A Polemic for Bayesian Statistics

Richard E. Neapolitan
Northeastern Illinois University, Chicago, Il 60625
RE-Neapolitan@neiu.edu

Abstract. In the early part of the 20th century, forefathers of current statistical methodology were largely Bayesians. However, by the mid-1930’s the Bayesian method fell into disfavor for many, and frequentist statistics became popular. Seventy years later the frequentist method continues to dominate. My purpose here is to compare the Bayesian and frequentist approaches. I argue for Bayesian statistics claiming the following: 1) Bayesian methods solve a wider variety of problems; and 2) It is sometimes difficult to interpret frequentist results.

2.1 Introduction

In the early part of the 20th century, forefathers of current statistical methodology were largely Bayesians (e.g. R.A. Fisher and Karl Pearson). However, by the mid-1930’s the Bayesian method fell into disfavor for many, and frequentist statistics, in particular the use of confidence intervals and the rejection of null hypotheses using $p$-values, became popular. Seventy years later the frequentist method dominates, and most university statistics courses present this approach without even a reference to Bayesian methods. However, there has recently been a resurgence of interest in Bayesian statistics, partly due to the use of Bayesian methods in Bayesian networks ([Pearl, 1988], [Neapolitan, 1990]) and in machine learning. Curiously, sometimes systems that learn using Bayesian methods are evaluated using frequentist statistics (e.g. in [Buntine, 1992]). This attests to the dominance of frequentist statistics. That is, it seems researchers in Bayesian machine learning learned frequentist evaluation methods first and therefore adhere to them.

My purpose here is to compare the Bayesian and frequentist approaches focusing on interval estimation and hypothesis testing. I make no effort to cover the frequentist or Bayesian method in any detail; rather I show only sufficient mathematics and examples to contrast the two approaches. I argue for Bayesian statistics claiming the following: 1) Bayesian methods solve a wider variety of problems; and 2) It is sometimes difficult to interpret frequentist results, particularly in the case of hypothesis testing, whereas the interpretation of Bayesian results is straightforward and meaningful. Bolstad [2004] solves many more problems using the two approaches, while Zabell [1992] presents some of the history of statistics over the 20th century, focusing on the life of R.A. Fisher.
assume the reader is familiar with both the Bayesian and frequentist approaches to probability. See [Neapolitan, 2004] for a brief, facile discussion of the two.

After methods for interval estimation are compared, hypothesis testing is discussed.

2.2 Interval Estimation

A random sample consists of independent, identically distributed random variables $X_1, X_2, \ldots, X_n$. In practice a random sample is obtained by selecting individuals (or items) at random from some population.

Example 1. Suppose we select 100 American males at random from the population of all American males. Let $X_i$ be a random variable whose value is the height of the $i$th individual. If, for example, the 10th individual is 71 inches, then the value of $X_{10}$ is 71 inches.

Mathematically, the population is really the collection of all values of the variable of interest. For example, if our set of entities is {George, Sam, Dave, Clyde}, their heights are 68 inches, 70 inches, 68 inches, and 72 inches, and the variable of interest is height, then the population is the collection [68, 70, 68, 72]. However, it is common to refer to the collection of all entities as the population.

A common task is to assume that the random variables are normally distributed, obtain values $x_1, x_2, \ldots, x_n$ of $X_1, X_2, \ldots, X_n$, estimate the true population mean (expected value) $\mu$ by

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i,$$

and obtain some measure of confidence as to how good this estimate is.

Example 2. Suppose we sample 100 American males and the average value of the height turns out to be 70 inches. Then $\bar{x} = 70$, and our goal is to obtain some measure of confidence as to how close the true average height $\mu$ is to 70.

Ordinarily, we don’t know the variance either. However, for the sake of simplicity, in the discussion here we will assume the variance is known. Although it is elementary statistics, we will review the classical statistical technique for obtaining a confidence interval so that we can compare the method to Fisher’s fiducial probability interval, and the Bayesian probability interval.

2.2.1 Confidence Intervals and Fiducial Probability Intervals

First denote the normal density function and normal distribution as follows:

$$\text{NormalDen}(x; \mu, \sigma^2) \equiv \frac{1}{2\pi\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$