New Developments on Bayesian Bootstrap for Unrestricted and Restricted Distributions

Reyhaneh Hosseini

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Mathematics[1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

[1]The Ph.D. program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Abstract

The recent popularity of Bayesian inference is due to the practical advantages of the Bayesian approach. The Bayesian analysis makes it possible to reflect ones prior beliefs into the analysis. In this thesis, we explore some asymptotic results in Bayesian nonparametric inference for restricted and unrestricted space of distributions. This thesis is divided into two parts. In the first part, we employ the Dirichlet process in a hypothesis testing framework to propose a Bayesian nonparametric chi-squared goodness-of-fit test. Our suggested method corresponds to Lo's Bayesian bootstrap procedure for chi-squared goodness-of-fit test. Indeed, our bootstrap rectifies some shortcomings of regular bootstrap which only counts number of observations falling in each bin in contingency tables. We consider the Dirichlet process as the prior for the distribution of data and carry out the test based on the Kullback-Leibler distance between the Dirichlet process posterior and the hypothesized distribution. We prove that this distance asymptotically converges to the same chi-squared distribution as the classical frequentist's chi-squared test. Moreover, the results are generalized to the chi-squared test of independence for contingency tables.

In the second part, our main focus is on Bayesian nonparametric inference for a restricted group of distributions called spherically symmetric distributions. We describe a Bayesian nonparametric approach to make inference for a bivariate spherically symmetric distribution. We place a Dirichlet invariant process prior on the set of all bivariate spherically symmetric distributions and derive the Dirichlet invariant process posterior. Indeed, our approach is an extension of the Dirichlet invariant process for the symmetric distributions on the real line to bivariate spherically symmetric distribution where the underlying distribution is invariant under a finite group of rotations. Further, we obtain the Dirichlet invariant process posterior for the infinite transformation group and we prove that it approaches a certain Dirichlet process. Finally, we develop our approach to obtain the Bayesian nonparametric posterior distribution for functionals of the distribution's support when the support satisfies certain symmetry conditions. When symmetry holds with respect to the parallel lines of axes (for example, in two dimensional space $x = a$ and $y = b$) we employ our approach to approximate the distribution of certain functionals such as area and perimeter for the support of the distribution. This suggests a Bayesian nonparametric bootstrapping scheme. The estimates can be derived based on posterior averaging. Then, our simulation results demonstrate that our suggested bootstrapping technique improves the accuracy of the estimates.

Dedications

To my parents Molouk and Esmaeil

You mean the world to me, but I dont tell you enough.

Acknowledgement

I would first like to thank my thesis supervisor Professor Mahmoud Zarepour for his valuable guidance and patience. Without his supervision and constant help, this thesis would not have been possible to be completed.

I wish to acknowledge and thank the University of Ottawa and the Department of Mathematics and Statistics for providing me with various financial opportunities and a great academic environment during my research process.

Finally, I must express my profound gratitude to my mother Molouk, my father Esmaeil, my sister Raheleh and my brother Ali for their love, unfailing support, patience, sacrifice and understanding. Thank you deeply.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The usage of Bayesian nonparametric statistics is rapidly becoming popular among the practitioners. Recently, the application of theoretical properties of Bayesian nonparametric methods has received noticeable attention in various fields of study like biomedical, environmental, econometric and many other areas. The use of Bayesian nonparametric priors like Indian buffet process (IBP) and Gaussian processes (GPs) in machine learning and deep neural network analysis are some examples. Lu and Tang (2015) used Gaussian processes as the models in face recognition analysis based on the deep learning techniques. Moreover, Griffiths and Ghahramani (2011) applied the Indian buffet process in latent feature modelling and learning overlapping clusters. The Dirichlet processes (DPs) as the most popular Bayesian nonparametric priors have a variety of applications in clustering for gene expression data analysis, see Medvedovic and Sivaganesan (2002) and Rasmussen et al. (2009), for instance. In addition, the innovative techniques of model selection and developing the Bayesian equivalents of hypothesis testing are extremely desirable.

This thesis is divided into two parts. In the first part, we suggest a Bayesian nonparametric chi-squared goodness-of-fit test based on the Kullback-Leibler distance between the Dirichlet process posterior and the hypothesized distribution. There are many one-sample and two-sample parametric goodness-of-fit tests in the literature. See for example, D'Agostino (1986) for a review. The chi-squared test examines whether the data has a specified distribution F_0, i.e., the null hypothesis is given as $H_0 : F = F_0$, where F_0 is the true distribution for the observed data. Some extensions of chi-squared goodness-of-fit test to Bayesian model assessment where the test statistic is based on the posterior distribution, are described by Johnson (2004, 2007).

The aim is to introduce a Bayesian nonparametric counterpart for the regular chi-squared goodness-of-fit test. The Bayesian nonparametric analysis of the problems requires the construction of proper prior distributions on infinite dimensional spaces and computation of the posterior distribution. The Dirichlet process introduced by Ferguson (1973) perhaps is the most popular prior in Bayesian nonparametric statistics which is defined as a distribution over the space of probability distributions. In fact, the Dirichlet prior is the empirical process of the Bayesian paradigm. In other words, in the Bayesian paradigm, the use of the plug-in principle of a frequentist should be modified to the use of the Dirichlet process posterior. There are extensive applications of the Dirichlet process in different areas of statistical inference. Some common applications of the Dirichlet process are in density estimation, clustering via mixture models and hypothesis testing. See for example, Neal (1992), Lo (1984) and Escobar and West (1995).

In Bayesian nonparametric inference, there are two strategies for goodness-of-fit tests. The first strategy considers a prior for the true distribution of data and constructs the test based on the distance between the posterior distribution and the proposed one. For example, Muliere and Tardella (1998), Swartz (1999), Al Labadi and Zarepour (2013, 2014) considered the Dirichlet process prior and the Kolmogorov distance. Al Labadi and Zarepour (2014) and Al Labadi and Zarepour (2017) carried out a goodness-of-fit test and a two-sample goodness-of-fit test, respectively by placing the Dirichlet process as a prior and constructing the test statistic based on the Kolmogorov distance. Viele (2000) used the Dirichlet process and the Kullback-Leibler distance for testing discrete distributions. Hsieh (2013) considered the Polya tree model as the prior and measured the Kullback-Leibler distance for testing the continuous distributions. The second strategy is conducted by embedding the hypothesized model H_0 in an alternative model H_1 and placing a prior on that. To examine the hypothesized model, the Bayes factor is used as a measure of evidence against the hypothesized model. For example, Florens et al. (1996) used a Dirichlet process prior for the alternative model. Tokdar and Martin (2011) carried out a Bayesian test for normality by considering a Dirichlet process mixture for the alternative model. Some authors used other Bayesian nonparametric priors. For instance, Holmes et al. (2015) described a Bayesian nonparametric two sample hypothesis testing based on a Polya tree prior. In order to test for the normal distribution, Berger and Guglielmi (2001) considered a mixture of Polya trees for the alternative model distribution, while Verdinelli and Wasserman (1998) suggested a mixture of Gaussian processes.

Our new proposed chi-squared goodness-of-fit test is based on the first approach discussed above. We place a Dirichlet process prior on the distribution of the observed data and define the chi-squared test statistic based on the Kullback-Leibler distance between the Dirichlet process posterior and the hypothesized distribution. In fact, in our Bayesian nonparametric approach, the test proceeds by constructing the chi-squared test statistic based on the distance between the observed probabilities obtained by the Dirichlet process posterior and the expected probabilities. Indeed, instead of counting the observed frequencies in each bin, we place a prior on the distribution of data. The probability of each bin is obtained by the exact posterior probability of that bin. Then, our new test statistic compares the posterior probabilities with the probabilities under the null hypothesis. In this procedure, based on the suggested Dirichlet prior, we know the exact distribution of the test statistic. This idea reminds us with some similarity to Fisher's exact test for contingency tables.

It is important to emphasize that placing Dirichlet prior on the unknown distribution corresponds to Bayesian bootstrap. In fact, our method is equivalent to the Bayesian bootstrap that was introduced first by Rubin (1981) and later was developed more carefully in Lo (1987). Throughout this thesis, we use both terminology as they are equivalent. Like Lo (1987), the asymptotic theory is also provided only for justification of the consistency of our method and the asymptotic result was never used directly and is discussed only for the purpose of verification. Let $\mathbf{X} = (X_1, \ldots, X_n)$, where $X_1, \ldots, X_n$ is a sequence of i.i.d. random variables having an unknown distribution F. Suppose $\theta(F, \mathbf{X})$ is a functional of both F and $\mathbf{X}$. Rubin (1981) suggested that for a given specific functional $\theta(D_n, \mathbf{X})$, the conditional distribution of $(\theta(D_n, \mathbf{X}) \mid \mathbf{X} = \mathbf{x})$ estimates the posterior distribution of $(\theta(F, \mathbf{X}) \mid \mathbf{X} = \mathbf{x})$, where D_n follows the Dirichlet distribution with parameters $(1, 1, \ldots, 1)$. This choice of prior makes Rubin's bootstrap somehow similar to regular bootstrap as $E(D_n) = (1/n, \ldots, 1/n)$. Lo (1987) proved the validity of the Bayesian bootstrap in general case where instead, the prior is a Dirichlet process and the posterior distribution (the updated Dirichlet process; the posterior) is used to evaluate the distribution of $(\theta(F, \mathbf{X}) \mid \mathbf{X} = \mathbf{x})$. This method provides a more flexible prior, as statisticians can influence their inference by their existing prior knowledge and also the strength of their knowledge in their prior beliefs. For comparison of different types of bootstrap and Bayesian bootstrap see Muliere and Secchi (1996). Using a similar approach, we extend our results to the chi-squared test of independence for contingency tables and finally we determine a proper concentration parameter for the Dirichlet process prior.

In the second part of this thesis, we narrow the space of all probability distribution to a restricted group of distributions called symmetric distributions. In fact, symmetry is a highly prominent feature that has an essential role in many areas of study. Spherical symmetry and symmetry with respect to straight lines are two important examples of symmetry. Further, these two cases of symmetry play a key role in various areas of human studies like biometric identification, iris recognition and face recognition. Indeed, the human's iris and face are the examples of the circular symmetry and symmetry with respect to a straight line (i,e, symmetry with respect to the line $y = b$ in mathematical consideration), respectively. Then, the iris and face recognition involves different statistical and mathematical modelling. These models are constructed on the basis of the circular symmetry of the iris and the midline symmetry of the face and the symmetric distance of facial key points. Figures 1.1 and 1.2 demonstrate some of these key points in the face and eye. Therefore, for a statistical modelling based on these data points, considering the probability distribution functions holding these symmetric properties is crucial.

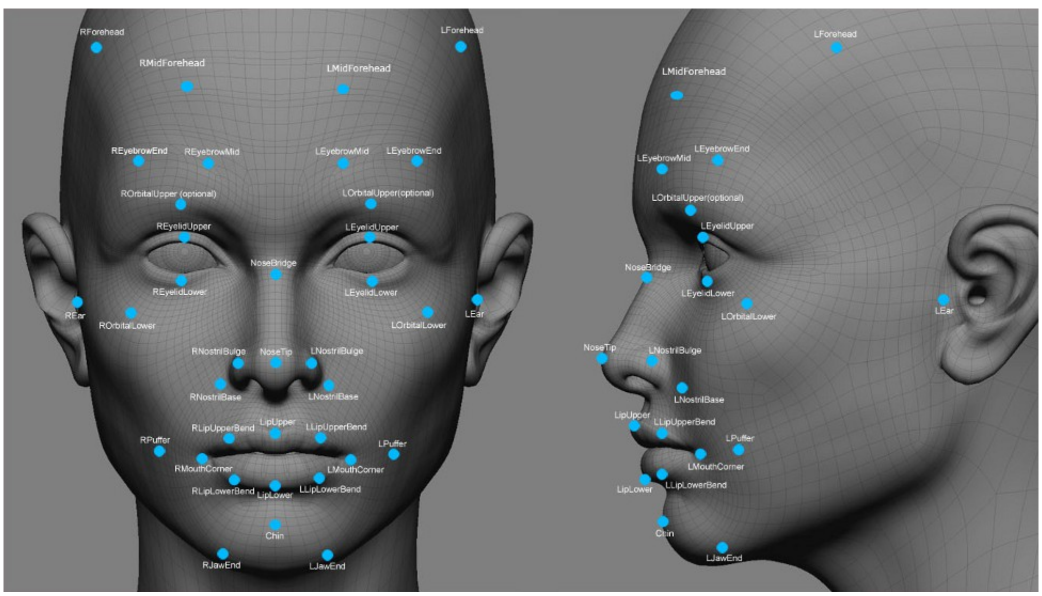

Figure 1.1: Some demonstrated key points around the midline symmetry of the face which are used in face recognition.(Photo from EndTheLie.com)

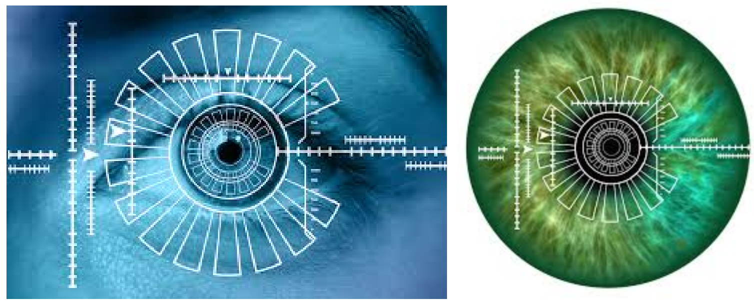

Figure 1.2: The circular symmetry of the iris used in iris recognition.

Then, the goal of the second part of this thesis is to provide a Bayesian nonparametric inference for these distributions. First, we describe a Bayesian nonparametric approach to draw an inference for a bivariate spherically symmetric distribution. As we mentioned earlier, the Dirichlet process is defined as the prior over the space of all probability measures. Antoniak (1974) also proposed the mixtures of Dirichlet processes which are extensively applied as a prior in bioassay and regression problems. These Bayesian nonparametric priors can be used in the case when the subset of interest is the set of all probability measures on a given probability space.

However, in many situations, based on our assumptions, we may need to consider a certain class of priors. In other words, it is required to tackle priors that are focused on smaller subsets. For instance, consider the problem of estimating some functional of a symmetric distribution, for example., the median, the mode of a symmetric distribution or a test for independence assuming permutation symmetry, exchangeability or spherical symmetry. Here, we need to use the priors on the class of distributions which are restricted to the property mentioned above.

Therefore, for a valid Bayesian formulation of the problems in this framework, where the underlying distributions are invariant under a specific group of transformations, employing the priors with invariant sample paths will be more reliable. Dalal (1979a) introduced a class of random processes with invariant sample paths called Dirichlet invariant processes. These processes which are directly related to

Dirichlet processes, generate probability distributions that are invariant under a specified transformation group. Specifically, let $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$ be any p-dimensional Euclidean space with the associated Borel σ-field. Let $\mathscr{G} = \{g_1, ..., g_k\}$ be any finite group of measurable transformations $\mathscr{X} \to \mathscr{X}$. Dalal (1979a) constructed a random probability measure such that for any set B, the probability of the events B and $g_j(B) = \{g_j(x) : x \in B\}$, $(j = 1, ..., k)$ are identical. In other words, Dirichlet invariant processes consider a subset of the set of all probability measures on a given probability space.

Suppose that we have some knowledge about the invariant properties of the distribution of interest. In this case, instead of using the Dirichlet process as the prior for the target distribution, based on our knowledge and assumptions, we can use a proper Dirichlet invariant process. Indeed, our knowledge as an invariant group will be embedded in Dirichlet invariant process prior. Similar to the Dirichlet process, these processes can be used as priors for Bayesian analysis of a wide range of problems. Fore instance, Yamato (1986, 1987) derived the Bayes estimates in the sample case by evaluating the expectation of random functional of a Dirichlet invariant process. Dalal (1979a) also obtained the Bayes estimator for the symmetric distribution with a known center of symmetry. Moreover, Dalal (1979b, 1980) employed the Dirichlet invariant process for estimating the unknown center of symmetry for the symmetric distributions.

Specifically, consider the set of all probability measures on real line $\mathbb{R}$ which are symmetric about μ. Dalal (1979a) discussed symmetry as the invariant property by defining the transformation group $\mathscr{G} = \{g_1, g_2\}$ with $g_1 = x$ and $g_2 = 2\mu - x$. Then, by using Dirichlet invariant process as the prior over the set of all univariate symmetric distributions with a known center of symmetry, the corresponding Dirichlet invariant process posterior was computed. We extend the Dirichlet invariant process to develop a prior over the set of spherically symmetric distributions and derive the corresponding posterior process for this class of distributions. As a special case, we place a Dirichlet invariant process prior over the space of bivariate spherically symmetric distributions, i.e, the underlying distribution is invariant under rotations. We first begin with a finite group of rotations $\mathscr{G}$ such that $\mid \mathscr{G} \mid = k$. Then, we consider the case when $k \to \infty$ and we prove that our Dirichlet invariant process converges weakly to Dirichlet process.

Finally, we carry out the simulation studies to illustrate the sample posterior paths generated from the Dirichlet invariant process for spherically symmetric distributions. We also compare our results with the paths obtained by Dirichlet process. Furthermore, we extend our approach to obtain the Bayesian nonparametric posterior paths for distributions which are symmetric with respect to the parallel lines of axes, i.e, $x = a$ and $y = b$. An algorithm for explaining the computation of the posterior distribution is also provided. Let K denote the support of the symmetric distribution F where symmetry holds with respect to the parallel lines of axes. As an application, we provide an asymptotically valid bootstrap to approximate the distribution of area and perimeter of K as two functionals of the support. This approximation has a variety of practical applications in real life and developed in many directions of research such as environmental or medical studies. For example, Macdonald et al. (1980) estimated the territorial range of a wildlife species by recording the locations $X_1, \ldots, X_m$ of an individual of this species which was tagged with a radio transmitter. Ripley and Rasson (1977) reconstructed the domain of the homogeneous Poisson process by compact convex set of the observed points. Here, we use Dirichlet invariant process to provide an approximation for the distribution of area and perimeter of the support. Specifically, let $S_m = \{\mathbf{X}_i\}_{i=1}^m$ be a sequence of i.i.d. random points drawn from unknown symmetric distribution F with support $K \subseteq \mathbb{R}^2$, where symmetry holds with respect to the parallel lines of axes, i.e, $x = a$ and $y = b$. We provide an asymptotically valid Bayesian nonparametric bootstrap for the distribution of area and perimeter of K. In fact, our method proceeds by the Bayesian nonparametric bootstrapping for the distribution of certain functionals of K on the basis of the sample data.

The thesis is organized as follows. In Chapter 2, we give an essential background on Dirichlet process and Dirichlet invariant process and their properties. Chapter 3 reviews the main bootstrap techniques. In Chapter 4, we first compute the Kullback-Leibler distance between the Dirichlet process and a continuous distribution. We then propose a Bayesian nonparametric chi-squared goodness-of-fit test based on the Kullback-Leibler distance between the Dirichlet process posterior and the hypothesized distribution. Further, we extend our suggested chi-squared test to present a Bayesian nonparametric chi-squared test of independence of two random variables. In addition, a method to obtain an appropriate concentration parameter based on the Kullback-Leibler distance between the Dirichlet process and the proposed distribution is described. Chapter 4 will be ended by some examples and simulation studies.

Chapter 5 discusses the Bayesian nonparametric inference for spherically symmetric distributions. We place the Dirichlet invariant process prior on the spherically symmetric distributions and derive the Dirichlet invariant process posterior for a finite group of rotations. Then, we obtain the Dirichlet invariant process posterior for the case where k approaches to infinity.

The aim of Chapter 6 is to carry out the simulation studies for the approach discussed in Chapter 5. We provide some examples with simulation studies in order to illustrate the posterior sample paths obtained from the Dirichlet invariant process for spherically symmetric distributions. Moreover, we extend our approach in Chapter 5 to obtain the Bayesian nonparametric posterior distribution for distributions which are symmetric with respect to both lines $x = a$ and $y = b$, where in fact the point (a, b) is the center of symmetry. As an application, we provide an asymptotically valid Bayesian nonparametric bootstrap for the distribution of area and perimeter for the support of these symmetric distributions.

In the final chapter, we conclude with a brief summary of the thesis and a discussion for the future research. The appendix also, contains some useful concepts that are used throughout this thesis as well as the R codes used for the simulation studies.

Chapter 2

Dirichlet Process and Dirichlet Invariant Process

In this chapter, we first review the definition, construction and various properties and some important series representations of Dirichlet process. Then, we review Dirichlet invariant process and some of its important properties. The notations and the results presented in this chapter will be used throughout this thesis. The proofs and details of the results covered in this chapter are omitted and may be found in the original sources and references included.

2.1 Dirichlet process

The Dirichlet process was initially formalized by Ferguson (1973) for general Bayesian statistical modeling as a distribution over probability distributions. We first recall the definition of Dirichlet distribution which is the finite-dimensional case of Dirichlet process. The Dirichlet distribution is known to Bayesians as the conjugate prior for the parameters of the Multinomial distribution.

Definition 2.1.1. *(Dirichlet distribution) Let $Z_1, \ldots, Z_k$ be independent random vari-*

ables with $Z_i \sim Gamma(\alpha_i, 1)$*, where* $\alpha_i > 0$*,* $i = 1, \ldots, k$*. Then, the distribution of the random vector* $\mathbf{P} = (p_1, \ldots, p_k)$*, where*

$$p_i = \frac{Z_i}{\sum_{i=1}^{k} Z_i},\ i = 1, 2, \ldots, k$$

is a Dirichlet distribution with parameter $(\alpha_1, \ldots, \alpha_k)$ *denoted by* $Dir(\alpha_1, \ldots, \alpha_k)$ *and with density function*

$$f(p_1, \ldots, p_k \mid \alpha_1, \ldots, \alpha_k) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} p_i^{\alpha_i - 1} I_{\mathbb{S}}(p_1, \ldots, p_k),$$

where $\mathbb{S}$ *is the simplex*

$$\mathbb{S} = \left\{ (p_1, \ldots, p_k) : p_i \geq 0,\ \sum_{i=1}^{k} p_i = 1 \right\}$$

and $\Gamma(x)$ *denotes the Gamma function defined by*

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt,\ x > 0.$$

For $k = 2$, the Dirichlet distribution reduces to the Beta distribution.

2.1.1 Definition and basic properties

Definition 2.1.2. *(Ferguson (1973)) Let* $\mathscr{X}$ *be a set,* $\mathscr{A}$ *be a* $\sigma-$*field of subsets of* $\mathscr{X}$*,* H *be a probability measure on* $(\mathscr{X}, \mathscr{A})$ *and* $\alpha > 0$*. A random probability measure* P *with parameters* α *and* H *is called a Dirichlet process (denoted by* $P \sim DP(\alpha H)$*) on* $(\mathscr{X}, \mathscr{A})$ *if for any finite measurable partition* $\{A_1, \ldots, A_k\}$ *of* $\mathscr{X}$*, the joint distribution of the random variables* $P(A_1), \ldots, P(A_k)$ *is a k-dimensional Dirichlet distribution with parameters* $\alpha H(A_1), \ldots, \alpha H(A_k)$*, where* $k \geq 2$*.*

We assume that if $H(A_k) = 0$, then $P(A_k) = 0$ with probability one. Then, a Dirichlet process is parameterized by α and H which are called the concentration parameter and the base distribution, respectively. The base distribution is also the mean of the Dirichlet process, i.e., for any measurable set $A \subset \mathscr{X}$,

$$E\left(P(A)\right) = H(A)$$

and

$$Var\left(P(A)\right) = \frac{H(A)(1 - H(A))}{1 + \alpha}. \tag{2.1.1}$$

These are obtained by this fact that for any measurable set $A \subset \mathscr{X}$, $P(A)$ has a beta distribution with parameters $\alpha H(A)$ and $\alpha(1 - H(A))$. One of the most remarkable properties of the Dirichlet process is that it satisfies the conjugacy property. The next theorem shows that the Dirichlet process has the conjugacy property. That is, the posterior distribution given the data is again a Dirichlet process. In the following theorem and throughout this thesis, we use a "*" as a superscript to denote posterior quantities.

Theorem 2.1.3. Let $X_1, \ldots, X_m$ be an i.i.d. sample from $P \sim DP(\alpha H)$. The posterior distribution of P given $X_1, \ldots, X_m$ is a Dirichlet process with parameters

$$\alpha_m^* = \alpha + m \tag{2.1.2}$$

and

$$H_m^* = \frac{\alpha}{\alpha + m} H + \frac{m}{\alpha + m} \frac{\sum_{i=1}^m \delta_{X_i}}{m}, \tag{2.1.3}$$

where $\delta_X(\cdot)$ is the Dirac measure, i.e., $\delta_X(A) = 1$ if $X \in A$ and 0 otherwise.

We denote the Dirichlet process posterior by $P_m^* = (P \mid X_1, \ldots, X_m) \sim DP(\alpha_m^* H_m^*)$. As it is seen in (2.1.3), the posterior base distribution H_m^* is a weighted average of H and the empirical distribution $F_m = \frac{\sum_{i=1}^m \delta_{X_i}}{m}$. Thus, for large values of α, $H_m^* \overset{a.s.}{\to} H$.

On the other hand, as $\alpha \to 0$ or as the number of observations m grows large, H_m^* becomes non-informative in the sense that H_m^* is just given by the empirical distribution and it is a close approximation of the true underlying distribution of X_i, $i = 1, \ldots, m$. This confirms the consistency property of the Dirichlet process, i.e., the posterior Dirichlet process approaches the true underlying distribution. For a discussion about the consistency property of Dirichlet process, see Ghosal (2010) and James (2008).

2.1.2 Sum representations for Dirichlet process

There are various sum representations for Dirichlet process. Here, we review some most important representations. A sum representation of Dirichlet process is presented by Ferguson (1973) based on the work of Ferguson and Klass (1972). Specifically, let $(\theta_i)_{i\geq 1}$ be a sequence of i.i.d. random variables with common distribution H and $(E_k)_{k\geq 1}$ be a sequence of i.i.d. random variables from the exponential distribution with mean 1. If $\Gamma_i = E_1 + \cdots + E_i$ and $(\Gamma_i)_{i\geq 1}$ are independent from $(\theta_i)_{i\geq 1}$, then,

$$P = \sum_{i=1}^{\infty} \frac{L^{-1}(\Gamma_i)}{\sum_{i=1}^{\infty} L^{-1}(\Gamma_i)} \delta_{\theta_i} = \sum_{i=1}^{\infty} p_i \delta_{\theta_i} \tag{2.1.4}$$

is a Dirichlet process with parameters α and H, where

$$L(x) = \alpha \int_x^{\infty} t^{-1} e^{-t} dt, \ x > 0 \tag{2.1.5}$$

and

$$L^{-1}(y) = \inf\{x > 0 : L(x) \geq y, y > 0\}. \tag{2.1.6}$$

Ishwaran and Zarepour (2002) introduced a finite sum approximation for the Dirichlet process which is easier to work with. Let $\mathbf{p} = (p_{1,n}, \ldots, p_{n,n})$ have a Dirichlet distribution with parameters $(\alpha/n, \ldots, \alpha/n)$ denoted by $\mathrm{Dir}(\alpha/n, \ldots, \alpha/n)$ and $(\theta_i)_{1\leq i\leq n}$ be a sequence of i.i.d. random variables with distribution H and independent of $(p_{i,n})_{1\leq i\leq n}$. Also, let $(\mathcal{G}_{i,n})_{1\leq i\leq n}$ be i.i.d. random variables from Gamma$(\alpha/n, 1)$ distribution and $p_{i,n} = \mathcal{G}_{i,n}/\mathcal{G}_n$, where $\mathcal{G}_n = \mathcal{G}_{1,n} + \cdots + \mathcal{G}_{n,n}$. Then,

$$P_n = \sum_{i=1}^{n} p_{i,n}\delta_{\theta_i} = \sum_{i=1}^{n} \frac{\mathcal{G}_{i,n}}{\mathcal{G}_n}\delta_{\theta_i} \tag{2.1.7}$$

is called a finite-dimensional Dirichlet process and approximates the Ferguson's Dirichlet process weakly. Another finite sum representation of the Dirichlet process with monotonically decreasing weights is presented in Zarepour and Al Labadi (2012). Specifically, let $(\theta_i)_{1\leq i\leq n}$ be a sequence of i.i.d. random variables with values in $\mathscr{X}$ and common distribution H and independent of $(\Gamma_i)_{1\leq i\leq n+1}$. Let $X_n \sim$ Gamma$(\alpha/n, 1)$ and define

$$G_n(x) = \Pr(X_n > x) = \frac{1}{\Gamma(\alpha/n)} \int_x^{\infty} t^{(\alpha/n)-1} e^{-t} dt$$

and

$$G_n^{-1}(y) = \inf\{x : G_n(x) \geq y, y \in [0, 1)\}.$$

Then, as $n \to \infty$,

$$P_n = \sum_{i=1}^{n} \frac{G_n^{-1}(\frac{\Gamma_i}{\Gamma_{n+1}})}{\sum\limits_{i=1}^{n} G_n^{-1}(\frac{\Gamma_i}{\Gamma_{n+1}})}\delta_{\theta_i} \xrightarrow{a.s.} P = \sum_{i=1}^{\infty} \frac{L^{-1}(\Gamma_i)}{\sum\limits_{i=1}^{\infty} L^{-1}(\Gamma_i)}\delta_{\theta_i}.$$

If we define

$$p_{i,n} = \frac{G_n^{-1}(\frac{\Gamma_i}{\Gamma_{n+1}})}{\sum\limits_{i=1}^{n} G_n^{-1}(\frac{\Gamma_i}{\Gamma_{n+1}})},$$

then, P_n can be written as

$$P_n = \sum_{i=1}^{n} p_{i,n}\delta_{\theta_i}. \tag{2.1.8}$$

Zarepour and Al Labadi (2012) proved that this finite sum representation converges almost surely to Ferguson's representation and empirically converges faster than the

Algorithm 1 Generating a sample distribution from the approximate Dirichlet process $DP(\alpha H)$

1. For a fixed relatively large positive integer n generate $\theta_i \overset{i.i.d.}{\sim} H$ for $i = 1, \ldots, n$.
2. Let $\Gamma_i = E_1 + \cdots + E_i$, where E_i, $(i = 1, \ldots, n+1)$ are generated from an exponential distribution with mean 1, independent of $(\theta_i)_{1 \le i \le n}$.
3. Compute $G_n^{-1}\left(\frac{\Gamma_i}{\Gamma_{n+1}}\right)$, for $i = 1, \ldots, n$ which are simply the quantile functions of the $Gamma(\frac{\alpha}{n}, 1)$ distribution evaluated at $1 - \frac{\Gamma_i}{\Gamma_{n+1}}$.
4. Construct the distribution P_n as given in equation (2.1.8).

---

other representations. Algorithm 1 by Zarepour and Al Labadi (2012) explains the steps of generating a sample from the approximate Dirichlet process with parameter αH. For other sum representations of Dirichlet process, see for example, Sethuraman (1994) and Bondesson (1982).

2.2 Dirichlet invariant process and invariant random probability

In this section, we review the definition, construction and various properties of Dirichlet invariant process. The Dirichlet invariant process was initially formalized by Dalal (1979a) as a prior over invariant probability distributions. Despite the slight complexity of the Dirichlet invariant process compared to the Dirichlet process, most of their properties are similar. We first recall the definition of the Dirichlet invariant process.

2.2.1 Definition and properties

Dalal (1979a) applied Ferguson's approach to define the Dirichlet invariant process as a prior over the space of all distributions invariant under a group of transformations. Let $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$ be any p-dimensional Euclidean space with the associated Borel σ-field. Let $\mathscr{G} = \{g_1, ..., g_k\}$ be any finite group of measurable transformations $\mathscr{X} \to \mathscr{X}$. Dalal (1979a) defined a random probability measure such that for any particular realization, sets B and $g_j(B), (j = 1, ..., k)$ have the same probabilities. We begin by recalling the definition of an invariant random probability measure.

Definition 2.2.1. *(Dalal (1979a)) Let $(\Omega, \mathscr{F}, Q)$ be any probability space, and let $\mathscr{P}(\mathscr{X})$ and $\mathscr{B}(\mathscr{P}(\mathscr{X}))$ denote the space of all probability measures on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$ and the corresponding Borel σ-field generated by the weak topology. Further let P : $(\Omega, \mathscr{F}) \to (\mathscr{P}(\mathscr{X}), \mathscr{B}(\mathscr{P}(\mathscr{X})))$ be any measurable mapping. Then, the random probability measure P is a $\mathscr{G}$-invariant random probability measure if $P(A) = P(g(A))$ for all g in $\mathscr{G}$, where $P(A)$ is the probability of set A under probability measure P.*

Definition 2.2.2. *(Dalal (1979a)) An invariant random probability measure P is a Dirichlet invariant process denoted by $P \sim DIP(\alpha H)$ if for any $\alpha > 0$, there exists a $\mathscr{G}$-invariant measure H on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$ such that for every $\mathscr{G}$-invariant measurable partition (that is the sets of the partitions are $\mathscr{G}$-invariant and $\mathscr{B}(\mathscr{X})$ measurable) $B_1, ..., B_k$ of $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$, the joint distribution of $(P(B_1), \ldots, P(B_k))$ is $Dir\,(\alpha H(B_1), \ldots, \alpha H(B_k))$.*

The proofs of all results in this section are discussed by Dalal (1979a) and are similar to those for Dirichlet process as presented in Ferguson (1973).

Theorem 2.2.3. (Dalal (1979a)) P is a Dirichlet invariant process with associated parameters α and H denoted by $P \sim DIP(\alpha H)$.

Proposition 2.2.4. *(Dalal (1979a)) Let P be a Dirichlet invariant process on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$ with parameters α and H and let X be a sample of size 1 from P. Then, for $A \in \mathscr{A}$*

$$P(X \in A) = P(X \in g(A)) = H(A), \text{ for any } g \text{ in } \mathscr{G}.$$

Proposition 2.2.5. *(Dalal (1979a)) Suppose the conditions of Proposition 2.2.4 hold. Then, X and $g(X)$ are identically distributed for any g in $\mathscr{G}$.*

Theorem 2.2.6. (Dalal (1979a)) Let $P \sim DIP(\alpha H)$ and $X_1, ..., X_m$ be a sample of size m from P. Then, the conditional distribution of P given $X_1, ..., X_m$ is $DIP(\alpha H + \sum\limits_{i=1}^{m} \delta_{X_i}^{g})$, where $\delta_{X_i}^{g} = \frac{1}{k} \sum\limits_{j=1}^{k} \delta_{g_j(X_i)}$ and δ_X is a measure degenerate at X and $k = |G|$.

Chapter 3

Bootstrap Methods

In this chapter, we review two main bootstrap techniques: Efron's regular bootstrap and Bayesian bootstrap with its various extensions.

3.1 Efron's regular bootstrap

Efron (1992) introduced a resampling procedure which is also called bootstrap sampling. The idea behind the bootstrap is to use the observed data for the purpose of approximating the sampling distribution of particular statistic of an unknown distribution. Indeed, the bootstrap proceeds by resampling with replacement from the sample data to generate a large number of bootstrap samples. Therefore, we can use the bootstrap estimator to make inference for the parameter of interest. Suppose F is the unknown distribution of the sample data $X_1, \ldots, X_m$. The aim of the regular bootstrap is to draw random samples with replacement from the empirical distribution F_m. Let $\theta = \theta(F)$ be the unknown parameter, $\hat{\theta}_m = \theta(F_m)$ be an estimator for θ and $\hat{\theta}^*_m = \theta(F^*_m)$ be the bootstrap estimator. The following algorithm states the regular bootstrap technique:

1. Construct the empirical distribution F_m of the sample data by assigning a prob-

ability of $\frac{1}{m}$ to each sample point $X_1, \ldots, X_m$.

2. Draw a random sample $X_1^*, \ldots, X_m^*$ of size m with replacement from the empirical distribution F_m. Then, $X_1^*, \ldots, X_m^*$ is a bootstrap sample with the bootstrap empirical distribution $F_m^* = \frac{1}{m}\sum_{i=1}^{m}\delta_{X_i^*}$.

3. Calculate the statistic of interest $\hat{\theta}_m$ for this bootstrap sample denoted by $\hat{\theta}_m^*$.

4. Repeat 1 to 3 for a large number of times, N, to obtain N values of the statistic $\hat{\theta}_m$ denoted by $\hat{\theta}_{m1}^*, \ldots, \hat{\theta}_{mN}^*$.

5. Construct the empirical distribution of $\hat{\theta}_{m1}^*, \ldots, \hat{\theta}_{mN}^*$ by assigning a probability of $1/N$ to each point $\hat{\theta}_{mi}^*, i = 1, \ldots, N$. Use the empirical distribution of $\hat{\theta}_{m1}^*, \ldots, \hat{\theta}_{mN}^*$ to make inferences about θ.

In this bootstrap procedure, since the sampling is carried out with replacement, it is unlikely to draw the bootstrap samples identical to the original sample data where m is large enough.

It is unnecessary to check whether or not both $a_m(\hat{\theta}_m - \theta)$ and $a_m(\hat{\theta}_m^* - \hat{\theta}_m)$ converge to the same distribution. See for example, Singh (1981) and Bickel and Freedman (1981).

3.2 Bayesian bootstrap

3.2.1 Rubin's Bayesian bootstrap

Rubin (1981) introduced a Bayesian bootstrap. Bayesian bootstrap is the Bayesian analogue of the regular bootstrap and are asymptotically equivalent. The regular bootstrap simulates the distribution of a statistic that is estimating the parameter of interest while the Bayesian bootstrap simulates the posterior distribution of the parameter. In other words, the Bayesian bootstrap assigns a posterior probability to each sample point while in regular bootstrap the probabilities are fixed and equal to $1/m$. Specifically, let $(u_{(1)}, \ldots, u_{(m)})$ be m ordered uniform(0,1) random variates. Then, a Bayesian bootstrap replication is generated by the vector $(g_1, \ldots, g_m)$ of random probabilities $g_i = u_{(i)} - u_{(i-1)}, \quad (i = 1, \ldots, m)$ attaching to the data values $x_1, \ldots, x_m$. It should be mentioned that both methods are applied based on consideration of some model specifications and the results are sensitive to these assumptions.

Suppose $X_1, \ldots, X_m$ is a sequence of i.i.d. random variables having an unknown distribution F. Let $\theta = \theta(F, \mathbf{X})$ be any given functional depending on F and $\mathbf{X} = (X_1, \ldots, X_m)$. A Bayes approach to estimate $\theta(F, \mathbf{X})$ given $\mathbf{X} = \mathbf{x}$ is carried out by considering a prior on the space of all distributions F. Then, $\theta(F, \mathbf{X})$ can be estimated by the posterior distribution of $\theta(F, \mathbf{X})$ given $\mathbf{X} = \mathbf{x}$. Rubin's Bayesian bootstrap is carried out based on a random distribution D_m constructed by random probabilities $g_1, \ldots, g_m$ attaching to $X_1, \ldots, X_m$. On the other hand, Rubin's bootstrap is defined by replacing the jump-size of the empirical distribution of $X_1, \ldots, X_m$ by the random gaps denoted by $g_1, \ldots, g_m$. Then, the conditional distribution of $\hat{\theta}_m = \theta(D_m, \mathbf{X}) \mid \mathbf{X} = \mathbf{x}$ is used as the posterior distribution of $\theta(F, \mathbf{X}) \mid \mathbf{X} = \mathbf{x}$. The Monte Carlo steps of the Rubin's Bayesian bootstrap to simulate the approximated posterior distribution of $\theta(F, \mathbf{X}) \mid \mathbf{X} = \mathbf{x}$ can be explained as follows:

1. Generate a random sample $U_1, \ldots, U_m$ of size m from uniform(0,1) distribution. Denote the ordered statistics of U's by $0 = U_{(0)}, U_{(1)}, \ldots, U_{(m-1)}, U_{(m)} = 1$ and define the gaps $g_i = U_{(m)} - U_{(m-1)}$, $i = 1, \ldots, m$. Construct a random discrete distribution function D_m by the weights g_i, $i = 1, \ldots, m$ assigning to x_i, $i = 1, \ldots, m$.

2. Repeat 1 for a large number of times, N, to obtain the posterior distributions $D_{m1}, \ldots, D_{mN}$. Compute the statistic of interest (Bayesian bootstrap estimators) $\hat{\theta}_{m1}, \ldots, \hat{\theta}_{mN}$ for these posterior distributions denoted by $\hat{\theta}^*_{m1}, \ldots, \hat{\theta}^*_{mN}$, where $\hat{\theta}^*_{mi} = \theta(D_{mi}, \mathbf{X})$, $i = 1, \ldots, N$.

3. The distribution of $\theta(D_m, \mathbf{X}) \mid \mathbf{X} = \mathbf{x}$ is approximated by empirical distribution function of $\hat{\theta}^*_{m1}, \ldots, \hat{\theta}^*_{mN}$.

Remark 3.2.1. (Wilks (1962)). Suppose $X_{(1)}, \ldots, X_{(m)}$ denote the order statistics of a sample from a continuous distribution F. Define the gaps by $U_1 = F(X_{(1)})$, $U_2 = F(X_{(2)}) - F(X_{(1)}), \ldots, U_m = F(X_{(m)}) - F(X_{(m-1)})$. Then, $U_1, \ldots, U_n$ are random variables following a Dirichlet distribution $Dir(1, \ldots, 1)$.

Remark 3.2.1 shows that the corresponding prior distribution for the parameter $\theta(F, \mathbf{X}) \mid \mathbf{X} = \mathbf{x}$ in the Rubin's Bayesian bootstrap follows a flat Dirichlet distribution $Dir(1, \ldots, 1)$. The fact that Rubin's Bayesian bootstrap is simulating the posterior distribution of the parameter $\theta(F, \mathbf{X}) \mid \mathbf{X} = \mathbf{x}$ under a specific model is proved in the following Theorem.

Theorem 3.2.2. (Rubin (1981)) Let $d = (d_1, \ldots, d_k)$ be the vector of all possible distinct values of X and let $\theta = (\theta_1, \ldots, \theta_k)$ be the associated vector of probabilities, where

$$P(X = d_i \mid \theta) = \theta_k, \quad i = 1, \ldots, k \tag{3.2.1}$$

and $\sum_{i=1}^{k} \theta_i = 1$. Let $x_1, \ldots, x_n$ be an i.i.d. sample of X and n_k be the number of x_i equal to d_k, i.e., $n_i = \sum_{j=1}^{n} I(X_j = d_i) = 1$, $i = 1, \ldots, k$. If θ has a Dirichlet prior distribution $Dir(l_1 + 1, \ldots, l_k + 1)$, then, the posterior distribution of θ is a $(k-1)$ dimensional Dirichlet distribution $Dir(l_1 + n_1 + 1, \ldots, l_k + n_k + 1)$.

In other words, the posterior distribution can be simulated by $m - 1$ independent uniform random numbers, where $m = n + k + \sum_{i=1}^{k} l_i$. Let $U_1, \ldots, U_{m-1}$ be i.i.d. random sample from uniform(0,1) distribution with corresponding order statistics $U_{(1)}, \ldots, U_{(m-1)}$. Denote the m gaps by $g_1, \ldots, g_m$, where $g_r = U_{(r)} - U_{(r-1)}$, $r = 1, \ldots, m$. If we divide $g_1, \ldots, g_m$ into k groups, then, the kth partition has $l_k + n_k + 1$ elements. If p_k denote the sum of the g_is in the kth partition, Wilks

(1962) proved that $(p_1, \dots, p_k)$ follows a $(k-1)$ dimensional Dirichlet distribution $Dir(l_1 + n_1 + 1, \dots, l_k + n_k + 1)$. Consequently, the simulated posterior distribution of θ is obtained by the weight p_i assigned to each d_i for $i = 1, \dots, k$.

3.2.2 Lo's Bayesian bootstrap

Lo (1987) extended Rubin's Bayesian bootstrap to a more general Bayesian bootstrap technique. Namely, assume $X_1, \dots, X_m$ is a random sample from an unknown distribution F with empirical distribution F_m. Let θ and $\hat{\theta}_n$ be the parameter of interest and the estimator for θ given F_m, respectively. Lo's Bayesian bootstrap is carried out by placing a Dirichlet process prior $DP(\alpha H)$ on distribution F and calculating the Bayesian bootstrap estimators by the Dirichlet process posterior $DP(\alpha_m^* H_m^*)$, where $\alpha_m^* = \alpha + m$ and $H_m^* = \frac{\alpha}{\alpha+m} H + \frac{m}{\alpha+m} F_m$. Moreover, Lo (1987) proved that the Bayesian bootstrap estimator is valid when the prior is a Dirichlet process. However, this result is not surprising as the prior distribution in the regular Bayesian bootstrap follows the flat Dirichlet distribution. The following describes the steps of Lo's Bayesian bootstrap:

1. Generate a sample P_m^* from Dirichlet process posterior $DP(\alpha_m^* H_m^*)$, where $\alpha_m^* = \alpha + m$ and $H_m^* = \frac{\alpha}{\alpha+m} H + \frac{m}{\alpha+m} F_m$ as explained in Chapter 2. Then, P_m^* is a random discrete posterior distribution.

2. Compute the statistic of interest $\hat{\theta}_m$ by the posterior distribution P_m^* denoted by $\hat{\theta}_m^*$.

3. Repeat steps 2 and 3 N times to generate N posterior distributions and obtain N values of the statistic denoted by $\hat{\theta}_{m1}^*, \dots, \hat{\theta}_{mN}^*$. Here N is large.

4. Approximate the distribution of $\hat{\theta}_m$ by empirical distribution of $\hat{\theta}_{m1}^*, \dots, \hat{\theta}_{mN}^*$.

Lo (1987) proved that the regular bootstrap and the Bayesian bootstrap are asymptotically equivalent. In other words, they have identical asymptotic distributions for the large enough sample size m.

3.2.3 Muliere and Secchi's bootstrap

A generalization of the Bayesian bootstrap associated with a Dirichlet process prior $DP(\alpha H)$ is introduced by Muliere and Secchi (1996) that can be considered as an extension of Efron's and Rubin's bootstraps. Their new resampling method has some advantages over the other methods. First, the prior opinions are taken into account. Second, the predictive distribution of the future observations is not necessarily concentrated on observed sample data. For $\alpha = 0$, the procedure is equivalent to Efron's regular bootstrap and Rubin's bootstrap.

Muliere and Secchi (1996) introduced the bootstrap resampling techniques of Efron (1992) and Rubin (1981) in a general Bayesian nonparametric setting where the aim is to approximate the posterior distribution of a statistical functional $\phi(F)$ of a random distribution function F. In Bayesian bootstrap, the procedures yield the approximation for the conditional distribution

$$\mathcal{L}(\phi(F, \mathbf{X}) \mid X_1, \ldots, X_m), \tag{3.2.2}$$

where, $\phi(F, \mathbf{X})$ is a functional depending on F and $\mathbf{X}$. The regular bootstrap technique approximates (3.2.2) by

$$\mathcal{L}(\phi(F_m^*, \mathbf{X}) \mid X_1, \ldots, X_m),$$

where F_m^* is the empirical distribution of an i.i.d. sample $X_1^*, \ldots, X_m^*$ from the empirical distribution function F_m of $X_1, \ldots, X_m$. In other words, for every Borel set B, the conditional distribution (3.2.2) is approximated by

$$\mathcal{L}(F_m^*(B) \mid X_1, \dots, X_m) = \frac{1}{m} Binomial(m, F_m(B)). \tag{3.2.3}$$

The Bayesian bootstrap technique proposes to approximate (3.2.2) by

$$\mathcal{L}(\phi(F_m^R, \mathbf{X}) \mid X_1, \dots, X_m), \tag{3.2.4}$$

where F_m^R is the random distribution function defined by

$$F_m^R(x) = \frac{1}{\sum_{i=1}^{m} V_i} \sum_{i=1}^{m} V_i I_{[X_i, \infty)}(x),$$

and $V_1, \dots, V_m$ are i.i.d. random variables with exponential distribution of parameter 1, $Exp(1)$ and independent of $X_1, \dots, X_m$. Here, for every Borel set B, the conditional distribution (3.2.2) can be approximated by

$$\mathcal{L}(F_m^R(B) \mid X_1, \dots, X_m) = Beta(mF_m(B), (1 - mF_m(B))).$$

The following Lemma states a trivial case to simulate a Dirichlet process when the base distribution H is a discrete distribution function with finite support.

Lemma 3.2.3. *(Muliere and Secchi (1996)). Let H be a discrete distribution function with a finite support $\{z_1, \dots, z_r\}$in $\mathbb{R}$ and p_i be the probability which H assigns to z_i, $i = 1, \dots, r$. Let $V_1, \dots, V_r$ be r independent random variables from a Gamma distribution $Gamma(\alpha p_i, 1)$, $i = 1, \dots, r$, where $\alpha > 0$. Suppose F is the random distribution function such that for every $x \in \mathbb{R}$,*

$$F(x) = \frac{1}{\sum_{i=1}^{r} V_i} \sum_{i=1}^{r} V_i I_{[z_i, \infty)}(x).$$

Then F is a Dirichlet process $DP(\alpha H)$.

Consider a random sample $X_1, \dots, X_m$ from an unknown distribution F. By placing a Dirichlet process prior $DP(\alpha H)$ on F, the steps of the new bootstrap technique are presented as follows:

1. Generate N i.i.d. observations $x_1^*, \dots, x_N^*$ from $H_m^* = \frac{\alpha}{\alpha+m} H + \frac{m}{\alpha+m} F_m$.
2. Generate v_i, $i = 1, \dots, N$ from the Gamma distribution
$$\text{Gamma} \left(\frac{\alpha + m}{N}, 1 \right).$$
3. Calculate the quantity (the parameter of interest)
$$t = \phi \left(\frac{1}{\sum_{i=1}^{N} v_i} \sum_{i=1}^{N} v_i I_{[x_i^*, \infty)}(.) \right).$$
4. Repeat steps 1 to 3, k times to obtain the quantities $t_1, \dots, t_k$.
5. Approximate the conditional distribution function $\mathcal{L}(\phi(F) \mid X_1, \dots, X_m)$ by the empirical distribution function of $t_1, \dots, t_k$.

Chapter 4

A Consistent Bayesian Bootstrap for Chi-squared Goodness-of-Fit Test Using a Dirichlet Prior

The main objective of this chapter is to construct a Bayesian nonparametric goodness-of-fit test based on the Kullback-Leibler distance between the Dirichlet process and a continuous distribution. Following this, we extend our approach to present a Bayesian nonparametric chi-squared test of independence for a contingency table. The examples with simulation studies are provided at the end of the chapter. We begin with recalling the definition of the Kullback-Leibler distance.

4.1 Kullback-Leibler distance

The Kullback-Leibler distance that measures the distance between two distributions introduced by Kullback and Leibler (1951). Suppose $\mathcal{P}$ and $\mathcal{Q}$ are two probability measures for discrete random variables on a measurable space $(\Omega, \mathcal{F})$. The Kullback-

Leibler distance between $\mathcal{P}$ and $\mathcal{Q}$ is defined as

$$D_{KL}(\mathcal{P} \parallel \mathcal{Q}) = \sum_i \mathcal{P}(i) \log\left(\frac{\mathcal{P}(i)}{\mathcal{Q}(i)}\right).$$

For continuous probability measures $\mathcal{P}$ and $\mathcal{Q}$ with $\mathcal{P}$ absolutely continuous with respect to $\mathcal{Q}$, the Kullback-Leibler distance is written as

$$D_{KL}(\mathcal{P} \parallel \mathcal{Q}) = \int \log\left(\frac{d\mathcal{P}}{d\mathcal{Q}}\right) d\mathcal{P}$$

where $\frac{d\mathcal{P}}{d\mathcal{Q}}$ is the Radon-Nikodym derivative of $\mathcal{P}$ with respect to $\mathcal{Q}$. Let $\mathcal{P} \ll \lambda$ and $\mathcal{Q} \ll \lambda$ where λ is the Lebesgue measure. If the densities of $\mathcal{P}$ and $\mathcal{Q}$ with respect to Lebesgue measure are denoted by $p(x)$ and $q(x)$, respectively, then the Kullback-Leibler distance is written as

$$D_{KL}(\mathcal{P} \parallel \mathcal{Q}) = \int_{\mathbb{R}} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx. \tag{4.1.1}$$

We compute the distance between the random distribution P_n as given in (2.1.8) and a continuous distribution F with density $f(x)$. Since P_n is a discrete measure and F is continuous, we estimate the density $f(x)$ by its histogram estimator on a partitioned space. Also, since the Kullback-Leibler distance is not symmetric, we compute both distances $\mathrm{D}_{KL}(P_n \parallel F)$ and $\mathrm{D}_{KL}(F \parallel P_n)$.

4.1.1 Kullback-Leibler distance between the Dirichlet process and a continuous distribution

Lemma 4.1.1. *Let H and F be two distributions defined on the same space $\mathcal{X}$ and $P_n = \sum_{i=1}^{n} p_{i,n}\delta_{\theta_i}$ be a random distribution as defined in* (2.1.8), *i.e., $\theta_1, \ldots, \theta_n$ are i.i.d. generated from H with corresponding order statistics $\theta_{(1)}, \ldots, \theta_{(n)}$. Suppose that the sample space is partitioned as $x_{(1)} < \cdots < x_{(n+1)}$ such that $x_{(i)} < \theta_{(i)} < x_{(i+1)}, i =$*

$1,\ldots,n$. *We have*

$$D_{KL}(P_n \parallel F) = -\mathcal{H}(\mathbf{p}) - \sum_{i=1}^{n} p_{i,n} \log(q_i) \tag{4.1.2}$$

and

$$D_{KL}(F \parallel P_n) = -\mathcal{H}(\mathbf{q}) - \sum_{i=1}^{n} q_i \log(p_{i,n}), \tag{4.1.3}$$

where $\mathcal{H}(\mathbf{p}) = -\sum_{i=1}^{n} p_{i,n} \log(p_{i,n})$ *is the entropy of* P_n *and* $\mathcal{H}(\mathbf{q}) = -\sum_{i=1}^{n} q_i \log(q_i)$ *with* $q_i = \frac{\triangle F(x_i)}{\triangle x_i}$.

Proof: Suppose that the sample space is partitioned as $x_{(1)} < \cdots < x_{(n+1)}$ such that $x_{(i)} < \theta_{(i)} < x_{(i+1)}, i = 1,\ldots,n$. By definition of the Kullback-Leibler distance, we have

$$\begin{aligned} D_{KL}(P_n \parallel F) &= \sum_{i=1}^{n} \triangle P_n(x_i) \log\left(\frac{\triangle P_n(x_i)}{\triangle F(x_i)/\triangle x_i}\right) \\ &= \sum_{i=1}^{n} \triangle P_n(x_i) \log(\triangle P_n(x_i)) - \sum_{i=1}^{n} \triangle P_n(x_i) \log\left(\frac{\triangle F(x_i)}{\triangle x_i}\right) \\ &= \sum_{i=1}^{n} p_{i,n} \log(p_{i,n}) - \sum_{i=1}^{n} p_{i,n} \log\left(\frac{\triangle F(x_i)}{\triangle x_i}\right) \\ &= -\mathcal{H}(\mathbf{p}) - \sum_{i=1}^{n} p_{i,n} \log\left(\frac{\triangle F(x_i)}{\triangle x_i}\right), \end{aligned} \tag{4.1.4}$$

where $\triangle F(x_i) = F(x_{(i+1)}) - F(x_{(i)}), \triangle x_i = x_{(i+1)} - x_{(i)}$,
$p_{i,n} = P_n(x_{(i+1)}) - P_n(x_{(i)}) = P_n(\theta_{(i)})$ and $\mathcal{H}(\mathbf{p}) = -\sum_{i=1}^{n} p_{i,n} \log(p_{i,n})$ is the entropy of P_n. Similarly, we get

$$D_{KL}(F \parallel P_n) = -\mathcal{H}(\mathbf{q}) - \sum_{i=1}^{n} q_i \log(p_{i,n}), \tag{4.1.5}$$

where $q_i = \frac{\triangle F(x_i)}{\triangle x_i}$ and $\mathcal{H}(\mathbf{q}) = -\sum_{i=1}^{n} q_i \log q_i$. ∎

4.2 A consistent Bayesian bootstrap for chi-squared goodness-of-fit test

The null hypothesis of the goodness-of-fit test is given as $H_0 : F = F_0$, where F is the true underlying distribution of the observed data and F_0 is some specified distribution. Suppose $X_1, \ldots, X_m$ is a sample of size m from the distribution F and the sample space is partitioned into k non-overlapping bins. Let O_i and E_i, $i = 1, \ldots, k$ denote the observed counts and the expected counts under the hypothesized distribution F_0 for bin k, respectively. The Pearson's goodness-of-fit test statistic is defined as

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

and X^2 asymptotically converges to a chi-squared distribution with $k-1$ degrees of freedom. To derive a counterpart Bayesian nonparametric test statistic similar to X^2, we consider a Dirichlet process with parameters α and $H = F_0$ as a prior for the true distribution of data, i.e., $X_1, \ldots, X_m \sim P$, where $P \sim DP(\alpha H)$. Then, given $X_1, \ldots, X_m$, the posterior distribution of P is a Dirichlet process $P_m^* = (P \mid X_1, \ldots, X_m) \sim DP(\alpha_m^* H_m^*)$, where α_m^* and H_m^* are as given in (2.1.2) and (2.1.3). We carry out the test based on the chi-squared distance between the posterior Dirichlet process P_m^* and the hypothesized distribution F_0. Note that for the large sample size, both the Pearson's goodness-of-fit test and the likelihood ratio test (the Kullback-Leibler distance) are asymptotically equivalent as it can easily be shown that

$$2\sum_{i=1}^{k} O_i \log \frac{O_i}{E_i} \approx \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}. \tag{4.2.1}$$

For simplicity, we only consider Pearson's goodness-of-fit test. To study asymptotic behavior of this distances results should rely on functional Bernstein-von Mises Theorem. This asymptotic property is studied in Al Labadi (2012) and Lo (1987), but for convenience we include simple calculations to derive the asymptotic distribution of $D_{\alpha m}(A) = \sqrt{m}(P_m^*(A) - H_m^*(A))$, where $A \in \mathcal{X}$. Notice that by having the partition $\{A, A^c\}$ and the definition of Dirichlet process,

$$P_m^*(A) \sim Beta\left(\alpha_m^* H_m^*(A), \alpha_m^* H_m^*(A^c)\right).$$

Set $Y = P_m^*(A)$ and $v = H_m^*(A)$, where H_m^* is defined in (2.1.3). Then, for $0 < y < 1$, the random variable Y has the probability density function

$$f(y) = \frac{\Gamma(\alpha_m^*)}{\Gamma(\alpha_m^* v)\Gamma(\alpha_m^*(1-v))} y^{\alpha_m^* v - 1}(1-y)^{\alpha_m^*(1-v)-1}.$$

Thus, the probability density function of $Z = \sqrt{m}(Y - v)$ is

$$f_Z(z) \quad = \quad \frac{\Gamma(\alpha_m^*)}{\sqrt{m}\Gamma(\alpha_m^* v)\Gamma(\alpha_m^*(1-v))} \left(\frac{z}{\sqrt{m}} + v\right)^{\alpha_m^* v - 1} \left(1 - \frac{z}{\sqrt{m}} - v\right)^{\alpha_m^*(1-v)-1} \tag{4.2.2}$$

where $-\sqrt{m}v < z < \sqrt{m}(1-v)$. By Scheffé's theorem (Billingsley (2013), page 29), we need to show that

$$f_Z(z) \to \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{z^2}{2\sigma^2}\right\},$$

where $\sigma^2 = F(A)(1 - F(A))$. By Stirling's formula, we have

$$\Gamma(x) \approx \sqrt{2\pi} x^{x-\frac{1}{2}} e^{-x} \text{ as } x \to \infty,$$

where we use the notation $f(x) \approx g(x)$ as $x \to \infty$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$. From (2.1.3) and (2.1.2), as $m \to \infty$, $H_m^* \overset{a.s.}{\to} F$ and $\alpha_m^* = \alpha + m \approx m$. Then, the equation (4.2.2) can be rewritten as

$$f_Z(z) \;=\; \frac{\Gamma(m)}{\sqrt{m}\Gamma(mv)\Gamma(m(1-v))}\left(\frac{z}{\sqrt{m}}+v\right)^{mv-1}\left(1-\frac{z}{\sqrt{m}}-v\right)^{m(1-v)-1},$$

where $v = F(A)$. Then,

$$\begin{aligned}\lim_{m\to\infty} f_Z(z) &= \frac{1}{\sqrt{2\pi}}\lim_{m\to\infty}\left\{\frac{\left(\frac{z}{\sqrt{m}}+v\right)^{mv-1}\left(1-\frac{z}{\sqrt{m}}-v\right)^{m(1-v)-1}}{v^{mv-\frac{1}{2}}(1-v)^{m(1-v)-\frac{1}{2}}}\right\} \\ &= \frac{1}{\sqrt{2\pi v(1-v)}}\exp\left\{\lim_{m\to\infty} m\ln(\eta_m)\right\}, \qquad (4.2.3)\end{aligned}$$

where

$$\eta_m = \left(1+\frac{z}{\sqrt{mv}}\right)^{v}\left(1-\frac{z}{\sqrt{m}(1-v)}\right)^{1-v}.$$

Therefore,

$$\lim_{m\to\infty} m\ln(\eta_m) \;=\; \lim_{m\to\infty}\frac{1}{1/m}\left\{v\ln\left(1+\frac{z}{\sqrt{mv}}\right)+(1-v)\ln\left(1-\frac{z}{\sqrt{m}(1-v)}\right)\right\}.$$

Since

$$\ln\left(1+\frac{z}{\sqrt{mv}}\right) \;=\; \frac{z}{\sqrt{mv}}-\frac{z^2}{2mv^2}+o(\frac{1}{m})$$

and

$$\ln\left(1-\frac{z}{\sqrt{m}(1-v)}\right) = -\frac{z}{\sqrt{m}(1-v)}-\frac{z^2}{2m(1-v)^2}+o(\frac{1}{m}),$$

we have

$$\lim_{m\to\infty} m \ln(\eta_m) = \frac{-z^2}{2v(1-v)}. \tag{4.2.4}$$

Substituting (4.2.4) in (4.2.3) completes the proof of normality of $D_{\alpha m}(A) = \sqrt{m}(P_m^*(A) - H_m^*(A))$. A similar method proves that as $m \to \infty$, for any partition $\{A_1, \ldots, A_k\}$ of the space $\mathcal{X}$,

$$(D_{\alpha m}(A_1), D_{\alpha m}(A_2), \ldots, D_{\alpha m}(A_k)) \xrightarrow{d} (B_F(A_1), B_F(A_2), \ldots, B_F(A_k)),$$

where B_F is the Brownian bridge.

Remark 4.2.1. A Gaussian process $\{B_F(A), A \in \mathcal{X}\}$ is called a Brownian bridge if $E(B_F(A)) = 0$ and $Cov(B_F(A_i), B_F(A_j)) = F(A_i \cap A_j) - F(A_i) \cap F(A_j)$, where $A_i, A_j \in \mathcal{X}$.

Now we can conclude the following Lemma.

Lemma 4.2.2. *Let $X_1, \ldots, X_m$ be a random sample from the distribution H. If P_m^* is the Dirichlet process posterior given $X_1, \ldots, X_m$, then, as $m \to \infty$,*

$$D_{\alpha m}(\cdot) = \sqrt{m}(P_m^*(\cdot) - H_m^*(\cdot)) \xrightarrow{d} B_F(\cdot).$$

For a detailed proof similar to what we presented here, see Al Labadi (2012). Also, see James (2008), Ghosal (2010) and Lo (1987). Al Labadi (2012) proved that as $\alpha \to \infty$, $D_\alpha(\cdot) = \sqrt{\alpha}(P(\cdot) - H(\cdot)) \xrightarrow{d} B_H(\cdot)$. The following Theorem describes this connection and the asymptotic distribution for the law of the posterior distance for large sample size which is equivalent to the frequentist's chi-squared test.

Theorem 4.2.3. Suppose $X_1, \ldots, X_m$ is a random sample from a distribution F on sample space $\mathcal{X}$. Let $P \sim DP(\alpha H)$ and $P_m^* = (P \mid X_1, \ldots, X_m) \sim DP(\alpha_m^* H_m^*)$, where $\alpha^*_m = \alpha + m$ and $H^*_m = \frac{\alpha}{\alpha+m}H + \frac{m}{\alpha+m}\frac{\sum_{i=1}^m \delta_{X_i}}{m}$. Let $D_{KL}(P^*_m \parallel H^*_m)$ denotes the Kullback-Leibler distance between P^*_m and H^*_m. For any finite partition $\{A_1, \ldots, A_k\}$ of $\mathcal{X}$, define

$$\mathcal{D}(P^*_m, H^*_m) := \alpha^*_m \sum_{i=1}^{k} \frac{(P^*_m(A_i) - H^*_m(A_i))^2}{H^*_m(A_i)}. \tag{4.2.5}$$

Then, as $m \to \infty$, we have

$$2\alpha^*_m D_{KL}(P^*_m \parallel H^*_m) \simeq \mathcal{D}(P^*_m, H^*_m) \xrightarrow{d} \chi^2_{(k-1)}.$$

Proof: We basically mimic the proof for the asymptotic frequentist's chi-squared goodness-of-fit test. Define

$$\mathcal{D}^* = (\alpha + m) \sum_{i=1}^{k} \frac{(P^*_m(A_i) - H^*_m(A_i))^2}{H^*_m(A_i)}. \tag{4.2.6}$$

Let $\mathbf{Y}^T_m = (Y_{1,m}, \ldots, Y_{k,m}) = (P^*_m(A_1), \ldots, P^*_m(A_k))$ and $\mathbf{v}^T_m = (v_{1,m}, \ldots, v_{k,m}) = (H^*_m(A_1), \ldots, H^*_m(A_k))$. By Lemma 4.2.2, as $m \to \infty$,

$$\sqrt{\alpha + m}(\mathbf{Y}_m - \mathbf{v}_m)^T \xrightarrow{d} N_k(\mathbf{0}, \Sigma). \tag{4.2.7}$$

In here, $\Sigma = (\sigma_{ij})_{k \times k}$ is the covariance matrix with $\sigma^2_{ii} = \text{var}(Y_{i,m}) = F(A_i)(1 - F(A_i))$, $i = 1, \ldots, k$ and $\sigma_{ij} = cov(Y_{i,m}, Y_{j,m}) = -F(A_i)F(A_j)$. Then, (4.2.6) can be written as

$$\mathcal{D}^* = (\alpha + m)(\mathbf{Y}_m - \mathbf{v}_m)^T \Sigma^{-1} (\mathbf{Y}_m - \mathbf{v}_m). \tag{4.2.8}$$

Note that the sum of the jth column of Σ is $F(A_j) - F(A_j)(F(A_1) + \cdots + F(A_k)) = 0$, that implies the sum of the rows of Σ is the zero vector, therefore Σ is not invertible. To avoid dealing with this singular matrix, we define $\mathbf{Y}^{*T}_m = (Y_{1,m}, \ldots, Y_{k-1,m})$. Let

$\mathbf{Y}_m^*$ be the vector consisting of the first $k-1$ components of $\mathbf{Y}_m$. Then, the covariance matrix of $\mathbf{Y}_m^*$ is the upper-left $(k-1) \times (k-1)$ sub-matrix of Σ which is denoted by Σ^*. Similarly, let $\mathbf{v}_m^{*T}$ denotes the vector $\mathbf{v}_m^{*T} = (v_{1,m}, \dots, v_{k-1,m})$. It can be verified simply that Σ^* is invertible. Furthermore, (4.2.8) can be rewritten as

$$\mathcal{D}^* = (\alpha + m)(\mathbf{Y}_m^* - \mathbf{v}_m^*)^T (\Sigma^*)^{-1} (\mathbf{Y}_m^* - \mathbf{v}_m^*). \tag{4.2.9}$$

Define

$$\mathbf{Z}_m^T = \sqrt{\alpha + m}(\Sigma^*)^{-1/2} (\mathbf{Y}_m^* - \mathbf{v}_m^*)^T.$$

The central limit theorem implies $\mathbf{Z}_m^T \xrightarrow{d} N_{k-1}(\mathbf{0}, I)$. By definition, the $\chi^2_{(k-1)}$ distribution is the distribution of the sum of the squares of $k-1$ independent standard normal random variables. Therefore,

$$\mathcal{D}^* = \mathbf{Z}_m^T \mathbf{Z}_m \xrightarrow{d} \chi^2_{(k-1)}.$$

∎

Note that as the sample size m increases, $H_m^* \xrightarrow{a.s.} F$ and therefore the posterior Dirichlet process P_m^* converges to the true underlying distribution F of the observed data $X_1, \dots, X_m$. In our methodology, we compute the observed probability for bin A_i, $i = 1, \dots, k$ of the partition $\{A_1, \dots, A_k\}$ by calculating the posterior probability $P_m^*(A_i)$, $i = 1, \dots, k$. Notice that in our Bayesian paradigm, we need to embed our prior information in our test statistic. In other words, the base distribution and the concentration parameter plays the role of the prior knowledge. Moreover, we do not count the observed frequencies in each bin. Instead, we calculate the exact posterior probability for each bin. As a result, k can be very large. Then, the X^2 distance in (4.2.5) compares the posterior probabilities with the hypothesized ones. Additionally, there is no need to apply the asymptotic distribution as we know the exact distribution of the X^2 distance via a Monte Carlo simulation. As we mentioned before the technique suggested here is equivalent to the Bayesian bootstrap introduced in Lo (1987). Theorem 4.2.3 in fact confirms the asymptotic validity and consistency of this Bayesian bootstrap of X^2 test using the Bayesian nonparametric posterior instead of the empirical distribution. Like cases in Lo (1987), the suggested Bayesian bootstrap asymptotically behaves like regular bootstrap. There are several advantages to Bayesian bootstrap (see for example; Rubin (1981) and Lo (1987)). To apply regular bootstrap we need to draw samples with replacement from vectors $\mathbf{e}_i = (0, 0, \ldots, 0, 1^{i-th}, 0, \ldots, 0),\ i = 1, \ldots, n$ (if an observation falls in the i-th bin.) This will be computationally expensive and the method will not work if the expected number of observations is small. Another advantage here is that the Statistician influences the decision making process with his prior knowledge. In addition, asymptotically, poor prior assumptions will be corrected when the sample size is large. There are many discussions for choosing the number of bins in the literature and different criteria are suggested by various authors. See, for example, Hamdan (1963), Dahiya and Gurland (1973), The approach here allow k to be as large as we like. Quine and Robinson (1985) and Johnson (2004). Here, we use the formula $k \simeq (n-1)^{0.4}$ suggested by Koehler and Gan (1990). Other commonly used nonparametric priors such as normalized inverse-Gaussian process and two-parameter Poisson-Dirichlet process do not have the conjugacy property. See, Lijoi et al. (2005) and Carlton (1999), for example. Despite the fact that these priors show similar asymptotic behavior when the concentration parameter is large, we do not use them as priors for our goodness-of-fit test due to the lack of their conjugacy property; See, James (2008) and Al Labadi (2012). In the following subsections, we first use the distance (4.2.5) to find an appropriate concentration parameter for the Dirichlet process. Then, we carry out a Bayesian nonparametric chi-squared goodness-of-fit test.

4.2.1 Selection of the concentration parameter of Dirichlet process

A challenging question in Bayesian nonparametric statistics is to determine α, the concentration parameter of the prior. To suggest an appropriate concentration parameter α, fix c and q such that

$$Pr(\mathcal{D}(P, F_0) \leq c) = q, \tag{4.2.10}$$

where

$$\mathcal{D} = \mathcal{D}(P, F_0) = \alpha \sum_{i=1}^{k} \frac{(P(A_i) - F_0(A_i))^2}{F_0(A_i)}. \tag{4.2.11}$$

Throughout this paper, $\mathcal{D} = \mathcal{D}(P, F_0)$ denotes the prior distance. Also, let $\mathcal{D}^* = \mathcal{D}(P_m^*, F_0)$ stand for the posterior distance as given in (4.2.11), replacing P by P_m^*. We can approximate the distribution of the prior distance $\mathcal{D} = \mathcal{D}(P, F_0)$ by the empirical distribution of N randomly generated values from $\mathcal{D}$. Thus, (4.2.10) can be approximated by the proportion of $\mathcal{D}$ values that are less than or equal to c. We start with an initial value of α and then we compute the probability in (4.2.10). If the probability is close to the value of q, we choose α, otherwise, we repeat this procedure by increasing or decreasing the value of α to reach the value of q. The results of a simulation study for an illustrated example are summarized in Table 1 in Section 6.

4.2.2 Bayesian nonparametric chi-squared goodness-of-fit test

Suppose $X_1, \ldots, X_m$ is a random sample from a distribution F. In order to test the null hypothesis $H_0 : F = F_0$, we place the Dirichlet process prior with parameters α and F_0 on F. Then, since under the null hypothesis, the true distribution of data is F_0, we calculate the distance between the Dirichlet process prior and F_0. The appropriate concentration parameter α of the Dirichlet process can be calculated by the method explained in Subsection 4.1. We follow the approach of Swartz (1999).

That is, for a fixed value of q and c, we obtain α by (4.2.10). Having α, we generate a random sample of size N from the Dirichlet process posterior with parameters α_m^* and H_m^* as given earlier to get N random samples of $\mathcal{D}^* = \mathcal{D}(P_m^*, F_0)$. The distribution of $\mathcal{D}^*$ can be estimated by the empirical distribution of $\mathcal{D}^*$ values. Hence, the posterior probability $Pr(\mathcal{D}(P_m^*, F_0) \leq c)$ can be estimated by the proportion of $\mathcal{D}^*$ which are less than or equal to c. Here, our decision making is based on the comparison of the posterior probability and the prior probability q, where q represents the prior belief that the underlying distribution F is practically equivalent to F_0. Usually $q = 0.5$ is considered. If the empirical posterior probability $Pr(\mathcal{D}(P_m^*, F_0) \leq c)$ is less than q, we reject the null hypothesis, otherwise there is no evidence to reject the null hypothesis. Similar to the frequentist's chi-squared goodness-of-fit test, we can also generalize the test to a family of distributions.

Now, consider the null hypothesis $H_0 : F = F_\theta$ for some $\theta \in \Theta$. Therefore, the true underlying distribution F is a member of a family of distributions indexed by the parameter θ. Our approach for this case is similar to the simple hypothesis with the addition of a prior distribution $\pi(\theta)$ on θ. Thus, the distance $\mathcal{D}(P_m^*, F_\theta)$ depends on the unknown parameter θ. In order to conduct the test, we first generate a random sample from the posterior distribution of θ given $X_1, \ldots, X_m$ that is given as

$$g(\theta \mid X_1, \ldots, X_m) \propto \left(\prod_{i=1}^{m} f_\theta(x_i) \right) \pi(\theta), \tag{4.2.12}$$

where $f_\theta(x)$ is the density function corresponding to F_θ. By having a specified c and q, we find the parameter α such that $Pr(\mathcal{D}(P, F_{\widehat{\theta}}) \leq c) = q$, where $\widehat{\theta} = E(\theta)$. Then, we generate a random sample θ_i^*, $i = 1, \ldots, M$ from the posterior distribution $g(\theta \mid X_1, \ldots, X_m)$. We obtain $\theta_{Min} = \arg\min_{\theta_i^*} D(P_m^*, F_{\theta_i^*})$, $i = 1, \ldots, M$, where P_m^* is the posterior Dirichlet process with the base distribution $H_{\theta_i^*}^*$ as given in (2.1.3) with H replaced by $H_{\theta_i^*}$. We then generate a sample of size N from $\mathcal{D}(P_m^*, F_{\theta_{Min}})$. Similar to the case of testing for the simple hypothesis, the decision is made by comparing the posterior probability $Pr(\mathcal{D}(P_m^*, F_{\theta_{Min}}) \leq c)$ and q. Note that in the case of a non-standard distribution in (4.2.12), in order to sample from the posterior distribution, we need to apply some specialized techniques such as Metropolis-Hastings algorithm. In Section 6, some examples with simulation studies are illustrated for the simple hypothesis $H_0 : F = N(0,1)$ and also the null hypothesis $H_0 : F = \mathcal{E}xp(\theta)$ with a Gamma (1.7, 2550) prior distribution for θ.

4.3 Bayesian nonparametric chi-squared test of independence

Here, we describe a Bayesian nonparametric chi-squared test of independence of two random variables. The null hypothesis of the chi-squared test of independence is given as $H_0 : F_{X,Y}(x,y) = F_X(x)F_Y(y)$ against the alternative $H_0 : F_{X,Y}(x,y) \neq F_X(x)F_Y(y)$ and hence it examines whether there is a significant relationship between two random variables X and Y. Suppose $\{A_j\}_{j=1,\ldots,r}$ is a partition of the space $\mathcal{X}$ of the random variable X and $\{B_k\}_{k=1,\ldots,s}$ is a partition of the space $\mathcal{Y}$ of the random variable Y, i.e., $\mathcal{X} = \overset{r}{\underset{j=1}{\cup}} A_j$ and $\mathcal{Y} = \overset{s}{\underset{k=1}{\cup}} B_k$. Let $(X_l, Y_l) \overset{i.i.d}{\sim} F(x,y)$, $l = 1,\ldots,m$ be the sample data and H be a bivariate distribution. Then, the Dirichlet process posterior with parameters H_m^* and α_m^* is written as $P_m^* = \sum_{i=1}^{\infty} p_i^{(m)} \delta_{(X_i^*,Y_i^*)}$, where $p_i^{(m)}$ is as given in (2.1.4), α is replaced by α_m^* and (X_i^*, Y_i^*), $i = 1,\ldots,n$ are generated from $H_m^* = \frac{\alpha}{\alpha+m}H + \frac{m}{\alpha+m}\frac{\sum_{i=1}^m \delta_{(X_i,Y_i)}}{m}$. In our new approach, we compute the observed probability at level j of the random variable X and at level k of the random variable Y by $P_m^*(A_j \times B_k)$ and the corresponding expected probability is computed as $P_m^*(A_j \times \mathcal{Y})P_m^*(\mathcal{X} \times B_k)$, where

$$P_m^*(A_j \times B_k) = \sum_{i=1}^{\infty} p_i^{(m)} \delta_{(X_i^*,Y_i^*)}(A_j \times B_k) \tag{4.3.1}$$

and

$$P_m^*(A_j \times \mathcal{Y}) = \sum_{i=1}^{\infty} p_i^{(m)} \delta_{(X_i^*, Y_i^*)}(A_j \times \mathcal{Y}) = \sum_{i=1}^{\infty} p_i^{(m)} \delta_{X_i^*}(A_j)$$
$$P_m^*(\mathcal{X} \times B_k) = \sum_{i=1}^{\infty} p_i^{(m)} \delta_{(X_i^*, Y_i^*)}(\mathcal{X} \times B_k) = \sum_{i=1}^{\infty} p_i^{(m)} \delta_{Y_i^*}(B_k). \qquad (4.3.2)$$

Then, test statistic is given as

$$\mathcal{D}^* = \alpha_m^* \sum_{k=1}^{s} \sum_{j=1}^{r} \frac{(P_m^*(A_j \times B_k) - P_m^*(A_j \times \mathcal{Y}) P_m^*(\mathcal{X} \times B_k))^2}{P_m^*(A_j \times \mathcal{Y}) P_m^*(\mathcal{X} \times B_k)} \qquad (4.3.3)$$

which asymptotically converges to $\chi^2_{(r-1)\times(s-1)}$. In order to carry out the test, we proceed a similar process as explained in Section 4 for the goodness-of-fit test. We generate a random sample of size N from the prior distance $\mathcal{D}$, where $\mathcal{D}$ is computed by (4.3.3) replacing α_m^* by α and the Dirichlet process posterior P_m^* by the Dirichlet process prior P. By having a fixed value c and a fixed probability q, an appropriate concentration parameter α is obtained by the equation $Pr(\mathcal{D} \leq c) = q$. Then, by generating a sample of size N from $\mathcal{D}^*$, we can approximate the distribution of $\mathcal{D}^*$ by the empirical distribution of $\mathcal{D}^*$ values. Our decision is made by comparing the probabilities $Pr(\mathcal{D}^* \leq c)$ and q and we reject the null hypothesis if $Pr(\mathcal{D}^* \leq c)$ is less than q. An illustrative example with a simulation study is discussed in the following section.

4.4 Simulation study

This section provides some examples with simulation studies for the Bayesian nonparametric tests described in the previous sections of this chapter. For all the simulations, we use the finite sum representation (2.1.8) to approximate the Dirichlet process.

Example 4.4.1. We consider a Dirichlet process with the base distribution $H = N(0,1)$ and $n = 2000$ terms in the finite sum representation (2.1.8). We partition the space into $k = 7$ bins. Table 4.1 represents the probability (4.2.10) when $F_0 = N(0,1)$. The probabilities are computed for various values of α and c and for a simulation of size $N = 2000$. As the Table 4.1 shows, for example, if we set $q = 0.48$ and $c = 3$, $\alpha = 10$ is an appropriate concentration parameter.

Table 4.1: The computed probability $Pr\left(\mathcal{D}(P, F_0) < c\right)$ for different choices of α and c in Example 1.

α	$c=1$	$c=2$	$c=3$	$c=4$	$c=5$	$c=6$
1	0.298	0.745	0.812	0.857	0.893	0.933
10	0.068	0.273	0.480	0.624	0.717	0.781
50	0.029	0.143	0.311	0.474	0.612	0.696
100	0.027	0.116	0.258	0.409	0.540	0.648
200	0.020	0.094	0.219	0.353	0.492	0.595
300	0.011	0.073	0.179	0.297	0.432	0.542
500	0.009	0.057	0.150	0.263	0.368	0.484

Example 4.4.2. Suppose $X_1, \ldots, X_{150}$ is a random sample from a standard Cauchy distribution. We want to test the null hypothesis $H_0 : F = N(0,1)$. We divide the sample space into $k = 7$ bins A_i, $i = 1, \ldots, 7$ and $Pr(A_i)$ shows the observed probability of each bin. We consider $H = N(0,1)$ as the base measure and $n = 2000$ terms in the finite sum representation of Dirichlet process as given in (2.1.8). Then, from Table 4.1, an appropriate concentration parameter $\alpha = 100$ is obtained when $q = 0.54$ and $c = 5$. By sampling $N = 2000$ times from the Dirichlet process posterior P_m^* and then $N = 2000$ realizations of $\mathcal{D}^*$, we obtain $Pr(\mathcal{D}(P_m^*, F_0) \leq c) \simeq 0$. Thus, we reject the normality hypothesis of the data. Our decision is consistent with the classical chi-squared test which gives a p-value of 2.2×10^{-16}. Also, our decision

is consistent with other choices of the base measure H, since the Dirichlet process posterior converges to the true underlying distribution as the data size increases. Table 4.2 illustrates the observed probabilities obtained by counting the data points in each bin and the corresponding probabilities computed by the Dirichlet process posterior.

Table 4.2: The computed probabilities $Pr(A_i)$, $P_m^*(A_i)$ and $F_0(A_i)$, where $Pr(A_i)$ is the observed probability obtained by counting the data points in ith bin, $P_m^*(A_i)$ is the corresponding probability computed by the Dirichlet process posterior for one simulation and $F_0(A_i)$ shows the corresponding expected probability under the null hypothesis

	$Pr(A_i)$	$P_m^*(A_i)$	$P_m^*(A_i)$
$A_1 = (-\infty, -2]$	0.133	0.072	0.023
$A_2 = (-2, -1]$	0.100	0.131	0.136
$A_3 = (-1, 0]$	0.313	0.342	0.341
$A_4 = (0, 1]$	0.240	0.310	0.341
$A_5 = (1, 2]$	0.060	0.069	0.136
$A_6 = (2, 3]$	0.067	0.030	0.022
$A_7 = (3, \infty)$	0.087	0.046	0.001

Figures 4.1, 4.2 and 4.3 show the Q-Q plot, the empirical distribution and the histogram of $N = 2000$ randomly generated from the prior distance $\mathcal{D} = \mathcal{D}(P, F_0)$ compared with a $\chi^2_{(4)}$ distribution, respectively.

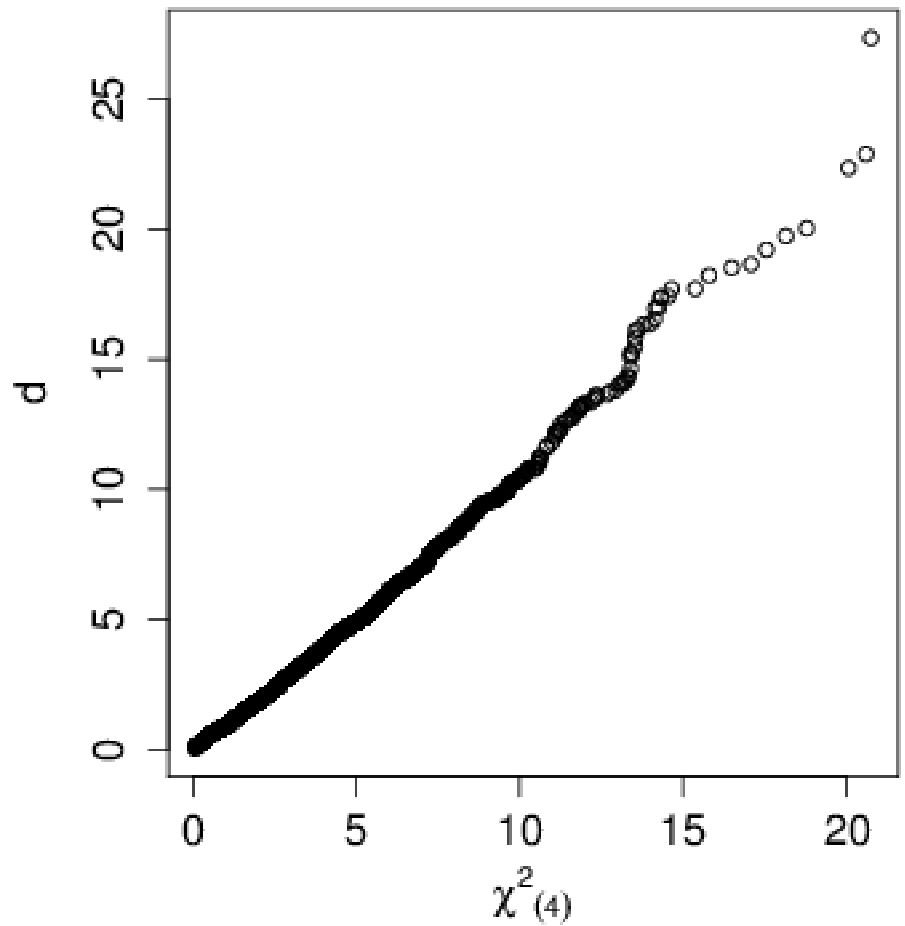

Figure 4.1: The Q-Q plot of $N = 2000$ realizations of $\mathcal{D} = \mathcal{D}(P, H)$ with $\alpha = 100$, $H = N(0, 1)$, $k = 5$ and $n = 3000$ versus a $\chi^2_{(4)}$ distribution.

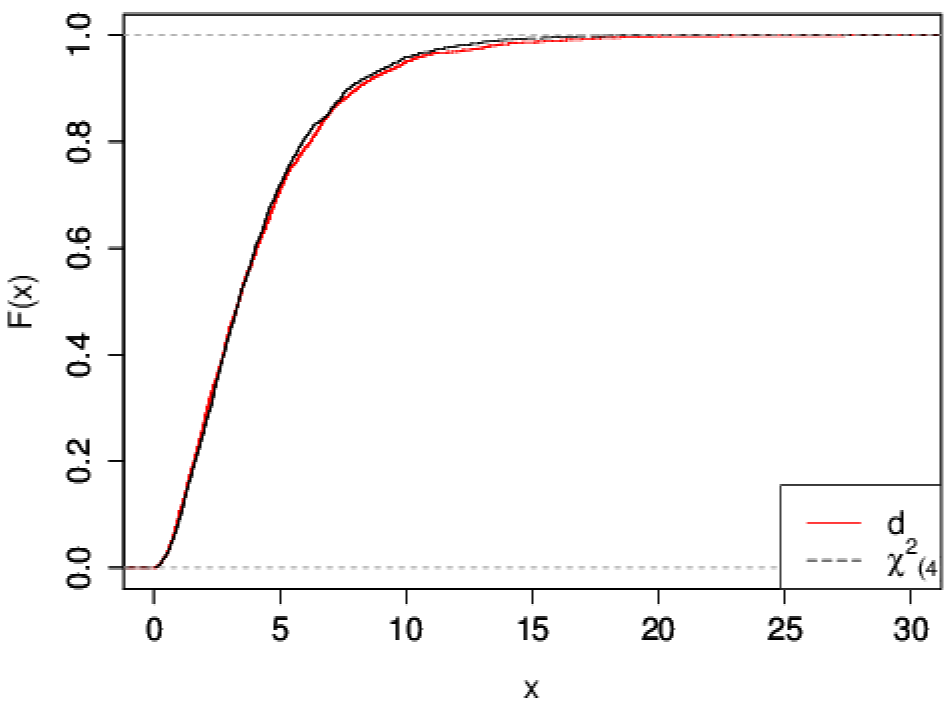

Figure 4.2: The empirical distribution function of $N = 2000$ realizations of $\mathcal{D} = \mathcal{D}(P, H)$ with $\alpha = 100$, $H = N(0, 1)$, $k = 5$ and $n = 3000$ and the cdf of a $\chi^2_{(4)}$ distribution.

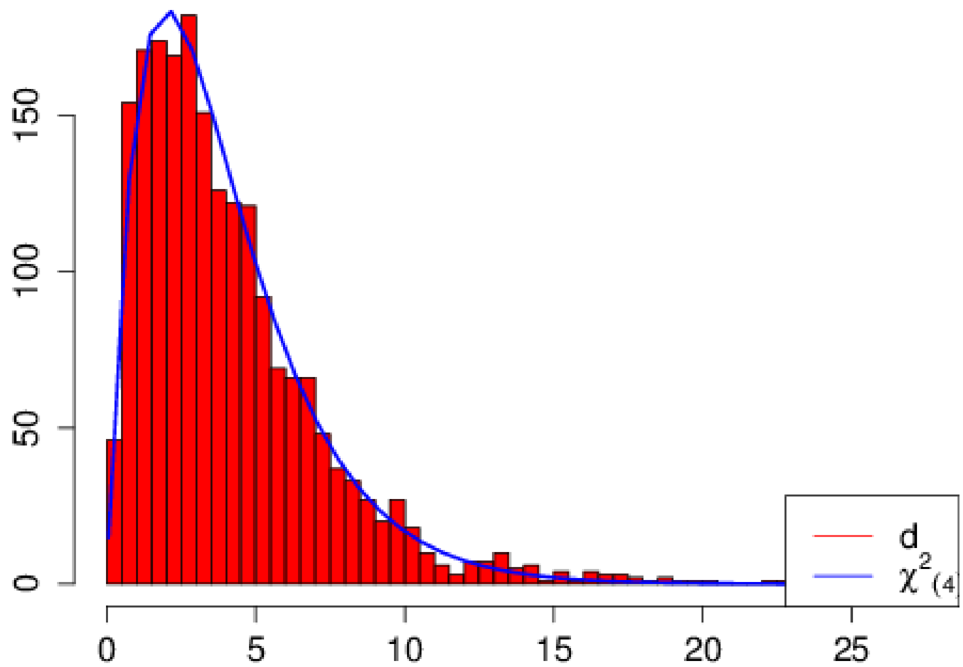

Figure 4.3: The histogram of $N = 2000$ realizations of $\mathcal{D} = \mathcal{D}(P, H)$ with $\alpha = 100$, $H = N(0, 1)$, $k = 5$ and $n = 3000$ and the pdf of a $\chi^2_{(4)}$ distribution.

Example 4.4.3. (Example 3.6. Hamada et al. (2008)) Suppose we have an observed data of size $m = 31$ for the lifetime of the liquid crystal display (LCD) projector lamps. We want to test if the lifetime distribution of the liquid crystal display (LCD) projector lamps is an exponential distribution with parameter $\theta > 0$. That is, we want to test the null hypothesis $H_0 : F_\theta = \mathcal{E}xp(\theta)$, where θ has a Gamma (1.7, 2550) prior distribution. Hence, the posterior distribution of θ given data is a Gamma (32.7, 20457) distribution. We consider $k = 4$ bins $A_1 = (-\infty, 0.25], A_2 = (0.25, 0.5], A_3 = (0.5, 0.75]$ and $A_4 = (0.75, \infty)$. By specifying the values $q = 0.51$ and $c = 3$, the appropriate $\alpha = 100$ is obtained. We obtain $\theta_1^*, \ldots, \theta_M^*$ as realizations from the distribution of $(\theta \mid X_1, \ldots, X_{31})$ and we get $\theta_{Min} = 0.00136$. By generating $N = 2000$ times from $\mathcal{D}^* = \mathcal{D}(P_m^*, F_{\theta_{Min}})$, we obtain $Pr(\mathcal{D}(P_m^*, F_{\theta_{Min}}) \leq c) = 0.71$. Hence, there is no evidence to reject the null hypothesis.

Example 4.4.4. Suppose we have a random sample (X_i, Y_i), $i = 1, \ldots, 150$ from a bivariate standard Cauchy distribution. We consider five levels of variable X and four levels of variable Y as given in Table 4.3. We want to test the null hypothesis of independence as given in Section 5. Consider a Dirichlet process prior with the base standard Student's t distribution with 4 degrees of freedom, i.e., $H = t_{(4)}$. For $q = 0.34$ and $c = 20$, by generating $N = 2000$ times from $\mathcal{D}$ and solving the equation $Pr(\mathcal{D} < c) = q$, we obtain an appropriate concentration parameter $\alpha = 100$. By generating a sample of size $N = 2000$ from the posterior distance $\mathcal{D}^*$, we have $Pr(\mathcal{D}^* < c) \simeq 0$. Therefore, we reject the null hypothesis of independence. The p-value of 7.5×10^{-3} obtained by the Fisher chi-squared test of independence results in the same conclusion. Table 4.3 represents the probability of each category calculated by the Dirichlet process posterior.

Table 4.3: A sample table of probabilities computed by the Dirichlet process posterior in Example 4.

	$A_1 = (-\infty, -1]$	$A_2 = (-1, 0]$	$A_3 = (0, 1]$	$A_4 = (1, 2]$	$A_5 = (2, \infty)$
$B_1 = (-\infty, -1]$	0.105	0.028	0.006	0.044	0.025
$B_2 = (-1, 0]$	0.077	0.090	0.088	0.038	0.000
$B_3 = (0, 1]$	0.092	0.026	0.067	0.067	0.006
$B_4 = (1, \infty)$	0.031	0.043	0.068	0.053	0.046

Chapter 5

Bayesian Nonparametric Inference for Spherically Symmetric Distributions

This chapter discusses the extension of Dirichlet invariant processes to bivariate spherically symmetric distributions, where the underlying distribution is invariant under rotations. We first begin with a finite group of rotations $\mathscr{G}$ such that $\mid \mathscr{G} \mid= k$. Then, we use Dirichlet invariant process to consider the case when $k \to \infty$.

5.1 Dirichlet invariant process for spherically symmetric distributions

In various situations, some evidence shows that the underlying distribution of interest is symmetric. In this case, assuming the priors that impose additional assumption of symmetry about the distribution leads to more reliable and accurate estimation of the parameter of interest.

Dalal (1979a) considered a special case of this problem when the unknown underlying distribution F is assumed to be symmetric about a known center of symmetry μ. Specifically, suppose that $X_1, \ldots, X_m$ is a sample from an unknown distribution F symmetric about μ, where μ is known. Here, the set $\mathscr{X}$ is the space of all distribution functions F on the real line $\mathbb{R}$ and symmetric about μ. Place a Dirichlet invariant process $P \sim DIP(\alpha H)$ as the prior on F, where H is a finite measure symmetric about μ. Then, under the invariant transformation group $\mathscr{G}(x) = \{x, 2\mu - x\}$, the Dirichlet invariant process posterior is obtained with the base distribution

$$H_m^* = \frac{\alpha H + \frac{1}{2}\sum_{i=1}^{m}\left(\delta_{X_i} + \delta_{2\mu - X_i}\right)}{\alpha + m},$$

where $F_m = \frac{1}{2m}\sum_{i=1}^{m}\left(\delta_{X_i} + \delta_{2\mu - X_i}\right)$ is the F-symmetrized version of the empirical distribution function. Further, from the Glivenko-Cantelli theorem and the fact that $p_m = \frac{\alpha}{\alpha+m} \to 0$ as $m \to \infty$, it follows that H_m^* converges to the true distribution function uniformly almost surely. The aim is to mimic Dirichlet invariant process for constructing a prior on the space of spherically symmetric distributions on $\mathbb{R}^p$. Indeed, our approach is an extension of Dirichlet invariant process for the symmetry to the spherical symmetry case. For simplicity, we consider $p = 2$. However, a generalization is discussed in forthcoming section. We first recall the definition of the spherically symmetric distribution.

Definition 5.1.1. *A random vector $\boldsymbol{X} \in \mathbb{R}^p$ is spherically symmetric if $\boldsymbol{X} \stackrel{d}{=} A\boldsymbol{X}$ i.e., $\boldsymbol{X}$ and $A\boldsymbol{X}$ have the same distribution for all $p \times p$ orthogonal matrices A. In other words, $\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|}$ is independent from $\| \boldsymbol{X} \|$.*

Theorem 5.1.2. Let $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$ be any two-dimensional Euclidean space with associated Borel σ-field. Let $\mathcal{P} = \{(0, \theta_1], (\theta_1, \theta_2], \ldots, (\theta_{k-1}, 2\pi]\}$ be a partition of the interval $[0, 2\pi]$ formed by k equally spaced subintervals. Then, $\mathscr{G}(\boldsymbol{X}) = \{g_1(\boldsymbol{X}) = A_{\theta_1}\boldsymbol{X}, g_2(\boldsymbol{X}) = A_{\theta_2}\boldsymbol{X}, \ldots, g_k(\boldsymbol{X}) = A_{\theta_k}\boldsymbol{X}\}$ is a finite group of orthogonal transformations corresponding to the partition $\mathcal{P}$, where $\theta_k = 2\pi$ and $A_{\theta_j} = \begin{pmatrix} \cos\theta_j & -\sin\theta_j \\ \sin\theta_j & \cos\theta_j \end{pmatrix}$, $j = 1, \ldots, k$. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ be a sample from an unknown bivariate spherically symmetric distribution F on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$. Consider a Dirichlet invariant process prior $P \sim DIP(\alpha H)$ on F. Then, the Dirichlet invariant process posterior is obtained with parameters $\alpha_m^* = \alpha + m$ and

$$H_{k,m}^* = \frac{\alpha H + \frac{1}{k}\sum_{j=1}^{k}\sum_{i=1}^{m}\delta_{A_{\theta_j}\boldsymbol{X}_i}}{\alpha + m}. \tag{5.1.1}$$

Proof: Since the random vector $\boldsymbol{X}$ is spherically symmetric, we have $\boldsymbol{X} \stackrel{d}{=} A_\Theta \boldsymbol{X}$, where

$$A_\Theta = \begin{pmatrix} \cos\Theta & -\sin\Theta \\ \sin\Theta & \cos\Theta \end{pmatrix}$$

for all $\Theta \sim U[0, 2\pi]$. Consider the partition $\mathcal{P}$ and the invariant group of transformations $\mathscr{G}$. Notice that $A_{2\pi} = I$ and $A_{\theta_i}A_{\theta_j} = A_{\theta_i+\theta_j}$. We place a Dirichlet invariant process prior $P \sim DIP(\alpha H)$ on F, where H is a bivariate invariant measure on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$. Then, by Theorem 2.2.6, the conditional distribution of P given $\boldsymbol{X}_1, ..., \boldsymbol{X}_m$ is $DIP(\alpha H + \sum_{j=1}^{m}\delta^g_{\boldsymbol{X}_i})$, where $\delta^g_{\boldsymbol{X}_i} = \frac{1}{k}\sum_{j=1}^{k}\delta_{g_j(\boldsymbol{X}_i)}$. By the group $\mathscr{G}$, we have $g_j(\boldsymbol{X}_i) = A_{\theta_j}\boldsymbol{X}_i$, $(j = 1, \ldots, k)$ for each $\boldsymbol{X}_i$, $(i = 1, \ldots, m)$ and

$$\delta^g_{\boldsymbol{X}_i} = \frac{1}{k}\sum_{j=1}^{k}\delta_{g_j(\boldsymbol{X}_i)} = \frac{1}{k}\sum_{j=1}^{k}\delta_{A_{\theta_j}\boldsymbol{X}_i}.$$

Therefore, we have

$$H_{k,m}^* = \frac{\alpha H + \frac{1}{k}\sum_{j=1}^{k}\sum_{i=1}^{m}\delta_{A_{\theta_j}\boldsymbol{X}_i}}{\alpha + m}.$$

∎

The next lemma proves the limiting distribution of Dirichlet invariant process for any finite Borel sets $B_1, \ldots, B_N \in \mathscr{B}(\mathscr{X})$. In this thesis, " $\xrightarrow{d}$" denotes convergence in distribution.

Lemma 5.1.3. *Suppose the assumptions of Theorem 5.1.2 hold. Let $P^*_{k,m} \sim DIP(\alpha^*_m H^*_{k,m})$ be the Dirichlet process posterior. Then, as $k \to \infty$, for any fixed sets $B_1, \ldots, B_N \in \mathscr{B}(\mathscr{X})$,*

$$\left(P^*_{k,m}(B_1), \ldots, P^*_{k,m}(B_N)\right) \xrightarrow{d} \left(P^*_m(B_1), \ldots, P^*_m(B_N)\right),$$

where $P^*_m \sim DP(\alpha^*_m H^*_m)$ is Dirichlet process posterior with parameters $\alpha^*_m = \alpha + m$ and

$$H^*_m = \frac{\alpha H + \sum_{j=1}^{m} F_{A_\Theta \boldsymbol{X}_j}}{\alpha + m}.$$

Proof: Without loss of generality, let $N = 2$. Consider any arbitrary partition $\{B_1, B_2\}$ of the space $\mathscr{X}$. Note that

$$\begin{aligned} & \left(P^*_{k,m}(B_1), P^*_{k,m}(B_2)\right) \\ \sim\ & Dir\left(\alpha^*_m H^*_{k,m}(B_1), \alpha^*_m H^*_{k,m}(B_2), \alpha^*_m\left(1 - H^*_{k,m}(B_1) - H^*_{k,m}(B_2)\right)\right) \end{aligned}$$

Set $Y_{k,i} = P^*_{k,m}(B_i)$ and $v_{k,i} = H^*_{k,m}(B_i)$, $i = 1, 2$. Then, the joint density function of $\boldsymbol{Y} = (Y_{k,1}, Y_{k,2})$ is

$$\begin{aligned} f_{Y_{k,1},Y_{k,2}}(y_{k,1}, y_{k,2}) \;=\; & \frac{\Gamma\left(\alpha^*_m\right)}{\Gamma(\alpha^*_m v_{k,1})\Gamma(\alpha^*_m v_{k,2})\Gamma(\alpha^*_m(1 - v_{k,1} - v_{k,2}))} y_{k,1}^{\alpha^*_m v_{k,1}-1} y_2^{\alpha^*_m v_{k,2}-1} \\ & \times(1 - y_{k,1} - y_{k,2})^{\alpha^*_m(1-v_{k,1}-v_{k,2})-1}. \end{aligned}$$

By using Scheffé's lemma, we only need to prove that

$$\begin{aligned} f_{Y_{k,1},Y_{k,2}}(y_{k,1}, y_{k,2}) \;\to\; & \frac{\Gamma\left(\alpha^*_m\right)}{\Gamma(\alpha^*_m t_1)\Gamma(\alpha^*_m t_2)\Gamma(\alpha^*_m(1 - t_1 - t_2))} y_1^{\alpha^*_m t_1-1} y_2^{\alpha^*_m t_2-1} \\ & \times(1 - y_1 - y_2)^{\alpha^*_m(1-t_1-t_2)-1} \end{aligned}$$

as $k \to \infty$, where $t_i = H^*_m(B_i)$, $i = 1, 2$. We need to find $\lim\limits_{k\to\infty} f_{Y_{k,1},Y_{k,2}}(y_{k,1}, y_{k,2})$. Therefore, we only need to show that $v_{k,i} = H^*_{k,m}(B_i) \to H^*_m(B_i) = t_i$ as $k \to \infty$. We have

$$\lim_{k\to\infty} H^*_{k,m} = \frac{\alpha H + \sum\limits_{i=1}^{m} \lim\limits_{k\to\infty} \sum\limits_{j=1}^{k} \delta_{A_{\theta_j}\boldsymbol{X}_i}/k}{\alpha + m},$$

where limit holds with respect to weak topology. Let $G = \{A_\theta \mid 0 \le \theta \le 2\pi\}$, where $A_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. Define

$$G_m = \left\{ A_\theta \mid \theta = \frac{2k\pi}{2^m},\ k = 0, 1, \dots, 2^m \right\},$$

where $m \ge 1$. Then, G_m is a finite subgroup of G and $\mid G_m \mid = 2^m$. Define $K = \bigcup\limits_{m=1}^{\infty} G_m$. Notice that for any $A_\theta \in G$, there is a $B \in K$ such that $\parallel A_\theta - B \parallel < \varepsilon$. We have

$$[0, 2\pi] = \bigcup_{k=0}^{2^m-1} \left[\frac{2k\pi}{2^m}, \frac{2(k+1)\pi}{2^m} \right].$$

Take $m \in \mathbb{N}$ such that $\frac{\pi}{2^{m-1}} < \frac{\varepsilon}{2}$. Hence, there exists an r, $(0 \le r \le 2^m)$ such that $\theta \in \left[\frac{2r\pi}{2^m}, \frac{2(r+1)\pi}{2^m}\right]$. Now let $B = \begin{pmatrix} \cos\delta & -\sin\delta \\ \sin\delta & \cos\delta \end{pmatrix}$, where $\delta = \frac{2r\pi}{2^m}$. Then, by using the Euclidean norm, we have

$$\parallel A_\theta - B \parallel = \sqrt{2} \sqrt{\left(\cos\theta - \cos\frac{2r\pi}{2^m}\right)^2 + \left(\sin\theta - \sin\frac{2r\pi}{2^m}\right)^2} \tag{5.1.2}$$

On the other hand, we have

$$\mid \cos\theta - \cos\frac{2r\pi}{2^m} \mid \le \mid \theta - \frac{2r\pi}{2^m} \mid \le \frac{\pi}{2^{m-1}} < \frac{\varepsilon}{2}. \tag{5.1.3}$$

Similarly, $\mid \sin\theta - \sin\frac{2r\pi}{2^m} \mid < \frac{\varepsilon}{2}$. Then, by (5.1.2),

$$\parallel A_\theta - B \parallel < \varepsilon.$$

Therefore, for any closed set C from strong law of large numbers we have

$$\lim_{k\to\infty}\sum_{j=1}^{k}\delta_{A_{\theta_j}\boldsymbol{X}_i}(C)/k = E\left(\delta_{A_\Theta \boldsymbol{X}_i}(C)\right) = F_{A_\Theta \boldsymbol{X}_1}(C).$$

Hence,

$$\sum_{j=1}^{k}\delta_{A_{\theta_j}\boldsymbol{X}_i}/k \xrightarrow{d} F_{A_\Theta \boldsymbol{X}_1},$$

where $F_{A_\Theta \boldsymbol{X}_1}(\cdot)$ is the probability measure for the random variable $A_\Theta \boldsymbol{X}_1$. ∎

Lemma 5.1.3 proves that the finite-dimensional distributions of the process $P^*_{k,m}$ converge to the corresponding finite-dimensional distribution of P^*_m. In the following theorem, we obtain the Dirichlet invariant process posterior for the infinite group, i.e., when $\mid \mathscr{G} \mid = k \to \infty$. Indeed, we prove that the Dirichlet invariant process approaches to Dirichlet process as $k \to \infty$ with respect to the weak topology.

Theorem 5.1.4. Suppose the assumptions of Theorem 5.1.2 hold. Consider a Dirichlet invariant process prior $P \sim DIP(\alpha H)$ on F and denote the corresponding posterior Dirichlet invariant process by $P^*_{k,m} \sim DIP(\alpha^*_m H^*_{k,m})$. Then, as $k \to \infty$,

$$P^*_{k,m} \xrightarrow{d} P^*_m$$

Proof: The convergence for the finite dimensional distributions is proved in Lemma 5.1.3. Now, let $\beta = \beta_1 + \beta_2 > 1$, $\gamma = \gamma_1 + \gamma_2 > 0$ and μ be a finite nonnegative measure on $T = [0,1] \times [0,1]$. Consider two neighboring blocks $C = (s,t] \times (a,b]$ and $D = (t,u] \times (a,b]$ in T, where $s \le t \le u$. By (Bickel and Wichura, 1971) and Theorem

13.5 of (Billingsley, 2013), we need to prove that

$$E\left[(P^*_{k,m}(C))^{\gamma_1}(P^*_{k,m}(D))^{\gamma_2}\right] \le (\mu(C))^{\beta_1}(\mu(D))^{\beta_2},$$

where β_1, β_2, γ_1 and γ_2 satisfy $\beta_1+\beta_2 > 1$ and $\gamma_1+\gamma_2 > 0$. Take $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = 1$ and let $\mu(\cdot) = \frac{\alpha\lambda(\cdot)+m}{\alpha+m}$, where λ is the Lebesgue measure on $T = [0,1] \times [0,1]$. Then, it is enough to show that

$$\begin{aligned}&E\left[P^*_{k,m}(C)P^*_{k,m}(D)\right] \le \\ &\left(\frac{\alpha}{\alpha^*}\right)^2\left((b-a)^2(t-s)(u-t) + \frac{m}{\alpha}(b-a)(u-s) + \left(\frac{m}{\alpha}\right)^2\right).\end{aligned}$$

Without loss of generality, we only consider the distribution H on $T = [0,1] \times [0,1]$, where $H((u_1,u_2] \times (w_1,w_2]) = (u_2-u_1)(w_2-w_1)$. Note that

$$\begin{aligned}&\left(P^*_{k,m}(C), P^*_{k,m}(D)\right)\\ \sim\ &Dir\left(\alpha^*_m H^*_{k,m}(C), \alpha^*_m H^*_{k,m}(D), \alpha^*_m\left(1 - H^*_{k,m}(C) - H^*_{k,m}(D)\right)\right)\end{aligned}$$

and

$$E\left[P^*_{k,m}(C)P^*_{k,m}(D)\right] = \frac{\alpha^*}{\alpha^*+1}H^*_{k,m}(C)H^*_{k,m}(D).$$

We have

$$\begin{aligned}&H^*_{k,m}(C)H^*_{k,m}(D) \le \\ &\left(\frac{\alpha}{\alpha^*}\right)^2\left((b-a)^2(t-s)(u-t) + \frac{m}{\alpha}(b-a)(u-s) + \left(\frac{m}{\alpha}\right)^2\right).\end{aligned}$$

Then,

$$E\left[P^*_{k,m}(C)P^*_{k,m}(D)\right] \le$$

$$\left(\frac{\alpha^*}{\alpha^*+1}\right)\left(\frac{\alpha}{\alpha^*}\right)^2\left((b-a)^2(t-s)(u-t)+\frac{m}{\alpha}(b-a)(u-s)+\left(\frac{m}{\alpha}\right)^2\right).$$

∎

In the following, we provide an algorithm which can be used in computational studies. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ be a sample distributed with an unknown bivariate spherically symmetric distribution F. Consider a Dirichlet invariant process with parameters α and H, where H is a bivariate spherically symmetric distribution. The algorithm below explains the steps of generating sample distributions from a Dirichlet invariant process posterior $DP(\alpha_m^* H_{k,m}^*)$.

Algorithm 2 Generating a sample posterior distribution from a Dirichlet invariant process posterior for a spherically symmetric distribution.

Let k be a large positive integer. Follow the steps 1 to 5 below:

1. Consider the jump points $\theta_i = 2\pi i/k, (i = 0, \ldots, k-1)$ with equal jump sizes from uniform distribution $U[0, 2\pi]$.

2. For each vector $\boldsymbol{X}_i$, $(i = 1, \ldots, m)$, compute the values $A_{\theta_j}\boldsymbol{X}_i$, $(j = 1, \ldots, k)$, where $A_{\theta_j} = \begin{pmatrix} \cos\theta_j & -\sin\theta_j \\ \sin\theta_j & \cos\theta_j \end{pmatrix}$.

3. Compute the empirical distribution of $A_{\theta_j}\boldsymbol{X}_i$'s, where $(j = 1, \ldots, k)$ and $(i = 1, \ldots, m)$ by assigning the equal weight $\frac{1}{km}$ to each vector $A_{\theta_j}\boldsymbol{X}_i$.

4. Generate n observations $\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_n^*$ from $H_{k,m}^*$, where n is large enough and

$$H_{k,m}^* = \frac{\alpha H + \frac{1}{k}\sum\limits_{j=1}^{k}\sum\limits_{i=1}^{m}\delta_{A_{\theta_j}\boldsymbol{X}_i}}{\alpha + m}.$$

5. Generate a sample posterior distribution $P_{k,m}^*$ from Dirichlet invariant process posterior with parameters $\alpha_m^* = \alpha + m$ and $H_{k,m}^*$ using Algorithm 1 as given in Chapter 2 by replacing the parameters α and H with α_m^* and $H_{k,m}^*$, respectively.

5.2 Symmetry in high-dimensional Euclidean space

The approach of Section 5.1 can be generalized to high-dimensional Euclidean space. Consider the distribution which is symmetric about one of the axes of a coordinate system. Then the rotation is performed around a certain axis. The following three basic orthogonal matrices A_{θ_x}, A_{θ_y} and A_{θ_z} rotate vectors counterclockwise by angles

θ_x, θ_y and θ_z about the x, y, or z-axis, respectively, in three dimensional space.

$$A_{\theta_x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix}, A_{\theta_y} = \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{bmatrix}$$

$$A_{\theta_z} = \begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Then, other rotation matrices can be obtained from these three matrices using matrix multiplication. For example, the product

$$A_{\theta_x} A_{\theta_y} A_{\theta_z}$$

$$= \begin{bmatrix} \cos\theta_y \cos\theta_z & -\cos\theta_y \sin\theta_z & \sin\theta_y \\ \cos\theta_x \sin\theta_z + \sin\theta_x \sin\theta_y \cos\theta_z & \cos\theta_x \cos\theta_z - \sin\theta_x \sin\theta_y \sin\theta_z & -\sin\theta_x \cos\theta_y \\ \sin\theta_x \sin\theta_z - \cos\theta_x \sin\theta_y \cos\theta_z & \sin\theta_x \cos\theta_z + \cos\theta_x \sin\theta_y \sin\theta_z & \cos\theta_x \cos\theta_y \end{bmatrix}$$

represents a rotation whose yaw, pitch, and roll angles are A_{θ_x}, A_{θ_y} and A_{θ_z}, respectively. Therefore, the equation (5.1.1) can be generalized based on the orthogonal matrix corresponding to the angles of the rotation for any multidimensional spherically symmetric distribution. Also, Algorithm 2 can be generalized to the multidimensional case easily.

5.3 Notes and remarks

Assuming orthogonal symmetry, the empirical part of the base distribution of the posterior will be augmented by uncountable many points. In regular symmetry for univariate case, the empirical part includes points of the form $\{X_i : i = 1, 2, \ldots, m\} \cup$

$\{2\mu - X_i; i = 1, 2, \ldots, m\}$. Therefore, the assumption of spherical symmetry imposes much greater impact in making inference compared to the regular symmetry. On the other hand, it seems that the general group requirement can be relaxed. For example, in the case where X has an unknown distribution F with $E(X) = 0$, the Dirichlet invariant process posterior should be obtained using the base distribution

$$H_m^* = \frac{\alpha H + \sum_{i=1}^{m} \delta_{X_i - \overline{X}}}{\alpha + m}.$$

Chapter 6

Simulation Study

In this chapter, in order to demonstrate the performance of our Bayesian nonparametric approach in Chapter 5, we carry out some simulation studies to generate Bayesian nonparametric posterior sample paths for spherically symmetric distributions. In addition, we provide an algorithm for generating posterior paths for distributions symmetric with respect to both lines $x = a$ and $y = b$, where $a, b \in \mathbb{R}$. Finally, we explain a Bayesian nonparametric bootstrapping scheme to approximate the distributions of certain functionals of the support for these symmetric distributions.

6.1 Simulated posterior paths for spherically symmetric distribution

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ be sample data from a spherically symmetric distribution F. We draw different sample paths from Dirichlet invariant posterior process by considering various parameters for our model. We also compare our results obtained by $DIP(\alpha H)$ with the ones generated by $DP(\alpha H)$ for different choices of the parameters α and H. It should be mentioned that the approach presented in Chapter 5 can be applied in hypothesis testing for the symmetric distributions and the spherical symmetry of a distribution. Indeed, like the goodness-of-fit test which was introduced in Chapter 4, we can perform a similar test for the spherical symmetric distributions on the basis of the Dirichlet process posterior extracted for these distributions. Then, the simulations studies and the posterior probabilities provided in this section, can be used for constructing and calculating the test statistics analogues to those given in equations (4.2.3) and (4.3.3) in chapter 4.

Example 6.1.1. Suppose we have a random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{150}$ from a bivariate Student's t-distribution with 4 degree of freedom, i.e., $F = t_{(4)}(\mathbf{0}, \boldsymbol{I}_2)$ with density

$$f(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\Gamma\left(\frac{\nu}{2}+1\right)}{\Gamma\left(\frac{\nu}{2}\right)\nu\pi \mid \boldsymbol{\Sigma} \mid^{\frac{1}{2}}} \left(1 + \frac{1}{\nu}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)^{\left(\frac{\nu}{2}+1\right)} \tag{6.1.1}$$

with $\nu = 4$, $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_2$. Note that when $\nu = 1$, it is equivalent to the bivariate Cauchy distribution denoted by $C_2(\mathbf{0}, \boldsymbol{I}_2)$. Figure 6.1 shows the distribution function of $F = t_{(4)}(\mathbf{0}, \boldsymbol{I}_2)$ distribution. We place a Dirichlet invariant process prior $DIP(\alpha H)$, where H is chosen to be a bivariate spherically symmetric distribution. Then, we generate the sample posterior distributions from the posterior process. We perform the simulations with various base distributions H and for different values of the concentration parameters α as illustrated in Table 6.1. We partition the interval $[0, 2\pi]$ into $k = 10$ equally spaced intervals. For the purpose of comparing our approach with Lo's approach, we generated the sample posterior distributions P_m^* and $P_{k,m}^*$ drawn from $DP(\alpha H)$ and $DIP(\alpha H)$, respectively. Figures 6.2 - 6.7 show the sample posterior distributions P_m^* and $P_{k,m}^*$. Moreover, the results related to these simulations are reported in Tables 6.2 - 6.8 which show the probabilities obtained by P_m^*, $P_{k,m}^*$, F and F_m, where F_m is the empirical distribution of data. As it is seen in Figures 6.2 - 6.5, the posterior distributions $P_{k,m}^*$ present a better estimation for the distribution of data than distributions P_m^*. In addition, the probabilities calculated by $P_{k,m}^*$ and P_m^* are summarized in Tables 6.2 - 6.8. As it is shown in Tables 6.2 - 6.8, the probabilities computed by $P^*_{k,m}$ are closer to the real values than the ones computed by P^*_m, (see also Chapter 5).

α	$2, 100, 200$
H	$N_2(\mathbf{0}, \boldsymbol{I}_2), C_2(\mathbf{0}, \boldsymbol{I}_2)$

Table 6.1: Parameter setting for the simulation studies.

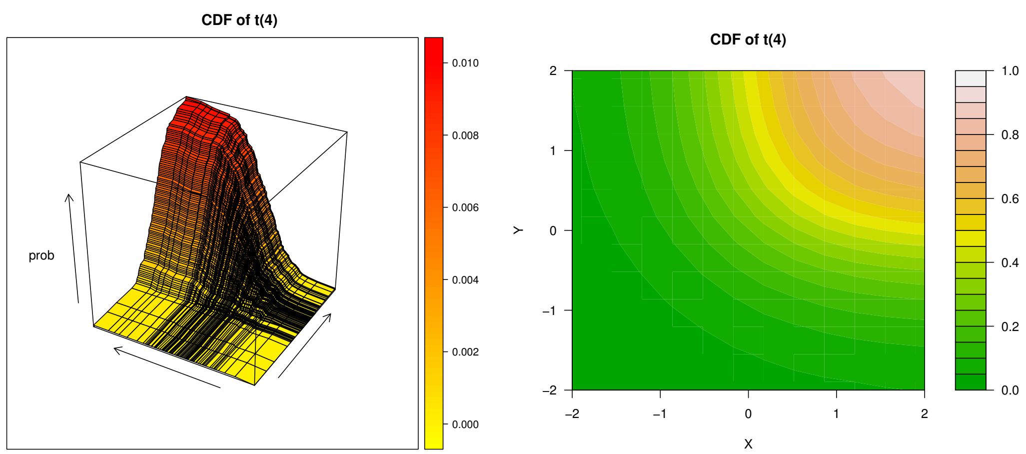

Figure 6.1: The distribution function of the bivariate Student's t-distribution with 4 degrees of freedom, i.e., $F = t_{(4)}(\mathbf{0}, \boldsymbol{I}_2)$.

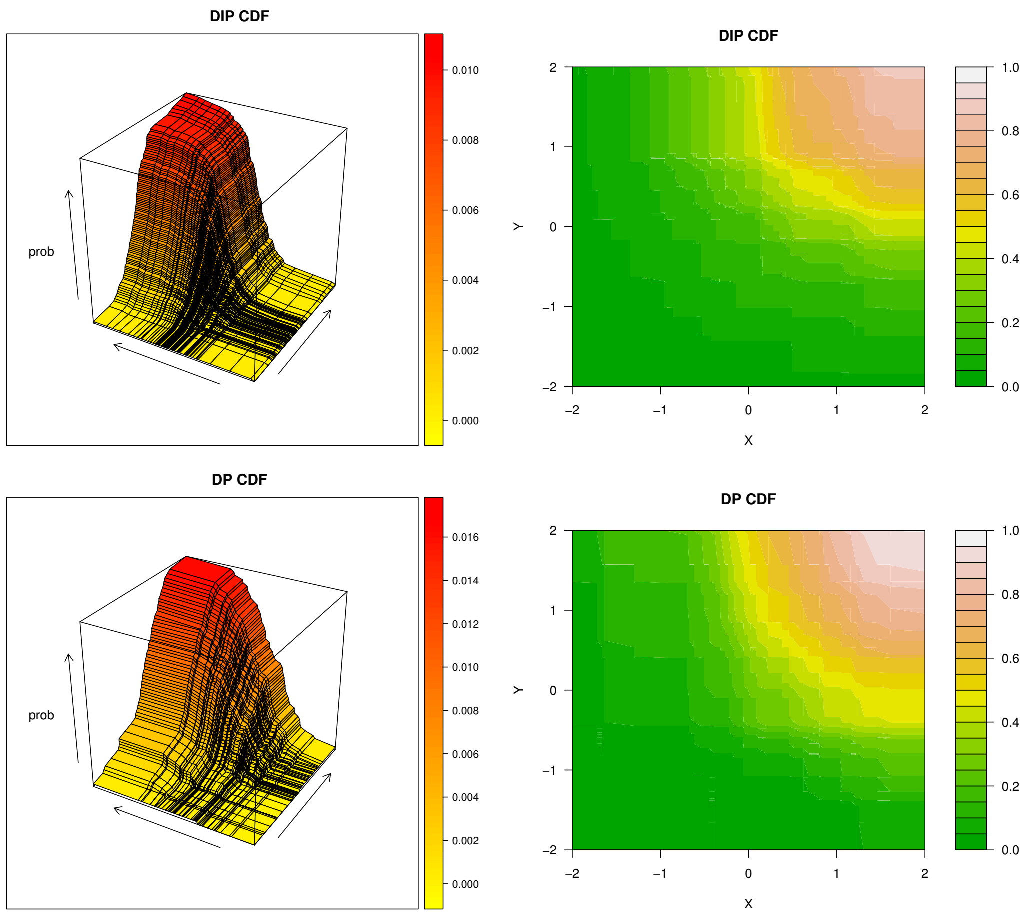

Figure 6.2: The distribution functions of the posterior distributions P_m^* and $P_{k,m}^*$ with parameters $\alpha = 2$ and $H = N_2(\mathbf{0}, \boldsymbol{I}_2)$.

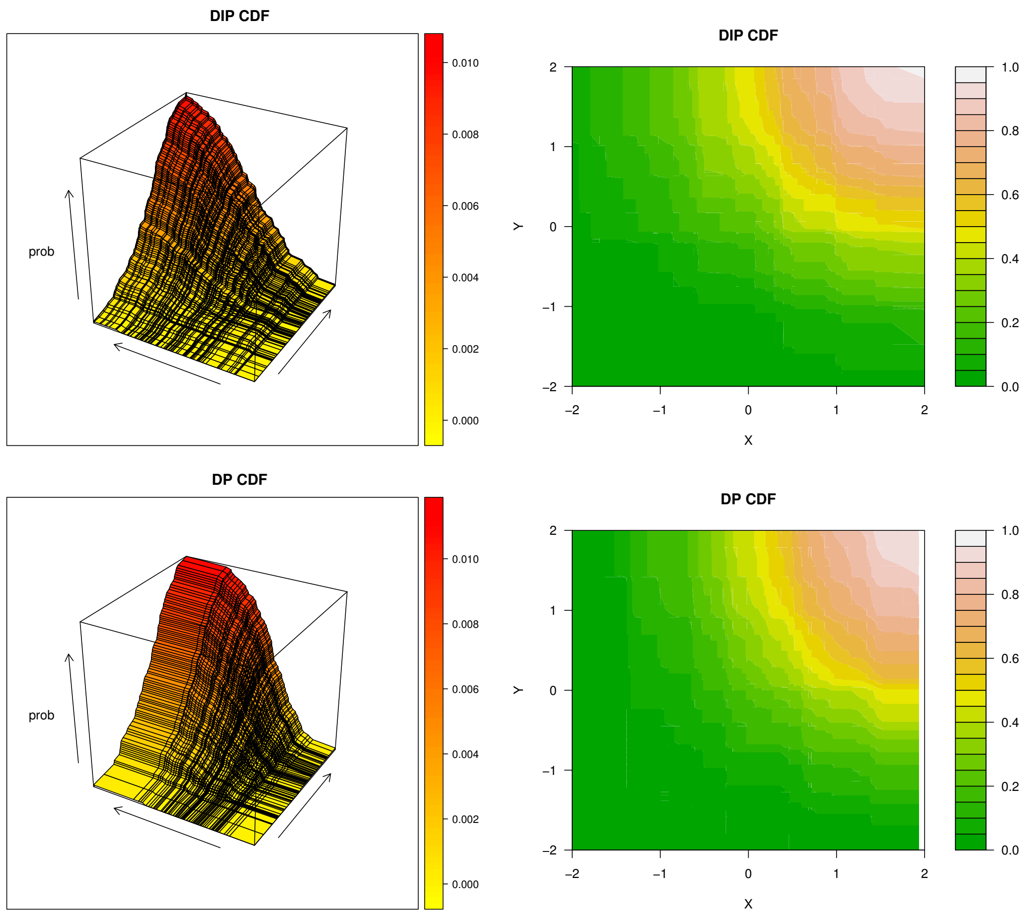

Figure 6.3: The distribution functions of the posterior distributions P_m^* and $P_{k,m}^*$ with parameters $\alpha = 100$ and $H = N_2(\mathbf{0}, \boldsymbol{I}_2)$.

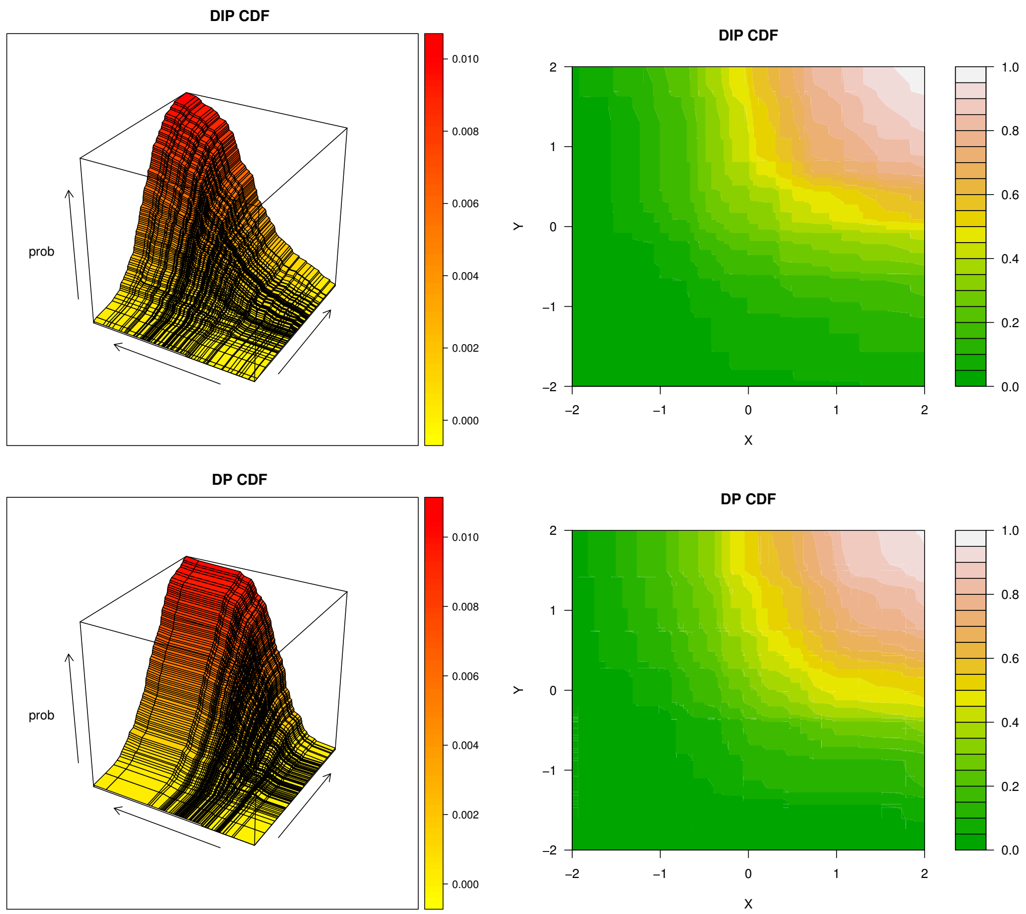

Figure 6.4: The distribution functions of the posterior distributions P^*_m and $P^*_{k,m}$ with parameters $\alpha = 200$ and $H = N_2(\mathbf{0}, \boldsymbol{I}_2)$.

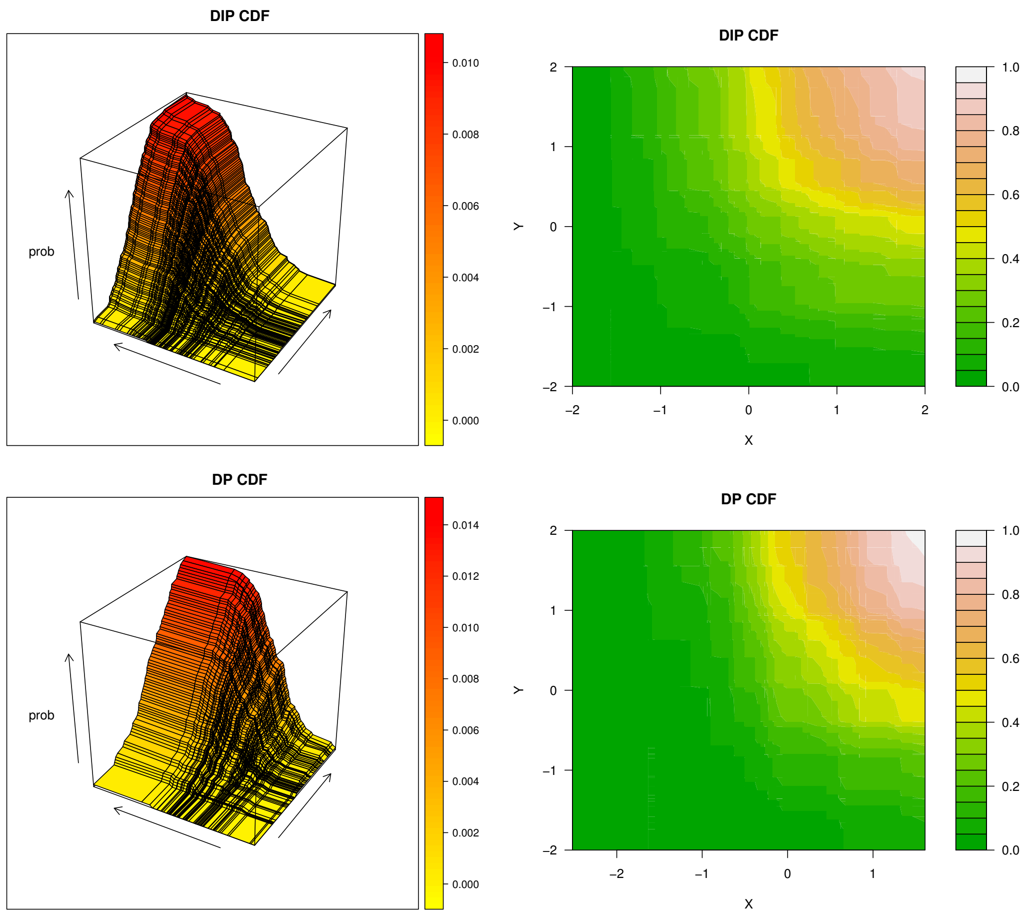

Figure 6.5: The distribution functions of the posterior distributions P_m^* and $P_{k,m}^*$ with parameters $\alpha = 2$ and $H = C_2(\mathbf{0}, \boldsymbol{I}_2)$.

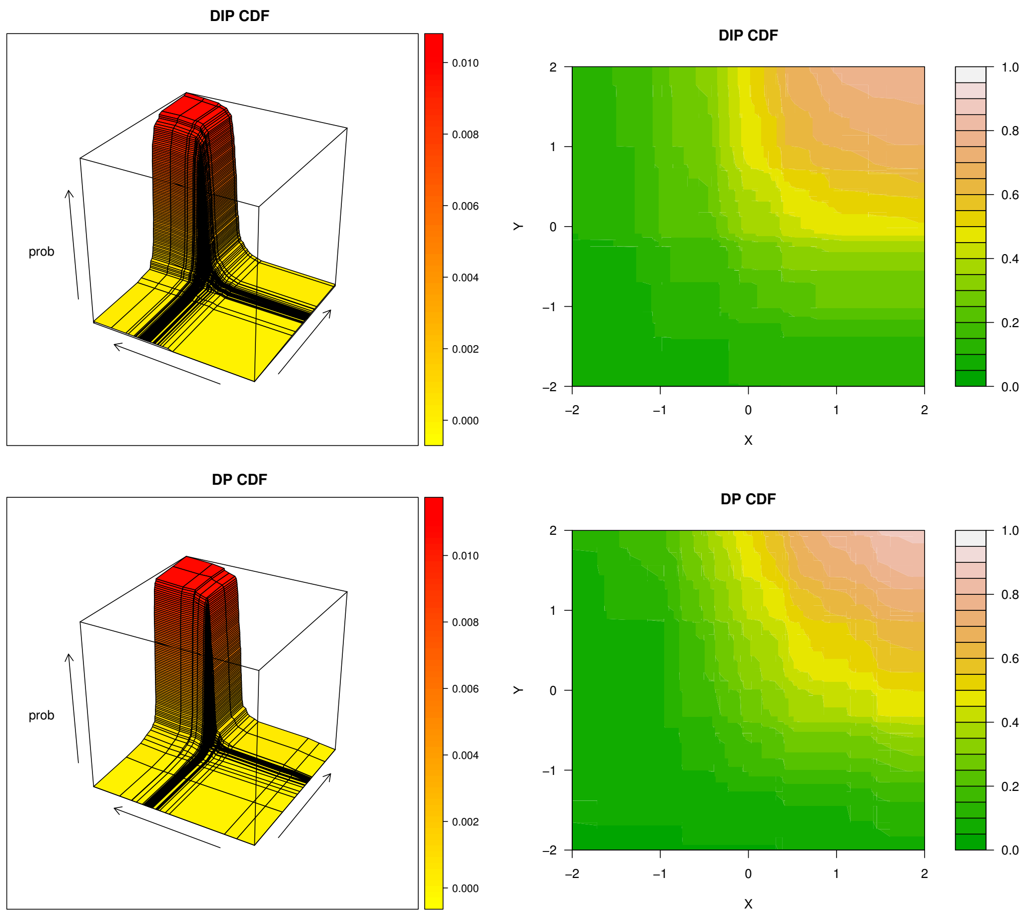

Figure 6.6: The distribution functions of the posterior distributions P_m^* and $P_{k,m}^*$ with parameters $\alpha = 100$ and $H = C_2(\mathbf{0}, \boldsymbol{I}_2)$.

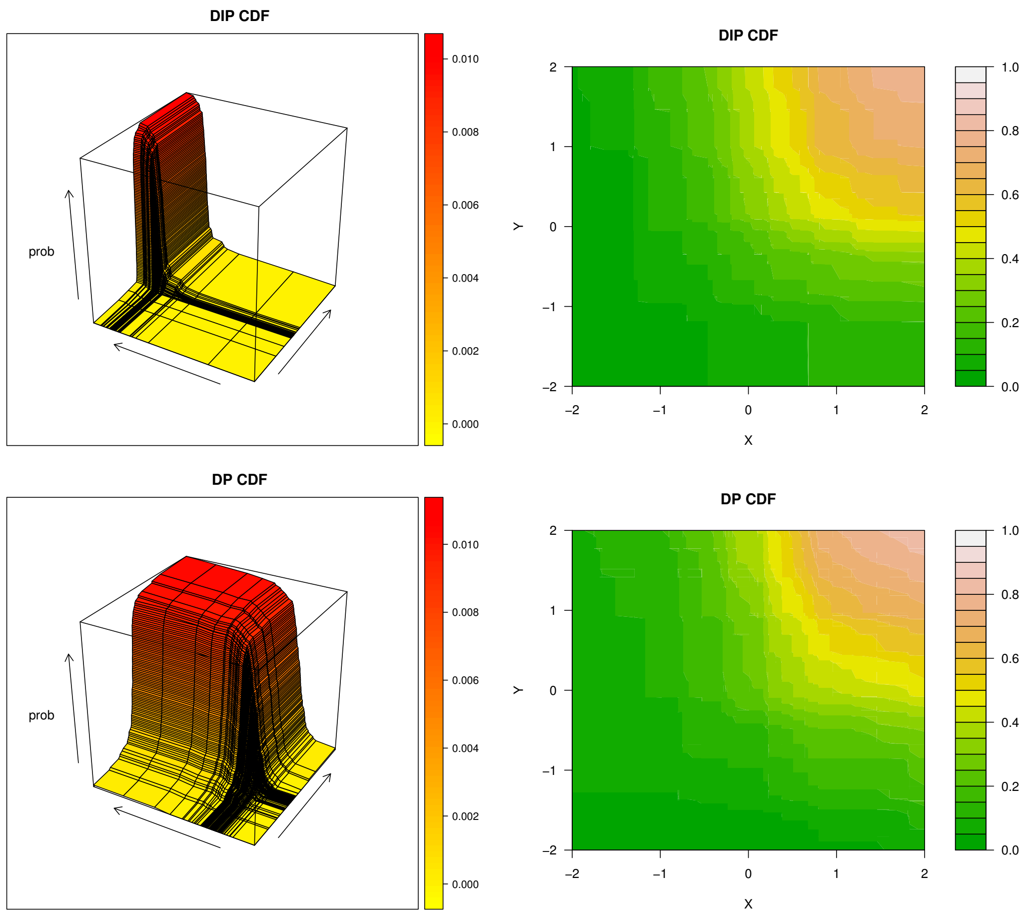

Figure 6.7: The distribution functions of the posterior distributions P_m^* and $P_{k,m}^*$ with parameters $\alpha = 200$ and $H = C_2(\mathbf{0}, \boldsymbol{I}_2)$.

		X_2			
	$F(A_i \times B_i)$ $F_n(A_i \times B_i)$	$B_1 = (-\infty, -1]$	$B_2 = (-1, 0]$	$B_3 = (0, 1]$	$B_4 = (1, \infty)$
X_1	$A_1 = (-\infty, -1]$	.042 .03	.051 .07	.051 .02	.042 .02
	$A_2 = (-1, 0]$	.051 .09	.105 .13	.105 .05	.051 .05
	$A_3 = (0, 1]$	.051 .04	.105 .15	.105 .1	.051 .07
	$A_4 = (1, 2)$	.026 .06	.038 .01	.038 .03	.026 .02
	$A_5 = (2, \infty)$	.016 .01	.013 .01	.013 .03	.016 .01

Table 6.2: Probabilities computed by F and F_m.

		X_2			
	$P^*_{m,k}(A_i \times B_i)$ $P^*_m(A_i \times B_i)$	$B_1 = (-\infty, -1]$	$B_2 = (-1, 0]$	$B_3 = (0, 1]$	$B_4 = (1, \infty)$
X_1	$A_1 = (-\infty, -1]$	.069 .019	.031 .062	.093 .042	.058 .00
	$A_2 = (-1, 0]$	.021 .105	.103 .153	.129 .34	.055 .034
	$A_3 = (0, 1]$	.017 .003	.141 .143	.081 .132	.071 .080
	$A_4 = (1, 2)$	.037 .044	.033 .011	.017 .070	.037 .033
	$A_5 = (2, \infty)$	.00 .00	.007 .010	.00 .025	.00 .00

Table 6.3: Probabilities computed by P^*_m and $P^*_{k,m}$ when $\alpha = 2$ and $H = N_2(\mathbf{0}, \boldsymbol{I}_2)$.

		X_2			
	$P^*_{m,k}(A_i \times B_i)$ $P^*_m(A_i \times B_i)$	$B_1 = (-\infty, -1]$	$B_2 = (-1, 0]$	$B_3 = (0, 1]$	$B_4 = (1, \infty)$
X_1	$A_1 = (-\infty, -1]$	.006 .008	.078 .142	.018 .102	.005 .078
	$A_2 = (-1, 0]$	.008 .142	.142 .159	.102 .107	.078 .033
	$A_3 = (0, 1]$	.098 .059	.113 .155	.101 .045	.065 .048
	$A_4 = (1, 2)$	.084 .005	.014 .020	.025 .048	.037 .008
	$A_5 = (2, \infty)$	.027 .007	.00 .00	.00 .020	.00 .00

Table 6.4: Probabilities computed by P^*_m and $P^*_{k,m}$ when $\alpha = 100$ and $H = N_2(\mathbf{0}, \boldsymbol{I}_2)$.

		X_2			
	$P^*_{m,k}(A_i \times B_i)$ $P^*_m(A_i \times B_i)$	$B_1 = (-\infty, -1]$	$B_2 = (-1, 0]$	$B_3 = (0, 1]$	$B_4 = (1, \infty)$
X_1	$A_1 = (-\infty, -1]$	.014 .044	.142 .050	.058 .028	.007 .020
	$A_2 = (-1, 0]$	.079 .090	.145 .119	.059 .138	.049 .071
	$A_3 = (0, 1]$	.067 .018	.047 .087	.131 .108	.056 .069
	$A_4 = (1, 2)$	.071 .008	.021 .006	.028 .106	.024 .025
	$A_5 = (2, \infty)$	.00 .006	.005 .00	.00 .005	.00 .00

Table 6.5: Probabilities computed by P^*_m and $P^*_{k,m}$ when $\alpha = 200$ and $H = N_2(\mathbf{0}, \boldsymbol{I}_2)$.

		X_2			
	$P^*_{m,k}(A_i \times B_i)$ $P^*_m(A_i \times B_i)$	$B_1 = (-\infty, -1]$	$B_2 = (-1, 0]$	$B_3 = (0, 1]$	$B_4 = (1, \infty)$
X_1	$A_1 = (-\infty, -1]$	.033 .007	.071 .036	.041 .053	.016 .064
	$A_2 = (-1, 0]$	.074 .071	.101 .146	.108 .144	.013 .026
	$A_3 = (0, 1]$	.071 .100	.062 .082	.142 .123	.023 .030
	$A_4 = (1, 2)$	.016 .056	.058 .026	.076 .010	.009 .025
	$A_5 = (2, \infty)$	.015 .000	.000 .000	.013 .000	.059 .000

Table 6.6: Probabilities computed by P^*_m and $P^*_{k,m}$ when $\alpha = 2$ and $H = C_2(\mathbf{0}, \boldsymbol{I}_2)$.

		X_2			
	$P^*_{m,k}(A_i \times B_i)$ $P^*_m(A_i \times B_i)$	$B_1 = (-\infty, -1]$	$B_2 = (-1, 0]$	$B_3 = (0, 1]$	$B_4 = (1, \infty)$
X_1	$A_1 = (-\infty, -1]$	.134 .060	.055 .016	.014 .027	.020 .071
	$A_2 = (-1, 0]$	.087 .073	.128 .095	.064 .147	.042 .103
	$A_3 = (0, 1]$	.014 .036	.110 .052	.043 .062	.079 .095
	$A_4 = (1, 2)$	.000 .037	.010 .025	.032 .020	.030 .006
	$A_5 = (2, \infty)$	.053 .028	.014 .000	.018 .000	.055 .048

Table 6.7: Probabilities computed by P^*_m and $P^*_{k,m}$ when $\alpha = 100$ and $H = C_2(\mathbf{0}, \boldsymbol{I}_2)$.

	$P^*_{m,k}(A_i \times B_i)$ / $P^*_m(A_i \times B_i)$	X_2: $B_1 = (-\infty, -1]$	$B_2 = (-1, 0]$	$B_3 = (0, 1]$	$B_4 = (1, \infty)$
X_1	$A_1 = (-\infty, -1]$	.059 .041	.079 .033	.052 .021	.040 .080
	$A_2 = (-1, 0]$	.019 .050	.083 .085	.098 .072	.061 .015
	$A_3 = (0, 1]$	.016 .078	.135 .122	.075 .150	.034 .108
	$A_4 = (1, 2)$	.007 .029	.000 .010	.033 .008	.050 .025
	$A_5 = (2, \infty)$	.087 .024	.018 .003	.004 .003	.050 .043

Table 6.8: Probabilities computed by P^*_m and $P^*_{k,m}$ when $\alpha = 200$ and $H = C_2(\mathbf{0}, \boldsymbol{I}_2)$.

6.2 Dirichlet invariant process for symmetric distributions with respect to the lines $x = a$ and $y = b$

In this section, we provide a Bayesian nonparametric approach to obtain the posterior distributions for distributions that are symmetric with respect to both lines $x = a$ and $y = b$. We also explain an algorithm that describes the steps of generating the paths from the Dirichlet invariant process posterior for these distributions. Consider a bivariate distribution which is symmetric with respect to the lines $x = a$ and $y = b$. In this case, the data transformation group is defined as

$$\mathscr{G}(x, y) = \{(x, y), (2a - x, y), (x, 2b - y), (2a - x, 2b - y)\}, \tag{6.2.1}$$

where the center of symmetry is $\mu = (a, b)$. Then, the Dirichlet invariant process posterior is obtained with the base distribution

$$H^*_{k,m} = \frac{\alpha H + \frac{1}{4}\sum_{i=1}^{m}\left(\delta_{(X_i,Y_i)} + \delta_{(2a-X_i,Y_i)} + \delta_{(X_i,2b-Y_i)} + \delta_{(2a-X_i,2b-Y_i)}\right)}{\alpha + m}.$$

The simulation algorithm below shows the steps of generating posterior paths from Dirichlet invariant process:

Algorithm 3 Generating a posterior distribution from Dirichlet invariant process posterior for a distribution that is symmetric with respect to both lines $x = a$ and $y = b$.

1. Consider the i.i.d. sample points $\boldsymbol{X}_i = (X_i, Y_i)$, $i = 1, \ldots, m$ from a bivariate distribution that is symmetric with respect to the lines $x = a$ and $y = b$.

2. For each $\boldsymbol{X}_i = (X_i, Y_i)$, $i = 1, \ldots, m$, compute $g_1(\boldsymbol{X}_i) = (X_i, Y_i), g_2(\boldsymbol{X}_i) = (2a - X_i, Y_i), g_3(\boldsymbol{X}_i) = (X_i, 2b - Y_i)$ and $g_4(\boldsymbol{X}_i) = (2a - X_i, 2b - Y_i)$.

3. Compute the empirical distribution of $g_j(\boldsymbol{X}_i)$'s, $j = 1, \ldots, 4$, for each $i = 1, \ldots, m$ by assigning the equal weight $\frac{1}{4m}$ to each point $g_j(\boldsymbol{X}_i)$.

4. Generate n observations $\boldsymbol{\theta}^*_1, \ldots, \boldsymbol{\theta}^*_n$ from $H^*_{k,m}$ for a large enough n, where $\boldsymbol{\theta}^*_l = (\theta^*_{1l}, \theta^*_{2l}), l = 1, \ldots, n$ and

$$H^*_{k,m} = \frac{\alpha H + \frac{1}{4}\sum_{i=1}^{m}\left(\delta_{(X_i,Y_i)} + \delta_{(2a-X_i,Y_i)} + \delta_{(X_i,2b-Y_i)} + \delta_{(2a-X_i,2b-Y_i)}\right)}{\alpha + m}.$$

5. Generate a sample posterior distribution $P^*_{k,m}$ from Dirichlet invariant process posterior with parameters $\alpha^*_m = \alpha + m$ and $H^*_{k,m}$ using Algorithm 1 as given in Chapter 2 by replacing the parameters α and H with α^*_m and $H^*_{k,m}$, respectively.

Note that the special cases of this symmetry can be considered with respect to one of the lines $x = a$ or $y = b$. Then, the transformation group in (6.2.1) is reduced to $\mathscr{G}_1(x, y) = \{(x, y), (x, 2b - y)\}$ and $\mathscr{G}_2(x, y) = \{(x, y), (2a - x, y))\}$, respectively.

6.3 Bayesian nonparametric bootstrap for area and perimeter of the rectangular support

We use the approach discussed in Section 6.2 to provide an approximate for the distribution of the area (A) and perimeter (S) of a rectangle which is centered at any point $\mu = (a, b)$, where $(a, b \in \mathbb{R})$ and based on the given points drawn from the symmetric distribution F with a rectangular support of K. In other words, we apply our approach in Chapter 5 to introduce a Bayesian nonparametric resampling technique for bootstrapping the specific functionals of K. Then, the bootstrapped data is used to estimate the quantity of interest which is the area and perimeter of K, for example. Specifically, let $\boldsymbol{X}_i = (X_i, Y_i)$, $1 \leq i \leq m$ be a sequence of i.i.d. random points distributed with distribution F on a rectangle $K = [2a - c, c] \times [2b - d, d]$, where F is symmetric with respect to both lines $x = a$ and $y = b$. Therefore, the rectangle is centered at the known point $\mu = (a, b)$, where $(a, b \in \mathbb{R})$ and c and d are two unknown parameters. Notice that

$$K_m : = \left[\min_{1 \leq i \leq m} X_i, \max_{1 \leq i \leq m} X_i \right] \times \left[\min_{1 \leq i \leq m} Y_i, \max_{1 \leq i \leq m} Y_i \right]$$

is the smallest rectangle which circumscribes all the sample points, and all the edges are parallel to the coordinate axes. It is easy to show that K_m is the maximum likelihood estimator for K. Obviously,

$$\min_{1 \leq i \leq m} X_i \xrightarrow{a.s.} 2a - c, \ \max_{1 \leq i \leq m} X_i \xrightarrow{a.s.} c, \ \min_{1 \leq i \leq m} Y_i \xrightarrow{a.s.} 2b - d, \ \max_{1 \leq i \leq m} Y_i \xrightarrow{a.s.} d.$$

Define

$$\Delta_{m,1}:=\max_{1\leq i\leq m} X_i-\min_{1\leq i\leq m} X_i,\ \Delta_{m,2}:=\max_{1\leq i\leq m} Y_i-\min_{1\leq i\leq m} Y_i.$$

It can be shown that

$$\Delta_{m,1}\xrightarrow{a.s.}2(c-a),\ \Delta_{m,2}\xrightarrow{a.s.}2(d-b).$$

Then, the estimators for area and perimeter of the rectangle can be obtained by

$$A_m=\Delta_{m,1}\times\Delta_{m,2} \tag{6.3.1}$$

and

$$S_m=2\left(\Delta_{m,1}+\Delta_{m,2}\right), \tag{6.3.2}$$

respectively. Let $\boldsymbol{X}_i=(X_i,Y_i)$, $i=1,\ldots,m$ be the i.i.d. random points distributed on a rectangle. In regular bootstrapping, in order to bootstrap any functional of K denoted by $\phi(K)$, the resampling procedure proceeds by sampling from the sample points. In Bayesian approach, by Rubin's bootstrapping, a Dirichlet process prior $DP(\alpha H)$ is considered as the prior over the distribution of sample data. Then, the posterior distribution obtained from the Dirichlet process posterior is used to bootstrap the functional of interest. Here, we apply our approach to bootstrap $\phi(K)$ by Dirichlet invariant process $DIP(\alpha H)$. Specifically, we consider a Dirichlet invariant process prior $DIP(\alpha H)$ on distribution of data. Then, we approximate the distribution of the functionals $A=\phi_1(K)$ and $S=\phi_2(K)$, by using the posterior distribution extracted from the Dirichlet invariant process posterior $DIP(\alpha_m^* H_{k,m}^*)$. Then, the bootstrapped data is used to compute the estimators $A_m=\phi_1(K_m)$ and $S_m=\phi_2(K_m)$. In the following, the procedures described above is applied to a data set in terms of an example with simulation studies.

Example 6.3.1. Let $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_m$ be the data that shows the locations that an animal has seen in its area of habitat within the total study area which is a rectangle $K = [0, c] \times [0, d]$, where c and d are unknown. The aim is to calculate the area A and the perimeter S of habitat which is obtained by $S = 2(c + d)$ and $A = c \times d$. Assume the data is distributed with symmetric distribution F over K. We place a Dirichlet invariant process prior $DIP(\alpha H)$ on F, where F is symmetric with respect to both lines $x = c/2$ and $y = d/2$. Then, we extract the posterior process by considering different values of the parameters α and H. Algorithm 4 indicates the steps of the simulation.

Algorithm 4 Bootstrapping the area and perimeter for the rectangular support of the symmetric distribution by Dirichlet invariant process posterior.

1. Consider the i.i.d. sample points $\boldsymbol{X}_i = (X_i, Y_i)$, $i = 1, \ldots, m$ distributed with distribution F on the rectangle $K = [0, c] \times [0, d]$ centered at the point $(c/2, d/2)$, where F is symmetric with respect to both lines $x = c/2$ and $y = d/2$.

2. For each $\boldsymbol{X}_i = (X_i, Y_i)$, $i = 1, \ldots, m$, compute $g_1(\boldsymbol{X}_i) = (X_i, Y_i), g_2(\boldsymbol{X}_i) = (c - X_i, Y_i), g_3(\boldsymbol{X}_i) = (X_i, d - Y_i)$ and $g_4(\boldsymbol{X}_i) = (c - X_i, d - Y_i)$.

3. Compute the empirical distribution of $g_j(\boldsymbol{X}_i)$'s, $j = 1, \ldots, 4$ for each $i = 1, \ldots, m$ by assigning the equal weight $\frac{1}{4m}$ to each point $g_j(\boldsymbol{X}_i)$.

4. Generate n observations $\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_n^*$ from $H_{k,m}^*$, where $\boldsymbol{\theta}_l^* = (\theta_{1l}^*, \theta_{2l}^*), l = 1, \ldots, n$ and

$$H_{k,m}^* = \frac{\alpha H + \frac{1}{4}\sum_{i=1}^{m}\left(\delta_{(X_i,Y_i)} + \delta_{(c-X_i,Y_i)} + \delta_{(X_i,d-Y_i)} + \delta_{(c-X_i,d-Y_i)}\right)}{\alpha + m},$$

   for large enough n.

5. Generate a sample posterior distribution $P_{k,m}^*$ from Dirichlet invariant process posterior with parameters $\alpha_m^* = \alpha + m$ and $H_{k,m}^*$ using Algorithm 1 as given in Chapter 2 by replacing the parameters α and H with α_m^* and $H_{k,m}^*$, respectively.

6. Compute A_m and S_m by

$$A_m = \Delta_{n,1} \times \Delta_{n,2} \quad \text{and} \quad S_m = 2\left(\Delta_{n,1} + \Delta_{n,2}\right),$$

   where

$$\Delta_{n,1} = \max_{1 \le l \le n} \theta_{1l}^* - \min_{1 \le l \le n} \theta_{1l}^*, \ \Delta_{n,2} = \max_{1 \le l \le n} \theta_{2l}^* - \min_{1 \le l \le n} \theta_{2l}^*.$$

7. Repeat Steps 4 to 6, N times to obtain values $A_{m,1}, \ldots, A_{m,N}$ and $S_{m,1}, \ldots, S_{m,N}$.
8. Approximate the distributions of A and S by empirical distributions of $A_{m,1}, \ldots, A_{m,N}$ and $S_{m,1}, \ldots, S_{m,N}$, respectively.
9. Estimate A and S by $\widehat{A} = \sum_{l=1}^{N} A_{m,l}/N$ and $\widehat{S} = \sum_{l=1}^{N} S_{m,l}/N$, respectively.

Suppose we have a random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ from a truncated bivariate Student's t-distribution with 2 degree of freedom, i.e., where $F = t_{(2)}(\mathbf{0}, \boldsymbol{I}_2)$ on the rectangle $[0,7] \times [0,5]$. Table 6.9 shows the number of data points and the parameter values that are used for the simulation studies. For comparison purposes, we also report the estimations obtained by means of Dirichlet process $DP(\alpha H)$. Figures 6.8 and 6.9 illustrate the data points (black pluses), the points bootstrapped by $DIP(\alpha H)$ (blue triangles) and $DP(\alpha H)$ (green triangles) along with corresponding approximated rectangular supports for two sample sizes of m. The analytical results are reported in Tables 6.10 and 6.11. Tables 6.10 - 6.11 show the true values of S and A and the estimates $\widehat{S}$ and $\widehat{A}$ obtained by the empirical distribution of data points F_m, the sample distributions $P^*_{k,m}$ from Dirichlet invariant process posterior and the sample distributions P^*_m from Dirichlet process posterior. We performed the simulations for two sizes of data points (sample sizes) $m = 10$ and $m = 50$. The procedure was then iterated $N = 1000$ times. The histogram of the estimators $\widehat{S}$ and $\widehat{A}$ are demonstrated in Figure 6.10. As it is seen in Table 6.10, when $H = U([0,7] \times [0,5])$, for all values of the parameter α, the estimated values of S and A obtained by P^*_m and $P^*_{k,m}$ are closer to the real values compared to those computed by the empirical distribution F_m. Moreover, $P^*_{k,m}$ provides more accurate estimates than P^*_m. Here, for the first choice of the base distribution H, a uniform distribution $U([0,7] \times [0,5])$ is considered. Then, as seen in Tables 6.10 and 6.11, as the concentration parameter α increases, the parameter H^* of the posterior distribution approaches the base distribution H which is an appropriate guess for the support of the true distribution of data. Therefore, we obtain a better estimation as α increases.

By considering the second choice of the parameter H as a uniform $U([0,8]\times[0,4])$ for the prior distribution which is further from the true distribution, we discover a different trend. In this case, as seen in Tables 6.10 and 6.11, for $\alpha = 2$, we still have a reasonable estimation, since the more weight in the posterior base distributions H^*_m and $H^*_{k,m}$ is given to the empirical distribution of data. On the other hand, as α increases, the estimations diverge from the true values.

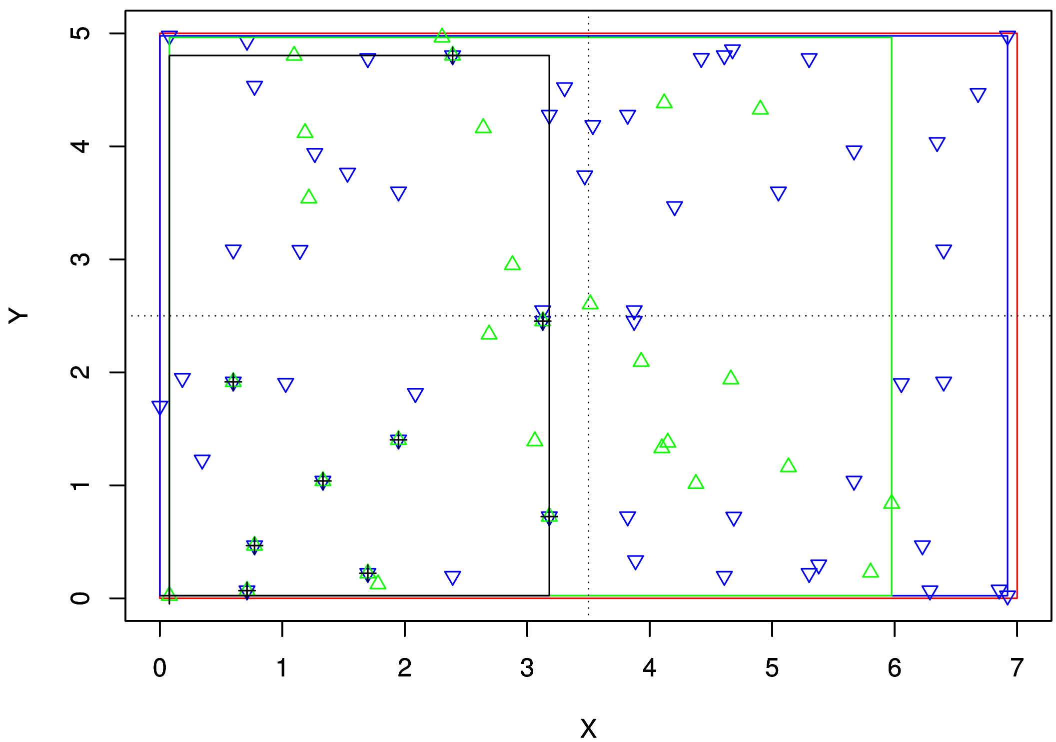

Figure 6.8: The data points (black plus spots) with bootstrapped rectangle obtained by F_m (the interior black rectangle), the points from P^*_m (green triangles) with corresponding bootstrapped rectangle (green), the points from $P^*_{k,m}$ (blue triangles) with corresponding bootstrapped rectangle (blue) and the true rectangle (red) where $\alpha = 2$, $H = U([0, 7] \times [0, 5])$ and $m = 10$.

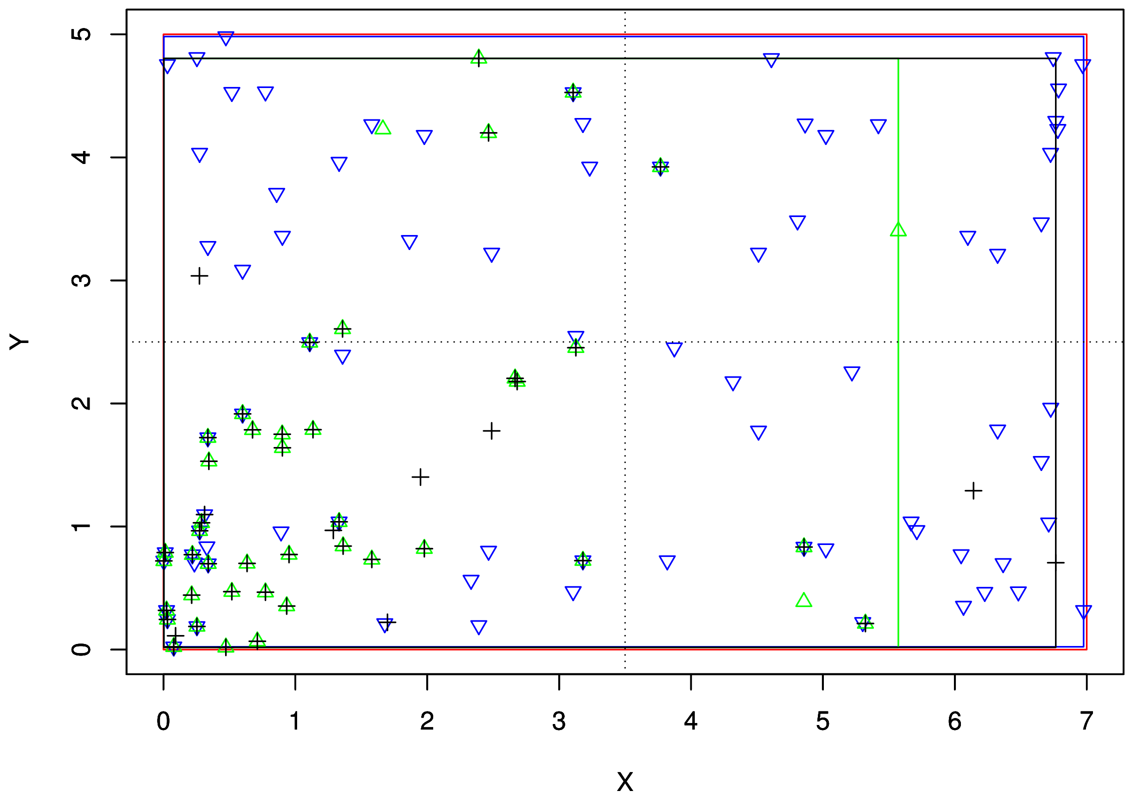

Figure 6.9: The data points (black plus spots) with bootstrapped rectangle obtained by F_m (the interior black rectangle), the points from P_m^* (green triangles) with corresponding bootstrapped rectangle (green), the points from $P_{k,m}^*$ (blue triangles) with corresponding bootstrapped rectangle (blue) and the true rectangle (red) where $\alpha = 2$, $H = U([0, 7] \times [0, 5])$ and $m = 50$.

m	$10, 50$
α	$2, 10, 50$
H	$U([0, 7] \times [0, 5])$, $U([0, 8] \times [0, 4])$

Table 6.9: The parameter setting used in example 6.3.1.

H	α	F	F_m	$P^*_{k,m}$	P^*_m
$U([0,7]\times[0,5])$	2	$S=24, A=35$	$14.8, 15.7$	$23.59, 33.89$	$22.64, 31.37$
	10			$23.64, 34.08$	$22.27, 33.24$
	50			$23.69, 34.14$	$23.53, 33.69$
$U([0,8]\times[0,4])$	2			$24.93, 35.50$	$24.41, 37.21$
	10			$25.19, 37.16$	$24.89, 37.46$
	50			$25.28, 37.36$	$25.44, 37.91$

Table 6.10: The perimeter ($\widehat{S}$) and area ($\widehat{A}$) computed by the true distribution F, the empirical distribution F_m of data points, a sample distribution $P^*_{k,m}$ from Dirichlet invariant process posterior and a sample distribution P^*_m from Dirichlet process posterior for different values of the parameters α and H when $m = 10$.

H	α	F	F_m	$P^*_{k,m}$	P^*_m
$U([0,7]\times[0,5])$	2	$S=24, A=35$	$23.09, 32.35$	$23.83, 34.46$	$22.82, 31.68$
	10			$23.83, 34.45$	$23.37, 34.39$
	50			$23.80, 34.37$	$23.78, 34.32$
$U([0,8]\times[0,4])$	2			$24.17, 35.29$	$23.43, 33.08$
	10			$24.93, 36.15$	$24.43, 36.29$
	50			$24.48, 38.44$	$24.90, 36.15$

Table 6.11: The perimeter ($\widehat{S}$) and area ($\widehat{A}$) computed by the true distribution F, the empirical distribution F_m of data points, a sample distribution $P^*_{k,m}$ from Dirichlet invariant process posterior and a sample distribution P^*_m from Dirichlet process posterior for different values of the parameters α and H when $m = 50$.

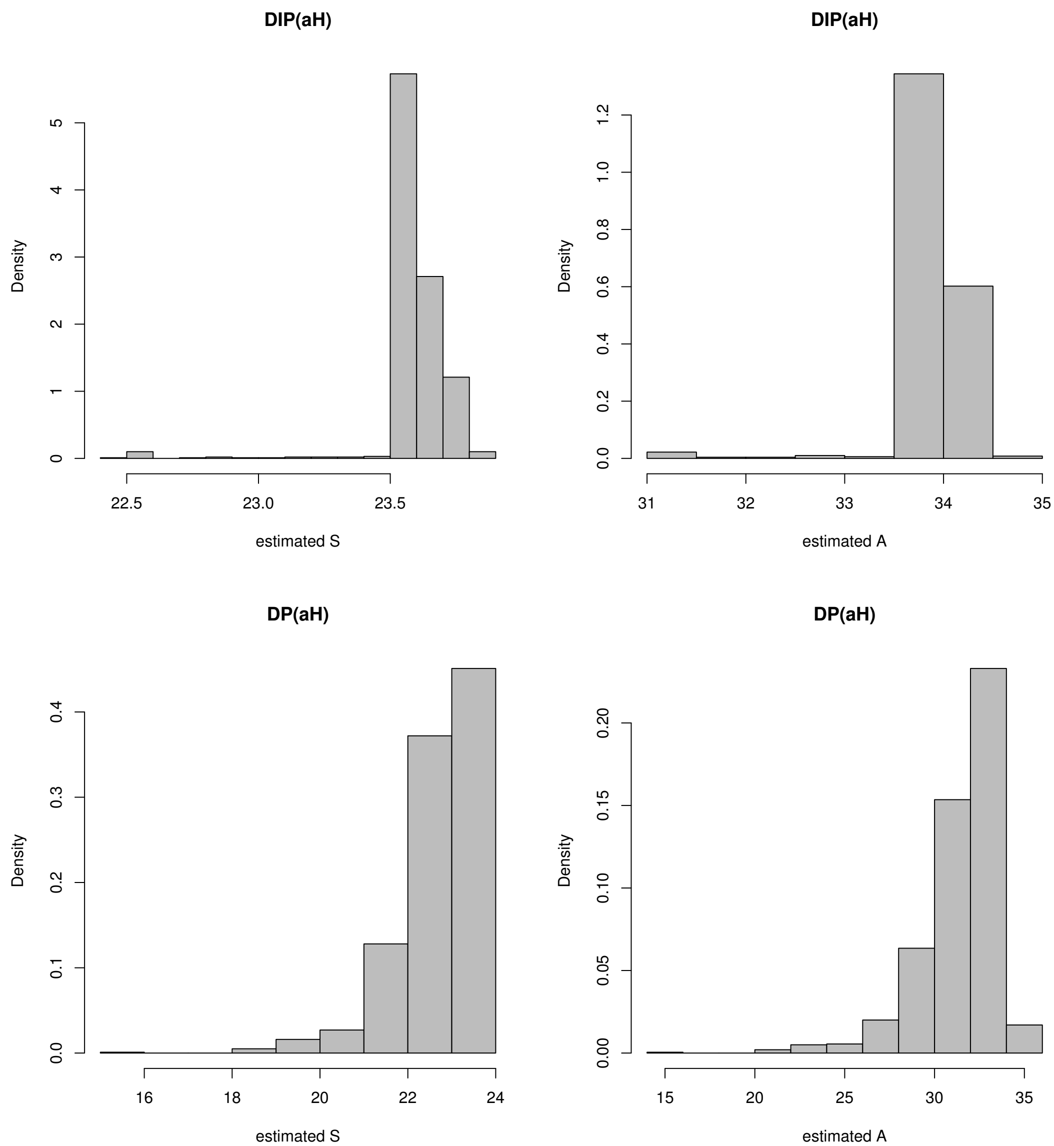

Figure 6.10: The histograms of the estimators $\widehat{S}$ (left) and $\widehat{A}$ (right) extracted by $DIP(\alpha H)$ (top) and $DP(\alpha H)$ (bottom) when $\alpha = 2$, $H = U([0,7] \times [0,5])$ and $m = 10$.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

This thesis was divided into two parts. In the first part, as discussed in Chapter 4, we proposed a Bayesian nonparametric chi-squared goodness of-fit-test based on the Kullback-Leibler distance between the Dirichlet process posterior and the hypothesized distribution. Our method proceeds by placing a Dirichlet process prior on the distribution of observed data and computing the probability of each bin of the partition from the Dirichlet process posterior. The suggested method is in contrast with the frequentist's Pearson's chi-squared goodness-of-fit test which is based on counting the observations in each bin of the partition. In fact, our suggested method is equivalent to the Bayesian bootstrap developed by Lo (1987) and our results confirm the asymptotic validity of Bayesian bootstrap for the chi-squared goodness-of-fit test similar to Lo (1987). As explained in Lo (1987) and as expected the Bayesian bootstrap (use of nonparametric Bayesian prior) asymptotically behaves similar to Efron's bootstrap. There are several advantages to our technique. For example, both the prior knowledge and the strength of the belief can be embedded in the process of decision making. Also, the large enough sample sizes rectify the poor prior assumptions and computationally easier to be carried out. We also extended our method to test of independence. Like the classical chi-squared test, we can generalize our goodness-of-fit test to several variables. For categorical observations with finite many variables, placing a Dirichlet distribution prior on the probabilities of categories and deriving the posterior Dirichlet distribution can establish similar tests. For example, the test of independence and conditional independence of qualitative observations follow easily. We commented before on the advantages of our method but a comparison between our method and the regular chi-square goodness-of-fit test is not relevant. A Bayesian procedure almost always can be made artificially superior to frequentist's approach by an artificial choice of prior near the actual distribution.

In the second part of this thesis, the goal was to draw Bayesian inference for restreicted populations. An important restricted population is when the distribution is spherically symmetric. This objective has been accomplished in Chapter 5. In order to draw Bayesian nonparametric inference for a spherically symmetric distributions, we placed a Dirichlet invariant process prior on the space of this class of distributions and we derived the corresponding posterior process. Indeed, our approach is an extension of Dirichlet invariant processes to a spherically symmetric distribution. Then, the knowledge and assumption about the invariant properties of the distribution of interest can be taken into account as a specific transformation group. We first obtained the Dirichlet invariant process posterior for a finite group of transformation. Then, we considered an infinite group of transformations. We proved that for an infinite group, the Dirichlet invariant process approaches the Dirichlet process.

We also extracted the Bayesian nonparametric posterior paths for distributions which are symmetric with respect to both lines $x = a$ and $y = b$. Then, we applied our approach for bootstrapping certain functionals of the support for these symmetric distributions. Our simulation studies showed that replacing the Dirichlet process by the Dirichlet invariant process in Bayesian bootstrapping can improve the estimation of the distribution of the functional of interest. Moreover, the proposed bootstrapping technique can be used for approximating the distribution of any functional of the support for spherically symmetric distributions discussed in Chapter 5.

7.2 Research extensions

The ideas introduced and discussed in this thesis, can be used for solving a variety of problems. Here are some suggestions:

1. Like the classical chi-squared test, we can generalize our goodness-of-fit test in chapter 4 to several variables. For categorical observations with finite many variables, placing a Dirichlet distribution prior on the probabilities of categories and deriving the posterior Dirichlet distribution can establish similar tests. For example, the test of independence, conditional independence of qualitative observations and log-linear models follow easily.

2. The approach developed in Chapter 5 can be applied to other forms of invariant transformation groups. Also, it can be used for estimating a density function, a unimodal density function, the mode of a distribution and performing goodness-of-fit tests. Specifically, based on the obtained posterior distribution for a spherically symmetric distribution, we can construct a chi-squared test statistic like the one introduced in Chapter 4 for testing spherical symmetry.

3. Additionally, our proposed Bayesian nonparametric bootstrapping technique explained in Chapter 6 can be applied to other symmetric distributions with symmetric support. For example, for bootstrapping the functionals of the circular supports by using the Dirichlet invariant process obtained for the spherically symmetric distributions in Chapter 5.

4. The posterior distribution we obtained for the symmetric distributions with respect to the y axis, i.e., when $y = b$, can be employed as the model for the

face recognition analysis and biometric identification.

Appendix A

More Concepts

In this appendix, we collect some definitions which are used throughout this thesis.

A.1 Group

Definition A.1.1. *(Group) A set G together with an operation $\bullet$ (called the group law of G) that combines any two elements a and b into another element, denoted by $a \bullet b$, is called a group G with operation $\bullet$ $(G, \bullet)$, if it satisfies four requirements below known as the group axioms:*

1. ***Closure:*** *For all a, b in G, $a \bullet b$ is also in G.*
2. ***Associativity:*** *For all a, b and c in G, $(a \bullet b) \bullet c = a \bullet (b \bullet c)$.*
3. ***Identity element**: There exists a unique element e in G such that for every element a in G, the equation $e \bullet a = a \bullet e = a$ holds.*
4. ***Inverse element**: For each a in G, there exists an element b in G commonly denoted by a^{-1} (or $-a$, if the operation is " $+$ ") such that $a \bullet b = b \bullet a = e$, where e is the identity element.*

Appendix B

R Programs

In this section, we give the R codes used in Chapter 4 and Chapter 6.

B.1 Bayesian nonparametric chi-squared goodness-of-fit test

```
#############    expected prob. by H=NORMAL
x<- seq(-3,4, by=1)
H=numeric()
for (k in 2:(length(x))){
  if(k==2){H[k-1]<- pnorm(x[2],0,1)}
  else if (2< k & k<= length(x)-1 )
  {H[k-1]<- pnorm(x[k],0,1)-pnorm(x[k-1],0,1)}
  else if (k >= length(x)-1)
  {H[k-1]<- 1-pnorm(x[k-1],0,1)}
}
############### DP prior #####################
N=2000;  n=2000; alpha=0.1
```

```
aa=c(500)
prior=numeric();prior1=numeric()
a<- aa; d0=numeric(); d00=numeric()
for (j in 1:N){
  g=numeric(); g1=numeric(); cmp=numeric(); mp=numeric()
  t<- rnorm(n,0,1) #H sample  rt(n,5)
  y<- rexp(n+1, 1)
  for (i in 1:n ){
    g[i]<- sum(y[1:i])/sum(y)
    g1[i]<- qgamma(1-g[i], a/n, 1)
  }
  p1<- g1/sum(g1)
  dp<- data.frame(t,p1)
  sort.dp<-dp[order(dp$t),]#  sort theta's
  t.sort<- sort.dp$t
  p<- sort.dp$p1
  cp1<- cumsum(p)
  x1<- unique(t.sort)
  for (i in 1: length(x1)) {
    cmp[i]= max(cp1[t.sort ==x1[i]])
  }
  P=numeric(0)
  for (k in 2:(length(x))){
    if( k<= length(x)-1){
      if (sum((which(x1<= x[k])))==0) { P[k-1]=0}
      else if(k-1==1) {P[k-1]<- cmp[max(which( x1< x[k]))]}
      else  {P[k-1]<- cmp[max(which( x1< x[k]))]
      -sum(P[1:(k-2)])}
    }
```

```
    else if (k > length(x)-1){P[k-1]<- 1
    -cmp[max(which(x1< x[k-1]))]}
  }
  d0[j]<- a*sum(((P-H)^2)/H)
  d00[j]<- (a+m)*sum(((P-H)^2)/H)
}
length(d0[!is.na(d0)])
range(d0)
range(rchisq(N,k-2))
epsi=c(1,2,3,4,5,6)
for (i in 1:length(epsi)){
    prior[i]<- sum(d0< epsi[i])/length(d0)
    prior1[i]<- sum(d00< epsi[i])/length(d00)
  }
pchisq(epsi[4],k-2)
##################
GG<- function(t){(a/(a+m))*pnorm(x[t],0,1)
+sum(data <= x[t])/(a+m) }
Hstar=numeric()
for (k in 2:(length(x))){
  if(k==2){Hstar[k-1]<- GG(2) }
  else if (2< k & k<= length(x)-1 )
  {Hstar[k-1]<- GG(k)-GG(k-1)}
  else if (k >= length(x)-1)
  {Hstar[k-1]<- 1-GG(k-1)}   # expected prob. by H*
}
############## DP posterior  #####################
count=numeric(0)
posterior=numeric(); post1=numeric(); post2=numeric()
```

```
for (k in 1:length(aa)){
  a<- aa[k]; d1=numeric(); d2=numeric(); dstar=numeric()
for (j in 1:N){
  g=numeric(); tstar=numeric()
  g2=numeric(); cmpp=numeric(); mpp=numeric()
  y<- rexp(n+1, 1)
  for (i in 1:n ){
    g[i]<- sum(y[1:i])/sum(y)
    g2[i]<- qgamma(1-g[i], (a+m)/n, 1)
    u<- runif(1)
    if  (u< a/(a+m)){
      tstar[i]<- rnorm(1,0,1)  # H   rt(1,5)
    } else{
      tstar[i]<- sample(data,1,replace=T)
    }
  }
  p2<- g2/sum(g2)
  dpp<- data.frame(tstar,p2)
  sort.dpp<-dpp[order(dpp$tstar),]#  sort theta's
  t2.sort<- sort.dpp$tstar
  pp<- sort.dpp$p2
  cp2<- cumsum(pp)
  x2<- unique(t2.sort)
  for (i in 1: length(x2)) {
    cmpp[i]= max(cp2[t2.sort ==x2[i]])   #cdf
  }
  P=numeric()
  for (k in 2:(length(x))){
    if( k<= length(x)-1){
```

```
      if (sum((which(x2<= x[k])))==0) { P[k-1]=0}
      else if(k-1==1) {P[k-1]<- cmpp[max(which( x2< x[k]))]}
      else  {P[k-1]<- cmpp[max(which( x2< x[k]))]
      -sum(P[1:(k-2)])}
    }
    else if (k > length(x)-1){P[k-1]<- 1
    -cmpp[max(which(x2< x[k-1]))]}
  }
dstar[j]<- m*sum(((P-H)^2)/H)
d1[j]<- a*sum(((P-H)^2)/H)
d2[j]<- (a+m)*sum(((P-H)^2)/H)
}
}
length(d2[!is.na(d2)])
range(d2)
range(rchisq(N,k-2))
for (i in 1:length(epsi)){
  posterior[i]<- sum(dstar< epsi[i])/length(dstar)
  post1[i]<- sum(d1< epsi[i])/length(d1)
  post2[i]<- sum(d2< epsi[i])/length(d2)
  power2[i]<- sum(d2 > epsi[i])/length(d2)
}
round(P, 3)
d<- mean(d2)
```

B.2 Dirichlet invariant process posterior for spherically symmetric distribution

```
library(MASS)
```

```
library(mvtnorm)
library(Emcdf)
library(rgl)
library(bivariate)
library(plot3D)
library(LaplacesDemon)
library(fMultivar)
#-------------------------  data
m=100; x<- c(-Inf,seq(-1,2, by=1),Inf);
y<- c(-Inf,seq(-1,1, by=1),Inf)
set.seed(3)
data<- rt2d(m, rho = 0, nu = 4)
#------------  probability computed by data
DD=matrix(0,nrow = length(x)-1, ncol = length(y)-1)
for (k in 1:(length(x)-1)){
  for (l in 1:(length(y)-1)){
    DD[k,l]<- sum(y[l]<= data[,2] &  data[,2]< y[l+1]
    & x[k]<= data[,1] &  data[,1]< x[k+1])/m
  }
}
RDD<- round(DD, 3)
#------  True probability by distribution of data
D=matrix(0,nrow = length(x)-1, ncol = length(y)-1)
for (k in 1:(length(x)-1)){
  for (l in 1:(length(y)-1)){
    D[k,l]<-  pt2d(x=x[k+1], y=y[l+1], rho = 0, nu = 4)
    +pt2d(x=x[k], y=y[l], rho = 0, nu = 4)
    -pt2d(x=x[k+1], y=y[l], rho = 0, nu = 4)
    -pt2d(x=x[k], y=y[l+1], rho = 0, nu = 4)
```

```
  }
}
RD<- round(D, 3)
##################  Spherical symmetry  ###############
k=10; teta=numeric()
n= m*k; X<- matrix(0,n,2)
for (i in 0:(k-1)){
  for (j in 1:m){
    r<- i*m+j
    t<- teta[i+1]<- 2*pi*i/k
    xdat<- data[j,]
    A<- matrix(c(cos(t),sin(t),-sin(t),cos(t)),2,2)
    X[r,]<- A%*%xdat
  }
}
######### probability computed by DIP #############
a=200; n=100
g=numeric(); tstar=matrix(0,n,2); p2=numeric()
g2=numeric(); pp=numeric()
y1<- rexp(n+1, 1)
for (i in 1:n ){
  g[i]<- sum(y1[1:i])/sum(y1)
  g2[i]<- qgamma(1-g[i], (a+m)/n, 1)
  u<- runif(1)
  if  (u< a/(a+m)){
    #tstar[i,]<- mvrnorm(1, rep(0, 2), diag(2))
    tstar[i,]<- rcauchy2d(1, rho = 0)
     } else{
    r<- sample(1:nrow(X),1,replace=T)
```

```
    tstar[i,]<- X[r,]
  }
}
p2<- g2/sum(g2)
dpp<- cbind(tstar,p2)
mpp<- dpp[,3]
xx<- dpp[, 1:2]
t<- x_uniq<- unique(xx)
#----------------------------------------------
for (i in 1: nrow(x_uniq)) {
  ind_p=numeric()
  for (j in 1: nrow(xx)){
    if (identical(xx[j,],x_uniq[i,])==TRUE)
    { ind_p<- cbind(ind_p,mpp[j])}}
  if (length(ind_p[!is.na(ind_p)]) > 0)
  {pp[i]<- sum(ind_p[!is.na(ind_p)]) }
  else {pp[i]<- 0}
}
DIP<- cbind(t,pp)
PP=matrix(0,nrow = length(x)-1, ncol = length(y)-1)
for (k in 1:(length(x)-1)){
  for (l in 1:(length(y)-1)){
   PP[k,l]<- sum(pp[which(y[l]<= t[,2] &  t[,2]< y[l+1]
   & x[k]<= t[,1] &  t[,1]< x[k+1])])
  }
}
RPP<- round(PP,3)
######### probability computed by DP ############
g=numeric(); tstar=matrix(0,n,2)
```

```
g2=numeric(); pp=numeric()
y1<- rexp(n+1, 1)
for (i in 1:n ){
  g[i]<- sum(y1[1:i])/sum(y1)
  g2[i]<- qgamma(1-g[i], (a+m)/n, 1)
  u<- runif(1)
  if  (u< a/(a+m)){
    #tstar[i,]<- mvrnorm(1, rep(0, 2), diag(2))
    tstar[i,]<- rcauchy2d(1, rho = 0)
  } else{
    r<- sample(1:nrow(data),1,replace=T)
    tstar[i,]<- data[r,]
  }
}
p2<- g2/sum(g2)
dpp<- cbind(tstar,p2)
mpp<- dpp[,3]
xx<- dpp[, 1:2]
t<- x_uniq<- unique(xx)
#-----------------------------------------------
if (identical(x_uniq,xx)== FALSE){
  for (i in 1: nrow(x_uniq)) {
    ind_p=numeric()
    for (j in 1: nrow(xx)){
      if (identical(xx[j,],x_uniq[i,])==TRUE)
      { ind_p<- cbind(ind_p,mpp[j])}}
    if (length(ind_p[!is.na(ind_p)]) > 0)
    {pp[i]<- sum(ind_p[!is.na(ind_p)]) }
    else {pp[i]<- 0}
```

```
  }
} else {t<- xx & pp<- mpp}
DP<- cbind(t,pp)
PP1=matrix(0,nrow = length(x)-1, ncol = length(y)-1)
for (k in 1:(length(x)-1)){
  for (l in 1:(length(y)-1)){
    PP1[k,l]<- sum(pp[which(y[l]<= t[,2] &  t[,2]< y[l+1]
    & x[k]<= t[,1] &  t[,1]< x[k+1])])
  }
}
RPP1<- round(PP1, 3)
#################  Plots  #################################
#-----------------  Plot true distribution of data
fdejong <- function (x,y) {
  return (pt2d(x, y, rho = 0, nu = 4))
}
x <- seq(-5, 5, length= 30)
y <- x
z <- outer(x, y, fdejong)
filled.contour(x,y,z,color=terrain.colors,main="CDF␣of␣t(4)"
, xlab='X',ylab='Y',xlim=c(-2,2), ylim=c(-2,2))
#--------------------------
N=1000; datan<- rt2d(N, rho = 0, nu = 4)
frame()
plotcdf(datan, type = "wireframe",angle = 60
,main = paste("CDF␣of␣t(4)"))
#-------------  plot DP and DIP posterior distributions
x_value<- DIP[,1:2]; prob<- DIP[,3]
# x_value<- DP[,1:2]; prob<- DP[,3]  # for DP posterior
```

```
s1<- sample(1:nrow(x_value),N,prob,replace=T)
x_value<- x_value[s1,]
frame()
plotcdf(x_value, type = "wireframe",angle = 60
,main = paste("DP␣CDF") )
#----------------------------------------
xv<- x_value[,1]; yv<- x_value[,2]
B<- Biemcdf(x_value)
rownames(B)<- colnames(B) <- numeric()
z<- B;  xv<- sort(xv); yv<- sort(yv)
filled.contour(xv,yv,z,color=terrain.colors,
main="DP␣CDF", xlab='X',ylab='Y', xlim=c(-2,2), ylim=c(-2,2))
################# Rectangular symmetry #############
m=50; a=0; b=7; c=0; d=5; midx1=(a+b)/2 ; midx2=(c+d)/2
set.seed(1)
data=matrix(0,m,2)
i <- 1
while (i < m+1) {
x1v<- rt2d(1, rho = 0, nu = 2)
if (a < x1v[,1] & x1v[,1] < b  &  c < x1v[,2] & x1v[,2] < d)
{data[i,]<- x1v; i = i+1}
}
#---------------- Data by DIP
data1= data2= data3= data
data3[,1]<- data1[,1]<- 2*midx1-data[,1]
data3[,2]<- data2[,2]<- 2*midx2-data[,2]
datafin<- rbind(data,data1,data2,data3)
x_rec<- datafin[,1]; y_rec<- datafin[,2]
#-------------  DIP posterior distribution
```

```
N=1000;aa=2;A=numeric(); S=numeric()
for (l in 1:N){
n=100; d2=numeric()
g=numeric(); tstar=matrix(0,n,2); p2=numeric()
g2=numeric(); pp=numeric()
y1<- rexp(n+1, 1)
for (i in 1:n ){
  g[i]<- sum(y1[1:i])/sum(y1)
  g2[i]<- qgamma(1-g[i], (a+m)/n, 1)
  u<- runif(1)
  if  (u< aa/(aa+m)){
    x1=runif(1,0,7)
    x2=runif(1,0,5)
    tstar[i,]<- c(x1,x2)
  } else{
    r<- sample(1:nrow(datafin),1,replace=T)
    tstar[i,]<- datafin[r,]
  }
}
p2<- g2/sum(g2)
dpp<- cbind(tstar,p2)
mpp<- dpp[,3]
xx<- dpp[, 1:2]
t<- x_uniq<- unique(xx)

if (identical(x_uniq,xx)== FALSE){
  for (i in 1: nrow(x_uniq)) {
    ind_p=numeric()
    for (j in 1: nrow(xx)){
```

```
      if (identical(xx[j,],x_uniq[i,])==TRUE)
      { ind_p<- cbind(ind_p,mpp[j])}}
    if (length(ind_p[!is.na(ind_p)]) > 0)
    {pp[i]<- sum(ind_p[!is.na(ind_p)]) }
    else {pp[i]<- 0}
  }
} else {t<- xx & pp<- mpp}
#------   Computing A and S by DIP  ---------
xmin<- min(t[,1]); xmax<- max(t[,1])
ymin<- min(t[,2]); ymax<- max(t[,2])
A[l]= 2*((xmax-xmin)+(ymax-ymin))
S[l]= (xmax-xmin)*(ymax-ymin)
}
AA<- mean(A);AS<- mean(S)
#----------------------  DP
A1=numeric(); S1=numeric()
for (l in 1:N){
n=100; d2=numeric()
  g=numeric(); tstar=matrix(0,n,2); p2=numeric()
  g2=numeric(); pp=numeric()
  y1<- rexp(n+1, 1)
  for (i in 1:n ){
    g[i]<- sum(y1[1:i])/sum(y1)
    g2[i]<- qgamma(1-g[i], (a+m)/n, 1)
    u<- runif(1)
    if  (u< aa/(aa+m)){
      x1=runif(1,0,7)
      x2=runif(1,0,5)
      tstar[i,]<- c(x1,x2)
```

```
    } else{
      r<- sample(1:nrow(data),1,replace=T)
      tstar[i,]<- data[r,]
    }
  }
  p2<- g2/sum(g2)
  dpp<- cbind(tstar,p2)
  mpp<- dpp[,3]
  xx<- dpp[, 1:2]
  t1<- x_uniq<- unique(xx)

  if (identical(x_uniq,xx)== FALSE){
    for (i in 1: nrow(x_uniq)) {
      ind_p=numeric()
      for (j in 1: nrow(xx)){
        if (identical(xx[j,],x_uniq[i,])==TRUE)
        { ind_p<- cbind(ind_p,mpp[j])}}
      if (length(ind_p[!is.na(ind_p)]) > 0)
      {pp[i]<- sum(ind_p[!is.na(ind_p)]) }
      else {pp[i]<- 0}
    }
  } else {t1<- xx & pp<- mpp}
#-------   Computing A and S by DP--------------
xmin1<- min(t1[,1]); xmax1<- max(t1[,1])
ymin1<- min(t1[,2]); ymax1<- max(t1[,2])
A1[l]= 2*((xmax1-xmin1)+(ymax1-ymin1))
S1[l]= (xmax1-xmin1)*(ymax1-ymin1)
}
AA1<- mean(A1);AS1<- mean(S1)
```

```
#--------------- Computing A and S  by data
xmin<- min(data[,1]); xmax<- max(data[,1])
ymin<- min(data[,2]); ymax<- max(data[,2])
A3= 2*((xmax-xmin)+(ymax-ymin))
S3= (xmax-xmin)*(ymax-ymin)
#---------------  histogram of S and A
hist(A, freq=FALSE,col="grey",xlab="estimated␣A"
,ylab="Density",main="DIP(aH)")
hist(S,freq=FALSE,col="grey",xlab="estimated␣S"
,ylab="Density",main="DIP(aH)")
hist(A1,freq=FALSE,col="grey",xlab="estimated␣A"
,ylab="Density",main="DP(aH)")
hist(S1,freq=FALSE,col="grey",xlab="estimated␣S"
,ylab="Density",main="DP(aH)")
#-----------------   plot true rectangular
B<- matrix(c(0,0,7,0,0,5,7,5),4,2, byrow=T)
b1=max(B[,1]);b2=max(B[,2])
border<- matrix(c(a,c,b,c,a,d,b,d),4,2, byrow=T)
frame()
plot( border, type='n', xlim=c(0,b1), ylim=c(0,b2)
, xlab='X', ylab='Y')
rect(a,c,b,d, xlim=c(0,b), ylim=c(0,d), xlab='X', ylab='Y'
,border='red')
#----------------  plot DIP and DIP symmetric data
par(new=TRUE)
xmin<- min(t[,1]); xmax<- max(t[,1])
ymin<- min(t[,2]); ymax<- max(t[,2])
plot(t, xlim=c(0,b1), ylim=c(0,b2), xlab='X', ylab='Y'
,pch=6, col='blue')
```

```
rect(xmin,ymin,xmax,ymax,border='blue')
par(new=TRUE)
xmin<- min(t1[,1]); xmax<- max(t1[,1])
ymin<- min(t1[,2]); ymax<- max(t1[,2])
plot(t1, xlim=c(0,b1), ylim=c(0,b2), xlab='X', ylab='Y'
,pch=2, col='green')
rect(xmin,ymin,xmax,ymax,border='green')
#------------------------- plot data
par(new=TRUE)
plot(data,xlim=c(0,b1), ylim=c(0,b2), xlab='X', ylab='Y', pch=3)
xmin<- min(data[,1]); xmax<- max(data[,1])
ymin<- min(data[,2]); ymax<- max(data[,2])
rect(xmin,ymin,xmax,ymax,border='black')
abline(h=midx2,col='black',lty=3)
abline(v=midx1,col='black',lty=3)
```

Bibliography

Al Labadi, L. (2012). *On new constructive tools in Bayesian nonparametric inference.* PhD thesis, Université d'Ottawa/University of Ottawa.

Al Labadi, L. and Zarepour, M. (2013). A bayesian nonparametric goodness of fit test for right censored data based on approximate samples from the beta-stacy process. *Canadian Journal of Statistics*, 41(3):466–487.

Al Labadi, L. and Zarepour, M. (2014). Goodness-of-fit tests based on the distance between the dirichlet process and its base measure. *Journal of Nonparametric Statistics*, 26(2):341–357.

Al Labadi, L. and Zarepour, M. (2017). Two-sample kolmogorov-smirnov test using a bayesian nonparametric approach. *Mathematical Methods of Statistics*, 26(3):212–225.

Antoniak, C. E. (1974). Mixtures of dirichlet processes with applications to bayesian nonparametric problems. *The annals of statistics*, 2(6):1152–1174.

Berger, J. O. and Guglielmi, A. (2001). Bayesian and conditional frequentist testing of a parametric model versus nonparametric alternatives. *Journal of the American Statistical Association*, 96(453):174–184.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 9(6):1196–1217.

Bickel, P. J. and Wichura, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *The Annals of Mathematical Statistics*, 42(5):1656–1670.

Billingsley, P. (2013). *Convergence of Probability Measures.* John Wiley & Sons.

Bondesson, L. (1982). On simulation from infinitely divisible distributions. *Advances in Applied Probability*, 14(4):855–869.

Carlton, M. A. (1999). *Applications of the two-parameter Poisson-Dirichlet distribution.* PhD thesis, Department of Statistics, University of California, Los Angeles. unpublished thesis.

D'Agostino, R. B. (1986). *Goodness-of-Fit Techniques.* Marcel Dekker, New York.

Dahiya, R. C. and Gurland, J. (1973). How many classes in the pearson chi-square test? *Journal of the American Statistical Association*, 68(343):707–712.

Dalal, S. (1979a). Dirichlet invariant processes and applications to nonparametric estimation of symmetric distribution functions. *Stochastic Processes and their Applications*, 9(1):99–107.

Dalal, S. (1979b). Nonparametric and robust bayes estimation of location. optimizing methods in statistics (ed. by j. s. rustagi), 141-166.

Dalal, S. R. (1980). Nonparametric bayes decision theory. *Trabajos de estadística y de investigación operativa*, 31(1):523.

Efron, B. (1992). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics*, pages 569–593. Springer.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90(430):577–588.

Ferguson, T. S. (1973). A bayesian analysis of some nonparametric problems. *Annals of Statistics*, 1(2):209–230.

Ferguson, T. S. and Klass, M. J. (1972). A representation of independent increment processes without gaussian components. *Annals of Mathematical Statistics*, 43(5):1634–1643.

Florens, J.-P., Richard, J.-F., and Rolin, J. M. (1996). Bayesian encompassing specification tests of a parametric model against a nonparametric alternative. Technical Report 9608, Université catholique de Louvain, Institut de Statistique.

Ghosal, S. (2010). The dirichlet process, related priors and posterior asymptotics. In *Bayesian Nonparametrics*, pages 35–79. Cambridge University Press, Cambridge.

Griffiths, T. L. and Ghahramani, Z. (2011). The indian buffet process: An introduction and review. *Journal of Machine Learning Research*, 12(Apr):1185–1224.

Hamada, M. S., Wilson, A., Reese, C. S., and Martz, H. (2008). *Bayesian Reliability*. Springer, New York.

Hamdan, M. (1963). The number and width of classes in the chi-square test. *Journal of the American Statistical Association*, 58(303):678–689.

Holmes, C. C., Caron, F., Griffin, J. E., and Stephens, D. A. (2015). Two-sample bayesian nonparametric hypothesis testing. *Bayesian Analysis*, 10(2):297–320.

Hsieh, P.-H. (2013). A nonparametric assessment of model adequacy based on kullback-leibler divergence. *Statistics and Computing*, 23(2):149–162.

Ishwaran, H. and Zarepour, M. (2002). Exact and approximate sum representations for the dirichlet process. *Canadian Journal of Statistics*, 30(2):269–283.

James, L. F. (2008). Large sample asymptotics for the two-parameter poisson–dirichlet process. volume 3, pages 187–199. Institute of Mathematical Statistics.

Johnson, V. E. (2004). A bayesian χ^2 test for goodness-of-fit. *Annals of Statistics*, 32(6):2361–2384.

Johnson, V. E. (2007). Bayesian model assessment using pivotal quantities. *Bayesian Analysis*, 2(4):719–733.

Koehler, K. J. and Gan, F. (1990). Chi-squared goodness-of-fit tests: Cell selection and power. *Communications in Statistics-Simulation and Computation*, 19(4):1265–1278.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86.

Lijoi, A., Mena, R. H., and Prünster, I. (2005). Hierarchical mixture modeling with normalized inverse-gaussian priors. *Journal of the American Statistical Association*, 100(472):1278–1291.

Lo, A. Y. (1984). On a class of bayesian nonparametric estimates: I. density estimates. *Annals of Statistics*, 12(1):351–357.

Lo, A. Y. (1987). A large sample study of the bayesian bootstrap. *Annals of Statistics*, 15(1):360–375.

Lu, C. and Tang, X. (2015). Surpassing human-level face verification performance on lfw with gaussianface. In *AAAI*, pages 3811–3819.

Macdonald, D., Ball, F., and Hough, N. (1980). The evaluation of home range size and configuration using radio tracking data. In *A handbook on biotelemetry and radio tracking*, pages 405–424. Elsevier.

Medvedovic, M. and Sivaganesan, S. (2002). Bayesian infinite mixture model based clustering of gene expression profiles. *Bioinformatics*, 18(9):1194–1206.

Muliere, P. and Secchi, P. (1996). Bayesian nonparametric predictive inference and bootstrap techniques. *Annals of the Institute of Statistical Mathematics*, 48(4):663–673.

Muliere, P. and Tardella, L. (1998). Approximating distributions of random functionals of ferguson-dirichlet priors. *Canadian Journal of Statistics*, 26(2):283–297.

Neal, R. M. (1992). Bayesian mixture modeling. In *Maximum Entropy and Bayesian Methods*, pages 197–211. Springer, New York.

Quine, M. and Robinson, J. (1985). Efficiencies of chi-square and likelihood ratio goodness-of-fit tests. *Annals of Statistics*, 13(2):727–742.

Rasmussen, C., de la Cruz, B., Ghahramani, Z., and Wild, D. (2009). Modeling and visualizing uncertainty in gene expression clusters using dirichlet process mixtures. *IEEE/ACM transactions on computational biology and bioinformatics*, 6(4):615–628.

Ripley, B. and Rasson, J.-P. (1977). Finding the edge of a poisson forest. *Journal of Applied probability*, 14(3):483–491.

Rubin, D. B. (1981). The bayesian bootstrap. *The annals of statistics*, 9(1):130–134.

Sethuraman, J. (1994). A constructive definition of dirichlet priors. *Statistica Sinica*, 4(2):639–650.

Singh, K. (1981). On the asymptotic accuracy of efron's bootstrap. *The Annals of Statistics*, 9(6):1187–1195.

Swartz, T. (1999). Nonparametric goodness-of-fit. *Communications in Statistics-Theory and Methods*, 28(12):2821–2841.

Tokdar, S. T. and Martin, R. (2011). Bayesian test of normality versus a dirichlet process mixture alternative. *arXiv:1108.2883*.

Verdinelli, I. and Wasserman, L. (1998). Bayesian goodness-of-fit testing using infinite-dimensional exponential families. *Annals of Statistics*, 26(4):1215–1241.

Viele, K. (2000). Evaluating fit using dirichlet processes. Technical Report 384, Department of Statistics, University of Kentucky.

Wilks, S. S. (1962). *Mathematical Statistics*. Wiley, New York.

Yamato, H. (1986). Bayes estimates of estimable parameters with a direchlet invariant process. *Communications in Statistics-Theory and Methods*, 15(8):2383–2390.

Yamato, H. (1987). Nonparametric bayes estimates of estimable parameters with a dirichlet invarant process and inveriant u-statistics. *Communications in Statistics-Theory and Methods*, 16(2):525–543.

Zarepour, M. and Al Labadi, L. (2012). On a rapid simulation of the dirichlet process. *Statistics & Probability Letters*, 82(5):916–924.