I. Introduction

MA480 – Introduction to Bayesian Inference

Academic Year 2014-2015

1 Statistical Inference

Let's begin by reviewing the fundamental ideas of statistical inference that form the backbone of the methodology discussed in other courses. We begin with a thought exercise adapted from Ellenberg's *How Not to Be Wrong: The Power of Mathematical Thinking.*

Example 1. *(Terrorists and Social Media)* Suppose a collaborative team from Facebook and Twitter has developed an algorithm for flagging possible terrorists based on their interaction with social media. Based on their findings, those flagged by the algorithm are "twice as likely to have ties to a terrorist organization compared to the typical social media user." Now, suppose you get a hold of the algorithm and determine your professor is on the list, how do you react?

Before addressing this problem directly, let's address a related question.

Example 2. *(Estimating Terrorist Activity)* What fraction of social media users are associated with terrorist activity? In particular, suppose an extremist group claims that at least 0.1% of all social media users has ties to a terrorist organization. Is there evidence to refute their claim?

★ ***Fundamental Idea I***

By framing a question of interest in terms of a _____ of the population, we can use data collected from a _____ and corresponding _____ to make statements about the population.

Example 3. *(Example 2 Cont.)* For the Estimating Terrorist Activity example, set up the null and alternative hypothesis that captures the question of interest.

★ ***Fundamental Idea II***

If data is to be useful for making conclusions about the _____ (a process referred to as _____ _____), proper data collection is crucial. _____ is the critical component to ensuring the sample is _____ of the population.

Example 4. *(Example 2 Cont.)* Continuing with the Estimating Terrorist Activity example, describe a data collection method for assessing the question of interest.

★ ***Fundamental Idea III***

The use of numerical data for decision making requires that the data be _____ and _____ in ways that allow you to address the question of _____.

★ ***Fundamental Idea IV***

Variability is inherit in any process, and as a result, _____ vary across _____ in a _____ way. that is, they have a _____ that can be modeled. Constructing such models will allow us to draw inference on the _____.

Example 5. *(Example 2 Cont.)* Continuing with our Estimating Terrorist Activity example, identify a statistic and its corresponding model for its sampling distribution that allows us to address the question of interest.

★ ***Fundamental Idea V***

With a model of the sampling distribution, we can make _____ statements about the value of a _____; that is, we can determine which values are likely, and which are not under a specified model. This allows us to draw conclusions about the corresponding _____ (and therefore population) of interest.

Example 6. *(Example 2 Cont.)* Using the model you constructed for the Estimating Terrorist Activity example, address the question of interest at the 0.05 significance level.

Now that we have reviewed the basic ideas, we are ready to reconsider that rogue instructor!

Example 7. *(Example 1 Cont.)* For the Terrorists and Social Media example, evaluate the evidence presented by the data to make a decision regarding your instructor.

2 Two Questions, Two Approaches

The Two Questions Posed in Decision-Making:

There are basically two questions we could ask when addressing a question for the purpose of decision-making:

- Given the _____ _____ is true, how likely is our _____ _____?
- Given the _____ _____, how likely is the _____ of interest?

Example 8. *(Example 1 Cont.)* For our Terrorists and Social Media example, given the observed data, how likely is it that your professor is a terrorist?

- Bayesian inference relies on _____ knowledge to make inference.
- Bayesian inference captures _____ in _____ or _____ through probability distributions.
- Bayesian inference relies on _____ probabilities.