

# Preface

The foundation of the subject of nonparametric Bayesian inference was laid in two technical reports: a 1969 UCLA report by Thomas S. Ferguson (later published in 1973 as a paper in the *Annals of Statistics*) entitled “A Bayesian analysis of some nonparametric problems”; and a 1970 report by Kjell Doksum (later published in 1974 as a paper in the *Annals of Probability*) entitled “Tailfree and neutral random probabilities and their posterior distributions”. In view of simplicity with which the posterior distributions were calculated (by updating the parameters), the Dirichlet process became an instant hit and generated quite an enthusiastic response. During the decades of 1970s and 1980s, hundreds of papers were published in developing nonparametric Bayesian procedures to handle many inferential problems. These publications may be considered as “pioneers” in championing the Bayesian methods and opening a vast unexplored area in solving nonparametric problems. A review article (Ferguson et al. 1992) summarized the progress of the two decades. However, the paper was not meant to provide details but just an overview. Moreover, since then several new prior processes and their applications have appeared in technical publications. Also in the last decade there has been a renewed interest in the applications of variants of the Dirichlet process in modeling large scale data (see for example the recent papers by Chung and Dunson 2011, and Rodriguez et al. 2010 and references cited therein; and a volume of essays “Bayesian Nonparametric” edited by Hjort et al. 2010). For these reasons there seems to be a need for a single source of the material published on this topic during the earlier decades. This is a prime motivator for undertaking the present task.

The objective of this monograph is to assemble and consolidate the scattered material on various prior processes, their properties and their numerous applications, in solving Bayesian inferential problems based on data that may possibly be right censored, sequential or quantal response data. Emphasis is placed on the Dirichlet process as well as other prior processes that have been discovered through 1990s and their applications. We anticipate that it would serve as a one-stop resource for future researchers. In that spirit, first various processes are introduced and their properties are stated. Thereafter, the focus is to present various applications in estimation of distribution and survival functions, estimation of density functions and hazard rates,

empirical Bayes, hypothesis testing, covariate analysis, and many other applications. A major requirement of Bayesian analysis is its analytical tractability. Since the Dirichlet process possesses the conjugacy property, it has simplicity and ability to get results in a closed form. Therefore, most of the applications that were published soon after Ferguson's paper, are based on the Dirichlet process. Unlike the trend in recent years where computational procedures are developed to handle large and complex data sets, the earlier procedures relied mostly on developing procedures in closed forms.

In addition, several new and interesting processes, such as, the Chinese restaurant process, Indian buffet process, and hierarchical processes have been introduced in the last decade with an eye toward applications in the fields outside mainstream statistics, such as machine learning, ecology, document classification, etc. Similarly, dependent and spatial Dirichlet processes are proposed to incorporate covariates and handle random effects models. They have roots in the Ferguson-Sethuraman infinite sum representation of the Dirichlet process and shed new light on the robustness of this approach. They are included here without going into much details but a long list of references is included for the reader to explore relevant areas of interest further.

This material is an outgrowth of my lecture notes developed during the week long lectures I gave at Zhongshen University in China in 2007 on this topic, followed by lectures at universities in India, Singapore and Jordan. Obviously, the choice of material included and the style of presentation solely reflects my preferences. This manuscript is not expected to include all the applications, but references are given, wherever possible for additional applications. The mathematical rigor is limited as it has already been dealt with in the theoretical book by Ghosh and Ramamoorthi (2003). Therefore, many theorems and results are stated without proofs and the questions regarding existence, consistency and convergences are skipped. To conserve space, numerical examples are not included but referred to the papers originating those specific topics. For these reasons, the notations of the originating papers are preserved so that the reader may find it easy to migrate to the original publications as needed.

Computational procedures that make nonparametric Bayesian analysis feasible when closed forms of solutions are impossible or complex, are becoming increasingly popular in view of the availability of inexpensive and fast computation power. In fact they are indispensable tools in modeling large scale and high dimensional data. There are numerous papers published in the last two decades that discuss them in great details and algorithms are developed to simulate the posterior distributions so that the Bayesian analysis can proceed. These aspects are covered extensively in books by Ibrahim et al. (2001) and Dey et al. (1998). To avoid duplication, they are not discussed here. Some newer applications are also discussed in the book of essays edited by Hjort et al. (2010). We refer the reader to these books. The papers by Chung and Dunson (2011) and Rodriguez et al. (2010) and references cited therein, should also prove useful in this regard.

Since this book discusses various prior processes, their properties and inferential procedures in solving problems encountered in practice, it is ideal to serve as

a comprehensive introduction to the subject of nonparametric Bayesian inference. It is to be considered as a complement to the book authored by Ghosh and Ramamoorthi (2003) but at a less rigorous level. It may be viewed as something in between their theoretical book and the books by Ibrahim et al. (2001) and Dey et al. (1998).

The first chapter is devoted to introducing various prior processes, their formulation and their properties. The sequencing of these priors reflects mostly the order in which they were developed. The Dirichlet process and its immediate generalizations are presented first. The neutral to the right processes and the processes with independent increments, which form the basis for other processes are discussed next. They are key in the development of processes that include beta, gamma and extended gamma processes, which are proposed primarily to address specific applications in the reliability theory. Beta-Stacy process which generalizes the Dirichlet process is discussed thereafter. Following that, tailfree and Polya tree processes are presented which are especially convenient for estimating density functions, and to place greater weights, where it is deemed appropriate, by selecting suitable partitions in developing the prior. Lijoi and Prünster's (2010) recent paper tie many of these processes in presenting a general unifying framework in terms of the completely random measures (Kingman 1967). Finally, some additional processes that have been discovered in recent years (mostly variants of existing processes) and found to be useful in practice are mentioned. They have origin in the Ferguson-Sethuraman infinite sum representation in which the weights are constructed by a stick-breaking construction. They are collectively called here as *Ferguson-Sethuraman processes* and include dependent and spatial Dirichlet processes, Pitman-Yor process, Chinese restaurant and Indian buffet processes, etc.

The second chapter contains various applications that cover multitudes of fields such as, estimation, hypothesis testing, empirical Bayes, density estimation, bioassay, etc. They are grouped according to the inferential task they signify. Since, a major part of efforts have been devoted to the estimation of the distribution function and its functional, they receive significant attention. This is followed by confidence bands, two-sample problems and other applications.

The third chapter is devoted to presenting inferential procedures based on censored data. Heavy emphasis is given to the estimation of survival function since it plays an important role in the survival data analysis. Estimation procedures based on different priors and under various sampling schemes are also included. This is followed by other examples which include estimation procedures in certain stochastic process models, Markov Chains, and competing risks models. Finally, estimation of the survival function in the presence of covariates is presented.

Since this book avoids deeper technical details, it should therefore be accessible to first time researchers and graduate students venturing into this interesting, fertile and promising field. As evident by the recent increased interest in using nonparametric Bayesian methods in modeling data, the field is wide open for new entrants. As such, it is my hope that this attempt will serve the purpose it was intended for, namely, to make such techniques readily available via this comprehensive but sim-

ple monograph. At the least, the reader will gain familiarity with many successful attempts in solving nonparametric problems from a Bayesian point of view in wide ranging areas of applications.

Wayne, USA

Eswar G. Phadia



<http://www.springer.com/978-3-642-39279-5>

Prior Processes and Their Applications

Nonparametric Bayesian Estimation

Phadia, E.G.

2013, XIV, 207 p., Hardcover

ISBN: 978-3-642-39279-5