The Frequentist School of Statistics
Class 17, 18.05, Spring 2014
Jeremy Orloff and Jonathan Bloom

1 Learning Goals

1. Be able to explain the difference between the frequentist and Bayesian approaches to statistics.

2. Know our working definition of a statistic and be able to distinguish a statistic from a non-statistic.

2 Introduction

After much foreshadowing, the time has finally come to switch from Bayesian statistics to frequentist statistics. For much of the twentieth century, frequentist statistics has been the dominant school. If you've ever encountered confidence intervals, p-values, t-tests, or χ^2-tests, you've seen frequentist statistics. With the rise of high-speed computing and big data, Bayesian methods are becoming more common. After we've studied frequentist methods we will compare the strengths and weaknesses of the two approaches.
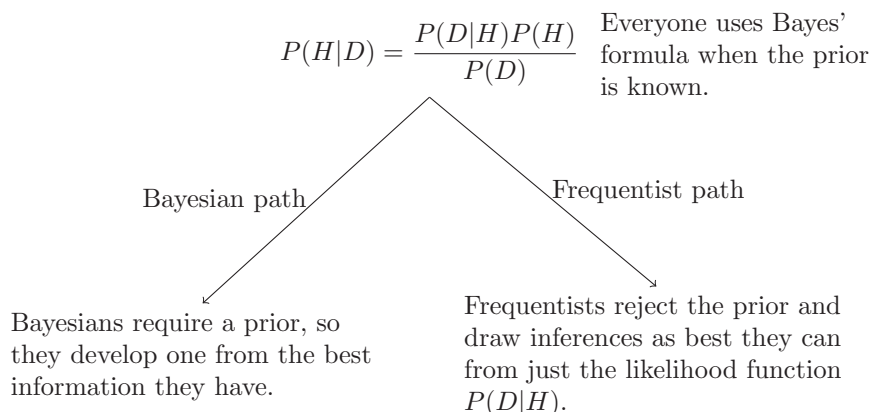
2.1 The fork in the road

Both schools of statistics start with probability. In particular both know and love Bayes theorem:

$$P(H|D) = \frac{P(D|H)\,P(H)}{P(D)}.$$

When the prior is known exactly all statisticians will use this formula. For Bayesian inference we take H to be a hypothesis and D some data. Over the last few weeks we have seen that, given a prior and a likelihood model, Bayes theorem is a complete recipe for updating our beliefs in the face of new data. This works perfectly when the prior was known perfectly. We saw this in our dice examples. We also saw examples of a disease with a known frequency in the general population and a screening test of known accuracy.

In practice we saw that there is usually no universally-accepted prior – different people will have different *a priori* beliefs – but we would still like to make useful inferences from data. Bayesians and frequentists take fundamentally different approaches to this challenge, as summarized in the figure below.

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

Everyone uses Bayes' formula when the prior is known.

Bayesian path

Frequentist path

Bayesians require a prior, so they develop one from the best information they have.

Frequentists reject the prior and draw inferences as best they can from just the likelihood function $P(D|H)$.

The reasons for this split are both practical (ease of implementation and computation) and philosophical (subjectivity versus objectivity and the nature of probability).

2.2   What is probability?

The main philosophical difference concerns the meaning of probability. The term *frequentist* refers to the idea that probabilities represent longterm frequencies of repeatable random experiments. For example, 'a coin has probability 1/2 of heads' means that the relative frequency of heads (number of heads out of number of flips) goes to 1/2 as the number of flips goes to infinity. This means the frequentist finds it non-sensical to specify a probability distribution for a parameter with a fixed value. While Bayesians are happy to use probability to describe their incomplete knowledge of a fixed parameter, frequentists reject the use of probability to quantify degree of belief in hypotheses.

Example 1. Suppose I have a bent coin with unknown probability θ of heads. The value of θ may be unknown, but it is a fixed value. Thus, to the frequentist there can be no prior pdf $f(\theta)$. By comparison the Bayesian may agree that θ has a fixed value, but interprets $f(\theta)$ as representing *uncertainty* about that value. Both the Bayesian and the frequentist are perfectly happy with $p(\text{heads}\,|\,\theta) = \theta$, since the longterm frequency of heads given θ is θ.

In short, Bayesians put probability distributions on everything (hypotheses and data), while frequentists put probability distributions on (random, repeatable, experimental) data given an hypothesis. For the frequentist when dealing with data from an unknown distribution only the likelihood has meaning. The prior and posterior do not.

3   Working definition of a statistic

Our view of statistics is that it is the art of drawing conclusions (making inferences) from data. With that in mind we can make a simple working definition of a statistic. There is a more formal definition, but we don't need to introduce it at this point.

Statistic. A *statistic* is anything that can be computed from data. Sometimes to be more precise we'll say a statistic is a *rule* for computing something from data and the *value* of the statistic is what is computed. This can include computing likelihoods where we hypothesize

values of the model parameters. But it does not include anything that requires we know the true value of a model parameter with unknown value.

Examples. 1. The mean of data is a statistic. It is a rule that says given data $x_1, \ldots, x_n$ compute $\frac{x_1 + \ldots + x_n}{n}$.

2. The maximum of data is a statistic. It is a rule that says to pick the maximum value of the data $x_1, \ldots, x_n$.

3. Suppose $x \sim N(\mu, 9)$ where μ is unknown. Then the likelihood

$$p(x|\mu = 7) \quad = \quad \frac{1}{3\sqrt{2\pi}} \, \mathrm{e}^{-\frac{(x-7)^2}{18}}$$

is a statistic. However, the distance of x from the true mean μ is *not* a statistic since we cannot compute it without knowing μ

Point statistic. A *point statistic* is a single value computed from data. For example, the mean and the maximum are both point statistics. The maximum likelihood estimate is also a point statistic since it is computed directly from the data based on a likelihood model.

Interval statistic. An *interval statistic* is an interval computed from data. For example, the range from the minimum to maximum of $x_1, \ldots, x_n$ is an interval statistic, e.g. the data .5, 1, .2, 3, 5 has range [.2, 5].

Set statistic. A *set statistic* is a set computed from data.

Example. Suppose we have five dice: 4, 6, 8, 12 and 20-sided. We pick one at random and roll it once. The value of the roll is the data. The set of dice for which this roll is possible is a set statistic. For example, if the roll is a 10 then the value of this set statistic is {12, 20}. If the roll is a 7 then this set statistic has value {8, 12, 20}.

It's important to remember that a statistic is itself a random variable since it is computed from random data. For example, if data is drawn from $N(\mu, \sigma^2)$ then the mean of n data points follows $N(\mu, \sigma^2/n)$).

Sampling distribution. The probability distribution of a statistic is called its *sampling distribution*.

Point estimate. We can use statistics to make a *point estimate* of a parameter θ. For example, if the parameter θ represents the true mean then the data mean $\bar{x}$ is a point estimate of θ.

18.05 Introduction to Probability and Statistics

Spring 2014