1 Introduction

In this thesis, we study the c-characteristic function, random functionals of Ferguson-Dirichlet processes, compatibility of given conditionals, and inverse Bayes formula. First, we give a brief introduction to the above four issues in Section 1. In Section 2, we introduce some notations and equations which will be used throughout this thesis. Next, we discuss the c-characteristic function and construct its inversion formula in Section 3. In Section 4, we use the c-characteristic function and its inversion formula to study random functionals of Ferguson-Dirichlet processes. Furthermore, we provide a fully and cleanly unified theory of compatibility of the given conditionals in Section 5. In Section 6, we use the compatibility theory to establish the generalized inverse Bayes formula. Finally, conclusions are given in Section 7.

1.1 The c-characteristic function

Jiang (1988) first gave the univariate c-characteristic function (or the univariate c-transformation), which has many properties similar to those of the traditional characteristic function (or the Fourier transformation). Using this new characteristic function, one can solve many problems that are difficult to manage with the traditional characteristic function. Examples can be seen in Jiang (1988, 1991). Recently, Jiang, Dickey, and Kuo (2004) further extended the univariate c-characteristic function to the multivariate case, and provided some useful properties. In Subsections 3.1 and 3.2, we review some well-known results, which will be used in this thesis, of the univariate and multivariate c-characteristic functions, respectively. We also provide important additional properties of the multivariate c-characteristic function of a spherical distribution in Subsection 3.2. One of these properties can be used to determine whether a distribution is spherical when the c-characteristic function is known. Moreover, we can easily obtain the c-characteristic function of any marginal of a spherical distribution through its multivariate c-characteristic function.

Although the univariate c-characteristic function is applicable to many problems, its usefulness has been somewhat limited due to the lack of its inversion formula. In Subsection 3.3, we construct inversion formulas of a univariate c-characteristic function for any $c > 0$. Applying them, the exact probability density function (PDF) or cumulative distribution function (CDF) of a random variable may be obtained easily when the associated univariate c-characteristic function is known.

A relation between the univariate c-characteristic function and the Fourier transformation is given in Subsection 3.4. With this relation, we derive the second approach to obtain the PDF from its known univariate c-characteristic function, via the inversion formula of a Fourier transformation.

Provost and Cheong (2000) noted that several test statistics, such as Moran and Geary's indices, the Cliff-Ord statistic for spatial correlation, the sample coefficient of determination, F-ratios and the sample autocorrelation coefficient, can be expressed in terms of linear combinations of the components of Dirichlet random vectors or process a similar structure. In Subsection 3.5, we give the c-characteristic function expression of such linear combination, and then use our inversion formulas to derive the associated explicit PDF.

1.2 The Ferguson-Dirichlet process

Ferguson (1973) first introduced the Ferguson-Dirichlet process and studied its applications to Bayesian nonparametric statistics. Before stating the definition of a Ferguson-Dirichlet process, we shall give the definition of a Dirichlet distribution.

Definition 1.1 *If the random vector $\boldsymbol{X}$ has the density in any $L-1$ of its coordinates,*

$$f(\boldsymbol{x}; \boldsymbol{b}) = \frac{1}{B(\boldsymbol{b})} \cdot \prod_{j=1}^{L} x_j^{b_j - 1},$$

for all $\boldsymbol{x}$ in the probability simplex $\{\boldsymbol{x} = (x_1, \ldots, x_L)' \mid$ each $x_j \geq 0,\ x_+ = 1\}$, where $x_+ = \sum_{j=1}^{L} x_j$ and the parameter vector $\boldsymbol{b} = (b_1, \ldots, b_L)'$, each $b_j > 0$, and

$$B(\boldsymbol{b}) = \frac{\prod_{j=1}^{L} \Gamma(b_j)}{\Gamma(b_+)},$$

then $\boldsymbol{X}$ *is said to follow a Dirichlet distribution with parameter* $\boldsymbol{b}$*, and is denoted by* $\boldsymbol{X} \sim Dir(\boldsymbol{b})$.

Now, we state the definition of a Ferguson-Dirichlet process.

Definition 1.2 *Let* μ *be a finite non-null measure on* $(\Omega, \mathcal{B})$*, where* $\mathcal{B}$ *is the* σ*-field of Borel subsets of Euclidean space* Ω*; and let* U *be a stochastic process indexed by elements of* $\mathcal{B}$*. We say that* U *is a Ferguson-Dirichlet process with parameter* μ*, denoted by* $U \sim D(\mu)$ *on* Ω*, if for every finite measurable partition* $\{B_1, \ldots, B_m\}$ *of* Ω *(i.e., the* B_i*'s are measurable, disjoint, and* $\bigcup_{i=1}^m B_i = \Omega$*), the random vector* $(U(B_1), \ldots, U(B_m))$ *has a Dirichlet distribution with parameter* $(\mu(B_1), \ldots, \mu(B_m))$*.*

Several researchers have studied the following random functional of a Ferguson-Dirichlet process,

$$\xi_\mu(h) = \int_\Omega h(y)\, dU(y),\ U \sim D(\mu), \tag{1.1}$$

where $h(y)$ is any bounded measurable function defined on Ω. Some earlier results can be seen in Hannum, Hollander, and Langberg (1981), Yamato (1984), and Jiang (1988). Cifarelli and Regazzini (1990) and Regazzini, Guglielmi, and Di Nunno (2002) showed that the distribution function of $\xi_\mu(h)$ can be expressed in terms of a Lebesgue integral. Jiang (1991) gave the distribution of $\int y\, dU(y)$ when the parameter measure μ of U has the uniform probability distribution on the unit circle. Diaconis and Kemperman (1996) provided an explicit PDF for $\int y\, dU(y)$ when μ is any probability measure on $[0, 1]$. Cifarelli and Melilli (2000) and Epifani, Guglielmi, and Melilli (2006) studied the distribution of the random variable $\int y^2\, dU(y) - (\int y\, dU(y))^2$. Lijoi and Regazzini (2004) used multiple hypergeometric functions to obtain a new and direct procedure for determining the exact form of $\xi_\mu(h)$. Hjort and Ongaro (2005) obtained more properties of $\xi_\mu(h)$. In Subsection 4.1, we apply the inversion formulas of a c-characteristic function to study $\xi_\mu(h)$. First, we give the univariate c-characteristic function expression of $\xi_\mu(h)$. With this expression and using our inversion formula, we then provide explicit PDFs of random means of a Ferguson-Dirichlet process with several interesting parameter measures, e.g., generalized beta distribution on $(-1, 1)$ with parameters $(L-1)/2$ and $(L-1)/2$, where L is any integer that is greater than 1. In addition, the explicit PDF of each random moment of a Ferguson-Dirichlet process with uniform probability measure on $(-1, 1)$ as its parameter is also given. More generally, we also obtain the Lebesgue integral expression, which is consistent with that given by Cifarelli and Regazzini (1990) and Regazzini, Guglielmi, and Di Nunno (2002), of the distribution of $\xi_\mu(h)$.

Only a few distributional results for a vector of functionals of a Ferguson-Dirichlet process are available in the literature; for instance, Jiang (1991) studied the distribution of the random mean of a Ferguson-Dirichlet process over a unit circle. In Subsection 4.2, we give the multivariate c-characteristic function expression of any functional of a Ferguson-Dirichlet process over any high-dimensional region. With this expression, we show that the random mean of a Ferguson-Dirichlet process over a spherical surface in n dimensions has a spherical distribution on the n-dimensional ball. The exact PDF of the random mean will also be given. Moreover, we will obtain the exact distribution of the random mean of a Ferguson-Dirichlet process over any ellipsoidal surface in n-space.

1.3 Compatible conditional distributions

Multivariate distributions that only their conditional and/or marginal distributions available are often used in probability modelling and Bayesian statistics. In particular, there is great interest in determining the joint distribution or marginal distributions when only conditional distributions are specified. The Gibbs sampler and Markov Chain Monte Carlo methods are important research areas that may involve the characterization of a joint distribution by given conditional distributions (Liu, 1996). When both the conditional distribution functions of $X|Y$ and $Y|X$ are specified, the issue of compatibility is whether there exists a joint distribution function of (X, Y) with these specified functions as conditional distribution functions. It is important to provide some criteria which are simple and easy to use for compatibility checking.

In bivariate discrete cases, there are several versions of necessary and sufficient conditions for compatibility given by Arnold and Press (1989) and Arnold, Castillo, and Sarabia (2002, 2004). However, an incompatible example (see Example 5.1 in Subsection 5.1) satisfying Arnold and Press' (1989) compatible condition is found. In some cases, the Arnold, Castillo, and Sarabia's (2004) condition for compatibility checking was found to be an uneasy task and less efficient. By investigating the structure of a ratio matrix, we successfully solve both the "existence" and the "uniqueness" problems. After interchanging rows and/or columns, the ratio matrix can be rearranged to an "irreducible block diagonal matrix." We found that the ranks of all blocks on the diagonal determine the compatibility of finite discrete conditional distributions. More precisely, the necessary and sufficient condition for compatibility is that every block is of rank one. This new method, which needs only some elementary operations of matrices, provides a simpler and more efficient approach. Equivalent criteria and details are presented in Subsection 5.1.

When the given conditional distributions are compatible, it is natural to ask whether the associated joint distribution is unique. This issue has been addressed by Amemiya (1975), Gourieroux and Montfort (1979), Nerlove and Press (1986), and Arnold and Press (1989). Arnold and Press (1989) pointed out that the condition for uniqueness is generally difficult to check. Through the structure of the ratio matrix, we give some quite simple criteria for uniqueness checking. Moreover, we obtain all the associated joint distributions even they are not unique.

Other related works such as near compatibility and compatibility of partial conditional distributions can be seen in Arnold, Castillo, and Sarabia (2002, 2004).

In Subsection 5.2, we follow the similar procedure of the bivariate case to study the compatibility of trivariate discrete conditionals. General multivariate results are presented in Subsection 5.3. In Subsections 5.4 and 5.5, we study the compatibility of continuous conditionals.

1.4 Inverse Bayes formula

Finding the associated joint distribution when full conditionals are given is an interesting probability issue and is important in several areas of statistics. For references concerning this issue and related applications, please see Tanner and Wong (1987), Arnold and Press (1989), Ng (1997), Tan, Tian and Ng (2003), Tian, Tan and Ng (2007), and references therein. Ng (1997) first introduced the inverse Bayes formula (IBF), which is an explicit formula for prior density given the posterior density and the likelihood. In distribution theory, the IBF says that the associated marginal density can be expressed explicitly in terms of the given conditional densities. Tian and Tan (2003) obtained extensive applications of the IBF method in a wide variety of statistical problems, including bivariate distributions, a genetic linkage model, sample surveys with nonresponse, misclassified multinomial data, a reliability growth model with missing data, and hierarchical models. Tian, Ng, and Geng (2003) further demonstrated that Bayesian computation can be routinely performed by the IBF method when the posterior has a grouped or a nested Dirichlet distribution.

However, users might likely experience difficulties in using IBF to discover the associated marginal density for nonproduct measurable space, see our Example 6.1 (for discrete cases) and Example 6.5 (for continuous cases). Tian and Tan (2003) extended the IBF in the product measurable space to the nonproduct measurable space having a transition point. This modified IBF is called MIBF in the paper. Although MIBF embodies an important improvement, it does not work in all cases in nonproduct space, as shown in our Example 6.5. Both IBF and MIBF are mainly used to produce the unique associated marginal density when the given compatible conditional densities are known. However, neither of them can be used to derive all possible marginal densities when they are not unique. To overcome the above shortcomings, we apply the compatibility theory to give a fully and cleanly unified theory of IBF and construct a generalized IBF (GIBF) that is applicable in the more general measurable space and can be used to obtain all associated marginal densities. Using GIBF, we offer a density fitting algorithm, which avoids the problems of convergence in iterative procedures such as the Gibbs sampler.