Bayesian Statistics

Debdeep Pati
Florida State University

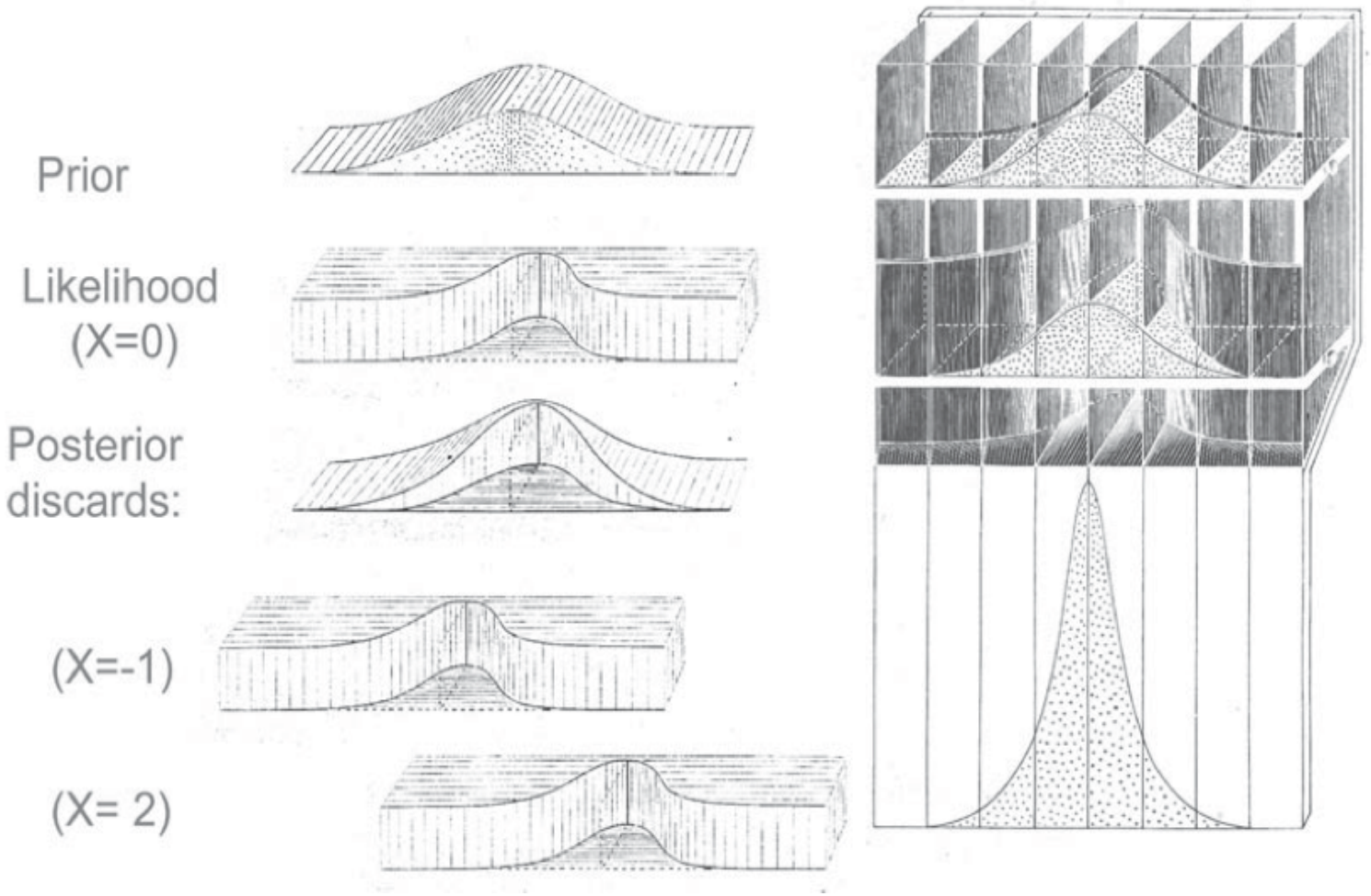
April 9, 2013

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

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

1877 Algorithm: Normal Prior-Posterior



Galton's 1877 machine

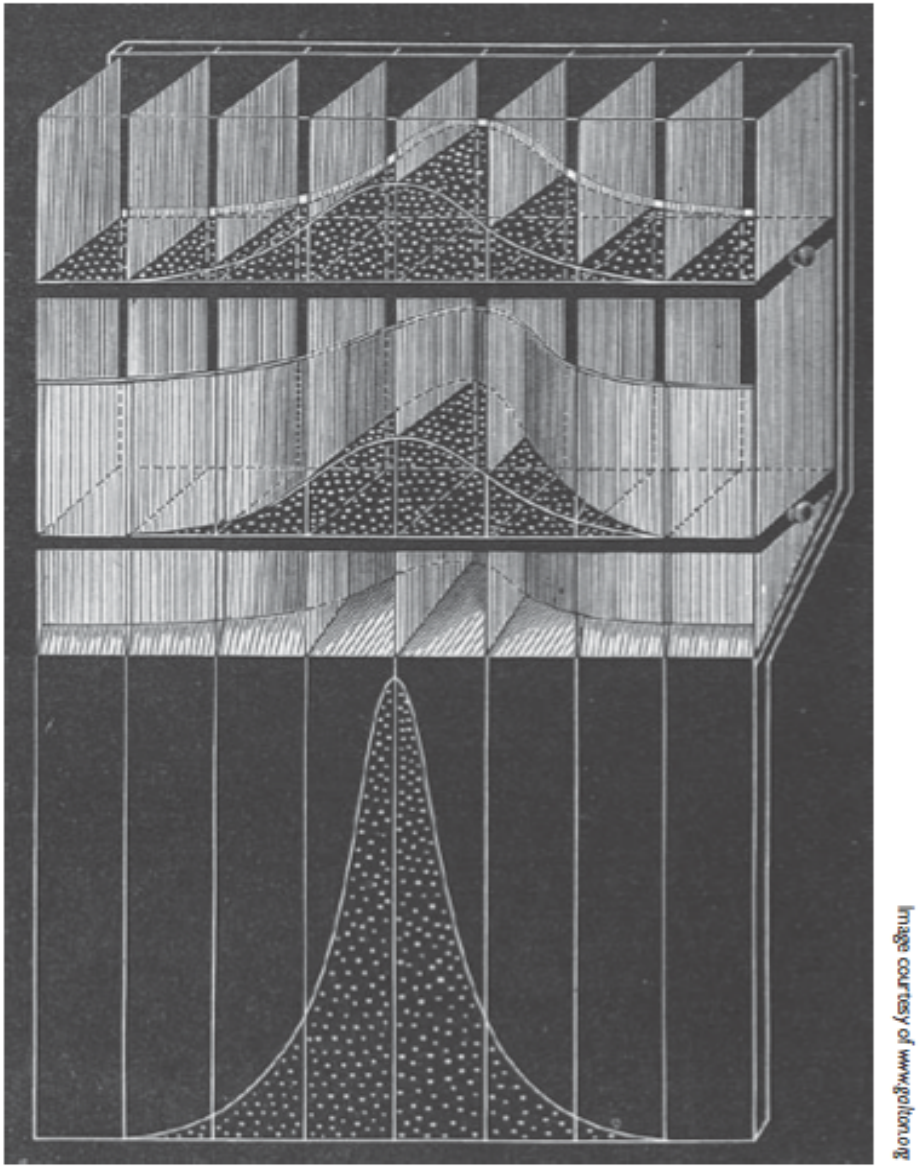
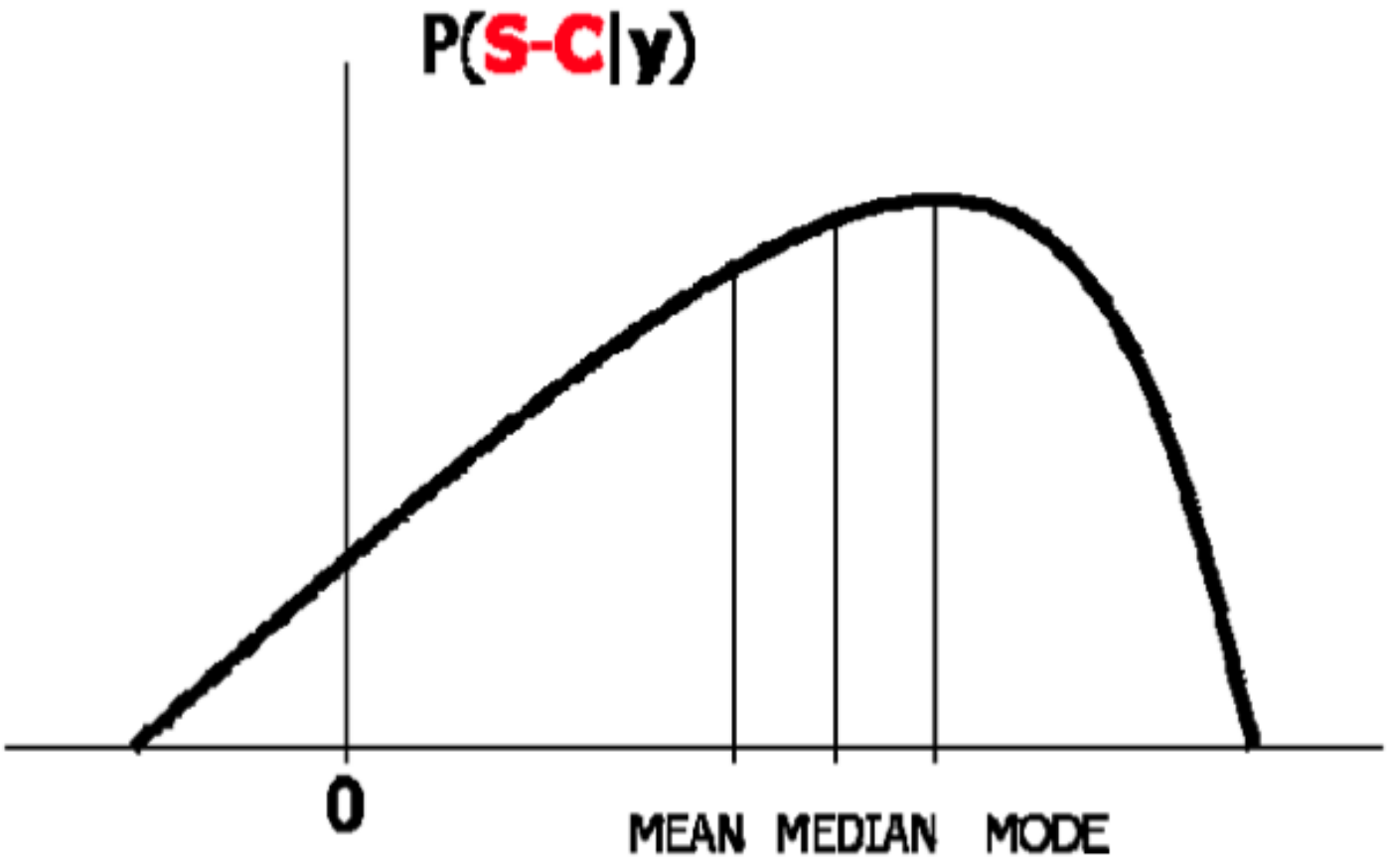


Figure 1. Galton's 1877 machine

Bayes estimates - measures of discrepancy

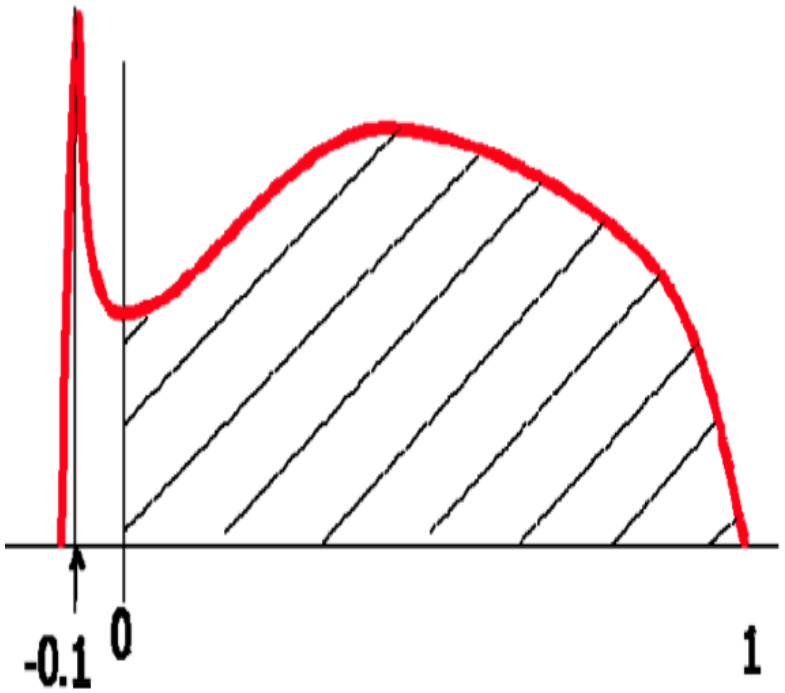
- Measure of Discrepancy - $R = E_{\theta, y_1, \ldots, y_n} L(\hat{\theta}(y_1, \ldots, y_n), \theta)$
  1. Posterior mean: minimizes R with squared error L
  2. Posterior median: minimizes R with L as absolute deviation
  3. Posterior mode minimizes R with L as the $0 - 1$ loss.

Bayes estimates

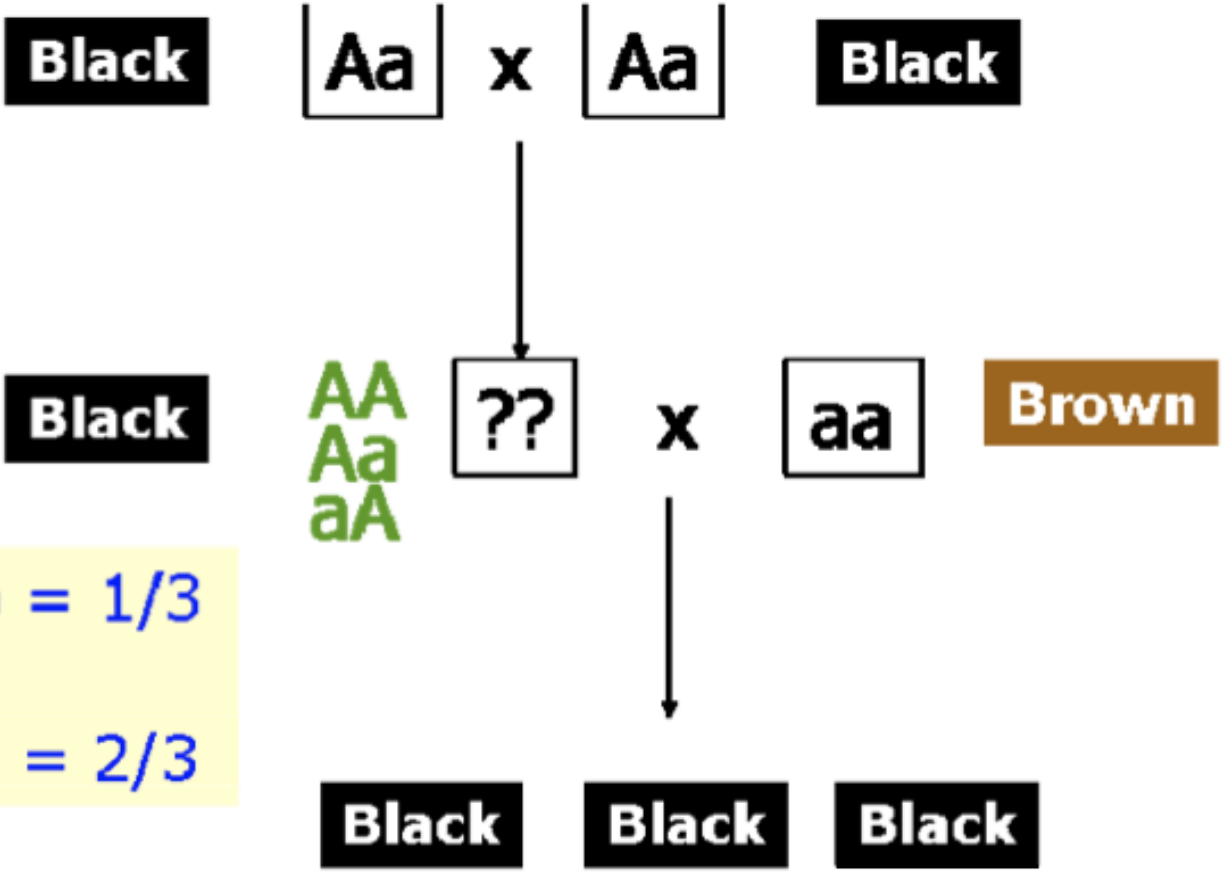


Loss functions

- 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



The Mouse example:



P(AA) = 1/3

P(Aa) = 2/3

Solution: The Bayes Theorem:

- $P(AA) = 1/3$
- By Bayes' Theorem

$$P(AA \mid y = 3) = \frac{P(y = 3 \mid AA)P(AA)}{P(y = 3)}$$

Solution: The Bayes Theorem:

- $P(y = 3 \mid AA) = 1$
- $P(y = 3) = P(y = 3 \mid AA)P(AA) + P(y = 3 \mid Aa)P(Aa)$
- $P(Aa) = 2/3$ and $P(y = 3 \mid Aa) = (1/2)^3 = 1/8$
- Plugging in Bayes' Theorem, $P(AA \mid y = 3) = 0.80$.
- $P(Aa \mid y = 3) = 0.20$.

Example: Binomial-Beta model

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

Posterior mean for the Beta-Binomial model

- Posterior mean:

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

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

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

The linear model

- $Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \epsilon_{n\times 1}$
- $\epsilon \sim \mathrm{N}(0, \sigma^2)$.
- What are the unknowns? (β and σ)
- Usually $\beta \sim N(\beta_0, \Sigma_0)$ and $\sigma^{-2} \sim \mathrm{IG}(a, b)$
- $\beta \mid \sigma^2, Y$ is a Normal distribution and $\sigma^2 \mid \beta, Y$ is an IG distribution.
- But we want $\beta, \sigma^2 \mid Y$

Gibbs sampling

σ_0^2 arbitrary

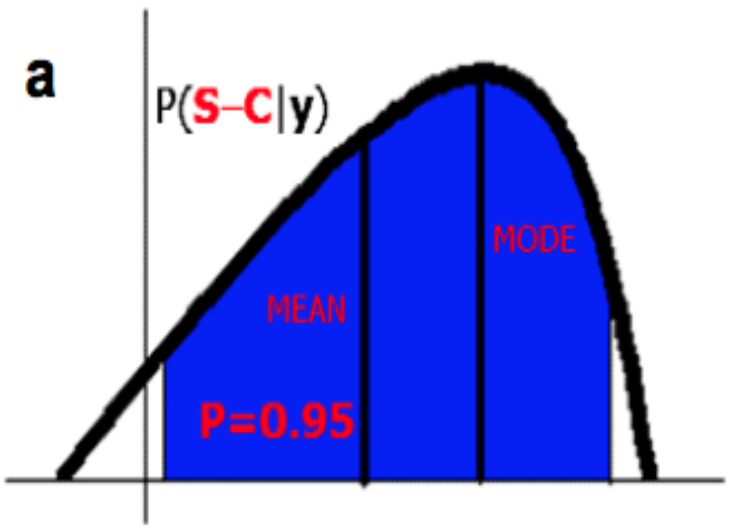
$f(\mathbf{b} \mid \sigma^2 = \sigma_0^2, \mathbf{y})$

$f(\sigma^2 \mid \mathbf{b} = \mathbf{b}_0, \mathbf{y})$

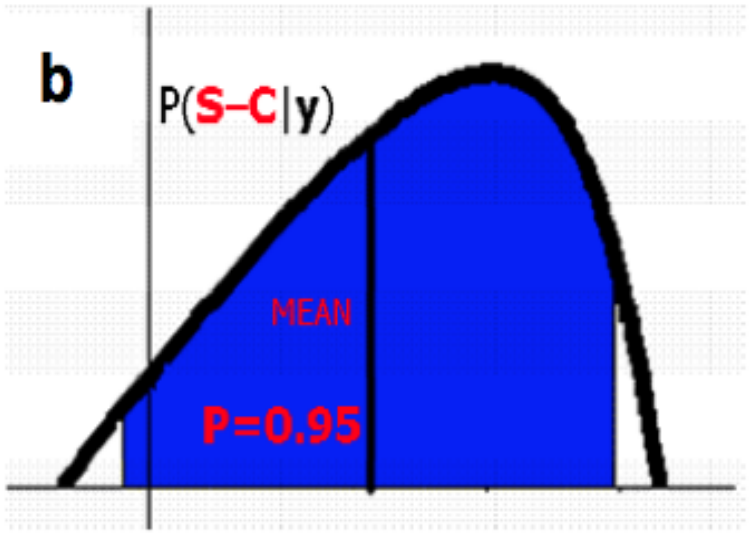
Precision of estimation - credible intervals

- 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



Shortest interval with P=0.95



Symmetric interval with P=0.95