Bayesian Statistics

Debdeep Pati
Florida State University

August 30, 2016

Historical Background

The Reverend Thomas Bayes, began the objective Bayesian theory, by solving a particular problem

- Suppose X is Binomial (n,p); an 'objective' belief would be that each value of X occurs equally often.
- The only prior distribution on p consistent with this is the uniform distribution.
- Along the way, he codified Bayes theorem.
- Alas, he died before the work was finally published in 1763.

REV. T. BAYES

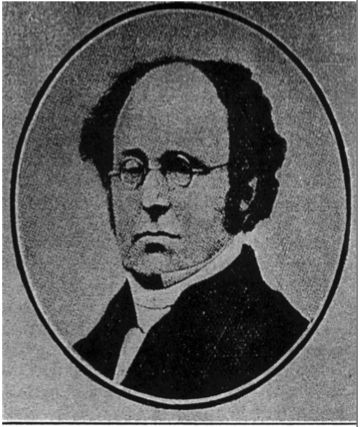
Historical Background

The real inventor of Objective Bayes was Simon Laplace (also a great mathematician, astronomer and civil servant) who wrote *Théorie Analytique des Probabilité* in 1812

- He established the 'central limit theorem' showing that, for large amounts of data, the posterior distribution is asymptotically normal (and the prior does not matter).
- He virtually always utilized a 'constant' prior density (reasons: CLT; parameter choice; robustness).
- He solved very many applications, especially in physical sciences.
- He had numerous methodological developments, e.g., a version of the Fisher exact test.
- Later in his life he invented frequentist statistics.

What's in a name, part I

- It was called *probability theory* until 1838.
- From 1838-1950, it was called *inverse probability*, apparently so named by Augustus de Morgan.
- From 1950 on it was called *Bayesian analysis* (as well as the other names); for why, see Fienberg (2006).



Augustus de Morgan

Brief History of Bayesian Statistics

1955 - : Emergence of Bayesian analysis, and development of Bayesian testing and model selection. $\downarrow$

Harold Jeffreys: fixed the logical flaw in inverse probability (objective Bayesian analysis)

$\downarrow$

Bruno de Finetti and others: developed the logically sound subjective Bayes school.

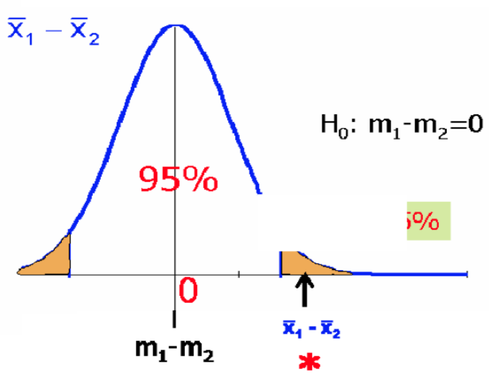
Motivating example

- Assess whether a selected population for has a higher growth rate than a control population.
- Classical statistics: the hypothesis to be tested is that there is no difference between the two treatments
- Before making the experiment, the error of rejecting this hypothesis when it is actually true is fixed at a level of 5%)
- Data ($\bar{x}_1 - \bar{x}_2$)
- true value of the difference between selected and control populations $(m_1 - m_2)$

Motivation

If our sample lies in the shadow area,

- There is no difference between treatments, our sample is a rare one
- The treatments are different, and repeating an infinite number of times the experiment, ($\bar{x}_1 - \bar{x}_2$) will not be distributed around zero but around an unknown value different from zero.



Bases of Bayesian inference

- Natural to find the most probable value of a parameter based on our data rather than to find which value of this parameter, if it would be the true value, would produce our data with a highest probability.
- To make probability statements based on our data we need some prior information and it is not clear how to introduce this prior information in our analysis or how to express lack of information using probability statements.
- Apply Bayes Theorem!

Bayes' Theorem

- A, B are 2 events

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

- Interested in assessing effect of a drug on growth rate of a rabbit population
- Selected group of rabbits and a control group in which growth rate has been measured.
- S : Effect of the drug the selected group, C: Effect of the control group, Interested in assessing $(S - C)$.
- Want to find the probabilities of all possible values of $(S - C)$ according to the information provided by our data.
- This can be expressed as $P(S - C \mid y)$

Components of a Bayesian machinery

- The posterior distribution

$$P(S-C \mid y) = \frac{P(y \mid S-C)P(S-C)}{P(y)}$$

- $P(y \mid S-C)$: distribution of the data for given value of the unknown, often known or assumed to be known from reasonable hypotheses.
- $P(S-C)$: Prior probability of the difference between selected and control group independent of data.
- $P(y)$: the probability of the sample.

Summarizing Bayesian inference: The general set-up

- General set up: $y_i \sim f(y \mid \theta), \theta \sim \Pi(\theta)$ &
- Obtain posterior distribution $\Pi(\theta \mid y_1, \ldots, y_n)$ as

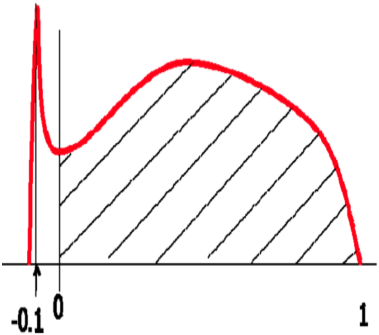
$$\Pi(\theta \mid y_1, \ldots, y_n) = \frac{\prod_{i=1}^{n} f(y_i \mid \theta)\Pi(\theta)}{\int_{\theta} \prod_{i=1}^{n} f(y_i \mid \theta)\Pi(\theta)}$$

Bayes estimates - measures of discrepancy

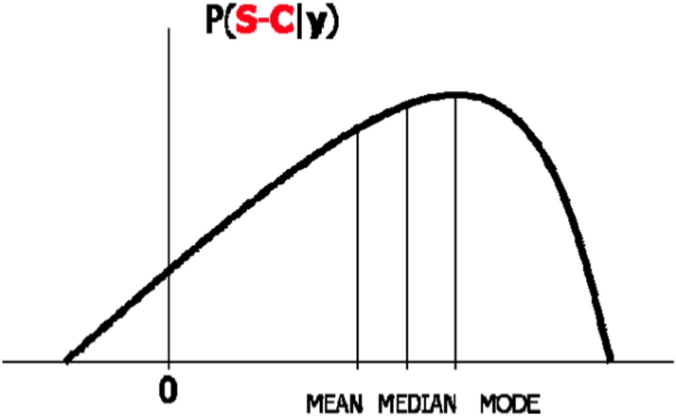
- Measure of Discrepancy - $R = E_{\theta,y_1,\ldots,y_n} L(\hat{\theta}(y_1, \ldots, y_n), \theta)$
  1. Posterior mean: minimizes R with squared error L
  2. Posterior median: minimizes R with L as absolute deviation
  3. Posterior mode minimizes R with L as the $0 - 1$ loss.

Loss functions

- Mean: 2-fold inconvenience: penalizes high errors, this risk function is not invariant to transformations
- Mode: signifies the most probable value, easier to calculate in the pre-MCMC era - may not be representative
- Median: true value has a 50% of probability of being higher or lower than the median. Attractive loss - invariant to one-to-one transformations



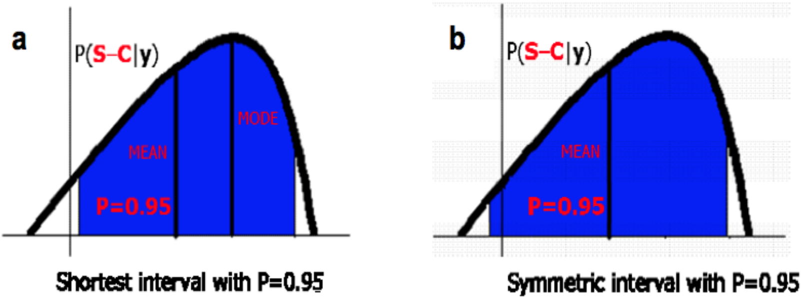
Bayes estimates



Precision of estimation - credible intervals

- Confidence interval: How often the interval contains the parameter if the samples are generated according to the truth
- Bayesian credible intervals: Given the data, want to construct interval that encloses 95% of the posterior probability of the parameter
- $P(\theta \in [L(y), U(y)] \mid y) = 0.95$
- Can find the shortest interval with a 95% credible interval (called the Highest posterior density interval at 95%).

Highest posterior density credible interval

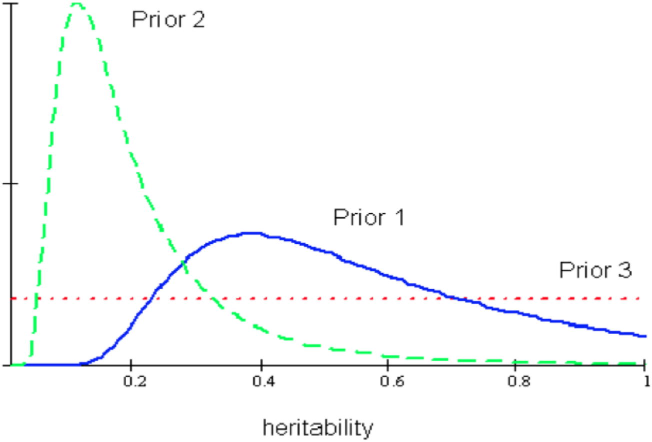


Example 1: Binomial-Beta model

- $X \sim \text{Bin}(n, p), p \sim \text{Beta}(a, b)$
- $p \mid X \sim \text{Beta}(a + X, n - X + b)$

How to choose a and b, Vague / No prior information

- Toss a coin n times - want to estimate the probability of head
- Three states of beliefs were tested
- Prior 3 called objective or non-informative by Bayesian statisticians



Posterior mean for the Beta-Binomial problem

- Posterior mean:

$$
\begin{aligned}
E(p \mid X) &= \frac{a+X}{a+b+n} \\
&= \frac{X}{n}\frac{n}{\alpha+\beta+n} + \frac{a}{a+b}\frac{a+b}{a+b+n}
\end{aligned}
$$

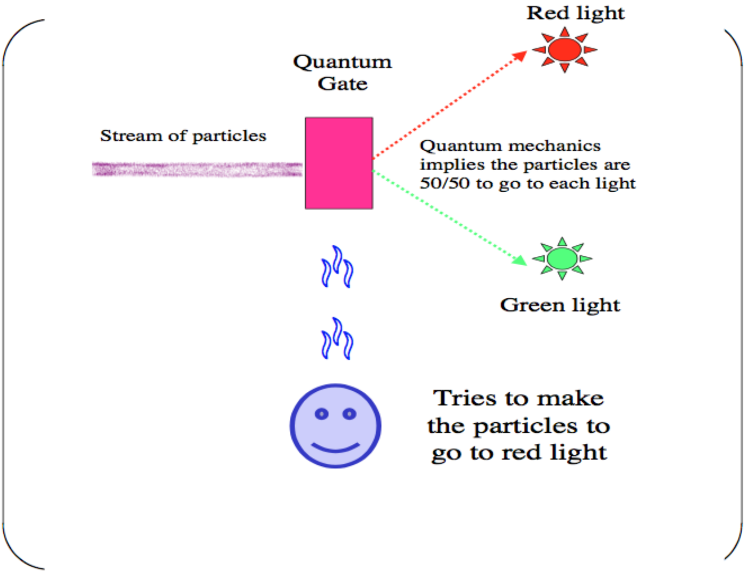
Posterior mean for the Normal-Normal model

- $Y_1, \ldots, Y_n \sim N(\mu, 1), \mu \sim N(\mu_0, 1)$
- $\mu \mid Y_1, \ldots, Y_n \sim N\left(\frac{n\bar{X}+\mu_0}{n+1}, \frac{1}{n+1}\right)$
- Posterior mean for $\mu = \frac{n\bar{X}+\mu_0}{n+1}$.
- Posterior mean $=$ MLE $* \frac{\text{(precision of MLE)}}{\text{(precision of MLE + prior precision)}} +$ prior mean $* \frac{\text{(prior precision)}}{\text{(precision of MLE + prior precision)}}$
- Conjugate prior: Posterior distribution has the same family as the prior distribution

Bayesian testing: Some Anomalies

- Psychokinesis Example
- Does a subject possess Psychokinesis?
- Schmidt, Jahn and Radin (1987) used electronic and quantum-mechanical random event generators with visual feedback; the subject with alleged psychokinetic ability tries to "influence" the generator.

Some Anomalies: Psychokinesis Example



Data and model

- Each "particle" is a Bernoulli trial (red = 1, green = 0), $\theta =$ probability of "1"
- $n = 104,490,000$ trials
- $X = \#$ successes
- $X \sim Binomial(n, \theta)$, $x = 52,263,470$ is the actual observation.
- To test $H_0 : \theta = 1/2$ (subject has no influence) versus $H_1 : \theta \neq 1/2$ (subject has influence)
- P-value $= P_{\theta=1/2}(|X - n| > |x - n|) \approx 0.0003$.
- Is there strong evidence against H_0 (i.e., strong evidence that the subject influences the particles) ?

Bayesian Analysis: (Jefferys, 1990)

- Prior distribution for hypothesis: $P(H_i) =$ prior probability that H_i is true, $i = 0, 1$;
- On $H_1 : \theta \neq 1$, let $\pi(\theta)$ be the prior density for θ.
- Subjective Bayes: choose the $P(H_i)$ and $\pi(\theta)$ based on personal beliefs.
- Objective (or default) Bayes: choose $P(H_0) = P(H_1) = 0.5$, $\pi(\theta) = 1$ on $0 < \theta < 1$.

Posterior probability of hypotheses:

$$
\begin{aligned}
P(H_0 \mid x) &= \text{probability that } H_0 \text{ true, given data, } x \\
&= \frac{f(x \mid \theta = 1/2)P(H_0)}{f(x \mid \theta = 1/2)P(H_0) + Pr(H_1) \int f(x \mid \theta)\pi(\theta)d\theta}.
\end{aligned}
$$

For the objective prior,

$$P(H_0 \mid x = 52,263,470) \approx 0.92$$

(recall, p-value $\approx .0003$) Posterior density on $H_1 : \theta \neq 1/2$ is

$$\pi(\theta \mid x, H_1) \propto \pi(\theta) f(x \mid \theta) \propto 1 \times \theta^x (1-\theta)^{n-x},$$

which is $\mathrm{Be}(\theta \mid 52263470, 52226530)$.

Bayes factor

Bayes Factor: An objective alternative to choosing $P(H_0) = P(H_1) = 1$ is to report the

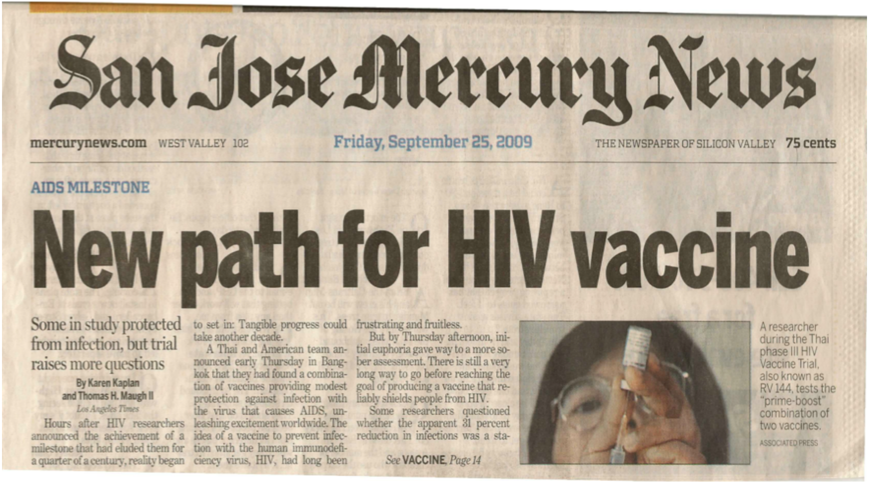
$$\begin{aligned}
\text{Bayes factor} &= \frac{\text{likelihood of observed data under } H_0}{\text{`average' likelihood of observed data under } H_1} \\
&= \frac{f(x \mid \theta = 1/2)}{\int_0^1 f(x \mid \theta)\pi(\theta)d\theta} \approx 12. \\
\frac{P(H_0 \mid x)}{P(H_1 \mid x)} &= \frac{P(H_0)}{P(H_1)} \times BF_{01} \\
\text{Posterior odds} &= \text{Prior odds} \times \text{Bayes factor}(BF_{01})
\end{aligned}$$

so BF_{01} is often thought of as "the odds of H_0 to H_1 provided by the data".

Clash between p-values and Bayes answers

- In the example, p-value is $.0003$, but
- The objective posterior probability of the null is 0.92; (Bayes factor gives 12 to 1 odds in favor of the null).
- Was the prior on θ inappropriate? (Few would say it was unfair to give H_0 prior probability of $1/2$.)
- But it was a neutral, objective prior.
- Any prior produces Bayes factors orders of magnitude larger than the p-value.

Another Example: A New(?) path for HIV vaccine



San Jose Mercury News

mercurynews.com WEST VALLEY 102 Friday, September 25, 2009 THE NEWSPAPER OF SILICON VALLEY 75 cents

AIDS MILESTONE

New path for HIV vaccine

Some in study protected from infection, but trial raises more questions

By Karen Kaplan and Thomas H. Maugh II
Los Angeles Times

Hours after HIV researchers announced the achievement of a milestone that had eluded them for a quarter of a century, reality began to set in: Tangible progress could take another decade.

A Thai and American team announced early Thursday in Bangkok that they had found a combination of vaccines providing modest protection against infection with the virus that causes AIDS, unleashing excitement worldwide. The idea of a vaccine to prevent infection with the human immunodeficiency virus, HIV, had long been frustrating and fruitless.

But by Thursday afternoon, initial euphoria gave way to a more sober assessment. There is still a very long way to go before reaching the goal of producing a vaccine that reliably shields people from HIV.

Some researchers questioned whether the apparent 31 percent reduction in infections was a sta-

See VACCINE, Page 14

A researcher during the Thai phase III HIV Vaccine Trial, also known as RV 144, tests the "prime-boost" combination of two vaccines.

ASSOCIATED PRESS

Hypotheses and Data

- Alvac had shown no effect, Aidsvax had shown no effect
- Question: Would Alvac as a primer and Aidsvax as a booster work?
- The Study: Conducted in Thailand with $16,395$ individuals from the general (not high-risk) population:
  - 71 HIV cases reported in the 8198 individuals receiving placebos
  - 51 HIV cases reported in the 8197 individuals receiving the treatment

The test that was likely performed:

- Let p_1 and p_2 denote the probability of HIV in the placebo and treatment populations, respectively.
- Test $H_0 : p_1 = p_2$ versus $H_1 : p_1 > p_2$ (vaccines were not live, so $p_1 < p_2$ can probably be ignored)
- Normal approximation okay, so

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sigma_{\hat{p}_1 - \hat{p}_2}} = \frac{0.00866 - 0.00622}{0.00134} = 1.82.$$

  is approximately $N(\theta, 1)$, where $\theta = (p_1 - p_2)/(0.00134)$.
- We thus test $H_0 : \theta = 0$ versus $H_1 : \theta > 0$, based on z.
- Observed $z = 1.82$, so the (one-sided) p-value is 0.034.

Bayesian analysis

- Prior distribution: $P(H_i) =$ prior probability that H_i is true, $i = 0, 1$.
- On $H_1 : \theta > 0$, let $\pi(\theta)$ be the prior density for θ.
- Objective (or default) Bayes: choose $P(H_0) = P(H_1) = 1/2$
- $\pi(\theta) = \text{Uniform}(0, 6.46)$, which arises from assigning: uniform for p_2 on $0 < p_2 < p_1$, and plug in for p_1.
- For the objective prior, $P(H_0 \mid z = 1.82) \approx 0.336$, whereas one-sided p-value is 0.034.

Bayesian analysis

- Robust Bayesian theory suggests a general and simple way to calibrate p-values. (Sellke, Bayarri and Berger, 2001 Am. Stat.; Sellke (2012)).
- Theorem: If $p < e^{-1}$, then $B_{01} \geq -ep\log p$ (or $B_{10} \leq 1/\{-ep\log(p)\}$.

p	.2	.1	.05	.01	.005	.001	.0001	.00001
$-ep\log(p)$	.879	.629	.409	.123	.072	.0189	.0025	.00031