GOODNESS-OF-FIT TESTS FOR DIRICHLET DISTRIBUTIONS WITH APPLICATIONS

Yi Li

A Dissertation
Submitted to the Graduate College of Bowling Green State University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2015

Committee:

Maria Rizzo, Advisor

Pamela Bechtel,
Graduate Faculty Representative

James Albert

Wei Ning
We present five new goodness-of-fit tests for Dirichlet distributions. Since beta, Dirichlet, and generalized Dirichlet distributions are widely used in statistics such as Dirichlet regression, data mining, and machine learning. Thus, goodness-of-fit tests for beta and Dirichlet distribution are important applications. We develop new test procedures for testing beta and Dirichlet distribution.

The first test, called energy test, is based on energy statistics. The second test, called distance covariance test, is based on the property of complete neutrality of the Dirichlet distribution and distance covariance tests. We also expand the distance covariance test into testing the mutual independence of all components of arbitrary random vector. The third, fourth and fifth tests, called the triangle tests, are based on the interpoint distances. Simulation studies of power performance for the new beta and Dirichlet goodness-of-fit tests are presented. Results show that distance covariance goodness-of-fit test have good performance in contaminated Dirichlet distributions which contain the small perturbation in the Dirichlet distribution.

The parameter estimation based on the maximum likelihood is derived for the generalized Dirichlet distribution and the initial value for the iteration of the Newton-Raphson algorithm is also proposed. Applications includes the goodness-of-fit tests of Dirichlet and generalized Dirichlet distributions, model evaluation of Dirichlet regression models, and influence diagnostics of Dirichlet regression models.
ACKNOWLEDGMENTS

I would like to express my special gratitude to my advisor Dr. Rizzo, for her excellent guidance, support, patience. I will always remember my discussion with Dr. Rizzo about the new results and algorithm of this dissertation. I want to thank my committee members, Dr. James Albert and Dr. Wei Ning, whose work are so important for this dissertation. I would also like to thank my family. They were always supporting me and encouraging me with their best wishes.

In the end, I thank Bowling Green State University, which provided me a chance to study here.
# TABLE OF CONTENTS

## CHAPTER 1 INTRODUCTION

1

## CHAPTER 2 LITERATURE REVIEW

2.1 EDF Tests for Univariate Distribution 3

2.2 Chi-squared Test and Likelihood Ratio Tests 4

2.3 Two-Dimensional Kolmogorov-Smirnov Goodness-of-Fit Tests 6

2.4 Energy Tests 6

2.5 Triangle Tests 8

## CHAPTER 3 BETA AND DIRICHLET DISTRIBUTION

3.1 Beta Distribution 12

3.2 Stochastic Representation of Beta Distribution 14

3.3 Dirichlet Distribution 15

3.4 Stochastic Representation of Dirichlet Distribution 16

3.5 Estimation for the Beta and Dirichlet Parameter Vector 17

3.5.1 MLE Based on the Newton-Raphson Algorithm 17

3.5.2 Method of Moments Estimates 19

## CHAPTER 4 ENERGY AND TRIANGLE TESTS OF BETA AND DIRICHLET DISTRIBUTION

4.1 Univariate Energy Goodness-of-Fit Statistics 20

4.2 Energy Goodness-of-Fit Test for Uniform Distribution 21

4.3 Data Transformation and Standardization in the Dirichlet Distribution 26

4.4 Energy Test of Beta or Dirichlet Distribution 30

4.5 Triangle Tests of Beta or Dirichlet Distribution 31
CHAPTER 5  DISTANCE CORRELATION TEST OF DIRICHLET DISTRIBUTIONS 33

5.1 Distance Covariance and Distance Correlation ................. 33

5.2 Characterization of Beta and Dirichlet Distributions .......... 37
   5.2.1 Complete Neutrality ........................................ 37

5.3 Testing of Mutual Independence of All Components in a Random Vector .... 40

5.4 Distance Covariance Goodness-of-Fit Tests ...................... 41

CHAPTER 6  SIMULATION STUDY 44

6.1 Simulation Study .................................................. 44
   6.1.1 Energy and Triangle Tests ................................. 44
   6.1.2 Distance Covariance Test ................................... 45

6.2 Contaminated Dirichlet Distribution ............................ 46

6.3 Logistic Normal Distribution .................................... 54

CHAPTER 7  MAIN RESULTS FOR GENERALIZED DIRICHLET DISTRIBUTIONS 58

7.1 Generalized Dirichlet Distribution .............................. 58

7.2 Estimation for the Generalized Dirichlet Distribution Parameter Vector .... 62
   7.2.1 MLE Based on the Newton-Raphson Algorithm .......... 63
   7.2.2 Direct Proof of Global Convexity of Log Likelihood Function of the Generalized Dirichlet ......................... 65
   7.2.3 The Initial Values of Iteration .............................. 67
   7.2.4 Iteration Contingency Criteria .............................. 67

7.3 Generalized Dirichlet Distribution Model vs. Dirichlet Distribution Model .... 68
   7.3.1 The Likelihood Ratio Test .................................. 69
   7.3.2 The Wald Test ................................................ 69
   7.3.3 Rao’s Score Test ............................................. 70
   7.3.4 Bootstrap Hypothesis Testing ............................... 71
<table>
<thead>
<tr>
<th>CHAPTER 8 APPLICATIONS</th>
<th>74</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1 Applications of the Goodness-of-Fit Test of Dirichlet and Generalized Dirichlet Distribution</td>
<td>74</td>
</tr>
<tr>
<td>8.2 Application to Goodness-of-Fit Testing of Dirichlet Regression Models</td>
<td>75</td>
</tr>
<tr>
<td>8.3 Application to the Influence Diagnostics of the Dirichlet Regression Models</td>
<td>80</td>
</tr>
</tbody>
</table>

| CHAPTER 9 SUMMARY AND FURTHER DIRECTIONS | 86 |

| BIBILOGRAPHY | 88 |
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Regions defining $h_k$, $k = 1, 2, 3$ .......................... 10</td>
</tr>
<tr>
<td>5.1</td>
<td>Probability of family-wise significance level based on Bonferroni correction. . . 42</td>
</tr>
<tr>
<td>6.1</td>
<td>Contour plot of density function of variables $Y_1$ and $Y_2$. .......................... 48</td>
</tr>
<tr>
<td>6.2</td>
<td>Contour plot of density function of Dir(1, 3/4, 5/4). ................................. 49</td>
</tr>
<tr>
<td>6.3</td>
<td>Empirical power of five tests of Dirichlet against a contaminated Dirichlet alternative in Example 6.1 ................................. 50</td>
</tr>
<tr>
<td>6.4</td>
<td>Contour plot of density function of variables $Y_1$, $Y_2$ and $Y_3$ in (5.9). ................................. 52</td>
</tr>
<tr>
<td>6.5</td>
<td>Contour plot of density function of Dir(1, 1, 2). ................................. 52</td>
</tr>
<tr>
<td>6.6</td>
<td>Contour plot of density function of Dir(1, 0.7495, 1.2505). ................................. 53</td>
</tr>
<tr>
<td>6.7</td>
<td>Empirical power of five tests of Dirichlet against a contaminated Dirichlet alternative in Example 6.2 ................................. 54</td>
</tr>
<tr>
<td>6.8</td>
<td>Empirical power of five tests of Dirichlet against a contaminated Dirichlet alternative in Example 6.3 ................................. 55</td>
</tr>
<tr>
<td>6.9</td>
<td>Empirical power of five tests of Dirichlet against logistic normal alternative in Example 6.4 ................................. 57</td>
</tr>
<tr>
<td>8.1</td>
<td>Arctic lake dataset and corresponding scatterplots. ................................. 81</td>
</tr>
<tr>
<td>8.2</td>
<td>Observed and fitted compositions in the Arctic lake sediments using Dirichlet quadratic regression model. ................................. 82</td>
</tr>
<tr>
<td>8.3</td>
<td>Residual plots of three compositions in the Arctic lake sediments using Dirichlet quadratic model. ................................. 83</td>
</tr>
<tr>
<td>8.4</td>
<td>Scatterplot of $T_1$ and $T_2$ from the transformed Arctic lake sediments using Dirichlet quadratic model. ................................. 84</td>
</tr>
</tbody>
</table>
8.5 Scatterplot of $V_1$ and $V_2$ from the transformed Arctic lake sediments using Dirichlet quadratic model.
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>Empirical Power of Five Tests of Dirichlet Against a Contaminated Dirichlet Alternative in Example 6.1</td>
<td>49</td>
</tr>
<tr>
<td>6.2</td>
<td>Empirical Power of Five Tests of Dirichlet Against a Contaminated Dirichlet Alternative in Example 6.2</td>
<td>53</td>
</tr>
<tr>
<td>6.3</td>
<td>Empirical Power of Five Tests of Dirichlet Against a Contaminated Dirichlet Alternative in Example 6.3</td>
<td>53</td>
</tr>
<tr>
<td>6.4</td>
<td>Empirical Power of Five Tests of Dirichlet Against Logistic Normal Alternative in Example 6.4</td>
<td>56</td>
</tr>
</tbody>
</table>
CHAPTER 1 INTRODUCTION

Beta, Dirichlet, and generalized Dirichlet distributions are widely used in statistics, data mining, and machine learning. The Dirichlet distribution is the multivariate beta distribution. The generalized Dirichlet distribution has more flexible variance-covariance structure than the Dirichlet distribution, which makes it more practical in data analysis. The beta distribution is also frequently used as the prior probability distribution in Bayesian analysis. The beta distribution is a good model for the prior distribution of a proportion. A Dirichlet distribution and generalized Dirichlet distribution can be fitted as the prior multinomial distribution in Bayesian analysis. The Dirichlet and generalized Dirichlet distributions can also be applied to compositional data. A Dirichlet regression model regresses the compositional data on the corresponding covariates. Before we estimate the parameters of the beta, Dirichlet or generalized Dirichlet distributions, we need to test whether the data came from a population that has a beta, Dirichlet, or generalized Dirichlet distribution. In Dirichlet regression models, we also need goodness-of-fit tests of the Dirichlet or generalized Dirichlet distributions. The null hypothesis of interest for a goodness-of-fit test is that a given random variable or vector $X$ follows a specified beta or Dirichlet distribution $F(x)$. The parameters of the hypothesized distribution may be known or unknown.

If $X_1, \cdots, X_n$ is a random sample with common Dirichlet distribution function $F$, then we would like to test $F = F_0$ against $F \neq F_0$, where $F_0$ is a specified distribution with known or unknown parameters.

There are no goodness-of-fit tests or standardized goodness-of-fit tests developed for the general case of the Dirichlet distribution or generalized Dirichlet distribution in the literature. We propose five goodness-of-fit tests for the Dirichlet distribution. The first test, which is called energy test, is based on energy statistics. The second test, called distance covariance test, is based on the property of complete neutrality of the Dirichlet distribution and distance covariance tests. The third, fourth and fifth tests, called the triangle tests, are based on the interpoint distances. The power of the five proposed tests in simulation studies is presented. The simulation results show that
the distance covariance goodness-of-fit tests have best performance for the contaminated Dirichlet distribution. The distance covariance test also has good performance for the logistic normal distribution.

Mutual independence of a set of random variables or random vectors is a frequently used assumption in statistical inference. Therefore, testing mutual independence of a set of variables has very important applications. We propose a new method based on the distance covariance and permutation to test the mutual independence of all components of random vector in arbitrary dimensions. The energy, and triangle goodness-of-fit tests can also be used for the generalized Dirichlet distribution, which seems more widely used in data mining and machine learning because of its more flexible variance-covariance structure. For the principle of the parsimony, to determine whether the data can be better fitted by Dirichlet or generalized Dirichlet distribution, we apply the likelihood ratio test, Wald test, Rao’s score test to decide whether the generalized Dirichlet distribution is necessary. Some derivation about these three tests are also provided in the dissertation. The limiting distribution of the likelihood ratio, Wald or Rao’s test statistics is a chi-squared distribution, which is based on the large sample theory. However, no general theory provides a practical guidance about how large the sample size should be to apply the chi-squared approximation. Hence, we propose the parametric bootstrap likelihood ratio test, parametric bootstrap Wald test, and parametric bootstrap Rao’s score test for the small or medium sample size. For the generalized Dirichlet distribution, the derivation of the MLE method for the parameter estimation, the initial values of the iteration and the criteria of convergence for the Newton-Raphson algorithm has been proposed. The application of the dissertation includes goodness-of-fit tests of the Dirichlet and generalized Dirichlet distribution, model evaluation of the Dirichlet regression analysis, outlier and influential observation identification in the Dirichlet regression model. The generalized Dirichlet regression model is introduced. Copulas are widely used in finance, biology, insurance. The goodness-of-fit test of copulas is still a challenging problem. The copula goodness-of-fit tests based on the distance covariance test of mutual independence of all components of random vector in arbitrary dimensions and Rosenblatt transformation are also introduced.
CHAPTER 2  LITERATURE REVIEW

A historical overview of several important goodness-of-fit tests of univariate and multivariate distributions follows.

2.1 EDF Tests for Univariate Distribution

The empirical distribution function is the cumulative distribution function of the sample. This empirical distribution function is a piece-wise function that increases by $1/n$ at each of the observations when the sample size is $n$. Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed (iid) real random variables with the common cumulative distribution function $F(x)$, then the empirical distribution function is defined as

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} 1\{x_i \leq t\},$$

where $1\{}$ is the indicator function. The indicator $1\{X \leq x\}$ is a Bernoulli random variable with parameter $p = F(x)$, and $\hat{F}_n(x)$ is an unbiased estimator for $F(x)$. By the central limit theorem, the empirical $\hat{F}_n(x)$ has an asymptotically normal distribution with mean 0 and variance $F(x)(1 - F(x))$. That is,

$$\sqrt{n}(\hat{F}_n(x) - F(x)) \overset{d}{\to} \mathcal{N}(0, F(x)(1 - F(x))).$$

By the strong law of large numbers, the estimator $\hat{F}_n(x)$ converges to $F(x)$ almost surely, thus $\hat{F}_n(x)$ is a consistent estimator of $F(x)$. This establishes the pointwise convergence of the empirical distribution function to the true cdf. There is a stronger result, called the Glivenko Cantelli theorem, which states that the convergence happens uniformly over $x$. That is,

$$\|\hat{F}_n - F\|_\infty \equiv \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow{a.s.} 0.$$ (2.1)
The sup norm in (2.1) is the Kolmogorov-Smirnov statistic for testing the goodness-of-fit between the empirical distribution \( \hat{F}_n(x) \) and the assumed null hypothesis \( F(x) \). In (2.1), if we replace the sup norm with the \( L^2 \) norm,

\[
\| \hat{F}_n - F \|_a = \int \left| \hat{F}_n(x) - F(x) \right|^2 \, dx.
\]

Then the corresponding statistic is called the Cramer-von Mises statistic. Another norm may be reasonably used here instead of the sup norm or \( L^2 \) norm.

The statistic of the Kolmogorov-Smirnov test for a given null distribution function \( F(x) \) is

\[
D = \sup \left| \hat{F}_n(x) - F(x) \right|.
\]

If the observations come from the hypothesized null distribution \( F(X) \), the test statistic \( D \) converges to 0 almost surely as \( n \to \infty \). The critical value of Kolmogorov-Smirnov test comes from the Kolmogorov distribution. When \( K_n \) is greater than the critical value \( K_\alpha \), we reject the null hypothesis at the significance level \( \alpha \). In other words, we have strong evidence that the data do not come from the null distribution.

The statistic of the Cramer-von Mises test is

\[
C_n = \frac{1}{12n} + \sum_{i=1}^{n} \left[ \frac{2i - 1}{2n} - F(x_{(i)}) \right]^2,
\]

where \( X_{(i)} \) are the ordered observations.

If the test statistic \( C_n \) is greater than the critical value \( C_\alpha \), which can be obtained from the table of critical values for the Cramer-von Mises test, then we reject the null hypothesis at the level \( \alpha \). That is, we have enough evidence that the data do not come from the null distribution.

2.2 Chi-squared Test and Likelihood Ratio Tests

Suppose the multinomial distribution has \( k \) cells with unknown probabilities \( p_1, \ldots, p_k \). Let \( n_1, \ldots, n_k \) be a set of observed cell frequencies with \( n_1 + \cdots + n_k = n \). The null hypothesis is
given by
\[ H_0 : p_i = p^*_i, \quad 1 \leq i \leq k, \]
where \( p^*_i \) are a specified probability distribution on the \( k \) cells. The chi-squared goodness-of-fit statistic is defined as
\[ \chi^2 = \sum_{i=1}^{k} \frac{(n_i - e_i)^2}{e_i}, \]
where
\[ e_i = np^*_i, \quad 1 \leq i \leq k. \]

The null hypothesis is rejected for large \( \chi^2 \). The chi-squared goodness-of-fit test may also be applied to continuous distributions. The range of the observed data are grouped into mutually exclusive bins. Then \( n_i \) is the number of observations that fall into the \( i \)-th bin, and \( p_i \) is the probability that an observation will fall into the \( i \)-th bin under the null hypothesis. Based on the same null hypothesis, the likelihood ratio goodness-of-fit statistic is given by
\[ G^2 = 2L = 2 \sum_{i=1}^{k} n_i \log \left( \frac{n_i}{np_i} \right). \]

Fisher (1928) has proven that
\[ L = \sum_{i=1}^{k} \left( \frac{(n_i - np_i)^2}{2np_i} + \frac{(n_i - np_i)^3}{6np_i^2} + \cdots + \frac{(i^2 + i) (n_i - np_i)^{(i+1)}}{np_i^i} + \cdots \right). \]

The likelihood ratio test statistic has an asymptotic \( \chi^2_{k-1} \) distribution. The first term of \( L \) is half of the \( \chi^2 \), and therefore the likelihood ratio test statistic is asymptotically equivalent to chi-squared statistic.

Basically, we do not know the population parameter vector before we conduct the chi-squared test. We must first estimate the parameter vector of the null distribution, then chooses bins based on the estimated parameters. The degrees of freedom of the chi-square statistic is adjusted for the estimated parameters. Here it would be \( k \) (the number of bins) minus one minus the size of
the parameter vector. Both the chi-squared test and likelihood ratio test have relatively low power because we lose much information of the actual data when we bin the data. Chi-squared tests are generally less powerful than EDF tests (D’Agostino and Stephens [1986]).

2.3 Two-Dimensional Kolmogorov-Smirnov Goodness-of-Fit Tests

Peacock [1982] proposed a two-dimensional Kolmogorov-Smirnov goodness-of-fit test. For the two-dimensional distribution, given a specific observation \((x, y)\), four quadrants of the plane are defined by \((x < X, y < Y)\), \((x < X, y > Y)\), \((x > X, y < Y)\), and \((x > X, y > Y)\). The four quadrants are equally valid areas for the definition of the cumulative probability distribution. The algorithm is to consider each definition and adopt the largest of the four differences in empirical and theoretical cumulative distribution as the final statistics. This test computes in \(O(n^3)\) and the power is low. If we use his idea to develop a test for \(k\)-dimensional data, the test will compute in \(O(n^{k+1})\), and the computational efficiency is low.

2.4 Energy Tests

Energy statistics are based on Newton’s potential energy which is determined by the relative location of two bodies in a gravitational space. The potential energy is the function of the distance between two objects in a gravitational space. The longer the distance between the two objects, the bigger the potential energy. The potential energy of one object to itself is zero because the distance of one object to itself is zero. The energy statistics regard distributions or sample observations as the spatial objects. The energy distance between two distributions is zero if and only if the two distributions are identical, which is similar to the fact that the potential energy of one object to itself is zero.

If the observations play a symmetric role, then it makes sense to suppose that energy statistics are symmetric functions of distance between observations. Energy statistics are \(V\)-statistics or \(U\)-statistics based on the Euclidean distance. Let \(X_1, X_2, \ldots, X_n\) be a \(k\)-dimensional random sample from the distribution \(F\). The kernel function \(h: R^k \times R^k \to R\) is symmetric \(h(X, Y) = h(Y, X)\),
and the energy statistics are defined as

\[ V_n = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} h_k(X_i, X_j), \]

or

\[ U_n = \frac{1}{n(n - 1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} h_k(X_i, X_j). \]

In this dissertation, we consider the energy statistic as \( V \)-statistics. Consider the kernel

\[ h(x, y) = E|x - Y| + E|y - Y| - E|Y - Y'| - |x - y|, \]

where \(|x - y|\) is the Euclidean distance between sample observations and \( Y' \) is an independent, identically distributed copy of the random variable \( Y \) from the distribution \( F \). The kernel function \( h(x, y) \) is a function of distance \( \phi(z) \), where \( z = |x - y| \). Under the null hypothesis \( F = G \), the expected value of \( h(x, Y) \) is zero, so the kernel \( h \) is degenerate.

If \( F_n \) denotes the empirical cumulative distribution function (cdf) of \( X \), then the sample energy \( \mathcal{E} \) was introduced by Székely (2003) is

\[ \mathcal{E}_n^{(\alpha)}(F_n, G) = \left( \frac{2}{n} \sum_{i=1}^{n} E(|x_i - Y|^\alpha) - E|Y - Y'|^\alpha - \frac{1}{n^2} \sum_{i,j=1}^{n} |x_i - x_j|^\alpha \right), \]

where \( \alpha \) is in \((0, 2)\). The energy distance between two distributions \( F \) and \( G \) is defined by

\[ \mathcal{E}(F, G) = \varepsilon(X, Y) = 2E(|X - Y|) - E|X - X'| - E|Y - Y'|, \]

where \( X \) is a random variable with the distribution \( F \), and \( Y \) is a random variable with \( G \). Then \( \mathcal{E}(F, G) \) is nonnegative and equals zero if and only if \( F = G \), which matches the fact that the potential energy between two different objects are positive and the potential energy of an object to itself is zero. The goodness-of-fit test statistic for the null hypothesis that data come from \( Y \)
distribution based on energy statistics is

\[ Q_n = n \left( \frac{2}{n} \sum_{i=1}^{n} E(|x_i - Y|) - E |Y - Y'| - \frac{1}{n^2} \sum_{i,j=1}^{n} |x_i - x_j| \right). \]

Based on Euclidean distances between observations, the family of statistics and tests is indexed by
an exponent in \((0, 2)\) on Euclidean distances

\[ Q_n^{(\alpha)} = n \left( \frac{2}{n} \sum_{i=1}^{n} E(|x_i - Y|^{\alpha}) - E |Y - Y'|^{\alpha} - \frac{1}{n^2} \sum_{i,j=1}^{n} |x_i - x_j|^{\alpha} \right). \]

According to the theory of \(V\)-statistics, under the null hypothesis, if \(E(h^2(X, X'))\) is finite,
the limiting distribution of \(Q_n\) is a quadratic form \(Q = \sum \lambda_k Z_k^2\) of independent and identically
distributed standard normal random variables \(Z_k\), \(k = 1, 2, \ldots\). The nonnegative coefficients \(\lambda_k\)
are eigenvalues of the integral operator with kernel \(h(u_1, u_2)\), i.e.

\[ \int_{-\infty}^{+\infty} h(u_1, u_2) \psi(u_2) dF(u_2) = \lambda \psi(u_2), \]

where \(\psi\) is the eigenfunction. We reject the null hypothesis for the large values of \(Q_n\). Thus, it
is easy to find that \(E(Q_n) = \sum k \lambda_k\) and \(Var(Q_n) = 2 \sum k \lambda_k^2\). The energy tests are consistent
against general alternatives, and they are rotation and translation invariant omnibus tests.

2.5 Triangle Tests

The triangle goodness-of-fit test is a kind of test based on the geometry figure, triangle. The
length of each side of the triangle is measured by the interpoint distance. The triangle test was
proposed by [Bartoszynski and Lawrence (1997)]. The interpoint distance can be the Euclidean dis-
tance, the Manhattan distance, or other distances. The interpoint distance distribution of a specific
distribution \(G\) is defined by \(G^*(z) = P(\delta(X, Y) \leq z)\), where \(X\) and \(Y\) are independent variables
from the distribution \(G\) and \(\delta\) is any properly chosen distance function. [Maa and Bartoszynski (1996)]
posited that two distributions, \(F\) and \(G\), are identical if and only if two corresponding
interpoint distance variables $\delta(X_1, X_2)$ and $\delta(X_1, Y)$ have the identical interpoint distance distribution, where the random variables $X_1$ and $X_2$ follow the $F$ distribution and the random variable $Y$ follows $G$ distribution and three variables are mutually independent. Let $X_1, X_2, \ldots, X_n$ be a $k$-dimensional random sample from the distribution $F$. We form a random triangle from a pair of randomly selected data points $X_i$ and $X_j$ and a variable $Y$ from the hypothesized distribution $G$. According the theorem by [Maa and Bartoszynski (1996)], the null hypothesis $F = G$ holds if and only if the lengths of the three sides of this random triangle have the same interpoint distance distribution $G^*$. Under the null hypothesis, the side joining the two sample points is equally likely to be the shortest, middle, or the longest side in the random triangle. The triangle goodness-of-fit test is based on statistics which estimate the probability that the side joining the two sample points is the smallest, middle, or the largest side of the triangle. [Bartoszynski and Lawrence (1997)] proposed the following three kernel functions:

$$h_1(x, y) = P[Z\{\delta(x, y) < \min[\delta(x, Z), \delta(y, Z)]\}] + \frac{1}{2}P[Z\{\delta(x, y) = \delta(x, Z) < \delta(y, Z)\}] + \frac{1}{2}P[Z\{\delta(x, y) = \delta(y, Z) < \delta(x, Z)\}] + \frac{1}{3}P[Z\{\delta(x, y) = \delta(x, Z) = \delta(y, Z)\}],$$

$$h_3(x, y) = P[Z\{\delta(x, y) > \max[\delta(x, Z), \delta(y, Z)]\}] + \frac{1}{2}P[Z\{\delta(x, y) = \delta(x, Z) > \delta(y, Z)\}] + \frac{1}{2}P[Z\{\delta(x, y) = \delta(y, Z) > \delta(x, Z)\}] + \frac{1}{3}P[Z\{\delta(x, y) = \delta(x, Z) = \delta(y, Z)\}],$$

$$h_2(x, y) = 1 - h_1(x, y) - h_3(x, y),$$
where $x$ and $y$ are the sample points from the distribution $F$, and $Z$ is a random variable from the distribution $G$. For a continuous distribution, $h_1$, $h_2$, and $h_3$ denote the probabilities that the side joining the sample points $x$ and $y$ is the smallest, middle, or largest in the triangle with two vertices $x$, $y$ and the third vertex $Z$ from the distribution $G$. For a two-dimensional distribution, sample points $x$ and $y$ divide the two-dimensional data space into three different regions 1, 2, and 3. The functions $h_1$, $h_2$, and $h_3$ based on the Euclidean distance denote the probabilities of $Z$ falling into regions 1, 2, and 3 in Figure 2.1 where the probabilities on the boundaries have been split evenly among the adjacent regions.

Based on the above three kernel functions, Bartoszynski and Lawrence (1997) proposed the following three statistics that the triangle goodness-of-fit test is based on:

$$U_k = \frac{1}{\binom{n}{2}} \sum_{i<j} h_k (X_i, X_j) \quad for \quad k = 1, 2, 3.$$ 

These three statistics $U_1$, $U_2$, and $U_3$ estimate the chance that the side joining the two data points is the smallest, middle, or largest side of the triangle formed from two randomly selected data points and the hypothetical point. $U_k$ estimates the chance of a point falling into region $k$ in Figure 2.1. Thus, it is obvious that $U_1 + U_2 + U_3 = 1$. 

Figure 2.1: Regions defining $h_k$, $k = 1, 2, 3$
\[ Cov (U_r, U_s) = \frac{2(n-2)\theta_{rs}}{\binom{n}{2}} + \frac{\sigma_{rs}}{\binom{n}{2}} \text{ for } 1 \leq r, s \leq 3 \]

where,

\[ \theta_{rs} = cov [h_r (X_1, X_2), h_s (X_1, X_3)] , \]

and

\[ \sigma_{rs} = cov [h_r (X_1, X_2), h_s (X_1, X_2)] . \]

Under the null hypothesis and from the symmetry point of view, \( E(U_k) = \frac{1}{3} \) for \( k = 1, 2, 3 \). If \( h_1, h_2, \) and \( h_3 \) have finite variances, the asymptotic distribution of \( (U_r, U_s) \) is bivariate normal for \( r, s = 1, 2, 3 \) and \( r \neq s \) under the null distribution. We can use any of the statistics \( U_1, U_2, \) or \( U_3 \) individually or use a combination of these three statistics to give the most power against a specific family of alternatives. For example, one test statistic based on \( U_1 \) and \( U_3 \) can be given by:

\[ Q = \frac{Z_1^2 + Z_3^2 - 2\rho Z_1 Z_3}{1 - \rho^2}, \]

where \( Z_i = \frac{(U_i - \frac{1}{3})}{\sqrt{\text{var}(U_i)}}, \ i = 1, 2, \) and \( \rho = \text{cov}(Z_1, Z_2). \) Under the null hypothesis, \( Q \) has a chi-squared distribution with degrees of freedom 2.
CHAPTER 3 BETA AND DIRICHLET DISTRIBUTION

Beta and Dirichlet distributions are widely used in statistics. The beta distribution is frequently used as the prior probability distribution in Bayesian analysis. For example, the beta distribution is a good model for the prior distribution of a proportion.

3.1 Beta Distribution

**Definition 3.1.** A random variable $x$ follows a beta distribution with shape parameters $\alpha, \beta > 0$, if it has the pdf

$$f(x \mid \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 < x < 1,$$

where $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha, \beta)$ denotes the beta function. A random variable $X$ that is beta-distributed with shape parameters $\alpha, \beta$ is denoted by $X \sim \text{Beta}(\alpha, \beta)$.

**Definition 3.2.** The incomplete beta function is defined by

$$B(x \mid \alpha, \beta) = \int_0^x x^{\alpha-1}(1-x)^{\beta-1} dx.$$  

**Definition 3.3.** The regularized incomplete beta function is defined by

$$I(x \mid \alpha, \beta) = \int_0^x \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} dx.$$  

**Definition 3.4.** The conditional regularized incomplete beta function given $x_1, \cdots, x_{i-1}$ is defined by

$$I(x_i \mid \alpha, \beta, x_1, \cdots, x_{i-1}) = \int_0^{\frac{x_i}{1-x_1-\cdots-x_{i-1}}} x^{\alpha-1}(1-x)^{\beta-1} \frac{dx}{B(\alpha, \beta)}.$$
**Definition 3.5.** The hypergeometric function is defined for $|x| < 1$ by the power series

$$
{2F_1}(a, b; c; x) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{x^n}{n!}.
$$

**Definition 3.6.** $(a)_n$ is the Pochhammer symbol, which is defined by

$$(a)_n = \begin{cases} 
  1 & \text{if } n = 0; \\
  a(a + 1) \cdots (a + n - 1) & \text{if } n > 0.
\end{cases}$$

Incomplete beta functions $B(x; \alpha, \beta)$ is given by the hypergeometric function

$$
B(x; \alpha, \beta) = \frac{x^{\alpha}}{\alpha} {2F_1}(\alpha, 1 - \beta; \alpha + 1; x).
$$

It is also given by the power series

$$
B(x; \alpha, \beta) = x^{\alpha} \sum_{n=0}^{\infty} \frac{(1 - \beta)_n}{n!(\alpha + n)} x^n.
$$

For $x = 1$, the incomplete beta function coincides with the complete beta function. The beta function is symmetric, that is,

$$
B(\alpha, \beta) = B(\beta, \alpha).
$$

When $\alpha$ and $\beta$ are positive integers, it is related to factorial

$$
B(\alpha, \beta) = \frac{(\alpha - 1)! (\beta - 1)!}{(\alpha + \beta - 1)!}.
$$

When $\alpha$ and $\beta$ are positive numbers, it is related to the gamma function

$$
B(\alpha, \beta) = \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)}.
$$

**Definition 3.7.** The regularized incomplete beta function is defined by
When \(a\) and \(b\) are integers,

\[
I_x(a, b) = \sum_{i=a}^{a+b-1} \binom{a+b-1}{i} x^i (1-x)^{a+b-1-i}.
\]

The cumulative distribution function of beta distribution is

\[
F(x; \alpha, \beta) = \frac{B(x; \alpha, \beta)}{B(\alpha, \beta)} = I_x(\alpha, \beta),
\]

where \(B(x; \alpha, \beta)\) is the incomplete beta function, and \(I_x(\alpha, \beta)\) is the regularized incomplete beta function.

### 3.2 Stochastic Representation of Beta Distribution

If \(X\) is a random variable whose distribution is identical to the distribution of a function of other random variables \(Y_1, \ldots, Y_n\), the stochastic representation of \(X\) is defined as

\[
X \overset{d}{=} g(Y_1, \ldots, Y_n).
\]

In practice, the random variables \(Y_1, \ldots, Y_n\) are mutually independent and identically distributed, and the distribution of random variables \(Y_1, \ldots, Y_n\) are familiar to us and easier to be simulated. We can call the \(d\) operator the identical distribution operator. If \(X = Y\), then \(X \overset{d}{=} Y\), but the inverse may not be true. When the distribution of a random variable \(x\) or its properties are hard to derive or be simulated, the stochastic representation is a very powerful method.

**Definition 3.8.** A random variable \(x\) follows a gamma distribution with parameters \(\alpha, \beta > 0\), if it has the pdf

\[
f(x|\alpha, \beta) = \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)} \quad x > 0 \quad \alpha, \beta > 0,
\]
where \( \Gamma(\alpha) \) denotes the \( \Gamma \) function. A random variable \( X \) that is gamma-distributed with shape parameter \( \alpha \) and scale parameter \( \beta \) is denoted by \( X \sim \Gamma(\alpha, \beta) \).

Let \( X_1 \) and \( X_2 \) be independent random variables, each having a gamma distribution with \( \beta = 1 \). The joint pdf of the two variables can be written as

\[
f(x_1, x_2) = \begin{cases} \prod_{i=1}^{2} \frac{1}{\Gamma(\alpha_i)} x_i^{\alpha_i-1} e^{-x_i}, & 0 < x_i < \infty, \\ 0, & \text{otherwise}. \end{cases}
\]

The random variable \( Y \) is given by

\[
Y = \frac{X_1}{X_1 + X_2}.
\]

The random variable \( Y \) follows the beta distribution with shape parameters \( \alpha_1 \) and \( \alpha_2 \). The associated transformation maps the space

\[
\mathcal{A} = \{(x_1, x_2) : 0 < x_i < \infty, i = 1, 2\}
\]

onto the space

\[
\mathcal{B} = \{y : 0 < y < 1\}.
\]

In what follows, the stochastic representation \( Y \) of beta distribution by gamma variables is key.

3.3 Dirichlet Distribution

The Dirichlet distribution can be thought of as the multivariate beta distribution. The Dirichlet distribution can be fitted as the prior multinomial distribution in Bayesian analysis. The Dirichlet distribution is the conjugate prior for the probability vector from multinomial sampling.

**Definition 3.9.** A random variable \( X \) follows the Dirichlet distribution with parameters \( \alpha_1, \alpha_2, \ldots, \alpha_{k+1} > 0 \), if it has the pdf

\[
f(x_1, x_2, \ldots, x_k) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_{k+1})}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_{k+1})} x_1^{\alpha_1-1} \cdots x_k^{\alpha_k-1} (1 - x_1 - \cdots - x_k)^{\alpha_{k+1}-1},
\]
where $0 < x_i < 1$, $i = 1, \ldots, k$, and $x_1 + x_2 + \cdots + x_k < 1$.

The Dirichlet distribution with parameters $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_{k+1})^T$ and probability density function is denoted by $\text{Dirichlet}(\alpha_1, \alpha_2, \ldots, \alpha_{k+1})$ or $\text{Dir}(\alpha_1, \alpha_2, \ldots, \alpha_{k+1})$. When $k = 1$, the Dirichlet distribution becomes a beta distribution. The marginal distributions of Dirichlet distribution are each a beta distribution. Suppose $X$ has a Dirichlet$(\alpha_1, \alpha_2, \ldots, \alpha_{k+1})$ distribution and $X_1$ is any subvector of $X$, then $X_1$ has a Dirichlet distribution too. For example, for a given $k$-dimensional random vector $(X_1, X_2, \ldots, X_k)$ from the Dirichlet$(\alpha_1, \alpha_2, \ldots, \alpha_{k+1})$ distribution, the first component of the vector, $X_1$, follows Beta$(\alpha_1, \sum_{i=2}^{k+1} \alpha_i)$ distribution. The third component of the vector, $X_3$, follows Beta$(\alpha_3, \sum_{i=1}^{k+1} \alpha_i - \alpha_3)$ distribution. The joint distribution of the subvector $(X_1, X_3)$ follows the Dirichlet$(\alpha_1, \alpha_3, \sum_{i=1}^{k+1} \alpha_i - \alpha_1 - \alpha_3)$.

3.4 Stochastic Representation of Dirichlet Distribution

Let $X_1, X_2, \ldots, X_{k+1}$ be independent random variables, each having a gamma distribution with $\beta = 1$. The joint pdf of these variables is written as

$$f(x_1, x_2, \ldots, x_{k+1}) = \begin{cases} \prod_{i=1}^{k+1} \frac{1}{\Gamma(\alpha_i)} x_i^{\alpha_i-1} e^{-x_i}, & 0 < x_i < \infty, \\ 0, & \text{otherwise.} \end{cases}$$

Let

$$Y_i = \frac{X_1}{X_1 + X_2 + \cdots + X_{k+1}}, \quad i = 1, 2, \ldots, k.$$ 

The random variables $Y_1, Y_2, \ldots, Y_k$ have a Dirichlet pdf

$$f(y_1, \ldots, y_k) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_{k+1})}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_{k+1})} y_1^{\alpha_1-1} \cdots y_k^{\alpha_k-1} (1 - y_1 - \cdots - y_k)^{\alpha_{k+1}-1},$$

where $0 < y_i, i = 1, \ldots, k, y_1 + y_2 + \cdots + y_k < 1$. Therefore, the random variables $Y_1, Y_2, \ldots, Y_k$ follows a Dirichlet distribution. The associated transformation maps the space

$$\mathcal{A} = \{(x_1, \ldots, x_{k+1}) : 0 < x_i < \infty, \ i = 1, \ldots, k+1\}.$$
onto the space

\[ \mathcal{B} = \{(y_1, \ldots, y_k) : 0 < y_i, \ i = 1, 2, \ldots, k, \ y_1 + \cdots + y_k < 1\}. \]

The R gtools package by [Warners, Bolker, and Lumley (2014)](https://www.r-project.org) provides the rdirichlet function to generate the random samples following a specific Dirichlet distribution.

3.5 Estimation for the Beta and Dirichlet Parameter Vector

For the energy and triangle goodness-of-fit tests, we need to estimate the parameter vector. We have \( n \) independent and identically distributed variables from the Beta(\( \alpha, \beta \)) or Dir(\( \alpha_1, \ldots, \alpha_{k+1} \)) distribution. For \( i = 1, \ldots, n \), let \( x_i = (x_{i1}, \ldots, x_{i(k+1)}) \) be \( n \) observations.

3.5.1 MLE Based on the Newton-Raphson Algorithm

The log-likelihood function of the parameter vector \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_{k+1}) \) for the \( n \) observations is

\[
l(\alpha) = m \left\{ \log \Gamma \left( \sum_{i=1}^{k+1} \alpha_i \right) - \sum_{i=1}^{k+1} \log \Gamma (\alpha_i) + \sum_{j=1}^{k+1} (\alpha_j - 1) \log \prod_{i=1}^{n} x_{ij} \right\}.
\]

The Newton-Raphson method is used in the following iteration to calculate the MLE of the parameter vector (Narayanan, 1990):

\[
\begin{pmatrix}
\hat{\alpha}_1 \\
\vdots \\
\hat{\alpha}_{k+1}
\end{pmatrix}_i = \begin{pmatrix}
\hat{\alpha}_1 \\
\vdots \\
\hat{\alpha}_{k+1}
\end{pmatrix}_{i-1} + \begin{pmatrix}
\text{var}(\hat{\alpha}_1) & \ldots & \text{cov}(\hat{\alpha}_1, \hat{\alpha}_{k+1}) \\
\vdots & \ddots & \ddots \\
\text{cov}(\hat{\alpha}_{k+1}, \hat{\alpha}_1) & \ldots & \text{var}(\hat{\alpha}_{k+1})
\end{pmatrix}_{(i-1)} \begin{pmatrix}
g_1(\hat{\alpha}) \\
\vdots \\
g_{k+1}(\hat{\alpha})
\end{pmatrix}_{(i-1)}
\]
where $\tilde{\alpha}_{(0)} = (\tilde{\alpha}_{1(0)}, \ldots, \tilde{\alpha}_{k(0)})'$ are the initial estimates. The gradient vector is

$$
g = \nabla l(\alpha) = \begin{pmatrix} \frac{\partial \log \Gamma(\sum_{i=1}^{k+1} \alpha_i)}{\partial \alpha_1} - \frac{\partial \log \Gamma(\alpha_1)}{\partial \alpha_1} + \log \prod_{j=1}^{n} x_{1j} \\ \frac{\partial \log \Gamma(\sum_{i=1}^{k+1} \alpha_i)}{\partial \alpha_2} - \frac{\partial \log \Gamma(\alpha_2)}{\partial \alpha_2} + \log \prod_{j=1}^{n} x_{2j} \\ \vdots \\ \frac{\partial \log \Gamma(\sum_{i=1}^{k+1} \alpha_i)}{\partial \alpha_{(k+1)}} - \frac{\partial \log \Gamma(\alpha_{(k+1)})}{\partial \alpha_{(k+1)}} + \log \prod_{j=1}^{n} x_{(k+1)j} \end{pmatrix}.
$$

The Hessian matrix is

$$
H = \nabla^2 l(\alpha) = \begin{pmatrix} \frac{\partial^2 \log \Gamma(\sum_{i=1}^{k+1} \alpha_i)}{\partial \alpha_1^2} - \frac{\partial^2 \log \Gamma(\alpha_1)}{\partial \alpha_1^2} & & \frac{\partial^2 \log \Gamma(\sum_{i=1}^{k+1} \alpha_i)}{\partial \alpha_1 \partial \alpha_2} - \frac{\partial^2 \log \Gamma(\alpha_2)}{\partial \alpha_1 \partial \alpha_2} & & \cdots & & \frac{\partial^2 \log \Gamma(\sum_{i=1}^{k+1} \alpha_i)}{\partial \alpha_1 \partial \alpha_{(k+1)}} - \frac{\partial^2 \log \Gamma(\alpha_{(k+1)})}{\partial \alpha_1 \partial \alpha_{(k+1)}} \\
\frac{\partial^2 \log \Gamma(\sum_{i=1}^{k+1} \alpha_i)}{\partial \alpha_2^2} - \frac{\partial^2 \log \Gamma(\alpha_2)}{\partial \alpha_2^2} & & \frac{\partial^2 \log \Gamma(\sum_{i=1}^{k+1} \alpha_i)}{\partial \alpha_2 \partial \alpha_1} - \frac{\partial^2 \log \Gamma(\alpha_1)}{\partial \alpha_2 \partial \alpha_1} & & \cdots & & \frac{\partial^2 \log \Gamma(\sum_{i=1}^{k+1} \alpha_i)}{\partial \alpha_2 \partial \alpha_{(k+1)}} - \frac{\partial^2 \log \Gamma(\alpha_{(k+1)})}{\partial \alpha_2 \partial \alpha_{(k+1)}} \\
\vdots & & \vdots & & \ddots & & \vdots \\
\frac{\partial^2 \log \Gamma(\sum_{i=1}^{k+1} \alpha_i)}{\partial \alpha_{(k+1)} \partial \alpha_1} - \frac{\partial^2 \log \Gamma(\alpha_{(k+1)})}{\partial \alpha_{(k+1)} \partial \alpha_1} & & \frac{\partial^2 \log \Gamma(\sum_{i=1}^{k+1} \alpha_i)}{\partial \alpha_{(k+1)} \partial \alpha_2} - \frac{\partial^2 \log \Gamma(\alpha_{(k+1)})}{\partial \alpha_{(k+1)} \partial \alpha_2} & & \cdots & & \frac{\partial^2 \log \Gamma(\sum_{i=1}^{k+1} \alpha_i)}{\partial \alpha_{(k+1)} \partial \alpha_{(k+1)}} - \frac{\partial^2 \log \Gamma(\alpha_{(k+1)})}{\partial \alpha_{(k+1)} \partial \alpha_{(k+1)}} \end{pmatrix}.
$$

The $j$-th component of the gradient vector is

$$
g_j(\alpha) = n \Psi\left(\sum_{m=1}^{k+1} \alpha_m\right) - n \Psi(\alpha_j) + n \log G_{j} \quad j = 1, 2, \ldots, k + 1,
$$

where $\Psi$ is the digamma function. The variance-covariance matrix $V$ which is the inverse of the Fisher information matrix is found as

$$
V = D + \beta a a',
$$

where $D = \text{diag}(1/n\Psi'(\alpha_1), \ldots, 1/n\Psi'(\alpha_{k+1}))$, $a' = [1/\Psi'(\alpha_1), \ldots, 1/\Psi'(\alpha_{k+1})]$ and

$$
\beta = n \Psi\left(\sum_{j=1}^{k+1} \alpha_j\right)\left(1 - \Psi\left(\sum_{j=1}^{k+1} \alpha_j\right)\sum_{j=1}^{k+1} \frac{1}{\Psi'(\alpha_j)}\right)^{-1},
$$
where $\Psi'$ is the trigamma function. One criteria for convergence of this algorithm is developed by Choi and Wette (1967). The statistic $S$ for the convergence is given by

$$S = (g_1(\hat{\alpha}), \ldots, g_{k+1}(\hat{\alpha})) \left( \begin{array}{cccc} \text{var}(\hat{\alpha}_1) & \cdots & \text{cov}(\hat{\alpha}_1, \hat{\alpha}_{k+1}) \\ \vdots & \ddots & \vdots \\ \text{cov}(\hat{\alpha}_{k+1}, \hat{\alpha}_1) & \cdots & \text{var}(\hat{\alpha}_{k+1}) \end{array} \right) \left( \begin{array}{c} g_1(\hat{\alpha}) \\ \vdots \\ g_{k+1}(\hat{\alpha}) \end{array} \right).$$

The iteration can be continued until $S$ becomes less than $\chi^2_{k+1}(c)$ for a fixed real number $c$ in the lower tail or the number of iterations required exceeds the maximum number of iterations we set. In the computer program of this dissertation, we set $c = 0.0001$ and maximum 100. When the initial estimates are calculated by the method of moments, some components of the parameter vector $\alpha$ are negative numbers during the iterations. Thus the initial estimates are set to be the $\min\{x_{ij}\}$, which makes the parameter estimates positive during the iteration (Ronning, 1986).

3.5.2 Method of Moments Estimates

The method of moments for the estimates of the parameter vector is given by

$$\hat{\alpha}_j = \frac{(x'_{11} - x'_{21})x'_{1j}}{x'_{21} - x'_{11}}, \quad j = 1, 2, \ldots, k+1,$$

where

$$x'_{1j} = \frac{1}{n} \sum_{i=1}^{n} x_{ij}, \quad j = 1, 2, \ldots, k+1,$$

and

$$x'_{21} = \frac{1}{n} \sum_{i=1}^{n} x_{i1}^2.$$

In the MLE algorithm, we need to provide initial values for the iteration, and the initial values can be the parameter estimates based on the method of moments.
CHAPTER 4 ENERGY AND TRIANGLE TESTS OF BETA AND DIRICHLET DISTRIBUTION

4.1 Univariate Energy Goodness-of-Fit Statistics

We want to test if a given data set is from the Dirichlet family. The goodness-of-fit test statistic for the null hypothesis based on energy distance is

$$Q_n = n \left( \frac{2}{n} \sum_{i=1}^{n} E(|x_i - Y|) - E|Y - Y'| - \frac{1}{n^2} \sum_{i,j=1}^{n} |x_i - x_j| \right).$$

Based on Euclidean distances between observations, the family of statistics and tests is indexed by an exponent in $(0, 2)$ on Euclidean distances, which we denote by $\alpha$. The corresponding statistics are

$$Q_n^{(\alpha)} = n \left( \frac{2}{n} \sum_{i=1}^{n} E(|x_i - Y|^\alpha) - E|Y - Y'|^\alpha - \frac{1}{n^2} \sum_{i,j=1}^{n} |x_i - x_j|^\alpha \right), \quad 0 < \alpha < 2.$$  

Energy goodness-of-fit tests are derived for the beta distribution in this section. The third component of the energy statistic $Q_n$ is $\sum_{i,j=1}^{n} |x_i - x_j|$, which can be simplified into a linear combination of the ordered sample. The linear combination will reduce the calculation of $\sum_{i,j=1}^{n} |x_i - x_j|$ from $n^2$ terms to $n$ terms, and therefore it reduces the computation time. Suppose $X_1, X_2, \ldots, X_n$ is a univariate random sample, and the corresponding ordered sample is $X_{(1)}, \ldots, X_{(n)}$.

Then (Rizzo 2002)

$$\sum_{i,j=1}^{n} |x_i - x_j| = 2 \sum_{k=1}^{n} \left( (2k - 1) - n \right) X_{(k)}.$$
Proof.

\[
\sum_{i,j=1}^{n} |x_i - x_j| = 2 \sum_{i,j=1}^{n} (x_{(i)} - x_{(j)}) \\
= 2 \left[ (n-1)X_{(n)} + (n-2)X_{(n-1)} + (n-3)X_{(n-2)} + \cdots + (1)X_{(2)} \right] \\
- 2 \left[ (n-1)X_{(1)} + (n-2)X_{(2)} + (n-3)X_{(3)} + \cdots + (1)X_{(n)} \right] \\
= 2 \left[ (1-n)X_{(1)} + (3-n)X_{(2)} + (5-n)X_{(3)} + \cdots + (n-1)X_{(n)} \right] \\
= 2 \sum_{k=1}^{n} ((2k-1) - n) X_{(k)}. 
\]

4.2 Energy Goodness-of-Fit Test for Uniform Distribution

For energy and triangle goodness-of-fit tests, we transform arbitrary beta or Dirichlet distributions into standard beta or Dirichlet distributions. We regard Beta(1,1) as the standard beta distribution, which is the same as uniform distribution with the parameters 0 and 1.

**Proposition 4.1.** If \( Y \) is distributed as Beta(1,1), then

\[
E|x - Y| = x^2 - x + 1/2.
\]

**Proof.**

\[
E|x - Y| = \int_{0}^{x} (x - y)dy + \int_{x}^{1} (y - x)dy \\
= x^2 - x + 1/2.
\]

**Proposition 4.2.** If \( X \) and \( Y \) is independent and distributed as Beta(1,1), then

\[
E|X - Y| = 1/3.
\]
Proof.

\begin{align*}
E|X - Y| &= E(E(|X - Y| | x)) \\
&= E \left( \int_0^x (x - y)dy + \int_x^1 (y - x)dy \right) \\
&= E \left( x^2 - x + 1/2 \right) \\
&= 1/3.
\end{align*}

\[\Box\]

If \( X_1, \ldots, X_n \) is a random sample, then the energy goodness-of-fit test statistic for testing \( H_0 : X \sim Beta(1, 1) \) is given by

\[ Q_n = n \left( \frac{2}{n} \sum_{i=1}^n \left( X_i^2 - X_i + \frac{1}{2} \right) - \frac{1}{3} - \frac{2}{n^2} \sum_{i=1}^n (2k - 1 - n) X_{(i)} \right). \]

Since the cumulative distribution function of the beta distribution is the regularized incomplete beta function, when we apply the regularized incomplete beta function \( I(x) \) as the transformation function, Beta\( (\alpha, \beta) \) can be transformed into the standard beta distribution. If \( X_1, \ldots, X_n \) is a random sample from the beta distribution, \( X_{(1)}, \ldots, X_{(n)} \) is the ordered sample. Let \( Z_{(1)}, \ldots, Z_{(n)} \) be the transformed sample, where \( Z_{(i)} = I(X_{(i)}), i = 1, \ldots, n \). The goodness-of-fit test statistic for testing the null hypothesis \( H_0 : X \sim Beta(\alpha, \beta) \) is given by

\[ Q_n = n \left( \frac{2}{n} \sum_{i=1}^n \left( Z_{(i)}^2 - Z_{(i)} + \frac{1}{2} \right) - \frac{1}{3} - \frac{2}{n^2} \sum_{i=1}^n (2k - 1 - n) Z_{(i)} \right). \]

When we would like to use the actual sample without any transformation to construct the goodness-of-fit test, and when we would like to apply the family of statistics and tests which is indexed by an exponent in \((0, 2)\) on Euclidean distances, we need to have the following four propositions.
Proposition 4.3. If \( Y \) is distributed as \( \text{Beta}(\alpha, \beta) \), where \( \alpha > 0, \beta > 0 \) are fixed, then

\[
E|a - X| = \frac{a^{\alpha+1}}{B(\alpha, \beta)} \sum_{n=0}^{\infty} \frac{a^n (1 - \beta)_n}{n!(\alpha + n)(\alpha + n + 1)} + \frac{(1 - a)^\beta}{B(\alpha, \beta)} \sum_{n=0}^{\infty} \frac{(1 - a)^n (-\alpha)_n - a(1 - \alpha)_n}{n!(\beta + n)}
\]

where \( (a)_n = a(a+1)...(a+n-1) \) is the ascending factorial.

Proposition 4.4. If \( Y \) is distributed as \( \text{Beta}(\alpha, \beta) \), where \( \alpha > 0, \beta > 0 \) are fixed, and \( 0 < k < 2 \), then

\[
E|a - X|^k = \sum_{n=0}^{\infty} \sum_{i=1}^{n} \binom{n}{i} (-k)_{n-i} [\frac{(1 - \beta)_i a^{k+i} + \frac{(1 - \alpha)_i (1 - a)^{k+i}}{\beta + n}] B(\alpha, \beta) n! \alpha + n + (1 - \alpha)_i (1 - a)^{k+i} \beta + n].
\]

Proof.

\[
E|a - x|^k = \frac{1}{B(\alpha, \beta)} \int_0^1 (a - x)^k x^{\alpha-1}(1-x)^{\beta-1} dx = \frac{1}{B(\alpha, \beta)} \int_0^a (a - x)^k x^{\alpha-1}(1-x)^{\beta-1} dx + \frac{1}{B(\alpha, \beta)} \int_a^1 (x - a)^k x^{\alpha-1}(1-x)^{\beta-1} dx.
\]
Let \( f(x) \) be \((a - x)^k(1 - x)^{\beta-1}\), so that
\[
\frac{1}{B(\alpha, \beta)} \int_0^a (a - x)^k x^{\alpha-1} (1 - x)^{\beta-1} \, dx
\]
\[
= \frac{1}{B(\alpha, \beta)} \sum_{n=0}^{\infty} f^{(n)}(0) \frac{a^{\alpha+n}}{(\alpha + n)n!}
\]
\[
= \frac{1}{B(\alpha, \beta)} \left( f(0) \frac{a^\alpha}{\alpha} + f'(0) \frac{a^{\alpha+1}}{\alpha + 1} + \cdots + \frac{f^{(n)}(0) a^{\alpha+n}}{(\alpha + n)!} + \cdots \right)
\]
\[
= \frac{1}{B(\alpha, \beta)} \sum_{n=0}^{\infty} \sum_{i=0}^{n} C_i^k (-k)_{n-i} (1 - \beta)_i a^{k-n+i} \frac{a^{\alpha+n}}{(\alpha + n)n!}
\]
\[
= \frac{1}{B(\alpha, \beta)} \sum_{n=0}^{\infty} \sum_{i=0}^{n} C_i^k (-k)_{n-i} (1 - \beta)_i a^{k+n+i} \frac{a^{\alpha+n}}{(\alpha + n)n!}
\]

Let \( g(x) \) be \((1 - x - a)^k(1 - x)^{\alpha-1}\). We replace \( g(x) \) in the above integral with its Maclaurin Series at \( x = 0 \).

\[
\frac{1}{B(\alpha, \beta)} \int_0^1 (x - a)^k x^{\alpha-1} (1 - x)^{\beta-1} \, dx
\]
\[
= \frac{1}{B(\alpha, \beta)} \int_0^{1-a} x^{\beta-1} \sum_{n=0}^{\infty} \sum_{i=0}^{\infty} g^{(n)}(0) \frac{x^n}{n!} \, dx
\]
\[
= \frac{1}{B(\alpha, \beta)} \sum_{n=0}^{\infty} \sum_{i=0}^{n} C_i^k (-k)_{n-i} (1 - \alpha)_i (1 - a)^{k-n+i} \frac{1 - a^{n+\beta}}{n + \beta},
\]

where \( g^{(n)}(0) \) denotes the \( n \)-th derivative of \( g \) at \( y = 0 \).

After simplification, we have
\[
E|a - X|^k = \sum_{n=0}^{\infty} \sum_{i=1}^{n} \frac{C_i^k (1 - \beta)_i a^{k+n+i}}{B(\alpha, \beta)n!} \left[ \frac{(1 - \beta)_i a^{\alpha+n+i}}{\alpha + n} + \frac{(1 - \alpha)_i (1 - a)^{k+\beta+i}}{\beta + n} \right].
\]

\( \square \)

**Proposition 4.5.** If \( X \) and \( Y \) are independent and identically distributed as \( Beta(\alpha, \beta) \), where
\( \alpha > 0, \beta > 0 \) are fixed, then

\[
E|X - Y| = \frac{2}{B^2(\alpha, \beta)} \sum_{n=0}^{\infty} \frac{(1 - \beta)_n B(2\alpha + n + 1, \beta)}{(\alpha + n)_2!},
\]

where \((a)_n = a(a+1)\ldots(a+n-1)\) is the ascending factorial, and \(B(\alpha, \beta)\) denotes the complete beta function.

**Proposition 4.6.** If \(X\) and \(Y\) is independent and identically distributed as Beta\((\alpha, \beta)\), and \(\alpha, \beta\) are fixed, and \(0 < k < 2\), then

\[
E|X - Y|^k = \frac{2}{B^2(\alpha, \beta)} \sum_{n=0}^{\infty} \sum_{i=0}^{n} \frac{\binom{n}{i} (-k)_{n-i}(1 - \beta)_i B(2\alpha + k + i, \beta)}{(\alpha + n)_n!(\alpha + n)n!}.
\]

**Proof.**

\[
E|X - Y|^k = \int_0^1 \int_0^1 |x - y|^k x^{\alpha-1}(1-x)^{\beta-1} y^{\alpha-1}(1-y)^{\beta-1} \frac{B(\alpha, \beta)}{B(\alpha, \beta)} \frac{B(\alpha, \beta)}{B(\alpha, \beta)} dx dy
\]

\[
= \frac{2}{B^2(\alpha, \beta)} \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx \int_0^x (x-y)^k y^{\alpha-1}(1-y)^{\beta-1} dy
\]

Let \(f(y) = (a - y)^k(1-y)^{\beta-1}\). We use the Maclaurin Series of \(f(y)\) at \(y = 0\) to replace \(f(y)\) in the above integral, so that

\[
E|X - Y|^k = \frac{2}{B^2(\alpha, \beta)} \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} \sum_{n=0}^{\infty} f^n(0) \frac{x^{\alpha+n}}{\alpha + n} dx
\]

\[
= \frac{2}{B^2(\alpha, \beta)} \sum_{n=0}^{\infty} \int_0^1 \frac{x^{2\alpha+n+1}}{(\alpha + n)!} (1-x)^{\beta-1} f^{(n)}(0),
\]

where \(f^{(n)}(0)\) denotes the \(n\)-th derivative of \(f\) at \(y = 0\).
After simplification, we have

\[ E|X - Y|^k = \frac{2}{B^2(\alpha, \beta)} \sum_{n=0}^{\infty} \sum_{i=0}^{n} \binom{n}{i} (-k)^{n-i} (1 - \beta)^i B(2\alpha + k + i, \beta) \frac{(\alpha + n)n!}{(\alpha + n)!}. \]

A random variable \( X \) follows the Dirichlet distribution with parameters \( \alpha_1, \alpha_2, \ldots, \alpha_{k+1} > 0 \), if it has the pdf

\[ f(x_1, x_2, \ldots, x_k) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_{k+1})}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_{k+1})} x_1^{\alpha_1-1} \cdots x_k^{\alpha_k-1} (1 - x_1 - \cdots - x_k)^{\alpha_{k+1}-1}, \]

where \( 0 < x_i < 1, i = 1, \ldots, k \), and \( 0 < x_1 + x_2 + \cdots + x_k < 1 \). If all parameters of a Dirichlet distribution are one, the Dirichlet distribution is called the standard Dirichlet distribution.

For a two-dimensional Dirichlet distribution, the support is a right triangle with vertices (0,0), (0,1), and (1,0). For three-dimensional Dirichlet distributions, the support is a pyramid with vertices (0, 0, 0), (1, 0, 0), (0, 1, 0), and (0, 0, 1). We can regard the support of a higher-dimensional Dirichlet distribution as the superpyramid. For the goodness-of-fit test for the Dirichlet distribution, translation and rotation will change the support, which is bounded and fixed in the area of \( 0 < x_i < 1, i = 1, \ldots, k \), and \( 0 < x_1 + x_2 + \cdots + x_k < 1 \). Therefore, we do not need to consider the affine invariant properties of goodness-of-fit tests for the Dirichlet distribution, even though affine invariant properties are very important for goodness-of-fit tests for normal distributions. However, we still need one specific kind of affine transformation to change arbitrary \( k \)-dimensional Dirichlet distributions to the corresponding standard Dirichlet distribution. We calculate our test statistics of the energy or triangle test based on the transformed standard Dirichlet distribution, which makes both tests easier to use and improves their efficiency.

4.3 Data Transformation and Standardization in the Dirichlet Distribution

For a random variable \( X \) which follows the multivariate normal distribution \( N(\mu, \Sigma) \), we can use the transformation \( (X - \mu)^T \Sigma^{-1} (X - \mu) \) to change the multivariate normal random variable into
the corresponding multivariate standard normal variable. However, for the Dirichlet distribution, we do not have such a simple formula of transformation. We can use a two-step transformation technique to transform the \(k\)-dimensional Dirichlet distribution into the \(k\)-dimensional standard Dirichlet distribution.

**Theorem 4.7.** Suppose that the random vector \((X_1, \ldots, X_n)\) is from an \(n\)-dimensional Dirichlet distribution with the parameter vector \((\alpha_1, \ldots, \alpha_{n+1})\). Then \((X_1^*, \ldots, X_n^*)\) defined by the following two-step transformation is jointly distributed as the standard \(n\)-dimensional Dirichlet distribution. The two-step transformation is defined as follows. The first step of transformation is given by

\[
Y_1 = I \left( X_1 | \alpha_1, \sum_{i=2}^{n+1} \alpha_i \right),
\]

\[
Y_2 = I \left( \frac{X_2}{1 - X_1} | \alpha_2, \sum_{i=3}^{n+1} \alpha_i, x_1 \right),
\]

\[
Y_3 = I \left( \frac{X_3}{1 - X_1 - X_2} | \alpha_3, \sum_{i=4}^{n+1} \alpha_i, x_1, x_2 \right),
\]

\[\vdots\]

\[
Y_i = I \left( \frac{X_i}{1 - X_1 - X_2 - \cdots - X_{i-1}} | \alpha_i, \sum_{j=i+1}^{n+1} \alpha_i, x_1, \cdots, x_{i-1} \right),
\]

\[\vdots\]

\[
Y_n = I \left( \frac{X_n}{1 - X_1 - \cdots - X_{n-1}} | \alpha_n, \alpha_{n+1}, x_1, \cdots, x_{n-1} \right),
\]

where \(I \left( \frac{X_i}{1 - X_1 - X_2 - \cdots - X_{i-1}} | X_1, \cdots, X_{i-1} \right)\), function is the conditional regularized incomplete be-
A function of $X_i$ given $X_1, \cdots, X_{i-1}$. The second step of transformation is given by

$$
X_1^* = 1 - (1 - Y_1)^{\frac{1}{n}},
$$
$$
X_2^* = (1 - Y_1)^{\frac{1}{n}} \left[ 1 - (1 - Y_2)^{\frac{1}{n-1}} \right],
$$
$$
X_3^* = (1 - Y_1)^{\frac{1}{n}} (1 - Y_2)^{\frac{1}{n-1}} \left[ 1 - (1 - Y_3)^{\frac{1}{n-2}} \right],
$$
$$
\ldots \ldots
$$
$$
X_i^* = \prod_{j=1}^{i-1} (1 - Y_j)^{\frac{1}{n-1+i}} \left[ 1 - (1 - Y_i)^{\frac{1}{n-i+1}} \right],
$$
$$
\ldots \ldots
$$
$$
X_n^* = \prod_{i=1}^{n-1} (1 - Y_i)^{\frac{1}{n-i+1}} Y_n.
$$

The random vector $(X_1^*, \ldots, X_n^*)$ follows the standard $n$-dimensional Dirichlet distribution.

**Proof.** The first step is that we change the $k$-dimensional Dirichlet distribution into the $k$-dimensional uniform distribution with the support $0 < x_i < 1$, $i = 1, \ldots, k$, which can be called the supercube uniform distribution because for the two-dimensional support, it is a square with the unity side and for the 3-dimensional support, it is a cube with the unity side. The second step is that we change the $k$-dimensional supercube uniform distribution into the $k$-dimensional standard Dirichlet distribution. During the first step, the non-uniform density is transformed into the uniform density which is what we desire, but the support is transformed from superpyramid into supercube which is not what we desire. The second step is to keep the uniform density but change the support from supercube back into superpyramid again.

The Jacobian determinant is the determinant of the Jacobian matrix which is given by

$$
\begin{vmatrix}
\frac{\partial x_1^*}{\partial x_1} & \frac{\partial x_1^*}{\partial x_2} & \cdots & \frac{\partial x_1^*}{\partial x_n} \\
\frac{\partial x_2^*}{\partial x_1} & \frac{\partial x_2^*}{\partial x_2} & \cdots & \frac{\partial x_2^*}{\partial x_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial x_n^*}{\partial x_1} & \frac{\partial x_n^*}{\partial x_2} & \cdots & \frac{\partial x_n^*}{\partial x_n}
\end{vmatrix}.
$$
However, it is hard to directly derive the each element of the Jacobian matrix. On the other hand, the Jacobian determinant is the inverse of the determinant of the matrix which is given by

\[
\begin{bmatrix}
\frac{\partial x^*_1}{\partial x_1} & \frac{\partial x^*_1}{\partial x_2} & \cdots & \frac{\partial x^*_1}{\partial x_n} \\
\frac{\partial x^*_2}{\partial x_1} & \frac{\partial x^*_2}{\partial x_2} & \cdots & \frac{\partial x^*_2}{\partial x_n} \\
\cdots & \cdots & \cdots & \cdots \\
\frac{\partial x^*_n}{\partial x_1} & \frac{\partial x^*_n}{\partial x_2} & \cdots & \frac{\partial x^*_n}{\partial x_n}
\end{bmatrix}.
\]

We call the above matrix the inverse Jacobian matrix. Since \( X^*_i \) is a function of \( X_1, X_2, \ldots \), and \( X_i, \frac{\partial x^*_i}{\partial x_j} \) are all zero when the subscript \( j > i \), so the inverse Jacobian matrix becomes

\[
\begin{bmatrix}
\frac{\partial x^*_1}{\partial x_1} & 0 & \cdots & 0 \\
\frac{\partial x^*_2}{\partial x_1} & \frac{\partial x^*_2}{\partial x_2} & 0 & 0 \\
\cdots & \cdots & \cdots & 0 \\
\frac{\partial x^*_n}{\partial x_1} & \frac{\partial x^*_n}{\partial x_2} & \cdots & \frac{\partial x^*_n}{\partial x_n}
\end{bmatrix}.
\]

Therefore, we only need to calculate the diagonal element of the inverse Jacobian matrix to obtain the determinant of the inverse Jacobian matrix. The \( i \)-th term of the diagonal elements is given by

\[
\frac{\partial x^*_i}{\partial x_i} = \frac{\partial x^*_i}{\partial y_1} \frac{\partial y_1}{\partial x_i} + \frac{\partial x^*_i}{\partial y_2} \frac{\partial y_2}{\partial x_i} + \cdots + \frac{\partial x^*_i}{\partial y_n} \frac{\partial y_n}{\partial x_i}, \quad i = 1, 2, \ldots, n.
\]

Since \( Y_i \) is a function of \( X_1, X_2, \ldots \), and \( X_i, \frac{\partial y}{\partial x_j} \) are all zero when the subscript \( j > i \). Since \( X^*_i \) is a function of \( Y_1, Y_2, \ldots, Y_i \), \( \frac{\partial x^*_i}{\partial y_j} \) are all zero when the subscript \( j > i \), so the \( i \)-th term of the diagonal elements becomes

\[
\frac{\partial x^*_i}{\partial x_i} = \frac{\partial x^*_i}{\partial y_i} \frac{\partial y_i}{\partial x_i} = c \prod_{j=1}^{i-1} (1 - y_j)^{\frac{1}{i-j+1}} (1 - y_i)^{-\frac{1}{i+1}} \left( \frac{\alpha_i}{\alpha_i} \right) \left( 1 - \sum_{j=1}^{i-1} x_j \right) \left( 1 - \sum_{j=1}^{i-1} x_j \right) \left( \sum_{j=i}^{n+1} \alpha_j \right)^{-1},
\]

where \( c \) is a constant.
where \( c \) is \( \frac{1}{n-i+1} \). The determinant of the inverse Jacobian matrix is given by

\[
\prod_{i=1}^{n} \frac{\partial x_i^*}{\partial x_i} = \prod_{i=1}^{n} \frac{\partial x_i^*}{\partial y_i} \frac{\partial y_i}{\partial x_i} = \frac{\Gamma(\alpha_1 + \cdots + \alpha_{n+1})}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_{n+1})} \prod_{i=1}^{n} \frac{1}{i} x_i^{\alpha_i-1} \left(1 - \sum_{i=1}^{n} x_i\right)^{\alpha_{n+1}-1}.
\]

The Jacobian determinant is the inverse of the determinant of the inverse Jacobian matrix, so the density function of the vector \((X_1^*, X_2^*, \ldots, X_n^*)\) is

\[
f(x_1^*, x_2^*, \ldots, x_n^*) = f(x_1, x_2, \ldots, x_n) \times \text{Jacobian}
\]

\[
= \frac{\Gamma(\alpha_1 + \cdots + \alpha_{n+1})}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_{n+1})} n! \prod_{i=1}^{n} x_i^{\alpha_i-1} (1 - \sum_{i=1}^{n} x_i)^{\alpha_{n+1}-1}
\]

where \( 0 < x_i^*, i = 1, \ldots, k, x_1^* + x_2^* + \cdots + x_k^* < 1 \). The support of the vector \((X_1^*, X_2^*, \ldots, X_n^*)\) is the same as the support of the vector \((X_1, X_2, \ldots, X_n)\). Therefore, the vector \((X_1^*, X_2^*, \ldots, X_n^*)\) follows the standard Dirichlet distribution.

4.4 Energy Test of Beta or Dirichlet Distribution

Formally the energy test of beta or Dirichlet Distribution follows.

\[
Q_{n,d} = n \left( \frac{2}{n} \sum_{i=1}^{n} E(|x_i - X|) - E|X - X'| - \frac{1}{n^2} \sum_{i,j=1}^{n} |x_i - x_j|^\alpha \right).
\]

We transform the observed sample \( X \) into the standardized sample \( Y \) by the formula \( (4.3.1) \). The standardized sample depends on the unknown parameters. However, the maximum likelihood estimate is consistent, which converges in probability to the true population parameter vector. The limiting distribution of the transformed sample is the standard Dirichlet distribution. We will
ignore the dependence of the standardized sample. The test statistic we will use is

$$Q_{n,d} = n \left( \frac{2}{n} \sum_{i=1}^{n} E(|y_i - Z|) - E|Z - Z'| - \frac{1}{n^2} \sum_{i,j=1}^{n} |y_i - y_j| \alpha \right), \quad (4.4.1)$$

where $Z$ is the standard beta or Dirichlet distributed variable. The first component of $Q_{n,d}$ for the Dirichlet distribution is obtained by simulation. The null hypothesis that the sample is from the beta or Dirichlet family is rejected for large values of $Q_{n,d}$. Please see Section 6.1 for the details of implementation of the energy test.

4.5 Triangle Tests of Beta or Dirichlet Distribution

The triangle goodness-of-fit test for the null hypothesis that the data is from the beta or Dirichlet family is based on

$$U_k = \frac{1}{\binom{n}{2}} \sum_{i<j} h_k(X_i, X_j) \quad for \quad k = 1, 2, 3,$$

where $U_1, U_2,$ and $U_3$ estimate the chance that the side joining the two data points is the smallest, middle, or largest side of the triangle formed from two randomly selected data points and the hypothetical point. We first transform the observed sample $X$ into the standardized sample $Y$ by the formula. The standardized sample depends on the unknown parameters. However, the maximum likelihood estimate is consistent. The standardized sample converges in distribution to the Dirichlet distribution. We will ignore the dependence of the standardized sample. Thus,

$$U_k = \frac{1}{\binom{n}{2}} \sum_{i<j} h_k(Y_i, Y_j) \quad for \quad k = 1, 2, 3, \quad (4.5.1)$$

where $U_1, U_2,$ and $U_3$ estimate the chance that the side joining the two selected data points is the smallest, middle, or largest side of the triangle formed from two randomly selected data points from the standardized sample and the hypothetical point from the standard beta or Dirichlet distribution. The three triangle statistics we will use are $U_3, (U_1 - 1/3)^2$ and $(U_1 - 1/3)^2 + (U_3 - 1/3)^2$. The null hypothesis that the sample is from the beta or Dirichlet family is rejected for large values or small values of $U_3$, for large values of $(U_1 - 1/3)^2$, and for large values of $(U_1 - 1/3)^2 + (U_3 - 1/3)^2$. 
See Section 6.1 for details of implementation of the triangle test.
CHAPTER 5  DISTANCE CORRELATION TEST OF DIRICHLET DISTRIBUTIONS

5.1 Distance Covariance and Distance Correlation

Distance correlation is a new measure of dependence between the random vectors $X$ and $Y$ in arbitrary dimension, recently introduced by Székely, Rizzo, and Bakirov (2007).

For all distributions with finite first moments, distance correlation $\mathcal{R}$ generalizes the idea of correlation in two fundamental ways:

1. $\mathcal{R}(\mathcal{X}, \mathcal{Y})$ is defined for $X$ and $Y$ in arbitrary dimensions;

2. $\mathcal{R}(\mathcal{X}, \mathcal{Y}) = 0$ characterizes independence of $X$ and $Y$.

Pearson’s product-moment covariance measures linear dependence between two variables. If the Pearson correlation coefficient $\rho = 1$, there is a pure positive linear relationship between the variables $X$ and $Y$. If $\rho = -1$, there is a pure negative linear relationship between $X$ and $Y$. If $\rho = 0$, there is no linear relationship between $X$ and $Y$. For the bivariate normal distribution, $\rho = 0$ is equivalent to independence of the variables $X$ and $Y$. However, it can be shown that except for the bivariate normal distribution, $\rho = 0$ is not a characterization of independence of two random variables.

Distance correlation has properties of a true dependence measure, analogous to, but more general than the product-moment Pearson’s correlation coefficient. Distance correlation satisfies $0 \leq \mathcal{R} \leq 1$, and $\mathcal{R} = 0$ only if $X$ and $Y$ are independent. Distance covariance and distance correlation are applicable for random vectors in arbitrary, not necessarily equal dimensions.

The following definitions are given by Székely et al. (2007).

**Definition 5.1.** The distance covariance (dCov) between random vectors $X$ and $Y$ with finite first
moments is the nonnegative number

\[ d\text{Cov}^2(X,Y) = ||\hat{f}_{X,Y}(s,t) - \hat{f}_X(s)\hat{f}_Y(t)||^2 \]

\[ = \frac{1}{c_pc_q} \int_{\mathbb{R}^{p+q}} \left| \frac{\hat{f}_{X,Y}(s,t) - \hat{f}_X(s)\hat{f}_Y(t)}{|s|^{1+p}|t|^{1+q}} \right|^2 dt \, ds. \]

where \( \hat{f}_{X,Y}(s,t) \) and \( \hat{f}_Y(t) \) are the characteristic functions of random vectors \((X, Y)\), \(X\), and \(Y\), respectively, \(||\cdot||\) is the function \(L_2\) norm, \(|\cdot|\) is the Euclidean norm in \(\mathbb{R}^p\), \(p, q\) denote the Euclidean dimension of \(X\) and \(Y\), \(|\cdot|_p\) denotes the Euclidean norm in \(\mathbb{R}^p\), \(c_p = \frac{\pi^{(p+1)/2}}{\Gamma\left(\frac{p+1}{2}\right)}\) and \(c_q = \frac{\pi^{(q+1)/2}}{\Gamma\left(\frac{q+1}{2}\right)}\), and the integral above exists.

The weight function \((c_pc_q|s|^{1+p}|t|^{1+q})^{-1}\) is chosen to produce a scale equivariant and rotation invariant measure that does not become zero for dependent variables (Szekely et al., 2007). By the definition of the norm \(||\cdot||\), it is clear that \(d\text{Cov}\) is nonnegative and \(d\text{Cov} = 0\) if and only if \(X\) and \(Y\) are independent.

**Theorem 5.1.** (Szekely et al., 2007) The population value of distance covariance is the square root of


\[ - 2 E[|X - X'||Y - Y'']|, \]

where \(E\) denotes the expected value, \(|\cdot|\) is the Euclidean norm, and \((X,Y), (X',Y'), (X'',Y'')\) are independent and identically distributed, provided the expectations are finite.

**Definition 5.2.** The distance variance of the random vector \(X\) with finite first moments is the
nonnegative number

\[ dV\text{ar}^2(X) = dV\text{ar}^2(X, X) = ||\hat{f}_{X,X}(s, t) - \hat{f}_X(s)\hat{f}_X(t)||^2 \]

\[ = \frac{1}{c_p c_q} \int_{\mathbb{R}^{p+q}} \frac{||\hat{f}_{X,X}(s, t) - \hat{f}_X(s)\hat{f}_X(t)||^2}{|s|_p^{1+p}|t|_q^{1+q}} \, dt \, ds, \]

provided the integral is finite.

**Theorem 5.2.** The population value of distance variance is the square root of

\[ d\text{Var}^2(X) = E[|X - X'|^2] + E^2[|X - X'|] - 2 E[|X - X'||X - X''|], \]

where \( E \) denotes the expected value, \( X, X', \) and \( X'' \) are iid, and the expectations are finite (Székely et al., 2007).

**Definition 5.3.** The distance correlation of two random vectors is obtained by dividing their distance covariance by the product of their distance standard deviations. The distance correlation is

\[ d\text{Cor}^2(X, Y) = \begin{cases} \frac{d\text{Cov}^2(X, Y)}{\sqrt{d\text{Var}^2(X) \, d\text{Var}^2(Y)}}, & d\text{Var}(X) \, d\text{Var}(X) > 0, \\ 0, & d\text{Var}(X) \, d\text{Var}(X) = 0. \end{cases} \]

Let \( \hat{f}_X^n(s), \hat{f}_X^n(t), \) and \( \hat{f}_{X,Y}^n(s, t) \) be the empirical characteristic functions of the samples \( X, Y, \) and \( (X, Y) \), respectively. We substitute the empirical characteristic functions for the characteristic functions in the definition of the distance covariance. The following result has been proved by Székely et al. (2007). If \( (X, Y) \) is a random sample from the joint distribution of \((X, Y)\), then

\[ ||\hat{f}_{X,Y}^n(s, t) - \hat{f}_X^n(s)\hat{f}_Y^n(t)||^2 = S_1 + S_2 - 2S_3, \]

where

\[ S_1 = \frac{1}{n^2} \sum_{k, l=1}^n |X_k - X_l|_p|Y_k - Y_l|_q, \]
\[ S_2 = \frac{1}{n^2} \sum_{k,l=1}^{n} |X_k - X_l| \frac{1}{m^2} \sum_{k,l=1}^{m} |Y_k - Y_l|, \]
\[ S_3 = \frac{1}{n^3} \sum_{k=1}^{n} \sum_{l,m=1}^{n} |X_k - X_l| |Y_k - Y_m|, \]
and
\[ \text{dCov}_n^2 (X, Y) = S_1 + S_2 - 2S_3, \quad (5.1.1) \]

where \((5.1.1)\) is algebraically equivalent to \(dCov_n^2(X, Y)\) in the following theorem by Székely and Rizzo (2007).

**Theorem 5.3.** \((\text{Székely et al., 2007})\) The empirical distance covariance \(dCov_n^2 (X, Y)\) is the non-negative number
\[ dCov_n^2 (X, Y) = \frac{1}{n^2} \sum_{j,k=1}^{n} A_{j,k} B_{j,k}, \quad (5.1.2) \]

where \(A_{j,k}\) and \(B_{j,k}\) are given by
\[ A_{j,k} = a_{j,k} - \bar{a}_j - \bar{a}_k + \bar{a}_., \]
\[ B_{j,k} = b_{j,k} - \bar{b}_j - \bar{b}_k + \bar{b}_., \]
\[ a_{j,k} = \|X_j - X_k\|, \quad j, k = 1, 2, \ldots, n, \]
\[ b_{j,k} = \|Y_j - Y_k\|, \quad j, k = 1, 2, \ldots, n, \]
\[ \bar{a}_i = \frac{1}{n} \sum_{j=1}^{n} a_{ij}, \quad \bar{a}_j = \frac{1}{n} \sum_{i=1}^{n} a_{ij}, \quad \bar{a}_. = \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} a_{jk}, \]
\[ \bar{b}_i = \frac{1}{n} \sum_{j=1}^{n} b_{ij}, \quad \bar{b}_j = \frac{1}{n} \sum_{i=1}^{n} b_{ij}, \quad \bar{b}_. = \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} b_{jk}, \]
where subscript ". " denotes that the mean is computed for the index that is replaced.

**Definition 5.4.** The empirical distance correlation $d\text{Cov}_n(X, Y)$ is the square root of

$$d\text{Cor}_n^2(X, Y) = \begin{cases} \frac{d\text{Cov}_n^2(X, Y)}{\sqrt{d\text{Var}_n(X) d\text{Var}_n(Y)}}, & d\text{Var}_n(X) d\text{Var}_n(Y) > 0, \\ 0, & d\text{Var}_n(X) d\text{Var}_n(Y) = 0, \end{cases}$$

where the sample distance variance is defined by

$$d\text{Var}_n^2(X) = d\text{Cov}_n^2(X, X) = \frac{1}{n^2} \sum_{k, \ell} A_{k, \ell}^2.$$ 

The statistic $d\text{Var}_n(X)$ is zero if and only if each sample observation is identical.

**Theorem 5.4.** [Székely et al., 2007] If $E|X| < \infty$, $E|Y| < \infty$ then almost surely

$$\lim_{n \to \infty} d\text{Cov}_n(X, Y) = d\text{Cov}(X, Y),$$

$$\lim_{n \to \infty} d\text{Cor}_n^2(X, Y) = d\text{Cor}^2(X, Y).$$

For the beta and Dirichlet distribution, since their support are bounded, it is always true that $E|X| < \infty$ and $E|Y| < \infty$, and therefore the $d\text{Cov}$ and $d\text{Cor}$ coefficients exist in that case.

Under independence, $n d\text{Cov}_n^2(X, Y)$ converges in distribution to a quadratic form $Q \overset{D}{=} \sum_{j=1}^{\infty} \lambda_j Z_j^2$, where $Z_j$ are independent standard normal random variables, and $\lambda_j$ are nonnegative constants that depend on the distribution of $(X, Y)$, which is proven by Székely et al. (2007). Under dependence, $n d\text{Cov}_n^2(X, Y)$ tends to infinity stochastically as $n \to \infty$. Hence, the test which rejects the null hypothesis of independence for large $n d\text{Cov}_n^2(X, Y)$ is consistent against the alternative hypothesis of dependence.

5.2 Characterization of Beta and Dirichlet Distributions

5.2.1 Complete Neutrality

Connor and Mosiman (1969) introduced the concept of neutrality. $X_2$ is defined to be neutral
with respect to \( X_1 \) if \( X_1/(1 - X_2), X_2 \) are independent. The set \( X_1, X_2, \cdots, X_k \) is defined to be completely neutral for the order \( 1, 2, \cdots, k \) if

\[
\begin{align*}
X_1, \frac{X_2}{1 - X_1}, \frac{X_3}{1 - X_1 - X_2}, \cdots, \frac{X_k}{1 - X_1 - X_2 - \cdots - X_{k-1}}
\end{align*}
\]

are jointly independent with a corresponding property for every permutation of \( \{1, 2, \cdots, k\} \).

**Theorem 5.5.** The random variable set \( \{X_1, X_2, \cdots, X_k\} \) is completely neutral for all permutations if and only if \( X_1, X_2, \cdots, X_k \) have a Dirichlet distribution.

**Proof.** Darroch and Ratcliff (1971) has shown that if the set \( \{X_1, X_2, \cdots, X_k\} \) is completely neutral for all permutations, then \( X_1, X_2, \cdots, X_k \) have a Dirichlet distribution. Here, we will show that if \( X_1, X_2, \cdots, X_k \) have a Dirichlet distribution with the parameters \( (\alpha_1, \alpha_2, \cdots, \alpha_{k+1}) \), then

\[
\begin{align*}
X_1, \frac{X_2}{1 - X_1}, \frac{X_3}{1 - X_1 - X_2}, \cdots, \frac{X_k}{1 - X_1 - X_2 - \cdots - X_{k-1}}
\end{align*}
\]

are jointly independent.

Let \( Y_1 = X_1, Y_2 = \frac{X_2}{1 - X_1}, Y_3 = \frac{X_3}{1 - X_1 - X_2}, \cdots, \) and \( Y_k = \frac{X_k}{1 - X_1 - X_2 - \cdots - X_{k-1}} \). The single-valued inverse functions are \( X_1 = Y_1, X_2 = Y_2(1 - Y_1), X_3 = Y_3(1 - Y_2)(1 - Y_1), \cdots, \) and \( X_k = Y_k \prod_{i=1}^{k-1} (1 - Y_{k-1}) \), so that the Jacobian matrix is

\[
\begin{bmatrix}
1 & 0 & \cdots & 0 & 0 \\
-y_2 & 1 - y_1 & \cdots & 0 & 0 \\
y_3(1 - y_2) & y_3(1 - y_1) & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
Y_k \prod_{i=1}^{k-2} (1 - Y_{k-1}) & Y_k \prod_{i=1}^{k-2} (1 - Y_{k-1}) & \cdots & Y_k \prod_{i=1}^{k-2} (1 - Y_{k-1}) & k-1 \prod_{i=1}^{k-2} (1 - Y_{k-1})
\end{bmatrix}
\]
The support of $X_1, X_2, \cdots, X_k$ maps onto

$$0 \leq Y_1 \leq 1, \quad 0 \leq Y_2(1 - Y_1) \leq 1, \quad 0 \leq Y_3(1 - Y_2)(1 - Y_1) \leq 1, \cdots,$$

$$0 \leq Y_k \prod_{i=1}^{k-1} (1 - Y_{k-i}) \leq 1,$$

which is equivalent to the support given by

$$\mathcal{T} = \{(y_1, y_2, \cdots, y_k) : 0 \leq y_i \leq 1, i = 1, 2, \cdots, k\}.$$ 

Hence the joint pdf of $Y_1, Y_2, \cdots, Y_{k+1}$ is

$$f(y_1, \cdots, y_k) = \frac{\Gamma\left(\sum_{i=1}^{k+1} \alpha_i\right)}{\Gamma(\alpha_1)\Gamma\left(\sum_{i=2}^{k+1} \alpha_i\right)} y_1^{\alpha_1-1}(1 - y_1)^{\sum_{i=2}^{k+1} \alpha_i-1} \times \frac{\Gamma\left(\sum_{i=2}^{k+1} \alpha_i\right)}{\Gamma(\alpha_2)\Gamma\left(\sum_{i=3}^{k+1} \alpha_i\right)} y_2^{\alpha_2-1}(1 - y_2)^{\sum_{i=3}^{k+1} \alpha_i-1} \times \cdots \times \frac{\Gamma\left(\sum_{i=k-1}^{k} \alpha_i\right)}{\Gamma(\alpha_{k-1})\Gamma\left(\sum_{i=k}^{k+1} \alpha_i\right)} y_{k-1}^{\alpha_{k-1}-1}(1 - y_{k-1})^{\sum_{i=k}^{k+1} \alpha_i-1} \times \frac{\Gamma(\alpha_k + \alpha_{k+1})}{\Gamma(\alpha_k)\Gamma(\alpha_{k+1})} y_k^{\alpha_k-1},$$

(5.2.1)

$(y_1, y_2, \cdots, y_k) \in \mathcal{T}$. The marginal pdf of $Y_i$ is

$$f_i(y_i) = \frac{\Gamma\left(\sum_{j=i+1}^{k+1} \alpha_j\right)}{\Gamma(\alpha_i)\Gamma\left(\sum_{j=i+1}^{k+1} \alpha_j\right)} y_i^{\alpha_i-1}(1 - y_i)^{\sum_{j=i+1}^{k+1} \alpha_j-1}, \quad y_i \in [0, 1].$$

Therefore, $Y_i$ follows the beta distribution with the left parameter $\alpha_i$ and the right parameter $\sum_{j=i+1}^{k+1} \alpha_j$ and in (5.2.1)

$$f(y_1, y_2, \cdots, y_{k+1}) = f_1(y_1)f_2(y_2) \cdots f_k(y_k).$$

Therefore the variables $Y_1, Y_2, \cdots, Y_k$ are mutually independent.
5.3 Testing of Mutual Independence of All Components in a Random Vector

To apply Theorem 5.5, for a random vector \((Y_1, Y_2, \cdots, Y_k)\), we would like to test the mutual independence of \(Y_1, Y_2, \cdots, Y_k\). We construct the following algorithm to test the mutual independence of \(k\) variables based on the following \(k - 1\) independent hypotheses:

\[
\begin{align*}
\mathcal{H}_{0,1} : & \ Y_1 \text{ is independent of the vector } (Y_2, Y_3, \cdots, Y_k), \\
\mathcal{H}_{0,2} : & \ Y_2 \text{ is independent of the vector } (Y_3, Y_4, \cdots, Y_k), \\
& \cdots \cdots \\
\mathcal{H}_{0,i} : & \ Y_i \text{ is independent of the vector } (Y_{i+1}, Y_{i+2}, \cdots, Y_k), \\
& \cdots \cdots \\
\mathcal{H}_{0,k-1} : & \ Y_{k-1} \text{ is independent of } Y_k.
\end{align*}
\]

The distance covariance is applicable for random vectors in arbitrary dimensions, so the \(d\text{Cov}\) test is a statistically consistent test of each hypothesis \(\mathcal{H}_{0,1}, \mathcal{H}_{0,2}, \cdots, \mathcal{H}_{0,i}, \cdots, \mathcal{H}_{0,k-1}\). The \(i\)-th hypothesis tests the independence of \(Y_i, i = 1, \cdots, k - 1\) and \(Y_{i+1}, Y_{i+2}, \cdots, Y_k\), which does not test any dependence or independence information among the variables \(Y_{i+1}, Y_{i+2}, \cdots, Y_k\). If the \(i\)-th hypothesis holds,

\[
f_{ik}(y_i, y_{i+1}, \cdots, y_{k+1}) = f_i(y_i)f_{(i+1)k}(y_{i+1}, \cdots, y_k).
\]

The \(i\)-th hypothesis is independent of the \((i+1)\)-th hypothesis, \(i = 1, \cdots, k - 1\).
If all $k - 1$ hypotheses hold, the joint pdf of the variables $y_1, y_2, \ldots, y_k$

$$f(y_1, \cdots, y_k) = f_1(y_1)f_2(y_2, \cdots, y_k)$$

$$= f_1(y_1)f_2(y_2)f_3(y_3, \cdots, y_k)$$

$$= f_1(y_1)f_2(y_2)\cdots f_i(y_i)f_{i+1}(y_{i+1}, \cdots, y_k)$$

$$= f_1(y_1)f_2(y_2)\cdots f_k(y_k).$$

Thus, the above $k - 1$ hypotheses as a whole test whether the variables $Y_1, Y_2, \ldots, Y_{k-1}$ are jointly independent. Since we have $k - 1$ independent hypotheses, the Bonferroni correction is chosen to control the familywise type I error rate. The correction is based on the idea that if $n$ dependent or independent hypotheses are tested, then one way of maintaining the familywise error rate $\alpha$ is to test each individual hypothesis at a statistical significance level of $\frac{\alpha}{n}$. The Bonferroni correction tends to be less conservative when all tests are independent of each other, but more conservative when some or all tests are dependent.

For the $k + 1$ independent hypotheses, the probability of at least one significant test result is $1 - (1 - \alpha/k)^k$. Figure 5.1 shows that as the number of independent tests increases from 2 to 10, the family-wise significance level decreases but still very close to $\alpha = 0.05$, which is why the Bonferroni correction is conservative.

5.4 Distance Covariance Goodness-of-Fit Tests

We propose a new goodness-of-fit test for the Dirichlet distribution based on the distance covariance. For an observed $k$-dimensional sample of $n$ observations $(x_{i,1}, x_{i,2}, \cdots, x_{i,k-1}, x_{i,k}), i = 1, 2, \cdots, n$, we would like to test whether the sample comes from a Dirichlet distribution:

$\mathbf{H_0}$: The sampled population has a Dirichlet distribution,

$\mathbf{H_1}$: The sampled population does not have a Dirichlet distribution.
Let $Y_1 = X_1, Y_i = \frac{X_i}{1 - X_1 - \cdots - X_{i-1}}, i = 2, \cdots, k$. By Theorem 5.5 under the null hypothesis of the sample from a Dirichlet distribution, $Y_1, Y_2, \cdots, Y_k$ are mutually independent. Under the alternative hypothesis, $Y_1, Y_2, \cdots, Y_k$ are not mutually independent.

We can transform the original goodness-of-fit problem of testing if the sample comes from the Dirichlet distribution into another test whether $Y_1, Y_2, \cdots, Y_k$ are mutually independent or not. Since $d\text{Cov}(X, Y) = 0$ if and only if $X$ and $Y$ are independent, we can transform the test if $Y_1, Y_2, \cdots, Y_k$ are mutually independent into testing whether the population distance covariance coefficient of $Y_1, Y_2, \cdots, Y_k$ is 0 or not. Thus, our test statistic is based on $d\text{Cov}_n^2(X, Y)$.

For example, if the random vector $(X_1, X_2)$ follows the two-dimensional Dirichlet distribution, then the variables $Y_1 = X_1, Y_2 = \frac{X_2}{1 - X_1}$ are independent, and $Y_3 = X_2, Y_4 = \frac{X_1}{1 - X_2}$ are also independent. Therefore, the population distance covariance of $Y_1, Y_2$ is 0, and the population distance covariance of $Y_3, Y_4$ is also 0. If the random vector $(X_1, X_2)$ does not follow the two-dimensional Dirichlet distribution, then the variables $Y_1$ and $Y_2$ are not independent or the variables $Y_1$ and $Y_2$ are not independent.
are not independent. Therefore, the population distance covariance of $Y_1, Y_2$ is positive, or the population distance covariance of $Y_3, Y_4$ is positive. If the variables $Y_1, Y_2$ are independent and $Y_3, Y_4$ are not independent, the random vector $(X_1, X_2)$ does not follow the Dirichlet distribution, but it may follow the generalized Dirichlet distribution, which is discussed in Chapter 8.

For the random vector $(X_1, X_2, \cdots, X_k)$ from $k$-dimensional Dirichlet distribution, the transformed $k$ variables $Y_1 = X_1, Y_2 = \frac{X_2}{1-X_1}, Y_3 = \frac{X_3}{1-x_1-x_2}, \cdots, \text{and } Y_k = \frac{x_k}{1-x_1-x_2-\cdots-x_{k-1}} \text{ for all permutations of the set } \{X_1, X_2, \cdots, X_k\} \text{ are mutually independent because of the complete neutrality of Dirichlet Distributions.}$

Therefore, we need to conduct the distance covariance test for each of $k!$ permutations. For each permutation of the $k$ variables, we conduct $(k - 1)$ null hypotheses. For all $k!$ permutations, we need to construct $k! \times (k - 1)$ null hypotheses. For the Dirichlet distribution, many of all $k! \times (k - 1)$ hypotheses are dependent. For the trivariate Dirichlet distribution, we construct $3! \times (2) = 12$ hypotheses. For the four-dimensional Dirichlet distribution, we construct $4! \times (3) = 72$ hypotheses. In the Bonferroni procedure, the nominal familywise $\alpha$ level is divided by the number of hypothesis tests. So when the number of hypotheses is too big and many of the hypotheses are dependent, the Bonferroni correction will make the empirical familywise type I error rate much less than the nominal familywise type I error rate $\alpha$. We will propose an improved algorithm to consider the dependence structure of all hypotheses to control the familywise type I error rate for higher dimensional ($k > 4$) Dirichlet Distributions. An R function `dcov.test` for distance covariance test of any two random samples are provided in the energy package by [Rizzo and Székely](2007). We can apply the same logic to conduct the test of mutual independence of a collection of random vectors.
CHAPTER 6 SIMULATION STUDY

6.1 Simulation Study

The empirical power of a goodness-of-fit test against a specified alternative distribution at a fixed significance level is obtained by drawing random samples from the alternative distribution, and computing the percentage of samples for which the test is significant at that significance level. Empirical critical values for each of the tests are estimated from 20,000 random samples from the standard Dirichlet distribution.

6.1.1 Energy and Triangle Tests

First, we compute the empirical critical value for the energy or triangle tests.

1. Generate random sample $X_1, \ldots, X_n$ from the standard Dirichlet or Dirichlet distribution.
2. Compute estimates of parameters based on MLE.
4. Compute the energy statistic (or triangle statistics) on the transformed data.
5. Get critical values from the empirical 95th percentile of energy statistic (or triangle statistics) in step 4.

For a given alternative distribution, the empirical power is computed as follows.

1. Generate 1,000 random samples of size $n$ from the specified alternative distribution.
2. Standardize the sample using the formula (4.3.1).
3. Calculate the energy statistic $Q_n$ using formula (4.4.1), three triangle statistics $U_3$, $(U_1 - 1/3)^2$ and $(U_1 - 1/3)^2 + (U_3 - 1/3)^2$ using formula (4.5.1).
4. Set $N$ equal to the number of the test statistics which are greater than the critical value.
5. Compute the estimated power by $\frac{N}{1000}$.

For energy or triangle tests, we calculate the estimated power but not the p-values.
6.1.2 Distance Covariance Test

For the random vector \((X_1, X_2, \cdots, X_k)\) from \(k\)-dimensional Dirichlet distribution, the transformed \(k\) variables \(Y_1 = X_1, Y_2 = \frac{x_2}{1-x_1}, Y_3 = \frac{x_3}{1-x_1-x_2}, \cdots, Y_k = \frac{x_k}{1-x_1-x_2-\cdots-x_{k-1}}\) for all permutations of the set \(\{X_1, X_2, \cdots, X_k\}\) are mutually independent. Let \(\hat{\theta}\) be a two sample statistic for testing multivariate independence. Before we construct the distance covariance goodness-of-fit test, we need to introduce the permutation test of independence of two random vectors. Suppose that \(X\) is a \(p\)-dimensional sample and \(Y\) is a \(q\)-dimensional sample and \(Z = (X, Y)\). Then \(Z\) is a \(p+q\) dimensional sample. Let \(v_1\) be the row labels of the \(X\) sample and let \(v_2\) be the row labels of the \(Y\) sample. Then \((Z, v_1, v_2)\) is the sample from the joint distribution of \(X\) and \(Y\). The permutation test procedure permutes the row indices of \(Y\) sample.

Approximate permutation test procedure for independence of two random vectors is as follows.

Let \(\hat{\theta} (dCov^2_n)\) be a two sample statistic using formula of (5.1.2) for testing independence of two random vectors.

1. Compute the observed test statistic \(\hat{\theta}(X, Y) = \hat{\theta}(Z, v_1, v_2)\).
2. For each replicate, indexed \(b = 1, \cdots, B\):
   (a) Generate a random permutation \(\pi_b = \pi(v_2)\).
   (b) Calculate the statistic \(\hat{\theta}^{(b)}(X, Y^*, \pi(v_2))\).
   (c) Calculate the ASL by
      \[
      \hat{p} = \frac{\left\{1 + \sum_{b=1}^{B} I(\hat{\theta}^{(b)} \geq \hat{\theta})\right\}}{B + 1}.
      \]

Permutation test procedure for distance covariance goodness-of-fit tests is as follows.

1. Generate 1,000 random samples of size \(n\) from the specified alternative distribution. For each sample \(\{X_1, X_2, \cdots, X_k\}\), repeat step 2 to step 5.
2. For each of all \(k!\) permutations of each sample \(\{X_1, X_2, \cdots, X_k\}\), repeat step 3 to step 4.
3. Compute \(Y_1 = X_1, Y_2 = \frac{x_2}{1-x_1}, Y_3 = \frac{x_3}{1-x_1-x_2}, \cdots, Y_k = \frac{x_k}{1-x_1-x_2-\cdots-x_{k-1}}\).
4. Compute the \(k-1\) observed test statistics \(\hat{\theta}(Y_1, (Y_2, \cdots, Y_k)), \hat{\theta}(Y_2, (Y_3, \cdots, Y_k)), \cdots, \hat{\theta}\)
(Y_{k-1}, Y_k)$, and get the ASL based on the above approximate permutation test procedure for independence.

5. Reject $H_0$ at significance level $\alpha$ if any of $(k - 1)k!$ ASL (achieved significance level) is less than $\frac{\alpha}{(k-1)k!}$.

6. Set $N$ equal to the number of times $H_0$ is rejected.

7. Compute the estimated power by $\frac{N}{1000}$.

For more details about permutation test for independence, please see Rizzo (2008). For the three and four-dimensional Dirichlet distribution, based on the above procedures, the empirical type I error rate of the distance covariance test is 0.046.

6.2 Contaminated Dirichlet Distribution

We introduce a class of alternative distributions below that have the same support as a Dirichlet distribution.

**Definition 6.1.** A random vector $(Y_1, Y_2, \cdots, Y_k)$ has a contaminated Dirichlet distribution with parameters $(\alpha_1, \alpha_2, \cdots, \alpha_{k+1})$ if the pdf of the random vector is

$$f(y_1, y_2, \cdots, y_k) = f_1(y_1, y_2, \cdots, y_k^*)y_k^*, \quad 0 \leq y_1, y_2, \cdots, y_{k-1} < 1, \quad 0 < y_k \leq \frac{(1 - y_1 - \cdots - y_{k-1})^2}{1 - y_1 - \cdots - y_{k-2}},$$

where $f_1(y_1, y_2, \cdots, y_k)$ is the density function of a Dirichlet distribution with parameters $(\alpha_1, \alpha_2, \cdots, \alpha_{k+1})$ and $y_k^* = \frac{y_k(1-y_1-\cdots-y_{k-2})}{1-y_1-\cdots-y_{k-1}}$.

The following transformation method can be applied to generate a sample from a contaminated Dirichlet distribution with parameters $(\alpha_1, \alpha_2, \cdots, \alpha_{k+1})$. If the random variables $X_1, \cdots, X_k$ follow a Dirichlet distribution with the parameters $(\alpha_1, \alpha_2, \cdots, \alpha_{k+1})$, then

$$Y_i = X_i, \quad i = 1, 2, \cdots, k - 1,$$

$$Y_k = \frac{1 - x_1 - x_2 - \cdots - x_{k-1}}{1 - x_1 - x_2 - \cdots - x_{k-2}} x_k$$
is a contaminated Dirichlet distribution with parameters \((\alpha_1, \alpha_2, \ldots, \alpha_{k+1})\). The first \(k - 1\) components have the same corresponding marginal beta distribution as the Dirichlet distribution with parameters \((\alpha_1, \alpha_2, \ldots, \alpha_{k+1})\). Only the last component is contaminated; compared with the marginal distribution of the last component of the corresponding Dirichlet distribution, its values become relatively smaller, which have high density. By the density contour plots of Example 6.1 and Example 6.2 it is easy to see that small values of the last component have high density.

**Example 6.1** (Power comparison of tests of bivariate Dirichlet distributions).

Let two continuous random variables \(Y_1\) and \(Y_2\) have a joint density given by

\[
f(y_1, y_2) = \frac{2}{1 - y_1}, \quad 0 < y_1 < 1, \quad 0 < y_2 \leq (1 - y_1)^2.
\]

The marginal distribution of \(Y_1\) is the beta distribution with parameters 1 and 2. The marginal distribution of \(Y_2\) has the density function

\[
f(y_2) = 2 \log \frac{1}{\sqrt{y_2}}, \quad 0 < y_2 < 1.
\]

The random variable \(Y_2\) has a mean of \(\frac{1}{4}\) and a variance of \(\frac{7}{144}\). The marginal distribution of \(Y_1\) is the same as the marginal distribution of the first component of standard Dirichlet distribution with parameters \((1, 1, 1)\) and only the marginal distribution of \(Y_2\) is different from the marginal distribution of the second component; that is, the last component is contaminated, so we call the distribution of \(Y_1\) and \(Y_2\) a contaminated Dirichlet distribution. We have the following transformation method for generating the random variables \(Y_1\) and \(Y_2\). If the random variables \(X_1\) and \(X_2\) follow the bivariate Dirichlet distribution with the parameters \((1, 1, 1)\), then

\[
Y_1 = X_1, \\
Y_2 = X_2(1 - X_1)
\]
have the joint pdf

\[ f(y_1, y_2) = \frac{2}{1 - y_1}, \quad 0 \leq y_1 \leq 1, \quad 0 \leq y_2 \leq (1 - y_1)^2. \]

If the variables \( Z_1 \) and \( Z_2 \) follow the Dirichlet distribution with parameters \((1, \frac{3}{4}, \frac{5}{4})\), then they have the same mean and variance as the variables \( Y_1 \) and \( Y_2 \).

Figure 6.1: Contour plot of density function of variables \( Y_1 \) and \( Y_2 \).

Figure 6.1 shows the contour plot of the density function of variables \( Y_1 \) and \( Y_2 \).

Figure 6.2 shows the contour plot of the density function of a Dirichlet distribution with parameters \((1, 3/4, 5/4)\).

We compare the empirical power of the distance covariance test, energy test and triangle tests. If we denote the observed Type I error rate by \( \hat{p} \), then an estimate of \( \text{se}(\hat{p}) \) is

\[ \hat{\text{se}}(\hat{p}) = \sqrt{\frac{\hat{p}(1 - \hat{p})}{m}} \leq \frac{0.5}{\sqrt{m}}. \]

Here \( m \) is the number of replicates for the empirical power and we set \( m = 1000 \). Let \( \mathcal{N} \) denote the family of bivariate Dirichlet distributions. Then the test hypotheses are

\[ H_0 : F_x \in \mathcal{N} \quad \text{vs} \quad H_1 : F_x \notin \mathcal{N}. \]
The power comparison for Example 6.1 is summarized in Table 6.1 and in Figure 6.3. The sample size is 25, 50, 100, respectively.

Table 6.1: Empirical Power of Five Tests of Dirichlet Against a Contaminated Dirichlet Alternative in Example 6.1

<table>
<thead>
<tr>
<th>n</th>
<th>DCov</th>
<th>Energy</th>
<th>Triangle.1</th>
<th>Triangle.2</th>
<th>Triangle.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.643</td>
<td>0.349</td>
<td>0.070</td>
<td>0.051</td>
<td>0.049</td>
</tr>
<tr>
<td>50</td>
<td>0.949</td>
<td>0.726</td>
<td>0.071</td>
<td>0.061</td>
<td>0.057</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>0.930</td>
<td>0.079</td>
<td>0.087</td>
<td>0.064</td>
</tr>
</tbody>
</table>

Since the contaminated Dirichlet distribution is only the last component different from its counterpart of the Dirichlet distribution, it's shape is similar to that of the Dirichlet, and the triangle tests fail to detect the small change in the shape. However, the distance covariance test measure the independence. Even though there is small shape change in the last component, there is big change in the complete neutrality, that is, the big change in mutual independence. The distance covariance test is sensitive to the change of mutual independence of data structure. That is the reason why the distance covariance test has big powers. The simulation results suggest that the distance covariance test is the most powerful among all five tests against is alternative. The energy test is more powerful than all three triangle tests. The three triangle tests have comparable power.
Figure 6.3: Empirical power of five tests of Dirichlet against a contaminated Dirichlet alternative in Example 6.1.

to each other.

Example 6.2 (Power comparison of tests of trivariate Dirichlet distributions).

Three continuous random variables \(Y_1, Y_2\) and \(Y_3\) have a joint density which is given by

\[
    f(y_1, y_2, y_3) = 6 \frac{1 - y_1}{1 - y_1 - y_2}, \quad 0 < y_1 < 1, \quad 0 < y_2 < 1, \quad 0 < y_3 < \frac{(1 - y_1 - y_2)^2}{1 - y_1}.
\]

The marginal distributions of \(Y_1\) and \(Y_2\) both have the beta distribution with parameters \(\alpha = 1\) and \(\beta = 3\). The joint distribution of \(Y_1\) and \(Y_3\) has the density function

\[
    f(y_1, y_3) = 6(1 - y_1) \log \sqrt{\frac{1 - y_1}{y_3}},
\]
and its support is given by 

\[ 0 < y_1 < 1, \ 0 < y_3 < 1 - y_1. \]

The random variable \( Y_2 \) has mean \( \frac{1}{4} \) and variance \( \frac{7}{144} \). The marginal distribution of \( Y_1 \) and \( Y_2 \) is the same as the marginal distribution of the first component of the standard Dirichlet distribution with parameters \( (1, 1, 1) \) and only the marginal distribution of \( Y_3 \) is different from the marginal distribution of the last component of the standard Dirichlet distribution; that is, the last component is contaminated, so we call the joint distribution of \( Y_1, Y_2 \) and \( Y_3 \) a trivariate contaminated Dirichlet distribution. We have the following transformation method for generating random variables \( Y_1, Y_2 \) and \( Y_3 \). If the random variables \( X_1, X_2 \) and \( X_3 \) follow the standard Dirichlet distribution with the parameters \( (1, 1, 1, 1) \), then

\[
Y_1 = X_1, \\
Y_2 = X_2, \\
Y_3 = \frac{X_3}{1 - X_1} (1 - X_1 - X_2)
\]

have the joint pdf

\[
f(y_1, y_2, y_3) = 6 \frac{1 - y_1}{1 - y_1 - y_2}, \quad 0 < y_1 < 1, \ 0 < y_2 < 1, \ 0 < y_3 \leq \frac{(1 - y_1 - y_2)^2}{1 - y_1}.
\]

The Dirichlet distribution with parameters \( (\alpha_1 = 1, \alpha_2 = 0.750, \alpha_3 = 1.251) \) have the same mean and variance as the joint distribution of \( Y_1 \) and \( Y_3 \). Figure 6.4 shows the contour plot of the density function of variables \( Y_1, Y_2 \) and \( Y_3 \) in (5.9). The contour plot of the density function of a Dirichlet distribution with parameters \( (1, 1, 2) \) is shown in Figure 6.5.

Figure 6.6 shows the contour plot of the density function of a Dirichlet distribution with parameters \( (1, 0.7495, 1.2505) \).

We compare the empirical power of the distance covariance test, energy test and triangle tests. The number of replicates for the empirical power is 1000. Let \( \mathcal{N} \) denote the family of trivariate
Figure 6.4: Contour plot of density function of variables $Y_1$, $Y_2$ and $Y_3$ in (5.9).

Figure 6.5: Contour plot of density function of Dir(1, 1, 2).

Dirichlet distributions. Then the test hypotheses are

$$H_0 : F_x \in \mathcal{N} \quad \text{vs} \quad H_1 : F_x \notin \mathcal{N}.$$  

The power comparison for Example 6.2 is summarized in Table 6.2 and Figure 6.7.
Figure 6.6: Contour plot of density function of \( \text{Dir}(1, 0.7495, 1.2505) \).

Table 6.2: Empirical Power of Five Tests of Dirichlet Against a Contaminated Dirichlet Alternative in Example 6.2

<table>
<thead>
<tr>
<th>n</th>
<th>DCov</th>
<th>Energy</th>
<th>Triangle.1</th>
<th>Triangle.2</th>
<th>Triangle.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.367</td>
<td>0.261</td>
<td>0.050</td>
<td>0.048</td>
<td>0.051</td>
</tr>
<tr>
<td>50</td>
<td>0.839</td>
<td>0.508</td>
<td>0.053</td>
<td>0.054</td>
<td>0.058</td>
</tr>
<tr>
<td>100</td>
<td>0.998</td>
<td>0.773</td>
<td>0.061</td>
<td>0.071</td>
<td>0.061</td>
</tr>
</tbody>
</table>

The simulation results suggest that the distance covariance test is the most powerful among all five tests against this alternative. The energy test is more powerful than all three triangle tests. The three triangle tests have comparable power to each other.

We can generalize the distributions in Example 6.1 and Example 6.2 to multivariate contaminated Dirichlet distributions.

Table 6.3: Empirical Power of Five Tests of Dirichlet Against a Contaminated Dirichlet Alternative in Example 6.3

<table>
<thead>
<tr>
<th>n</th>
<th>DCov</th>
<th>Energy</th>
<th>Triangle.1</th>
<th>Triangle.2</th>
<th>Triangle.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.887</td>
<td>0.437</td>
<td>0.074</td>
<td>0.050</td>
<td>0.048</td>
</tr>
<tr>
<td>50</td>
<td>0.996</td>
<td>0.849</td>
<td>0.080</td>
<td>0.061</td>
<td>0.053</td>
</tr>
<tr>
<td>100</td>
<td>1.000</td>
<td>0.985</td>
<td>0.126</td>
<td>0.148</td>
<td>0.089</td>
</tr>
</tbody>
</table>

Example 6.3 (Power comparison of tests of trivariate Dirichlet distributions (3,5,4,2)).
Following Definition 6.1, we compared the power of the five tests against multivariate contaminated Dirichlet distribution with the parameter vector \((3, 5, 4, 2)\).

Again, the number of replicates for the empirical power is 1000. The results are summarized in Figure 6.8. The simulation results suggest that the distance covariance test is the most powerful among all five tests against this alternative. The energy test is more powerful than all three triangle tests. The three triangle tests have comparable power to each other.

6.3 Logistic Normal Distribution

The support of the beta distribution is from 0 to 1. The logit normal distribution has the same support as the beta distribution. If the random variable \(Y\) follows the normal distribution, then \(X = \frac{e^Y}{1+e^Y}\) follows the logit normal distribution. If \(X\) follows the logit normal distribution, then \(Y = \text{logit}(X) = \log(\frac{X}{1-X})\) follows the normal distribution.

The multivariate logit distribution is known as the logistic normal distribution. If the random
Figure 6.8: Empirical power of five tests of Dirichlet against a contaminated Dirichlet alternative in Example 6.3

The vector \( \mathbf{Y} \) follows the multivariate normal distribution \( N(\mathbf{\mu}, \mathbf{\Sigma}) \), then the transformed random vector

\[
\mathbf{X} = \left( \frac{e^{y_1}}{1 + \sum_{i=1}^{d-1} e^{y_i}}, \ldots, \frac{e^{y_{d-1}}}{1 + \sum_{i=1}^{d-1} e^{y_i}}, \frac{1}{1 + \sum_{i=1}^{d-1} e^{y_i}} \right)
\]

follows the logistic normal distribution.

If the \( d \)-dimensional random vector \( \mathbf{X} \) follows the logistic multivariate normal distribution, then the transformed random vector

\[
\mathbf{Y} = \left( \log \left( \frac{x_1}{x_d} \right), \ldots, \log \left( \frac{x_{d-1}}{x_d} \right) \right)
\]

follows the multivariate normal distribution.

The logit normal distribution is distinct from the beta distribution (Hinde, 2014). However, the logistic normal distribution is an alternative to the Dirichlet distribution. The logit normal
distribution has been applied in random effects models for binary data and the logistic normal distribution was applied by Aitchison and Begg (1976) for statistical diagnosis/discrimination.

**Example 6.4** (Power comparison of tests of logistic normal distributions).

We compare the empirical power of five proposed goodness-of-fit tests when the sampled population has a logistic normal distribution and our null distribution is the Dirichlet distribution.

In this experiment, we generated 1000 random samples of size 25, 50 and 100 from the four-dimensional logistic normal distribution with the corresponding trivariate normal distribution with mean vector $(0, 0, 0)$ and identity covariance matrix.

Results for this example are summarized in Table 6.4 and Figure 6.9.

Table 6.4: Empirical Power of Five Tests of Dirichlet Against Logistic Normal Alternative in Example 6.4

<table>
<thead>
<tr>
<th>n</th>
<th>DCov</th>
<th>Energy</th>
<th>Triangle.1</th>
<th>Triangle.2</th>
<th>Triangle.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.622</td>
<td>0.551</td>
<td>0.098</td>
<td>0.125</td>
<td>0.138</td>
</tr>
<tr>
<td>50</td>
<td>0.951</td>
<td>0.877</td>
<td>0.139</td>
<td>0.159</td>
<td>0.184</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>0.996</td>
<td>0.223</td>
<td>0.249</td>
<td>0.311</td>
</tr>
</tbody>
</table>

The simulation results suggest that the distance covariance test is the most powerful among all five tests against this alternative. The energy test is more powerful than all three triangle tests. The three triangle tests have comparable power to each other.
Figure 6.9: Empirical power of five tests of Dirichlet against logistic normal alternative in Example 6.4
CHAPTER 7 MAIN RESULTS FOR GENERALIZED DIRICHLET DISTRIBUTIONS

7.1 Generalized Dirichlet Distribution

A random vector \((X_1, X_2, \ldots, X_k)\) follows the generalized Dirichlet distribution with parameters \(\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_k, \beta_k > 0\), if it has the pdf

\[
f(x_1, x_2, \ldots, x_k) = \prod_{i=1}^{k} \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i) \Gamma(\beta_i)} \times \prod_{i=1}^{k-1} x_i^{\alpha_i-1} (1 - \sum_{j=1}^{i} x_j)^{\beta_i - (\alpha_i+1+\beta_i+1)} x_k^{\alpha_k-1} (1 - \sum_{j=1}^{k} x_j)^{\beta_k-1},
\]

where \(0 < x_i, i = 1, \ldots, k, x_1 + x_2 + \cdots + x_k < 1\). If \(\beta_i = \alpha_{i+1} + \beta_{i+1}, i = 1, 2, \ldots, k - 1\), then the generalized Dirichlet distribution becomes the Dirichlet distribution. The variables of the generalized Dirichlet distribution are neutral but not completely neutral (Connor and Mosiman, 1969). For two-dimensional generalized Dirichlet distributions, the support is a right triangle with vertices \((0,0), (0,1), \) and \((1,0)\). For three-dimensional generalized Dirichlet distributions, the support is a pyramid with vertices \((0, 0, 0), (1, 0, 0), (0, 1, 0), \) and \((0, 0, 1)\). We can regard the support of a higher-dimensional generalized Dirichlet distribution as the superpyramid. For the goodness-of-fit test for the generalized Dirichlet distribution, translation and rotation will change the support, which is bounded and fixed in the area of \(0 < x_i, i = 1, \ldots, k, x_1 + x_2 + \cdots + x_k < 1\). Therefore, we do not need to consider the affine invariant properties of goodness-of-fit tests for the generalized Dirichlet distribution, even though affine invariant properties are very important for goodness-of-fit tests for normal distributions. However, we still need one specific kind of affine transformation to change arbitrary \(k\)-dimensional generalized Dirichlet distributions to the corresponding standard Dirichlet distribution with all parameters 1. We calculate our test statistics of the energy or triangle test based on the transformed standard Dirichlet distribution, which makes
both tests easier to use and improves their efficiency.

For a random variable $X$ which follows the multivariate normal distribution $N(\mu, \Sigma)$, we can use the transformation $(X - \mu)^T \Sigma^{-1} (X - \mu)$ to change the multivariate normal random variable into the corresponding multivariate standard normal variable. However, for the generalized Dirichlet distribution, we do not have such a simple formula of transformation. Similar to the transformation applied in Chapter 4, we can use a two-step transformation technique to transform the $k$-dimensional generalized Dirichlet distribution into the $k$-dimensional standard Dirichlet distribution.

**Theorem 7.1.** Suppose that the random vector $(X_1, \ldots, X_n)$ is from an n-dimensional generalized Dirichlet distribution with the parameter vector $(\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_n, \beta_n)$. Then $(X^*_1, \ldots, X^*_n)$ defined by the following two-step transformation is jointly distributed as the standard n-dimensional Dirichlet distribution. The two-step transformation is defined as follows. The first step of transformation is given by

$$
Y_1 = I(X_1 | \alpha_1, \beta_1)
$$

$$
Y_2 = I\left(\frac{X_2}{1 - X_1} | \alpha_2, \beta_2, X_1\right),
$$

$$
Y_3 = I\left(\frac{X_3}{1 - X_1 - X_2} | \alpha_3, \beta_3, X_1, X_2\right),
$$

\ldots

$$
Y_i = I\left(\frac{X_i}{1 - X_1 - X_2 - \cdots - X_{i-1}} | \alpha_i, \beta_i, X_1, \ldots, X_{i-1}\right),
$$

\ldots

$$
Y_n = I\left(\frac{X_n}{1 - X_1 - \cdots - X_{n-1}} | \alpha_n, \beta_n, X_1, \ldots, X_{n-1}\right),
$$

where $I(\frac{X_i}{1 - X_1 - X_2 - \cdots - X_{i-1}} | \alpha_i, \beta_i, X_1, \ldots, X_{i-1})$, function is the conditional regularized incomplete beta function of $X_i$ given $X_1, \ldots, X_{i-1}$.
The second step of transformation is given by

\begin{align*}
X_1^* &= 1 - (1 - Y_1)^{\frac{1}{n}}, \\
X_2^* &= (1 - Y_1)^{\frac{1}{n}} \left[ 1 - (1 - Y_2)^{\frac{1}{n-1}} \right], \\
X_3^* &= (1 - Y_1)^{\frac{1}{n}} (1 - Y_2)^{\frac{1}{n-1}} \left[ 1 - (1 - Y_3)^{\frac{1}{n-2}} \right], \\
&\quad \ldots \ldots \ldots \ldots \\
X_i^* &= \prod_{j=1}^{i-1} (1 - Y_j)^{\frac{1}{n-i+1}} \left[ 1 - (1 - Y_i)^{\frac{1}{n-i+1}} \right], \\
&\quad \ldots \ldots \ldots \ldots \\
X_n^* &= \prod_{i=1}^{n-1} (1 - Y_i)^{\frac{1}{n-i+1}} Y_n.
\end{align*}

The random vector \((X_1^*, \ldots, X_n^*)\) follows the standard \(n\)-dimensional Dirichlet distribution.

**Proof.** The first step is that we change the \(k\)-dimensional generalized Dirichlet distribution into the \(k\)-dimensional uniform distribution with the support \(0 < x_i < 1, i = 1, \ldots, k\), which can be called the supercube uniform distribution because for the 2-dimensional support, it is a square with the unity side and for the 3-dimensional support, it is a cube with the unity side. The second step is that we change the \(k\)-dimensional supercube uniform distribution into the \(k\)-dimensional standard Dirichlet distribution. During the first step, the non-uniform density is transformed into the uniform density which is what we desire, but the support is transformed from superpyramid into supercube which is not what we want. During the second step, we keep the uniform density but change the support from supercube back into superpyramid again.

The Jacobian determinant is the determinant of the Jacobian matrix which is given by

\[
\begin{bmatrix}
\frac{\partial x_1^*}{\partial x_1} & \frac{\partial x_1^*}{\partial x_2} & \cdots & \frac{\partial x_1^*}{\partial x_n} \\
\frac{\partial x_2^*}{\partial x_1} & \frac{\partial x_2^*}{\partial x_2} & \cdots & \frac{\partial x_2^*}{\partial x_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial x_n^*}{\partial x_1} & \frac{\partial x_n^*}{\partial x_2} & \cdots & \frac{\partial x_n^*}{\partial x_n}
\end{bmatrix}.
\]
However, it is hard to directly derive each element of the Jacobian matrix. On the other hand, the Jacobian determinant is the inverse of the determinant of the matrix which is given by

$$
\begin{bmatrix}
\frac{\partial x^*_1}{\partial x_1} & \frac{\partial x^*_1}{\partial x_2} & \cdots & \frac{\partial x^*_1}{\partial x_n} \\
\frac{\partial x^*_2}{\partial x_1} & \frac{\partial x^*_2}{\partial x_2} & \cdots & \frac{\partial x^*_2}{\partial x_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial x^*_n}{\partial x_1} & \frac{\partial x^*_n}{\partial x_2} & \cdots & \frac{\partial x^*_n}{\partial x_n}
\end{bmatrix}.
$$

We call the above matrix the inverse Jacobian matrix. Since $X^*_i$ is a function of $X_1, X_2, \ldots, X_i$, $\frac{\partial x^*_i}{\partial x_j}$ are all zero when the subscript $j > i$, so the inverse Jacobian matrix becomes

$$
\begin{bmatrix}
\frac{\partial x^*_1}{\partial x_1} & 0 & \cdots & 0 \\
\frac{\partial x^*_2}{\partial x_1} & \frac{\partial x^*_2}{\partial x_2} & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial x^*_n}{\partial x_1} & \frac{\partial x^*_n}{\partial x_2} & \cdots & \frac{\partial x^*_n}{\partial x_n}
\end{bmatrix}.
$$

Therefore, we only need to calculate the diagonal element of the inverse Jacobian matrix to obtain the determinant of the inverse Jacobian matrix. The $i$-th term of the diagonal elements is given by

$$
\frac{\partial x^*_i}{\partial x_i} = \frac{\partial x^*_i}{\partial y_1} \frac{\partial y_1}{\partial x_i} + \frac{\partial x^*_i}{\partial y_2} \frac{\partial y_2}{\partial x_i} + \cdots + \frac{\partial x^*_i}{\partial y_n} \frac{\partial y_n}{\partial x_i}, \quad i = 1, 2, \ldots, n.
$$

Since $Y_i$ is a function of $X_1, X_2, \ldots, X_i$, $\frac{\partial y_i}{\partial x_j}$ are all zero when the subscript $j > i$. Since $X^*_i$ is a function of $Y_1, Y_2, \ldots, Y_i$, $\frac{\partial x^*_i}{\partial y_j}$ are all zero when the subscript $j > i$, so the $i$-th term of the diagonal elements becomes

$$
\frac{\partial x^*_i}{\partial x_i} = \frac{\partial x^*_i}{\partial y_i} \frac{\partial y_i}{\partial x_i} = c \prod_{j=1}^{i-1} \frac{1}{1 - y_j} \frac{1}{1 - y_i} \frac{1}{\alpha_i} x_i^{\alpha_i - 1} \left( \frac{1 - \sum_{j=1}^{i} x_j^{\beta_i - 1}}{1 - \sum_{j=1}^{i-1} x_j^{\beta_i}} \right) \frac{\sum_{j=1}^{i} x_j^{\beta_i - 1}}{\sum_{j=1}^{i-1} x_j^{\beta_i}},
$$
where \( c = \frac{1}{n-i+1} \frac{\Gamma(\alpha_i+\beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)}. \)

The determinant of the inverse Jacobian matrix is given by

\[
\prod_{i=1}^{n} \frac{\partial x_i^*}{\partial x_i} = \prod_{i=1}^{n} \frac{\partial x_i^*}{\partial y_i} \frac{\partial y_i}{\partial x_i} = \prod_{i=1}^{k} \frac{\Gamma(\alpha_i + \beta_i)}{i!\Gamma(\alpha_i)\Gamma(\beta_i)} \prod_{i=1}^{k-1} x_i^{\alpha_i-1} (1 - \sum_{j=1}^{i} x_j)^{\beta_i-(\alpha_{i+1}+\beta_{i+1})} x_k^{\alpha_k-1} (1 - \sum_{j=1}^{k} x_j)^{\beta_k-1} = \frac{f(x_1, x_2, \ldots, x_n)}{n!},
\]

where \( f(x_1, x_2, \ldots, x_n) \) is the density function of the generalized Dirichlet distribution with the parameter \( \alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_k, \beta_k \). The Jacobian determinant is the inverse of the determinant of the inverse Jacobian matrix, so the density function of the vector \((X_1^*, X_2^*, \ldots, X_n^*)\) is

\[
f(x_1^*, x_2^*, \ldots, x_n^*) = f(x_1, x_2, \ldots, x_n) \times \text{Jacobian} = f(x_1, x_2, \ldots, x_n) \times \frac{n!}{f(x_1, x_2, \ldots, x_n)} = n!,
\]

where \( 0 < x_i^*, i = 1, \ldots, k, x_1^* + x_2^* + \cdots + x_k^* < 1 \). The support of the vector \((X_1^*, X_2^*, \ldots, X_n^*)\) is the same as the support of the vector \((X_1, X_2, \ldots, X_n)\). Therefore, the vector \((X_1^*, X_2^*, \ldots, X_n^*)\) follows the standard Dirichlet distribution.

7.2 Estimation for the Generalized Dirichlet Distribution Parameter Vector

We have \( n \) independent and identically distributed variables from the generalized Dirichlet \((\alpha_1, \beta_1, \ldots, \alpha_k, \beta_k)\). For \( i = 1, \ldots, n \), let \( x_i = (x_{i1}, \ldots, x_{ik}) \) be \( n \) observations. Chang and Richards (1991) discussed the parameter estimation based on the maximum likelihood. We propose the maximum likelihood parameter estimation based on the Newton-Raphson Algorithm, the initial value of iterations and convergency criteria.
7.2.1 MLE Based on the Newton-Raphson Algorithm

The log-likelihood function of the parameter vector \( \alpha = (\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_k, \beta_k) \) for the \( n \) observations is

\[
l(\alpha) = n \left[ \sum_{i=1}^{k} \log \Gamma(\alpha_i + \beta_i) - \log \Gamma(\alpha_i) - \log \Gamma(\beta_i) \right]
+ (\alpha_1 - 1) \log \prod_{i=1}^{n} x_{i1} + \beta_1 \log \prod_{i=1}^{n} (1 - x_{i1})
+ \sum_{j=2}^{k-1} \left[ (\alpha_j - 1) \log \prod_{i=1}^{n} \frac{x_{ij}}{1 - \sum_{m=1}^{j-1} x_{im}} + \beta_j \log \prod_{i=1}^{n} \frac{1 - \sum_{m=1}^{j-1} x_{im}}{1 - \sum_{m=1}^{j-1} x_{im}} \right]
+ (\alpha_k - 1) \log \prod_{i=1}^{n} x_{ik} + (\beta_k - 1) \log \prod_{i=1}^{n} \left( \frac{1 - \sum_{m=1}^{k-1} x_{im}}{1 - \sum_{m=1}^{k-1} x_{im}} \right).
\]

The Newton-Raphson method is used in the following iteration to calculate the MLE of the parameter vector (Narayanan [1990]):

\[
\begin{pmatrix}
\hat{\alpha}_1 & \hat{\beta}_1 \\
\vdots & \vdots \\
\hat{\alpha}_k & \hat{\beta}_k
\end{pmatrix}_{(i)}
= \begin{pmatrix}
\hat{\alpha}_1 & \hat{\beta}_1 \\
\vdots & \vdots \\
\hat{\alpha}_k & \hat{\beta}_k
\end{pmatrix}_{(i-1)}
+ \begin{pmatrix}
\text{var}(\hat{\alpha}_1) & \cdots & \text{cov}(\hat{\alpha}_1, \hat{\beta}_k) \\
\vdots & \ddots & \vdots \\
\text{cov}(\hat{\beta}_k, \hat{\alpha}_1) & \cdots & \text{var}(\hat{\beta}_k)
\end{pmatrix}_{(i-1)} \begin{pmatrix}
g_1(\hat{\alpha}_1, \hat{\beta}_1) \\
g_2(\hat{\alpha}_1, \hat{\beta}_1) \\
\vdots \vdotss \\
g_{2k-1}(\hat{\alpha}_k, \hat{\beta}_k) \\
g_{2k}(\hat{\alpha}_k, \hat{\beta}_k)
\end{pmatrix}_{(i-1)}
\]
where $\hat{\alpha}(0) = (\hat{\alpha}_1(0), \hat{\beta}_1(0), \ldots, \hat{\alpha}_k(0), \hat{\beta}_k(0))'$ are the initial estimates. The gradient vector is

$$g = \nabla l(\alpha) = $$

$$\begin{pmatrix}
n\Psi(\alpha_1 + \beta_1) - n\Psi(\alpha_1) + \log \prod x_{i1} \\
n\Psi(\alpha_1 + \beta_1) - n\Psi(\beta_1) + \log \prod (1 - x_{i1}) \\
n\Psi(\alpha_2 + \beta_2) - n\Psi(\alpha_2) + \log \prod \frac{x_{i2}}{1-x_{i1}} \\
n\Psi(\alpha_2 + \beta_2) - n\Psi(\beta_2) + \log \prod \frac{1-x_{i1} - x_{i2}}{1-x_{i1}} \\
\vdots \\
n\Psi(\alpha_j + \beta_j) - n\Psi(\alpha_j) + \log \prod \frac{x_{ij}}{1-\sum_{m=1}^{j-1} x_{im}} \\
n\Psi(\alpha_j + \beta_j) - n\Psi(\beta_j) + \log \prod \frac{1-\sum_{m=1}^{j-1} x_{im}}{1-\sum_{m=1}^{j-1} x_{im}} \\
\vdots \\
n\Psi(\alpha_k + \beta_k) - n\Psi(\alpha_k) + \log \prod \frac{x_{ik}}{1-\sum_{m=1}^{k-1} x_{im}} \\
n\Psi(\alpha_k + \beta_k) - n\Psi(\beta_k) + \log \prod \frac{1-\sum_{m=1}^{k-1} x_{im}}{1-\sum_{m=1}^{k-1} x_{im}}
\end{pmatrix}.$$

The information matrix $I$ is

$$\begin{pmatrix}
I_{\alpha_1\beta_1} & 0 & \ldots & 0 \\
0 & I_{\alpha_2\beta_2} & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & I_{\alpha_k\beta_k}
\end{pmatrix}.$$

The $2 \times 2$ matrix $I_{\alpha_i\beta_i}$ is

$$\begin{pmatrix}
n\Psi'(\alpha_i) - n\Psi'(\alpha_i + \beta_i) & -n\Psi'(\alpha_i + \beta_i) \\
-n\Psi'(\alpha_i + \beta_i) & n\Psi'(\beta_i) - n\Psi'(\alpha_i + \beta_i)
\end{pmatrix},$$

where $\Psi$ is the digamma function.

The variance-covariance matrix is the inverse of the Fisher information matrix. Therefore, the
variance-covariance matrix has the block diagonal following form:

\[
\begin{pmatrix}
V_{\alpha_1\beta_1} & 0 & \ldots & 0 \\
0 & V_{\alpha_2\beta_2} & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & V_{\alpha_k\beta_k}
\end{pmatrix},
\]

where \(V_{\alpha_i\beta_i}\) is a \(2 \times 2\) matrix:

\[
c \begin{pmatrix}
n\Psi'(\beta_i) - n\Psi'(\alpha_i + \beta_i) & n\Psi'(\alpha_i + \beta_i) \\
n\Psi'(\alpha_i + \beta_i) & n\Psi'(\alpha_i) - n\Psi'(\alpha_i + \beta_i)
\end{pmatrix},
\]

and the constant \(c\) is \((n^2 [\Psi'(\alpha_i)\Psi'(\beta_i) - \Psi'(\alpha_i)\Psi'(\alpha_i + \beta_i) - \Psi'(\beta_i)\Psi'(\alpha_i + \beta_i)])^{-1}\).

7.2.2 Direct Proof of Global Convexity of Log Likelihood Function of the Generalized Dirichlet

The digamma function is denoted by \(\psi(x)\), which is the first derivative of log \(\Gamma(x)\). The trigamma function denoted by \(\psi_1(x)\) is the second derivative of log \(\Gamma(x)\).

Although the global concavity of the likelihood function could indirectly be seen from the fact that the generalized Dirichlet distribution belongs to the exponential family, no direct proof of global concavity has been given. We will directly prove the global concavity by the inequality (Ronning, 1986).

\[
\psi_1(x) < a\psi_1(ax), \quad x > 0, \quad 0 < a < 1.
\]

A sufficient condition for global concavity of the log likelihood function is that \(A = -\frac{\partial^2 l}{\partial a \partial a}\) is positive definite. The first proof that \(A\) is positive definite is based on the \(i\)th principal submatrix.

**Proof.** Since \(A\) is a symmetric matrix, the necessary and sufficient condition for \(A\) to be a positive definite matrix is that the determinant of the \(i\)th principal submatrix is positive. Let \(A_i\) be the \(i\)th
principal submatrix.

Let \( b_i = \Pi_{j=1}^{(i-1)/2} (\psi_i(\alpha_i) - \psi_1(\alpha_j + \beta_j))(\psi_1(\beta_j) - \psi_1(\alpha_j + \beta_j)) - \psi_i(\alpha_i)(\psi_1(\beta_j) - \psi_1(\alpha_i))(\psi_1(\beta_j)). \)

Then

\[
\det(A_i) = \begin{cases} 
(\psi_1(\alpha_i) - \psi_1(\alpha_i + \beta_i))b_i, & \text{if } i \text{ is odd,} \\
 b_i, & \text{if } i \text{ is even.}
\end{cases}
\]

When \( x = \alpha_j + \beta_j \) and \( a = \alpha_j/x, \psi_1(x) = \psi_1(\alpha_j + \beta_j) < a\psi_1(ax) = a\psi_1(\alpha_j) < \psi_1(\alpha_j), i = 1, \ldots, k. \) When \( x = \alpha_j + \beta_j \) and \( a = \beta_j/x, \psi_1(x) = \psi_1(\alpha_j + \beta_j) < a\psi_1(ax) = a\psi_1(\beta_j) < \psi_1(\beta_j), i = 1, \ldots, k. \)

Hence

\[
(\psi_1(\alpha_i) - \psi_1(\alpha_i + \beta_i))(\psi_1(\beta_j)) - \psi_1(\alpha_j + \beta_j) - \psi_1(\alpha_j)\psi_1(\beta_j) \\
\geq \left( 1 - \frac{\alpha_j + \beta_j}{\alpha_j + \beta_j} - \frac{\beta_j}{\alpha_j + \beta_j} \right) \psi_1(\alpha_j)\psi_1(\beta_j) = 0,
\]

so \( \det(A_i) \) is always positive.

The second proof that \( A \) is positive definite is based on the characteristic roots.

**Proof.** Since \( A \) is a symmetric matrix, the necessary and sufficient condition for \( A \) to be a positive definite matrix is that the characteristic roots of \( A \) are all positive. Let \( B_i(i = 1, 2, \ldots, k) \) be the matrix.

\[
\begin{bmatrix}
\psi_1(\alpha_i) - \psi_1(\alpha_i + \beta_i) & -\psi_1(\alpha_i + \beta_i) \\
-\psi_1(\alpha_i + \beta_i) & \psi_1(\beta_i) - \psi_1(\alpha_i + \beta_i)
\end{bmatrix}
\]

The characteristic roots of the matrix \( B_i \) are \( \lambda_{i1} \) and \( \lambda_{i2} \). It is easy to prove that both roots are positive. The characteristic roots of the matrix \( A \) are the characteristic roots of \( B_i(i = 1, 2, \ldots, k) \), \( \lambda_{i1} \) and \( \lambda_{i2}(i = 1, 2, \ldots, k) \), thus all \( 2k \) characteristic roots of \( A \) are positive.
7.2.3 The Initial Values of Iteration

The estimates from method of moments was suggested for the initial values of the Newton-Raphson method by [Dishon and Weiss (1980)] for the parameter estimation of the beta distribution. [Ronning (1986)] has shown that the initial values from the moment method led to negative values during the iterations of the maximum likelihood estimation of Dirichlet distributions.

Ronning’s suggestion is to set all $\alpha_j = \min\{X_{ij}\}, i = 1, \ldots, n, j = 1, \ldots, k$ to guarantee that the gradient vector is always positive for the parameter estimation of the Dirichlet distribution. Our experience is that this is not only true for the Dirichlet distribution and but also true for the generalized Dirichlet distribution. Thus the initial values for the iteration of the maximum likelihood method for the generalized Dirichlet distribution is to set all $\alpha_j = \min\{X_{ij}\}, i = 1, \ldots, n, j = 1, \ldots, k$.

7.2.4 Iteration Contingency Criteria

The statistic $S$ for the convergence is given by

$$S = \begin{pmatrix}
g_1(\hat{\alpha}) \\
g_2(\hat{\alpha}) \\
\vdots \\
g_{2k-1}(\hat{\alpha}) \\
g_{2k}(\hat{\alpha})
\end{pmatrix}^T \begin{pmatrix}
\text{var}(\hat{\alpha}_1) & \cdots & \text{cov}(\hat{\alpha}_1, \hat{\beta}_k) \\
\cdots & \ddots & \cdots \\
\text{cov}(\hat{\beta}_k, \hat{\alpha}_1) & \cdots & \text{var}(\hat{\beta}_k)
\end{pmatrix} \begin{pmatrix}
g_1(\hat{\alpha}) \\
g_2(\hat{\alpha}) \\
\vdots \\
g_{2k-1}(\hat{\alpha}) \\
g_{2k}(\hat{\alpha})
\end{pmatrix}.$$ 

Since $S$ has a quadratic form, it can be considered similar to a chi-squared random variable and its value can be applied as a criteria for convergence. The gradient vector of MLE is always zero because the maximum likelihood maximizes the loglikelihood. At the beginning of the iteration, the estimated vector is a little more different from MLE, the gradient vector is far from zero, so $S$ is relatively big at the beginning. When the iteration continued, the estimated parameter vector becomes closer to the MLE of the parameter vector. Thus, the element of the gradient vector generally becomes smaller, and closer to zero. Hence, $S$ becomes smaller as the iteration
is continued. The iteration is continued until $S$ becomes less than $\chi^2_{2k}(c)$ for a fixed $c$ in the right tail of the chi-squared distribution with $2k$ degrees of freedom. This approach was also used by [Choi and Wette (1967)] for the gamma distribution and by [Narayanan (1991)] for the Dirichlet distribution.

In practice, in order to avoid a large number of iterations, we can specify a maximum number of iterations. The iteration is continued until $S$ becomes less than $\chi^2_{2k}(c)$ for a fixed significance level $c$ or the number of iterations exceeds the maximum number of iteration we have set. In this paper, $c$ is set to 0.0001 and the maximum number of iterations is set to 100.

We propose another convergence criteria based on the partitioned variance matrix. Since the variance-covariance matrix $V$ is the inverse of the Fisher information matrix for the generalized Dirichlet distribution, we divide the parameter vector into $k$ mutually exclusive parts. The criteria based on the partitioned variance matrix is to guarantee that the iteration of each part of parameter vector $(\alpha_i, \beta_i)$ is continued until each $S_i$ becomes less than $\chi^2_{2}(v)$ for a fixed $v$ in the right tail of the chi-squared distribution with 2 degrees of freedom or exceeds the maximum number of iterations we have set. The criteria based on the whole variance matrix is to guarantee that the iteration of the whole parameter vector is continued until $S$ becomes less than $\chi^2_{2k}(v)$ for a fixed $v$ or the number of the iterations exceeds the maximum number.

7.3 Generalized Dirichlet Distribution Model vs. Dirichlet Distribution Model

For $i = 1, \ldots, n$, let $x_i = (x_{i1}, \ldots, x_{i(k+1)})$ be $n$ independent observations from the generalized Dirichlet distribution with parameters $(\alpha_1, \beta_1, \ldots, \alpha_k, \beta_k)$. The $k$-dimensional generalized Dirichlet distribution has $2k$ parameters, but the $k$-dimensional Dirichlet distribution has only $k+1$ parameters. Given a data set, based on principle of the parsimony, one may want to test whether a generalized Dirichlet distribution model is better than a Dirichlet distribution model to fit the observed data. The hypothesis of interest is

$$H_0 : \alpha \in \omega \text{ versus } H_1 : \alpha \in \Omega \cap \omega^c,$$
where $\alpha$ is the parameter vector, $\omega$ is the restricted $k + 1$ dimensional parameter space for the Dirichlet distribution, and $\Omega$ is the full $2k$ parameter space for the generalized Dirichlet distribution. Let $\hat{\alpha}$ denote the maximum likelihood estimator when the parameter space is the full space $\Omega$ and let $\tilde{\alpha}$ denote the maximum likelihood estimator when the parameter space is the reduced space $\omega$. Here $\tilde{\alpha}$ is the restricted maximum estimate, which can be obtained by the maximum likelihood estimates based on the assumption that sampled population has the Dirichlet distribution, because the constraints make the generalized Dirichlet distribution into the Dirichlet distribution.

7.3.1 The Likelihood Ratio Test

For a generalized Dirichlet distribution with the parameter vector $(\alpha_1, \beta_1, \cdots, \alpha_{k-1}, \beta_{k-1})$, it becomes a Dirichlet distribution when $\beta_i = \alpha_{i+1} + \beta_{i+1}, i = 1, 2, \cdots, k - 1$.

Thus, the null and alternative hypotheses are given by

$H_0 : \beta_i = \alpha_{i+1} + \beta_{i+1}, i = 1, 2, \cdots, k - 1$

$H_1 : \text{at least one } \beta_i \neq \alpha_{i+1} + \beta_{i+1}, i = 1, 2, \cdots, k - 1$

The likelihood ratio test statistic is

$$X^2_{LR} = -2 \log \Lambda \overset{D}{\rightarrow} \chi^2(k - 1),$$

where

$$\Lambda = \frac{L(\tilde{\alpha})}{L(\hat{\alpha}).}$$

When the null hypothesis is true, $-2 \log \Lambda$ has approximately a chi-squared distribution with $k - 1$ degrees of freedom. A proof of the above result can be found in [Rao (1973)].

7.3.2 The Wald Test

The $k$-dimensional generalized Dirichlet distribution has $2k$ parameters, but the $k$-dimensional Dirichlet distribution has only $k + 1$ parameters. Therefore, there are $k - 1$ constraints from the
generalized Dirichlet distribution to the Dirichlet distribution. The corresponding constraint matrix \( R \) is given by

\[
\begin{pmatrix}
0 & 1 & -1 & 0 & \ldots & 0 & 0 \\
0 & 0 & 0 & 1 & -1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \ddots & 0 & 1 \\
\end{pmatrix}
\]

\((k-1) \times 2k\).

The null and alternative hypotheses are given by

\[ H_0 : R\alpha = 0 \quad vs \quad H_1 : R\alpha \neq 0 \]

where \( R \) is the \((k - 1) \times 2k\) constraint matrix and \( \alpha \) is the parameter vector of the \( k \)-dimensional generalized Dirichlet Distribution. The rank of the constraint matrix \( R \) is \( k - 1 \). The Wald test statistic is

\[
X^2_W = (R\hat{\alpha})^T (RI^{-1}(\hat{\alpha})R^T)^{-1} (R\hat{\alpha})
\]

where \( I \) is the Fisher information matrix. When the null hypothesis is true, \( X^2_W \) has approximately a chi-squared distribution with \( k - 1 \) degrees of freedom.

7.3.3 Rao’s Score Test

Rao’s score test is also referred to as the Lagrange multiplier test. The Rao’s score test statistic is

\[
X^2_R = s(\hat{\alpha})^T (I(\hat{\alpha}))^{-1} s(\hat{\alpha})
\]

where \( s(\alpha) = \frac{1}{n} \frac{\partial L(\alpha)}{\partial \alpha} \) and \( L(\alpha) \) is the likelihood function. If the null hypothesis is true, the restricted maximum likelihood estimates \( \hat{\alpha} \) will be near the unrestricted maximum likelihood estimates. Here \( s(\hat{\alpha}) = 0 \) is always true because the unrestricted maximum likelihood maximizes the log-likelihood. The bigger the Fisher score, the more likely the null hypothesis is not true. When the null hypothesis is true, \( X^2_R \) has approximately a chi-squared distribution with \( k - 1 \) degrees of freedom.
freedom (Rao, 1973).

7.3.4 Bootstrap Hypothesis Testing

A comprehensive discussion of bootstrap methods can be found in the books by Efron and Tibshiranni (1993), and Davison and Hinkley (1997). When the null hypothesis is true, the likelihood ratio test statistic, the Wald test and Rao’s score test statistic each have approximately a chi-squared distribution. However, the general theory provides no practical guidance on how large the sample size should be for the chi-squared approximation to be reliable. Bootstrap methods provide an alternative approach to assessing the variance of an estimator and, in many cases, are more accurate for small samples.

To test a null hypothesis \( H_0 \), we specify a test statistic \( T \) such that large absolute value of \( T \) are evidence against \( H_0 \). Let \( T \) have observed value \( t \) which is calculated from the observed sample. We generally want to calculate the p-value

\[
p = P(|T| > t|H_0).\]

We need the distribution of \( T \) when the null hypothesis is true to evaluate this probability. However, the null hypothesis is that sample is from the Dirichlet family. We do not know the parameter vector in advance, so we use the MLE to replace the population parameter vector to construct the parametric bootstrap testing.

1. Evaluate the observed test statistic \( t = T(x_1, \cdots, x_n) \).
2. Find the maximum likelihood estimate of the parameter vector.
3. Generate \( B \) random samples from the target distribution with MLE as its the parameter vector and compute \( t^*_1, t^*_2, \cdots, t^*_n \).
4. Compute the Monte Carlo p-value.

For a given data set, the procedure of the parametric bootstrap method to the likelihood ratio test, the Wald test and Rao’s score test are as follows.
Likelihood Ratio Test Based on Parametric Bootstrapping

Suppose that $\theta$ is the parameter of interest, and $\hat{\theta}$ is an estimator of $\theta$.

Then the bootstrap likelihood ratio test is obtained as follows.

1. For a given data set, estimate the parameter vector $\hat{\theta}$ of the Dirichlet distribution and compute the likelihood ratio test statistic $X_{LR}^2$ and the corresponding $p$-value.

2. For each bootstrap replicate, indexed $b = 1, \ldots, B$:
   (a) Generate sample $x^{*(b)} = x_1^*, x_2^*, x_3^*, \ldots, x_n^*$ from a Dirichlet distribution with the parameter vector $\hat{\theta}$ from the first step.
   (b) Compute for the $b^{th}$ replicate the likelihood ratio test statistic $X_{LR}^{2(g)}$ from the $b^{th}$ bootstrap sample.

3. Set $N$ equal to the number of the values $g$ for which $X_{LR}^{2(g)} \geq X_{LR}^2$.

4. Compute $\hat{p} = \frac{N}{B}$.

5. Reject $H_0$ at the significance level $\alpha$ if $\hat{p} \leq \alpha$.

Wald Test Based on Parametric Bootstrapping

Suppose that $\theta$ is the parameter of interest, and $\hat{\theta}$ is an estimator of $\theta$.

Then the bootstrap Wald test is obtained as follows.

1. For a given data set, estimate the parameter vector $\hat{\theta}$ of the Dirichlet distribution and compute the Wald test statistic $X^2_W$ and the corresponding $p$-value.

2. For each bootstrap replicate, indexed $b = 1, \ldots, B$:
   (a) Generate sample $x^{*(b)} = x_1^*, x_2^*, x_3^*, \ldots, x_n^*$ from a Dirichlet distribution with the parameter vector $\hat{\theta}$ from the first step.
   (b) Compute for the $b^{th}$ replicate the likelihood ratio test statistic $X_{W}^{2(g)}$ from the $b^{th}$ bootstrap sample.

3. Set $N$ equal to the number of the values $g$ for which $X_{W}^{2(g)} \geq X_{W}^2$.

4. Compute $\hat{p} = \frac{N}{B}$.

5. Reject $H_0$ at the significance level $\alpha$ if $\hat{p} \leq \alpha$. 
Rao’s Score Test Based on Parametric Bootstrapping

Suppose that \( \theta \) is the parameter of interest, and \( \hat{\theta} \) is an estimator of \( \theta \).

Then the bootstrap Rao’s score test is obtained as follows.

1. For a given data set, estimate the parameter vector \( \hat{\theta} \) of the Dirichlet distribution and compute the Rao’s score test statistic \( X^2_R \).

2. For each bootstrap replicate, indexed \( b = 1, \ldots, B \):
   (a) Generate sample \( x^{(*)}_b = x^*_1, x^*_2, x^*_3, \ldots, x^*_n \) from a Dirichlet distribution with the parameter vector \( \hat{\theta} \) from the first step.
   (b) Compute for the \( b^{th} \) replicate the likelihood ratio test statistic \( X^{2(g)}_R \) from the \( b^{th} \) bootstrap sample.

3. Set \( N \) equal to the number of the values \( g \) for which \( X^{2(g)}_R \geq X^2_R \).

4. Compute \( \hat{p} = \frac{N}{B} \).

5. Reject \( H_0 \) at significance level \( \alpha \) if \( \hat{p} \leq \alpha \).
8.1 Applications of the Goodness-of-Fit Test of Dirichlet and Generalized Dirichlet Distribution

Example 8.1 (Proportions of Serum Proteins).

The data set came from a study of spurious correlations among proportions by Mosimann (1962) and is assumed to follow the Dirichlet distribution by Mosimann (1962) and Narayanan (1990). The maximum likelihood estimator of the parameter vector was given by Narayanan (1990). Dr. Jean-Marie Demers and Mr. Roch Carbonneau of the Department of Biology, University of Montreal, provided the original data set of the measurements of serum proteins in three-week-old white Peking ducklings. Each set of three measurements represents seven or twelve individual ducklings. The ducklings in each set were fed the same diet, but different sets represents different diets. The paper-electrophoresis with subsequent readings were used to produce the following proportional measurements. $P_1$, $P_2$, and $P_3$ are the proportions of pre-albumin, albumin and globulins of serum proteins. We will apply the proposed goodness-of-fit tests to the proportion of serum protein data.

The $p$-value of the energy, three triangle goodness-of-fit tests for the generalized Dirichlet distribution are 0.765, 0.445, 0.798, and 0.823, respectively. Hence we cannot reject the null hypothesis that the data come from a generalized Dirichlet distribution. We apply the generalized Dirichlet distribution with the parameter vector $(\alpha_1, \beta_1, \alpha_2, \beta_2)$ to fit data from the duck study.

The $p$-value of the energy, three triangle goodness-of-fit tests for the Dirichlet distribution are 0.413, 0.445, 0.693, and 0.752, respectively. Thus we cannot reject the null hypothesis for data that came from a Dirichlet distribution. We can apply the Dirichlet distribution $Dir(\alpha_1, \alpha_2, \alpha_3)$ to fit data from the duck study.

According to the principle of parsimony, is it necessary to apply the generalized Dirichlet distribution model rather than the simpler Dirichlet distribution model? We construct the following
hypothesis:

\[ H_0 \quad \beta_1 = \alpha_2 + \beta_2 \quad vs \quad H_\alpha \quad \beta_1 \neq \alpha_2 + \beta_2. \]

When the null hypothesis holds, the generalized Dirichlet distribution is the Dirichlet distribution. The \( p \)-value of the likelihood ratio, Wald, and Rao’s score test are 0.797, 0.800, 0.795, respectively, so the generalized Dirichlet distribution is not necessary for the duck data.

For the distance covariance goodness-of-fit test, according to the property of the complete neutrality of the Dirichlet distribution, we need to test the independence between \( p_1 \) and \( \frac{p_2}{1-p_1} \) and independence between \( p_2 \) and \( \frac{p_1}{1-p_2} \); so we construct two hypotheses:

\[ H_{01} \quad p_1 \text{ is independent of } \frac{p_2}{1-p_1}, \]
\[ H_{02} \quad p_2 \text{ is independent of } \frac{p_1}{1-p_2}. \]

We apply the Bonferroni correction to control the familywise type I error at 0.05 and finally we cannot reject the null hypothesis that the duck data \((p_1, p_2, p_3)\) come from the Dirichlet distribution at the significance level of 0.05.

8.2 Application to Goodness-of-Fit Testing of Dirichlet Regression Models

Campbell and Mosimann (1987) originally proposed the Dirichlet distribution for compositional data. Campbell and Mosimann extended the Dirichlet distribution to the Dirichlet covariate models, where the composition of data from the Dirichlet distribution are influenced by some common covariates. In this model, let \( y_i = (y_{i1}, \ldots, y_{id}) \) be the \( i \)-th observed composition from the Dirichlet distribution with the parameter vector \((\alpha_{i1}, \ldots, \alpha_{id}), i = 1, \ldots, n. \) Let \( x_i = (x_{i1}, \ldots, x_{im}), i = 1, 2, \ldots, n \) be the corresponding observed covariate vector with \( m \) components.

The parameters \( \alpha_i = (\alpha_{i1}, \ldots, \alpha_{id}), \ i = 1, \ldots, n \) of the conditional Dirichlet distribution is a function of the corresponding observed covariate vector \( x_i = (x_{i1}, \ldots, x_{id}), \ i = 1, \ldots, n. \) For a given covariate row vector \( x_i = (x_{i1}, \ldots, x_{id}) \), each parameter \( \alpha_{ij} \) is a positive-valued function
The response variable vector follows a conditional Dirichlet distribution,

\[
(y_1, \cdots, y_d)| (x_{i1}, \cdots, x_{im}) \sim Dir(\alpha_{i1}, \cdots, \alpha_{id}).
\]

The benchmark application of the Dirichlet regression model is the Arctic lake sediments, introduced by Coakley and Rust (1968) and adapted by Aitchison (1986). It consists of composition of sand, silt and clay for 39 sediment observations at different water depths in the lake. Each observation represents the composition of sand, silt and clay at different water depths. The response (dependent) variable is a vector, which contains three compositional variables of sand, silt, and clay. The response variable has 39 compositions, so the response matrix \( \mathbf{y} \) is a \( 39 \times 3 \) matrix. The covariate is the water depth and the corresponding covariate matrix is a \( 39 \times 1 \) matrix with Figure 8.1.

First, we take a look at the scatterplot of three observed compositional components (sand, silt and clay) and water depth in Figure 8.1.

Figure 8.1 is the scatterplot of sand, silt and clay with the water depth, which shows that there is some dependency between sediment composition and water depth. We can find that negative association between the proportion of the sand and water depths; the deeper the water, the smaller proportions of the sand. There is a positive association between the clay and water depth. Aitchison (1986) proposed the following log ratio model. The covariate in this model is the logarithm of the water depth.

\[
\log (\text{sand/clay}) = 9.70 - 2.74 \log (\text{depth}),
\]

\[
\log (\text{silt/clay}) = 4.80 - 1.10 \log (\text{depth}).
\]

However, Hijazi and Jerbigan (2007) stated that the application of the logarithm of the water is based only on Aichilson’s modeling method but not on any geological background, so the log transformation does not make the water depth more interpretable.

Hijazi and Jerbigan (2007) proposed the following Dirichlet quadratic regression model. The
sediments compositional data come from the conditional Dirichlet distribution with the parameter vector \((\alpha_{i1}, \alpha_{i2}, \alpha_{i3}), i = 1, \cdots, 39\). In order to keep the parameter positive, the model applies the exponentiation of the quadratic regression term.

\[
\begin{align*}
\alpha_{i1} &= \exp \left(5.240 - 0.072 \text{depth} + 0.001 \text{depth}^2\right), \\
\alpha_{i2} &= \exp \left(3.426 - 0.203 \text{depth} + 0.011 \text{depth}^2\right), \\
\alpha_{i3} &= \exp \left(3.635 - 0.391 \text{depth} + 0.013 \text{depth}^2\right).
\end{align*}
\]

Hijazi and Jerbigan (2007) showed that according to the criteria of the sum of compositional errors by Aitchison (1986) and the total variability evaluated using the moving windows technique by Hijazi and Jerbigan (2007), the Dirichlet quadratic regression model has better performance than the log ratio model. The Dirichlet regression model is an informative alternative to the log ratio model.

Hijazi and Jerbigan (2007) also evaluated the model performance based on fitting of the marginal distribution of the sediment composition in the Arctic lake data.

Figure 8.2 is the observed and fitted compositions in the Arctic lake sediments, which shows that the Dirichlet quadratic model fit each compositional component well.

Figure 8.3 is the residual plot for each of three compositional components in the Arctic lake sediments, which also shows that the Dirichlet quadratic model fit each compositional component (marginal distribution) of the sediments well. However, a model that fits the marginal distribution of the sediment composition well, does not necessarily fit the joint distribution of the sediment composition well.

A similar example is that for a sample from a specific multivariate distribution, univariate normal distributions fit each component of the sample well, which does not imply that the sample as a whole came from a multivariate normal distribution.

We propose a new test which is based on transforming samples from the conditional \(\text{Dir}(\alpha_{i1}, \cdots, \alpha_{id})\) to the standard \(\text{Dir}(1, \cdots, 1)\) to evaluate the overall performance of fitting all composi-
Suppose $\mathbf{y}_i = (y_{i1}, \ldots, y_{id})$ is an observed observation from the Dirichlet distribution with the parameter vector $(\alpha_{i1}, \ldots, \alpha_{id}), i = 1, \ldots, n$. The parameters $(\alpha_{i1}, \ldots, \alpha_{id}), i = 1, 2, \ldots, n$ of the Dirichlet distribution are based on the corresponding observed covariate vector $\mathbf{x}_i = (x_{i1}, \ldots, x_{im})$.

1. Each observation from the conditional Dirichlet distribution $(y_{i1}, \ldots, y_{id})$ with the parameter vector $(\alpha_{i1}, \ldots, \alpha_{id})$ was transformed to the corresponding observation $(y^*_{i1}, \ldots, y^*_{id})$ following the standard Dirichlet distribution.

2. Conduct the goodness-of-fit test of the transformed sample from the standard Dirichlet distribution.

Since we do not know the parameter vector $(\alpha_{i1}, \ldots, \alpha_{id}), i = 1, \ldots, n$ of the conditional Dirichlet distribution, we apply the parameter estimates from the Dirichlet regression model to transform the corresponding observation.

However, the transformed observations based on the parameter estimates do not exactly follow the standard Dirichlet distribution.

We apply the following parametric bootstrap method based on the distance covariance goodness-of-fit test to show that replacing the conditional Dirichlet parameter vector with the corresponding estimated parameter is practical and has a very acceptable type I error.

1. Compute the estimated Dirichlet parameter vector $(\hat{\alpha}_{i1}, \ldots, \hat{\alpha}_{id}), i = 1, \ldots, n$ for each observation from the proposed Dirichlet regression model.

2. Set $G = 0$.

3. For each replicate, indexed $l = 1, \ldots, N$:
   (a) Draw a random data point from $\operatorname{Dir}((\hat{\alpha}_{i1}, \ldots, \hat{\alpha}_{id}), i = 1, \ldots, n$.
   (b) Pool all $n$ random data points to get the $l^{th}$ sample.
   (c) Apply the Dirichlet regression model to the $l^{th}$ random sample to obtain the estimated conditional Dirichlet parameter vector.
   (d) Transform each vector of proportions of the $l^{th}$ sample by the corresponding estimated
parameter vector from the above step.

(e) Conduct the distance covariance goodness-of-fit test of the Dirichlet distribution for the transformed sample.

(f) Increase $G$ by 1 if the above test is rejected at the fixed significance level.

3. Compute the achieved significance level $\hat{p} = \frac{G}{N}$.

We apply the above bootstrap method based on the distance covariance goodness-of-fit test to the Arctic lake data, set $N = 1000$ and the fixed significance level 0.05, and finally the achieved significance level $\hat{p} = 0.045$. So we can apply the goodness-of-fit test based on the transformed sample for the Dirichlet regression model evaluation and selection.

In the end, we apply the estimated conditional Dirichlet parameter vector to the Arctic lake data and obtain the transformed unconditional Dirichlet distributed data, then apply the distance covariance goodness-of-fit test to the transformed unconditional Dirichlet distributed data. The null hypothesis that the Arctic lake data come from the proposed Dirichlet quadratic regression model is rejected at the significance level of 0.05. The Dirichlet quadratic regression model fits the marginal distribution of the composition well, but it fails to fit the joint distribution of all three compositional components well. We can try different Dirichlet polynomial terms, or add more covariates, or apply some variable transformations in the model. We propose the generalized Dirichlet regression model to fit the Arctic lake data. For the Dirichlet distribution, one component is always negatively correlated with other components. The generalized Dirichlet distribution (Kotz and Johnson, 2000) has a more flexible covariance structure to fit the compositional data, so the generalized Dirichlet distribution can be more practical and useful. For example, two components of the generalized Dirichlet distribution can be positively correlated.

Let $S_2$ and $S_1$ be the corresponding sand and silt variables in the transformed data. Let $T_1 = S_1, T_2 = \frac{S_2}{1-S_1}, V_1 = S_2, V_2 = \frac{S_1}{1-S_2}$. We take a look at the scatterplots of $T_1$ and $T_2$, and $V_1$ and $V_2$. A scatterplot between $T_1$ and $T_2$ in Figure 8.4 suggests that $T_1$ and $T_2$ seem independent, with correlation coefficient -0.089. The $p$-value of the distance covariance independence test is 0.678, so we cannot reject the null hypothesis that $T_1$ and $T_2$ are independent.
A scatterplot between $V_1$ and $V_2$ in Figure 8.5 suggests that $V_1$ and $V_2$ are moderately associated, with correlation coefficient 0.646. The $p$-value of the distance covariance independence test [Rizzo 2008] is 0.001, so we reject the null hypothesis of independence of $V_1$ and $V_2$.

If $T_1$ and $T_2$ are independent, and $V_1$, $V_2$ are also independent, then $S_1$ and $S_2$ follow the Dirichlet distribution [Darroch and Ratcliff 1971].

When $T_1$ and $T_2$ are independent, but $V_1$ and $V_2$ are not independent, it is a typical feature of the generalized Dirichlet distribution. So we can consider applying the generalized Dirichlet distribution to fit the transformed Arctic lake data. Furthermore, we can consider applying the generalized Dirichlet regression model to fit the original Arctic lake data. The composition follows the conditional generalized Dirichlet distribution. The parameter of the conditional generalized Dirichlet distribution is a function of the corresponding covariates. We will consider the generalized Dirichlet regression model in future research. For other books on analysis of compositional data, please see [Gerald and Raimon 2013], and [Pawlowsky-Glahn and Tolosana-Delgado 2013].

8.3 Application to the Influence Diagnostics of the Dirichlet Regression Models

Different methods have been proposed to detect the outliers in compositional data by Barceló, Pawlowsky, and Grunsky (1996) and by Baxter (1999). Hijazi (2005) stated these methods are suitable for unconditional compositions. For the conditional composition, the joint distribution of the quantile residuals can be applied to identify the compositions with large Mahalanobis distance as outliers [Marida and Bibby 1979]. When compositions are influenced by some covariates, residuals can be used to identify outliers. We can apply Dirichlet standardized distribution transformation (4.2) we have proposed in this dissertation to transform the conditional Dirichlet distributed data(different compositions has different parameter vectors) into the unconditional Dirichlet distributed data(all transformed compositions have the same parameter vectors), then apply the outlier identification methods based on the unconditional Dirichlet distribution to identify the potential outliers. For example, one can apply the joint distribution of the transformed compositional data, which has an unconditional Dirichlet distribution, to identify the compositions with large Mahalanobis distance as the potential outliers.
Figure 8.1: Arctic lake dataset and corresponding scatterplots.
Figure 8.2: Observed and fitted compositions in the Arctic lake sediments using Dirichlet quadratic regression model.
Figure 8.3: Residual plots of three compositions in the Arctic lake sediments using Dirichlet quadratic model.
Figure 8.4: Scatterplot of $T_1$ and $T_2$ from the transformed Arctic lake sediments using Dirichlet quadratic model.
Figure 8.5: Scatterplot of $V_1$ and $V_2$ from the transformed Arctic lake sediments using Dirichlet quadratic model.
CHAPTER 9 SUMMARY AND FURTHER DIRECTIONS

The Dirichlet distribution is the multivariate Beta distribution. The Dirichlet distribution is a special case of the generalized Dirichlet distribution. We propose a goodness-of-fit test based on energy distances and a goodness-of-fit test based on interpoint distances. We also propose a goodness-of-fit test based on the property of complete neutrality of the Dirichlet distribution and distance covariance tests. A new algorithm based on the distance covariance and permutation to test mutual independence of all components of random vector in arbitrary dimensions is introduced. The formula for transforming a variable following the generalized Dirichlet distribution or Dirichlet distribution into another variable following Dirichlet distribution is constructed.

For the maximum likelihood estimation of the parameters in the generalized distribution, the initial values of the iteration and the criteria of convergence for the Newton-Raphson algorithm has been proposed. The application of the dissertation includes goodness-of-fit tests of the Dirichlet and generalized Dirichlet distribution, model evaluation of the Dirichlet regression analysis, outlier and influential observation identification in the Dirichlet regression model.

Since the distance covariance test has very good performance in measuring the independence and in the goodness-of-fit test in the Dirichlet distribution, we will apply the distance covariance goodness-of-fit test on the copula models. Copulas are a function that joins or couples multivariate distribution functions to their one-dimensional marginal distribution functions (Nelsen, 1999). Copulas enable the modelling of marginal distributions and the dependence structure among the marginal distributions in two steps (Joe, 1997), which make it widely used in economics, hydrology, insurance, reinsurance, finance and biology. Although the goodness-of-fit of the univariate distributions is well documented, the study of goodness-of-fit tests for the copulas emerged recently as a challenging problem (Berg, 2007).

Our future work also includes the derivation of the generalized Dirichlet regression model. The generalized Dirichlet distribution has more flexible variance-covariance structure than the Dirichlet distribution, so the generalized Dirichlet regression model will be more useful than the Dirichlet
regression model. The generalized Dirichlet distribution is widely used in data mining and machine
learning. The parameter estimation based on maximum likelihood methods and other methods will
be derived. Since there is not closed form MLE method, numerical methods will be applied and the
criteria of setting up the initial values for the iteration of the numerical methods will be discussed.
The comparison of the Dirichlet regression and generalized Dirichlet regression model will be
considered in future research.


