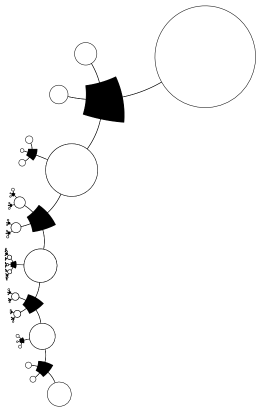
Part IV

Probabilities and Inference



About Part IV

The number of inference problems that can (and perhaps should) be tackled by Bayesian inference methods is enormous. In this book, for example, we discuss the decoding problem for error-correcting codes, the task of inferring clusters from data, the task of interpolation through noisy data, and the task of classifying patterns given labelled examples. Most techniques for solving these problems can be categorized as follows.

Exact methods compute the required quantities directly. Only a few interesting problems have a direct solution, but exact methods are important as tools for solving subtasks within larger problems. Methods for the exact solution of inference problems are the subject of Chapters 21, 24, 25, and 26.

Approximate methods can be subdivided into

1. **deterministic approximations**, which include maximum likelihood (Chapter 22), Laplace's method (Chapters 27 and 28) and variational methods (Chapter 33); and
2. **Monte Carlo methods** – techniques in which random numbers play an integral part – which will be discussed in Chapters 29, 30, and 32.

This part of the book does not form a one-dimensional story. Rather, the ideas make up a web of interrelated threads which will recombine in subsequent chapters.

Chapter 3, which is an honorary member of this part, discussed a range of simple examples of inference problems and their Bayesian solutions.

To give further motivation for the toolbox of inference methods discussed in this part, Chapter 20 discusses the problem of clustering; subsequent chapters discuss the probabilistic interpretation of clustering as mixture modelling.

Chapter 21 discusses the option of dealing with probability distributions by completely enumerating all hypotheses. Chapter 22 introduces the idea of maximization methods as a way of avoiding the large cost associated with complete enumeration, and points out reasons why maximum likelihood is not good enough. Chapter 23 reviews the probability distributions that arise most often in Bayesian inference. Chapters 24, 25, and 26 discuss another way of avoiding the cost of complete enumeration: marginalization. Chapter 25 discusses message-passing methods appropriate for graphical models, using the decoding of error-correcting codes as an example. Chapter 26 combines these ideas with message-passing concepts from Chapters 16 and 17. These chapters are a prerequisite for the understanding of advanced error-correcting codes.

Chapter 27 discusses deterministic approximations including Laplace's method. This chapter is a prerequisite for understanding the topic of complexity control in learning algorithms, an idea that is discussed in general terms in Chapter 28.

Chapter 29 discusses Monte Carlo methods. Chapter 30 gives details of state-of-the-art Monte Carlo techniques.

Chapter 31 introduces the Ising model as a test-bed for probabilistic methods. An exact message-passing method and a Monte Carlo method are demonstrated. A motivation for studying the Ising model is that it is intimately related to several neural network models. Chapter 32 describes 'exact' Monte Carlo methods and demonstrates their application to the Ising model.

Chapter 33 discusses variational methods and their application to Ising models and to simple statistical inference problems including clustering. This chapter will help the reader understand the Hopfield network (Chapter 42) and the EM algorithm, which is an important method in latent-variable modelling. Chapter 34 discusses a particularly simple latent variable model called independent component analysis.

Chapter 35 discusses a ragbag of assorted inference topics. Chapter 36 discusses a simple example of decision theory. Chapter 37 discusses differences between sampling theory and Bayesian methods.

A theme: what inference is about

A widespread misconception is that the aim of inference is to find *the most probable explanation* for some data. While this most probable hypothesis may be of interest, and some inference methods do locate it, this hypothesis is just the peak of a probability distribution, and it is the whole distribution that is of interest. As we saw in Chapter 4, the *most probable* outcome from a source is often not a *typical* outcome from that source. Similarly, the most probable hypothesis given some data may be atypical of the whole set of reasonably-plausible hypotheses.

About Chapter 20

Before reading the next chapter, exercise 2.17 (p.36) and section 11.2 (inferring the input to a Gaussian channel) are recommended reading.