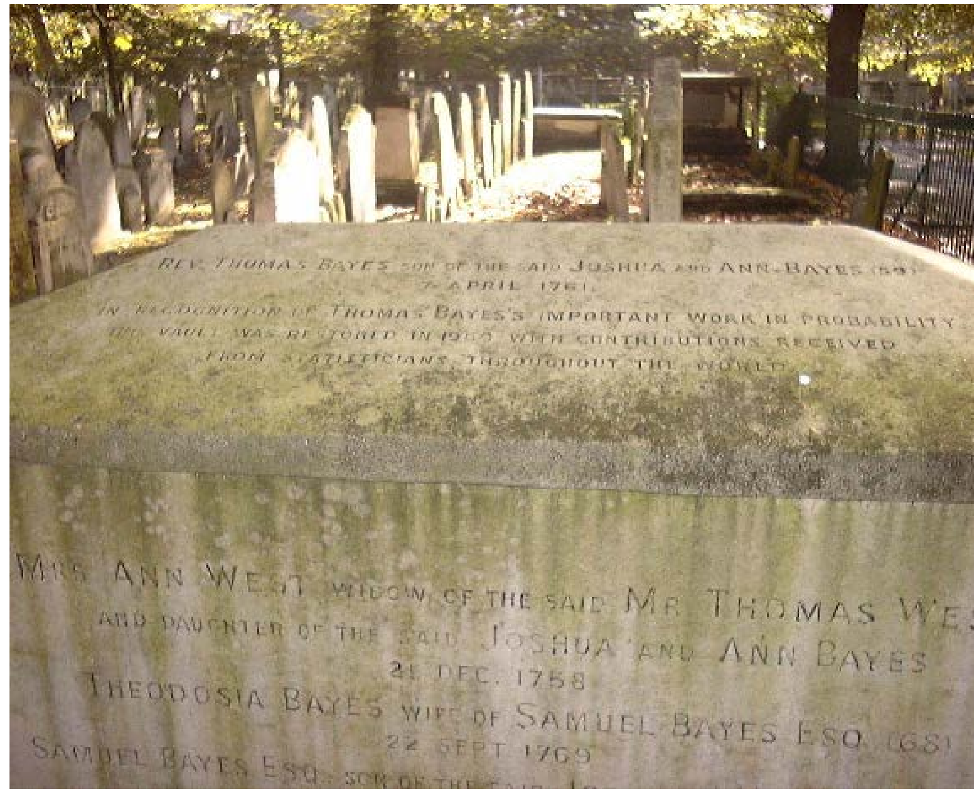
# Bayesian inference for sample surveys

# Approaches to Statistical Inference

- Classical (Randomization, Frequentist) Inference
  - parameters are treated as *fixed*
  - Inferences -- *P-Values, confidence intervals* -- are based on distribution of statistics in repeated sampling
- Bayesian Inference
  - Key elements contained in a posthumous essay by Rev. Thomas Bayes (1702?-1761)
  - Gaining popularity, particularly for complex statistical modeling
  - parameters are treated as *random*, and assigned a prior distribution to represent prior knowledge
  - Bayesian inference based on *posterior distribution* of parameters given data, computed via *Bayes Rule*

# Rev Thomas Bayes (1702?-1761)

Rev. T. Bayes



# Publications in his lifetime

1. Divine Benevolence, or an Attempt to Prove That the Principal End of the Divine Providence and Government is the Happiness of His Creatures (1731) (Theological Work).

2. An Introduction to the Doctrine of Fluxions, and a Defense of the Mathematicians Against the Objections of the Author of the Analyst (published anonymously in 1736) , Fellow of Royal Society based on this work in 1742 (speculative).

# Bayes’ famous posthumous essay

“I now send you an essay which I have found among the papers of our deceased friend Mr. Bayes, and which, in my opinion, has great merit, and deserves to be preserved”

Richard Price, read to the Royal Society Dec 23, 1763

# Bayes' rule

- Bayesian statistics is founded on Bayes' rule, which is a simple consequence of basic rules of probability:

  If $A$ and $B$ are two events, then the product rule is:

  $$\Pr(A,B) = \Pr(A) \times \Pr(B \mid A)$$

  Special case: $A$ and $B$ are independent if

  $$\Pr(A,B) = \Pr(A) \times \Pr(B)$$

- This rule can be applied to make inferences about (a) hypotheses, (b) parameters, or (c) predictions of nonsampled values

# Bayes' rule for hypotheses

- Applying the product rule with $A$ = hypothesis $H$, $B$ = data $D$:

$$\Pr(H, D) = \Pr(D) \times \Pr(H \mid D) = \Pr(H) \times \Pr(D \mid H)$$

$$\Pr(H \mid D) = \Pr(H) \times \Pr(D \mid H) / \Pr(D)$$

Hence, $$\frac{\Pr(H \mid D)}{\Pr(H' \mid D)} = \frac{\Pr(H)}{\Pr(H')} \times \frac{\Pr(D \mid H)}{\Pr(D \mid H')}$$

That is, posterior odds = prior odds $\times$ Bayes factor

# $\Pr(D|H)$ or $\Pr(H|D)$?

- Examples of *H*:
  - The existence of god, or life in other universes
  - Does environmental factor X cause disease Y?
  - Will Donald Trump win the election?
  - Will sales for product X meet some target value?
- Examples of *D*:
  - Astronomical measurements
  - Opinion polls
  - Market research data
  - National probability samples

# A simple application of Bayes: Screening Tests

- A friend is diagnosed by a screening test ($D$ = result of test, + or -) to have an extremely rare form of cancer ($H$ = has cancer). Only one out of a million people in his age group have the cancer.
- Naturally he is very upset as the test is pretty accurate:

  Sensitivity: Pr(+| has cancer)=0.99, implying

  Pr(-| has cancer)=0.01 (False negative)

  Specificity: P(-|no cancer)=0.999, implying

  Pr(+| no cancer)=0.001 (False positive)

# False Positive

- The probability that matters is the positive predictive value, which by Bayes Rule is

$$\Pr(\text{has cancer} \mid +) = \frac{\Pr(+ \mid \text{has cancer})\Pr(\text{has cancer})}{\Pr(+)}$$

$$= \frac{(0.99)(1/1000000)}{(0.99)(1/1000000) + (0.001)(999999/1000000)}$$

$$= 0.001 \quad (!)$$

# False Positive

Very likely, the friend does not have cancer.

# Screening Tests: output

Sensitivity: Pr(+| has cancer)=0.99,

Positive predictive value: Pr(has cancer|+)=0.001

Specificity: P(-|no cancer)=0.999,

Negative predictive value: Pr(no cancer|-)=0.999999

Sensitivity and specificity are properties of the test

PPV and NPV are what we really care about – though they require prior information about probability of disease (which may be hard to pin down)

# History of Bayes

- Much maligned in the last century, Bayesian statistics has since experienced a dramatic revival
- Excellent and fun book: “The theory that would not die” by Sharon McGrayne

# Statisticians Impacting Science: Bayesians in red

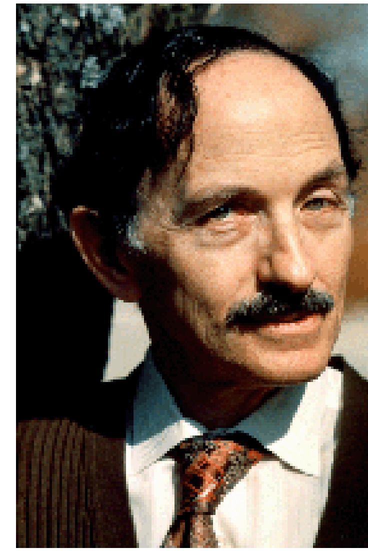
## Most-cited mathematicians in science **(Science Watch 02)**

**2** D. L. Donoho Stanford Stat;
**3** A.F.M. Smith London Stat
**4** E. A. Thompson Washington Biostat;
**5** I.M.Johnstone Stanford Stat
**6** J. Fan Hong Kong Stat;
**7** D.B. Rubin Harvard Stat.
**9** A. E. Raftery Washington Stat;
**10** A.E. Gelfand U. Conn Stat.
**11** S-W Guo Med. Coll. Wisc Biostat;
**12** S.L. Zeger JHU Biostat.
**13** P.J. Green Bristol Stat; **14** B.P. Carlin Minnesota Biostat
**15** J. S. Marron UNC Stat; **16** D.G. Clayton Cambridge Biostat
16 G.O. Roberts Lancaster Stat; 20. X-L Meng Chicago Stat
21. M. P. Wand Harvard Biostat; 22.W.R. Gilks MRC Biostat
**23** M. Chris Jones Open U Stat; 25.N. E. Breslow Washington Biostat

# Bayes Impacting Society

Alan Turing sculpture by Stephen Kettle, Bletchley Park. Photo by Jon Callas

Alan Turing and Jack Good’s Bayesian statistical methods helped decode German naval ciphers, arguably reducing the length of World War II by two years or more, saving millions of lives.



I. Jack Good (IJG)

# … and while on scientists who are subjects of current films…

"One can make the Anthropic Principle precise, by using **Bayes** statistics… One weights the a-priori probability (of a class of histories) … with the probability that the class of histories contain intelligent life… Stephen Hawking

# Bayes for comparing hypotheses

$$\Pr(H_1 \mid D) = \Pr(H_1) \times \Pr(D \mid H_1) / \Pr(D)$$

$$\Pr(H_2 \mid D) = \Pr(H_2) \times \Pr(D \mid H_2) / \Pr(D)$$

Hence $$\frac{\Pr(H_1 \mid D)}{\Pr(H_2 \mid D)} = \frac{\Pr(H_1)}{\Pr(H_2)} \times \frac{\Pr(D \mid H_1)}{\Pr(D \mid H_2)}$$

Posterior odds = prior odds x Bayes factor

"Prior odds are modified by the relative probability of observing the data under the two hypotheses"

Fun fact: according to Jack Good, Alan Turing coined the term "Bayes factor" (without the Bayes)

# Bayesian inference for population parameters

- Bayes rule also provides inferences for population quantities, like population means

$$\theta = \text{population parameters}$$

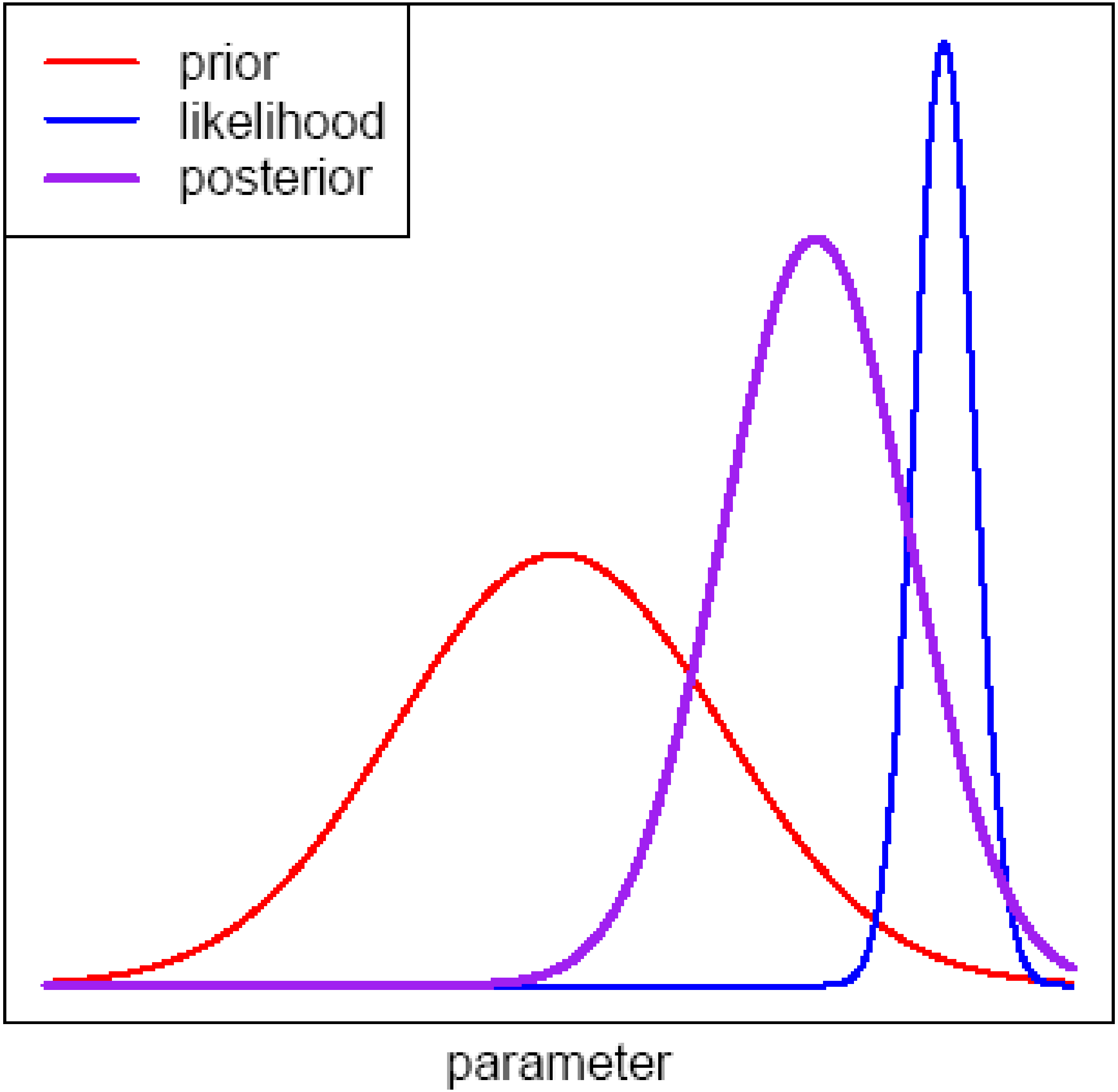
$$p(\theta \mid data) = \frac{\pi(\theta)\, p(data \mid \theta)}{p(data)}$$

Posterior distribution of $\theta$

Prior distribution of $\theta$

Likelihood $L(\theta \mid \text{data})$

Normalizing constant


prior
likelihood
posterior
parameter

# Bayesian inference for predictions

- Bayes rule also provides inferences for predictions of future values (or finite population quantities, which are functions of sampled and nonsampled values)

$$y^* = \text{ future value, } \theta = \text{population parameters}$$

$$p(y^* \mid \text{data}) = \int p(y^* \mid \theta)\, p(\theta \mid \text{data}) d\theta$$

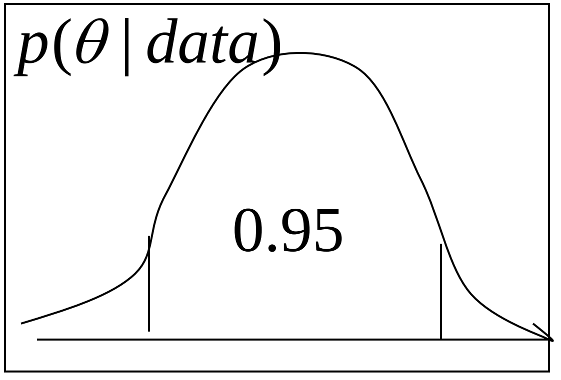
*Posterior predictive distribution*

- In a sense, all of statistics is about prediction, and quantifying the associated uncertainty
- This use of Bayes is a strong emphasis in this course

# Forms of Bayesian inference

- The posterior distribution summarizes information about an unknown quantity given the data
- Parameters can be *estimated* by the mean or median of the posterior distribution
- The spread of the posterior distribution indicates uncertainty, e.g. the posterior standard deviation
- A 95% <u>posterior probability or credibility </u>interval replaces the 95% confidence interval

  in frequentist statistics



# Prior distribution

- The novel feature of Bayesian statistics is the prior distribution, which formalizes prior knowledge about the parameters before the data are collected.
- Sometimes prior distributions are developed using information from previous studies (eg meta-analysis)
- Another approach is to use "noninformative" priors that correspond to limited prior information
  - There are standard "noninformative" or "reference" priors for common problems
- Subjective Bayesian approach "elicits" priors from experts

# Properties of Bayesian statistics

1. Bayes with noninformative reference priors and frequentist solutions agree on many standard problems.

Example: independent normal sample

$$y_1,\ldots,y_n \mid \theta \sim_{iid} N(\theta,\sigma^2)$$

Frequentist inference: The standard 95% confidence interval for $\theta$ is :

$$I = \overline{y} \pm t_{.975,n-1} s/\sqrt{n}, \text{ where}$$

$$\overline{y} = \text{sample mean, } s = \text{ sample sd}$$

We'll see that $I$ is also the 95% posterior credibility interval for $\theta$ ,with a particular choice of reference prior distribution

# Agreement of Bayes and classical inferences

- More generally, in large samples, Bayesian inference yields similar answers to <u>maximum likelihood</u> (ML) inference, a very important method in classical inference
- The underlying principle behind many standard analysis methods in epidemiology is ML – linear and logistic regression, survival analysis, etc.
- I'll review the basic ideas of maximum likelihood and how they relate to Bayes inference later

# Properties of Bayes

2. Bayesian inference is more direct, less mysterious than frequentist inference
   - confidence intervals have a tricky interpretation
   - hypothesis testing is worse:

     Frequentist: P-Value = Pr(data|H)

     P-value is not the probability that H is true!

     Bayes: Pr(H|data)

# Properties of Bayes

3. There are many problems that have no exact frequentist answers in small samples. For example
   (a) Comparing means of two independent samples with different means and variances. There is no exact frequentist solution, only approximate solutions (e.g. Welch approximation)
   (b) Inference for nonlinear models (e.g. logistic regression, Cox model) assume large samples.

   Bayesian methods yield answers to such problems

# Properties of Bayes

4. Bayesian inference allows for prior information to be incorporated in the analysis in a simple and clear way, via informative prior distributions.
   - In large-sample survey applications, we often use weak, relatively noninformative priors – most of the information is contained in the likelihood – this is sometimes called "objective Bayes"
   - However, informative priors can play a useful role in situations where some parameters of a model or not identified, or weakly identified
   - One example of this is for nonresponse that is missing not at random (as discussed in the material on nonresponse later)

# Properties of Bayes

5. Bayesian inferences under a carefully-chosen model should have good frequentist properties.
6. Theory says you are best to act (e.g. bet) like a Bayesian.
7. Modern computation tools make Bayesian analysis much more feasible than in the past.
   - Markov Chain Monte Carlo (MCMC) methods for computing posterior distributions

# Frequentist and Bayes methods have pluses and minuses...

| Frequentist | Bayes |
|---|---|
| No need for prior – limited specification | Needs detailed specification of prior and likelihood |
| Good repeated sampling properties | Bad model may have poor repeated sampling properties |
| Not prescriptive, can be ambigous | Prescriptive, unambigous |
| Not enough exact answers | Plenty of answers |
| Fails the likelihood principle | Satisfies the likelihood principle |

# The compromise: calibrated Bayes

| Activity | Bayes | Frequentist |
|---|---|---|
| Inference under assumed model | Strong | |
| Model formulation / assessment | | Strong |

Bayesian for inference

Frequentist for model assessment (enriched by Bayesian ideas)

Attempt to capitalize on strengths of both paradigms (Little, 2006 American Statistician)

# Bayes/frequentist compromises

"I believe that … sampling theory is needed for exploration and ultimate *criticism* of the entertained model in the light of the current data, while Bayes' theory is needed for *estimation* of parameters conditional on adequacy of the model."

George Box (1980)

# Calibrated Bayes

"… frequency calculations are useful for making Bayesian statements scientific, scientific in the sense of capable of being shown wrong by empirical test; here the technique is the calibration of Bayesian probabilities to the frequencies of actual events."

Don Rubin (1984 Annals of Statistics)

# Summary

- Bayes is flexible and principled
- Bayesian software is increasingly available
- Needs to specify prior distributions for parameters – approaches to be discussed
- For survey applications, emphasis is on predicting or "filling in" the nonsampled and nonresponding values – inference is based on the <u>posterior predictive distribution</u> of finite population quantities