Bayesian estimation

Bruce M. Boghosian

Department of Mathematics

Tufts University

Outline

Convergence in probability and almost sure convergence

- Recall MLE estimator $\hat{\sigma^2}$ for the variance of the normal is not unbiased,

$$E\left[\hat{\sigma^2}(\vec{Y})\right] = \frac{n-1}{n}\sigma^2$$

- but it is *asymptotically unbiased* since

$$E\left[\lim_{n\to\infty}\left(\hat{\sigma^2}(\vec{Y}) - \sigma^2\right)\right] = \lim_{n\to\infty}\left(\frac{n-1}{n}\sigma^2 - \sigma^2\right) = 0.$$

Bayesian estimation: The problem

Bruce M. Boghosian

- 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 may have an initial guesses for the *prior distribution*, $\{P(A_j)\}_{j=1}^n$.
- Now event let B be some new observation related to the above events.
- We wish to update the $\{P(A_j)\}_{j=1}^n$, given the observation B.

Bayesian estimation: The problem

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

Bayes' Theorem

Bruce M. Boghosian

- 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

- *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 a librarian or a 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?
- Kahneman: There are $20\times$ as many farmers as librarians in the US.
- Consider a sample of 105 people consisting of 100 farmers and 5 librarians.
- Say 40% of librarians have this personality trait and only 10% of farmers.
- That means 2 librarians and 10 farmers in our sample have the 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 and Farmers problem

- 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.

Bayesian search strategy

- $A_j :=$ Event that missing object is in region r_j
- $B_j :=$ Event that object would be found if it was in r_j, and r_j was searched
- $P\left(A_j \mid B_j^C\right)$ is probability that the item is in r_j, given
- Start with *prior* estimates for the probabilities $\{P(A_j)\}_{j=1}^n$. This could be as simple as the uniform distribution $\forall j \in \{1, \ldots, n\}: \; P(A_j) = \frac{1}{n}$.
- Suppose a search is undertaken in r_j, and nothing is found there.

Updating the prior for the searched region

Bruce M. Boghosian

- Apply Bayes' Theorem, and note that $P\left(B_j^C \mid A_j^C\right) = 1$

$$P\left(A_j \mid B_j^C\right) = \frac{P\left(B_j^C \mid A_j\right)P(A_j)}{P\left(B_j^C \mid A_j\right)P(A_j) + P\left(B_j^C \mid A_j^C\right)P(A_j^C)}$$
$$= \frac{\left[1 - P\left(B_j \mid A_j\right)\right]P(A_j)}{\left[1 - P\left(B_j \mid A_j\right)\right]P(A_j) + \left[1 - P(A_j)\right]} = \left[\frac{1 - P\left(B_j \mid A_j\right)}{1 - P\left(B_j \mid A_j\right)P(A_j)}\right]P(A_j)$$

- Hence, after unsuccessful search in r_j, prior for $P(A_j)$ is updated to obtain

$$\boxed{P^*(A_j) = P\left(A_j \mid B_j^C\right) = \left[\frac{1 - P\left(B_j \mid A_j\right)}{1 - P\left(B_j \mid A_j\right)P(A_j)}\right]P(A_j) < P(A_j).}$$

Updating the priors for the other regions

- Rescale all the other priors, $P(A_k)$ where $k \neq j$,

$$P^*(A_k) = P\left(A_k \mid B_j^C\right) = \alpha_j P(A_k)$$

- Demand normalization $1 = \sum_{m=1}^n P^*(A_m)$ to find

$$\alpha_j = \frac{1}{1 - P\left(B_j \mid A_j\right) P(A_j)}$$

- So the other priors update as follows

$$\boxed{P^*(A_k) = P\left(A_k \mid B_j^C\right) = \left[\frac{1}{1 - P\left(B_j \mid A_j\right) P(A_j)}\right] P(A_k) > P(A_k).}$$

Bayesian search strategy algorithm

- For all $m = 1, \ldots, n$:
  - Determine the probabilities that a search in r_m will find the object if it is in r_m.
  - Guess the priors $P(A_m)$.
- Now repeat until the object is found:
  - Suppose that the value of m for which $P(A_m)$ is highest is $m = j$.
  - Conduct a search in region r_j
  - If the search is unsuccessful, update the priors and repeat.

Application to estimation

- Let W be a statistic dependent on a parameter θ. Call its pdf $f_W(w \mid \theta)$.
- Suppose that θ is the value of a random variable Θ.
- Denote the prior distribution of Θ by
  - $p_\Theta(\theta)$ if Θ is discrete
  - $f_\Theta(\theta)$ if Θ is continuous
- Posterior distribution of Θ, given that $W = w$ is the quotient

$$g_\Theta(\theta \mid W = w) = \begin{cases} \dfrac{p_W(w \mid \theta) f_\Theta(\theta)}{\int d\xi \; p_W(w \mid \xi) f_\Theta(\xi)} & \text{if } W \text{ is discrete} \\[2ex] \dfrac{p_W(w \mid \theta) f_\Theta(\theta)}{\int d\xi \; f_W(w \mid \xi) f_\Theta(\xi)} & \text{if } W \text{ is continuous} \end{cases}$$

- If Θ is discrete
  - Replace the $f_\Theta(\theta)$ by $p_\Theta(\theta)$.
  - Replace the integrals over θ by sums.

Summary

- We have reviewed Bayes' Theorem
- We have given an example of Bayesian reasoning
- We have studied Bayesian search strategy
- We have studied Bayesian estimation