

## Chapter 2.4

# Bayesian Modelling for Machine Learning

**Paul Rippon**

*The University of Newcastle, Australia*

**Kerrie Mengersen**

*The University of Newcastle, Australia*

### INTRODUCTION

Learning algorithms are central to pattern recognition, artificial intelligence, machine learning, data mining, and statistical learning. The term often implies analysis of large and complex data sets with minimal human intervention. Bayesian learning has been variously described as a method of updating opinion based on new experience, updating parameters of a process model based on data, modelling and analysis of complex phenomena using multiple sources of information, posterior probabilistic expectation, and so on. In all of these guises, it has exploded in popularity over recent years.

General texts on Bayesian statistics include Bernardo and Smith (1994), Gelman, Carlin, Stern, and Rubin (1995), and Lee (1997). Texts that derive more from the information science discipline, such as Mitchell (1997) and Sarker,

Abbass, and Newton (2002), also include sections on Bayesian learning.

Given recent advances and the intuitive appeal of the methodology, Bayesian learning is poised to become one of the dominant platforms for modelling and analysis in the 21st century. This article provides an overview of Bayesian learning in this context.

### BACKGROUND

#### Bayesian Modelling

Bayesian learning aims to provide information about unknown characteristics of a population (such as a mean and/or a variance) or about relationships between characteristics (for example, via a regression equation or a neural network). We often have a set of alternative models or

hypotheses,  $H_1, H_2, \dots, H_m$ , that could describe these unknowns, such as possible values for the unknown mean or alternative neural network representations. The Bayesian approach allows prior beliefs about these models to be updated in the light of new data. The fundamental enabling mechanism is Bayes' rule:

$$p(H_i | D) = \frac{p(D | H_i)p(H_i)}{p(D)}, \quad (1)$$

which states that the *posterior* probability  $p(H_i | D)$  of a particular model,  $H_i$ , conditional on data,  $D$ , is proportional to the probability  $p(D | H_i)$  of the data, given the model multiplied by the *prior* probability  $p(H_i)$  of the model. The denominator  $p(D)$ , a normalizing constant designed to make the posterior probability sum or integrate to one, can be termed the probability of the data and is expressed as:

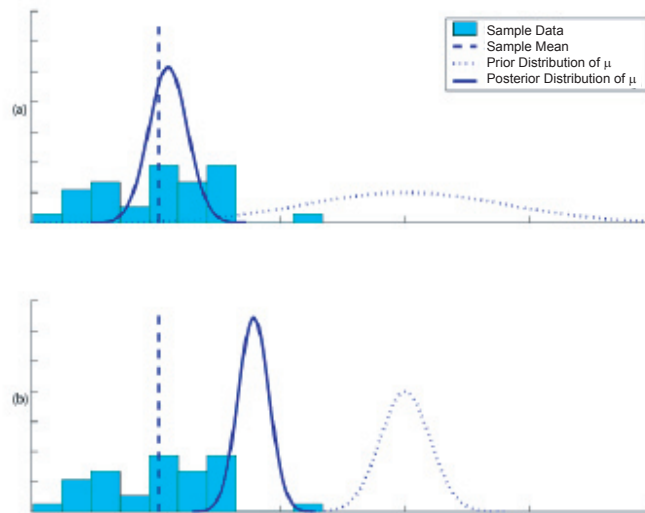
$$p(D) = \sum_{i=1}^m p(H_i)p(D | H_i).$$

The number of plausible models might be infinite, for example, when the different models are represented by unknown values of a continuously distributed population mean. In this case, probability distributions become densities and the summation in  $p(D)$  is replaced by an integral. In either case, it is this denominator,  $p(D)$ , that is often intractable. This motivates the development of numerical methods such as Markov chain Monte Carlo, described in the next section.

As a simple example, consider sampling  $n$  data points  $y_1, y_2, \dots, y_n$  from a population of normally distributed measurements in order to estimate an unknown mean,  $\mu$ , and assume that the population variance,  $\sigma^2$ , is known. Thus,  $H$  is the set of all possible values that  $\mu$  may take. The sample mean,  $\bar{y}$ , represents the information contained in the data so that  $p(D | H) = p(\mu) = N(\mu, \sigma^2/n)$ .

In practice, we often have some prior knowledge about  $\mu$ , such as, “ $\mu$  is known from experience to be around a value  $\mu_0$ .” We might express this prior knowledge as  $\mu \sim N(\mu_0, \tau_0^2)$ , where  $\tau_0^2$  represents the uncertainty around the best guess,  $\mu_0$ . Now, according to Bayes' rule:

Figure 1.



7 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/bayesian-modelling-machine-learning/24294](http://www.igi-global.com/chapter/bayesian-modelling-machine-learning/24294)

## Related Content

---

### Knowledge Management Within the Context of Organizational Innovation

M. Hanefi Calp (2020). *Advanced MIS and Digital Transformation for Increased Creativity and Innovation in Business* (pp. 26-49).

[www.irma-international.org/chapter/knowledge-management-within-the-context-of-organizational-innovation/237260](http://www.irma-international.org/chapter/knowledge-management-within-the-context-of-organizational-innovation/237260)

### A Novel Approach for Face Recognition under Varying Illumination Conditions

V Mohanraj, V. Vaidehi, S Vasuhiand Ranajit Kumar (2018). *International Journal of Intelligent Information Technologies* (pp. 22-42).

[www.irma-international.org/article/a-novel-approach-for-face-recognition-under-varying-illumination-conditions/205668](http://www.irma-international.org/article/a-novel-approach-for-face-recognition-under-varying-illumination-conditions/205668)

### Cloud-Based Big Data Analytics in Smart Educational System

Newlin Rajkumar Manokaran, Venkatesa Kumar Varathanand Shalinie Deepak (2018). *Deep Learning Innovations and Their Convergence With Big Data* (pp. 189-199).

[www.irma-international.org/chapter/cloud-based-big-data-analytics-in-smart-educational-system/186477](http://www.irma-international.org/chapter/cloud-based-big-data-analytics-in-smart-educational-system/186477)

### Cross-Layer Scheme for Meeting QoS Requirements of Flying Ad-Hoc Networks: QoS Requirements of Flying Ad-Hoc Networks

Bilal Muhammad Khanand Rabia Bilal (2017). *Handbook of Research on Recent Developments in Intelligent Communication Application* (pp. 255-281).

[www.irma-international.org/chapter/cross-layer-scheme-for-meeting-qos-requirements-of-flying-ad-hoc-networks/173247](http://www.irma-international.org/chapter/cross-layer-scheme-for-meeting-qos-requirements-of-flying-ad-hoc-networks/173247)

### Scaling Instant Messaging Communication Services: A Comparison of Blocking and Non-Blocking Techniques

Leigh Griffin, Kieran Ryan, Eamonn de Leastarand Dmitri Botvich (2012). *International Journal of Ambient Computing and Intelligence* (pp. 1-19).

[www.irma-international.org/article/scaling-instant-messaging-communication-services/68841](http://www.irma-international.org/article/scaling-instant-messaging-communication-services/68841)