On Estimated Projection Pursuit Type Crámer-von Mises Statistics *

Li-Xing Zhu[1], Kai-Tai Fang[2] and M. Ishaq Bhatti[3]
[1] Chinese Academy of Sciences and Catholic University of Louvain
[2] Hong Kong Baptist University and Chinese Academy of Sciences
[3] King Saud University and Griffith University

Abstract

This paper addresses the problem of testing for a multivariate distribution, which belongs to a known parametric distribution family. The estimated Crámer-Von Mises type test statistics are constructed using projection pursuit technique. Some interested properties of the test statistics, like asymptotics , bootstrap approximations and the tail behavior of the limits of test statistics are investigated. For computational reasons, an approximation via number theoretic method to extreme value and integral on super sphere surface is considered.

Key words and Phrases: Bootstrap, Number-theoretic method, Projection pursuit, Tail probability.

AMS (1991) subject classification: 62H99 62H40, 56C05.

1. Introduction

Suppose that G(.) is a d-dimensional probability measure and that $\underline{x}_1, \ldots, \underline{x}_n$ are i.i.d. sample from G(.). A statistical problem addressed in this article is how to test G(.) which is a member of a known family of probability measures with unknown parameter, that is, how to test $G(.) = P_\theta(.)$, where the functional form of P is known, and θ is an unknown parameter. For this problem the estimated Crámer-Von Mises test (see, Eastwood (1993) and the recent references therein) is often a primary choice. Let P_n be the empirical measure determined by the sample $\{\underline{x}_1, \ldots, \underline{x}_n\}$, the estimated Crámer-Von Mises test statistic is expressed as

$$n\int (P_n I(\underline{x} \le \underline{t}) - P_{\hat\theta}(\underline{t}))^2 dP_{\hat\theta}(\underline{t}), \tag{1}$$

*Supported by a Hong Kong UPGC grant, the NNSF of China and the URIS of Griffith University, Australia. Part of this work was done when the first author was visiting Insitut de Statistique, Catholique Universite Louvain, Belgium. Research support from the contract " Projet d'Actions de Recherche Concertées " (PARC No. 93/98-164) of the Belgian Government is grateful acknowledged.

where I(A) is the indicator function of a set A, $P_n f$ stands for the expected value of $f(\underline{x})$, i.e., $\int f(\underline{x}) dP_n(\underline{x})$ and $\hat{\theta}$ is an estimate of θ. When the dimension of a vector d is one, the properties of this statistic, i.e. especially that of its weak limit, have been well known. For an example, see Shorack and Wellner (1986).

However, when the dimension d is larger than one the sparseness of sample points in high-dimensional space, called " Curse of Dimensionality " (see, Huber (1985)), is an obstacle for an effective use of this sort of statistic. The main objective of this article is to overcome this difficulty by using the projection pursuit technique. That is, we suggest two estimated projection pursuit type Crámer-Von Mises test statistics W_{n1} and W_{n2} which are expressed as follows.

$$W_{n1} = n \sup_{a \in S^{d-1}} \int (P_n I(\underline{a}^\tau \underline{x} \leq t) - P_{\hat{\theta}} I(\underline{a}^\tau \underline{x} \leq t))^2 P_{\hat{\theta}} I(\underline{a}^\tau \underline{x} \leq t) \tag{2}$$

and

$$W_{n2} = n \int \int (P_n I(\underline{a}^\tau \underline{x} \leq t) - P_{\hat{\theta}} I(\underline{a}^\tau \underline{x} \leq t))^2 dP_{\hat{\theta}} I(\underline{a}^\tau \underline{x} \leq t) d\mu(\underline{a}), \tag{3}$$

where $S^{d-1} = \{\underline{a} \in R^d, ||\underline{a}|| = 1\}$ is the unit super sphere surface in R^d and μ is the uniform distribution on S^{d-1}.

In this paper we will further investigate the properties of W_{n1} and W_{n2} defined at (2) and (3) respectively. In section 2, we will illustrate the limiting properties of W_{n1} and W_{n2}. To determine the critical values, we will use bootstrap approximation (see, *e.g.* Gine and Zinn (1990)) and apply a number theoretic method to deal with the approximation to the extreme value and integral on S^{d-1} to be discussed in section 3. Noticing that the accurate distributions of the limits of W_{n1} and W_{n2} are not tractable, we in section 4 will present the tail behaviors of W_{n1} and W_{n2}, when $P_\theta(.)$ is a nonatomic, elliptically symmetric distribution with unknown location parameter $\underline{\theta}$ and dispersion matrix $\underline{\Sigma}$. The final section of this article contains some concluding remarks. Throughout this paper, $\underline{x}$ stands for the vector.

2. The asymptotics of W_{n1} and W_{n2}

We first investigate the convergence of W_{n1} and W_{n2} in distribution.

According to the statistical usefulness, a good estimate of θ is often in the form of a regular estimate, that is

$$\hat{\theta} = \theta + n^{-1} \sum_{j=1}^{n} L_\theta(\underline{x}_j) + O_p(n^{1/2}) \tag{4}$$

for some d-dimensional function L_θ satisfying $\int L_\theta(\underline{x}) dP_\theta(\underline{x}) = 0$ and $V = \int L_\theta(\underline{x}) L_\theta(\underline{x})^\tau dP_\theta(\underline{x})$ a positive definite matrix.

On the other hand, we provide $P_\theta(.)$ is uniformly differentiable to guarantee the convergence of W_{n1} and W_{n2} as follows:

$$\sup_{a \in S^{d-1}} \sup_{t \in R^1} \quad \{\int (P_{\theta'} I(\underline{a}^\tau \underline{x} \leq t) - P_\theta I(\underline{a}^\tau \underline{x} \leq t) - (\theta' - \theta)^\tau A(a,t))^2 dP_\theta(t)\}^{1/2}$$

$$= o(||\theta' - \theta||) \quad near\ \theta, \tag{5}$$

for some fixed d-dimensional function A(. , .) in (5) with all its components in the collection $[S^{d-1} \times R^1]$ consisting of all continuous functions. Hence

W_{n1} and W_{n2} can be written as

$$\begin{aligned} W_{n1} &= n \sup_{a \in S^{d-1}} \int \{(P_n I(\underline{a}^\tau \underline{x} \le t) - P_\theta I(\underline{a}^\tau \underline{x} \le t) - P_n L_\theta^\tau(\underline{x}) A(\underline{a}, t)\}^2 \\ & \quad dP_{\hat{\theta}} I(\underline{a}^\tau \underline{x} \le t) + o_p(1) \end{aligned} \tag{6}$$

and

$$\begin{aligned} W_{n2} &= n \int \int \{(P_n I(\underline{a}^\tau \underline{x} \le t) - P_\theta I(\underline{a}^\tau \underline{x} \le t) - P_n L_\theta^\tau(\underline{x}) A(\underline{a}, t)\}^2 \\ & \quad dP_{\hat{\theta}} I(\underline{a}^\tau \underline{x} \le t) d\mu(\underline{a}) + o_p(1). \end{aligned} \tag{7}$$

It is well known that (see Dudley (1978) or Pollard (1984)) the stochastic process $\{\sqrt{n}(P_n I(\underline{a}^\tau \underline{x} \le t) - P_\theta I(\underline{a}^\tau \underline{x} \le t) - P_n L_\theta^\tau(\underline{x}) A(\underline{a}, t))\} : a \in S^{d-1}, t \in R^1\}$ converges weakly to a centered Gaussian process, $\{W(\underline{a}, t) - Z_\theta^\tau A(a, t) : \underline{a} \in S^{d-1}, t \in R^1\}$, where sample paths continue uniformly and are bounded under $L^2(P_\theta(.))$ seminorm with the covariance function,

$$\begin{aligned} R((a,t),(a_1,t_1)) = \quad & P_\theta(I(\underline{a}^\tau \underline{x} \le t)(I(\underline{a}_1^\tau \underline{x} \le t_1) \\ & -P_\theta(I(\underline{a}^\tau \underline{x} \le t) P_\theta(I(\underline{a}_1^\tau \underline{x} \le t_1)) \\ & +P_\theta\{A(\underline{a}, t))^\tau L_\theta(\underline{x}) L_\theta^\tau(\underline{x}) A(\underline{a}, t_1))\} \\ & -P_\theta(I(\underline{a}^\tau \underline{x} \le t) L_\theta^\tau(\underline{x}) A(\underline{a}, t)) \\ & -P_\theta(I(\underline{a}_1^\tau \underline{x} \le t_1) - L_\theta^\tau(\underline{x}) A(\underline{a}_1, t_1)). \end{aligned}$$

Consequently, from (6) and (7) we have

$$W_{n1} \Longrightarrow W_1 \triangleq \sup_{a \in S^{d-1}} \int \{W(\underline{a}, t) - Z_\theta^\tau A(\underline{a}, t))^2 dP_\theta I(\underline{a}^\tau \underline{x} \le t) \tag{8}$$

and

$$W_{n2} \Longrightarrow W_2 \triangleq \int \int (W(\underline{a}, t) - Z_\theta^\tau A(\underline{a}, t))^2 dP_\theta I(\underline{a}^\tau \underline{x} \le t) d\mu(\underline{a}), \tag{9}$$

respectively, where the notation " $\Longrightarrow$ " means the weak convergence.

3. Bootstrap Approximations and Number Theoretic Method

3.1. Two Bootstrap Procedures

¿From equations (8) and (9) above we can see that the accurate distributions of W_{n1} , W_{n2} , W_1 and W_2 depend on the unknown parameter θ and are not tractable. Hence we suggest bootstrap approximations to W_{n1} and W_{n2}. To this end, there are two bootstrap procedures which may be applicable. The first is a naive bootstrap procedure. Independently generate bootstrap sample $\underline{x}_i^*$ from $\{\underline{x}_1, \ldots, \underline{x}_n\}$ and then compute

$$W_{n1}^* = n \sup_a \int (W_n^*(\underline{a}, t))^2 dP_{\hat{\theta}} I(\underline{a}^\tau \underline{x} \le t) \tag{10}$$

and

$$W_{n2}^* = n\int\int (W_n^*(\underline{a},t))^2 dP_{\hat\theta} I(\underline{a}^\tau \underline{x} \le t) d\mu(\underline{a}), \tag{11}$$

where

$$W_n^*(\underline{a},t) = P_n^* I(\underline{a}^\tau \underline{x}^* \le t) - P_n I(\underline{a}^\tau \underline{x} \le t) - (P_{\hat\theta *} I(\underline{a}^\tau \underline{x}^* \le t) - P_{\hat\theta} I(\underline{a}^\tau \underline{x} \le t))$$

$\hat\theta = \hat\theta(\underline{x}_1, \ldots, \underline{x}_n)$, $\hat\theta^* = \hat\theta(\underline{x}_1^*, \ldots, \underline{x}_n^*)$ and P_n^* is an empirical measure based on $\underline{x}_1^*, \ldots, \underline{x}_n^*$. By Gine and Zinn (1990, Theorem 2.4), we know the following asymptotic equivalences hold: for almost series $\{\underline{x}_1, \ldots, \underline{x}_n, \ldots\}$ under H_0,

$$\begin{aligned} &\{n(P_n^* I(\underline{a}^\tau \underline{x}^* \le t) - P_n I(\underline{a}^\tau \underline{x} \le t)) : a \in S^{d-1}, t \in R^1\} \\ \approx\ &\{n(P_n I(\underline{a}^\tau \underline{x} \le t) - P_\theta I(\underline{a}^\tau \underline{x} \le t)) : a \in S^{d-1}, t \in R^1\} \end{aligned} \tag{12}$$

and

$$\{P_{\hat\theta *}(\underline{t}) - P_{\hat\theta} I(\underline{t}) : \underline{t} \in R^d\} \approx (P_{\hat\theta}(\underline{t}) - P_\theta(\underline{t}), \underline{t} \in R^d\} \tag{13}$$

Consequently, the bootstrap test statistics W_{n1}^* and W_{n2}^* which base on the process $\{W_n^*(\underline{a},t) : \underline{a} \in S^d, t \in R^1\}$ have the same limits as W_{n1} and W_{n2}. On the other hand, under alternative, similar conclusions to (12) and (13) hold when P_θ is replaced by Q, the underlying distribution of $\underline{x}$. W_{n1}^* and W_{n2}^* will then have finite limits, while W_{n1} and W_{n2} will tend to infinity under any fixed alternative. Hence the test procedure based on the bootstrap approximation above is consistent.

We now turn to second bootstrap procedure. This will generate artificial data from the distribution $P_{\hat\theta}$ and then mimic the distribution of $W_{n1}(\ W_{n2})$. The algorithm is based on the following steps:

Step 1. Generate $\underline{z}_1, \cdots, \underline{z}_n$ from $P_{\hat\theta}$, where $\hat\theta = \hat\theta(\underline{x}_1, \cdots, \underline{x}_n)$;
Step 2. Let $\hat\theta_1 = \hat\theta(\underline{z}_1, \cdots, \underline{z}_n)$and define the statistics based on $\underline{z}_1, \cdots, \underline{z}_n$ by

$$W_{n11}^* = n \sup_a \int [(P_n I(\underline{a}^\tau \underline{z} \le t) - P_{\hat\theta_1} I(\underline{a}^\tau \underline{z} \le t)]^2 dP_{\hat\theta_1} I(\underline{a}^\tau \underline{z} \le t) \tag{14}$$

and

$$W_{n21}^* = n\int\int [P_n I(\underline{a}^\tau \underline{z} \le t) - P_{\hat\theta_1} I(\underline{a}^\tau \underline{z} \le t)]^2 dP_{\hat\theta_1} I(\underline{a}^\tau \underline{z} \le t) d\mu(\underline{a}). \tag{15}$$

Step 3. Repeat step 1 and 2 exactly c times to get values of W_{n11}^* and W_{n21}^* say $\{W_{n11(1)}^*, \cdots, W_{n11(c)}^*\}$ and $\{W_{n21(1)}^*, \cdots, W_{n21(c)}^*\}$ respectively, and

Step 4. Sort out these values in each set and use $[(1-\alpha)c]$-th largest values as the critical values of significant level α.

¿From the empirical process theory (e.g. Dudley (1978) or Pollard (1984)), it is easily derived that under H_0, $W_{n11}^*(W_{n21}^*)$ is asymptotically equivalent to $W_{n1}(W_{n2})$ and under alternative, if $\hat\theta$ converges in probability to with the rate $\sqrt{n}$, an $\hat\theta_A$ which may be different from θ, $W^*_{n11}(W^*_{n21})$ will converges weakly to a functional of certain Gaussian process. Actually, by using equations (4) to (7) of section 2, we can easily derive that

$$\begin{aligned} W^*_{n11} &= n \sup_{a \in S^{d-1}} \int \{P_n I(\underline{a}^\tau \underline{z} \le t) - P_{\hat{\theta}} I(\underline{a}^\tau \underline{z} \le t) - P_n L^\tau_{\hat{\theta}}(\underline{z}) A(\underline{a}, t)\}^2 \\ &\qquad dP_{\hat{\theta}} I(\underline{a}^\tau \underline{z} \le t) + o_p(1) \end{aligned} \tag{16}$$

and similarly for W^*_{n21}. Comparing to (6) ((7)), $W^*_{n11}(W^*_{n21})$ has the same limit behavior as W_{n1} (W_{n2}). Hence, in large sample sense, both bootstrap procedures share similar properties. However, for some cases the later will enjoy better properties. Especially, in the case where the parameter is location, we are able to show that the distribution of W^*_{n11} (W^*_{n21}) is the same as that of W_{n1} (W_{n2}) under null hypothesis. This means that, in the location case, one is able to approximate, when W^*_{n11} (W^*_{n21}) is used, the distribution of W_{n1} (W_{n2}) accurately as long as the repeated time c in the algorithm is large enough. We will perform some small simulation experiments to compare the performance of the power of W^*_{n11} (W^*_{n21}) to that of W^*_{n1} (W^*_{n2}). Actually, in location parameter case W^*_{n11} (W^*_{n21}) is a Monte Carlo test procedure which is quite similar to that proposed by Barnard (1963). Hall and Titterington (1989) pointed out some nice properties of Monte Carlo test. Detailed discussion is refered to their paper. We will see that from simulation experiments we conducted the performance of W^*_{n11} (W^*_{n21}) is better than that of W^*_{n1} (W^*_{n2}).

3.2. Number-theoretic method

We now turn to the computational aspects on evaluating W_{n1} (W_{n2}). There are two problems in computing W_{n1} (W_{n2}). One is that for each fixed direction $\underline{a} \in S^{d-1}$, there may be computational difficulty in evaluating

$$W_n(\underline{a}) = \int n\{(P_n I(\underline{a}_i^\tau \underline{x} \le t) - P_{\hat{\theta}} I(\underline{a}^\tau \underline{x} \le t)\}^2 dP_{\hat{\theta}} I(\underline{a}^\tau \underline{x} \le t). \tag{17}$$

As $P_{\hat{\theta}}$ is completely known, we then are able to solve this problem in terms of a Monte Carlo approximation. That is, generate random vectors $\{\underline{y}_1, \ldots, \underline{y}_m\}$ from $P_{\hat{\theta}}$ and substitute $W_n(\underline{a})$ by

$$\frac{1}{m} \sum_{j=1}^{m} \{(P_n I(\underline{a}^\tau \underline{x} \le \underline{a}^\tau \underline{y}_j) - P_{\hat{\theta}} I(\underline{a}^\tau \underline{x} \le \underline{a}^\tau \underline{y}_j))\}^2. \tag{18}$$

Clearly, the approximation should be accurate enough for large m. Next, note that W_{n1} and W_{n2} are respectively extreme value and integral on S^{d-1}. This means that when dimension d is larger the approximation to W_{n1} and W_{n2} then become attractive for computational reasons. We suggest an approximation via Number-theoretic method. Here by Number- theoretic method we choose a subset of S^{d-1}, say $\{\underline{a}_1, \ldots, \underline{a}_{k_n}\}$ and then to construct an approximation to W_{n1} and W_{n2} by

$$W_{nn1} = \sup_{1 \le i \le k_n} \int n\{(P_n I(\underline{a}_i^\tau \underline{x} \le t) - P_{\hat{\theta}} I(\underline{a}_i^\tau \underline{x} \le t)\}^2 dP_{\hat{\theta}} I(\underline{a}_i^\tau \underline{x} \le t), \tag{19}$$

and

$$W_{nn2} = \frac{n}{k_n}\sum_{i=1}^{k_n}\int\{P_n I(\underline{a}_i^\tau \underline{x} \le t) - P_{\hat{\theta}} I(\underline{a}_i^\tau \underline{x} \le t)\}^2 dP_{\hat{\theta}} I(\underline{a}_i^\tau \underline{x} \le t), \tag{20}$$

respectively.

It is important to note that $\{\underline{a}_1, \ldots, \underline{a}_{k_n}\}$ enjoys better uniformity on S^{d-1} than that of $\{\underline{a}'_1, \ldots, \underline{a}'_{k_n}\}$ on S^{d-1} chosen randomly such as Beran and Millar (1986) in the following sense,

$$\sup_{\delta\in\Delta} |\frac{1}{k_n}\sum_{i=1}^{k_n}\int I(\underline{a}_i \in \delta) - \mu(\delta)| = 0(k_n^{-1}(\log k_n)^{p-1})$$

but

$$\sup_{\delta\in\Delta} |\frac{1}{k_n}\sum_{i=1}^{k_n}\int I(\underline{a'}_i \in \delta) - \mu(\delta)| = 0_p(k_n^{-1/2}), \tag{21}$$

where in (21), Δ is a class of sets and δ is the set with the form

$$\begin{aligned}\delta(v_1, \ldots, v_{d-1}) = \{\underline{a} = \{(\cos(f_1(u_1)), \sin(f_1(u_1)), \cos(f_2(u_2)), \ldots,\\ \prod_{i=1}^{d-2}\sin(f_i(u_i))\cos(2\pi u_{d-1}), \prod_{i=1}^{d-2}\sin(f_i(u_i))\sin(2\pi u_{d-1})\} :\\ 0 \le u_i \le v_i \le 1, i = 1, \ldots, d-1\}\end{aligned} \tag{22}$$

and $f_i^{-1}(u) = F_i(u)$ is the distribution function with density function $g_i(u) = c(i)(\sin u)^{d-i-1}$, $u \in [0, Z], i = 1, \ldots, d-2$. The choosing procedure of $\{\underline{a}'_1, \ldots, \underline{a}'_{k_n}\}$ can be illustrated simply as follows.

(i) For a given integer $n(> p-1)$, choose an integer vector $\underline{h} = (\underline{h}_1, \ldots, \underline{h}_{d-1})$ which satisfy $1 \le h_i < n, \quad h_i \ne h_j, i \ne j$, let

$$\begin{cases} g_{ki} \equiv kh_i(mod\ n) & k = 1, \ldots, n,\ i = 1, \ldots, d-1 \\ c_{ki} \equiv (2g_{ki} - 1)/2n & k = 1, \ldots, n,\ i = 1, \ldots, d-1 \end{cases}$$

where we use the usual additive operation modulo n such that g_{ki} is confined by $1 \le g_{ki} < n$. Let $c_k = (c_{k-1}, \ldots, c_{k(d-1)})$. For more detail about the choice of $\underline{h} = (\underline{h}_1, \ldots, \underline{h}_{d-1})$ one may refer to Appendix of Hua and Wang (1987) for a table of $(\underline{h}_1, \ldots, \underline{h}_{d-1})$ for $1 < d \le 20$. For example, when $d = 4$ and $n = 11$, one can choose $h_1 = 1, h_2 = 5,\ h_3 = 7$, to generate $c_k,\ for\ k = 1, \ldots, 11$.

(ii) Let $\underline{a}_k = \underline{a}^\tau(\underline{c}_k)$ denoted by equation (22) be the point on S^{d-1} corresponding to $\underline{c}_k$.

For more detail explanation about this procedure one may refer to Wang and Fang (1990).

3.3. Simulation Experiment

We in this subsection perform some simulations to compare the performance of the power of the bootstrap test statistics W^*_{n1}, W^*_{n2}, W^*_{n11} and W^*_{n21}. The hypothesized null distribution is 4-dimensional normal distribution $N(\underline{\nu}, V_4)$. In the first case, $\underline{\nu}$ is unknown parameter and V_4 completely known. In the second case both $\underline{\nu}$ and V_4 are regarded as unknown parameters. To reveal the performance of the tests, we consider that the underlying distribution is a convolution of $N(\underline{\nu}, V_4)$ and the uniform distribution on [0,b]^4, $b = 0, 1, 2, 3, 4$. In the simulation, the sample of size 20 is generated from $N(\underline{\nu}, V_4)$ where $\underline{\nu} = (1, 1, 1, 1)^\tau$ and V_4 is the identity natrix. In the first simulation, let $\theta = \underline{\nu}$ where $\underline{\nu} = (1, 1, 1, 1)^\tau$ and for second case, let $\theta = (\underline{\nu}, V_4)$. The preassigned significant level was $\alpha = 0.05$. A set of 135 projection directions on S^3 were chosen by the number-theoretic method described above. To get the critical values, we run independently $c = 500$ times of bootstrap procedures. All reported values were based on 500 replications of the basic experiment. Table 1 below presents the empirical power of all four tests when $\underline{\theta} = \underline{\nu}$ and Table 2 for the case where $\underline{\theta} = (\underline{\nu}, V_4)$.

Table 1: Empirical power of the tests, $n = 20$, $\underline{\theta} = \underline{\nu}$

b	0	1	2	3	4
W^*_{n1}	0.044	0.374	0.848	0.976	1.000
W^*_{n11}	0.054	0.626	0.902	1.000	1.000
W^*_{n2}	0.044	0.528	0.944	1.000	1.000
W^*_{n21}	0.048	0.730	1.000	1.000	1.000

Table 2: Empirical power of the tests, $n = 20$, $\underline{\theta} = (\underline{\nu}, V_4)$

b	0	1	2	3	4
W^*_{n1}	0.048	0.334	0.774	0.938	1.000
W^*_{n11}	0.057	0.390	0.784	0.946	1.000
W^*_{n2}	0.044	0.426	0.936	0.992	1.000
W^*_{n21}	0.047	0.472	0.958	0.990	1.000

We see that under H_0 the actual percentages of times H_0 was rejected is close to the nominal level. The performance of power of W^*_{n11} (W^*_{n21}) is better than that of W^*_{n1} (W^*_{n2}) in the cases we conducted. On the other hand comparing the extreme value type statistic W^*_{n1} (W^*_{n11}) to the quadratic type statistic W^*_{n2} (W^*_{n21}), the later has better performance than the former.

4. The tail behavior of W_{n1} and W_{n2}

In this section we will discuss the tail behavior of W_{n1} and W_{n2}. We begin by viewing equations (8) and (9), and note that the accurate distribution of W_1 and W_2 can not be derived except for some special cases. Therefore, it is of our interest to investigate the tail probability bounds of W_{n1} and W_{n2}.

Suppose that $P_\theta(.) = P(. - \underline{\theta})$, where P(.) is a non-atomic, spherically symmetric distribution, and $\underline{\theta}$ is a d-dimensional unknown center of $P_\theta(.)$, which lies in an open convex set Ω in R^d. Define by $F_1(.)$ the marginal distribution of $P(.)$ at the projection direction $\underline{a}_0 = (1, 0, \ldots, 0)$. By the the spherical symmetry of P, we know that $\underline{a}^\tau(\underline{x} - \underline{\theta})$ has the distribution $F_1(.)$ for all projection directions $\underline{a}$, and $F_1(t)$ has the density function $f(t)$. Taking $\hat{\theta} = \frac{1}{n}\sum_{j=1}^n \underline{x}_j$, we have $A(\underline{a}, t) = \underline{a} f(t - \underline{a}^\tau \underline{\theta})$. Let $L(\underline{x}) = \underline{x} - \underline{\theta}$ and $V = cov(\underline{x}) = I_d$, the covariance matrix of $\underline{x}$. Let

$$W_1 = \sup_a \int \{W(\underline{a}, t) - f(t - \underline{a}^\tau \underline{\theta}) a^\tau \ \underline{z}\}^2 dF_1(t),$$

where $\underline{z}$ has the normal distribution $N(0, I_d)$.

We now present the following results which would be useful for our discussion in the rest of the paper.

Theorem 4.1. Assume that 1) $Cov(\underline{x}) < \infty$ and 2) $\int f^2(t) dF_1(t) < \infty$, where $f(\cdot)$ is the density function of $F(\cdot)$. Then for $\lambda > 0$,

$$\begin{aligned} P\{W_1 > \lambda\} \leq & \quad c\lambda^{2d-1/2} \exp(-\pi^2\lambda/4(e\pi^2 + 1)) \\ & +1 - \{\Phi(\{\pi^2\lambda/2(e\pi^2+1)\}^{1/2}) - \Phi(-\{\pi^2\lambda/2(e\pi^2+1)\}^2)\}^d, \end{aligned} \tag{23}$$

and

$$\begin{aligned} P\{W_2 > \lambda\} \leq \ & c\lambda^{-\frac{1}{2}} \log^+ \lambda \exp\{-(\pi^2(d+1)(d+3)\lambda)/(2(3e\pi^2 + (d+1)(d+3)))\} \\ & + P\{\chi_d^2 > (\pi^2(d+1)(d+3)\lambda/(2(3e\pi^2 + (d+1)(d+3)))\}, \end{aligned} \tag{24}$$

where $e = \int f^2 dF_1$ and $\log^+ \lambda = \max\{\underline{1}, \log \lambda\}$.

For the case that $P_\theta(.)$ is elliptically symmetric distribution with the density function $f_1(\underline{x}^\tau \ \underline{\Sigma} \ \underline{x})$, where $\underline{\Sigma}$ is is a positive definite $(d \times d)$-matrix. It is well known that $\underline{\Sigma}^{-1/2}\underline{x}$ has a spherical distribution. When $\underline{\Sigma}$ is unknown, one need to estimate it so as to construct the statistic as W_{n1} and W_{n2}, in (2) and (3), respectively. Let $F_1(.)$ be the marginal distribution of $\underline{\Sigma}^{-1/2}\underline{x}$ at the direction $\underline{a} = (1, 0, \ldots, 0)$. We have by the spherical symmetry $P_\theta I(\underline{a}^\tau \underline{x} \leq t) = F_1(t/\sqrt{\underline{a}^\tau \underline{\sum} a})$. For each $\underline{a} \in S^{d-1}$ let $\underline{a}^\tau \hat{\underline{\Sigma}} \underline{a} = n^{-1}\sum_{j=1}^n (\underline{a}^\tau \underline{x}_j)^2$. If $E||\underline{x}||^4 < \infty$, it is easy to see that $\{\sqrt{n}(\underline{a}^\tau \hat{\underline{\sum}} \underline{a} - \underline{a}^\tau \underline{\sum} \underline{a}) : \underline{a} \in S^{d-1}\}$ converges weakly to a centered, continuous Gaussian process $\{V(\underline{a}) : \underline{a} \in S^{d-1}\}$, with covariance function $R(\underline{a}, \underline{a}_1) = E(\underline{a}^\tau \underline{x} \ \underline{x}^\tau \underline{a}_1)^2 - (\underline{a}^\tau \sum \underline{a} \ \underline{a}^\tau \sum \underline{a}_1$. Along with the similar argument used in section 2, let $\underline{b}^\tau = \underline{a}^\tau \underline{\sum}^{1/2}/\sqrt{\underline{a}^\tau \underline{\sum} \underline{a}}$, $t_1 = t/\sqrt{\underline{a}^\tau \underline{\sum} \underline{a}}$, $\underline{x}' = \underline{\sum}^{-1/2}\underline{x}$, and let P_n' be the empirical measure based on $\{\underline{x'}_1, \ldots \underline{x'}_n\}$, then when $L_a(\underline{x}) = (\underline{a}^\tau \underline{x})^2 - E(\underline{a}^\tau \underline{x})^2$, we have

$$\begin{aligned} W_{n1} \quad = \quad & n \sup_{a \in S^{d-1}} \int \{P_n \ I(\underline{a}^\tau \underline{x} \leq t) - P_\theta I(\underline{a}^\tau \underline{x} \leq t) \\ & -f\left(t/\sqrt{\underline{a}^\tau \underline{\sum} \ \underline{a}}\right)\left(t/2\sqrt{\underline{a}^\tau \underline{\sum} \ \underline{a}}\right) \frac{1}{n}\sum_{j=1}^n \frac{\underline{a}^\tau \underline{\sum}^{1/2}}{\sqrt{\underline{a}^\tau \underline{\sum} \underline{a}}} \times \\ & \left[\underline{\sum}^{-1/2} \underline{x}_j \underline{x}_j^\tau \underline{\sum}^{-1/2} - E\left(\underline{\sum}^{-1/2} \underline{x} \ \underline{x}^\tau \underline{\sum}^{-1/2}\right)\right] \end{aligned}$$

$$\left.\frac{\underline{\sum}^{1/2}\underline{a}}{\sqrt{a^\tau \underline{\sum}\ \underline{a}}}\right\}^2 dP_n I(\underline{a}^\tau \underline{x} \le t)$$

$$= n \sup_{b \in S^{d-1}} \int \{(P'_n I(\underline{b}^\tau \underline{x}' \le t_1) - F_1(t_1)$$

$$-f(t_1)\frac{t_1}{2}\frac{1}{n}\sum_{j=1}^{n}[(\underline{b}^\tau \underline{x}'_j)^2 - E(\underline{b}^\tau \underline{x}')^2]\}^2 dP'_n I(\underline{b}^\tau \underline{x}' \le t_1)$$

$$\Rightarrow W_1 = \sup_{b\in S^{d-1}} \int \{W(\underline{b}, t_1) - f(t_1)t_1 \underline{b}^\tau \underline{V}\ \underline{b}\}^2 dF_1(t_1) \tag{25}$$

where $\underline{V}$ is the limit normal random matrix of $\frac{1}{\sqrt{n}}\sum_{j=1}^n\{\underline{x}_j\underline{x}_j^\tau - E\underline{x}\ \underline{x}^\tau\}$. W_2 also has the similar version. Hence we obtain the following result.

Theorem 4.2. Assume that $E||\underline{x}||^4 < \infty$, and that

$$I \triangleq \int f^2(t)t^2 dF_1(t) < \infty, \tag{26}$$

then for any $\lambda > 0$

$$\begin{aligned} P\{W_2 > \lambda\} \le P\{W_1 > \lambda\} &\le c\lambda^{2d-1/2}\exp(-\pi^2\lambda/4(I\pi^2+1)) \\ &+ \sum_{1\le i,k\le d} P\{|V_{ik}| > \pi^2\lambda/2(I\pi^2+1)\} \end{aligned} \tag{27}$$

where V_{ik} has the normal distribution $N(0, cov(x_i x_k))$ $1 \le i, k \le d$ and x_i's are the components of $\underline{x}$.

Furthermore, if $P_\theta(.)$ is a nonatomic, elliptically symmetric distribution with unknown location and dispersion matrix, it is easy to see that

$$W_1 = \sup_a \int \{W(\underline{a}, t) - f(t)\underline{a}^\tau \underline{Z} + f(t)\frac{t}{2}\underline{a}^\tau\ \underline{V}\ \underline{a}\}^2 dF_1(t) \tag{28}$$

$$W_2 = \int\int \{W(\underline{a}, t) - f(t)(\underline{a}^\tau \underline{Z} + \frac{t}{2}\underline{a}^\tau \underline{V}\ \underline{a})\}^2 dF_1(t)d\mu(\underline{a}). \tag{29}$$

Combining Theorem 4.1 with Theorem 4.2, we can derive the analogous tail probability bound for W_1 and W_2.

Before proving the proof of the theorems, we cite an important result which is useful in our proofs below.

Lemma 4.1. (Cheng and Zhu (1992)). Assume that P is a d-dimensional nonatomic spherically symmetric distribution. Then for all $\lambda > 0$,

$$P\left\{\sup_{a\in S^{d-1}} \int W^2(\underline{a}, t)dF_1(t) > \lambda\right\} \le c\lambda^{2d-1/2}\exp(-\pi^2\lambda/2) \tag{30}$$

and

$$P\left\{\int\int W^2(\underline{a}, t)dF(t)d\mu(\underline{a}) > \lambda\right\} \le c\lambda^{-1/2}\log^+\lambda\exp(-\pi^2\lambda/2). \tag{31}$$

Proof of Theorem 4.1. Note that $\sup_a |\underline{a}^\tau \underline{Z}|^2 = \max_{1\le i\le d} |Z_i|^2$, where Z_i are the components of Z. Then

$$\begin{aligned}\sup_a \int (f(t)\underline{a}^\tau Z)^2 dF_1(t) &= \int f^2(t)dF_1(t) \sup_a |\underline{a}^\tau \underline{Z}|^2 \\ &= \int f^2(t)dF_1(t) \max_{1\le i\le d} |Z_i|^2.\end{aligned}$$

Combining the independence of Z_i's, we have

$$\begin{aligned}&P\left\{\sup_a \int \{W(\underline{a},t) - f(t)\underline{a}^\tau \underline{Z}\}^2 dF_1(t) > \lambda\right\} \\ &\le P\left\{\sup_a \int W^2(\underline{a},t)dF_1(t) > \lambda/(2(e\pi^2+1))\right\} \\ &+P\{\max_{1\le i\le d} |Z_i|^2 > \pi^2\lambda/(2(e\pi^2+1))\} \\ &\le c\lambda^{2d-1/2}\exp(-\pi^2\lambda/(4(e\pi^2+1))) + 1 - \left\{\Phi(\{\pi^2\lambda/(2(e\pi^2+1))\}^{1/2})\right. \\ &\left.-\Phi(-\{\pi^2\lambda/(4(e\pi^2+1))\}^{1/2})\right\}^d \end{aligned} \tag{32}$$

(23) is proved. To prove (24), we only need to see that when $\underline{Z}$ is given, $\underline{a}^\tau \underline{Z}/||\underline{Z}||$ has the distribution density function,

$$\frac{1}{B(1/2, \frac{d}{2})}(1-y^2)^{\frac{d}{2}-1}, \quad -1 \le y \le 1. \tag{33}$$

For ready reference, see Fang, et. als (1990, Theorem 2, equation 2.5) where $B(p,q)$ is a β-function, then it is easy to see that, together with the properties of β-function,

$$\begin{aligned}\int (\underline{a}^\tau \underline{Z}/||\underline{Z}||)^2 d\mu(a) &= \frac{1}{B(1/2, \frac{d}{2})}\int_0^1 y^{3/2}(1-y)^{\frac{d}{2}-1}dy \\ &= \frac{B(\frac{5}{2},\frac{d}{2})}{B(\frac{1}{2},\frac{d}{2})} = \frac{3}{d(d+2)}\end{aligned} \tag{34}$$

Then by combining the independence of Z_i's, we can get

$$\begin{aligned}&P\left\{\int\int f^2(t)(\underline{a}^\tau \underline{Z})^2 dF_1(t)d\mu(\underline{a}) > \lambda\right\} \\ &= P\left\{||\underline{Z}||^2 > d\lambda/e\} = P\{\chi_d^2 > \frac{d(d+2)\lambda}{3e}\right\},\end{aligned} \tag{35}$$

where χ_d^2 has χ^2-distribution with d degrees of freedom. In a similar way, as we did for proving (23), we can easily derive (24), completing the proof of Theorem 4.1.

Proof of Theorem 4.2. Note that Schwarz inequality implies that $|\bar{Z}_{ij}a_i a_j| \le d$, where a_i's are the components of $\underline{a}$. Hence $\max_{a\in s^{d-1}} |\underline{a}^\tau \underline{V}\, \underline{a}| \le \max_{1\le i,k\le d} |v_{ik}|d$, where v_{ik} has the normal distribution $N(0, cov(x_i, x_k))$. Then it is easy to show (27) by an analog of (31).

5. Concluding Remarks

In this article we considered the testing problem associated with a multivariate distribution which is a member of a known parametric distribution family. We estimated Crámer-Von mises test statistics using projection pursuit technique. We investigated the limiting properties of our test statistics using bootstrapping approximation and Number Theoretic Method. We noted that the accurate limiting distribution of our test statistics are not tractable so, we presented the tail behaviors of our estimated test given that $P_\theta(.)$ is nonatomic and elliptically symmetric distribution matrix Σ. We conclude that our testing procedure work better than the existing techniques in the literature.

REFERENCES

[1] Barnard, G. (1963). Discussion of " The spectral analysis of point processes" by Bartlett, M. S. *Journal of the Royal Statistical Society* Ser. B, **25**, 294.

[2] Beran, R. and Miller, P. W. (1986). Confidence sets for a multivariate distribution. *Ann. Statist.*, **14**, 432-443.

[3] Bhatti, M. I. (1994). *Testing Regression Models Based on Sample Survey Data*, Aldershots Avebury, U. K. (forthcoming).

[4] Cheng, P. and Zhu, L. X. (1992). The tail probability inequalities for the PP type Crámer-Von Mises statistics. *Nankai Series: Probab. and Statist. and Statist.*, (Z. P. Jiang et. al eds 46-55, World-Scientific, Singapore.

[5] Dudley, R. M. (1978). Central limit theorems for empirical measure.*Ann. Probab.*, **6**, 899-929.

[6] Eastwood, V. R. (1993). Some nonparametric methods for change point problems. *The Canadian Journal of Statistics*, **21**, 209-222.

[7] Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distribution*, Chapman and Hall, London/New York.

[8] Gine, E. and Zinn, J. (1990). Bootstrapping general empirical measures, *Ann. Probab.*, **18**, 851-869.

[9] Hall, P. and Titterington, D. M. (1989). The effect of simulation order on level accuracy and power of Monte Carlo tests *Journal of the Royal Statistical Society* Ser. B, **51**, 459-467.

[10] Hua, L. G. and Wang, Y. (1987). *The Applications of Number Theory To approximate Analysis.* Science Press. Beijing

[11] Huber, P. (1985). Projection pursuit (with discussion). *Ann. Statist.*, **13**, 435-535.

[12] Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer, New York.

[13] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*, John Wiley & Sons. New York.

[14] Wang, Y. and Fang, K. T. (1990). Number theoretic methods in applied statistics. *Chinese Ann. Math.*, **11**(B), 41-55.

[15] Zhu, L. X., Fang, K. I., Bhatti, M. I. and Bentler, P. M. (1995). A Testing Approach Based on Projection Pursuit for Sphericity, *Pakistanian Journal of Statistics*, **11**, 49-65.

Institute of Applied Mathematics
Chinese Academy of Sciences
Beijing, China
and
Institute of Statistics
Catholic University of Louvain
Louvain-la-Neuve, Belgium

Department of Mathematics
Hong Kong Baptist University
Kowloon Tong, Hong Kong
and
Institute of Applied Mathematics
Chinese Academy of Sciences
Beijing, China

Department of Statistics & O. R.,
King Saud University, Saudi Arabia, and
Griffith University, Australia