Chapter 4
Bayesian Methods

Marco Ramoni* and Paola Sebastiani**
*Harvard University and **University of Massachusetts, Amherst, USA

4.1. Introduction

Classical statistics provides methods to analyze data, from simple descriptive measures to complex and sophisticated models. The available data are processed and then conclusions about a hypothetical population — of which the data available are supposed to be a representative sample — are drawn.

It is not hard to imagine situations, however, in which data are not the only available source of information about the population.

Suppose, for example, we need to guess the outcome of an experiment that consists of tossing a coin. How many biased coins have we ever seen? Probably not many, and hence we are ready to believe that the coin is fair and that the outcome of the experiment can be either head or tail with the same probability. On the other hand, imagine that someone would tell us that the coin is forged so that it is more likely to land head. How can we take into account this information in the analysis of our data? This question becomes critical when we are considering data in domains of application for which knowledge *corpora* have been developed. Scientific and medical data are both examples of this situation.

Bayesian methods provide a principled way to incorporate this external information into the data analysis process. To do so, however, Bayesian methods have to change entirely the vision of the data analysis process with respect to the classical approach. In a Bayesian approach, the data analysis process starts already with a given probability distribution. As this distribution is given *before* any data is considered, it is called *prior* distribution. In our previous example, we would represent the fairness of the coin as a uniform prior probability distribution, assigning probability 0.5 of landing to both sides of the coin. On the other hand, if we learn, from some external source of information, that the coin is biased then we can model a prior probability distribution that assigns a higher probability to the event that the coin lands head.

The Bayesian data analysis process consists of using the sample data to update this prior distribution into a *posterior* distribution. The basic tool for this updating is a theorem, proved by Thomas Bayes, an Eighteen century clergyman. The fundamental role of Bayes' theorem in this approach is testified by the fact that the whole approach is named after it.

The next section introduces the basic concepts and the terminology of the Bayesian approach to data analysis. The result of the Bayesian data analysis process is the posterior distribution that represents a revision of the prior distribution on the light of the evidence provided by the data. The fact that we use the posterior distribution to draw conclusions about the phenomenon at hand changes the interpretation of the typical statistical measures that we have seen in the previous chapters. Section 4.3 describes the foundations of Bayesian methods and their applications to estimation, model selection, and reliability assessment, using some simple examples. More complex models are considered in Section 4.4, in which Bayesian methods are applied to the statistical analysis of multiple linear regression models and Generalized Linear Models. Section 4.5 will describe a powerful formalism known as *Bayesian Belief Networks* (BBN) and its applications to prediction, classification and modeling tasks.

4.2. The Bayesian Paradigm

Chapters 2 and 3 have shown that classical statistical methods are usually focused on the distribution $p(\mathbf{y}|\boldsymbol{\theta})$ of data $\boldsymbol{y}$, where $p(\cdot|\boldsymbol{\theta})$ denotes either the probability mass function or the density function of the sample of n cases $\mathbf{y} = (y_1, \ldots, y_n)$ and is known up to a vector of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$. The information conveyed by the sample is used to refine this probabilistic model by estimating $\boldsymbol{\theta}$, by testing hypotheses on $\boldsymbol{\theta}$ and, in general, by performing statistical inference. However, classical statistical methods do not allow the possibility of incorporating external information about the problem at hand. Consider an experiment that consists of tossing a coin n times. If the results can be regarded as values of independent binary random variables Y_i taking values 1 and 0 — where $\theta = p(Y_i = 1)$ and $Y_i = 1$ corresponds to the event "head in trial i" — the likelihood function $L(\theta) = p(\mathbf{y}|\theta)$ (see Chapter 2) is

$$L(\theta) = \theta^{(\sum_i y_i)}(1-\theta)^{(n-\sum_i y_i)}$$

and the ML estimate of θ is

$$\hat{\theta} = \frac{\sum_i y_i}{n},$$

which is the relative frequency of heads in the sample. This estimate of the probability of head is only a function of the sample information.

Bayesian methods, on the other hand, are characterized by the assumption that it is also meaningful to talk about the conditional distribution of $\boldsymbol{\theta}$, given