"Frequentism and Bayesianism: A Python-driven Primer" Summary Advanced Methods in Applied Statistics

NJT478

March 7, 2018

Summary

Introduction

One of the first things we hear about in an introductory statistics or data science course is the existence of two schools in statistical inference, namely frequentist- and Bayesian inference.

This article [1] presents the fundamental differences between these two approaches and discusses the diverging results through simple examples with python code. Which also shows the practical differences between the two methods. Furthermore it explores how these disagreements affect the data analysis and interpretation of scientific results.

Frequentist and Bayesian definition of probability

What fundamentally separates the frequentist and Bayesian schools is their different views on the definition of probability.

For a frequentist probabilities are ultimately related to the frequency of an observed event. The example given in the article looks at repeated flux measurements F taken from a star, each value varying slightly because of statistical error. After many measurements it will then be the frequency of the values which will give us the probability of measuring it. There is no meaning in talking about the probability of the true value of F since the true value of F would be a fixed number with no probability distribution.

Whereas for Bayesians, probabilities are related to our own knowledge of an event or lack thereof. Therefore one can argue with what probability they claim know the true value of F or the true value of F is lying within some range.

This is the difference which in practice leads to the different approaches of the statistical analysis.

Practical examples

The articles goes through 4 problems, a simple parameter estimation problem, a nuisance parameter problem i.e what do we do about a parameter whose value is not of interest but it required to determine the relevant value, how do we handle uncertainties i.e confidence vs credibility and lastly a realistic scenario to show the difference in approach.

I will go through two of the examples in little more detail and summarize the other two in the conclusion.

In the simple parameter estimation example he creates some toy data which is to represent 50 measurements of a stars flux F where we know the mean and error. Next is to try to estimate these known parameters of the true flux F.

The frequentist approach would be to use classic maximum likelihood method on our data where we assume our data is normally distributed.

We then try to accomplish the same using Bayes' theorem assuming that our model prior is 1. This gives us that the probability of our true value is proportional to the likelihood i.e giving the same result.

But how do we know the model prior? It is here the frequentists would say the subjective choice of the model prior is problematic. He counters it by saying that frequentists doesn't solve the problem but simply avoids it.

The next example is about the handling of nuisance parameters. The setup is a gambling game in which Alice and Bob bet on an outcome they can't observe directly. 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. What is the probability that Bob will win?

The naive frequentist approach to determine this would be to first estimate the location of the marker. The probability of the first marker placement is given by p and the first placement for any throw should favor Alice since she already have 5 points. Assuming this he estimates $p = \frac{5}{8}$ using the maximum likelihood method and calculates the probability that Bob will win using $P(B) = (1-p)^3 = 0.053$ or odds against Bob wining 18 to 1.

It is here the Bayesian approach reaches different conclusion namely that the odds against bob winning to be 10 to 1. Who is right?

The problem can be solved simulating a large number of games which confirms that the Bayesian approach gives us the right answer.

So is the frequentist method wrong? No the frequentist answer were wrong because we didn't approach the problem in the right way. Were it would require special expertise in the area using the frequentist method in this case, he argues that these subtleties of handling nuisance parameters could be avoided using the Bayesian method.

Conclusion

This article in essence wanted to show the differences between the frequentist and Bayesian approach and that the fundamental difference in the definition of probability gives rise to vastly different results.

He concludes by arguing that the Bayesian method in the cases discussed is a more natural approach to statistical analysis. Where it was shown in the first example 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.

In the third example where we wanted to bind our true value within some range. He showed us that using the Bayesian method answers the question asked and you'd have to be careful applying the frequentist method as it answers another question if not applied properly.

Lastly he combined the different ideas of the frequentist and Bayesian method on a more realistic example to show how you could handle the different problems in the Python environment using frequentist and Bayesian statistical analysis tools.

References

[1] Jake VanderPlas *Frequentism and Bayesianism: A Python-driven Primer* arXiv:1411.5018