Preface

This book is our attempt to synthesize theoretical, methodological and computational aspects of Bayesian methods for infinite-dimensional models, popularly known as Bayesian nonparametrics. While Bayesian nonparametrics existed in an informal fashion as far back in time as Henri Poincaré, it became a serious methodology only after the introduction of the Dirichlet process by Thomas Ferguson in 1973. The development of computing algorithms especially suited for Bayesian analysis in the 1990s together with the exponential growth of computing resources enabled Bayesian nonparametrics to go beyond the simplest problems and made it a universally applicable paradigm for inference. Unfortunately, examples showed that seemingly natural priors can lead to uncomfortable properties of the posterior distribution, making a systematic study of the behavior of posterior distributions of infinite-dimensional parameters essential. The framework for such a study is frequentist. It assumes that the data are generated according to some “true” distribution, and the question is whether and how well the posterior distribution can recover this data-generating mechanism. This is typically studied when the sample size increases to infinity, although bounds for finite sample sizes are obtained implicitly within the analysis. Many significant results on such frequentist behavior of the posterior distribution have been obtained in the past 15–20 years building upon the seminal work of Lorraine Schwartz published in 1965. A fairly complete theory of the large sample behavior of the posterior distribution for dominated experiments is now presentable, covering posterior consistency, contraction rates, adaptation, and distributional approximations. A primary focus of this book is to describe this theory systematically, using examples of model and prior combinations, and to convey a general understanding of what type of priors make a Bayesian nonparametric method work. It is manifest that we emphasize positive aspects of Bayesian methods for infinite-dimensional models. We state and prove theorems that give precise conditions that ensure desirable properties of Bayesian nonparametric methods. Proofs are almost always given completely, often generalized or simplified compared with the existing literature. Historical notes at the end of each chapter point out the original sources. Some results improve on the literature or appear here for the first time.

The book is intended to give a treatment of all aspects of Bayesian nonparametrics: prior construction, computation and large sample behavior of the posterior distribution. It is important to consider these aspects together, as a prior is useful only if the posterior can be computed efficiently and has desirable behavior. These two properties of posteriors can thus throw light on how to construct useful priors and what separates a good prior from a not-so-good one. An introductory chapter describes the goals, advantages and difficulties of nonparametric Bayesian inference. Chapters 2 and 3 deal with methods of prior constructions on functions and distributions. In Chapter 4 the Dirichlet process is introduced and and studied. This continues in Chapter 5 with kernel mixtures with the Dirichlet process prior on the mixing distribution. This chapter also describes tools for computing the corresponding posterior distribution, such as Markov chain Monte Carlo techniques and variational methods. Posterior consistency theory is developed in Chapter 6 and applications are discussed in Chapter 7. Chapter 8 refines the theory to contraction rates, and applications are presented in Chapter 9. In Chapter 10 on Bayesian adaptation it is shown that an optimal contraction rate is often obtainable without knowing the complexity of the true parameter, by simply putting a good prior on this complexity. Related results on the behavior of Bayes factors are also presented in this chapter. Chapter 11 deals with properties of Gaussian process priors and the resulting posteriors for various inference problems, and gives an overview of methods for posterior computation. Bernstein–von Mises-type results for infinite-dimensional models are studied in Chapter 12. Priors based on independent increment processes, applied to survival models, are the subject of Chapter 13. As the corresponding models are not dominated, Bayes's theorem does not apply directly, so posterior theory and large sample theory is developed based on conjugacy. Finally Chapter 14 studies properties of discrete random distributions and the partition structures that arise when sampling from such distributions. Dependencies between the chapters are shown in Figure 1.

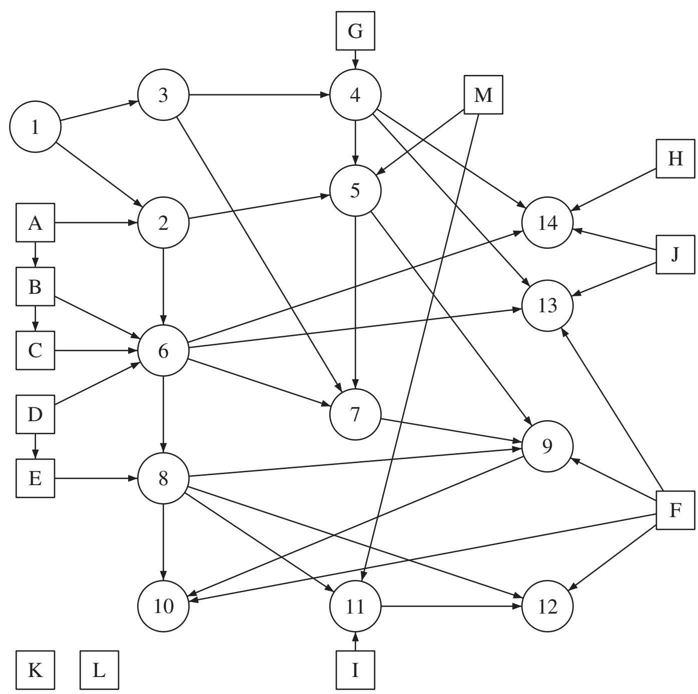

Figure 1 Dependencies between chapters. A chapter at the origin of an arrow must be read before the chapter at its end. Chapters from the main body of the book (given by numbers) are represented by circles, whereas chapters from the appendix (with letters) are given by squares.

Although the asymptotic behavior of posterior distributions is a main focus of the book, a considerable part of the book concerns properties of priors, posterior updating, and computation. This is true in particular for Chapters 1–5 and the biggest part of Chapters 13–14. Thus even those who (surprisingly) are not interested in large sample studies may find much useful material. The only aspect we do not cover is the application of the methods on specific data sets. There are several excellent books dedicated to applied Bayesian nonparametrics, for instance Dey et al. (1998) and the recent publication Müller et al. (2015).

Bayesian nonparametrics continues to develop, so that this book cannot be comprehensive. Recent developments that are not covereed or are barely covered are posterior contraction rates under stronger norms, Bayesian analysis of inverse problems, including deconvolution problems, the performance of empirical Bayes procedures and nonparametric Bernstein–von Mises results. Perhaps the biggest omission regards the coverage of (adaptive) credible regions, which is central to justifying (or not) the use of the full posterior distribution in a truly Bayesian sense. All these topics are still actively developing and their theory is incomplete at the present. Even without these topics the book has swelled considerably in size, more than we imagined in the beginning. Notwithstanding this growth, many interesting results within the book's scope had to be omitted because of space (and time) constraints. Problem sections at the end of each chapter describe many of the omitted results, with an original source cited. While this hardly does justice to the omitted results, readers can at least access the statements without going to the original sources. Most of these problems should be interpreted as pointers to supplementary reading rather than simple exercises. At the end of each chapter we also give a brief account of the historical development of the material, to the extent known to us. We apologize for (unintended) omissions, and any failures to give proper credit.

Reading the book requires familiarity with graduate-level mathematics, although some chapters and sections are more demanding in this respect than others. Probability theory and real analysis are needed almost throughout the book, and familiarity with measure theory is assumed, at least to the extent of understanding abstract integration. Measurability issues are addressed properly, but readers not concerned about measurability can typically skip these portions of the text and follow the rest of the material. Functional analysis beyond the conception of function spaces is not heavily used, except in the chapters on Gaussian processes and the Bernstein–von Mises theorem, and in the section on misspecification. Mathematical rigor is maintained, and formal proofs of almost all results are given. For the reader's convenience, we collected in 13 appendices many results, typically with proofs, which are needed in the main body of the book but not directly part of Bayesian nonparametrics. This concerns mathematical and probabilistic tools such as distances on measures, entropy, function approximation theory, Gaussian processes and completely random measures, as well as statistical tools from the theory of testing. Often prior reading of the relevant section of the appendix is advisable for better understanding a chapter or section in the main body of the book. The diagram in Figure 1 shows which sections of the appendix are used in which chapters in the main body.

One goal for writing this book was to provide a text appropriate for teaching a one-semester course on Bayesian nonparametrics. In our experience, Sections 1.2, 2.1–2.5, 3.2–3.5, 3.7, 4.1, 4.3, 4.5, 4.7, 5.1–5.3, 5.5, 6.1, 6.4, 6.7.1, 7.1.1, 7.1.2, 7.2.1, 7.4.1, 8.1, 8.2, 8.3.1, 9.1, 9.4 (without proof), 10.1, 10.3, 10.4.1, 10.5, 10.6.2, 11.1–11.3, 13.1, 13.3, 13.4, 13.4.1, 14.1, 14.2, 14.4, along with the relevant prerequisites from the appendix provide a good basis for a primarily theoretical course with some material omitted based on judgement. For a more methodology-oriented course, asymptotic properties other than the most basic consistency results may be omitted, and the material may be supplemented by the additional computational techniques in Section 11.7 and data applications from a good book on applied Bayesian nonparametrics.

The first draft of the book goes back nearly a decade, and parts of the book were written even earlier. It goes without saying that the landscape of Bayesian nonparametrics changed continually. The writing was nourished by opportunities to deliver graduate courses and short courses. The first author taught two courses at North Carolina State University as well as short courses at the Institut de Mathématiques de Luminy, France (2006), at the "Nonparametric Bayesian Regression" meeting in Cambridge, UK (2007), at Bilkent University, Ankara, Turkey (2007), and at Eurandom, Eindhoven, The Netherlands (2011), based on the drafts of the book available at the corresponding time. The second author used parts of the manuscript for master courses in the Netherlands and several short courses in Germany.

This book would not have materialized without contributions from various people. A large chunk of the presented material comes from our work with co-authors (including student co-authors): Eduard Belitser, Ismaël Castillo, Jayanta Ghosh, Jiezhun Gu, Bas Kleijn, Bartek Knapik, Willem Kruijer, Juri Lember, R.V. Ramamoorthi, Judith Rousseau, Anindya Roy, Weining Shen, Botond Szabó, Yongqiang Tang, Surya Tokdar, William Weimin Yoo, Yuefeng Wu, Harry van Zanten. Working with them has always been enjoyable and furthered our understanding of Bayesian nonparametrics. Igor Prünster was kind enough to read first drafts of the last two chapters of the manuscript. He provided us with many important references, answered many questions and supplied several missing proofs, especially in the last chapter. It would be difficult to imagine that chapter taking shape without his tireless help. S. M. Srivastava helped us to clarify matters related to descriptive set theory. Ismaël Castillo, Kolyan Ray and Botond Szabó read portions of the manuscript and provided useful resources. William Weimin Yoo and Julyan Arbel carefully read parts of the manuscript and pointed out many typos and small mistakes. Discussions with Peter Müller, Surya Tokdar, Yongdai Kim, Antonio Lijoi and Sasha Gnedin were instrumental in resolving many questions. The book would not have been possible without face-to-face meeting between us through occasional visits. Leiden University, VU Amsterdam, North Carolina State University and Eurandom hosted these visits. Funding from the Netherlands Organisation for Scientific Research (NWO), the Royal Netherlands Academy of Sciences (KNAW), the European Research Council (ERC Grant Agreement 320637), and the US National Science Foundation (NSF) were instrumental in supporting some of these visits, and also contributed to the research in general. We are indebted to the staff of Cambridge University Press, particularly Diana Gillooly, for their help, patience and constant encouragement throughout the long process.

Subhashis Ghosal
Aad van der Vaart