Bayesian Inference

Chapter 1. Introduction and basic concepts

M. Concepción Ausín
Department of Statistics
Universidad Carlos III de Madrid

Master in Business Administration and Quantitative Methods
Master in Mathematical Engineering
Objective

The aim of this course is to introduce the modern approach to Bayesian statistics, emphasizing the computational aspects and the differences between the classical and Bayesian approaches. The course includes Bayesian solutions to real problems.
Recommended reading

- *Bayesian Data Analysis*, Second Edition
  - Andrew Gelman, John B. Carlin, Hal S. Stern and Donald B. Rubin
  - Chapman & Hall/CRC

- *Introducing Monte Carlo Methods with R*
  - Christian P. Robert and George Casella
  - Springer

  - Dani Gamerman and Hedibert F. Lopes
  - Chapman & Hall/CRC
More references

Introduction

- Bayesian statistics is based on probability laws to quantify the degree of belief in an hypothesis.

- Two different views of probability:
  - Frequentist or classical approach: The limit of the relative frequencies of an event in an infinite series of repetitions of the same experiment.
  - Bayesian approach: Statements about our state of knowledge only based on the available information.

- Bayes theorem is the basic formula to update our beliefs about a hypothesis once we have additional observed information.

- In the Bayesian approach, the parameters of a distribution are considered as random variables, while the data are fixed.
Introduction

• Given a prior probability about a hypothesis and the observed information, the Bayes rule is used to obtain the posterior probability which is the conditional probability on the observed evidence.

• In Bayesian statistics, unlike the classical approach, inference is always based on Probability Theory.

• The main criticism to Bayesian inference is the usual subjective choice of the prior probabilities.

• In practice, posterior probabilities can be obtained analytically only in simple problems.

• Realistic complex problems usually require computational intensive methods based on numerical approximations or simulation methods. In particular, Markov Chain Monte Carlo (MCMC) methods have been largely developed in the last decades.
### Basic concepts

Given a hypothesis, $H$, the Bayes theorem is used to update our beliefs about it once the data have been observed:

$$
Pr(H \mid Data) = \frac{Pr(Data \mid H)Pr(H)}{Pr(Data)}
$$

where:

- $Pr(H)$ is our prior belief or **prior probability** on the hypothesis.
- $Pr(H \mid Data)$ is the **posterior probability** for $H$ once the data have been observed.
- $Pr(Data \mid H)$ is the **likelihood** for the observed data given that $H$ is true.
- $Pr(Data)$ is the **marginal likelihood** for the observed data independently of $H$ being true or not.
Main steps in Bayesian inference

Suppose now that we are interested in making inference about a parameter, $\theta$, given a data sample, $x = \{x_1, \ldots, x_n\}$.

Using the Bayesian approach, $\theta$ is not a fixed value, but a random variable. Then, the main steps are:

1. Select a **prior distribution** for $\theta$, denoted by $\pi(\theta)$, which should reflect our prior beliefs about $\theta$ before observing the data.

2. Collect some data, $x$, assume a statistical model for them given $\theta$ and obtain the **likelihood**, $f(x \mid \theta)$.

3. Using Bayes’ rule, update our prior beliefs about $\theta$ by calculating the **posterior probability**:

$$
\pi(\theta \mid x) = \frac{f(x \mid \theta) \pi(\theta)}{f(x)}
$$
Main steps in Bayesian inference

The marginal likelihood,

$$f(x) = \int f(x | \theta) \pi(\theta) d\theta$$

is an integration constant to ensure that the posterior distribution of $\theta$ integrates up to one and does not depend on $\theta$.

Therefore, this constant does not provide any additional information about the posterior distribution and then, this is usually expressed by,

$$\pi(\theta | x) \propto f(x | \theta) \pi(\theta)$$

which means that the posterior distribution is proportional to the likelihood times the prior distribution.

Therefore, a 95% credible interval for $\theta$ is simply an interval $(a, b)$ such that the posterior probability that $\theta$ is in $(a, b)$ is equal to 0.95.
To sum up...

<table>
<thead>
<tr>
<th>Frequentist</th>
<th>Bayesian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability can model <strong>frequency</strong>: how often an event occurs</td>
<td>Probability can model our <strong>subjective beliefs</strong>.</td>
</tr>
<tr>
<td>We assume that experiments are repeatable.</td>
<td>Experiments cannot be repeatable.</td>
</tr>
<tr>
<td>The <strong>parameters</strong> of a distribution are fixed.</td>
<td>The <strong>parameters</strong> of a distribution are random variables.</td>
</tr>
<tr>
<td>The <strong>data</strong> is random: it is drawn from a fixed distribution.</td>
<td>The <strong>data</strong> is the only thing that is fixed &amp; immutable.</td>
</tr>
<tr>
<td>I have a 95% <strong>confidence</strong> that the population mean is between -2.2 and +0.3</td>
<td>There is a 95% <strong>probability</strong> that the population mean is between 12 and 13</td>
</tr>
<tr>
<td>If H₀ is true, we would only see such <strong>extreme results</strong> 3.2% of the time, so we can reject H₀ at the 5% level.</td>
<td><strong>Probability</strong> of H₀ is 0.26%.</td>
</tr>
</tbody>
</table>

Summary from Brian Potetz web page
In the next chapter, we will introduce simple models where the posterior distributions can be obtained analytically assuming certain prior distribution models, known as conjugate distributions.