# FREQUENTIST-BAYES GOODNESS-OF-FIT TESTS

A Dissertation

by

QI WANG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2011

Major Subject: Statistics

# FREQUENTIST-BAYES GOODNESS-OF-FIT TESTS

A Dissertation

by

QI WANG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

| | |
|---|---|
| Chair of Committee, | Jeffrey D. Hart |
| Committee Members, | Faming Liang |
| | Uschi Müller-Harknett |
| | Ximing Wu |
| Head of Department, | Simon J. Sheather |

August 2011

Major Subject: Statistics

# ABSTRACT

Frequentist-Bayes Goodness-of-fit Tests. (August 2011)

Qi Wang, B.S., Zhejiang University;

M.S., Texas A&M University

Chair of Advisory Committee: Dr. Jeffrey D. Hart

In this dissertation, the classical problems of testing goodness-of-fit of uniformity and parametric families are reconsidered. A new omnibus test for these problems is proposed and investigated. The new test statistics are a combination of Bayesian and score test ideas. More precisely, singletons that contain only one more parameter than the null describing departures from the null model are introduced.

A Laplace approximation to the posterior probability of the null hypothesis is used, leading to test statistics that are weighted sums of exponentiated squared Fourier coefficients. The weights depend on prior probabilities and the Fourier coefficients are estimated based on score tests. Exponentiation of Fourier components leads to tests that can be exceptionally powerful against high frequency alternatives. Comprehensive simulations show that the new tests have good power against high frequency alternatives and perform comparably to some other well-known omnibus tests at low frequency alternatives.

Asymptotic distributions of the proposed test are derived under null and alternative hypotheses. An application of the proposed test to an interesting real problem is also presented.

To Grandma, Mom and Dad

# ACKNOWLEDGMENTS

I would like to take this opportunity to express my gratitude to all the people who have been influential in my time in graduate school. I find that I could not have made it through without the kind support of many people along my journey.

First, I would like to sincerely thank my committee chair, Dr. Jeffrey D. Hart, for being such a great advisor. Thank you for your guidance and support throughout the work leading to this dissertation, and thank you for giving me a great graduate experience and education. I cannot express enough gratitude for what you have done for me. I would also like to thank Dr. Faming Liang, Dr. Uschi Müller-Harknett, and Dr. Ximing Wu, for your time and suggestions while serving on my committee.

Thanks also go to my friends and colleagues, and the department faculty and staff, for making my time at Texas A&M University a great experience.

Last but not least, thanks to my family for their unconditional support and encouragement, and to my fiance for his patience and love. I am indebted to you for everything that you have done for me. I love you all.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

FIGURE Page

# CHAPTER I

## INTRODUCTION

This dissertation is composed of five chapters. Each is separate with its own goal, but the parts are related by the theme of frequentist-Bayes goodness-of-fit tests based on Laplace approximations. We propose goodness-of-fit tests motivated by a combination of Bayesian and score test ideas and apply them for testing simple hypotheses (no unspecified parameters). We then extend this idea to testing goodness-of-fit when the null hypothesis is composite (e.g.,for testing normality or exponentiality), these cases being of more practical interest.

When testing for goodness-of-fit, alternative hypotheses are often vague and an omnibus test is welcome. By an omnibus test, we mean a test that is consistent against essentially all alternatives. In this dissertation, we propose and investigate a new omnibus goodness-of-fit test. To motivate our choice, we start with some background.

We first consider the simple hypothesis. Let $X_1, ..., X_n$ be i.i.d observations with density $f$. We wish to test the null hypothesis $H_0 : f \equiv f_0$, where $f_0$ is some completely specified density. There are many consistent tests for testing $H_0$. The most popular ones are the Kolmogorov-Smirnov (KS) test proposed in 1933 and the Cramér-von Mises (CvM) test proposed by Cramér in 1928 and corrected by Smirnov in 1936. These tests are described in many textbooks and a lot of work has been done on their empirical and asymptotic powers, efficiencies and other properties. Thus, there is now strong evidence that, for moderate sample sizes, only a few types of deviations can be detected by these two tests with substantial power. This feature

---

The journal model is *Journal of the American Statistical Association.*

can be seen in simulations [cf. Quesenberry and Miller (1977), Miller and Quesenberry (1979) and Kim (1992)].

As reviewed by Inglot, Kallenberg and Ledwina (1997), there are some theoretical results due to Neuhaus (1976) and Milbrodt and Strasser (1990) that explain the deficiencies of KS and CvM tests. See also Janssen (1995) for some developments. These results show how the above-mentioned and some other tests distribute their power in the space of all alternatives when the sample size is large. In particular, they show that there are only a few directions of deviations from the null hypothesis for which the tests have reasonable asymptotic power. These directions correspond to some very smooth departures from the null distribution (low-frequency alternatives). Moreover, following from the "principal component representation" of the local asymptotic power, there is only one direction with highest asymptotic power that is possible. In each other direction the power is small. For a "bad" direction, the power is close to the significance level. As a result, Milbrodt and Strasser (1990) concluded that these tests behave very much like a parametric test for a one-dimensional alternative and not like a well-balanced test for higher-dimensional alternatives. Therefore, at least from a local point of view, the tests do not have the omnibus property usually attributed to them.

We also would like to mention the investigation of the relative efficiency of a given test with respect to the Neyman-Pearson test for an alternative of interest. Such an approach for the KS, CvM and other goodness-of-fit tests has been developed by Nikitin (1984, 1995). He used the notion of Bahadur efficiency and has shown that the tests mentioned before are usually less powerful than the Neyman-Pearson test when the alternatives differ from the null only with respect to location or scale. Inglot and Ledwina (1990) arrive at the same conclusion by exploiting the notion of intermediate efficiency. Some related results with regard to Bahadur slopes of goodness-of-fit tests and the local intermediate equivalence can be found in Koning (1992, 1993).

The above mentioned deficiency of the KS and CvM tests caused renewed interest in Neyman's smooth tests for goodness of fit, especially for higher-frequency alternatives. To be specific, we hypothesize that we have a random sample $X_1, ..., X_n$ with probability density function $f$ and cumulative distribution function $F$. Both of these are completely specified. We could apply, as did Neyman, the probability integral transformation. Therefore, it is sufficient to consider tests for uniformity. The "smooth" alternatives to uniformity are defined by

$$g_k(x;\boldsymbol{\theta}) = C(\boldsymbol{\theta}) \exp\{\sum_{i=1}^{k} \theta_i u_i(x)\}, \quad 0 < x < 1, \tag{1.1}$$

where $u_1, u_2, ...$ are an orthonormal system in $L_2([0,1])$ with bounded functions, $\boldsymbol{\theta} = (\theta_1, ..., \theta_k) \in \mathbb{R}^k$ and $C(\boldsymbol{\theta})$ is a constant depending on $\boldsymbol{\theta}$, introduced to ensure that the probability density function integrates to one. Of course, testing for uniformity is equivalent to testing $H_0 : \boldsymbol{\theta} = 0$ against $H_a : \boldsymbol{\theta} \neq 0$. The so-called smooth test statistics are given by

$$N_k = \sum_{j=1}^{k} \left( n^{-1/2} \sum_{i=1}^{n} u_j(X_i) \right)^2, \quad k = 1, 2, .... \tag{1.2}$$

See also Rayner and Best (1989, 2009), Milbrodt and Strasser (1990), Eubank and LaRiccia (1992) and Kaigh (1992) for details.

To enlarge the applicability of the original Neyman's smooth test and to make the test consistent against essentially any alternative, some data-driven versions of Neyman's smooth test have been proposed by Bickel and Ritov (1992), Eubank and LaRiccia (1992), Eubank, Hart and LaRiccia (1993), Ledwina (1994), Kallenberg and Ledwina (1995a) and Fan (1996). Worth special mention due to their fundamental nature are adaptive versions of the Neyman smooth test, which were introduced by

Ledwina (1994). In this work Ledwina proposed that the Schwarz criterion, i.e., BIC, be used to choose the number of components in a Neyman smooth statistic. The selection rule is seen as the first step, followed by the finishing touch of applying the smooth test in the selected dimension. Extensive simulations presented in Ledwina (1994), Kallenberg and Ledwina (1995a, 1995b), Bogdan (1995) and Bogdan and Ledwina (1996) show that the data-driven smooth test proposed by Ledwina (1994) and extended by Kallenberg and Ledwina (1995a) compares very well to classical tests and other competitors.

We are also interested in the composite hypothesis $H_0 : f(x) \in \{f(x; \boldsymbol{\beta}), \boldsymbol{\beta} \in \mathcal{B}\}$, where $\mathcal{B} \subset \mathbb{R}^q$ and $\{f(x; \boldsymbol{\beta}), \boldsymbol{\beta} \in \mathcal{B}\}$ is a given family of densities (for instance, the family of normal or exponential densities) with unknown parameter $\beta$.

Again a lot of work has been done to investigate KS and CvM test statistics in the case of a composite null hypothesis. As is well known, when a nuisance parameter $\boldsymbol{\beta}$ is present, the situation is more complicated. The reason is that a natural counterpart of the empirical process, on which these statistics are based, is no longer distribution free or even asymptotically distribution free. Refer to Durbin (1973), Neuhaus (1979), Khmaladze (1981) and D'Agostino and Stephens (1986) for more thorough discussions of the composite null case.

Two general solutions have been proposed to deal with the nuisance parameter $\boldsymbol{\beta}$. One is proposed by Khmaladze (1981), depending on modifying the natural empirical process with estimated parameters to get a martingale converging weakly to a Wiener process under the null hypothesis. This method makes it possible to construct some counterparts of the classical KS and CvM test statistics based on the new process. The other is given by Burke and Gombay (1988), consisting in taking a single bootstrap sample to estimate the nuisance parameter $\boldsymbol{\beta}$, which makes the KS and CvM statistics asymptotically distribution free, based on the related empirical process.

These two solutions, elegant mathematically, were proposed to enable the use of classical solutions in a more complicated situation when the nuisance parameter is present. However, one anticipates that these tests will have the same deficiency in the composite null case as in the simple null case. In fact, simulation studies by Angus (1982), Ascher (1990) and Gan and Koehler (1990) show that more specialized tests, such as Gini's test for exponentiality and Shapiro-Wilk's test for normality, dominate the composite null versions of the KS and CvM tests in most situations. Thus, as in the case of testing the simple hypothesis, it seems to be promising to consider smooth tests.

To be more specific, let $F(x;\boldsymbol{\beta})$ be the distribution function of $X_i$ when $\boldsymbol{\beta}$ is the true parameter value, and define exponential families (with respect to $\boldsymbol{\theta}$) by their densities

$$g_k(x;\boldsymbol{\beta},\boldsymbol{\theta}) = C(\boldsymbol{\beta},\boldsymbol{\theta})\exp\{\sum_{i=1}^{k}\theta_i u_i[F(x;\boldsymbol{\beta})]\}f(x;\boldsymbol{\beta}), \quad k = 1,2,..., \tag{1.3}$$

where $u_1, u_2, ...$ are a bounded orthonormal system in $L_2([0,1])$, $\boldsymbol{\theta} = (\theta_1, ..., \theta_k) \in \mathbb{R}^k$ and $C(\boldsymbol{\beta},\boldsymbol{\theta})$ is a constant depending on $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. The latter constant ensures that the probability density function integrates to one. Testing $H_0$ within the exponential family (1.3) means testing $H_0 : \boldsymbol{\theta} = 0$ against $\boldsymbol{\theta} \neq 0$. An obvious test statistic for this testing problem is the score statistic (see details in Javitz (1975), Kopecky and Pierce (1979), Thomas and Pierce (1979), Neyman (1980) and Rayner and Best (1989, 2009)). Denoting by $\boldsymbol{I}_k$ the $k \times k$ identity matrix, the score statistic is given by

$$W_k = nY_n^{\mathrm{T}}(\hat{\boldsymbol{\beta}})\{\boldsymbol{I}_k + \boldsymbol{R}(\hat{\boldsymbol{\beta}})\}Y_n(\hat{\boldsymbol{\beta}}), \tag{1.4}$$

where, writing $\mathrm{E}_{\boldsymbol{\beta}}$ for the expected value under $f(x;\boldsymbol{\beta})$,

$$
\begin{aligned}
Y_n(\boldsymbol{\beta}) &= (\bar{u}_1(\boldsymbol{\beta}), ..., \bar{u}_k(\boldsymbol{\beta}))^{\mathrm{T}} \\
&= n^{-1}\sum_{i=1}^{n}(u_1[F(X_i;\boldsymbol{\beta})], ..., u_k[F(X_i;\boldsymbol{\beta})])^{\mathrm{T}},
\end{aligned}
$$

$$
\boldsymbol{I}_{\boldsymbol{\beta}} = \left\{ -\mathrm{E}_{\boldsymbol{\beta}}\frac{\partial}{\partial\beta_t}u_j[F(X;\boldsymbol{\beta})]\right\}_{t=1,...,q;\ j=1,...,k}, \tag{1.5}
$$

$$
\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \left\{ -\mathrm{E}_{\boldsymbol{\beta}}\frac{\partial^2}{\partial\beta_t\partial\beta_u}\log f(X;\boldsymbol{\beta})\right\}_{t=1,...,q;\ u=1,...,q}, \tag{1.6}
$$

$$
\boldsymbol{R}(\boldsymbol{\beta}) = \boldsymbol{I}_{\boldsymbol{\beta}}^{\mathrm{T}}(\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} - \boldsymbol{I}_{\boldsymbol{\beta}}\boldsymbol{I}_{\boldsymbol{\beta}}^{\mathrm{T}})^{-1}\boldsymbol{I}_{\boldsymbol{\beta}}, \tag{1.7}
$$

and $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$ under $H_0$. However, $\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ often cannot be computed, in which case one could use the observed information matrices $\boldsymbol{J}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ and $\boldsymbol{J}_{\boldsymbol{\beta}}$:

$$
\boldsymbol{J}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \left\{ -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\beta_t\partial\beta_u}\log f(X_i;\boldsymbol{\beta})\right\}_{t=1,...,q;\ u=1,...,q},
$$

$$
\boldsymbol{J}_{\boldsymbol{\beta}} = \left\{ -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\beta_t}u_j[F(X_i;\boldsymbol{\beta})]\right\}_{t=1,...,q;\ j=1,...,k}.
$$

There are many smooth tests derived from this score statistic that have been recommended for testing goodness of fit (see,.e.g.,Milbrodt and Strasser (1990); Rayner and Best (1990)), but a poor choice for the number $k$ of components in the test statistic can result in a considerable loss of power. Therefore, a good procedure is needed for choosing a value for $k$ that can be used in practice. Research in this area shows that a deterministic procedure gives no simple answer (see Inglot, Kallenberg,and Ledwina (1994)). As we have mentioned before, Ledwina (1994) introduced a data-driven version of Neyman's test for testing uniformity. To apply this procedure for testing composite hypotheses (e.g., for testing normality or exponentiality), Kallenberg and Ledwina (1997a, b) extend Schwarz's selection rule by inserting an estimator of the parameters involved in the composite null hypothesis. They also show that the extended data-driven smooth test is consistent against essentially any alternative and competitive with well-known "special" tests, such as the Shapiro-Wilk test for normality and Gini's test for exponentiality. Moreover, Inglot and Ledwina (2006) proposed a method of extending the sensitivity of data driven smooth tests defined using a (simplified) Schwarz selection rule to determine the number of components, in which the type of penalty (AIC or BIC) is chosen on the basis of the data. They claim that the test is powerful in detecting both lowly and highly oscillating alternatives.

Since we will compare the performance of our proposed tests with Ledwina's data driven Neyman smooth tests in the simulations of Chapter IV, we would now like to introduce more details about their selection rules.

When testing a simple null hypothesis, let

$$L_s(\theta) = \log \prod_{i=1}^{n} g_s(X_i; \boldsymbol{\theta}), \tag{1.8}$$

where the model $g_s$ is defined by (1.1), and

$$L_s = \sup_{\theta \in \Omega_s} L_s(\boldsymbol{\theta}), \qquad \mathcal{L}_s = L_s - \frac{1}{2} s \log n. \tag{1.9}$$

Schwarz's technique selects the model with index $S$ defined by

$$S = \min\{j,\ 1 \leq j \leq K : \mathcal{L}_j = \max_{1 \leq s \leq K} \mathcal{L}_s\}. \tag{1.10}$$

So the family $g_k(x; \boldsymbol{\theta})$ is selected to choose a relatively (with respect to the sample size $n$) simple density that has high likelihood. Criterion (1.10) is an approximation of the

Bayes procedure for model choice under a special class of priors (Haughton (1988), Schwarz (1978)). On the other hand, (1.10) can be interpreted as an approximation of the selection rule based on a minimum description-length criterion (Barron and Cover 1991; Rissanen 1983) or as an approximation for the stochastic complexity (Rissanen 1987). Having chosen the model of dimension $S$, Ledwina (1994) proposed to use $N_S$, with $S$ in place of $k$ in (1.2) as a new version of Neyman's test.

Kallenberg and Ledwina (1997a, b) made some modifications to selection rules for testing composite hypotheses. As nuisance parameters $\boldsymbol{\beta}$ need to be estimated in composite cases, their selection rule is defined in terms of $W_k$, given by (1.4). The modified criterion of Kallenberg and Ledwina (1997a, b) is

$$S1 = S1(\hat{\boldsymbol{\beta}}) = \min\{k : 1 \leq k \leq K,\ W_k - k\log n \geqslant W_j - j\log n, j = 1, ..., K\}, \quad (1.11)$$

and the corresponding test statistic is

$$W_{S1} = W_{S1(\hat{\boldsymbol{\beta}})}. \quad (1.12)$$

A more simple modification, which is easier to calculate, is

$$\begin{aligned} S2 = S2(\hat{\boldsymbol{\beta}}) = \min\{k : 1 \leq k \leq K, n\|Y_n(\hat{\boldsymbol{\beta}})\|^2_{(k)} \quad (1.13) \\ -k\log n \geqslant n\|Y_n(\hat{\boldsymbol{\beta}})\|^2_{(j)} - j\log n, \quad j = 1, ..., K\}, \end{aligned}$$

where the index of the norm denotes the dimension. The corresponding test statistic is

$$W_{S2} = W_{S2(\hat{\boldsymbol{\beta}})}. \quad (1.14)$$

For testing simple null hypotheses, Inglot and Ledwina (2006) proposed a new selection criterion designed to work better than (1.10) for high frequency alternatives.

The penalty of this criterion is defined by

$$\pi(j,n) = \{j\text{log}n\}\{I_n(c_0)\} + \{2j\}\{1 - I_n(c_0)\}, \tag{1.15}$$

and the new selection rule is

$$T = \min\{k : 1 \leq k \leq K, N_k - \pi(k,n) \geqslant N_j - \pi(j,n), \quad j = 1, ..., K\}, \tag{1.16}$$

where

$$I_n(c) = \mathbf{1}\left( \max_{1 \leq j \leq K} |\sqrt{n}\hat{b}_j| \leq \sqrt{c\text{log}n} \right),$$

$\hat{b}_j = \frac{1}{n}\sum_{i=1}^{n} u_j(X_i)$ and $c_0$=2.4. The new data driven statistic is $N_T$.

Until now, all of the approaches mentioned are frequentist in nature. Verdinelli and Wasserman (1998) proposed a purely Bayesian nonparametric goodness-of-fit test. However, we would like to focus interest on what some have termed "hybrid Bayes-frequentist" methods, i.e., methods that combine Bayesian and frequentist thinking; for details see Bayarri and Berger (2004), Conrad, Botner, Hallgren and Perez de los Heros (2003), Aerts, Claeskens and Hart (2004) and Chang and Chow (2005). Our proposed tests are examples of such hybrids, as they are derived from Bayesian principles but used in frequentist fashion. We shall refer to such tests as *frequentist-Bayes*. Good (1957) proposed a frequentist-Bayes test based on a Bayes factor. Aerts, Claeskens and Hart (2004) appear to be the first to propose frequentist-Bayes lack-of-fit tests based on posterior probabilities. Hart (2009) proposed another frequentist-Bayes motivated test. He used the method of Laplace to approximate posterior probabilities, which is precisely the subject of the current dissertation. But we apply this method to test for probability density functions, whereas he tested for regression functions.

This dissertation proposes a frequentist-Bayes omnibus test that has good power at high frequencies and also performs comparably to some popular omnibus tests at low frequencies. The next chapter describes methodology and its motivation. Theoretical properties of the new tests are presented in Chapter III, and Chapter IV describes the performance of tests, including simulations and a real data example. The dissertation ends with some concluding remarks in Chapter V.

# CHAPTER II

# METHODOLOGY AND ITS MOTIVATION

It is assumed that $X_1, ..., X_n$ are a random sample from an unknown density $f$, and we wish to test the following null hypothesis:

$$H_0 : f \in \{f(\cdot;\boldsymbol{\beta}) : \boldsymbol{\beta} \in \mathcal{B}\} = \mathcal{F}_0,$$

where $f(\cdot;\boldsymbol{\beta})$ is a density for each $\boldsymbol{\beta} \in \mathcal{B}$ and $\mathcal{B} \subset \mathbb{R}^q$. The proposed tests of $H_0$ are motivated by a combination of Bayesian and score test ideas. We will derive the statistics and in the process provide motivation for them.

## 2.1. Derivation of Test Statistics

Let $u_1, u_2, ...$ be basis functions that are orthonormal on the interval [0,1] in the sense that

$$\int_0^1 u_j(x)u_k(x)dx = \delta_{jk} \tag{2.1}$$

and

$$\int_0^1 u_j(x)dx = 0, \tag{2.2}$$

where $\delta_{jk}$ is the Kronecker delta. Now define, for each $x$,

$$\phi_j(x;\boldsymbol{\beta}) = u_j(F(x;\boldsymbol{\beta})), \quad \boldsymbol{\beta} \in \mathcal{B}, \quad j = 1, 2, ...,$$

where $F(\cdot;\boldsymbol{\beta})$ is the cumulative distribution function corresponding to $f(\cdot;\boldsymbol{\beta})$.

Define, for $j = 1, 2, ...$, the class of densities $\mathcal{F}_j$ by

$$\mathcal{F}_j = \{f_j(\cdot;\boldsymbol{\beta}, \theta_j) : \boldsymbol{\beta} \in \mathcal{B}, -\infty < \theta_j < \infty\},$$

where

$$f_j(x;\boldsymbol{\beta},\theta_j) = C(\boldsymbol{\beta},\theta_j)\exp(\theta_j\phi_j(x;\boldsymbol{\beta}))f(x;\boldsymbol{\beta}), \tag{2.3}$$

and $C(\boldsymbol{\beta},\theta_j)$ is a positive constant ensuring that $f_j$ integrates to 1. Our test statistics are approximations to the posterior probability of $H_0$ assuming that the true density is in one of the classes $\mathcal{F}_0, \mathcal{F}_1, \ldots$. Using Bayes' theorem:

$$\begin{aligned} P(H_0|\boldsymbol{x}) &= \frac{P(\boldsymbol{x}|H_0)P(H_0)}{P(\boldsymbol{x})} \\ &= \frac{P(\boldsymbol{x}|H_0)P(H_0)}{\sum_{j=0}^{\infty} P(\boldsymbol{x}|H_j)P(H_j)} \\ &= \left(1 + \sum_{j=1}^{\infty} \frac{P(\boldsymbol{x}|H_j)P(H_j)}{P(\boldsymbol{x}|H_0)P(H_0)}\right)^{-1}, \end{aligned}$$

where $P(\boldsymbol{x}|H_j)$ is the marginal likelihood for model $\mathcal{F}_j$, $\;j = 1, 2, \ldots$ .

Let $\pi_j$ denote the prior probability that $f$ is in $\mathcal{F}_j$, $j = 0, 1, \ldots$. The prior distribution for $\boldsymbol{\beta}$ given that $f \in \mathcal{F}_0$ is denoted $\pi^0$. For any $j = 1, 2, \ldots$, given that $f \in \mathcal{F}_j$ it is assumed that $\theta_j$ and $\boldsymbol{\beta}$ have joint prior $\pi(\theta_j)\pi^0(\boldsymbol{\beta})$. Given observations $\boldsymbol{x} = (x_1, \ldots, x_n)$, define

$$\begin{aligned} m_0(\boldsymbol{x}) &= P(\boldsymbol{x}|H_0) \\ &= \int P(\boldsymbol{x}|\boldsymbol{\beta})P(\boldsymbol{\beta}|H_0)d\boldsymbol{\beta} \\ &= \int_{\mathcal{B}} \pi^0(\boldsymbol{\beta}) \prod_{i=1}^{n} f(x_i;\boldsymbol{\beta})d\boldsymbol{\beta}, \end{aligned}$$

and

$$\begin{aligned} m_j(\boldsymbol{x}) &= P(\boldsymbol{x}|H_j) \\ &= \int P(\boldsymbol{x}|\boldsymbol{\beta},\theta_j,H_j)P(\boldsymbol{\beta},\theta_j|H_j)d(\boldsymbol{\beta},\theta_j) \\ &= \int_{\mathcal{B}}\int_{-\infty}^{\infty} \pi(\theta_j)\pi^0(\boldsymbol{\beta}) \prod_{i=1}^{n} f_j(x_i;\boldsymbol{\beta},\theta_j)d\theta_j d\boldsymbol{\beta}, \quad j = 1, 2, \ldots. \end{aligned}$$

The posterior probability of $H_0$ is

$$P(H_0|\boldsymbol{x}) = \left(1 + \sum_{j=1}^{\infty} \frac{\pi_j}{\pi_0} \cdot \frac{m_j(\boldsymbol{x})}{m_0(\boldsymbol{x})}\right)^{-1}. \tag{2.4}$$

The null hypothesis is rejected if $P(H_0|\boldsymbol{x})$ is sufficiently small. The cutoff point for rejection of $P(H_0)$ is defined in the usual frequentist way, i.e., it is the $\alpha 100$th percentile of the distribution of $P(H_0|\boldsymbol{x})$ assuming $H_0$ to be true.

The last expression sheds light on the difference between the way Bayesians and frequentists would assess the evidence against $H_0$ given a value of $P(H_0|\boldsymbol{x})$. For a Bayesian, the prior probability of $H_0$ is crucial since $P(H_0|\boldsymbol{x})$ varies between 0 and 1 as $\pi_0$ varies in the same way. On the other hand, a frequentist would reject $H_0$ if and only if $P(H_0|\boldsymbol{x})$ is less than its $\alpha$ quantile under $H_0$, in which case the test is independent of the value of $\pi_0$. This can be seen by noting that a frequentist test based on $P(H_0|\boldsymbol{x})$ is equivalent to one based on $\sum_{j=1}^{\infty} \pi_j m_j(\boldsymbol{x})/m_0(\boldsymbol{x})$. So, to a frequentist, as long as $0 < \pi_0 < 1$, the choice of $\pi_0$ in $P(H_0|\boldsymbol{x})$ is arbitrary. A frequentist test based on $P(H_0|\boldsymbol{x})$ depends on $\pi_0, \pi_1, ...$, only through the relative sizes of $\pi_1, \pi_2, ....$

Only in very special circumstances can $m_0(\boldsymbol{x}), m_1(\boldsymbol{x}), ...$ be determined exactly. In the Bayesian world, the currently most popular means of approximating such quantities is to use MCMC. For a frequentist, computing $P(H_0|\boldsymbol{x})$ only solves a small part of the problem since the null sampling distribution of $P(H_0|\boldsymbol{x})$ is unknown. If the bootstrap were used to approximate the distribution of $P(H_0|\boldsymbol{x})$, then MCMC would have to be used on every bootstrap sample to approximate the test statistic. For this reason, we will propose various means of approximating $P(H_0|\boldsymbol{x})$ that can either be computed exactly or approximated quickly.

## 2.2. Approximations

As we discussed above, marginal likelihoods are generally difficult to compute. Exact solutions are known for a small class of distributions. In general, some kind of numerical integration method is needed, either a general method such as Gaussian integration or a Monte Carlo method, or a method specialized to statistical problems, such as the Laplace approximation, Gibbs sampling or the EM algorithm.

Our basic approximation of $P(H_0|\boldsymbol{x})$ is based on approximating each of the integrals $m_j(\boldsymbol{x})$ by the method of Laplace. Laplace approximation provides a general way to approach marginalization problems. The basic setting is to approximate an integral of the form:

$$I_n = \int b(x)e^{h_n(x)}dx,$$

where $n$ is typically the number of data points. Let $x$ denote a $d$-dimensional vector, $b(x)$ a function of $x$ alone, and $h_n(x)$ is a function of both $x$ and $n$. After performing a Taylor series expansion of both $h_n(x)$ and the exponential function and evaluating some elementary integrals, we obtain:

$$I_n \approx (2\pi)^{d/2} \det(H)^{-1/2} b(\hat{x})e^{h_n(\hat{x})}, \tag{2.5}$$

where $H = -D^2h(\hat{x})$ is the Hessian matrix of $h$ evaluated at $\hat{x}$ and $\hat{x} = \text{argmax}_x h(x)$.

Let $\hat{\boldsymbol{\beta}}_j$ and $\hat{\theta}_j$ be the maximum likelihood estimates of $\boldsymbol{\beta}$ and $\theta_j$, respectively, when it is assumed that $f \in \mathcal{F}_j, j = 0, 1, ....$ We may write $m_j(\boldsymbol{x})$ as

$$\int_{\mathcal{B}} \int_{-\infty}^{\infty} \pi(\theta_j)\pi^0(\boldsymbol{\beta}) \exp\left\{\log\left(\prod_{i=1}^{n} f_j(x_i; \boldsymbol{\beta}, \theta_j)\right)\right\} d\theta_j d\boldsymbol{\beta},$$

where $b(\boldsymbol{\beta}, \theta_j) = \pi(\theta_j)\pi^0(\boldsymbol{\beta})$ and $h_n(\boldsymbol{\beta}, \theta_j) = \log(\prod_{i=1}^{n} f_j(x_i; \boldsymbol{\beta}, \theta_j))$. Using the Laplace approximation up to the first order as in (2.5), we get,

$$\begin{aligned}\hat{m}_j(\boldsymbol{x}) &\approx (2\pi)^{(q+1)/2}|H_j(\hat{\boldsymbol{\beta}}_j,\hat{\theta}_j)|^{-1/2}\pi(\hat{\theta}_j)\pi^0(\hat{\boldsymbol{\beta}}_j)e^{\log(\prod_{i=1}^n f_j(x_i;\hat{\boldsymbol{\beta}}_j,\hat{\theta}_j))} \\ &\approx (2\pi)^{(q+1)/2}|H_j|^{-1/2}\pi(\hat{\theta}_j)\pi^0(\hat{\boldsymbol{\beta}}_j)\prod_{i=1}^n f_j(x_i;\hat{\boldsymbol{\beta}}_j,\hat{\theta}_j), \quad j=1,2,.... \qquad (2.6)\end{aligned}$$

A similar approximation holds for $m_0(\boldsymbol{x})$:

$$\begin{aligned}\hat{m}_0(\boldsymbol{x}) &= \int_{\mathcal{B}} \pi^0(\boldsymbol{\beta})\exp\left\{\log\prod_{i=1}^n f(x_i;\boldsymbol{\beta})\right\}d\boldsymbol{\beta} \\ &\approx (2\pi)^{q/2}|H_0(\hat{\boldsymbol{\beta}}_0)|^{-1/2}\pi^0(\hat{\boldsymbol{\beta}}_0)\prod_{i=1}^n f(x_i;\hat{\boldsymbol{\beta}}_0) \qquad (2.7)\end{aligned}$$

Substitution of $\hat{m}_j(\boldsymbol{x})$ for $m_j(\boldsymbol{x})$, using the fact that $P(H_0|\boldsymbol{x})$ is equivalent to the statistic $\sum_{j=1}^{\infty}\pi_j m_j(\boldsymbol{x})/m_0(\boldsymbol{x})$, as discussed at the end of subsection 2.1, and truncation of the series at, say, $k$ leads to a computationally feasible test statistic:

$$\sqrt{2\pi}\sum_{j=1}^{k}\pi_j\pi(\hat{\theta}_j)\cdot\frac{\pi^0(\hat{\boldsymbol{\beta}}_j)}{\pi^0(\hat{\boldsymbol{\beta}}_0)}\cdot\frac{|H_j(\hat{\boldsymbol{\beta}}_j,\hat{\theta}_j))|^{-\frac{1}{2}}}{|H_0(\hat{\boldsymbol{\beta}}_0))|^{-\frac{1}{2}}}\cdot\frac{\prod_{i=1}^n f_j(x_i;\hat{\boldsymbol{\beta}}_j,\hat{\theta}_j)}{\prod_{i=1}^n f(x_i;\hat{\boldsymbol{\beta}}_0)}. \qquad (2.8)$$

Rejecting $H_0$ for small values of $P(H_0|\boldsymbol{x})$ is equivalent to rejecting $H_0$ for large values of (2.8).

To compute (2.8), we need to find the maximum likelihood estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\theta}_j$, which can be done using a standard method such as gradient search. It also requires computing the second derivative matrix to obtain $H$. This is usually the harder quantity to calculate. Therefore, further simplifications are desirable from both computational and motivational standpoints.

## 2.3. Utilizing Score Statistics

Score tests (see, e.g., Rayner and Best 1989, pp. 77-81) achieve computational simplicity relative to likelihood ratio tests by

(i) computing the information matrix on the assumption that $H_0$ is true,

(ii) evaluating the information matrix and log-likelihood derivatives at *null* maximum likelihood estimates.

We will apply a similar approach to (2.8) and thereby obtain a simplified statistic that has motivational appeal.

Before further simplifications, we first make some clarifications. Define

$$\begin{aligned} l_j &= \log\left(\prod_{i=1}^{n} f_j(x_i;\boldsymbol{\beta},\theta_j)\right) \\ &= n\log(C(\boldsymbol{\beta},\theta_j)) + \theta_j \sum_{i=1}^{n} \phi_j(x_i;\boldsymbol{\beta}) + \sum_{i=1}^{n} \log f(x_i;\boldsymbol{\beta}). \end{aligned}$$

Then,

$$l_j{}' = \frac{\partial l_j}{\partial \theta_j} = n\frac{\partial \log(C(\boldsymbol{\beta},\theta_j))}{\partial \theta_j} + \sum_{i=1}^{n} \phi_j(x_i;\boldsymbol{\beta}),$$

and

$$l_j{}'' = \frac{\partial^2 l_j}{\partial {\theta_j}^2} = n\frac{\partial^2 \log(C(\boldsymbol{\beta},\theta_j))}{\partial {\theta_j}^2}.$$

We make note of some good properties of $C(\boldsymbol{\beta};\boldsymbol{\theta})$ that will be used later. Since

$$\int f_j(x;\boldsymbol{\beta},\theta_j)dx = 1,$$

(2.3) implies

$$C(\boldsymbol{\beta},\theta_j)\int \exp(\theta_j\phi_j) f(x;\boldsymbol{\beta})dx = 1. \tag{2.9}$$

Plugging in $\theta_j = 0$, it follows that

$$C(\boldsymbol{\beta},0)\int f(x;\boldsymbol{\beta})dx = 1.$$

As $\int f(x;\boldsymbol{\beta})dx \equiv 1$, we have $C(\boldsymbol{\beta},0) \equiv 1$. From (2.9), we observe that

$$C(\boldsymbol{\beta},\theta_j) = \frac{1}{\int \exp(\theta_j\phi_j)f(x;\boldsymbol{\beta})dx},$$

and so

$$\log C(\boldsymbol{\beta},\theta_j) = -\log\biggl(\int \exp(\theta_j\phi_j)f(x;\boldsymbol{\beta})dx\biggr),$$

$$\frac{\partial \log C(\boldsymbol{\beta},\theta_j)}{\partial\theta_j} = -\frac{\int \exp(\theta_j\phi_j)\phi_j f(x;\boldsymbol{\beta})dx}{\int \exp(\theta_j\phi_j)f(x;\boldsymbol{\beta})dx},$$

and

$$\frac{\partial \log C(\boldsymbol{\beta},\theta_j)}{\partial\beta_t} = -\frac{\frac{\partial}{\partial\beta_t}\int \exp(\theta_j\phi_j)f(x;\boldsymbol{\beta})dx}{\int \exp(\theta_j\phi_j)f(x;\boldsymbol{\beta})dx}.$$

Plugging in $\theta_j = 0$, it follows that

$$\left.\frac{\partial \log C(\boldsymbol{\beta},\theta_j)}{\partial\theta_j}\right|_{\theta_j=0} = -\frac{\int \phi_j f(x;\boldsymbol{\beta})dx}{\int f(x;\boldsymbol{\beta})dx} = 0.$$

Similarly,

$$\begin{aligned}
&\left.\frac{\partial^2 \log C(\boldsymbol{\beta},\theta_j)}{\partial\theta_j^2}\right|_{\theta_j=0} \\
= & -\left.\frac{\int \exp(\theta_j\phi_j){\phi_j}^2 f(x;\boldsymbol{\beta})dx \cdot \int \exp(\theta_j\phi_j)f(x;\boldsymbol{\beta})dx}{\Bigl(\int \exp(\theta_j\phi_j)f(x;\boldsymbol{\beta})dx\Bigr)^2}\right|_{\theta_j=0} \\
& + \left.\frac{\Bigl(\int \exp(\theta_j\phi_j)\phi_j f(x;\boldsymbol{\beta})dx\Bigr)^2}{\Bigl(\int \exp(\theta_j\phi_j)f(x;\boldsymbol{\beta})dx\Bigr)^2}\right|_{\theta_j=0} \\
= & -\int \phi_j^2 f(x;\boldsymbol{\beta})dx \\
= & -1,
\end{aligned}$$

$$
\begin{aligned}
&\left.\frac{\partial^2 \log C(\boldsymbol{\beta}, \theta_j)}{\partial \theta_j \partial \beta_t}\right|_{\theta_j=0} \\
= & -\left.\frac{\left(\frac{\partial}{\partial \beta_t} \int \exp(\theta_j \phi_j)\phi_j f(x;\boldsymbol{\beta})dx\right) \cdot \int \exp(\theta_j \phi_j) f(x;\boldsymbol{\beta})dx}{\left(\int \exp(\theta_j \phi_j) f(x;\boldsymbol{\beta})dx\right)^2}\right|_{\theta_j=0} \\
& + \left.\frac{\int \exp(\theta_j \phi_j)\phi_j f(x;\boldsymbol{\beta})dx \cdot \left(\frac{\partial}{\partial \beta_t} \int \exp(\theta_j \phi_j) f(x;\boldsymbol{\beta})dx\right)}{\left(\int \exp(\theta_j \phi_j) f(x;\boldsymbol{\beta})dx\right)^2}\right|_{\theta_j=0} \\
= & \; 0,
\end{aligned}
$$

and

$$
\begin{aligned}
&\left.\frac{\partial^2 \log C(\boldsymbol{\beta}, \theta_j)}{\partial \beta_t \partial \beta_u}\right|_{\theta_j=0} \\
= & -\left.\frac{\left(\frac{\partial^2}{\partial \beta_t \partial \beta_u} \int \exp(\theta_j \phi_j) f(x;\boldsymbol{\beta})dx\right) \cdot \int \exp(\theta_j \phi_j) f(x;\boldsymbol{\beta})dx}{\left(\int \exp(\theta_j \phi_j) f(x;\boldsymbol{\beta})dx\right)^2}\right|_{\theta_j=0} \\
& + \left.\frac{\left(\frac{\partial}{\partial \beta_t} \int \exp(\theta_j \phi_j) f(x;\boldsymbol{\beta})dx\right) \cdot \left(\frac{\partial}{\partial \beta_u} \int \exp(\theta_j \phi_j) f(x;\boldsymbol{\beta})dx\right)}{\left(\int \exp(\theta_j \phi_j) f(x;\boldsymbol{\beta})dx\right)^2}\right|_{\theta_j=0} \\
= & \; 0.
\end{aligned}
$$

### 2.3.1. Basic Ideas

Now, we start the simplification steps by applying score test ideas. Firstly, using $\hat{\boldsymbol{\beta}}_0$ and 0 as initial estimates of $\boldsymbol{\beta}$ and $\theta_j$, respectively, a one-step Newton's approximation of $\hat{\theta}_j$ is

$$\begin{aligned}
\tilde{\theta}_j &= 0 - \frac{l_j{}'|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_0,\theta_j=0}}{l_j{}''|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_0,\theta_j=0}} \\
&= 0 - \frac{n\frac{\partial \log(C(\hat{\boldsymbol{\beta}}_0,\theta_j))}{\partial\theta_j}\Big|_{\theta_j=0} + \sum_{i=1}^n \phi_j(x_i;\hat{\boldsymbol{\beta}}_0)}{n\frac{\partial^2 \log(C(\hat{\boldsymbol{\beta}}_0,\theta_j))}{\partial{\theta_j}^2}\Big|_{\theta_j=0}} \\
&= \frac{1}{n}\sum_{i=1}^n \phi_j(x_i;\hat{\boldsymbol{\beta}}_0). \qquad (2.10)
\end{aligned}$$

Now consider the ratio $|H_j(\hat{\boldsymbol{\beta}}_j,\hat{\theta}_j)|^{-1/2}\Big/|H_0(\hat{\boldsymbol{\beta}}_0)|^{-1/2}$. As defined in subsection 2.2, we have

$$H_j(\boldsymbol{\beta},\theta_j) = -\left(\sum_{i=1}^n \frac{\partial^2 \log f_j(x_i;\boldsymbol{\beta},\theta_j)}{\partial(\boldsymbol{\beta},\theta_j)\partial(\boldsymbol{\beta},\theta_j)^{\mathrm{T}}}\right)_{(q+1)\times(q+1)},$$

where

$$\frac{\partial^2 \log f_j(x;\boldsymbol{\beta},\theta_j)}{\partial\beta_t\partial\beta_u} = \frac{\partial^2 \log C(\boldsymbol{\beta},\theta_j)}{\partial\beta_t\partial\beta_u} + \theta_j\frac{\partial^2\phi_j(x;\boldsymbol{\beta})}{\partial\beta_t\partial\beta_u} + \frac{\partial^2 \log f(x;\boldsymbol{\beta})}{\partial\beta_t\partial\beta_u},$$

$$\frac{\partial^2 \log f_j(x;\boldsymbol{\beta},\theta_j)}{\partial\beta_t\partial\theta_j} = \frac{\partial^2 \log C(\boldsymbol{\beta},\theta_j)}{\partial\beta_t\partial\theta_j} + \frac{\partial\phi_j(x;\boldsymbol{\beta})}{\partial\beta_t},$$

$$\frac{\partial^2 \log f_j(x;\boldsymbol{\beta},\theta_j)}{\partial\theta_j^2} = \frac{\partial^2 \log C(\boldsymbol{\beta},\theta_j)}{\partial\theta_j^2},$$

and

$$H_0(\boldsymbol{\beta}) = -\left(\sum_{i=1}^n \frac{\partial^2 \log f(x_i;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}}\right)_{q\times q}.$$

Assuming $H_0$ is true and using the results at the beginning of subsection 2.3, we obtain that

$$H_j(\boldsymbol{\beta},0) = \begin{pmatrix} H_0(\boldsymbol{\beta}) & -\sum_{i=1}^n \frac{\partial\phi_j(x_i;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \\ \left(-\sum_{i=1}^n \frac{\partial\phi_j(x_i;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right)^{\mathrm{T}} & n \end{pmatrix}.$$

A property of determinants gives

$$\begin{aligned} |H_j(\boldsymbol{\beta},0)| &= |H_0(\boldsymbol{\beta})| \times \left| n - \left[-\sum_{i=1}^n \frac{\partial\phi_j(x_i;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right]^{\mathrm{T}} H_0(\boldsymbol{\beta})^{-1} \left[-\sum_{i=1}^n \frac{\partial\phi_j(x_i;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right]\right| \\ &= n|H_0(\boldsymbol{\beta})| \left(1 - \frac{1}{n}\left[-\sum_{i=1}^n \frac{\partial\phi_j(x_i;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right]^{\mathrm{T}} H_0(\boldsymbol{\beta})^{-1} \left[-\sum_{i=1}^n \frac{\partial\phi_j(x_i;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right]\right) \\ &= n|H_0(\boldsymbol{\beta})| \left(1 - \frac{1}{n}M_j\right) \quad \text{(say)}. \end{aligned}$$

Therefore,

$$\frac{|H_j(\hat{\boldsymbol{\beta}}_0,0)|^{-1/2}}{|H_0(\hat{\boldsymbol{\beta}}_0)|^{-1/2}} = \frac{1}{\sqrt{n}}\left(1 - \frac{\hat{M}_j}{n}\right)^{-1/2},$$

where $\hat{M}_j$ is $M_j$ evaluated at $\hat{\boldsymbol{\beta}}_0$. Under both null and alternative hypotheses, $\hat{M}_j/n$ converges in probability to a constant as $n \to \infty$. Hence, we may as well absorb $(1 - \hat{M}_j/n)^{-1/2}$ into the term $\pi_j$ to simplify matters.

Substitution of $\tilde{\theta}_j$ for $\hat{\theta}_j$ and $n^{-\frac{1}{2}}$ for $|H_j(\hat{\boldsymbol{\beta}}_j,\hat{\theta}_j)|^{-1/2} \big/ |H_0(\hat{\boldsymbol{\beta}}_0)|^{-1/2}$ in (2.8) leads to the following statistic that is computationally straightforward,

$$\sqrt{\frac{2\pi}{n}} \sum_{j=1}^k \pi_j \pi(\hat{\theta}_j) \exp\left[\tilde{\theta}_j \sum_{i=1}^n \phi_j(x_i;\hat{\boldsymbol{\beta}}_0) + n\log C(\hat{\boldsymbol{\beta}}_0,\tilde{\theta}_j)\right], \tag{2.11}$$

except perhaps for the quantities $C(\hat{\boldsymbol{\beta}}_0,\tilde{\theta}_j)$. Concerning these, the following remarks are relevant.

R1. By (2.2) and Jensen's inequality, it follows that $C(\boldsymbol{\beta},\theta) \leq 1$ for all $\boldsymbol{\beta}$ and $\theta$.

R2. Since

$$C(\boldsymbol{\beta},0)\equiv 1,\quad \left.\frac{\partial \log C(\boldsymbol{\beta},\theta)}{\partial\theta}\right|_{\theta=0}\equiv 0,\quad \text{and}\quad \left.\frac{\partial^2 \log C(\boldsymbol{\beta},\theta)}{\partial\theta^2}\right|_{\theta=0}\equiv -1,$$

it follows that under $H_0$, $\log C(\hat{\boldsymbol{\beta}}_0,\tilde{\theta}_j)=-\tilde{\theta}_j^2/2+o_p(n^{-1})$.

Remark R2 implies that $n\tilde{\theta}_j^2+n\log C(\hat{\boldsymbol{\beta}}_0,\tilde{\theta}_j)$ has the same asymptotic null distribution as $n\tilde{\theta}_j^2/2$. Furthermore, remark R1 implies that using $n\tilde{\theta}_j^2/2$ instead of $n\tilde{\theta}_j^2+n\log C(\hat{\boldsymbol{\beta}}_0,\tilde{\theta}_j)$ is not necessarily a power liability, and could even be beneficial in terms of power. We thus propose the following statistic:

$$S_k=\sum_{j=1}^{k}\pi_j\pi(\tilde{\theta}_j)\exp\left(\frac{n\tilde{\theta}_j^2}{2}\right), \tag{2.12}$$

where $\tilde{\theta}_j$ is defined by (2.10), and $H_0$ is rejected for large values of $S_k$. However, for the sake of normalization, $\tilde{\theta}_j$ still needs further investigation.

### 2.3.2. Further Discussion About $\tilde{\theta}_j$

In simple null hypotheses cases, the parameter $\boldsymbol{\beta}$ is known, and a one-step Newton's approximation leads to a score statistic. Note that $n\tilde{\theta}_j^2$ is just a component of the score statistic $N_k$ given by (1.2), and then $\sqrt{n}\tilde{\theta}_j \xrightarrow{\mathcal{D}} N(0,1)$ by the central limit theorem. But, in the composite case, plugging the MLE $\hat{\boldsymbol{\beta}}_0$ into $\phi_j(x_i;\boldsymbol{\beta})$ means that $\sqrt{n}\tilde{\theta}_j$ is no longer asymptotically distributed as standard normal. As a result, the limiting distribution will not be free of unknowns. In order to avoid this problem, we will add the proper normalizing factor to $\tilde{\theta}_j$. Simulations have shown that scaling $\tilde{\theta}_j$ so that it is asymptotically distribution-free can also yield a more powerful test.

Since $W_k$, the score statistic given by (1.4) for a composite null, has asymptotically a chi-square distribution with degrees of freedom $k$, we will take advantage of this property and use the statistic

$$\frac{1}{n}\sum_{i=1}^{n}\phi_j(x_i;\hat{\boldsymbol{\beta}}_0)\{1+R_j(\hat{\boldsymbol{\beta}}_0)\}^{\frac{1}{2}}, \quad j=1,2,\dots ,$$

where, writing $\mathrm{E}_{\boldsymbol{\beta}}$ for the expected value under the null hypothesis,

$$\boldsymbol{I}_{\boldsymbol{\beta}j} = \left\{ -\mathrm{E}_{\boldsymbol{\beta}}\frac{\partial}{\partial\beta_t}u_j[F(X;\boldsymbol{\beta})]\right\}_{t=1,\dots,q}, \tag{2.13}$$

$$\boldsymbol{I}_{\boldsymbol{\beta\beta}} = \left\{ -\mathrm{E}_{\boldsymbol{\beta}}\frac{\partial^2}{\partial\beta_t\partial\beta_u}\log f(X;\boldsymbol{\beta})\right\}_{t=1,\dots,q;\ u=1,\dots,q}, \tag{2.14}$$

$$R_j(\boldsymbol{\beta}) = \boldsymbol{I}_{\boldsymbol{\beta}j}^{\mathrm{T}}(\boldsymbol{I}_{\boldsymbol{\beta\beta}} - \boldsymbol{I}_{\boldsymbol{\beta}j}\boldsymbol{I}_{\boldsymbol{\beta}j}^{\mathrm{T}})^{-1}\boldsymbol{I}_{\boldsymbol{\beta}j}, \tag{2.15}$$

and $\hat{\boldsymbol{\beta}}_0$ is the maximum likelihood estimate of $\boldsymbol{\beta}$ assuming that $H_0$ is true.

In the case of a location-scale family, $R_j(\boldsymbol{\beta})$ defined above does not depend on $\boldsymbol{\beta}$. To simplify the presentation some additional notation is now introduced. Since

$$f(x;\boldsymbol{\beta}) = \frac{1}{\beta_2}f_0\left(\frac{x-\beta_1}{\beta_2}\right) \quad \text{and} \quad F(x;\boldsymbol{\beta}) = F_0\left(\frac{x-\beta_1}{\beta_2}\right)$$

with known $f_0$ and $F_0$, $R_j(\boldsymbol{\beta})$ depends on $X_1,...,X_n$ only through

$$\frac{X_i-\hat{\beta}_1}{\hat{\beta}_2}, \quad i=1,...,n,$$

where $(\hat{\beta}_1,\hat{\beta}_2)=\hat{\boldsymbol{\beta}}_0$. Because $(\hat{\beta}_1,\hat{\beta}_2)$ is location-scale equivariant, the distribution of

$$\left(\frac{X_1-\hat{\beta}_1}{\hat{\beta}_2},...,\frac{X_n-\hat{\beta}_1}{\hat{\beta}_2}\right)$$

does not depend on the location-scale parameter if $X_i$ comes from a location-scale family. The same remark applies to location families and to scale families. Statistic

(2.13) can be written in the form

$$\frac{1}{n}\sum_{i=1}^{n} u_j \left[F_0\left(\frac{x_i-\hat{\beta}_1}{\hat{\beta}_2}\right)\right]\{1+R_{0j}\}^{\frac{1}{2}}, \quad j=1,2,\dots \quad , \tag{2.16}$$

where $R_{0j}$ is free of unknowns. The proof that $R_j(\boldsymbol{\beta})$ does not depend on the location-scale parameter is presented in Appendix A.

In preparation for the later simulations, we summarize the forms of $\tilde{\theta}_j$ for the simple and composite null cases.

A1. $\tilde{\theta}_{j,\text{simple}} = \frac{1}{n}\sum_{i=1}^{n}\phi_j(x_i) = \frac{1}{n}\sum_{i=1}^{n}u_j[F_0(x_i)], \quad j=1,2,...,$

A2. $\tilde{\theta}_{j,\text{composite}} = \frac{1}{n}\sum_{i=1}^{n}\phi_j(x_i;\hat{\boldsymbol{\beta}}_0)\{1+R_{0j}\}^{\frac{1}{2}} = \frac{1}{n}\sum_{i=1}^{n}u_j\left[F_0\left(\frac{x_i-\hat{\beta}_1}{\hat{\beta}_2}\right)\right]\{1+R_{0j}\}^{\frac{1}{2}},$
$j=1,2,\dots$ ,

where in the simple case $F_0$ denotes the hypothesized distribution function. More details about the asymptotic properties of $\tilde{\theta}_j$ will be presented in Chapter III.

## 2.4. Choice of Priors

In a Bayesian analysis, the prior probabilities $\pi_j,\ j=0,1,...,k$ and the prior distribution $\pi(\theta_j),\ j=0,1,...,k$, are chosen to represent the investigator's degree of belief in the various alternatives and the parameters therein. A Bayesian who wishes to do an analysis independent of his own prior beliefs may wish to use noninformative priors. In our setting, very little is known about the underlying density. In such a case it would make sense to use vague prior probabilities over various densities and also noninformative priors for the parameters in these models.

Two possibilities for $\pi(\theta)$ are:

(1) the constant improper prior,

(2) the proper prior: $\pi(\theta_j) = C\exp(-\frac{1}{2}{\theta_j}^2)$.

The second prior may be regarded as a reference prior with information equivalent to that in a single observation. The difference between using (1) and (2) is negligible for all but very small sample sizes. There have been many arguments about what is the most appropriate noninformative prior in a given situation, and about whether or not any prior can truly express ignorance about the underlying parameters. Kass and Wasserman (1996) give a review of the problem and many relevant references.

We now turn to the problem of assigning vague prior probabilities to the density models. One possibility is to simply give each model the same probability of $1/(k+1)$. The problem with this choice is that it fails to reflect our knowledge that relatively few of a function's Fourier coefficients will substantially different from 0.

The sequence $\boldsymbol{\pi} = (\pi_1, \pi_2, ...)$ can be chosen to represent the experimenter's degree of belief about the relative sizes of $\mathrm{E}(\tilde{\theta}_j^2)$, $j = 1, 2, ...$ . Clearly, a test based on $S_k$ will benefit in terms of power if the largest probabilities are placed on those components $\tilde{\theta}_j$ with the largest values of $\mathrm{E}(\tilde{\theta}_j^2)$. The noninformative choice of $\boldsymbol{\pi}$ should at least reflect the facts that, in general, $\mathrm{E}(\tilde{\theta}_j)$ will tend to 0 as $j \to \infty$ and that "smooth" densities occur more frequently in practice than do wiggly ones. To this end, it seems reasonable to arrange basis functions $\phi_1, \phi_2, ...$ in order from lowest to highest frequency, and to choose $\boldsymbol{\pi}$ so that $\pi_j$ decreases monotonically to 0.

Taking $\pi_j = 1/j^c$ for any $c > 1$ satisfies the above criteria, and letting $c$ be fairly close to 1 will ensure vagueness of the prior probabilities. A choice for $\boldsymbol{\pi}$ that has proven useful in a regression context is such that $\pi_j \propto j^{-2}$ (Hart 2009). We will present more details about optimal $\pi_j$ in Chapter IV.

# CHAPTER III

# ASYMPTOTIC DISTRIBUTION THEORY

In this chapter, we consider the limiting distribution of $S_k$ under both the null hypothesis and local alternatives that converge to the null at rate $1/\sqrt{n}$. The local alternatives are obtained by putting $\delta_j = \theta_j\sqrt{n}$, which gives

$$f_l(x) = C(\boldsymbol{\beta}, \boldsymbol{\delta}) \exp\left\{ \sum_{r=1}^{\infty} \frac{\delta_r}{\sqrt{n}} \phi_r(x; \boldsymbol{\beta}) \right\} f(x; \boldsymbol{\beta}), \tag{3.1}$$

where $C(\boldsymbol{\beta}, \boldsymbol{\delta})$ is the appropriate normalizing constant, $\boldsymbol{\beta} \in \mathcal{B}$ and $\boldsymbol{\delta} = (\delta_1, \delta_2, ...)$.

## 3.1. Limiting Distribution for Simple Null Hypotheses

**<u>Theorem</u> 1.** *Let $u_1, u_2, ...$ be orthonormal basis functions defined as in subsection 2.1 and assume that $\sum_{r=1}^{\infty} \delta_r u_r(x)$ is uniformly bounded, i.e.,*

$$\sup_{x\in[0,1]} \left| \sum_{r=1}^{\infty} \delta_r u_r(x) \right| < \infty.$$

*Let $Z_1, Z_2, ...$ be i.i.d. standard normal random variables. Then under the local alternative $f_l$ defined by (3.1), the statistic $S_k = \sum_{j=1}^{k} \pi_j \exp(n\tilde{\theta}_j^2/2)$ converges in distribution to*

$$\mathcal{S}_1 = \sum_{j=1}^{k} \pi_j \exp[(Z_j + \delta_j)^2/2],$$

*where $k$ is an arbitrarily large but fixed number.*

*Proof.* For a simple null hypothesis, the parameter $\boldsymbol{\beta}$ is completely specified, and we deal with the limiting distribution of $\left\{\tilde{\theta}_j\right\}_{j=1,...,k}$ in the form of $\left\{\tilde{\theta}_{j,\text{simple}}\right\}_{j=1,...,k}$, defined in subsection 2.3.2. By definition we write

$$\begin{aligned}
\mathrm{E}(\tilde{\theta}_j) &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} u_j(F_0(X_i))\right] \\
&= \int_{\mathbb{R}} u_j(F_0(x)) f_l(x) dx \\
&= C(\boldsymbol{\beta}, \boldsymbol{\delta}) \int_{\mathbb{R}} u_j(F_0(x)) \exp\left\{\sum_{r=1}^{\infty} \frac{\delta_r}{\sqrt{n}} u_r(F_0(x))\right\} f_0(x) dx.
\end{aligned}$$

As in a simple null, there are no nuisance parameters. Making the change of variable $y = F_0(x)$, we get

$$\mathrm{E}(\tilde{\theta}_j) = C(\boldsymbol{\beta}, \boldsymbol{\delta}) \int_0^1 u_j(y) \exp\left\{\sum_{r=1}^{\infty} \frac{\delta_r}{\sqrt{n}} u_r(y)\right\} dy.$$

By a Taylor expansion we obtain

$$\begin{aligned}
\mathrm{E}(\tilde{\theta}_j) &= C(\boldsymbol{\beta}, \boldsymbol{\delta}) \int_0^1 u_j(y) \left(1 + \frac{1}{\sqrt{n}} \sum_{r=1}^{\infty} \delta_r u_r(y) + \frac{1}{2n}\left[\sum_{r=1}^{\infty} \delta_r u_r(y)\right]^2 \exp\left(\xi_n(y)\right)\right) dy \\
&= C(\boldsymbol{\beta}, \boldsymbol{\delta}) \left(\frac{\delta_j}{\sqrt{n}} + \frac{1}{2n} \int_0^1 u_j(y) \left[\sum_{r=1}^{\infty} \delta_r u_r(y)\right]^2 \exp\left(\xi_n(y)\right) dy\right),
\end{aligned}$$

where $\xi_n(y)$ is between 0 and $\sum_{r=1}^{\infty} \delta_r u_r(y)/\sqrt{n}$. So by the boundedness of $\sum_{r=1}^{\infty} \delta_r u_r(y)$, we have

$$\mathrm{E}(\tilde{\theta}_j) = \frac{\delta_j}{\sqrt{n}} + O(n^{-1}),$$

and therefore

$$\mathrm{E}(\sqrt{n}\tilde{\theta}_j) = \delta_j + O(n^{-1/2}).$$

For $j = 1, ..., k$,

$$
\begin{aligned}
\mathrm{Var}(\tilde{\theta}_j) &= \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} u_j(F_0(X_i))\right] \\
&= \frac{1}{n}\mathrm{Var}\left[u_j(F_0(X_1))\right] \\
&= \frac{1}{n}\left[\mathrm{E}u_j^2(F_0(X_1)) - \mathrm{E}u_j(F_0(X_1))^2\right].
\end{aligned}
$$

As in calculation of $\mathrm{E}(\tilde{\theta}_j)$,

$$
\begin{aligned}
\mathrm{E}u_j^2(F_0(X_1)) &= \int_{\mathbb{R}} u_j^2(F_0(x)) f_l(x)dx \\
&= C(\boldsymbol{\beta},\boldsymbol{\delta}) \int_{\mathbb{R}} u_j^2(F_0(x)) \exp\left\{\sum_{r=1}^{\infty} \frac{\delta_r}{\sqrt{n}} u_r(F_0(x))\right\} f_0(x)dx \\
&= C(\boldsymbol{\beta},\boldsymbol{\delta}) \int_0^1 u_j^2(y) \exp\left\{\sum_{r=1}^{\infty} \frac{\delta_r}{\sqrt{n}} u_r(y)\right\} dy \\
&= C(\boldsymbol{\beta},\boldsymbol{\delta}) \int_0^1 u_j^2(y) \left(1 + \sum_{i=1}^{\infty} \frac{\delta_r}{\sqrt{n}} u_r(y)\right. \\
&\quad \left. + \frac{1}{2n}\left[\sum_{r=1}^{\infty} \delta_r u_r(y)\right]^2 \exp\left(\xi_n(y)\right)\right) dy \\
&= C(\boldsymbol{\beta},\boldsymbol{\delta}) \left(1 + \frac{1}{\sqrt{n}} \int_0^1 u_j^2(y) \sum_{r=1}^{\infty} \delta_r u_r(y)\right. \\
&\quad \left. + \frac{1}{2n} \int_0^1 u_j^2(y) \left[\sum_{r=1}^{\infty} \delta_r u_r(y)\right]^2 \exp\left(\xi_n(y)\right) dy\right),
\end{aligned}
$$

and so

$$
\begin{aligned}
\mathrm{Var}(\sqrt{n}\tilde{\theta}_j) &= 1 + O(n^{-1/2}) - O(n^{-1}) \\
&= 1 + O(n^{-1/2}).
\end{aligned}
$$

Similarly, it is straightforward to show that $\mathrm{Cov}(\sqrt{n}\tilde{\theta}_j, \sqrt{n}\tilde{\theta}_l) = O(n^{-1/2})$ for any $j \neq l$.

It now follows immediately from the Multivariate Central Limit Theorem [cf. Theorem B, Page 30 of Serfling (1980)] that

$$(\sqrt{n}\tilde{\theta}_1, ..., \sqrt{n}\tilde{\theta}_k) \xrightarrow{\mathcal{D}} N(\boldsymbol{\delta}, \boldsymbol{I}_k) \quad \text{with} \quad \boldsymbol{\delta} = (\delta_1, ...\delta_k).$$

Using the fact that $\exp(\cdot)$ is a continuous function, the continuous mapping theorem implies that $S_k$ converges in distribution to $\mathcal{S}_1 = \sum_{j=1}^{k} \pi_j \exp[(Z_j + \delta_j)^2/2]$. $\square$

Note that the limiting distribution under the null hypothesis is a special case of Theorem 1 with $\delta_j = 0$ for all $j$. Therefore, $S_k$ converges in distribution to $\sum_{j=1}^{k} \pi_j \exp[Z_j^2/2]$ under $H_0$.

### 3.2. Limiting Distribution for Composite Null Hypothesis

We now consider the asymptotic properties of $S_k$ for a composite null hypothesis. We begin with the limiting distribution under $H_0$.

For the family $\{f(x;\boldsymbol{\beta}) : \boldsymbol{\beta} \in \mathcal{B}\}$ we need the following regularity conditions [cf. Inglot, Kallenberg and Ledwina (1997)]. These conditions are assumed to hold on any open subset $\mathcal{B}_0$ of $\mathcal{B}$. The true value of $\boldsymbol{\beta}$ is supposed to lie in $\mathcal{B}_0$.

C1. For $t, u = 1, ..., q$, $\frac{\partial}{\partial \beta_t} f(x;\boldsymbol{\beta})$ and $\frac{\partial^2}{\partial \beta_t \partial \beta_u} f(x;\boldsymbol{\beta})$ exist almost everywhere and are such that for each $\boldsymbol{\beta}_0 \in \mathcal{B}_0$, uniformly in a neighborhood of $\boldsymbol{\beta}_0$,

$$\left| \frac{\partial}{\partial \beta_t} f(x;\boldsymbol{\beta}) \right| \leq G_t(x)$$

and

$$\left| \frac{\partial^2}{\partial \beta_t \partial \beta_u} f(x;\boldsymbol{\beta}) \right| \leq K_{tu}(x),$$

where

$$\int_{\mathbb{R}} G_t(x)dx < \infty \quad \text{and} \quad \int_{\mathbb{R}} K_{tu}(x)dx < \infty.$$

C2. For $t, u = 1, ..., q$, $\frac{\partial}{\partial \beta_t}\log f(x;\boldsymbol{\beta})$ and $\frac{\partial^2}{\partial \beta_t \partial \beta_u}\log f(x;\boldsymbol{\beta})$ exist almost everywhere and are such that the Fisher information matrix,

$$\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \mathrm{E}_{\boldsymbol{\beta}}\left\{\left[\frac{\partial}{\partial \boldsymbol{\beta}}\log f(X;\boldsymbol{\beta})\right]\left[\frac{\partial}{\partial \boldsymbol{\beta}}\log f(X;\boldsymbol{\beta})\right]^{\mathrm{T}}\right\},$$

is finite, positive definite and continuous, and as $\gamma \to 0$, we have

$$\mathrm{E}_{\boldsymbol{\beta}}\left\{\sup_{\{h:\|h\|\leq\gamma\}}\left\|\frac{\partial^2}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^{\mathrm{T}}}\log f(X;\boldsymbol{\beta}+h) - \frac{\partial^2}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^{\mathrm{T}}}\log f(X;\boldsymbol{\beta})\right\|\right\} \to 0.$$

C3. For each $\boldsymbol{\beta}_0 \in \mathcal{B}_0$ there exists $\eta = \eta(\boldsymbol{\beta}_0) > 0$ with

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|<\eta}\ \sup_{x\in\mathbb{R}}\left|\frac{\partial^2}{\partial \beta_t \partial \beta_u}F(x;\boldsymbol{\beta})\right| < \infty, \quad t, u = 1, ..., q$$

and

$$\sup_{x\in\mathbb{R}}\left|\frac{\partial}{\partial \beta_t}F(x;\boldsymbol{\beta})\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} < \infty, \quad t, u = 1, ..., q.$$

The next conditions concern the orthonormal basis functions $\{\phi_j\}_{j=0}^{\infty}$ [cf. Inglot, Kallenberg and Ledwina (1997)].

S1. $\sup_{x\in[0,1]}|\phi_j'(x;\boldsymbol{\beta})| \leq c_1 j^{m_1}$ for any $j = 1, 2, ...k$ and some $c_1 > 0$, $m_1 > 0$.

S2. $\sup_{x\in[0,1]}|\phi_j''(x;\boldsymbol{\beta})| \leq c_2 j^{m_2}$ for any $j = 1, 2, ...k$ and some $c_2 > 0$, $m_2 > 0$.

**Theorem 2.** *Let $\phi_1, \phi_2, ...$ be orthonormal basis functions defined as in subsection 2.1. Assume R1-R3 and S1,S2. Suppose $\boldsymbol{Y} = (Y_1, ..., Y_k) \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{W}(\boldsymbol{\beta}_0)(\boldsymbol{I}_k - \boldsymbol{T}_{\boldsymbol{\beta}_0})\boldsymbol{W}(\boldsymbol{\beta}_0))$, where $\boldsymbol{W}(\boldsymbol{\beta}_0) = diag([1+R_1(\boldsymbol{\beta}_0)]^{1/2}, ..., [1+R_k(\boldsymbol{\beta}_0)]^{1/2})$, $R_1(\boldsymbol{\beta}_0), ..., R_k(\boldsymbol{\beta}_0)$ are defined by (2.15), $\boldsymbol{T}_{\boldsymbol{\beta}_0} = \boldsymbol{I}_{\boldsymbol{\beta}_0}^{\mathrm{T}}\boldsymbol{I}_{\boldsymbol{\beta}_0\boldsymbol{\beta}_0}^{-1}\boldsymbol{I}_{\boldsymbol{\beta}_0}$ and $\boldsymbol{I}_{\boldsymbol{\beta}_0}$ and $\boldsymbol{I}_{\boldsymbol{\beta}_0\boldsymbol{\beta}_0}$ are defined by (1.5) and*

*(1.6), respectively. Then under* $H_0$ *the statistic* $S_k = \sum_{j=1}^{k} \pi_j \exp(n\tilde{\theta}_j^2/2)$ *converges in distribution to*

$$\mathcal{S}_2 = \sum_{j=1}^{k} \pi_j \exp[Y_j^2/2],$$

*where k is an arbitrarily large but fixed number.*

*Proof.* For a composite null hypothesis, the parameter $\boldsymbol{\beta}$ is unknown, and we thus deal with the limiting distribution of $\left\{\tilde{\theta}_j\right\}_{j=1,\dots,k}$ in the form of $\left\{\tilde{\theta}_{j,\text{composite}}\right\}_{j=1,\dots,k}$, defined as A2 in subsection 2.3.2.

Let $\boldsymbol{W}(\hat{\boldsymbol{\beta}}_0) = \text{diag}([1 + R_1(\hat{\boldsymbol{\beta}}_0)]^{1/2}, \dots, [1 + R_k(\hat{\boldsymbol{\beta}}_0)]^{1/2})$ and $\underline{\tilde{\boldsymbol{\theta}}} = (\tilde{\theta}_1, \dots \tilde{\theta}_k)^{\mathrm{T}}$. Then

$$\widetilde{\Theta} = \left([1 + R_1(\hat{\boldsymbol{\beta}}_0)]^{\frac{1}{2}}\tilde{\theta}_1, \dots, [1 + R_k(\hat{\boldsymbol{\beta}}_0)]^{\frac{1}{2}}\tilde{\theta}_k\right)^{\mathrm{T}} = \boldsymbol{W}(\hat{\boldsymbol{\beta}}_0)\underline{\tilde{\boldsymbol{\theta}}}. \tag{3.2}$$

Referring to the score statistic given by Cox and Hinkley (1974), P. 324 and Thomas and Pierce (1979), P. 443, we have

$$\sqrt{n}\underline{\tilde{\boldsymbol{\theta}}} \xrightarrow{\mathcal{D}} \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{I}_k - \boldsymbol{I}_{\boldsymbol{\beta}_0}^{\mathrm{T}} \boldsymbol{I}_{\boldsymbol{\beta}_0\boldsymbol{\beta}_0}^{-1} \boldsymbol{I}_{\boldsymbol{\beta}_0}). \tag{3.3}$$

By the continuity of $\boldsymbol{I}_{\boldsymbol{\beta}_0 j}$, $\boldsymbol{I}_{\boldsymbol{\beta}_0\boldsymbol{\beta}_0}$ and the convergence in probability of $\hat{\boldsymbol{\beta}}_0$ to $\boldsymbol{\beta}_0$, we have

$$R_j(\hat{\boldsymbol{\beta}}_0) \xrightarrow{P} R_j(\boldsymbol{\beta}_0), \quad j = 1, \dots, k,$$

and so

$$\boldsymbol{W}(\hat{\boldsymbol{\beta}}_0) \xrightarrow{P} \boldsymbol{W}(\boldsymbol{\beta}_0). \tag{3.4}$$

Using (3.2), (3.3), (3.4) and Slutsky's theorem, it follows that

$$\sqrt{n}\widetilde{\Theta} \xrightarrow{\mathcal{D}} \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{W}(\boldsymbol{\beta}_0)(\boldsymbol{I}_k - \boldsymbol{T}_{\boldsymbol{\beta}_0})\boldsymbol{W}(\boldsymbol{\beta}_0)).$$

Since $\exp(\cdot)$ is a continuous function, the continuous mapping theorem implies that $S_k$ converges in distribution to $\mathcal{S}_2 = \sum_{j=1}^{k} \pi_j \exp[Y_j^2/2]$, where $\boldsymbol{Y} = (Y_1, ..., Y_k) \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{W}(\boldsymbol{\beta}_0)(\boldsymbol{I}_k - \boldsymbol{T}_{\boldsymbol{\beta}_0})\boldsymbol{W}(\boldsymbol{\beta}_0))$ with $\boldsymbol{T}_{\boldsymbol{\beta}_0} = \boldsymbol{I}_{\boldsymbol{\beta}_0}^{\mathrm{T}} \boldsymbol{I}_{\boldsymbol{\beta}_0\boldsymbol{\beta}_0}^{-1} \boldsymbol{I}_{\boldsymbol{\beta}_0}$. $\square$

Some remarks are in order concerning Theorem 2.

1. Since

$$\boldsymbol{T}_{\boldsymbol{\beta}} = \boldsymbol{I}_{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}{}^{-1}(\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} - \boldsymbol{I}_{\boldsymbol{\beta}}\boldsymbol{I}_{\boldsymbol{\beta}}^{\mathrm{T}})(\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} - \boldsymbol{I}_{\boldsymbol{\beta}}\boldsymbol{I}_{\boldsymbol{\beta}}^{\mathrm{T}})^{-1}\boldsymbol{I}_{\boldsymbol{\beta}} = \boldsymbol{R}(\boldsymbol{\beta}) - \boldsymbol{T}_{\boldsymbol{\beta}}\boldsymbol{R}(\boldsymbol{\beta}),$$

it follows that

$$(\boldsymbol{I}_k - \boldsymbol{T}_{\boldsymbol{\beta}})(\boldsymbol{I}_k + \boldsymbol{R}(\boldsymbol{\beta})) = \boldsymbol{I}_k + \boldsymbol{R}(\boldsymbol{\beta}) - \boldsymbol{T}_{\boldsymbol{\beta}} - \boldsymbol{T}_{\boldsymbol{\beta}}\boldsymbol{R}(\boldsymbol{\beta}) = \boldsymbol{I}_k.$$

The special case with $k = 1$ yields

$$(1 - \boldsymbol{I}_{\boldsymbol{\beta}j}^{\mathrm{T}} \boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}{}^{-1} \boldsymbol{I}_{\boldsymbol{\beta}j})(1 + R_j(\boldsymbol{\beta})) = 1, \quad j = 1, ..., k.$$

For any diagonal element $\Sigma_{jj}$ in covariance matrix $\Sigma = \boldsymbol{W}(\boldsymbol{\beta})(\boldsymbol{I}_k - \boldsymbol{T}_{\boldsymbol{\beta}})\boldsymbol{W}(\boldsymbol{\beta})$, $\Sigma_{jj} = \left[\boldsymbol{I}_k - \boldsymbol{I}_{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}{}^{-1} \boldsymbol{I}_{\boldsymbol{\beta}}\right]_{jj} (1 + R_j(\boldsymbol{\beta}))$, where $\left[\boldsymbol{I}_k - \boldsymbol{I}_{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}{}^{-1} \boldsymbol{I}_{\boldsymbol{\beta}}\right]_{jj}$ is the $j$th diagonal element of $\boldsymbol{I}_k - \boldsymbol{I}_{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}{}^{-1} \boldsymbol{I}_{\boldsymbol{\beta}}$.

A further investigation gives that $\left[\boldsymbol{I}_k - \boldsymbol{I}_{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}{}^{-1} \boldsymbol{I}_{\boldsymbol{\beta}}\right]_{jj} = 1 - \boldsymbol{I}_{\boldsymbol{\beta}j}^{\mathrm{T}} \boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}{}^{-1} \boldsymbol{I}_{\boldsymbol{\beta}j}$, and so $\Sigma_{jj}$=1 for $j = 1, ..., k$, and $\sqrt{n}[1 + R_j(\hat{\boldsymbol{\beta}}_0)]^{1/2}\tilde{\theta}_j$ indeed leads to a normalized statistic.

2. Let $u_1, u_2, ...$ be orthonormal basis functions defined as in subsection 2.1.

   – For testing exponentiality, each off diagonal element $\Sigma_{ij}$ of the covariance matrix $\Sigma = \boldsymbol{W}(\boldsymbol{\beta})(\boldsymbol{I}_k - \boldsymbol{T}_{\boldsymbol{\beta}})\boldsymbol{W}(\boldsymbol{\beta})$ is

$$-\frac{\left(\int_0^\infty y f_0{}'(y) u_i[F_0(y)]dy\right)\left(\int_0^\infty y f_0{}'(y) u_j[F_0(y)]dy\right)}{\left(1 - \left[\int_0^\infty y f_0{}'(y) u_i[F_0(y)]dy\right]^2\right)^{\frac{1}{2}} \left(1 - \left[\int_0^\infty y f_0{}'(y) u_j[F_0(y)]dy\right]^2\right)^{\frac{1}{2}}}.$$

– For testing normality, each off diagonal element $\Sigma_{ij}$ of the covariance matrix $\Sigma = \boldsymbol{W}(\boldsymbol{\beta})(\boldsymbol{I}_k - \boldsymbol{T}_{\boldsymbol{\beta}})\boldsymbol{W}(\boldsymbol{\beta})$ is

$$-\frac{2I_{\mu i}I_{\mu j} + I_{\sigma i}I_{\sigma j}}{\left(2 - 2I_{\mu i}^2 - I_{\sigma i}^2\right)^{\frac{1}{2}}\left(2 - 2I_{\mu j}^2 - I_{\sigma j}^2\right)^{\frac{1}{2}}},$$

where

$$I_{\mu i} = \int_{-\infty}^{\infty} {f_0}'(y)u_i[F_0(y)]dy, \quad j = 1, ..., k,$$

and

$$I_{\sigma i} = \int_{-\infty}^{\infty} y{f_0}'(y)u_i[F_0(y)]dy, \quad j = 1, ..., k.$$

3. Using an approach similar to that in the proof of Theorem 2, we find that the limiting distribution under local alternative $f_j$, defined by (3.1), is $\sum_{j=1}^{k} \pi_j \exp[(Y_j + (1 - \sum_{t=1}^{q} i_{jt}\boldsymbol{I}_{\boldsymbol{\beta}_{tj}})\delta_j)^2/2]$, where $Y_j$ is defined the same as in Theorem 2, $q$ is the dimension of $\boldsymbol{\beta}$, $i_{jt}$ is the element in the $j$th row and $t$th column of $\boldsymbol{I}_{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{I}_{\boldsymbol{\beta\beta}}{}^{-1}$, and $\boldsymbol{I}_{\boldsymbol{\beta}_{tj}}$ is the element in the $t$th row and $j$th column of $\boldsymbol{I}_{\boldsymbol{\beta}}$.

The limiting distributions demonstrate that $S_k$ can detect $1/\sqrt{n}$ alternatives whenever at least one of the Fourier coefficients $\delta_1, ..., \delta_k$ is nonzero.

# CHAPTER IV

## THE PERFORMANCE OF TESTS

In this chapter we present the results of an extensive Monte Carlo study to see how well the tests perform, including evaluating the choice of the number of Fourier coefficients, optimal weights, empirical critical values and power of the considered tests for testing simple and composite null hypotheses.

We first clarify the test statistic and related parameters that will be used in the simulations. The proposed test statistic is

$$S_k = \sum_{j=1}^{k} \pi_j \exp\left(\frac{n\tilde{\theta}_j^2}{2}\right), \tag{4.1}$$

and $H_0$ will be rejected for large values of $S_k$. The statistic $\tilde{\theta}_j$ will take the forms of A1 and A2 in the case of simple and composite null hypotheses, respectively. Let $u_j, j = 1, 2, ...,$ be orthonormal on the interval [0,1], defined as in subsection 2.1. Examples of basis functions $u_1, u_2, ...$ that could be used are Legendre polynomials, trigonometric functions and wavelets. In this chapter, we use orthonormal Legendre polynomials with respect to Lebesgue measure defined on [0,1] as $u_j$s. Then the basis functions $\phi_j(\cdot;\boldsymbol{\beta}) = u_j(F(x;\boldsymbol{\beta}))$, $j = 1, 2, ...,$ defined as in subsection 2.1.

For simulations concerning the proposed test statistic $S_k$ defined as in (4.1) we have to choose the number of Fourier components $k$ and the weights $\pi_j$. We start each subsection with comments on the choice of $k$ and $\pi_j$. We then present the resulting power of the proposed tests and compare them with some other commonly used tests.

## 4.1. Testing for Simple Hypotheses

In the simulation study of simple hypotheses, we consider the following two types of alternatives:

$$p_j(x;\rho) = 1 + \rho\cos(\pi j x), \quad \rho \in (0,1], \quad j = 1,2,..., \tag{4.2}$$

$$g_k(x;\boldsymbol{\theta}) = C_k(\boldsymbol{\theta}) \exp\left\{\sum_{j=1}^{k} \theta_j u_j(x)\right\}, \quad k = 1,2,..., \tag{4.3}$$

where $C_k(\boldsymbol{\theta})$ is the normalizing factor and $u_1,...,u_k$ are orthonormal Legendre polynomials on [0,1]. In simulations, we use rejection sampling to generate data from these alternates. Fourier coefficients are defined as:

$$\mathrm{E}(\tilde{\theta}_{\mathrm{j,simple}}) = \int_0^1 u_j(x) f_a(x) dx, \tag{4.4}$$

where $f_a(x)$ is the considered alternative. The alternative here is the density of $F_0(x;\boldsymbol{\beta})$ and $F_0(x;\boldsymbol{\beta})$ is the cumulative distribution function under $H_0$.

### 4.1.1. Number of Fourier Components $k$

In this subsection we investigate how $k$ (in $S_k$) affects critical values and power of the test when we take $\pi_j = 1/j^2$, $j = 1,...,k$. We do simulations under the null hypothesis based on 10,000 replications to determine 0.05 level critical values for different $k$ and different sample size $n$. The power as $k$ ranges from 5 to 45 by 5 is obtained by 10,000 replications at sample size $n = 100$ and significant level $\alpha = 0.05$.

There is empirical evidence that $S_k$ changes smoothly as $k$ increases. For illustration, see Figure 1, which shows the critical values of proposed test statistic $S_k$ as a function of $k$ with sample sizes $n = 50$ and $n = \infty$. We have shown in subsection 3.1 that the limiting distribution of $S_k$ under $H_0$ is that of $\sum_{j=1}^{k} \pi_j \exp(Z_j^2/2)$, where

$Z_1, ..., Z_k$ are i.i.d. standard normal random variables. Critical values for $n = \infty$ were obtained by simulating values of $\sum_{j=1}^{k} \pi_j \exp(Z_j^2/2)$. Figure 1 shows that when sample size $n$ goes to $\infty$, critical values increase at a slower rate than at $n = 50$.

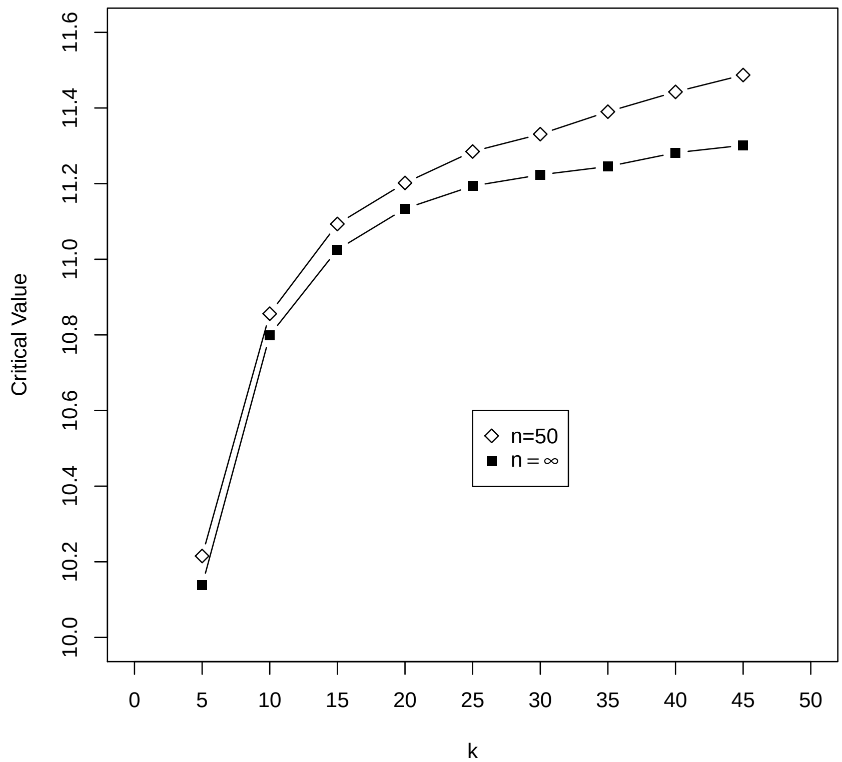


Figure 1.

The behavior of simulated critical values of proposed test statistic $S_k$ as a function of $k$. $\alpha = 0.05$, 10,000 Monte Carlo runs.

A good choice of $k$ is related to the $j$ in basis function $\phi_j$ that has the largest corresponding Fourier coefficient. Roughly speaking, if the only nonzero coefficient is at $j = 10$ or the largest Fourier coefficient appears at $j = 10$, then any choice of $k$ that is at least 10 should "work."

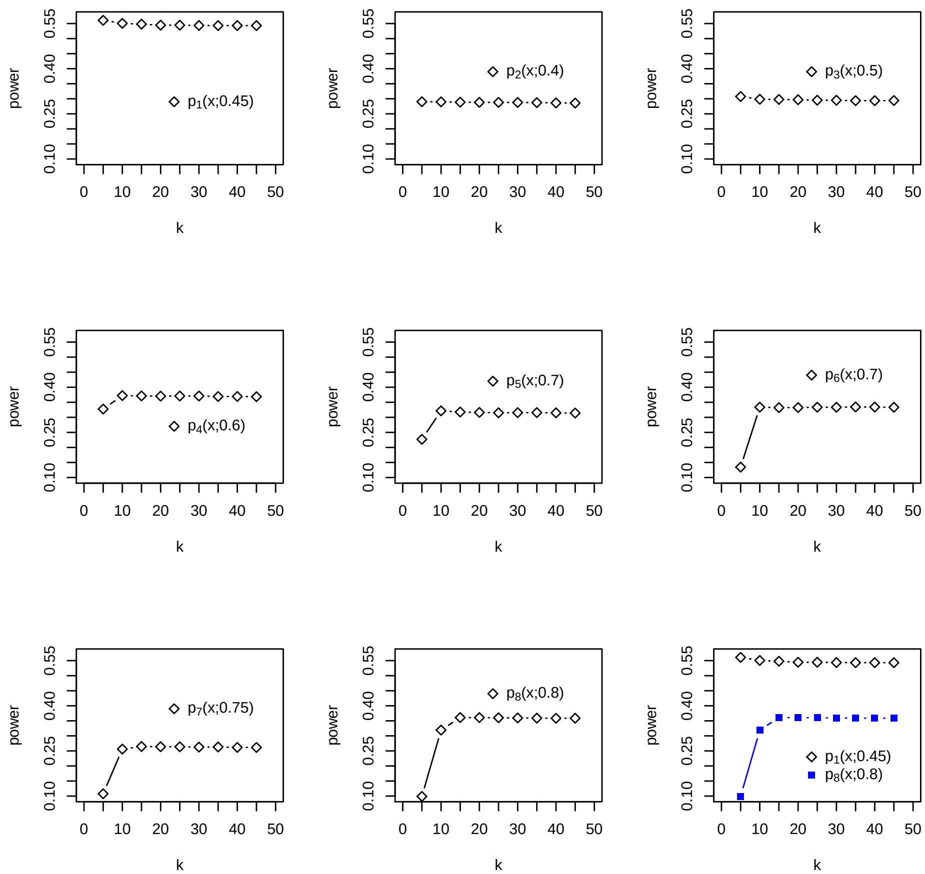


Figure 2.

The behavior of simulated powers of proposed test statistic $S_k$ as a function of $k$ under the alternative $p_j(x;\rho)$. $n = 100$, $\alpha = 0.05$, 10,000 Monte Carlo runs.

Figures 2 and 3 show the change of power as $k$ increases under the alternative densities (4.2) and (4.3) respectively. In each figure, the graphs are arranged in order of increasing frequency. It is shown that small $k = 5$ works just a little bit better than larger $k$ when the alternative densities are low frequency, e.g. $p_1$, $p_2$, $p_3$, $g_1$, $g_2$ and $g_3$. But for highly oscillating alternatives, e.g. $p_4 - p_8$, $g_6$ and $g_8$, $k = 5$ is not enough, and we need a larger $k$ like 10 to 20 to guarantee better power. This result is not surprising in light of our discussion in the previous paragraph. More details will be presented at the end of this subsection.

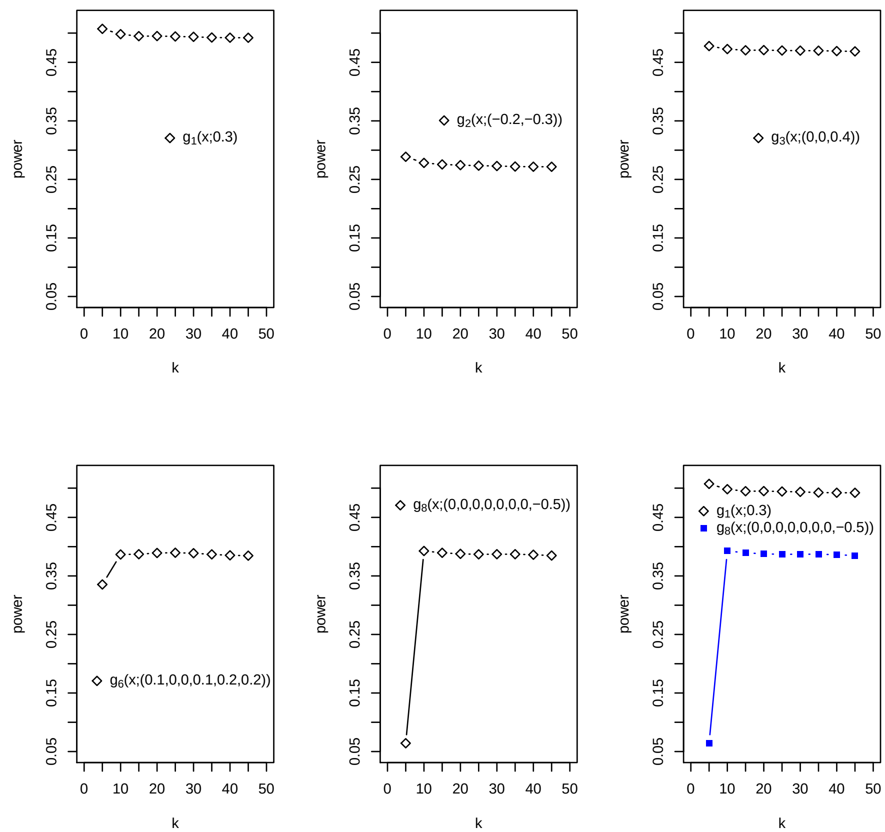


Figure 3.

The behavior of simulated powers of proposed test statistic $S_k$ as a function of $k$ under the alternative $g_k(x;\boldsymbol{\theta})$. $n = 100$, $\alpha = 0.05$, 10,000 Monte Carlo runs.

However, the powers do not vary much for different values of $k$ in the range of 15 to 45. As shown by the last graphs of both Figures 2 and 3, the powers when testing low frequency (i.e. $p_1$ and $g_1$) change little with $k$ in comparison with power when testing high frequency (i.e. $p_8$ and $g_8$), even if $k = 5$ works slightly better than larger $k$ for low frequency alternatives. These observations, along with the fact that alternatives of higher frequency than $p_8$ or $g_8$ are very uncommon in practice, suggest that a choice of $k$ around 20 would generally work well.

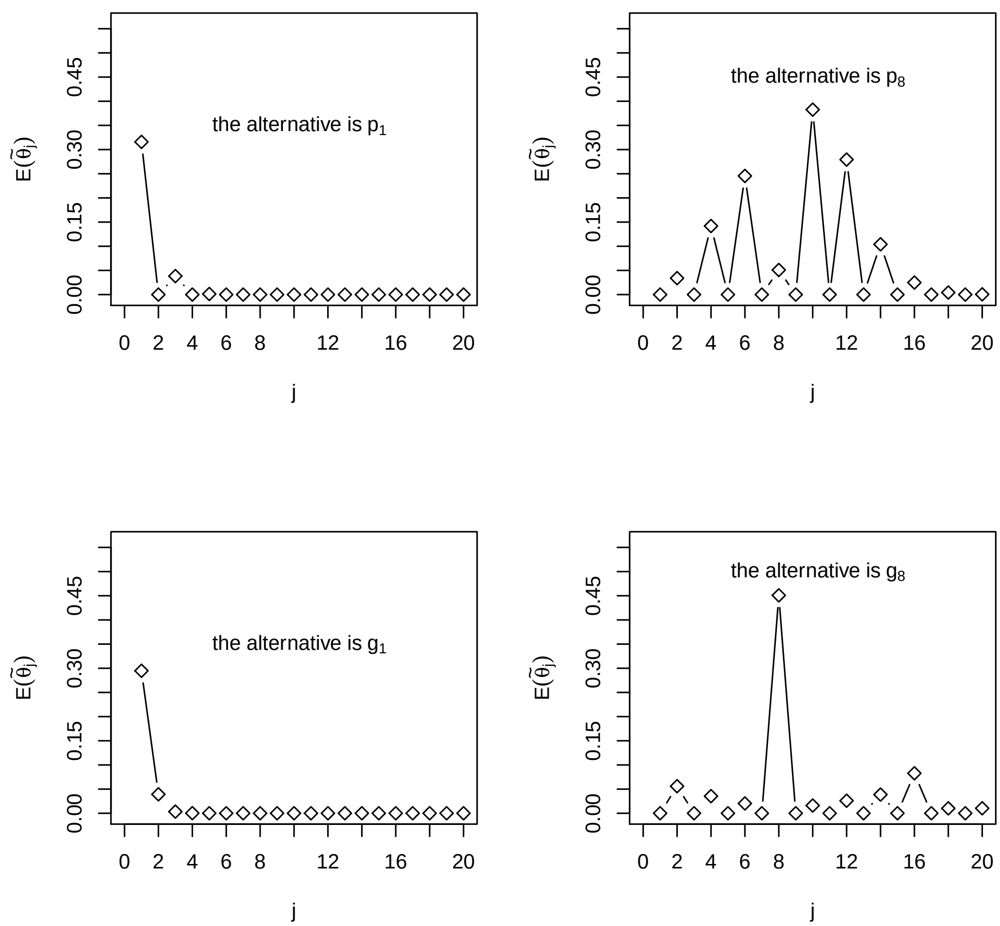


Figure 4.

The behavior of Fourier coefficients, $E(\tilde{\theta}_{j,\text{simple}})$, as a function of $j$ under the alternatives $p_1$, $p_8$, $g_1$ and $g_8$.

Figure 4 shows Fourier coefficients, $E(\tilde{\theta}_{j,\text{simple}})$, for alternatives $p_1, p_8, g_1, g_8$. The results coincide with our discussion above. We also find that

- Under the alternative $p_k$ given by (4.2), infinitely many of $\mathrm{E}(\tilde{\theta}_j)$ will be nonzero unless the test statistic uses cosine basis functions.

- Under the alternative $g_k$ given by (4.3), infinitely many of $\mathrm{E}(\tilde{\theta}_j)$ will be nonzero even if the $\phi_j$s in $g_k$ are the same as the basis functions in the test statistic.

We will use $\boldsymbol{k = 20}$ in our non-adaptive tests since the preliminary results of this subsection suggest that $k = 20$ has reasonably good power against both low and high frequency alternatives. As mentioned before, densities having largest $\left|\mathrm{E}(\tilde{\theta}_{j,\text{simple}})\right|$ for $j > 20$ are extremely unusual in practice.

### 4.1.2. Prior Probabilities $\pi_j$

The last subsection suggests that the number of Fourier components $k$ does not play a crucial role, since the power of proposed test statistic $S_k$ is almost stable for $k$ between 15 and 45. On the other hand, the choice of prior probabilities $\pi_1, \pi_2, \ldots$ may be more important.

Assume that the alternative is represented by linear combinations of polynomials, and $\phi_1, \phi_2, \ldots$ corresponds to the basis functions arranged in order of increasing frequency. One could argue that it is natural to place larger prior probabilities on the Fourier coefficient with lower index. Doing so will tend to increase the power of the resulting test if one's assumptions are justified. As argued in Chapter II, we consider $\pi_j = 1/j^c$ for $c > 1$. Our task turns now to a good choice of $c$. We use $k = 20$ as suggested by subsection 4.1.1.

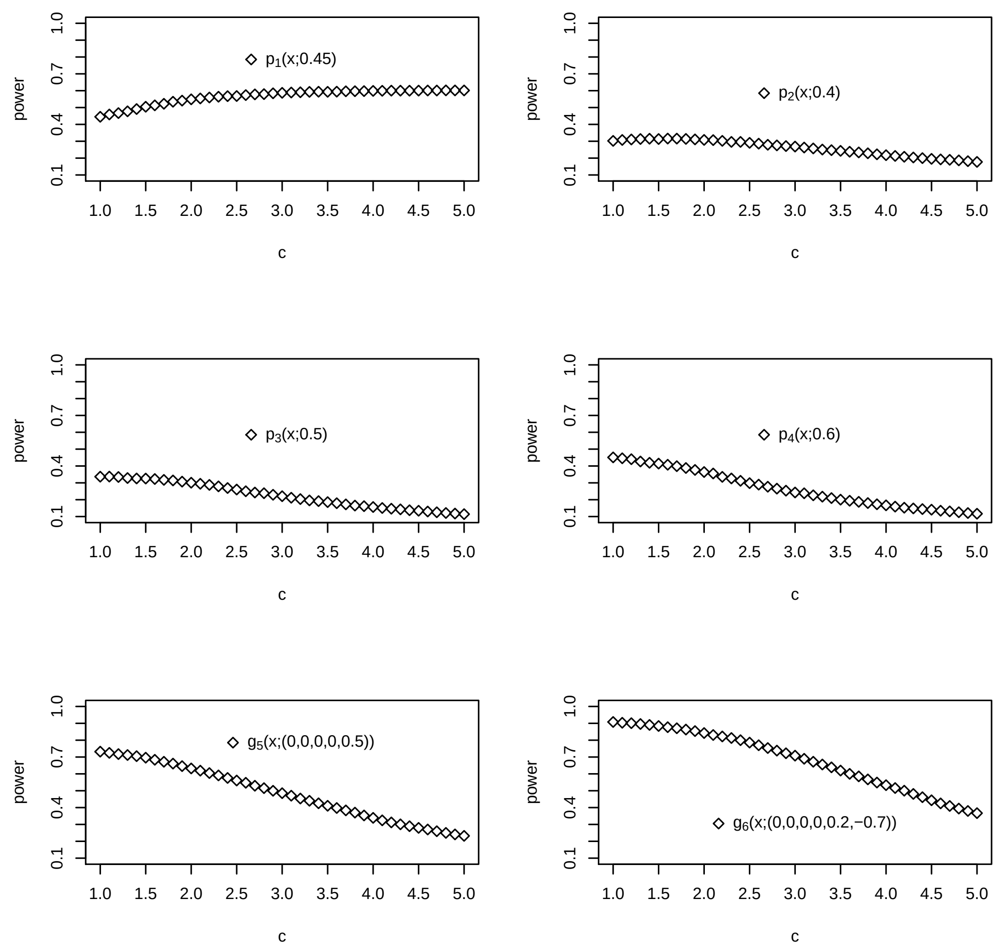


Figure 5.

The behavior of simulated powers of proposed test statistic $S_k$ as a function of $c$ under the alternatives $p_1$, $p_2$, $p_3$, $p_4$, $g_5$ and $g_6$, where $\pi_j = 1/j^c$. $n = 100$, $\alpha = 0.05$, 10,000 Monte Carlo runs.

Figures 5 and 6 show the performance of $S_k$ as $c$ ranges from 1 to 5 by 0.1. The graphs are placed in order of increasing frequency, as measured by $\mathrm{E}(\tilde{\theta}_j)$. We notice that the best power is at large $c$ when the alternative is low frequency, i.e. $c = 5$ for $p_1$ and $c = 1.6$ for $p_2$. For highly oscillating alternatives, the smaller the $c$

is, the higher the power. In $p_3, p_4, g_5, g_6, g_7, g_8, g_9$, power decreases with increasing $c$. These results coincide with our expectation, since large $c$ down-weights high frequency alternatives and small $c$ emphasizes higher frequency alternatives. However, the last graph in Figure 6 shows that average power over the various alternatives peaks at around $c = 2$. Subsequently, we will consider $\pi_j = 1/j^2$ as a good choice of prior probabilities.

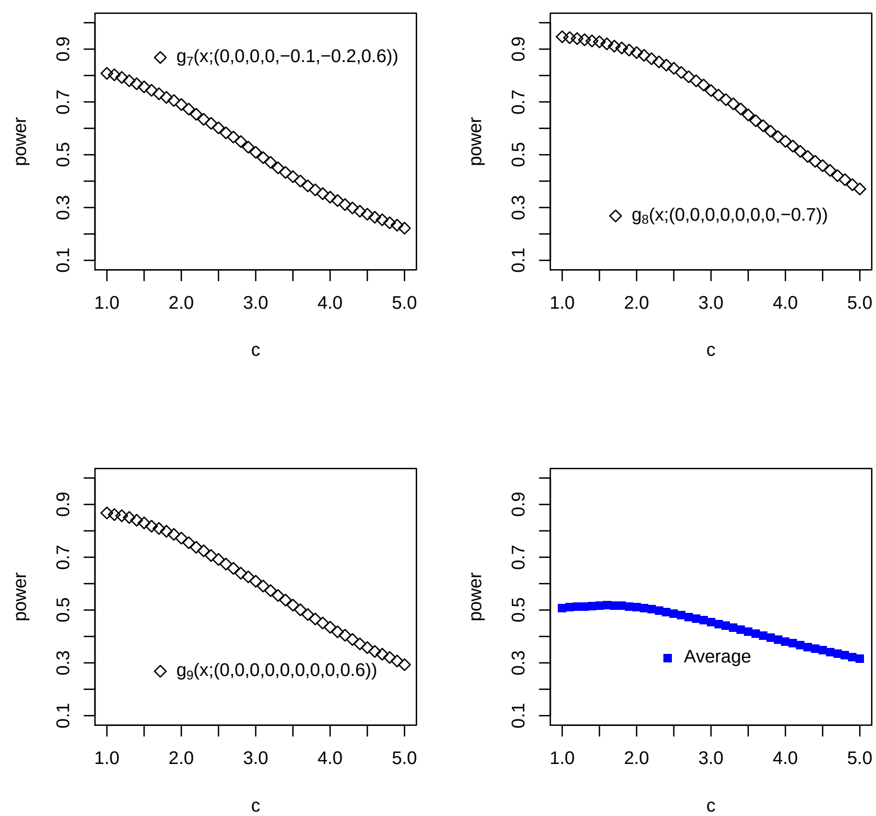


Figure 6.

The behavior of simulated powers of proposed test statistic $S_k$ as a function of $c$ under the alternatives $g_7$, $g_8$, $g_9$, and the average, where $\pi_j = 1/j^c$. $n = 100$, $\alpha = 0.05$, 10,000 Monte Carlo runs.

One point attracting our attention in our extensive power comparison is that the proposed test does not perform well against alternatives with $\phi_2$ having the largest Fourier coefficient. Further investigation discloses that the problem may be caused by the fact that the prior probabilities put on $\phi_j$ decay too quickly from $j = 1$ to $j = 2$ in comparison to the remaining weights. Thus, we would like to consider $\pi_j = 1/(1+j)^2$. In addition to being reasonably noninformative, these probabilities lead to a good compromise so that power will be improved at higher frequency alternatives without hurting too much at lower frequency alternatives.

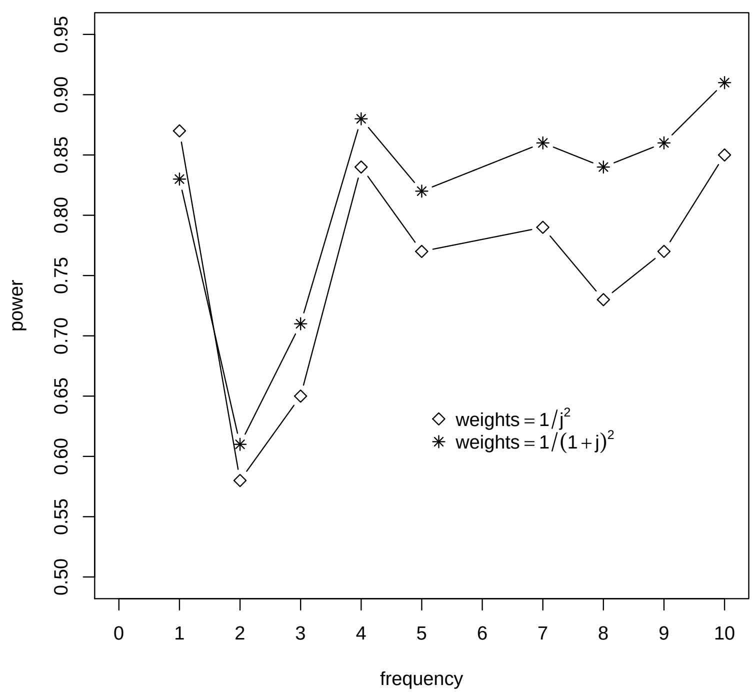


Figure 7.

The behavior of simulated average powers of proposed test statistic $S_k$ according to the different weights when testing for simple hypotheses. $n = 100$, $\alpha = 0.05$, 10,000 Monte Carlo runs.

Figure 7 presents the differences in power between $\pi_j = 1/j^2$ and $\pi_j = 1/(1+j)^2$, where the frequency is measured by the number $j$ corresponding to $j$ in basis function $\phi_j$ with the largest Fourier coefficient, each power is the average over the various alternatives, $g_k$ and $p_j$, with relative frequencies. As expected, the power with $\pi_j = 1/(1+j)^2$ increases somewhat at higher frequencies, including $j = 2$, but does not decrease too much at frequency $j = 1$. We will thus take $\boldsymbol{\pi_j = 1/(1+j)^2}$ as the prior probabilities used in the next subsection to compare with other omnibus tests.

#### 4.1.3. Power Comparisons in the Simulation Study

Based on the results in subsections 4.1.1 and 4.1.2, we will take $k = 20$ and $\pi_j = 1/(1+j)^2$ to do power comparisons in this subsection. From the enormous number of test statistics for testing uniformity available in the literature, we focus our attention on three that have proven to be powerful. One of these is $Z_A$, introduced by Zhang (2002) as an improved construction compared to traditional tests and defined as

$$Z_A = -\sum_{i=1}^{n} \left[ \frac{\log(F(X_{(i)}))}{n - i + \frac{1}{2}} + \frac{\log(1 - F(X_{(i)}))}{i - \frac{1}{2}} \right],$$

where $F(x)$ is a hypothesized distribution function and the $X_{(i)}$s are the order statistics from a random sample. The other two statistics are $N_S$ and $N_T$, adaptive statistics proposed by Ledwina (1994) and Inglot and Ledwina (2006) with different selection rules. The rules used by $N_S$ and $N_T$ are BIC and one designed for highly oscillating alternatives. Details about the selection rules have been introduced in Chapter I.

We do simulations under the null hypothesis based on 10,000 replications to determine the 0.05 level critical value for each test. Each replication has sample size $n = 100$. The critical values so determined for the simple hypothesis are 3.421 for $Z_A$, 5.636 for $N_S$, 5.987 for $N_T$ and 4.348 for $S_{20}$.

Table 1.

Powers of Zhang's test, Ledwina's tests based on $N_S$ and $N_T$ and one based on $S_{20}$ under alternative $g_k(x; \boldsymbol{\theta})$.

| Parameters | | The five largest (in absolute value) | | | | | Powers(%) | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $k$ | $\theta$ | Fourier coefficients × 1000 | | | | | $Z_A$ | $N_S$ | $N_T$ | $S_{20}$ |
| 1 | 0.3 | [**1**]295 | [**2**]39 | [**3**]3 | [**4**]1 | [**5**]1 | 70 | 74 | 71 | 77 |
| 2 | (-0.2,-0.3) | [**2**]255 | [**1**]151 | [**3**]40 | [**4**]30 | [**5**]5 | 70 | 75 | 73 | 64 |
| 3 | (0,0,0.4) | [**3**]393 | [**6**]66 | [**2**]47 | [**4**]44 | [**5**]13 | 53 | 87 | 87 | 89 |
| 4 | (0.1,0.15,-0.25, -0.35) | [**4**]335 | [**3**]235 | [**1**]150 | [**2**]137 | [**7**]66 | 47 | 85 | 86 | 88 |
| 5 | (0,0,0,0,0.4) | [**5**]397 | [**10**]66 | [**2**]46 | [**8**]40 | [**4**]38 | 31 | 56 | 76 | 82 |
| 6 | (0.1,0,0,0.1, 0.2,0.2) | [**5**]277 | [**6**]270 | [**4**]176 | [**1**]167 | [**7**]77 | 62 | 61 | 66 | 75 |
| 8 | (0,0,0,0,0,0,0, -0.5) | [**8**]451 | [**2**]56 | [**4**]35 | [**12**]26 | [**6**]23 | 7 | 30 | 90 | 92 |

$n = 100$, $\alpha$=0.05, 10,000 Monte Carlo runs.

In the simulation study we consider the alternatives given by (4.2) and (4.3). To have some insight into the structure and magnitude of the alternatives, in each case we calculate twenty Fourier coefficients (in Legendre basis) of the underlying distributions. The five largest (from these 20) Fourier coefficients are presented in Tables 1 and 2. Each bold face number $j$ corresponds to $j$ in basis function $\phi_j$. We also display the powers of the four tests considered.

The results are encouraging. The new test statistic based on $S_{20}$ has a stable and relatively high power for the whole range of alternatives considered here. It dominates

$Z_A$ in both smooth and highly oscillating cases. $S_{20}$ is much more powerful than $N_S$ for high frequency alternatives and is comparable to it for smooth alternatives. The performance of $S_{20}$ is even slightly better than $N_T$. These results are impressive since $S_{20}$ is not adaptive and does not choose the number of Fourier components through data driven means.

Table 2.

Powers of Zhang's test, Ledwina's tests based on $N_S$ and $N_T$ and one based on $S_{20}$ under alternative $p_j(x;\rho)$.

| Parameters | | The five largest (in absolute value) | | | | | Powers(%) | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $k$ | $\rho$ | Fourier coefficients × 1000 | | | | | $Z_A$ | $N_S$ | $N_T$ | $S_{20}$ |
| 1 | 0.45 | [**1**]316 | [**3**]38 | [**5**]2 | [**7**]1 | [**9**]1 | 76 | 81 | 78 | 83 |
| 2 | 0.40 | [**2**]272 | [**4**]78 | [**6**]7 | [**8**]1 | [**10**]1 | 18 | 70 | 68 | 61 |
| 3 | 0.50 | [**3**]317 | [**5**]149 | [**1**]39 | [**7**]25 | [**9**]2 | 34 | 65 | 66 | 71 |
| 4 | 0.60 | [**4**]335 | [**6**]233 | [**2**]102 | [**8**]58 | [**10**]8 | 17 | 64 | 72 | 82 |
| 5 | 0.70 | [**7**]319 | [**5**]317 | [**3**]173 | [**9**]109 | [**1**]20 | 41 | 60 | 78 | 86 |
| 6 | 0.70 | [**8**]346 | [**6**]231 | [**4**]208 | [**10**]155 | [**2**]53 | 14 | 46 | 77 | 84 |
| 7 | 0.75 | [**9**]377 | [**5**]238 | [**11**]216 | [**7**]147 | [**3**]96 | 33 | 33 | 82 | 86 |
| 8 | 0.80 | [**10**]383 | [**12**]279 | [**6**]245 | [**4**]142 | [**8**]51 | 13 | 34 | 90 | 92 |

$n = 100$, $\alpha$=0.05, 10,000 Monte Carlo runs.

### 4.2. Testing for Composite Hypotheses

Our simulations for composite hypotheses will focus on location-scale families (i.e. testing exponentiality and normality), but are fairly comprehensive in that setting. We consider the broad class of alternatives given in Table 3 and alternatives based on

$p_j$ and $g_k$, given by (4.2) and (4.3). In Table 3, $U$ denotes a $N(0,1)$ random variable and $R$ denotes a uniform random variable on (0,1). Note that the Weibull alternative is a scale family with respect to $b$, the Lognormal LN is a scale family with respect to $\exp(g/d)$, the Shifted exponential is a location-scale family with respect to $l$ and $b$.

Table 3.

Alternatives used for testing composite hypotheses.

| alternative | density/definition |
|---|---|
| Weibull($b$;$k$) | $bk(bx)^{k-1}\exp\{-(bx)^k\}, \quad x>0$ |
| $\chi^2_k$ | $\{2^{\frac{1}{2}k}\Gamma(k/2)\}^{-1}x^{\frac{1}{2}k-1}\exp(-\frac{1}{2}x), \quad x>0$ |
| LN($g$; $d$) | $d(x\sqrt{2\pi})^{-1}\exp-\frac{1}{2}(d\log x+g)^2, \quad x>0$ |
| Beta($p$; $q$) | $x^{p-1}(1-x)^{q-1}\{B(p,q)\}^{-1}, \quad 0\le x\le 1$ |
| Uniform($a$; $b$) | $(b-a)^{-1}, \quad a\le x\le b$ |
| Shifted exp.($l$; $b$) | $b\exp\{-(x-l)b\}, \quad x\ge l$ |
| Pareto($a$; $k$) | $ak^a x^{-a-1}, \quad x\ge k$ |
| Shifted Pareto | $2(1+x)^{-3}, \quad x>0$ |
| Logistic | $e^x(1+e^x)^{-2}, \quad -\infty<x<\infty$ |
| SU($g$; $d$) | $U=g+d\sinh^{-1}(X), \quad -\infty<X<\infty$ |
| TU($l$) | $X=R^l-(1-R)^l, \quad -1\le X\le 1$ |
| SC($p$; $d$) | $(2\pi)^{-\frac{1}{2}}[(p/d)\exp(-\frac{1}{2}x^2/d^2)+(1-p)\exp(-\frac{1}{2}x^2)], -\infty<x<\infty$ |
| LC($p$; $m$) | $(2\pi)^{-\frac{1}{2}}[p\exp\{-\frac{1}{2}(x-m)^2)\}+(1-p)\exp(-\frac{1}{2}x^2)], -\infty<x<\infty$ |
| SB($g$; $d$) | $U=g+d\log X/(1-X),\ 0<X<1$ |

The Fourier coefficients for testing exponentiality are defined as:

$$\int u_j\left[F(x/\mu)\right]f_a(x)dx, \quad j=1,2,..., \tag{4.5}$$

and for testing normality they are:

$$\int u_j \left[\Phi((x-\mu)/\sigma)\right] f_a(x)dx, \quad j = 1, 2, ... \tag{4.6}$$

where $F(x) = 1 - \exp(-x)$, $\Phi(x)$ is the cumulative distribution function of standard normal, $\mu$ and $\sigma$ are the mean and standard deviation of the considered alternative $f_a(x)$, respectively.

### 4.2.1. Number of Fourier Components $k$

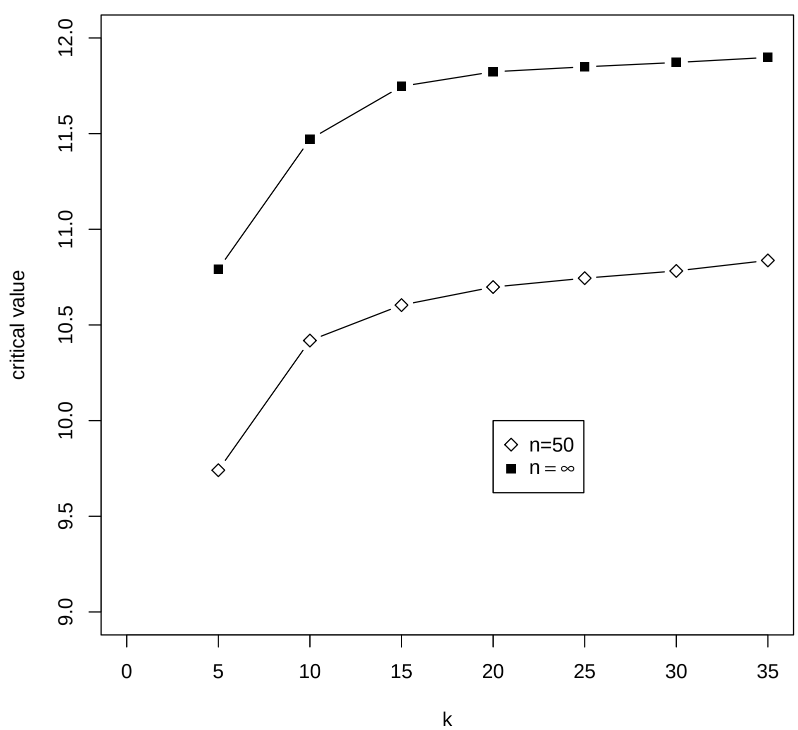


Figure 8.

The behavior of simulated critical values of proposed test statistic $S_k$ as a function of $k$ when testing for exponentiality. $\alpha = 0.05$, 10,000 Monte Carlo runs.

We now consider the effect of the number, $k$, of Fourier components when testing for composite null hypotheses. We use $\pi_j = 1/j^2$, $j = 1, ..., k$ as the prior probabilities in this subsection. We will see that under both the null and alternative hypotheses, the number of Fourier components $k$ affects the performance of proposed test statistic $S_k$. Similar to the simple null hypothesis, we also do simulations based on 10,000 replications to determine critical values and power as $k$ ranges from 5 to 35 by 5.

Figures 8 and 9 show empirical evidence that percentiles of $S_k$ change smoothly as $k$ increases when testing exponentiality and normality, respectively. In both cases, the trend of critical values with sample sizes $n = 50$ and $n = \infty$ are presented. We showed in subsection 3.2 that the limiting distribution of $S_k$ under $H_0$ is $\sum_{j=1}^{k} \pi_j \exp(Y_j^2/2)$ and $(Y_1, ..., Y_k) \sim N(\mathbf{0}, \boldsymbol{W}(\boldsymbol{\beta})(\boldsymbol{I}_k - \boldsymbol{T}_{\boldsymbol{\beta}})\boldsymbol{W}(\boldsymbol{\beta}))$, where the covariance matrix $\boldsymbol{W}(\boldsymbol{\beta})(\boldsymbol{I}_k - \boldsymbol{T}_{\boldsymbol{\beta}})\boldsymbol{W}(\boldsymbol{\beta})$, is defined in remark 2 of subsection 3.2. We simulate values of $\sum_{j=1}^{k} \pi_j \exp(Y_j^2/2)$ to get critical values for $n = \infty$. When sample size $n$ increases from 50 to $\infty$, critical values do not change a great deal.

We intend to find a value of $k$ so that the proposed tests will have good power under both low frequency alternatives (the largest Fourier coefficient corresponds to smaller $j$, i.e. $j = 1, 2, 3$) and high frequency alternatives (the largest Fourier coefficient corresponds to larger $j$, i.e. $j = 5, 6, 7, ...$).

As observed in the case of a simple null, the $j$ in the basis function $\phi_j$ with the largest corresponding Fourier coefficient has an effect on the choice of $k$. Roughly speaking, if the largest Fourier coefficient appears at $j = 10$ or the only nonzero coefficient is at $j = 10$, then any choice of $k$ that is at least 10 should result in a powerful test.

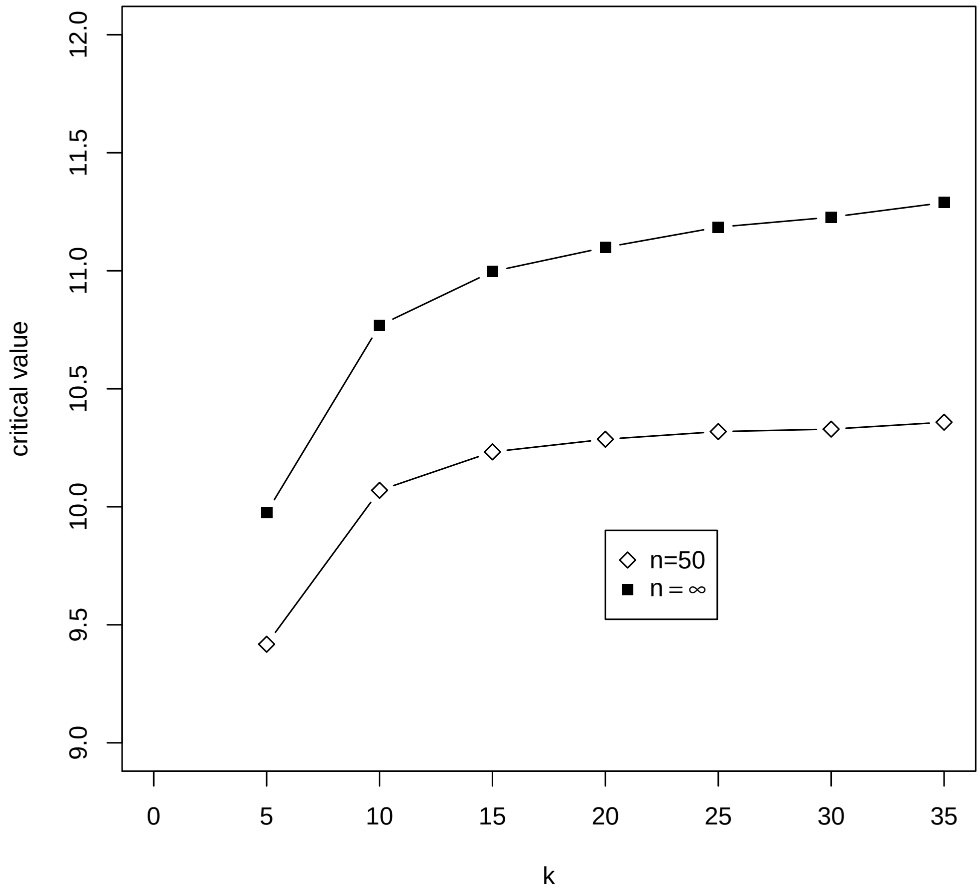


Figure 9.

The behavior of simulated critical values of proposed test statistic $S_k$ as a function of $k$ when testing for normality. $\alpha = 0.05$, 10,000 Monte Carlo runs.

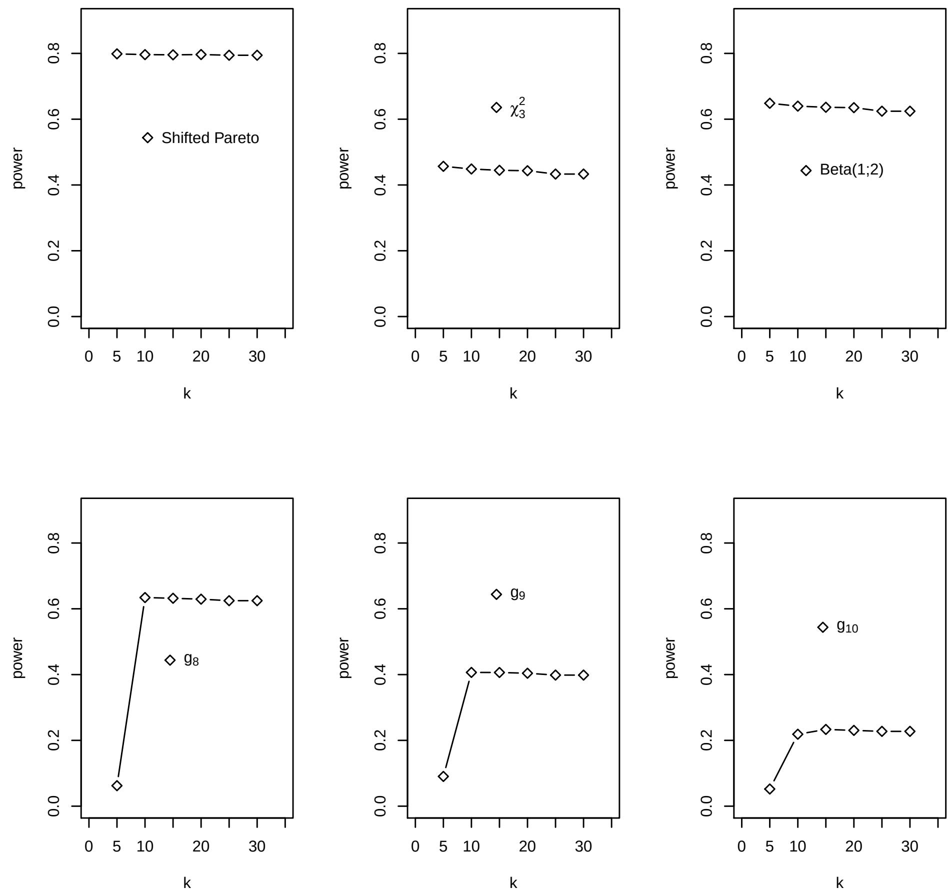


Figure 10.

The behavior of simulated powers of proposed test statistic $S_k$ as a function of $k$ when testing for exponentiality. $n = 50$, $\alpha = 0.05$, 10,000 Monte Carlo runs.

Figures 10 and 11 show the change in power as $k$ increases under the alternative densities given at the beginning of subsection 4.2. In each figure, the graphs are ordered by increasing frequency. For testing exponentiality, we notice that small $k = 5$ works the best among the ranges 5 to 45 when the alternative densities are low frequency, e.g. Shifted Pareto, $\chi^2_3$, Beta(1;2). However, $k = 5$ does not work well enough for highly oscillating alternatives, e.g. $g_8$, $g_9$, $g_{10}$. We need a larger $k$ like 10 to 20 to achieve better power. Similar arguments hold for testing normality.

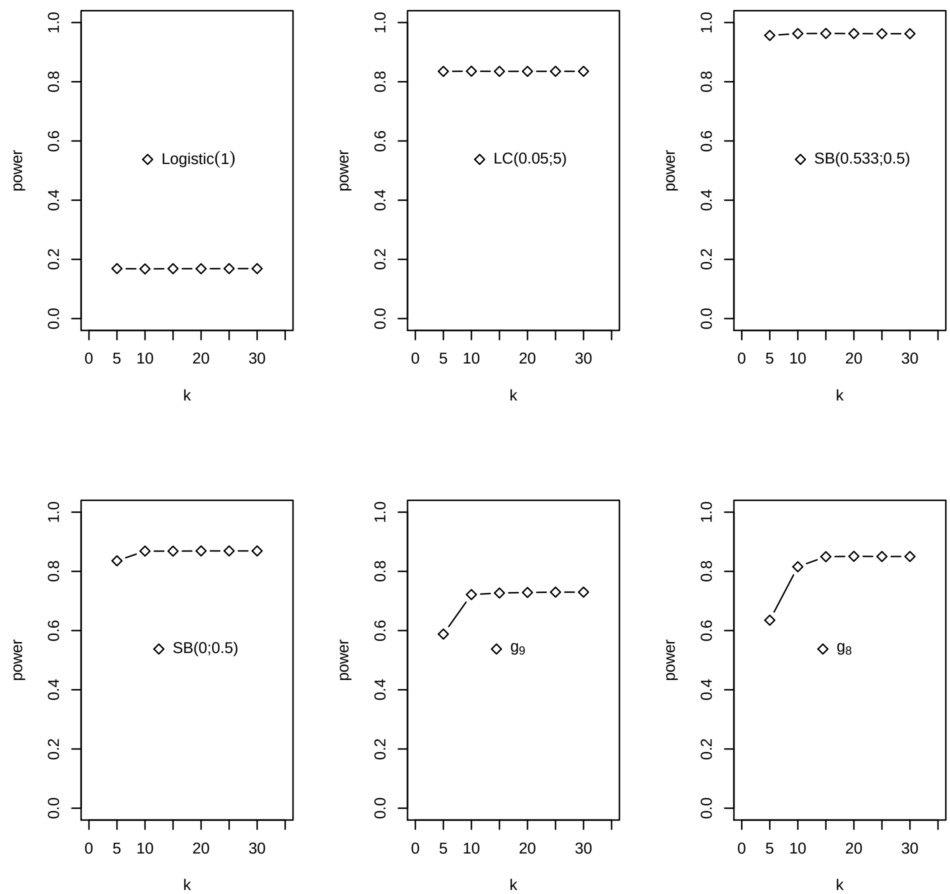


Figure 11.

The behavior of simulated powers of proposed test statistics $S_k$ as a function of $k$ when testing for normality. $n = 50$, $\alpha = 0.05$, 10,000 Monte Carlo runs.

With a further investigation of Figures 10 and 11, we find the powers do not change much for different $k$ in the range 10 to 35. Figure 12 shows that the powers for testing low frequency (i.e. $\chi^2_3$, LC(0.05;5)) just change slightly with increasing

$k$ compared with power when testing high frequency (i.e. $g_8$), even if $k = 5$ works slightly better than larger $k$ for low frequency alternatives.

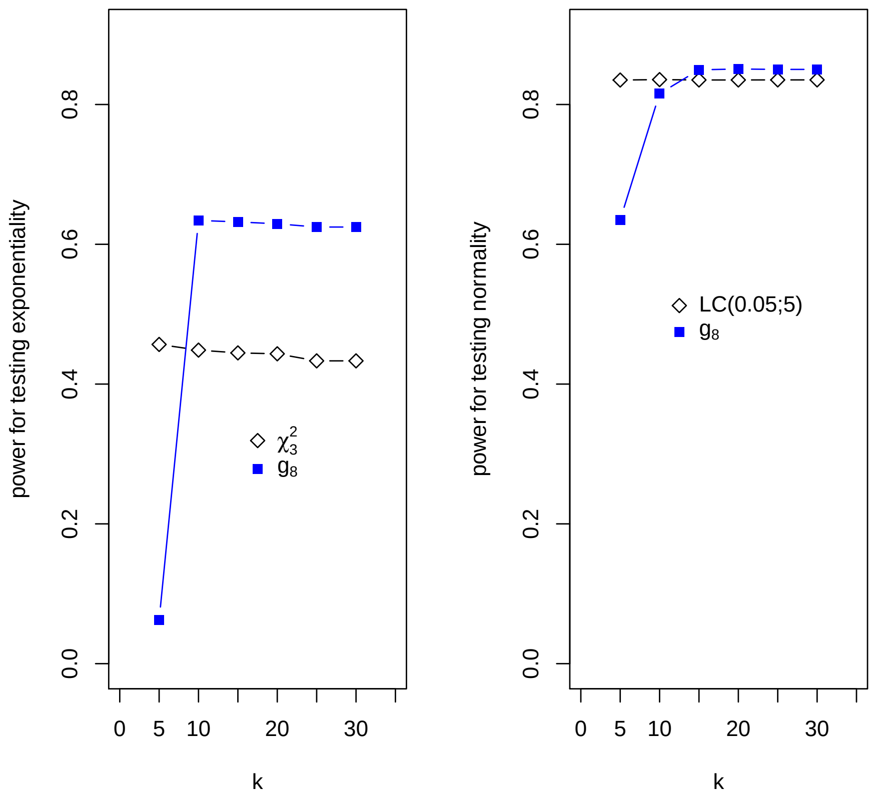


Figure 12.

Comparison of simulated power under low frequency alternatives with that under high frequency alternatives as a function of $k$. $n = 50$, $\alpha = 0.05$, 10,000 Monte Carlo runs.

Based on the previous results in this subsection, we still recommend $\boldsymbol{k = 20}$ in the composite hypothesis case to guarantee that the proposed test will be powerful against high frequency alternatives and perform comparably to some other popular omnibus tests at low frequency alternatives.

### 4.2.2. Prior Probabilities $\pi_j$

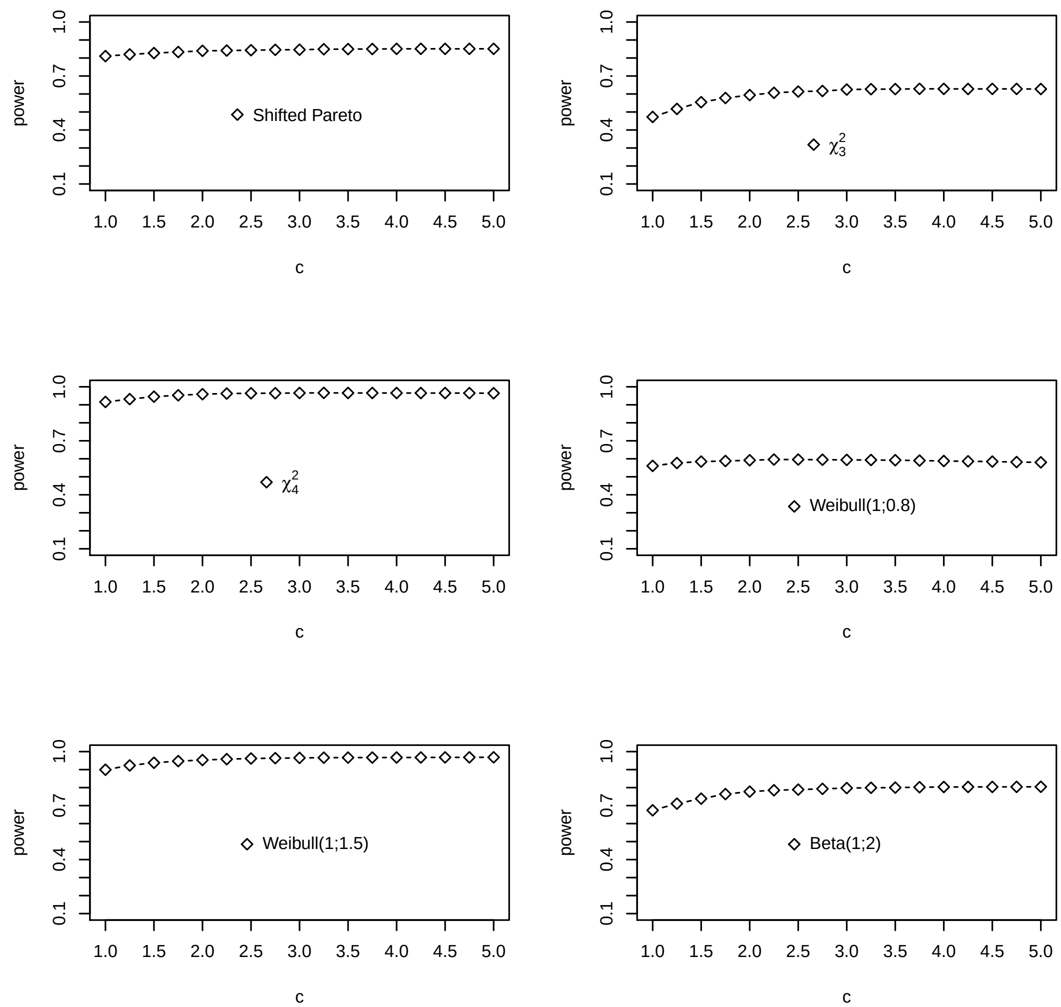


Figure 13.

The behavior of simulated power of proposed test statistic $S_k$ as a function of $c$ when testing exponentiality under the alternatives Shifted Pareto, Chi-square, Weibull and Beta. $n = 50$, $\alpha = 0.05$, 10,000 Monte Carlo runs.

The results in subsection 4.2.1 indicate that the number of Fourier components $k$ does not play a crucial role in the proposed test, since the power does not vary too much in a certain range of $k$ (i.e. 10-35). But the choice of prior probabilities $\pi_j$s may be significant. We now discuss this choice for composite cases. As argued in subsection 4.1.2, $\pi_j = 1/j^c$ for $c > 1$ is considered first and our goal turns to a good choice of $c$ at this step. We take $k = 20$ as recommended in last subsection.

Figures 13 and 14 show the performance of $S_k$ when testing exponentiality. The graphs are arranged from the lowest to highest frequency, as measured by Fourier coefficients, (see (4.5)). We notice that the power increases to a certain level and then stays almost flat as $c$ increases for low frequency alternatives, e.g. Shifted Pareto, $\chi_3^2$, $\chi_4^2$, Weibull(1;1.5), Beta(1;2) and LN(0;0.8). For the Weibull(1;0.8), $c$ around 2.2 yields the highest power. Thus, we may conclude $c = 2$ and above works well for low frequency alternatives. However, at highly oscillating alternatives, the smaller the $c$ is, the higher the power achieved. For alternatives $g_6$, $g_7$, $g_8$ and $g_9$, the powers decrease as $c$ increases when testing exponentiality. As a result, we may say smaller $c$ (below 2) performs well at high frequency alternatives.

The last paragraph agrees with our discussion in the simple null hypothesis. That small $c$ emphasizes higher frequency alternatives and large $c$ down-weights high frequency alternatives is also valid when the hypothesis is composite. The last graph in Figure 14 illustrates average power against the various alternatives, which still peaks at around $c = 2$. Therefore, $\pi_j = 1/j^2$ will be considered as a good choice of prior probabilities from now on.

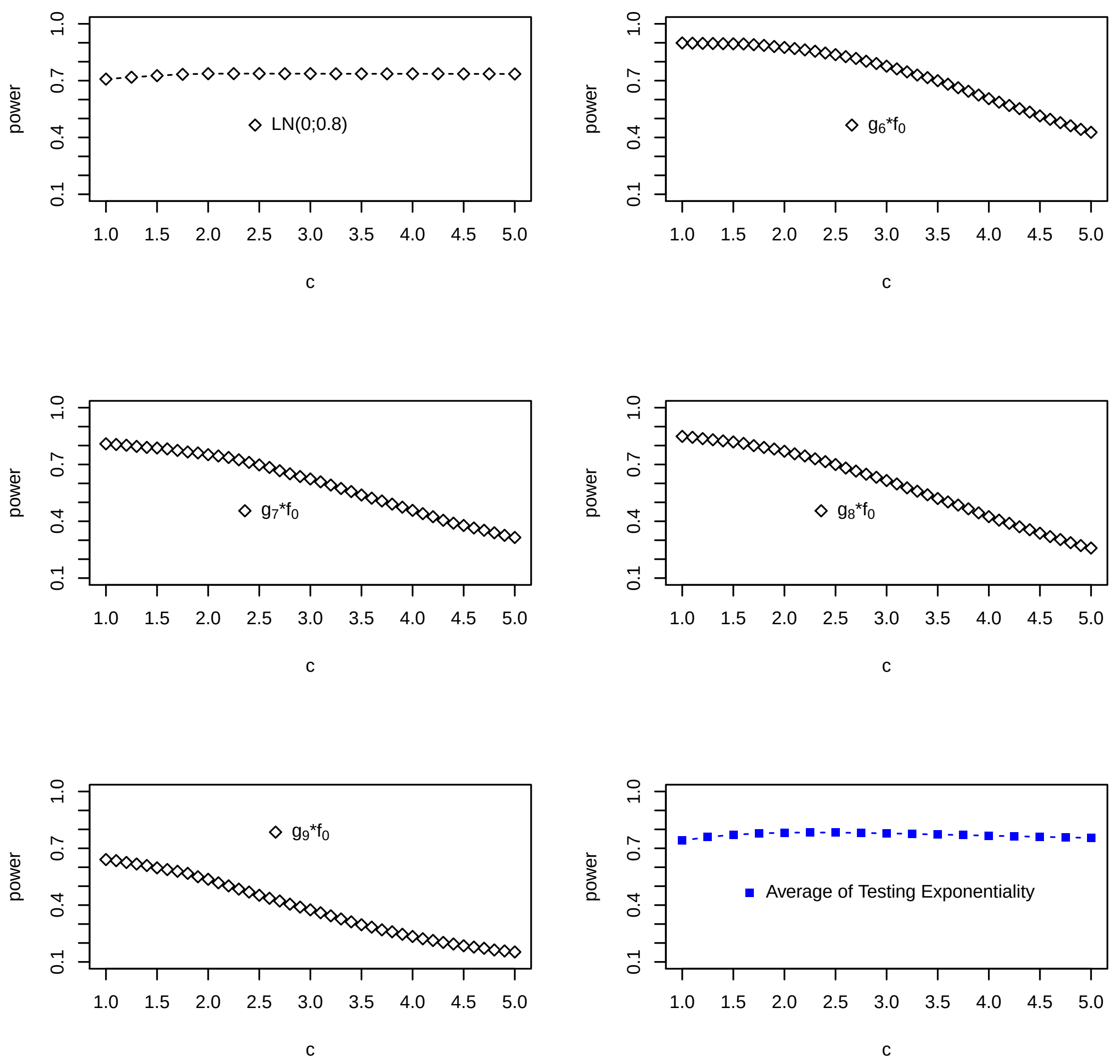


Figure 14.

The behavior of simulated power of proposed test statistic $S_k$ as a function of $c$ when testing exponentiality under the alternatives Lognormal, $g_6$, $g_7$, $g_8$, $g_9$, and the average. $n = 50$, $\alpha = 0.05$, 10,000 Monte Carlo runs.

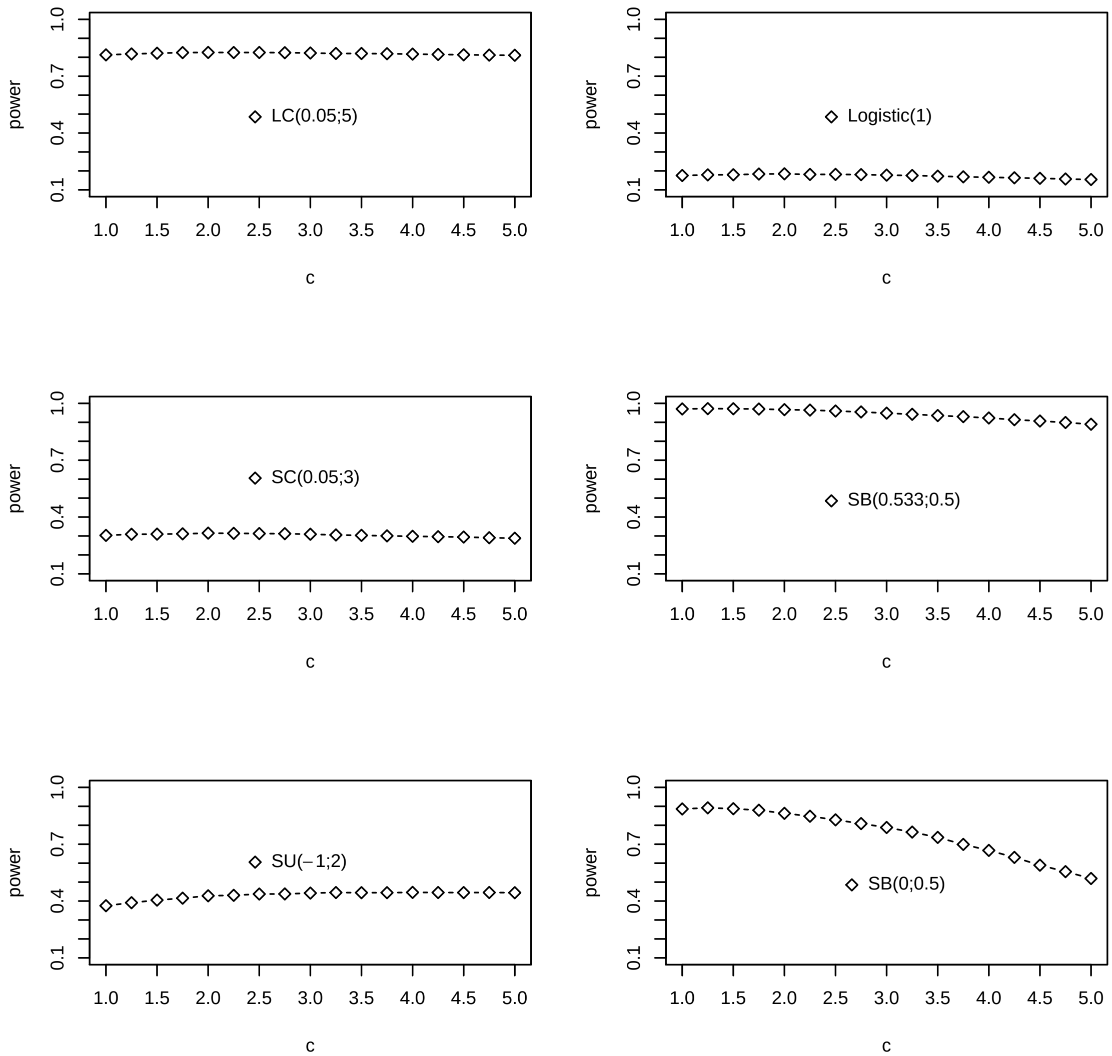


Figure 15.

The behavior of simulated power of proposed test statistic $S_k$ as a function of $c$ when testing normality under the alternatives LC, Logistic, SC, SB and SU. $n = 50$, $\alpha = 0.05$, 10,000 Monte Carlo runs.

Similar conclusions hold for testing normality. For illustration see Figures 15 and

16, where the graphs are also placed in order of increasing frequency.

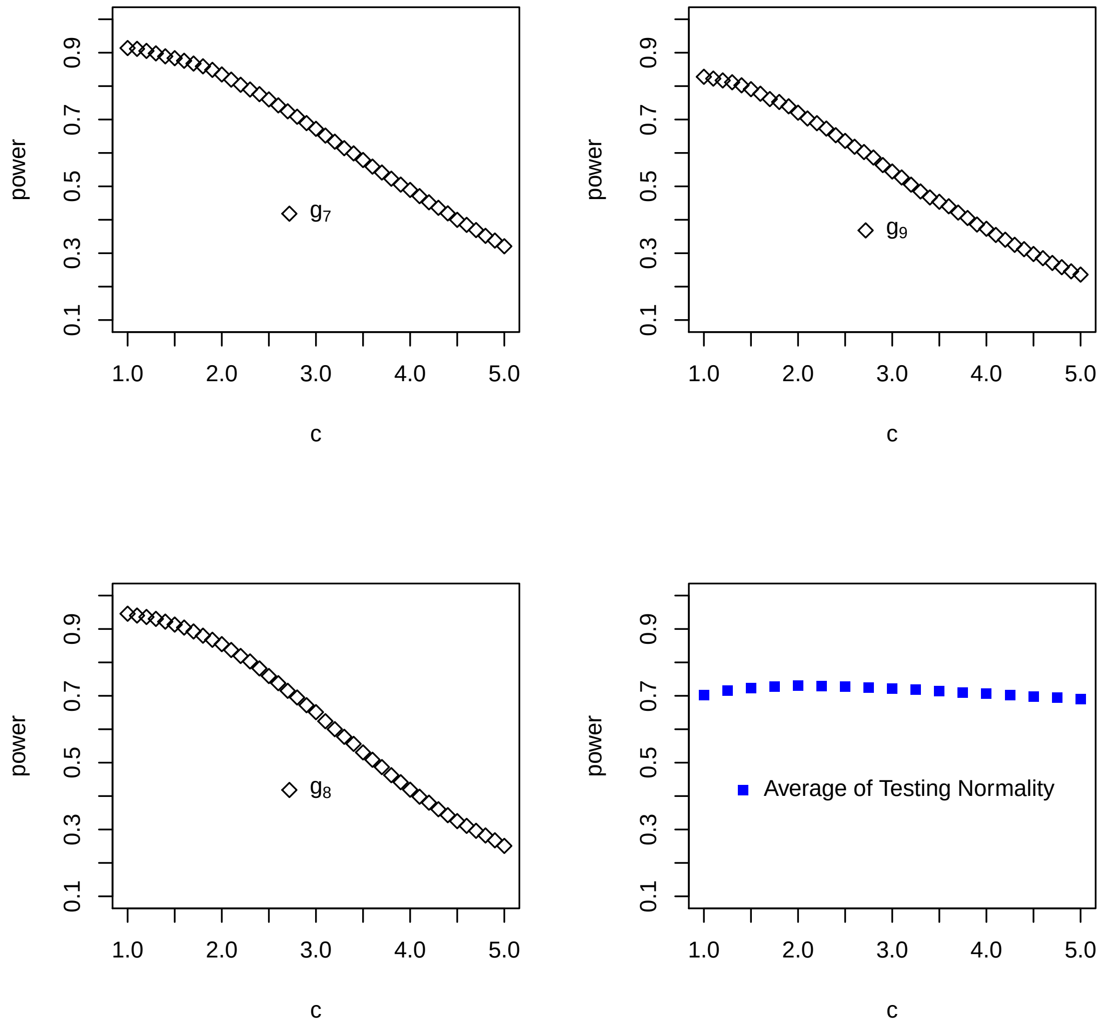


Figure 16.

The behavior of simulated power of proposed test statistic $S_k$ as a function of $c$ when testing normality under the alternatives $g_7$, $g_8$, $g_9$, and the average. $n = 50$, $\alpha = 0.05$, 10,000 Monte Carlo runs.

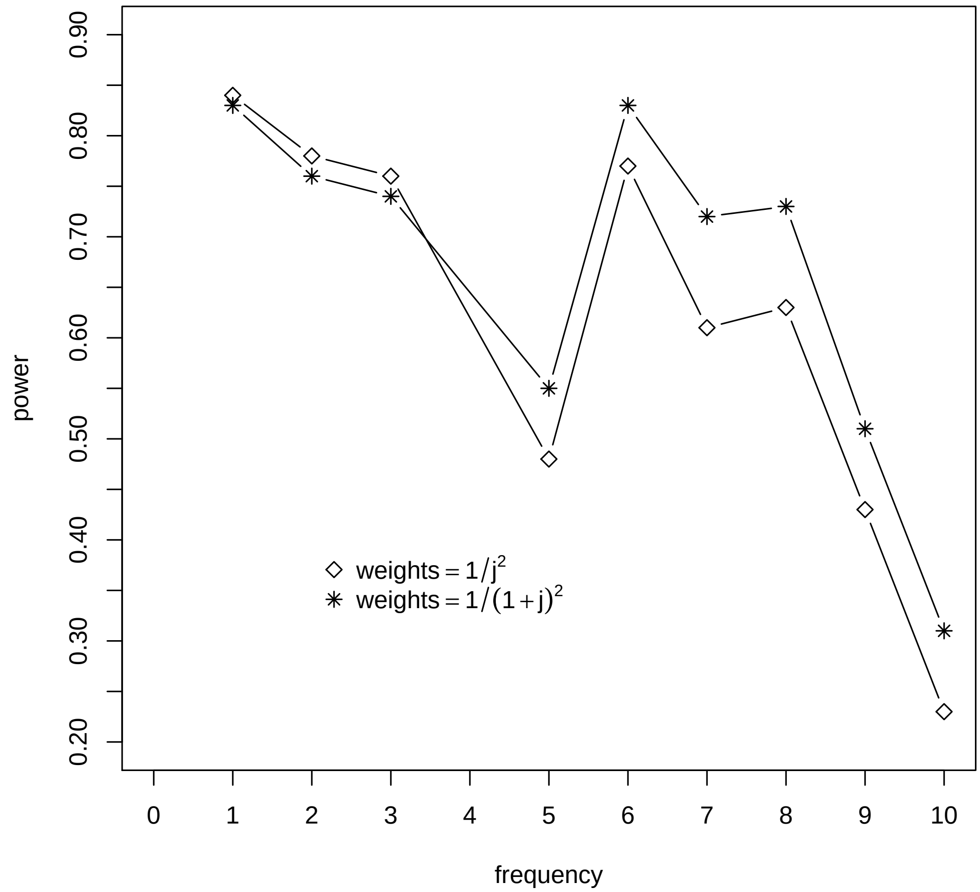


Figure 17.

The behavior of simulated average power of proposed test $S_k$ according to the different weights when testing a composite null hypothesis. $n = 50$, $\alpha = 0.05$, 10,000 Monte Carlo runs.

When testing a simple null hypothesis, we proposed $\pi_j = 1/(1+j)^2$ as an "optimal" prior. Here we compare the performances of $\pi_j = 1/(1+j)^2$ and $\pi_j = 1/j^2$ in testing for composite hypotheses. Figure 17 presents the differences in power between these two versions and the power presented is the average against the various alternatives. Similar to the simple hypotheses cases, the power under high frequency alternatives is improved by taking $\pi_j = 1/(1+j)^2$, but does not hurt too much under lower frequency alternatives.

Since the goal of this dissertation is to propose tests that are powerful against both low and high frequency alternatives, we recommend $\boldsymbol{\pi_j = 1/(1+j)^2}$ as prior probabilities for the Fourier coefficients and will use them in the next subsection to compare with other omnibus tests.

### 4.2.3. Power Comparisons in the Simulation Study

In the last two subsections, we determined good choices for $k$ and the prior probabilities $\pi_1, \pi_2, \ldots$. Now we will start power comparisons with other omnibus tests. We use $k = 20$ and $\pi_j = 1/(1+j)^2$ as recommended. The simulated critical values for composite hypotheses are presented in Table 4.

Table 4.

Approximate critical values of proposed test based on 10,000 Monte Carlo runs.

| Null | $n$ | $\alpha$ | Critical Value of $S_{20}$ |
|---|---|---|---|
| Exponentiality | 50 | 0.05 | 4.240 |
| | | 0.10 | 2.599 |
| | 100 | 0.10 | 2.671 |
| Normality | 50 | 0.05 | 3.794 |
| | 100 | 0.05 | 4.187 |

To see how well the proposed tests perform we show the result of an extensive Monte Carlo study of the power. The null hypothesis of exponentiality corresponds to $H_0$: $f \in \{f(x;\beta) : \beta > 0\}$, where $f(x;\beta)$ is defined as

$$f(x;\beta) = \beta^{-1} \exp(-\beta^{-1}x), \quad x \geq 0,$$

and the null MLE of $\beta$ is $\hat{\beta} = \dfrac{1}{n} \sum_{i=1}^{n} X_i$. For power comparison when testing exponentiality we consider the Gini statistic $G$ introduced by Gail and Gastwirth (1978) and $W_{S1}$ and $W_{S2}$ proposed by Kallenberg and Ledwina (1997) for composite hypotheses. Gini's test is "powerful against a variety of alternatives" [cf. Gail and Gastwirth (1978)] and turned out to perform well in the study of Ascher (1990). It was also used for comparative purposes by Rayner and Best (1989) and LaRiccia (1991). $W_{S1}$ and $W_{S2}$ have been introduced in Chapter I. The alternatives considered for simulations are shown at the beginning of subsection 4.2.

The following tables present the power for testing exponentiality. Note that several alternatives are used in more than one paper. In the cited papers one may find simulated power for other tests for these alternatives. Many authors show simulation results for $n = 20$. In our opinion this is an extreme situation when testing goodness-of-fit, so we present the more realistic choices $n = 50$ and $n = 100$ in Table 5.

Although motivated by general ideas, the proposed test based on $S_{20}$ can compare even with 'special' tests for exponentiality, like Gini's test, under the above alternatives. As a non-adaptive test, $S_{20}$ also performs comparably to the popular adaptive tests $W_{S1}$ and $W_{S2}$ on average when alternatives are these classical densities. As is seen in Table 5, for $n = 50$ the proposed test based on $S_{20}$, $W_{S1}$ and $W_{S2}$ often have higher power than Gini's test $G$ with great differences in LN(0;1), Shifted exp.(0.2;1), Shifted exp.(0.2;0.7) and Pareto(1;0.2). The proposed test improves considerably from $n = 50$ to $n = 100$.

Table 5.

Power of Gini's test, Ledwina's tests based on $W_{S1}$ and $W_{S2}$ and one based on $S_{20}$ when testing exponentiality under the alternatives given in Agnus (1982).

| Alternatives | Power(%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | $G$ | $W_{S1}$ | | $W_{S2}$ | | $S_{20}$ | |
| | $n = 50$ | $n = 50$ | $n = 100$ | $n = 50$ | $n = 100$ | $n = 50$ | $n = 100$ |
| $\chi_1^2$ | 93 | 96 | 100 | 97 | 100 | 97 | 100 |
| $\chi_3^2$ | 58 | 51 | 88 | 60 | 84 | 58 | 87 |
| $\chi_4^2$ | 95 | 93 | 100 | 96 | 100 | 96 | 100 |
| LN(0;0.8) | 74 | 76 | 95 | 74 | 94 | 75 | 94 |
| LN(0;1) | 22 | 62 | 86 | 46 | 71 | 42 | 75 |
| LN(0;1.2) | 46 | 81 | 99 | 83 | 99 | 79 | 99 |
| Weibull(1;0.8) | 59 | 56 | 82 | 60 | 85 | 60 | 85 |
| Weibull(1;1.2) | 43 | 34 | 64 | 42 | 69 | 41 | 66 |
| Weibull(1;1.5) | 97 | 93 | 100 | 96 | 100 | 96 | 100 |
| Beta(1;2) | 81 | 71 | 97 | 76 | 98 | 77 | 98 |
| Uniform(0;2) | 100 | 99 | 100 | 99 | 100 | 100 | 100 |
| Shifted exp.(0.2;1) | 68 | 83 | 100 | 90 | 100 | 90 | 100 |
| Shifted exp.(0.2;0.7) | 45 | 58 | 89 | 65 | 93 | 61 | 95 |
| Pareto(1;0.2) | 74 | 100 | 100 | 100 | 100 | 100 | 100 |
| Pareto(0.8;0.01) | 94 | 100 | 100 | 100 | 100 | 100 | 100 |
| Shifted Pareto | 86 | 84 | 98 | 84 | 98 | 84 | 98 |
| Average | 71 | 77 | 94 | 79 | 93 | 79 | 94 |

$\alpha$=0.1, 10,000 MC runs.

Since our goal in this dissertation is to develop a test which can be comparable to other popular omnibus tests at low frequencies and perform exceptionally well at high frequencies, we will compare the behavior of the considered tests under low frequency alternatives and high frequency alternatives in Tables 6 and 7 separately. For the sake of comparison, a modified data driven smooth test statistic $W_T$ developed from $N_T$ is introduced. This statistic uses the penalty for high frequency alternatives when testing a composite null, where the penalty was defined in Chapter I. In fact, Inglot and Ledwina (2006) restricted attention to testing uniformity. We combine the new selection rule with the data driven smooth test for composite hypotheses, $W_{S2}$, as the test statistic $W_T$.

Table 6.

Power of Ledwina's tests based on $W_{S2}$ and $W_T$ and proposed test statistic based on $S_{20}$ when testing exponentiality under low frequency alternatives.

| Alternatives | The five largest (in absolute value) Fourier coefficients × 1000 | | | | | Power(%) | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | $W_{S2}$ | $W_T$ | $S_{20}$ |
| LN(0;0.8) | [**1**]244 | [**3**]228 | [**5**]192 | [**7**]133 | [**9**]102 | 63 | 61 | 67 |
| Shifted Pareto | [**1**]334 | [**2**]210 | [**4**]99 | [**5**]93 | [**6**]92 | 76 | 73 | 79 |
| $\chi^2_3$ | [**2**]235 | [**1**]122 | [**4**]52 | [**3**]24 | [**6**]21 | 43 | 34 | 44 |
| $\chi^2_4$ | [**2**]373 | [**1**]192 | [**3**]88 | [**4**]33 | [**5**]16 | 90 | 83 | 90 |
| Weibull(1;0.8) | [**2**]224 | [**1**]148 | [**3**]115 | [**4**]77 | [**5**]56 | 50 | 47 | 51 |
| Weibull(1;1.5) | [**2**]340 | [**1**]199 | [**3**]151 | [**4**]73 | [**5**]13 | 87 | 81 | 91 |
| Beta(1;2) | [**3**]210 | [**4**]174 | [**1**]154 | [**2**]147 | [**5**]106 | 53 | 46 | 62 |
| Uniform(0;2) | [**3**]398 | [**2**]272 | [**4**]262 | [**1**]234 | [**7**]214 | 97 | 95 | 99 |

$n = 50$, $\alpha$=0.1, 10,000 MC runs.

For each case twenty Fourier coefficients (in Legendre basis) of the underlying distributions are calculated in order to illustrate some insight into the structure and magnitude of the alternatives. The five largest (from these 20) Fourier coefficients are presented. Each bold face number $j$ corresponds to $j$ in basis function $\phi_j$. The power of proposed test based on $S_{20}$ is comparable to $W_{S2}$ but higher than that of $W_T$ at low frequencies in Table 6. At high frequency alternatives in Table 7, $S_{20}$ outperforms $W_{S2}$ and works comparably to $W_T$.

Table 7.

Power of Ledwina's tests based on $W_{S2}$ and $W_T$ and proposed test statistic based one $S_{20}$ when testing exponentiality under high frequency alternatives $g_k(F_0(x);\boldsymbol{\theta}) \cdot f_0(x)$.

| Parameters | | The five largest (in absolute value) | | | | | Power(%) | | |
|---|---|---|---|---|---|---|---|---|---|
| $k$ | $\theta$ | Fourier coefficients × 1000 | | | | | $W_{S2}$ | $W_T$ | $S_{20}$ |
| 5 | (0,0,0,0,0.5) | [**5**]435 | [**6**]118 | [**4**]115 | [**1**]85 | [**10**]55 | 37 | 52 | 55 |
| 6 | (0,0,0,0,0.2, -0.7) | [**6**]529 | [**5**]273 | [**2**]123 | [**11**]108 | [**7**]107 | 62 | 80 | 83 |
| 7 | (0,0,0,0,-0.1, -0.2,0.6) | [**7**]525 | [**6**]198 | [**5**]91 | [**2**]57 | [**8**]56 | 38 | 74 | 72 |
| 8 | (0,0,0,0,0,0,0, -0.7) | [**8**]596 | [**2**]93 | [**4**]54 | [**9**]48 | [**12**]46 | 31 | 78 | 72 |
| 9 | (0,0,0,0,0,0,0, 0,0.6) | [**9**]417 | [**10**]279 | [**8**]225 | [**7**]114 | [**6**]102 | 15 | 56 | 51 |
| 10 | (0,0,0,0,0,0,0, 0,0,-0.5) | [**10**]454 | [**2**]53 | [**4**]36 | [**11**]31 | [**6**]23 | 8 | 26 | 31 |

$n = 50$, $\alpha$=0.05, 10,000 MC runs.

Next we consider the null hypothesis of normality, corresponding to $H_0$: $f \in \{f(x;\mu,\sigma) : \mu \in \mathbb{R}, \sigma \in \mathbb{R}\}$, where $f(x;\mu,\sigma)$ is defined as

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{ -\frac{1}{2}(x-\mu)^2/\sigma^2 \right\}$$

and the null MLE is $(\hat{\mu},\hat{\sigma}) = \left(\bar{X}, \left\{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right\}^{\frac{1}{2}}\right)$ with $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$. Here we consider the data driven smooth tests $W_{S1}$ and $W_{S2}$ mentioned before and the often recommended Shaprio-Wilk test, $SW$. According to Bowman (1992) $SW$ sets a high standard as an omnibus test of normality. The details about alternatives used for testing normality are presented at the beginning of subsection 4.2.

Table 8 presents the results for a variety of symmetric and skew alternatives for $n = 50$ and 100. As in Pearson *et al.* (1977) and Kallenberg and Ledwina (1997b) both symmetric and skew alternatives in Table 8 are ordered according to increasing kurtosis.

It turns out that except for the first 3 symmetric cases, which are close to the null hypothesis, the proposed test statistic based on $S_{20}$ performs comparably to $SW$, the 'special' test for normality. Comparing with $W_{S1}$ and $W_{S2}$, $S_{20}$ dominates the former in skewed cases and the latter in symmetric cases. But $W_{S1}$ and $W_{S2}$ work slightly better than $S_{20}$ in symmetric alternatives and skew alternatives respectively. On average, $S_{20}$ is more powerful than $W_{S1}$ and $W_{S2}$ and close to $SW$. We could conclude that the proposed test statistic based on $S_{20}$ outperforms $W_{S1}$ and $W_{S2}$ and is comparable to $SW$, the high standard omnibus test of normality.

Table 8.

Power of Shapiro-Wilk test, Ledwina's tests based on $W_{S1}$ and $W_{S2}$ and one based on $S_{20}$ when testing normality under the alternatives given in Pearson *et al.* (1977).

| Alternatives | Power(%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | $SW$ | $W_{S1}$ | | $W_{S2}$ | | $S_{20}$ | |
| | $n=50$ | $n=50$ | $n=100$ | $n=50$ | $n=100$ | $n=50$ | $n=100$ |
| Symmetric alternatives | | | | | | | |
| SB(0;0.5) | 99 | 93 | 100 | 55 | 92 | 88 | 100 |
| TU(1.5) | 92 | 74 | 99 | 26 | 61 | 65 | 98 |
| TU(0.7) | 62 | 45 | 88 | 9 | 19 | 33 | 78 |
| Logistic(1) | 13 | 21 | 35 | 12 | 13 | 18 | 26 |
| TU(10) | 99 | 100 | 100 | 99 | 100 | 100 | 100 |
| SC(0.05;3) | 31 | 38 | 57 | 24 | 32 | 32 | 47 |
| SC(0.2;5) | 95 | 98 | 100 | 92 | 99 | 98 | 100 |
| SC(0.05;5) | 62 | 66 | 87 | 55 | 74 | 62 | 83 |
| SC(0.05;7) | 74 | 77 | 94 | 70 | 88 | 74 | 92 |
| SU(0;1) | 68 | 81 | 96 | 61 | 83 | 77 | 96 |
| | | | | | | | |
| Skew alternatives | | | | | | | |
| SB(0.533;0.5) | 100 | 95 | 100 | 92 | 100 | 97 | 100 |
| SB(1;1) | 81 | 57 | 95 | 71 | 96 | 71 | 97 |
| LC(0.2;3) | 60 | 52 | 90 | 69 | 95 | 69 | 95 |
| Weibull(2) | 41 | 29 | 64 | 41 | 74 | 40 | 71 |
| LC(0.1;3) | 50 | 51 | 83 | 58 | 86 | 59 | 86 |
| $\chi^2_{10}$ | 57 | 48 | 85 | 62 | 91 | 61 | 89 |
| LC(0.05;3) | 32 | 37 | 58 | 34 | 54 | 37 | 57 |
| LC(0.1;5) | 98 | 98 | 100 | 97 | 100 | 98 | 100 |
| SU(-1;2) | 37 | 40 | 67 | 42 | 68 | 44 | 69 |
| $\chi^2_4$ | 95 | 86 | 100 | 93 | 100 | 93 | 100 |
| LC(0.05;5) | 85 | 87 | 97 | 78 | 95 | 84 | 97 |
| LC(0.05;7) | 92 | 92 | 99 | 90 | 98 | 92 | 99 |
| SU(1;1) | 96 | 97 | 100 | 98 | 100 | 98 | 100 |
| LN(0;1) | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Average | 72 | 69 | 87 | 64 | 78 | 71 | 87 |

$\alpha$=0.05, 10,000 Monte Carlo runs.

Table 9.

Power of Ledwina's tests based on $W_{S2}$ and $W_T$ and proposed test statistic based on $S_{20}$ when testing normality under low frequency alternatives.

| Alternatives | The five largest (in absolute value) Fourier coefficients × 1000 | | | | | Power(%) $W_{S2}$ | $W_T$ | $S_{20}$ |
|---|---|---|---|---|---|---|---|---|
| LC(0.05;5) | [**2**]277 | [**3**]221 | [**12**]178 | [**11**]177 | [**10**]174 | 78 | 78 | 84 |
| Logistic(1) | [**2**]83 | [**4**]66 | [**6**]51 | [**8**]47 | [**10**]43 | 12 | 12 | 18 |
| SC(0.05;3) | [**2**]135 | [**12**]58 | [**10**]55 | [**8**]50 | [**6**]39 | 24 | 23 | 32 |
| SU(0;1) | [**2**]363 | [**4**]272 | [**8**]116 | [**12**]93 | [**10**]79 | 61 | 62 | 77 |
| SB(0.533;0.5) | [**3**]458 | [**4**]309 | [**12**]283 | [**7**]273 | [**9**]272 | 92 | 93 | 97 |
| SU(-1;2) | [**3**]178 | [**2**]102 | [**5**]70 | [**6**]60 | [**7**]57 | 42 | 40 | 44 |

$n = 50$, $\alpha$=0.05, 10,000 MC runs.

As done for testing exponentiality, we would also like to compare the power of the considered tests under low frequencies and high frequencies in Tables 9 and 10. The considered tests $W_{S2}$ and $W_T$ are defined as before. Twenty Fourier coefficients (in Legendre basis) of the underlying distributions are calculated for each case in order to present some insight into the structure and magnitude of the alternatives. The five largest (from these 20) Fourier coefficients are shown. Each bold face number $j$ is corresponding to $j$ in basis function $\phi_j$. The power of proposed test based on $S_{20}$ dominates $W_{S2}$ and $W_T$ at both low frequencies in Table 9 and high frequencies in Table 10 for testing normality.

Table 10.

Power of Ledwina's tests based on $W_{S2}$ and $W_T$ and proposed test statistic based on $S_{20}$ when testing normality under high frequency alternatives $g_k(x;\boldsymbol{\theta})$.

| Parameters | | The five largest (in absolute value) | | | | | Power(%) | | |
|---|---|---|---|---|---|---|---|---|---|
| $k$ | $\theta$ | Fourier coefficients $\times$ 1000 | | | | | $W_{S2}$ | $W_T$ | $S_{20}$ |
| 4 | (0,-0.5,0,-0.2) | [**4**]191 | [**6**]179 | [**2**]140 | [**8**]102 | [**10**]38 | 8 | 11 | 27 |
| 5 | (0,0,0,0,0.5) | [**6**]333 | [**5**]304 | [**4**]243 | [**9**]233 | [**8**]195 | 51 | 64 | 80 |
| 6 | (0.1,0,0,0.1, 0.2,0.2) | [**6**]269 | [**8**]233 | [**4**]180 | [**12**]178 | [**2**]157 | 19 | 33 | 54 |
| 7 | (0,0,0,0,-0.1, -0.2,0.6) | [**6**]523 | [**7**]268 | [**10**]235 | [**11**]191 | [**5**]177 | 67 | 82 | 90 |
| 8 | (0,0,0,0,0,0,0, -0.7) | [**12**]501 | [**8**]490 | [**4**]313 | [**2**]194 | [**6**]118 | 48 | 82 | 92 |
| 9 | (0,0,0,0,0,0,0, 0,0.6) | [**6**]346 | [**9**]273 | [**4**]261 | [**7**]205 | [**2**]190 | 32 | 55 | 78 |
| 10 | (0,0,0,0,0,0,0, 0,0,-0.3) | [**6**]354 | [**4**]240 | [**2**]181 | [**12**]91 | [**10**]83 | 23 | 39 | 65 |
| SB(0;0.5) | | [**6**]393 | [**4**]362 | [**12**]245 | [**2**]225 | [**10**]198 | 55 | 65 | 88 |

$n = 50$, $\alpha$=0.05, 10,000 MC runs.

### 4.3. Further Discussion about Frequency

We have repeatedly used the terms low and high frequency, and so now we would like to clarify what we mean by these terms. The definition of high frequency is a little subjective. For any given alternative, if the first $m$ Fourier coefficients corresponding to basis functions $\phi_1$,...,$\phi_m$ are quite small, and the Fourier coefficients corresponding to $\phi_{m+1}$, $\phi_{m+2}$,... are larger and $m \geq 3$, we say that the density is a high frequency alternative. One thing we wish to point out is that not all oscillatory densities are high frequency and vice versa.

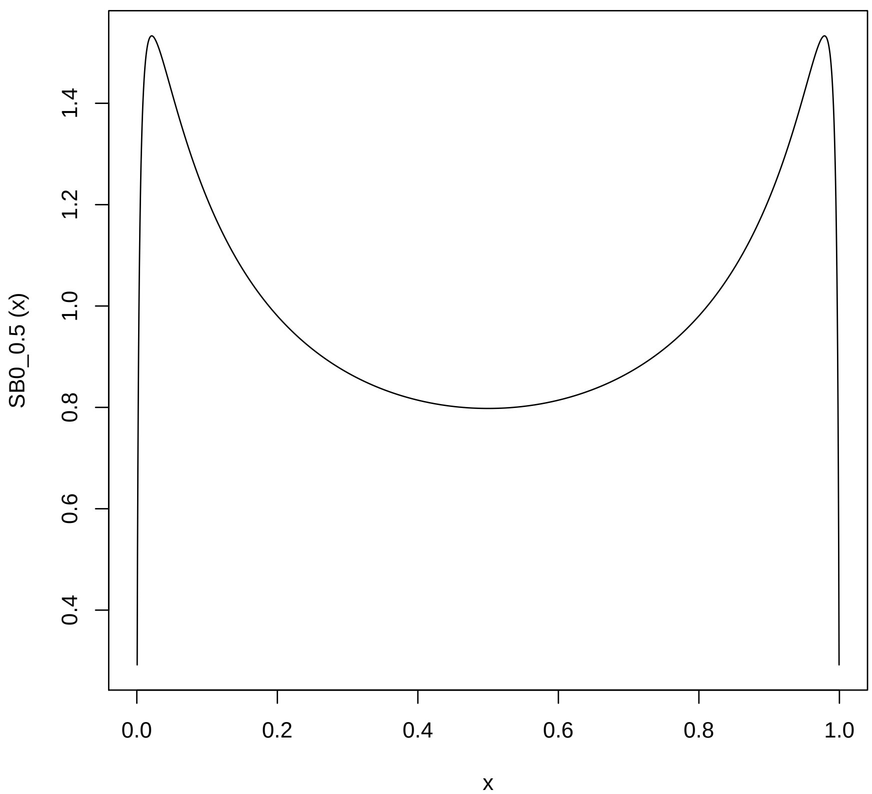


Figure 18.

The density of SB(0;0.5).

Figure 18 presents the density of SB(0:0.5), which is U-shaped. However, the largest Fourier coefficient of SB(0;0.5) corresponds to basis function $\phi_6$. In other words, SB(0;0.5) is a high frequency density even though it has only two peaks.

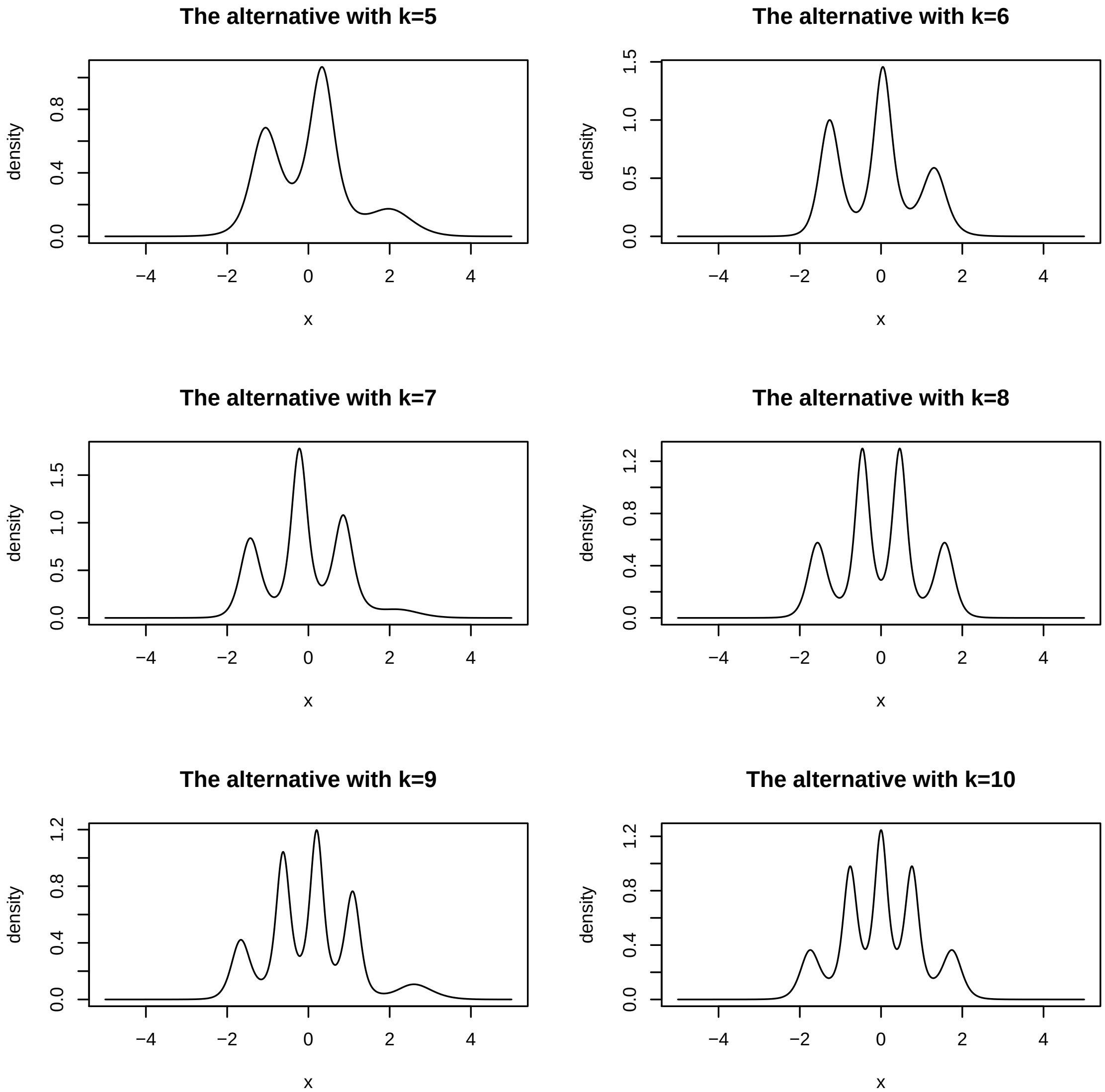


Figure 19.

The densities of alternatives used in Table 11.

The six graphs shown in Figure 19 are all highly oscillatory. However, a further investigation about them shown in Table 11 indicates that they are “low frequency” densities. The performances of $S_k$, $W_{S2}$ and $W_T$ are close to each other and quite good.

Table 11.

Power of Ledwina's tests based on $W_{S2}$ and $W_T$ and one based on $S_{20}$ when testing normality under the alternatives $g_k(x;\boldsymbol{\theta}) \cdot f_0(x)$.

| Parameters | | The five largest (in absolute value) | | | | | Power(%) | | |
|---|---|---|---|---|---|---|---|---|---|
| $k$ | $\theta$ | Fourier coefficients $\times$ 1000 | | | | | $W_{S2}$ | $W_T$ | $S_{20}$ |
| 5 | (0,0,0,0,0.5) | [**2**]658 | [**5**]647 | [**7**]291 | [**3**]224 | [**1**]183 | 100 | 100 | 100 |
| 6 | (0,0,0,0,0.2, -0.7) | [**2**]960 | [**6**]668 | [**8**]379 | [**4**]375 | [**7**]208 | 95 | 97 | 99 |
| 7 | (0,0,0,0,-0.1, -0.2,0.6) | [**2**]1115 | [**7**]1007 | [**5**]684 | [**9**]492 | [**6**]256 | 88 | 89 | 91 |
| 8 | (0,0,0,0,0,0,0, -0.7) | [**2**]940 | [**8**]741 | [**10**]427 | [**6**]422 | [**4**]186 | 99 | 99 | 100 |
| 9 | (0,0,0,0,0,0,0, 0,0.6) | [**2**]789 | [**9**]742 | [**11**]378 | [**7**]377 | [**1**]157 | 100 | 100 | 100 |
| 10 | (0,0,0,0,0,0,0, 0,0,-0.5) | [**2**]772 | [**10**]711 | [**12**]326 | [**8**]325 | [**4**]38 | 74 | 76 | 81 |

$n$ = 50, $\alpha$=0.05, 10,000 Monte Carlo runs.

## 4.4. Real Data Analysis

We end with an application of our methodology to a real problem. The data are dust concentrations taken from a manufacturing plant in Munich, Germany. We will analyze the natural log of the variable of interest. The sample size is $n = 1246$.

As shown in Figure 20, a kernel density estimate based on the S-J plug-in bandwidth=0.03438 [cf. Sheather and Jones (1991)] shows several modes, whereas a kernel density estimate with a "normal reference" bandwidth=0.1636 [cf. Silverman

(1986)] yields three modes. Therefore, we consider testing the null hypothesis that the log(data) come from a mixture of three normal distributions. This is an interesting example of testing a smooth density against a possibly high frequency alternative.

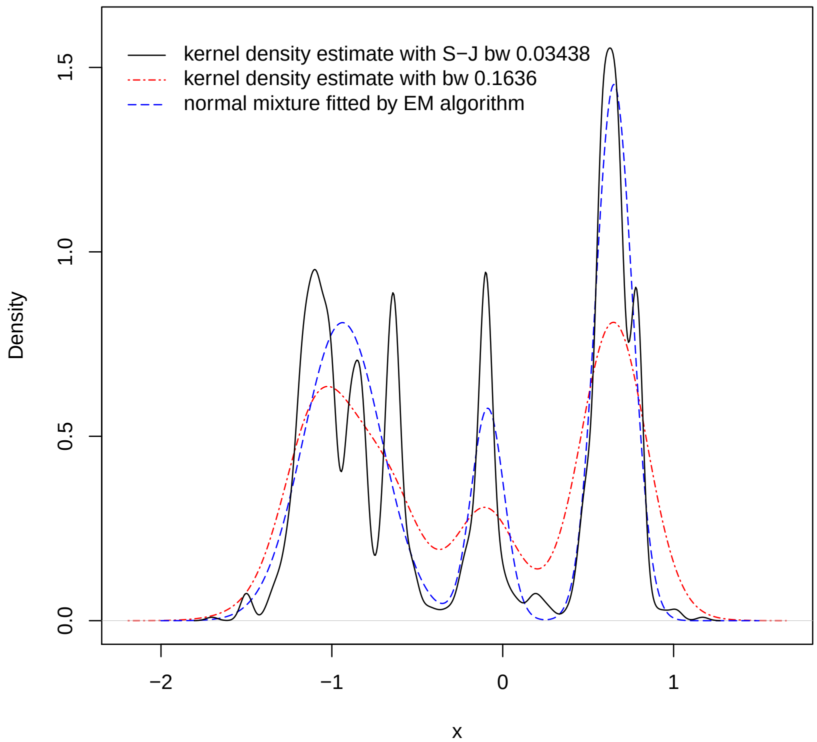


Figure 20.

The density estimates for log(dust concentration), computed by three methods.

The mixture of normals null hypothesis is given as:

$$H_0 : f(x) = \frac{p_1}{\sigma_1}\phi\left(\frac{x-\mu_1}{\sigma_1}\right) + \frac{p_2}{\sigma_2}\phi\left(\frac{x-\mu_2}{\sigma_2}\right) + \frac{1-p_1-p_2}{\sigma_3}\phi\left(\frac{x-\mu_3}{\sigma_3}\right), \qquad (4.7)$$

where $\phi(x)$ is the probability density function of the standard normal distribution, unknown parameters $p_i$, $\mu_i$ and $\sigma_i$, $i = 1, 2, 3$, are weights, means and standard deviations of three normal distributions, respectively. This model is estimated by the expectation-maximization (EM) algorithm, an iterative method which alternates between performing an expectation (E) step that computes the expectation of the log-likelihood evaluated using the current estimate for the latent variables, and a maximization (M) step that computes parameters maximizing the expected log-likelihood found on the E step. We then obtain: $\hat{p}_1 = 0.4693173$, $\hat{p}_2 = 0.1391294$, $\hat{p}_3 = 0.3915533$, $\hat{\mu}_1 = -0.93759518$, $\hat{\mu}_2 = -0.08797447$, $\hat{\mu}_3 = 0.65096132$, $\hat{\sigma}_1 = 0.23167941$, $\hat{\sigma}_2 = 0.09651196$ and $\hat{\sigma}_3 = 0.10743332$. This density is presented in Figure 20.

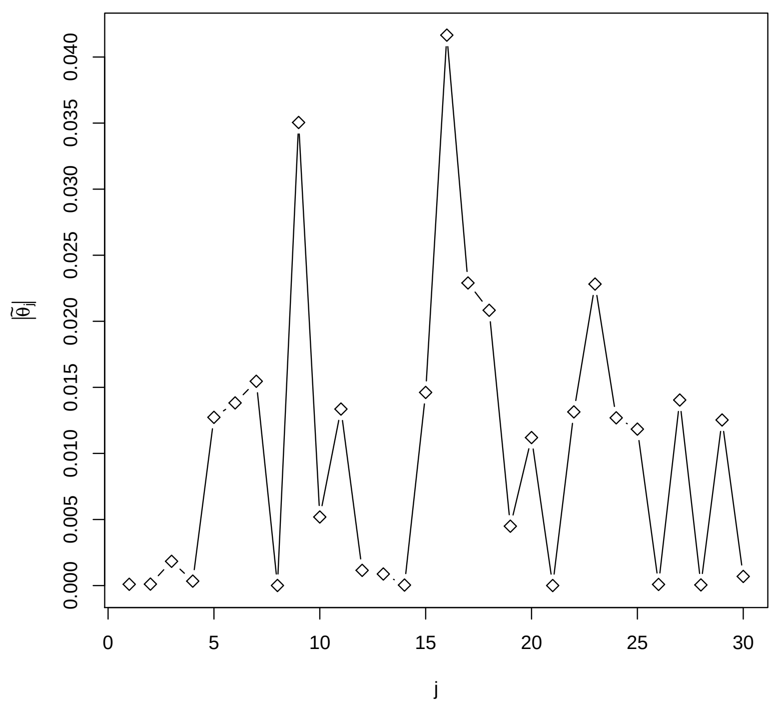


Figure 21.

The behavior of estimated Fourier coefficients, $\tilde{\theta}_j$, as a function of $j$.

Since the null hypothesis is composite, we will use (2.13) as the form for $\tilde{\theta}_j$, based on the arguments in Chapter II. In order to calculate the proposed test statistic $S_k$, we need $\boldsymbol{I}_{\boldsymbol{\beta}j}$ and $\boldsymbol{I}_{\boldsymbol{\beta\beta}}$. Unfortunately, these matrices do not have a closed form in this case, and so we use the observed information matrices $\boldsymbol{J}_{\boldsymbol{\beta}j}$ and $\boldsymbol{J}_{\boldsymbol{\beta\beta}}$ instead, as discussed in Chapter I.

Figure 21 presents values of $\tilde{\theta}_j$. Obviously, the largest $\tilde{\theta}_j$ among 30 Fourier coefficients corresponds to $j = 16$. Based on the arguments about definition of high frequency in subsection 4.3, it appears that the true density is high frequency in this case. Thus, it is a good choice to use $k = 20$ and $\pi_j = 1/(1+j)^2$ in our later work. For the Ledwina data driven smooth test, $W_{S2}$ and $W_T$ defined in Chapter I, the $S2$ and $T$ that optimize their selection criteria are both 20 with upper bound 20.

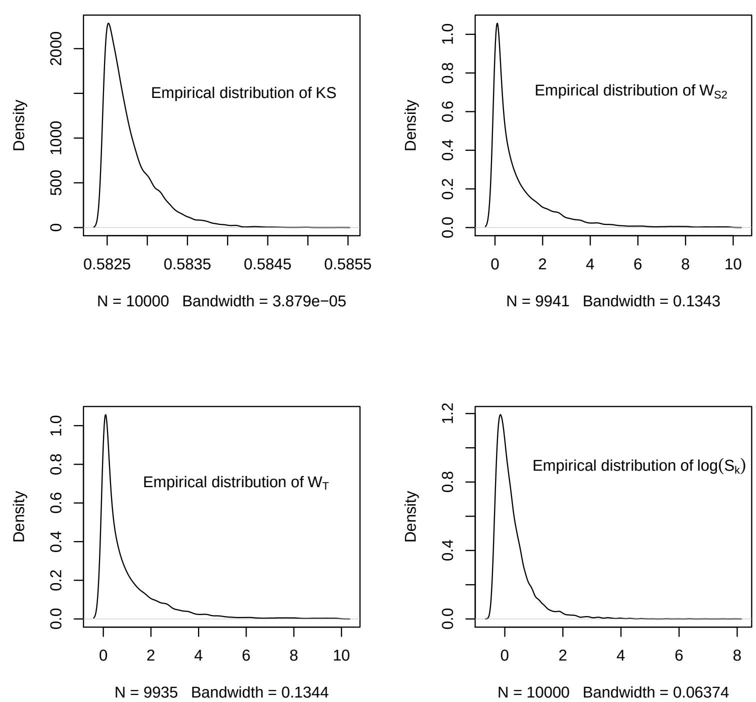


Figure 22.

The empirical distributions of the considered test statistics: $KS$, $W_{S2}$, $W_T$ and $S_k$.

Approximations to P-values are determined by assuming that the true distribution is a mixture of three normals. A random sample of size $n$ is generated from the fitted mixture of normals, where $n$ is 1246. The Kolmogorov-Smirnov test $KS$, Ledwina data-driven smooth test $W_{S2}$ and $W_T$, and proposed test $S_k$ are calculated from the data so generated. This process is repeated 10,000 times independently, and P-values for the considered tests are approximated by comparison of each statistic with the appropriate empirical distribution of 10,000 values. The empirical distributions of the four considered test statistics are presented in Figure 22. The corresponding 95th percentiles of $KS$, $W_{S2}$, $W_T$ and $\log(S_k)$ are 0.583, 4.067, 4.085 and 1.471 respectively.

In this process, the number of times $S2 = 1$ is 9911. This result agrees with the conclusions in Kallenberg and Ledwina (1997b) that the selection rule $S2$ concentrates on dimension 1 under $H_0$. The number of times $T = S2$ is 9992, which means Ledwina's test $W_T$ uses the BIC selection rule in most replications. The P-values of $KS$, $W_{S2}$, $W_T$ and $S_k$ obtained from the above process are 0.2067, 0, 0 and 0, respectively. The last three tests give the same results for these data. Apparently then the structure found by the S-J bandwidth=0.03438 is significant. Also we are not surprised that the $KS$ test fails to reject the null hypothesis due to its well known lack of power for high frequency alternatives.

# CHAPTER V

# CONCLUSIONS

## 5.1. Summary

In this dissertation, frequentist-Bayes goodness-of-fit tests are proposed. The key idea of the new test statistics is the combination of Bayesian and score test ideas. More precisely, the null hypothesis is rejected if the value of the proposed statistic, which corresponds to substituting score tests for log-likelihood ratios in a posterior probability, is large. The test is subsequently carried out in a frequentist way. Alternatives to the null hypothesis are modeled by a sequence of classical models, which need not be nested. A similar approach based on score tests is applied to achieve computational simplicity.

A Laplace approximation to the marginal likelihoods in the posterior probability of the null hypothesis is used, since only in very special circumstances can the marginal likelihoods be determined exactly. In the Bayesian world, the currently most popular means of approximating such quantities is to use MCMC, which is rather time consuming. Laplace approximation provides a general way to approach marginalization problems.

The proposed test statistics are weighted sums of exponentiated squared Fourier coefficients, where the weights depend on prior probabilities. A version of such a sum with the selected optimal weights has excellent power properties in simulation studies. These results suggest that it is not necessary to use adaptive test statistics dependent on data-driven smoothing parameters in order to obtain an omnibus goodness-of-fit test with good overall power. A simple weighted sum of independent Fourier components, as suggested in this dissertation, does the trick. An application of the proposed test to an interesting real data problem shows that the proposed test is powerful for high frequency alternatives.

In addition, theoretical work has been done to investigate properties of the proposed frequentist-Bayes tests. The asymptotic distribution of the test statistic is found, and it is shown that the test can detect $1/\sqrt{n}$ local alternatives.

## 5.2. Future Research

Our study shows that the proposed omnibus goodness-of-fit tests are powerful. Future research includes application of the proposed tests to various real-world problems. For example, goodness-of-fit tests are widely used in risk management. We would like to discuss the application of our frequentist-Bayes tests to this area below.

### 5.2.1. Validation of Default Probabilities

> "In conclusion, at present no really powerful tests of adequate calibration are currently available. Due to the correlation effects that have to be respected there even seems to be no way to develop such tests. Existing tests are rather conservative - such as the binomial test and the chi-square test - or will only detect the most obvious cases of miscalibration as in the case of the normal test."
>
> Basel Committee on Banking Supervision (2005)

The above quote from a study by the Basel Committee on Banking Supervision (BIS) relates to the current test statistics for validating probabilities of default (PD), which are used by banks to forecast credit default events. Banks are required by regulatory authorities, such as the BIS, to report the accuracy of their default probability estimates. They must demonstrate to their supervisor that the internal validation process allows assessing the performance of internal rating and risk estimation systems consistently and meaningfully. In particular, "banks must regularly compare realized default rates with estimated PDs for each grade and be able to demonstrate that the realized default rates are within the expected range for that grade." [cf. Basel Committee on Banking Supervision (2004).] Such a comparison asks for an adequate statistical test procedure. It is of interest to apply the ideas proposed in my dissertation to develop such test statistics that overcome the absence of sufficient historical default data and dependence of credit default events.

### 5.2.2. Goodness-of-fit Tests for Copulas

The multivariate normality of the latent variables is a core assumption of the KMV and CreditMetrics models in risk management, but there is no compelling reason to choose a multivariate normal (Gaussian) distribution for asset values. Moreover, even if individual default probabilities of obligors and the matrix of latent variable correlations are fixed, it is still possible to develop alternative models leading to much heavier-tailed loss distributions. In recent years, copulas have proved to be useful in understanding how a multivariate latent variable distribution determines the distribution of the number of defaults in a portfolio and with it, the need for a simple and reliable method to choose the right copula family.

Existing methods present numerous difficulties and none is completely satisfactory. Most of those rely on previous estimation of an optimal parameter set. As a result, comparisons are made between copulas with given parameters, and not between copula families. It would be of interest to investigate a model selection method independent of the parameter choice by utilizing our Bayesian formulation.

### 5.2.3. Extreme Value Distribution Selection

Extreme event risk is present in almost every area of risk management. No matter which type of risk we are concerned with, implementing risk management models which allow for rare but damaging events, and permitting the measurement of their consequences is one of the greatest challenges to the risk manager. The challenge of analyzing and modeling extreme values is that there are only a few observations for which to build a model, and there are ranges of extreme values that have yet to occur. To meet the challenge, researchers must assume a certain distribution. The extreme value distributions (EVD) are frequently used to develop appropriate probabilistic models and assess the risks caused by these events.

The selection among distributional forms is an important task. We can use goodness of fit tests to compare the fit of the extreme value distributions. There are a few tests for the extreme value distribution, notably the Sherman (1957) and an adaptation of Kolmogorov-Smirnov. However, most existing tests are frequentist and tend to overfit (i.e. be too lenient) or be conservative. It would also be of interest to investigate how to best select the fitting distribution by utilizing the combination of Bayesian and frequentist statistics to overcome the intricacies associated with sparseness.

# REFERENCES

Aerts, M., Claeskens, G., and Hart, J. D. (2004), "Bayesian-Motivated Tests of Function Fit and Their Asymptotic Frequentist Properties," *The Annals of Statistics*, 32, 2580-2615.

Angus, J. E. (1982), "Goodness-of-fit Tests for Exponentiality Based on A Loss-of-memory Type Functional Equation," *Journal of Statistical Planning and Inference*, 6, 241-251.

Ascher, S. (1990), "A Survey of Tests for Exponentiality," *Communications in Statistics - Theory and Methods*, 19, 1811-1825.

Barron, A. R., and Cover, T. M. (1991), "Minimum Complexity Density Estimation," *IEEE Transactions on Information Theory*, 37, 1034-1054.

Bayarri, M. J., and Berger, J. O. (2004), "The Interplay of Bayesian and Frequentist Analysis," *Statistical Science*, 19, 58-80.

Bickel, P. J., and Ritov, Y. (1992), "Testing Goodness Fit: New Approach," *Nonparametric Statistics and Related Topics*, Amsterdam: Elsevier Science Publishers B.V., 51-57.

Bogdan, M. (1995), "Data Driven Versions of Pearson's Chi-square Test for Uniformity," *Journal of Statistical Computation and Simulation*, 52, 217-237.

Bogdan, M., and Ledwina, T. (1996), "Testing Uniformity via Log-spline Modeling," *Statistics*, 28, 131-157.

Bowman, A. W. (1992), "Density-based Tests for Goodness of Fit," *Journal of Statistical Computation and Simulation*, 40, 1-13.

Burke, M. D., and Gombay, E. (1988), "On Goodness-of-fit and the Bootstrap," *Statistics & Probability Letters*, 6, 287-293.

Chang, M., and Chow, S.-C. (2005), “A Hybrid Bayesian Adaptive Design for Dose Response Trials,” *Journal of Biopharmaceutical Statistics*, 15, 677-691.

Conrad, J., Botner, O., Hallgren, A., and Pérez de los Heros, C. (2003), “Including Systematic Uncertainties in Confidence Interval Construction for Poisson Statistics,” *Physical Review D*, 67, 012002.

Cox, D. R., and Hinkley, D. V. (1974), *Theoretical Statistics*. London: Chapman and Hall.

D’Agostino, R. B., and Stephens, M. A. (1986), *Goodness-of-Fit Techniques*. New York: Dekker.

Durbin, J. (1973), *Distribution Theory for Tests Based on the Sample Distribution Function*, Philadelphia: SIAM.

Eubank, L., Hart, J. D., and LaRiccia, V. N. (1993) “Testing Goodness Fit via Nonparametric Function Estimation Techniques,” *Communications in Statistics - Theory and Methods*, 22, 3327-3354.

Eubank, L., and LaRiccia, N. (1992), “Asymptotic Comparison of Cramér-von Mises and Nonparametric Function Estimation Techniques for Testing Goodness-of-fit,” *The Annals of Statistics*, 20, 2071-2086.

Fan, J. (1996), “Tests of Significance Based on Wavelet Thresholding and Neyman’s Truncation,” *Journal of the American Statistical Association*, 91, 647-688.

Gail, M. H., and Gastwirth, J. L. (1978), “A Scale-free Goodness-of-fit Test for the Exponential Distribution Based on Gini Statistic,” *Journal of the Royal Statistical Society*, Ser. B, 40, 350-357.

Gan, F. F., and Koehler, K. J. (1990), "Goodness-of-fit Tests Based on P-P Probability Plots," *Technometrics*, 32, 289-303.

Good, I. J. (1957), "Saddle-point Methods for the Multinomial Distribution," *The Annals of Mathematical Statistics*, 28, 861-881.

Hart, J.D. (2009), "Frequentist-Bayes Lack-of-fit Tests Based on Laplace Approximations," *Journal of Statistical Theory and Practice*, 3, 681-704.

Haughton, D. M. A. (1988), "On the Choice of A Model to Fit Data from An Exponential Family," *The Annals of Statistics*, 16, 342-355.

Inglot,T., Kallenberg, W. C. M., and Ledwina, T. (1994), "Power Approximations to and Power Comparison of Certain Goodness-of-Fit Tests," *Scandinavian Journal of Statistics*, 21, 131-145.

Inglot,T., Kallenberg, W. C. M., and Ledwina, T. (1997), "Data Driven Smooth Tests for Composite Hypotheses," *The Annals of Statistics*, 25, 1222-1250.

Inglot,T., and Ledwina, T. (1990), "On Probabilities of Excessive Deviations for Kolmogorov-Smirnov, Cramér-von Mises and Chi-square Statistics," *The Annals of Statistics*, 18, 1491-1495.

Inglot,T., and Ledwina, T (2006), "Towards Data Driven Selection of A Penalty Function for Data Driven Neyman Tests," *Linear Algebra and Its Applications*, 417, 124-133.

Janssen, A. (1995), "Principal Component Decomposition of Non-parametric Tests," *Probability Theory and Related Fields*, 101, 193-209.

Javitz, H. S. (1975), "Generalized Smooth Tests of Goodness of Fit, Independence, and Equality of Distributions," Ph.D. dissertation, University California, Berkeley, CA.

Kaigh, W. D. (1992), “EDF and EQF Orthogonal Component Decompositions and Tests of Uniformity,” *Journal of Nonparametric Statistics*, 1, 313-334.

Kallenberg, W. C. M., and Ledwina, T. (1995a), “Consistency and Monte Carlo Simulation of A Data Driven Version Smooth Goodness-of-fit,” *The Annals of Statistics*, 23, 1594-1608.

Kallenberg, W. C. M., and Ledwina, T. (1995b), “On Data Driven Neyman's Tests,” *Probability and Mathematical Statistics* 15, 409-426.

Kallenberg, W. C. M., and Ledwina, T. (1997a), “Data Driven Smooth Tests for Composite Hypotheses: Comparison of Powers,” *Journal of Statistical Computation and Simulation*, 59, 101-121.

Kallenberg, W. C. M., and Ledwina, T. (1997b), “Data-driven Smooth Tests When the Hypothesis is Composite,” *Journal of the American Statistical Association*, 92, 1094-1104.

Kass, R. E., and Wasserman, L. (1996),“The Selection of Prior Distributions by Formal Rules,” *Journal of the American Statistical Association*, 91, 1343-1370.

Khamaladze, E. V. (1981), “Martingale Approach to the Theory of Goodness of Fit Tests,” *Theory of Probability and Its Applications*, 26, 240-257.

Kim, J. (1992), “Testing Goodness-of-fit via Order Selection Criteria,” Ph.D. dissertation, Texas A&M University, College Station, TX.

Koning, A. J. (1992), “Approximation of Stochastic Integrals with Applications to Goodness-of-fit tests,” *The Annals of Statistics*, 20, 428-454.

Koning, A. J. (1993), “Stochastic Integrals and Goodness-of-fit Tests,” *Mathematical*

*Centre Tracts*, 98, Mathematisch Centrum, Amsterdam.

Kopecky, K. J., and Pierce, D. A. (1979), "Efficiency of Smooth Goodness-of-Fit Tests," *Journal of the American Statistical Association*, 74, 393- 397.

LaRiccia, V. N. (1991), "Smooth Goodness-of-fit Tests: A Quantile Function Approach," *Journal of the American Statistical Association*, 86, 427-431.

Ledwina, T. (1994), "Data Driven Version of Neyman's Smooth Test of Fit," *Journal of the American Statistical Association*, 89, 1000-1005.

Milbrodt, H., and Strasser, H. (1990), "On the Asymptotic Power of the Two-sided Kolmogorov-Smirnov Test," *Journal of Statistical Planning and Inference*, 26, 1-23.

Miller, F. L., and Quesenberry, C. P. (1979), "Power Studies of Tests for Uniformity II," *Communications in Statistics - Simulation and Computation*, 8, 271-290.

Neuhaus, G. (1976), "Asymptotic Power Properties of the Cramér-von Mises Test under Contiguous Alternatives," *Journal of Multivariate Analysis*, 6, 95-110.

Neuhaus, G. (1979), "Asymptotic Theory of Goodness of Fit When Parameters are Present: A Survey," *Mathematische Operationsforschung und Statistik. Series Statistics*, 10, 479-494.

Neyman, J. (1937), " 'Smooth Test' for Goodness of Fit," *Skandinavian Aktuarietidskr*, 20, 149-199.

Neyman, J. (1980), "Some Memorable Incidents in Probabilistic/Statistical Studies," in *Asymptotic Theory of Statistical Tests and Estimation*, ed. I.M. Chakravarti, New York: Academic Press, pp. 1-32.

Nikitin, Y. Y. (1984), "Local Bahadur Optimality and Characterization Problems," *Theory of Probability and Its Applications*, 29, 79-92.

Nikitin, Y. Y. (1995), "Asymptotic Efficiency of Nonparametric Tests," Cambridge University Press.

Pearson, E. S., D'Agostino, R. B., and Bowman, K. O. (1977), "Tests for Departure from Normality: Comparison of Powers," *Biometrika*, 64, 231-246.

Quesenberry, C. P., and Miller, F. L. (1977), "Power Studies of Some Tests for Uniformity," *Journal of Statistical Computation and Simulation*, 5, 169-191.

Rayner, J. C. W., and Best, D. J. (1989), *Smooth Tests of Goodness of Fit*, New York: Oxford University Press.

Rayner, J. C. W., and Best, D. J. (1990), "Smooth Tests of Goodness of Fit: An Overview," *International Statistical Review*, 58, 9-17.

Rayner, J. C. W., Thas, O., and Best, D. J. (2009), *Smooth Tests of Goodness of Fit: Using R*, Singapore: John Wiley & Sons.

Rissanen, J. (1983), "A Universal Prior for Integers and Estimation by Minimum Description Length," *The Annals of Statistics*, 11, 416-431.

Rissanen, J. (1987), "Stochastic Complexity," *Journal of the Royal Statistical Society*, Ser. B, 49, 223-239.

Schwarz, G. (1978), "Estimating the Dimension of A Model," *The Annals of Statistics*, 6, 461-464.

Sen, P. K., and Singer, J. M. (1993), *Large Sample Methods in Statistics.* New York: Chapman and Hall.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

Sheather, S. J., and Jones M. C. (1991), " A Reliable Data-based Bandwidth Selection Method for Kernel Density Estimation," *Journal of the Royal Statistical Society*, Ser. B, 53, 683-690.

Sherman, L. K. (1957), "Percentiles of the $\Omega_n$ Statistic," *Annals of Mathematical Statistics*, 28, 259-268.

Silverman, B. W. (1986), *Density Estimation*. London: Chapman and Hall.

Thomas, D. R., and Pierce, D. A. (1979), "Neyman's Smooth Goodness-of-fit Test When the Hypothesis is Composite," *Journal of the American Statistical Association*, 74, 441-445.

Verdinelli, I., and Wasserman, L. (1998), "Bayesian Goodness-of-fit Testing Using Infinite-dimensional Exponential Families," *The Annals of Statistics*, 26, 1215-1241.

Zhang, J. (2002), "Powerful Goodness-of-fit Tests Based on Likelihood Ratio," *Journal of the Royal Statistical Society*, Ser. B, 64, 281-294.

# APPENDIX A

## LOCATION-SCALE INVARIANCE OF $R_J(\boldsymbol{\beta})$

The proof that $R_j(\boldsymbol{\beta})$ does not depend on the location-scale parameter is provided here.

In a location-scale family $\boldsymbol{\beta} = (\beta_1, \beta_2)$, let $I_{1j}$ and $I_{2j}$ denote elements of $I_{\boldsymbol{\beta}j} = (I_{1j}, I_{2j})^{\mathrm{T}}$ for $j = 1, 2, \ldots$. Since

$$\frac{\partial}{\partial \beta_1} f(x; \boldsymbol{\beta}) = -\frac{1}{\beta_2^2} \frac{\partial f_0\left(\frac{x-\beta_1}{\beta_2}\right)}{\partial\left(\frac{x-\beta_1}{\beta_2}\right)},$$

and

$$\frac{\partial}{\partial \beta_2} f(x; \boldsymbol{\beta}) = -\frac{1}{\beta_2^2} f_0\left(\frac{x-\beta_1}{\beta_2}\right) - \frac{1}{\beta_2^2} \frac{\partial f_0\left(\frac{x-\beta_1}{\beta_2}\right)}{\partial\left(\frac{x-\beta_1}{\beta_2}\right)} \left(\frac{x-\beta_1}{\beta_2}\right),$$

by the definition of $I_{\boldsymbol{\beta}j}$, $I_{1j}$ and $I_{2j}$ take the form

$$\begin{aligned} I_{1j} &= -\int \frac{\partial}{\partial \beta_1} u_j[F(x; \boldsymbol{\beta})] f(x; \boldsymbol{\beta}) dx \\ &= \int u_j[F(x; \boldsymbol{\beta})] \frac{\partial}{\partial \beta_1} f(x; \boldsymbol{\beta}), \end{aligned}$$

since

$$\frac{\partial}{\partial \beta_1} \left(u_j[F(x; \boldsymbol{\beta})] f(x; \boldsymbol{\beta})\right) = \frac{\partial}{\partial \beta_1} u_j[F(x; \boldsymbol{\beta})] f(x; \boldsymbol{\beta}) + u_j[F(x; \boldsymbol{\beta})] \frac{\partial}{\partial \beta_1} f(x; \boldsymbol{\beta})$$

and

$$\int_{-\infty}^{\infty} u_j[F(x; \boldsymbol{\beta})] f(x; \boldsymbol{\beta}) dx = 0.$$

Therefore,

$$I_{1j} = \int u_j \left[ F_0 \left( \frac{x-\beta_1}{\beta_2} \right) \right] \left( -\frac{1}{\beta_2^2} \frac{\partial f_0 \left( \frac{x-\beta_1}{\beta_2} \right)}{\partial \left( \frac{x-\beta_1}{\beta_2} \right)} \right) dx \tag{A.1}$$
$$= -\frac{1}{\beta_2} \int u_j [F_0(y)] f_0'(y) dy.$$

Similarly,

$$\begin{aligned} I_{2j} &= -\int \frac{\partial}{\partial \beta_2} u_j [F(x;\boldsymbol{\beta})] f(x;\boldsymbol{\beta}) dx \\ &= \int u_j [F(x;\boldsymbol{\beta})] \frac{\partial}{\partial \beta_2} f(x;\boldsymbol{\beta}) dx \\ &= \int u_j \left[ F_0 \left( \frac{x-\beta_1}{\beta_2} \right) \right] \left( -\frac{1}{\beta_2^2} \frac{\partial f_0 \left( \frac{x-\beta_1}{\beta_2} \right)}{\partial \left( \frac{x-\beta_1}{\beta_2} \right)} \left( \frac{x-\beta_1}{\beta_2} \right) \right) dx \\ &= -\frac{1}{\beta_2} \int u_j [F_0(y)] f_0'(y) y dy. \end{aligned} \tag{A.2}$$

For $t = 1, 2$ and $u = 1, 2$

$$\frac{\partial^2}{\partial \beta_t \partial \beta_u} \log f(x;\boldsymbol{\beta}) = \frac{1}{f(x;\boldsymbol{\beta})} \frac{\partial^2 f(x;\boldsymbol{\beta})}{\partial \beta_t \partial \beta_u} - \frac{1}{f(x;\boldsymbol{\beta})^2} \frac{\partial f(x;\boldsymbol{\beta})}{\partial \beta_t} \frac{\partial f(x;\boldsymbol{\beta})}{\partial \beta_u},$$

and so

$$\begin{aligned} E_{\boldsymbol{\beta}} \frac{\partial^2}{\partial \beta_t \partial \beta_u} \log f(x;\boldsymbol{\beta}) &= \int \frac{\partial^2 f(x;\boldsymbol{\beta})}{\partial \beta_t \partial \beta_u} dx - \int \frac{1}{f(x;\boldsymbol{\beta})} \frac{\partial f(x;\boldsymbol{\beta})}{\partial \beta_t} \frac{\partial f(x;\boldsymbol{\beta})}{\partial \beta_u} dx \\ &= -\int \frac{1}{f(x;\boldsymbol{\beta})} \frac{\partial f(x;\boldsymbol{\beta})}{\partial \beta_t} \frac{\partial f(x;\boldsymbol{\beta})}{\partial \beta_u} dx. \end{aligned}$$

It follows that

$$\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \frac{1}{\beta_2^2} \begin{pmatrix} \int \frac{[f_0'(y)]^2}{f_0(y)} dy & \int f_0'(y) dy + \int y \frac{[f_0'(y)]^2}{f_0(y)} dy \\ \int f_0'(y) dy + \int y \frac{[f_0'(y)]^2}{f_0(y)} dy & \int y^2 \frac{[f_0'(y)]^2}{f_0(y)} dy + 2 \int y f_0'(y) dy + 1 \end{pmatrix}. \tag{A.3}$$

The definition of $R_j(\boldsymbol{\beta})$, (A.1), (A.2) and (A.3) implies that $R_j(\boldsymbol{\beta})$ does not depend on the parameter $\boldsymbol{\beta}$.

# VITA



| | |
|---|---|
| Name: | Qi Wang |
| Address: | Department of Statistics |
| | c/o Dr. Jeffrey D. Hart |
| | Texas A&M University |
| | College Station, TX 77843-3143 |
| Email Address: | wangqi@stat.tamu.edu |
| Education: | B.S., Statistics, Zhejiang University, China, 2006 |
| | M.S., Texas A&M University, USA, 2008 |
| | Ph.D., Texas A&M University, USA, 2011 |