"Frequentism and Bayesianism: A Python-driven Primer"

Daniel Ramyar

Niels Bohr Institute
University of Copenhagen

March 8, 2018

Overview

The article

Frequentism and Bayesianism: A Python-driven Primer

Jake VanderPlas*†

Abstract—This paper presents a brief, semi-technical comparison of the essential features of the frequentist and Bayesian approaches to statistical inference, with several illustrative examples implemented in Python. The differences between frequentism and Bayesianism fundamentally stem from differing definitions of probability, a philosophical divide which leads to distinct approaches to the solution of statistical problems as well as contrasting ways of asking and answering questions about unknown parameters. After an example-driven discussion of these differences, we briefly compare several leading Python statistical packages which implement frequentist inference using classical methods and Bayesian inference using Markov Chain Monte Carlo.[1]

advanced Bayesian and frequentist diagnostic tests are left out in favor of illustrating the most fundamental aspects of the approaches. For a more complete treatment, see, e.g. [Wasserman2004] or [Gelman2004].

The Disagreement: The Definition of Probability

Fundamentally, the disagreement between frequentists and Bayesians concerns the definition of probability.

For frequentists, probability only has meaning in terms of

Jake VanderPlas *Frequentism and Bayesianism: A Python-driven Primer*, arXiv:1411.5018

The Disagreement

Frequentist

Probabilities are related to frequency of events

Example:

Imagine repeated flux measurements of a star.
The measurements will vary slightly due to statistical error.
The frequency of any measured value is the probability of measuring that value

Bayesian

Probabilities are related to our own knowledge of an event

Example:

We know with some probability the true flux of the star.
This probability can be estimated through repeated measurements and is a statement of researches knowledge.

Photon flux measurement

Given data estimate the value of the true flux?

We create some toy data which is 50 samples from some flux measurements F with mean 1000 and some known error.

```
>>> np.random.seed(2)  # for reproducibility
>>> e = np.random.normal(30, 3, 50)
>>> F = np.random.normal(1000, e)
```

The frequentist approach

We assume the true flux is distributed Gaussian.

$$P(D_i|F) = (2\pi e_i^2)^{-1/2} \exp\left(\frac{-(F_i - F)^2}{2e_i^2} \right) \tag{1}$$

Then we use the maximum likelihood method to estimate the flux we get

$$\hat{F} = 999 \pm 4 \tag{2}$$

The Bayesian approach

We apply Bayes' theorem

$$P(F|D) = \frac{P(D|F)P(F)}{P(D)} \tag{3}$$

We assume flat prior $P(F) \propto 1$ then we get

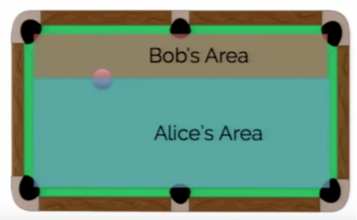
$$P(F|D) \propto \mathcal{L}(D|F) \tag{4}$$

So the estimates are equivalent! Though much more information can be incorporated in the prior.

Bayes' Billiards games

What is the probability that Bob will win?

Behind a curtain a third person Carol rolls a ball down a table and marks where it lands. Now the process is repeated and if the subsequent balls lands to the right of the mark Bob gets a point and if it lands to the left Alice gets a point. If either of them gets 6 points they win. The standing is now 5 to Alice and 3 to Bob.



The naive frequentist approach

We need to estimate the marker location with some probability p. We use maximum likelihood to do this. We get

$$\hat{p} = 5/8 \tag{5}$$

The probability that bob will win will then be

$$P(B) = (1 - \hat{p})^3 = 0.053 \tag{6}$$

or the odds against Bob wining is 18 to 1

The Bayesian approach

We want to calculate $P(B|D)$ recognizing this as a marginal probability

$$P(B|D) = \int_{-\infty}^{\infty} P(B,p|D)dp \tag{7}$$

Manipulating this expression using some Bayesian relations and Bayes' rule we get

$$P(B|D) = \frac{\int_0^1 (1-p)^6 p^5 dp}{\int_0^1 (1-p)^3 p^5 dp} = 0.091 \tag{8}$$

or the odds against Bob wining is 10 to 1.
So who's right?

Conclusion

Bayesian method in the cases discussed is a more natural approach to statistical analysis.

First example showed us that the frequentist approach was merely a special case of the Bayesian method and that the existence of the model prior would allow us to incorporate much more information into the calculation.

The second example showed us that subtleties of handling nuisance parameters in the frequentist method could be totally avoided using the Bayesian method.