Probability Theory and Statistics

Lecture 8

November 12, 2013

Robert Dahl Jacobsen

robert@math.aau.dk

Department of Mathematical Sciences
Aalborg University

AALBORG UNIVERSITY
DENMARK

Agenda

Bayesian statistics

Frequentists vs Bayesian statistics

- Frequentists approach:
  - Parameters are fixed
  - Data is random
  - Binomial example:

$$P(\text{success}) = \frac{\#\text{ successes}}{\#\text{ trials}} \quad \text{as trials grow}$$

- Bayesian statistics:
  - Parameters are random
  - Data is fixed
  - `http://xkcd.com/1132`

Examples

- The probability that it will rain tomorrow is 0.3
- Binomial experiment:

  $$X \sim b(n, \theta)$$

  What is $P(\theta > 0.7)$?

Bayesian inference

- Bayes formula:

$$P(\text{parameter} \mid \text{data}) = \frac{P(\text{data} \mid \text{parameter})P(\text{parameter})}{P(\text{data})}$$
$$\propto P(\text{data} \mid \text{parameter})P(\text{parameter})$$

- Bayesian inference:

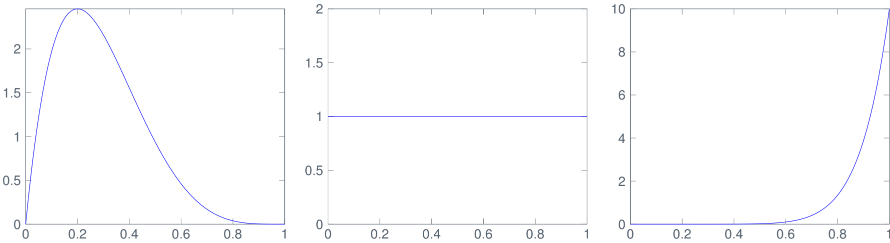
$$\text{posterior distribution} \propto \text{likelihood} \times \text{prior}$$

- Note:

$$P(\text{data}) = \int \cdots \int P(\text{data} \mid \text{parameter})P(\text{parameter})d\text{parameter}$$

Priors

- Preliminary knowledge about parameters.
- Example: Inference in binomial experiment.
    - Parameter: Probability of success $= \theta$.
    - Prior distribution for θ:



Likelihood

- Binomial experiment with observations $x_1, \ldots, x_n \in \{0, 1\}$:

$$P(x \mid \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

- Maximum likelihood estimate:

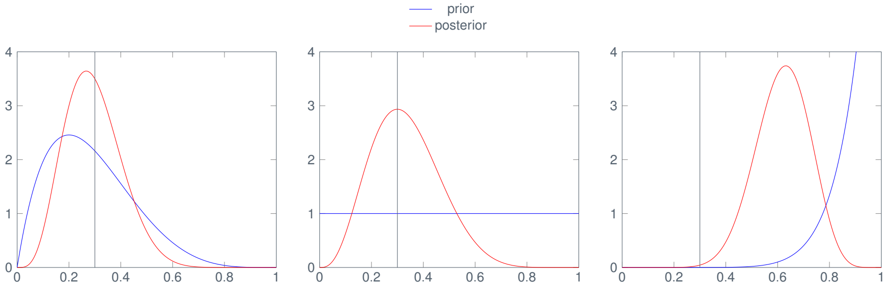
$$\widehat{\theta} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Posterior

- Binomial experiment:
  - Observations: $n = 10$
  - Successes: $= 3$
  - MLE:

$$\widehat{\theta} = \frac{3}{10}$$

- Posterior distributions:

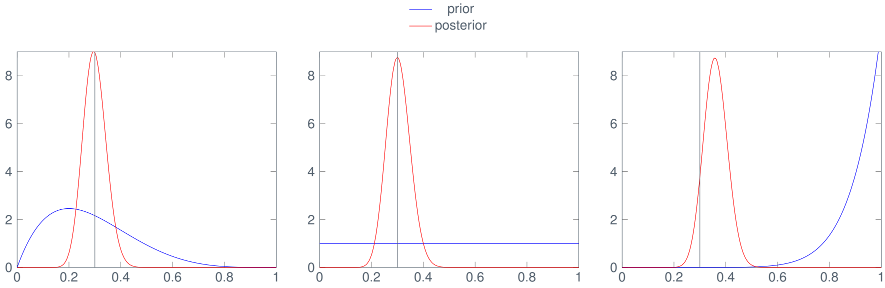


Posterior

- Binomial experiment:
  - Observations: $n = 100$
  - Successes: $= 30$
  - MLE:

$$\widehat{\theta} = \frac{3}{10}$$

- Posterior distributions:



Repeated experiments

- Sequentially observed independent data:

$$\overline{x}_1 = (x_{1,1}, \ldots, x_{1,n_1})$$
$$\overline{x}_2 = (x_{2,1}, \ldots, x_{2,n_2})$$

- Posterior distribution after 1st data set:

$$P(\theta \,|\, \overline{x}_1) \propto P(\overline{x}_1 \,|\, \theta) P(\theta)$$

- Posterior distribution after 2nd data set:

$$P(\theta \,|\, \overline{x}_1, \overline{x}_2) \propto P(\overline{x}_2 \,|\, \theta) P(\theta \,|\, \overline{x}_1)$$

Note: New prior is old posterior.

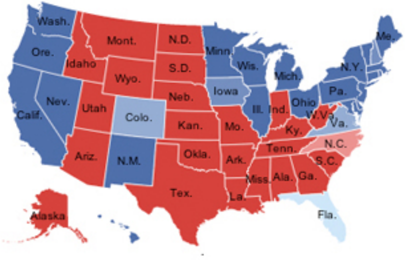
Pros and Cons

Pros:

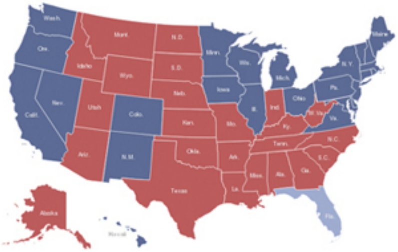
- Intuitive: Data is fixed.
- Control of uncertainty.
- Nate Silver:

Cons:

- How to specify priors?
- Computationally intensive



Nate Silver's Map



The Actual Map