Class 1: Introduction
18.05, Spring 2013

1 Introduction

In this introduction we will preview what we will be studying in 18.05. Don't worry if many of the terms are unfamiliar, they will be explained as the course proceeds.

Probability and statistics two different subjects which are deeply connected because all statistical statements are statements about probability. Despite this the two sometimes feel like two separate subjects. Probability is logically self-contained; there are a few rules and, though computations can be tricky, answers all follow logically from the rules. In statistics we apply probability to draw conclusions from data. This can be messy and often involves as much art as science.

Probability example
Toss a fair coin (probability 1/2 of heads). What is the probability of 60 or more heads in 100 tosses?

Here the probability is fully known and there is only one answer. We will learn how to compute the answer to this question.

Statistics example
You have have a coin of unknown origin and want to know if it is fair. To do this you run an experiment and record the data.
Experiment: you toss the coin 100 times.
Data: you get 60 heads in the 100 tosses.

Your job as a statistician is to draw a conclusion (inference) from this data. There many ways to proceed, both in terms of the form the conclusion takes and the probability computations used to justify the conclusion.

Note, that in the first example the random process (a fair coin) is fully known. The objective is to find the probability of certain outcomes (at least 60 heads) arising from the random process. In the second example, the outcomes are known (60 heads) and the objective is to illuminate the unknown random process (the probability of heads).

2 Frequentist vs. Bayesian Interpretations

In fact, there are two prominent and sometimes conflicting approaches to statistics: Bayesian and frequentist. The differences are rooted in differing interpretations of the meaning of probability.

Frequentist: Probability measures the frequency of various outcomes of an experiment. For example, saying a fair coin has a 50% chance of coming up heads means that if we toss it a lot of times then we expect about half the tosses to be heads.

Bayesian: Probability is an abstract concept that measures a state of knowledge or a degree of belief in a given proposition. In practice Bayesians do not assign a single value for the probability of a coin coming up heads. Rather they assign a range of values each with its own likelihood of being true.

In 18.05 we will study and compare these approaches. The frequentist approach has long been dominant in fields like biology, medicine, public health and social sciences. The Bayesian approach has enjoyed a resurgence in the era of powerful computers and big data. It is especially useful when incorporating new data into an existing statistical model, for example, when training a speech or face recognition system. More and more statisticians are using both approaches in complementary ways.

3   Applications and Toy Models

Probability and statistics are used widely in physical science, engineering, medicine, the social sciences and the life sciences. The list of applications is essentially endless. Examples include experiments testing one medical treatment against another (or a placebo), measures of genetic linkage, the search for elementary particles, machine learning for vision or speech, gambling probabilities and strategies, climate modeling, economic forecasting, epidemiology, marketing and googling.

It is important to learn how to see the simple essence inside a messy real situation. We will therefore spend quite a bit of time thinking about *toy models*. The most common one will be simply tossing a coin. Although we use the word 'toy' to descibe this example, it is a realistic model of many situations with two possibilities: success or failure of a treatment, an airplane engine, a bet, or even a class.