# Bayesian Statistics 

Debdeep Pati<br>Florida State University

April 9, 2013

## Historical Background

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

- Suppose X is Binomial ( $\mathrm{n}, \mathrm{p}$ ); 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.



## Historical Background

## 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

## Other opinions

- Stigler (1983) attributes it to Saunderson (1683-1739), a blind professor of Optics
- The first deduction of the least square method made by Gauss (1795) using Bayesian methods


## Motivating example

- Assess whether a selected population for growth rate 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 $5 \%$
- Repeating the experiment an infinite number of times difference between the averages of these samples ( $\bar{x}_{1}-\bar{x}_{2}$ )
- true value of the difference between selected and control populations $\left(m_{1}-m_{2}\right)$


## 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, ( $\left.\bar{x}_{1}-\bar{x}_{2}\right)$ 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.
- Information about the parameters we want to estimate that exists before we perform our experiment.
- It is almost impossible to do this formally, with some exceptions. We will distinguish three scenarios:

1. exact prior information
2. vague prior information
3. No prior information - describing ignorance

## Exact prior information



## 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



## Summarizing Bayesian inference: The general set-up

- General set up: $y_{i} \sim f(y \mid \theta), \theta \sim \Pi(\theta) \&$
- Obtain posterior distribution $\Pi\left(\theta \mid y_{1}, \ldots, y_{n}\right)$ as

$$
\Pi\left(\theta \mid y_{1}, \ldots, y_{n}\right)=\frac{\prod_{i=1}^{n} f\left(y_{i} \mid \theta\right) \Pi(\theta)}{\int_{\theta} \prod_{i=1}^{n} f\left(y_{i} \mid \theta\right) \Pi(\theta)}
$$

## Galton's 1877 machine

## I877 Algorithm: Normal Prior-Posterior



## Example: Binomial-Beta model

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


## Bayes estimates - measures of discrepancy

- Measure of Discrepancy - $R=E_{\theta, y_{1}, \ldots, y_{n}} L\left(\hat{\theta}\left(y_{1}, \ldots, y_{n}\right), \theta\right)$

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.

## Bayes estimates



## 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



## Posterior mean for the Beta-Binomial problem

- Posterior mean:

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

- MLE $* \frac{\text { (precision of MLE) }}{\text { (precision of MLE }+ \text { prior precision) }}+$ prior mean $*$ (prior precision)
$\overline{\text { (precision of MLE }+ \text { prior precision) }}$


## Precision of estimation - credible intervals

- Confidence interval: How often the interval contains the true value if the samples are generated according to the truth
- Credible intervals: Given the data, how often does the interval contains the true parameter
- $P(\theta \in[L(y), U(y)] \mid y)=0.95$
- Can find the shortest interval with a $95 \%$ probability of containing the true value (what is called the Highest posterior density interval at 95\%).


## Highest posterior density credible interval



Shortest interval with $\mathrm{P}=0.95$


Symmetric interval with $\mathrm{P}=0.95$

