Ph 21.4a: Introduction to Bayesian Data Analysis

-v20130102-

Introduction to Bayesian Analysis

In this assignment, we will explore some elementary concepts in Bayesian data analysis, also called "Bayesian Inference". We will only scratch the surface of this very important topic. For a readable, and much more extensive presentation of the subject, see the book by Sivia, *Data Analysis: A Bayesian Tutorial* [1]. The approach in this course is, as usual, to provide some exposure to the topic of Bayesian Analysis and point out some of the computational tools available.

The interest in and use of Bayesian analysis techniques has grown with the increasing availability of computing resources. In these assignments we will be working with rather simple examples involving models with a small number of parameters (e.g. 1 or 2), just to illustrate the basic principles. However, the true utility of Bayesian techniques only becomes evident for problems with large-dimensionality, i.e. ones involving models of the data containing many parameters. Basic objectives of Bayesian Analysis include determination of confidence intervals for parameter values and quantitative comparison of how well different models describe data.

Bayesian techniques are often contrasted with "frequentist" techniques. We will not attempt to address that discussion here. You should, however, be aware that before Bayesian techniques became popular, there was a large body of classical statistical techniques that were used to draw inferences from data. These involve definition of a quantity called a "statistic", often depending on model parameters, whose value is calculated from the data. *Numerical Recipes* [2] discusses several of these classical statistical approaches.

Bayes' Theorem

Bayesian analysis starts, as might be expected, from Bayes' Theorem. Suppose the system for which we have data can be well-described by a "model" characterized by M parameters: $\mathbf{X} = \{x_1, x_2, ..., x_M\}$. Further, suppose the we have N data: $\mathbf{D} = \{d_1, d_2, ..., d_N\}$. Bayes' Theorem is then:

$$\text{prob}(\mathbf{X}|\mathbf{D}, I) = \frac{\text{prob}(\mathbf{D}|\mathbf{X}, I) \times \text{prob}(\mathbf{X}|I)}{\text{prob}(\mathbf{D}|I)}$$

where I represents the background information that we have about the system. In words this is:

$$\text{prob}(\textit{hypothesis given data}\ \&\ info) \propto \text{prob}(\textit{data given hypothesis}\ \&\ info) \times \text{prob}(\textit{hypothesis given info})$$

that is, the probability of the parameters taking on a given set of values given a set of measured data is proportional to the probability of that data being realized in a system characterized by parameters with the given values, times the 'probability of those parameter values given our prior information about the system. Bayes' theorem follows from very general axioms of probability. It provides a prescription for converting a statement about the probability of measuring certain values for the data to a statement about the likelihood of certain values of the parameters given a measured set of data. This is often what is needed in data analysis: we have a model of a system, we have data, and we want to make inferences about the parameters that characterize the model of the system.

Some terminology: $\text{prob}(\mathbf{X}|I)$ is called the "prior"; $\text{prob}(\mathbf{X}|\mathbf{D}, I)$ is called the "posterior" ; $\text{prob}(\mathbf{D}|\mathbf{X}, I)$ is sometimes called the "likelihood'; and $\text{prob}(\mathbf{D}|I)$ is a normalization, called the "evidence", that is often disregarded in calculations, the proportionality being the important consideration.

The Assignment - Part I

Begin by calculating a very simple system: coin tossing with a possibly biased coin. This system is discussed in section 2.1 of [1]. Let H be the parameter which quantifies the probability of "heads". For an unbiased coin, we should have $H = 0.5$. A coin with heads on both sides of the coin will have $H = 1$.

Figure 1: Lighthouse Geomtery

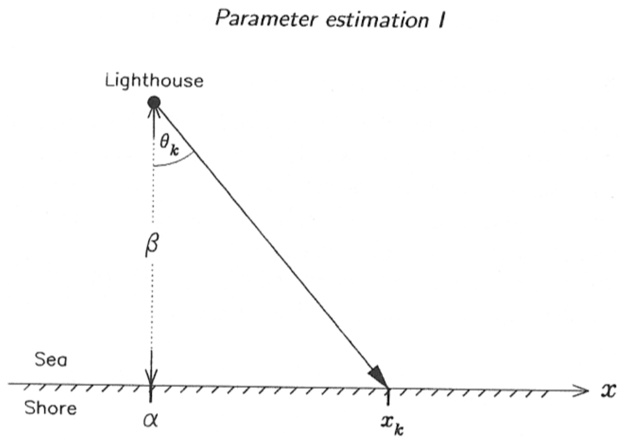

Fig. 2.7 A schematic illustration of the geometry of the lighthouse problem.

For coin flipping, the probability of a set of results (data) is given by a binomial distribution:

$$\text{prob}(\mathbf{D}|H, I) = \frac{n!}{h!(n-h)!} H^h (1-H)^{n-h}$$

where $\mathbf{D}$ is the data, i.e. h heads in n trials.

For this part of the assignment, develop python code that can simulate coin tossing and calculate posterior distributions, i.e. $\text{prob}(H|\mathbf{D}, I)$ for various assumed values of H. Plot $\text{prob}(H|\mathbf{D}, I)$ versus H for a variety of number of trials n and for several choices of the "true" value of H. For instance, see how $\text{prob}(H|\mathbf{D}, I)$ evolves by plotting it for $n = 1, 2, 4, 8, 16, 32, \ldots$ trials.

You will need to assume some priori distributions. First investigate a uniform prior, i.e. $\text{prob}(\mathbf{X}|I) =$ constant. Also calculate distributions for a gaussian prior. Plot distributions both for "true" values of H within 1-sigma of the peak of your chosen gaussian and for an H that is 3-sigma from the peak. Investigate how the posterior distributions evolve with time.

Section 2.1 of [1] discusses these various calculations and shows plots similar to those that you should obtain. See attached figure 3.

The Assignment - Part II

The coin tossing problem is a very simple 1-parameter system with familiar binomial and gaussian probability distributions. The next system is also quite simple, but begins to illustrate some of the complexities of data analysis and inference. In particular, it is a system for which one of the foundation assumptions of classical statistics, the central limit theorem, does not hold. As part of your challenge, you should be aware that this is a problem given to 1st-year Cambridge undergraduates.

The system is taken again from [1] (section 2.4), the lighthouse problem. The geometry of the problem is shown in figure 1 taken from [1].

The lighthouse is a distance α along the shore and a distance β out to sea. It emits highly collimated flashes at random times and at random azimuths (a good lighthouse emits flashes at regular intervals, usually due to rotation of the lighthouse beam). A dense array of photodetectors is set up along the shore and each flash

Figure 2: Flash position distribution

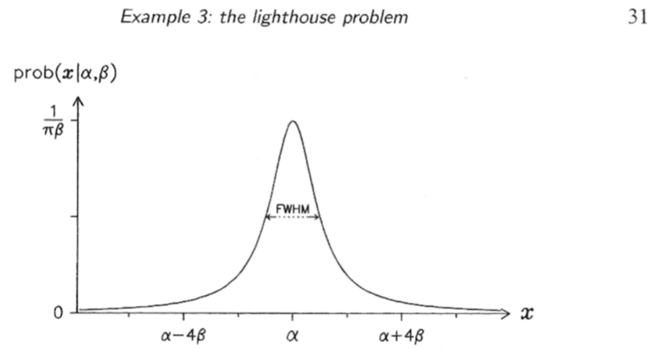

Fig. 2.8 The Cauchy, or Lorentzian, distribution. It is symmetric with respect to the maximum, at $x=\alpha$, and has a FWHM of twice β.

is recorded by a single photodetector. The data consist of the position x_k of the illuminated photodetector for the kth flash. The distribution is shown in figure 2.

For this part of the assignment, assume "true" values for α and β, say 1 km. Initially assume that β is known and that the problem is to infer the posterior probability distribution of α from the data $\{x_k\}$. [Hint: the best way to do this is to compute the logarithm of the posterior probability on a grid of values for α. Once this is done, you can re-normalize the values (since only the proportionality of the posterior distribution is really important). You can then take the exponential in order to plot the posterior probability distribution of α.] As in the coin tossing problem, plot the posterior distributions for various values n of the number of flashes recorded, n = 1,2,4,8,16,32, ... Also calculate the mean of $\{x_k\}$. You might expect that the mean of $\{x_k\}$ is a good estimator for the most probable value of α. Why isn't it?

Next consider the case where neither α nor β is known beforehand and the challenge is to infer the most probable value from the data. Calculate the distributions of the logarithm of the posterior probability for a grid of α and β values. Then plot the 2-dimensional distribution of the posterior distribution as a contour plot in the α-β plane.

The Next Assignment

You should have seen that Bayesian inference relies on significant amounts of computation, even for the simple systems explored in this assignment. However, many interesting systems that are encountered in research require many parameters for their characterization (i.e. have high dimensionality), have very complex probability distributions for the data (e.g. variable errors and bounds on the data), and have complex priors. To determine the most likely values of the parameters and the error estimates for these parameters requires additional significant computational machinery. That will be the subject of the next assignment.

References

[1] D.S. Sivia, Data Analysis: A Bayesian Tutorial (2006)

[2] W.H. Press et al., Numerical Recipes, various editions beginning in 1988.

Figure 3: Posterior distributions for coin tossing

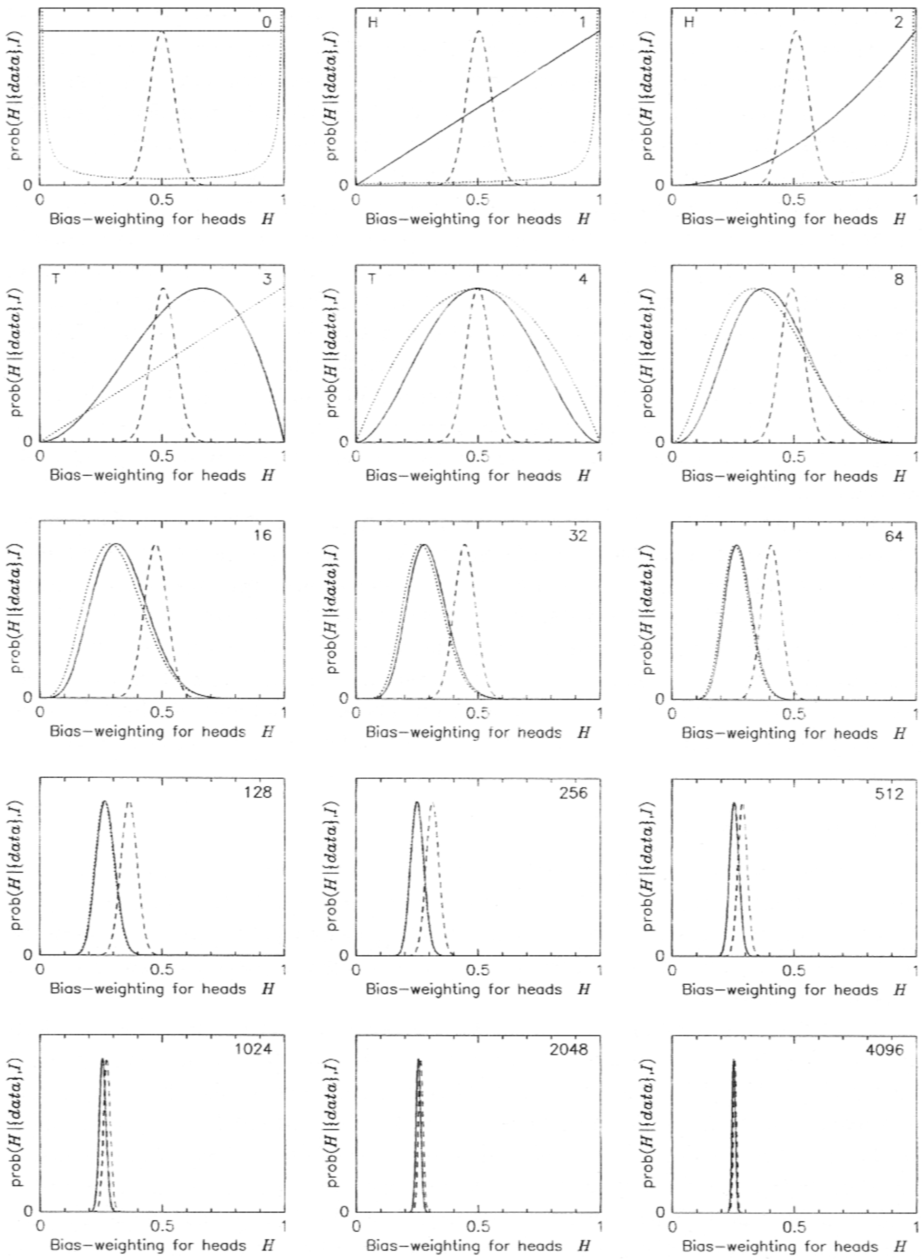

Fig. 2.2 The effect of different priors, prob$(H|I)$, on the posterior pdf for the bias-weighting of a coin. The solid line is the same as in Fig. 2.1, and is included for ease of comparison. The case for two alternative priors, reflecting slightly different assumptions in the conditioning information I, are shown with dashed and dotted lines.