Tufts UNIVERSITY

Bayesian estimation

Motivation and Bayes' Theorem

Bruce M. Boghosian

Tufts UNIVERSITY | School of Arts and Sciences

Department of Mathematics

Tufts University

Outline

Bruce M. Boghosian

Bayesian estimation: The problem

- Given n possible events $\{A_j\}_{j=1}^n$, each with some corresponding probability
- These events *partition* the set of all such possibilities A:

  $A = \bigcup_{j=1}^n A_j = A_1 \cup A_2 \cup \cdots \cup A_n$ (Every possible outcome is accounted for)

  $A_j \cap A_k = \emptyset$ if $j \neq k$ (Outcomes are mutually exclusive)

- We have initial guess for *prior distribution*, $\{P(A_j)\}_{j=1}^n$.
- Let event B be a new observation related to above events.
- We wish to update $\{P(A_j)\}_{j=1}^n$, given observation B.

Bayesian estimation: The problem

- We wish to update $\{P(A_j)\}_{j=1}^n$, given observation B.
- We wish to find *posterior distribution*, $\{P\left(A_j \mid B\right)\}_{j=1}^n$.
- If there is a subsequent guess, process could be repeated.
- In this way, the algorithm "learns" from new information.
- According to "Bayesians", you should not suppose that a new piece of information will give you all the answers you are seeking.
- Rather, new information should update prior beliefs.

Bayes' Theorem

- Recall from the definition of conditional probability

$$P(A_j, B) = P(A_j \mid B)P(B) = P(B \mid A_j)P(A_j).$$

$$\therefore \quad P(A_j \mid B) = \frac{P(B \mid A_j)P(A_j)}{P(B)}$$

- Then recall that, since the $\{A_j\}_{j=1}^{n}$ partition A,

$$P(B) = \sum_{k=1}^{n} P(B \mid A_k)P(A_k)$$

- This gives us *Bayes' Theorem*,

$$\boxed{P(A_j \mid B) = \frac{P(B \mid A_j)P(A_j)}{\sum_{k=1}^{n} P(B \mid A_k)P(A_k)}.}$$

Reasoning with Bayes' Theorem

Bruce M. Boghosian

- *Bayes' Theorem*

$$\boxed{P\left(A_j \mid B\right) = \frac{P\left(B \mid A_j\right)P(A_j)}{\sum_{k=1}^{n} P\left(B \mid A_k\right)P(A_k)}.}$$

- Note numerator is one term of denominator, and all terms are positive.
- Result will always be in $[0, 1]$, as expected.
- Gives the desired $P\left(A_j \mid B\right)$ in terms of things that we might actually know, or be able to measure, or at least be able to estimate.
- The results of thinking in this way can defy intuition.
- An enormous body of scientific conclusions are gleaned in exactly this way.
- Key question: How do we use this theorem in the context of a real problem?

Example: Librarians and Farmers

- Example due to D. Kahneman and A. Tversky, as presented in "Thinking Fast and Slow" by D. Kahneman, FSG Publishers, New York (2011) pp. 6-7.
- **Given information:** "An individual has been described by a neighbor as follows: 'Steve is very shy and withdrawn, invariably helpful but with very little interest in people or the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail.' "
- **Question:** "Is Steve more likely to be librarian or farmer?"
- According to Kahneman, the vast majority of people who are asked this question say that Steve is more likely to be a librarian, almost surely due to the stereotypical view of librarians as having these personality traits.
- Does this make sense? Are we missing any information?

Example: Librarians and Farmers (continued)

- Does this make sense? Are we missing information?
- There are $20\times$ as many farmers as librarians in the US.
- Sample of 105 people – 100 farmers and 5 librarians.
- Say 40% librarians have this trait & only 10% farmers.
- Hence 2 librarians & 10 farmers in our sample have trait.
- Given that a randomly selected person this trait, there is only a $^2/_{12} = {}^1/_6$ chance that they are a librarian, and a $^{10}/_{12} = {}^5/_6$ chance that they are a farmer.
- Kahneman: "Because there are so many farmers, it is almost certain that more 'meek and tidy' will be found on tractors than at library information desks."
- Clearly we *were* missing information. The ratio of farmers to librarians *matters* to the question as posed! Most people don't even consider this possibility!

Bayesian analysis of Librarians & Farmers problem

Bruce M. Boghosian

- Decide on your events and give them variable names:
  - $B =$ Steve is a "meek and tidy soul"
  - $A_1 =$ Steve is a librarian
  - $A_2 =$ Steve is a farmer
- We know that $P(A_1) = \frac{1}{21}$ and $P(A_2) = \frac{20}{21}$.
- We estimate $P(B \mid A_1) = \frac{4}{10}$, and $P(B \mid A_2) = \frac{1}{10}$.
- Then the probability that Steve is a librarian given that he is a "meek and tidy soul" is

$$P(A_1 \mid B) = \frac{P(B \mid A_1)P(A_1)}{P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2)} = \frac{\frac{4}{10} \cdot \frac{1}{21}}{\frac{4}{10} \cdot \frac{1}{21} + \frac{1}{10} \cdot \frac{20}{21}} = \frac{4}{24} = \frac{1}{6},$$

in agreement with our earlier calculation.

Summary

- We have reviewed Bayes' Theorem
- We have given an example of Bayesian reasoning
- We have motivated Bayesian parameter estimation