Light Introduction to Bayesian Statistics

Christine P. Chai

Department of Statistical Science
Duke University

April 5, 2016

Bayesian Statistics

- Conditional probability and the Bayes' theorem
  - A, B are events in the sample space $\mathbb{S}$
  - $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(A)P(B|A)}{P(B)} = \dfrac{P(A)P(B|A)}{\sum_{A \in \mathbb{S}} P(A)P(B|A)}$

- Bayesian method to estimate the unknown parameter θ
  - Prior: $p(\theta)$ reflects our belief **before** looking at the data
  - Likelihood: $p(y|\theta)$ is the probability of data given θ
  - Posterior: $p(\theta|y)$ is the updated belief **after** looking at the data
  - $p(\theta|y) = \dfrac{p(\theta)p(y|\theta)}{p(y)} = \dfrac{p(\theta)p(y|\theta)}{\int p(\theta)p(y|\theta)d\theta} \propto p(\theta)p(y|\theta)$
  - posterior $\propto$ prior $\times$ likelihood

Bayesian Statistics

- Conditional probability and the Bayes' theorem
  - A, B are events in the sample space $\mathbb{S}$
  - $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B|A)}{P(B)} = \frac{P(A)P(B|A)}{\sum_{A \in \mathbb{S}} P(A)P(B|A)}$
- Bayesian method to estimate the unknown parameter θ
  - Prior: $p(\theta)$ reflects our belief **before** looking at the data
  - Likelihood: $p(y|\theta)$ is the probability of data given θ
  - Posterior: $p(\theta|y)$ is the updated belief **after** looking at the data
  - $p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)} = \frac{p(\theta)p(y|\theta)}{\int p(\theta)p(y|\theta)d\theta} \propto p(\theta)p(y|\theta)$
  - posterior $\propto$ prior $\times$ likelihood

Bayesian Statistics

- Conditional probability and the Bayes' theorem
  - A, B are events in the sample space $\mathbb{S}$
  - $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(A)P(B|A)}{P(B)} = \dfrac{P(A)P(B|A)}{\sum_{A \in \mathbb{S}} P(A)P(B|A)}$
- Bayesian method to estimate the unknown parameter θ
  - Prior: $p(\theta)$ reflects our belief **before** looking at the data
  - Likelihood: $p(y|\theta)$ is the probability of data given θ
  - Posterior: $p(\theta|y)$ is the updated belief **after** looking at the data
  - $p(\theta|y) = \dfrac{p(\theta)p(y|\theta)}{p(y)} = \dfrac{p(\theta)p(y|\theta)}{\int p(\theta)p(y|\theta)d\theta} \propto p(\theta)p(y|\theta)$
  - posterior $\propto$ prior $\times$ likelihood

Bayesian Example

- A deck of 10 poker cards: red (success) and black (failure)
- Fact: The success probability θ is either 80% or 20%.
- Goal: Estimate the success probability θ

- Prior: $P(\theta \text{ is } 0.8) = P(\theta \text{ is } 0.2) = 0.5$
- Likelihood: Draw samples from the data **with** replacement

- Data collection comes at a cost
- How many datapoints are needed to make a decision?
- You can make the decision once and only once

Bayesian Example

- A deck of 10 poker cards: red (success) and black (failure)
- Fact: The success probability θ is either 80% or 20%.
- Goal: Estimate the success probability θ

- Prior: $P(\theta \text{ is } 0.8) = P(\theta \text{ is } 0.2) = 0.5$
- Likelihood: Draw samples from the data **with** replacement

- Data collection comes at a cost
- How many datapoints are needed to make a decision?
- You can make the decision once and only once

Bayesian Example

- A deck of 10 poker cards: red (success) and black (failure)
- Fact: The success probability θ is either 80% or 20%.
- Goal: Estimate the success probability θ

- Prior: $P(\theta \text{ is } 0.8) = P(\theta \text{ is } 0.2) = 0.5$
- Likelihood: Draw samples from the data **with** replacement

- Data collection comes at a cost
- How many datapoints are needed to make a decision?
- You can make the decision once and only once

Bayesian Example – Solution

- Prior: $P(\theta \text{ is } 0.8) = 0.5$
- Likelihood: Data contains r red (success) and b black (failure)
  - Data follows a binomial distribution $Bin(r+b, \theta)$
  - $x \sim Bin(n,p) : f(x) = \binom{n}{x} p^x (1-p)^{n-x}$
- Posterior: $P(\theta \text{ is } 0.8 | \text{ data })$

$$= \frac{P(\theta \text{ is } 0.8) P(\text{ data } | \theta = 0.8)}{P(\theta \text{ is } 0.8) P(\text{ data } | \theta = 0.8) + P(\theta \text{ is } 0.2) P(\text{ data } | \theta = 0.2)}$$

$$= \frac{0.8^r (1-0.8)^b}{0.8^r (1-0.8)^b + 0.2^r (1-0.2)^b}$$

- The higher success rate we have seen, the more likely that θ is 0.8

Bayesian Example – Solution

- Prior: $P(\theta \text{ is } 0.8) = 0.5$
- Likelihood: Data contains r red (success) and b black (failure)
  - Data follows a binomial distribution $Bin(r+b, \theta)$
  - $x \sim Bin(n,p) : f(x) = \binom{n}{x} p^x (1-p)^{n-x}$
- Posterior: $P(\theta \text{ is } 0.8 | \text{ data })$

$$= \frac{P(\theta \text{ is } 0.8)P(\text{ data } | \theta = 0.8)}{P(\theta \text{ is } 0.8)P(\text{ data } | \theta = 0.8) + P(\theta \text{ is } 0.2)P(\text{ data } | \theta = 0.2)}$$

$$= \frac{0.8^r(1-0.8)^b}{0.8^r(1-0.8)^b + 0.2^r(1-0.2)^b}$$

- The higher success rate we have seen, the more likely that θ is 0.8

Bayesian Example – Solution

- Prior: $P(\theta \text{ is } 0.8) = 0.5$
- Likelihood: Data contains r red (success) and b black (failure)
  - Data follows a binomial distribution $Bin(r+b, \theta)$
  - $x \sim Bin(n,p) : f(x) = \binom{n}{x} p^x (1-p)^{n-x}$
- Posterior: $P(\theta \text{ is } 0.8 | \text{ data })$

$$= \frac{P(\theta \text{ is } 0.8) P(\text{ data } | \theta = 0.8)}{P(\theta \text{ is } 0.8) P(\text{ data } | \theta = 0.8) + P(\theta \text{ is } 0.2) P(\text{ data } | \theta = 0.2)}$$

$$= \frac{0.8^r (1-0.8)^b}{0.8^r (1-0.8)^b + 0.2^r (1-0.2)^b}$$

- The higher success rate we have seen, the more likely that θ is 0.8

Next Class

- Bayesian inference for continuous random variables
- Conjugate priors: same distribution family as the posterior