STATS 700-001 **Fall 2017**

Lecture 2: September 11

Lecturer: Yang Chen *Scribes: Cameron Hollingshead*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

2.1 What is Bayesian inference?

To best illustrate what Bayesian inference means, it may suffice to distinguish it from the frequentist approach to statistical inference. Suppose we have a STATS 700 class roster, and we wish to figure out how many students who attended the first lecture will drop the class before the second lecture.

Frequentist: We can build a model, using such variables as counts on different assignment types, exams, or programming tools, that can be used to estimate either how many or what percentage of students drop the class.

Bayesian: If we had prior information on what caused students from previous semesters to drop this class, say through a survey sample, we could use that to predict the dropout rate for this semester's class, i.e. before it is certain how many students drop the class. Additionally, now that we know the dropout rate, we can use this information to update our model, in order to better predict dropout by the next semester.

2.1.1 Why choose Bayesian inference?

Models are built and restructured, based on a continuous feed of information. This procedural method of modeling allows for deeper learning into the underlying ”true” model of the data. When acting as an inductive reasoner, we must accept the uncertainty of particular parameters, regarding the population of consideration. A great advantage to this is that our estimators can work for a sample of any size, without as significant of a risk in overfitting a model.

2.2 How to Conduct Bayesian Analysis

1. Initial Setup of Full Probability Model

2. Obtaining Posterior Distribution After Observing Data

3. Evaluation / Interpretation of Posterior

2.2.1 Setup of Probability Model

Prior Distribution: $P(\theta)$ describes our belief that, prior to observing any data, parameter θ represents the true population characteristics of interest

Sampling Model: $P(y_1, \ldots, y_n|\theta)$ describes our belief that, given how we define our parameter of interest θ, each $y_i, 1 \leq i \leq n$, would be what we observe

$y_i \sim P(\bullet|\theta)$, conditional probability distribution

$\theta \sim P(\bullet)$, marginal distribution

2.2.2 The Posterior Distribution

In the frequentist approach, we would see θ as a well-defined parameter, and aim to build the most accurate sampling model after observing data. In the Bayesian approach, we have preemptively set up a sampling model and prior distribution. What we do next is observe some quantities, and then derive a new distribution that describes parameter θ. This is our posterior distribution.

Posterior Distribution: $P(\theta|y_1, \ldots, y_n) = \frac{P(y_1,\ldots,y_n|\theta)*P(\theta)}{P(y_1,\ldots,y_n)}$, where the denominator is a constant defined by each independent observation of the data

2.2.3 What We Do with Posterior

How we interpret the posterior distribution is that, now that we have observed some data, we have a better way to describe the parameter θ. That means, when we observe a new quantity y_{n+1}, our new "prior" distribution is the posterior we had obtained before.

New Prior: $P(\theta|y_1, \ldots, y_n)$

New Sampling Model: $P(y_1, \ldots, y_{n+1}|\theta)$

New Posterior to Obtain: $P(\theta|y_1, \ldots, y_{n+1}) = \frac{P(y_1,\ldots,y_{n+1}|\theta)*P(\theta|y_1,\ldots,y_n)}{P(y_1,\ldots,y_{n+1})}$

References

[lec2] Y. Chen, "Basics of Bayesian Analysis," *Canvas Slides*, 2017, pp. 1–17.