Statistica Sinica 3(1993), 35-51

SOME RECENT DEVELOPMENTS IN PROJECTION PURSUIT IN CHINA

Guo-Ying Li and Ping Cheng

Academia Sinica

Abstract: Projection pursuit (PP) is a class of methods for exploratory and confirmative analysis of high-dimensional data sets. In this paper, a brief introduction to PP is given. Recent contributions of Chinese statisticians in PP classification, tests, estimation and tail behavior are presented. Some other developments and applications are also discussed.

Key words and phrases: Projection pursuit, classification, goodness-of-fit, location, dispersion matrix, principal components, asymptotics, tail probability.

1. Introduction

Our interest in projection pursuit (PP) originated in the early 1980's when G. Li and Z. Chen were visiting Harvard University and working with Professor Peter Huber. Later, in 1984, Professor Ping Cheng visited University of Wisconsin and University of Manitoba. In preparing a discussion paper he obtained a copy of the preprints of Huber (1985) and the associated discussion papers. With these papers, we realized that PP was at a stage of development where practical experience and extension of its usage were needed, and where, more critically, some theoretical understanding of it to remedy the "gap developing between practice and theory (Miller (1985))" is especially required. Thus, starting from 1985, a long term seminar on PP and related topics was conducted by P. Cheng and G. Li at the Institute of Systems Science, Academia Sinica. Our research activities were extended to include compiling lecture notes and giving short courses throughout of China. More statisticians and students were attracted to this area. In the past seven years, about 10 statisticians and 20 students have been involved in research and applications of PP. So far, about forty papers have been completed, of which more than thirty have been published. Most of the papers are concerned with the theory and methodology in PP classification, tests, estimation and tail behavior of PP statistics; only a few are in applications. This paper summarizes the major results of these papers.

A brief introduction to PP is given in Section 2. Sections 3-7 are concentrated on the contributions of Chinese statisticians.

2. Projection Pursuit

Projection pursuit is a class of methods dealing with high-dimensional data analysis. It serves both exploratory and confirmative analyses. The basic idea is to project data points onto low- (one- or two-, mostly,) dimensional subspaces and to study the original data set by searching for interesting projections.

Analyzing high-dimensional data can be very difficult because (1) it is impossible to draw a visible scatter plot or other pictures of the data, which will be very helpful for exploratory purposes; (2) data points are extremely sparse in the high-dimensional space so that kernel smoothers and similar techniques do not even work; (3) some optimal methods in the low-dimensional case behave poorly when the dimension is high. For example, M-estimators for location and scale parameters can reach the largest breakdown point $1/2$ in the one-dimensional case, but in the p-dimensional case, all M-estimators have breakdown point less than $1/p$; the sample mean is admissible for the normal mean when the dimension is lower than three, but not so if the dimension is higher. PP is able to avoid this "curse of dimensionality" since it is actually working with the low-dimensional projections. Also, quite often there exist some irrelevant variables when the dimension of a data set is moderate or high.

PP is categorized by manual PP and automatic PP. The former is a graphical display system, which provides pictures of any two-dimensional projections of a data set. The latter is a class of statistical techniques which choose a projection index and obtain the interesting projections by maximizing the index successively. Manual PP becomes more and more time-consuming as the dimension increases. As Huber (1985) pointed out: "an exhaustive visual search is out of the question if d (dimension) exceeds 4", so "we need an automated procedure that ferrets out projections likely to be of interest to the data analyst". This paper discusses only automatic PP.

An index is a measure of the interestingness of projections. But what is meant by "interestingness"? This is a practical and important question, but it is difficult to answer. According to our understanding, in exploratory analysis an interesting projection should be able to show the structure/feature of the original data set; and for confirmative analysis, it should be helpful in making inference. The most important sturctures/features, which people frequently want to reveal, are, perhaps, groups (usually clusters), the relationship among variables for complicated data sets, and variation and shape for simple data sets. We need different types of indices to capture these structures. "Interestingness" varies according to the purposes of analyzing data sets. Mathematically, for example, an index of one-dimension-projection is a map Q from R^n to R^1. Once the index Q is determined, PP searches for the most interesting direction, say a_*, that

maximizes $Q(a^\tau X_1, \ldots, a^\tau X_n)$, where $X_1, \ldots, X_n$ are R^d-valued observations and a^τ is the transpose of a unit vector a. This unit vector a_* and the maximum value $Q(a_*^\tau X_1, \ldots, a_*^\tau X_n)$ may all be called the PP statistics. Frequently, not only one but several interesting projections are selected for inspection. The above indices are a "practical version". An abstract version index of one-dimensional projection is a functional from one-dimensional distribution space to the real number space. When an empirical distribution is used as the argument of this functional, the abstract version index is a practical version.

PP was first proposed in the early seventies. Friedman and Tukey (1974) successfully implemented this technique and coined its catchy name. Huber (1985) gave an excellent survey of PP that, for the first time, put the fascinating problems and ramifications of PP into a coherent perspective. Cheng and Li (1986), and Cheng, Li et al. (1986) also gave in-depth discussions.

3. Exploratory Projection Pursuit

Exploratory PP includes, up to now, methods for classification, regression and density estimation. Little progress has been made in the area of density estimation recently. We shall concentrate our discussion on classification and also include the regression problem.

PP actually started from multivariate classification. The pioneer work in this area was done by Kruskal (1969, 1972), Switzer (1970), Switzer and Wright (1971), Friedman and Tukey (1974) (cf. Huber (1985)). Research on PP classification basically proposes indices to measure the interestingness that serves certain purposes, and develops the corresponding algorithms, as the pioneers did. Huber (1985) heuristically pointed out that interestingness for classification goes together with non-normality, and listed three examples of projection indices that are all essentially statistic for testing normality. One of them is the standardized negative Shannon entropy. Jones and Sibson (1987) realized that if this entropy index is used, the computation is very intensive. They derived an approximation to the entropy index by expressing the probability density as a truncated Gram-Charlier expansion based on Hermite orthogonal polynomials, and obtained a moment index, which is a simple function of the third and fourth moments. Friedman (1987) also presented a PP algorithm for clustering. After centering and sphering the data, he performed a transformation on each projection

$$R(a) = 2\Phi(a^\tau X) - 1$$

where $\Phi(\cdot)$ is the standard normal cdf. With $p_a(r)$ the probability density of $R(a)$, $\int_{-1}^{1}(p_a(r) - \frac{1}{2})^2 dr$ measures the departure of $a^\tau X$ from normality. Expanding $p_a(r)$ in Legendre polynomials and truncating the resultant sum of the above

integral, Friedman obtained a projection index

$$I(a) = \sum_{j=1}^{J} \frac{2j+1}{2} E^2[p_j(R(a))], \tag{3.1}$$

where E denotes expectation and p_j $(j = 0, 1, \ldots)$ are Legendre polynomials on $(-1, 1)$. This is not really a new index, but is essentially Neyman's (1937) statistic for testing normality (cf. Section 4). Hall (1989b) discussed another index without any transformation, that is a moment approximation to

$$\int_{-\infty}^{\infty} \left[g^a(u) - \phi(u)\right]^2 du \tag{3.2}$$

(g^a is the probability density of $a^\tau X$ and ϕ that of the standard normal) by truncating the expansion of g^a in Hermite polynomials. This index was also proposed independently by Li (1989). Following Friedman and Tukey's (1974) idea that a point cloud with groups tends to be locally dense and globally spread out, Li added a measure of dispersion to Friedman's (1987) and Hall's (1989b) indices. Both their indices basically describe only the local density of the data sets.

The advantage of these indices is that they are more rapidly computable than the previous ones. On the other hand, they are all within the framework of testing normality. Why normality is one of the least interesting distributions for classification may be due to its simple structure with only one group. The uniform distribution is also one of the least interesting distributions. A data set drawn from a uniform distribution may be considered in which every data point is a group. This is another extreme case of the classification problem. Based on this idea, Zhang (1990a) proposed a new index that is the product of the trimmed variance and the mode of the probability density, i.e., following the notation in (3.2), $I(a) = \sigma_\alpha^2(a^\tau X) \sup_t g^a(t)$. He also discussed the asymptotic properties of the optimum direction and the mode in this direction. Suppose f_n^a is a kernel estimator of g^a and $g^{a_1}(t_1) = \sup_a \sup_t g^a(t)$, $f^{a_n}(t_n) = \sup_a \sup_t f_n^a(t)$; then the bivariable sequence $\sqrt{nh_n^3}((t_n - t_1), (a_n - a_1))$ is asymptotically multivariate normal under proper conditions, with h_n the bandwidth in kernel estimation.

The progress on PP regression is on its theoretical aspects. Hall (1989a) investigated the consistency property of the kernel-based PP regression estimator for the first projective approximation to the regression function. He showed, under suitable assumptions, that if the orientation estimate θ_n is sufficiently close to the "true" projective direction θ_0, that is

$$|\theta_n - \theta_0| \leq (nh)^{-1/2} n^\varepsilon \quad \text{for any fixed } \varepsilon < 1/[2(2r+1)] \tag{3.3}$$

($|\alpha|$ stands for the absolute value of a number α or the length of a vector α, h is the window size and r the order of the kernel function), then $|\theta_n - \theta_0| = O((nh)^{-1/2})$ a.s. and the curve estimate $\hat{g}_{\theta_n}(\theta_n^\tau x)$ of the first projection approximation $g_{\theta_0}(\theta_0^\tau x)$ satisfies $|\hat{g}_{\theta_n}(\theta_n^\tau x) - g_{\theta_0}(\theta_0^\tau x)| = O_p((nh)^{-1/2})$. Zhu and Fang (1992) further studied this problem and proved, under proper assumptions, that the estimate θ_n of the first optimum direction satisfies condition (3.3) needed by Hall's (1989a) argument.

4. Tests

Huber (1985) pointed out that PP emerged as the most powerful method yet invented to lift one-dimensional techniques to higher dimensions. This is especially true for statistical testing procedures. As a matter of fact, almost at the same time as PP was suggested, Malkovich and Afifi (1973) used exactly the same idea to construct two multivariate normality tests based on one-dimensional kurtosis and skewness. After the advent of PP, Beran and Miller (1986) discussed confidence sets for multivariate distributions, which are actually, in our terminology, the critical values of PP Kolmogorov-Smirnov tests for multivariate goodness-of-fit. They proved the feasibility of bootstrap construction for critical values.

We also start with goodness-of-fit problem. Suppose $X_1, \ldots, X_n$ are d-vectors which are iid with common cdf F. Let F_n be the empirical distribution of $X_1, \ldots, X_n$, P and P_n be the corresponding probability measures of F and F_n respectively. Denote the cdf of $a^\tau X$ by F^a and the associated probability measure by P^a. Similarly, define F_n^a and P_n^a. Zhang (1988) first built up a PP χ^2 test,

$$Z_n = \sup_{|a|=1} \sum_{i=1}^{m} \frac{n[P_n^a(S_i) - P^a(S_i)]^2}{P^a(S_i)}$$

with $\{S_1, \ldots, S_m\}$ a partition of the real number space R^1. He obtained the asymptotic distributions of Z_n for the null hypothesis completely known and with unknown parameters respectively. Then Cai (1991) constructed a test W_n based on the Cramér-Von Mises-Smirnov test,

$$W_n = \sup_{|a|=1} \int \left(F_n^a(t) - F^a(t)\right)^2 dF^a(t).$$

Li and Zha (1991) presented a PP Neyman test

$$K_n = \sup_{|a|=1} \sum_{j=1}^{m} \frac{1}{n}\left[\sum_{i=1}^{n} \pi_j\left(F^a(a^\tau X_i)\right)\right]^2,$$

where $\pi_0 = 1$, $\pi_1, \ldots, \pi_m$ are orthogonal polynomials on (0,1). Each of these two papers derived the asymptotic distribution of its PP test statistic, constructed two kinds of bootstrap approximations, and proved the consistency of the bootstrap procedures, i.e., the bootstrap statistics all have the same asymptotic distributions as their original PP statistic. If we write down the sample version of Friedman's (1987) projection index (3.1), that is,

$$I_n(a) = \sum_{j=1}^{J} \frac{2j+1}{2n^2} \Big[\sum_{i=1}^{n} p_j \big(2\Phi(a^\tau X_i) - 1 \big) \Big]^2,$$

then we see that this index is essentially the same as that of the PP Neyman test except that the F^a of a general F is replaced by the standard normal Φ and the orthogonal polynomials π_j on (0,1) replaced by the same kind of polynomials p_j on $(-1, 1)$. Sun (1989) discussed the p-value of this index and obtained some elegant results.

Practically, some data analysts consider a random p-vector to be normal if its p marginal distributions are all normal. To clarify this usage, Cui (1990) gave counterexamples to show that a p-vector is not necessarily normal even if its N marginal distributions of projections $a_1^\tau X, \ldots, a_N^\tau X$ are all normal for any given finite N and N directions $a_1, \ldots, a_N$.

Another type of tests being discussed is the location problem. Zhang (1989) proposed two location tests based on the Mann-Whitney test and a robust t-type test, and derived the asymptotic distributions of these two PP tests, respectively, under null hypotheses.

As we have seen, the above PP tests are mostly for goodness-of-fit, and otherwise for the location problem. Also, their asymptotic properties are studied individually. Can we establish some methods for other types of testing problems? Is it possible to study a class of PP statistics as a whole, like U-statistics and so on, in the one-dimensional case?

Note that:

(1) for two d-vectors m and m_0,

$$m = m_0 \Longleftrightarrow a^\tau m = a^\tau m_0 \text{ for all unit vectors } a;$$

(2) for $k \times k$ matrices V and V_0,

$$V = V_0 \Longleftrightarrow a^\tau V a = a^\tau V_0 a \text{ for all unit vectors } a;$$

(3) for random vectors X and Y,

$$X \stackrel{d}{=} Y \Longleftrightarrow a^\tau X \stackrel{d}{=} a^\tau Y \text{ for all unit vectors } a;$$

(4) for random s-vector X and t-vector Y,

$$X \text{ and } Y \text{ are independent} \iff a^\tau X \text{ and } b^\tau Y \text{ are independent for all unit vectors } a \text{ and } b;$$

where "$\stackrel{d}{=}$" means that both sides have the same distribution. The PP goodness-of-fit tests mentioned above are all built on the fact (3). Aware of the above foundation underlying PP tests based on one-dimensional procedures, Li and her two students studied PP L-, R- and U-statistics for tests as a whole (see Shi and Li (1992, 1991), Shi (1991), Tang and Li (1992)). They took centralized L-, R- and U-statistics as projection indices, and built up the corresponding PP statistics for the tests. The asymptotic distributions of each class of PP test statistics are derived. As special cases and applications of the general results, examples for one and/or two sample location and dispersion tests, and for testing independence of two random vectors are given. However, these asymptotic distributions depend on the unknown underlying populations, so Shi and Li (1991) discussed bootstrap approximations for PP L-statistics and reported some numerical results of simulation studies.

In the above papers, the asymptotic levels of the tests are obtained. What about the powers of these tests? Zhang and Cheng (1989) gave a general result from which the asymptotic power of most of the above PP tests can be derived.

Let Π be a set with a distance $d(\cdot,\cdot)$ on it. Let $V(t)$ be a continuous real-valued function on Π and $\{V_n(t) : t \in \Pi\}$ be a stochastic process. Put

$$S_n(t) = \sqrt{n}(V_n(t) - V(t)), \quad B = \left\{ s : V(s) = \sup_{\Pi} V(t) \right\}.$$

Theorem 4.1. *Assume that B is not empty and that there is a stochastic process, say $S = \{S(t) : t \in \Pi\}$, with uniformly continuous and bounded paths such that*

$$\sup_{\Pi} |S_n(t) - S(t)| \longrightarrow 0 \quad \text{a.s.} \tag{4.1}$$

Then

$$\sqrt{n}\left(\sup_{\Pi} V_n(t) - \sup_{\Pi} V(t) \right) \longrightarrow \sup_{B} S(t) \quad \text{a.s.}$$

To explain how to apply this theorem, let us look at the location problem. For simplicity, we assume the underlying covariance matrix is known. The null hypothesis is $m = m_0$, and the alternative is $m \neq m_0$. It is evident that the statistic

$$T_n = \sup_{a} \left| a^\tau \bar{X}_n - a^\tau m_0 \right|$$

provides a test, where $\bar{X}_n$ is the sample mean of $X_1, \ldots, X_n$ which are iid with common probability measure P. Let $Pf = \int f dP$. If the null hypothesis is true, then by simple argumentation we have

$$T_n = \sup_a \left|\sqrt{n}(P_n - P)(a^\tau x)\right| \xrightarrow{d} \sup_a |G(a)|,$$

where $\xrightarrow{d}$ stands for convergence in distribution and $G(a)$ is a Gaussian process. Now assume $m \neq m_0$. Put

$$V_n(a) = \left|a^\tau \bar{X}_n - a^\tau m_0\right| = \left|P_n(a^\tau x) - a^\tau m_0\right|,$$
$$V(a) = \left|P(a^\tau x) - a^\tau m_0\right|, \quad B = \left\{b : V(b) = \sup_a V(a)\right\}.$$

Note that now $(P_n(a^\tau x) - a^\tau m_0)$ is no longer centralized. The asymptotic distribution of $T_n = \sup_a V_n(a)$ is not obtainable by the usual theory of empirical process. But it can be derived by Theorem 4.1. Since for sufficiently large n,

$$S_n(a) = \sqrt{n}\Big(V_n(a) - V(a)\Big) = \sqrt{n}\Big(P_n(a^\tau x) - P(a^\tau x)\Big) sign\Big(P(a^\tau x) - a^\tau m_0\Big) \text{ a.s.}$$

we have

$$S_n \hat{=} \{S_n(a) : |a| = 1\} \xrightarrow{d} \Big\{G(a) sign(P(a^\tau x) - a^\tau m_0) : |a| = 1\Big\}$$
$$\hat{=} S = \{S(a) : |a| = 1\}.$$

By the Representation Theorem (Pollard (1984, p.71)), there exist $\bar{S}_n = \{\bar{S}_n(a) : |a| = 1\}$ and $\bar{S} = \{\bar{S}(a) : |a| = 1\}$ such that

$$\bar{S}_n \stackrel{d}{=} S_n, \quad \bar{S} \stackrel{d}{=} S,$$

$$\sup_{|a|=1} \left|\bar{S}_n(a) - \bar{S}(a)\right| \longrightarrow 0. \text{ a.s.}$$

Put $\bar{V}_n(a) = \bar{S}_n(a)/\sqrt{n} + V(a)$. Then, from **Theorem 4.1**, it follows that

$$\sqrt{n}\Big(\sup_a \bar{V}_n(a) - \sup_a V(a)\Big) \longrightarrow \sup_B \bar{S}(t) \text{ a.s.}$$

This yields

$$\sqrt{n}\Big(T_n - \sup_a V(a)\Big) = \sqrt{n}\Big(\sup_a V_n(a) - \sup_a V(a)\Big) \xrightarrow{d} \sup_B S(a).$$

5. Estimation

It was once a difficult problem to construct estimators for multivariate dispersion matrices that have most of the important properties in the light of equivariance, accuracy, breakdown point, positive definiteness and so on. The classical sample covariance matrix is affinely equivariant, positive definite and asymptotically normal; but its performance is not stable, and it has a very poor breakdown point 0. While existing robust estimators either have no equivariance, or have a low breakdown point; some of them have no asymptotic theory to back it up; some do not even guarantee a positive definite matrix.

Take advantage of PP, Li and Chen (1985) and Donoho (1982) proposed, respectively, two types of estimators that possess most of the desirable properties. A remarkable fact is that both types of estimators have high breakdown point, and certain equivariance as well.

Donoho showed that his PP-type estimators for multivariate location and dispersion are affinely equivariant and have sample-breakdown point close to $\frac{1}{2}$. Later, Li (1987) proved theoretically that Donoho's estimators are strongly consistent and qualitatively robust. Zhang (1987) verified that they are also asymptotically normal. A serious drawback is that they are computationally very expensive since maximization over the unit sphere is necessary for every data point of the observations.

Li and Chen (1985) presented another kind of PP procedures. Note that by its spectral decomposition, the sample covariance matrix can be built up by its principal components. The principal components are actually PP statistics with the sample variance as the projection index. Li and Chen took a robust scale as the projection index, firstly obtained robust principal components using the PP method, and then constructed a dispersion matrix based on these principal components. As shown in their theoretical and Monte-Carlo studies, this kind of robust estimators, namely robust PP estimators, for dispersion matrix and principal components are qualitatively robust, strongly consistent and rotationally equivariant. Their breakdown point can reach $\frac{1}{2}$ if the robust scale is properly selected. Most of the theory was given in Li (1984). The robust PP estimators for dispersion matrix and principal components together with their good properties were extended by Li (1986) to the case where the location is unknown and needs to be estimated simultaneously. One basic condition required in the above papers is that the underlying distribution belongs to an elliptic distribution family. Cui (1992) showed that this is not a necessary condition for the robust PP estimators of dispersion matrix and principal components being qualitatively robust. He proved, under suitable conditions, the necessary and sufficient condition for that is that the underlying distribution is of an orthogonal structure (see Cui (1992) for detail). Li (1986) and Cui (1992) also studied the convergence rate in probability of these robust PP estimators. Zhang (1990a) and Zhang, Zhu and

Cheng (1989) obtained, under proper conditions, the asymptotic distributions of the estimators for the dispersion matrices and the principal components when the eigenvalues of the underlying dispersion matrix are either all distinct or all identical, and also that for the largest and the smallest components in the general case. Zhang (1991) investigated the bootstrap approximation of these asymptotic distributions, and constructed confidence sets for the parameters.

6. Tail Behavior of Projection Pursuit Statistics

The asymptotic distributions of most PP statistics depend on the underlying distributions that are usually unknown. To apply these asymptotic results, we can either perform empirical bootstrap, or theoretically study their tail probabilities.

Typically, a PP statistic actually describes the largest difference between projections of the data and that of the (assumed) underlying distribution in a certain aspect. It may be regarded as a PP version and extension of Kolmogorov-Smirnov (K-S) distance. The tail probability of this type of statistics is an old and basic problem and has become quite attractive in recent years. Besides Cheng and his students in China, Alexander (1984), Adler and Brown (1986), Adler and Samorodnitsky (1987), Öhvrik (1987, 1988), Huber (1988), Sun (1989), etc. studied this kind of problem from different aspects. In the following we sketch some of their results. Let us follow the notation introduced in Section 4.

Assume $\mathcal{F}$ is a class of functions on R^d whose class of graphs has polynomial discrimination of degree v (cf. Pollard (1984, Ch.2)). Alexander (1984) proved that for $\lambda \geq 8$

$$Pr\left\{\sqrt{n}\sup_{\mathcal{F}}|P_nf - Pf| > \lambda\right\} \leq 16\lambda^{2^{12}v}\exp\{-2\lambda^2\}.$$

Zhu (1990, 1991a,b), Zhang and Cheng (1991), and Zhang (1990a, 1992) improved this inequality. For example, they showed that for any $\varepsilon \in (0, 1/4)$, if $\lambda \geq \lambda_0 = \lambda(v, \varepsilon)$, the factor $16\lambda^{2^{12}v}$ can be replaced by $\lambda^{4(v-1)(1+\varepsilon)/(1-2\varepsilon)}$. A more accurate result and other similar results are also given in their papers.

For PP K-S statistics, Öhvrik (1987, 1988) did many simulation experiments. He generated many samples of d-vectors from multivariate normal and other spherically symmetric distributions, which resulted in two empirical formulas,

$$\begin{aligned}
\xi(n,d,\lambda) &\doteq Pr\left\{\sqrt{n}\sup_{a,t}\left|F_n^a(t) - F^a(t)\right| \geq \lambda\right\} \\
&\approx 2\exp\left\{-2\lambda^2 + \frac{d-1}{\sqrt{\pi}}\ln(2en/d)\right\}, \\
\xi(n,d,\lambda) &\approx 2\exp\left\{-2\lambda^2 + 2.464(d-1)\right\}.
\end{aligned}$$

Huber (1988) studied this problem theoretically and improved his results of 1985. He verified, for $\lambda > d/\sqrt{n}$,

$$\xi(n,d,\lambda) \le 2\left(\frac{en}{d}\right)^d \exp\left\{-2(\lambda - d/\sqrt{n})^2\right\}$$

with the underlying distribution spherically symmetric. This is the best result so far for the finite sample case. Based on Öhvrik's simulation results, Huber made a conjecture that

$$\xi(n,d,\lambda) \le N \cdot 2\exp\{-2\lambda^2\}, \tag{6.1}$$

where $N = N(d)$ does not depend on λ and n.

Zhang (1990a,b, 1992), Zhang and Cheng (1991), Zhang, Zhu, and Cheng (1993) and Zhu (1990, 1991a,b) investigated the upper and lower bounds of $\xi(n,d,\lambda)$ for a wide range of underlying distributions. One conclusion from their results is that for $n \ge n_0 = n_0(\lambda)$,

$$c_1(d)\lambda^{2(d-1)}\exp\{-2\lambda^2\} \le \xi(n,d,\lambda) \le c_2(d)\lambda^{2(d-1)}\exp\{-2\lambda^2\}. \tag{6.2}$$

For fixed dimension d, the two sides have the same order $\lambda^{2(d-1)}\exp\{-2\lambda^2\}$. Hence, when n is large enough, (6.2) contradicts (6.1). The right side of (6.2), which improves Huber's (1988) inequality in large sample sense, holds not only for elliptically symmetric distributions, but also some symmetric and stable laws, etc. (Zhang (1990a, 1992)). The left side of (6.2), which leads to a contradiction of (6.1), holds, as well, for even a wider class of underlying distributions (cf. Zhang (1990a,b) and Zhu (1990)).

Cheng and Zhu (1992) discussed the upper bound for $\xi(n,d,\lambda)$ for the case that the underlying distribution is elliptically symmetric with unknown and estimated parameters. For example, when both location θ and dispersion matrix Σ are unknown, and estimated by $\hat{\theta}_n$ and $\hat{\Sigma}_n$ respectively, they obtained

$$\tilde{\xi}(n,d,\lambda) \le c\lambda^{2(d-1)+1}\exp\{-\lambda^2/2b^2\},$$

where $\tilde{\xi}(n,d,\lambda)$ is the same as $\xi(n,d,\lambda)$ but with θ and Σ replaced by $\hat{\theta}_n$ and $\hat{\Sigma}_n$ respectively in F, and $b^2 = b^2(F)$ is a constant.

Zhang and Cheng (1991) and Zhang, Zhu and Cheng (1993) studied PP K-S statistic with m-dimensional projections. Let

$$\mathcal{F}_m = \left\{I(Ax \le t) : A \text{ is an } m \times d \text{ matrix}, t \in R^m\right\},$$

$$\xi_m(n,d,\lambda) = Pr\left\{\sup_{\mathcal{F}_m}\left|\sqrt{n}(P_n f - Pf)\right| > \lambda\right\}.$$

They proved that, if P is elliptically symmetric with $P\{x = 0\} = 0$, then for any $\lambda \geq 2$ there exists an $n(\lambda)$ such that

$$\xi_m(n, d, \lambda) \leq c\lambda^{2(dm-1)} \exp\{-2\lambda^2\} \quad \text{for all} \quad n \geq n(\lambda).$$

They also discussed the upper and lower bounds of $\xi_m(n, d, \lambda)$ for general underlying P.

Combining the Von Mises statistic and the PP idea, Cheng and Zhu (1992) discussed two statistics for goodness of fit tests. One is

$$V_n = \int_{|a|=1} \int_{-\infty}^{\infty} \left[\sqrt{n}(F_n^a(t) - F^a(t))\right]^2 dF^a(t) dH(a),$$

where H is the uniform distribution on the d-dimensional unit sphere. They showed that, under mild conditions for $\lambda \geq 1$, there exists an $n(\lambda)$ such that

$$Pr\{V_n > \lambda\} \leq c\lambda^{-1/2}[\ln(\lambda)]^+ e^{-\lambda\pi^2/2},$$

where $[a]^+ = aI(a > 0)$. Another is

$$W_n = \sup_{|a|=1} \int_{-\infty}^{\infty} \left[\sqrt{n}(F_n^a(t) - F^a(t))\right]^2 dF^a(t).$$

If F is elliptically symmetric, then

$$Pr\{W_n > \lambda\} \leq c(d)\lambda^{2d-1/2} e^{-\lambda\pi^2/2}.$$

Zhang (1990a,b) studied the lower bounds for other K-S statistics. He extended the result of dimension $d = 2$ by Adler and Brown (1986) to general d, and proved that the lower bound in Adler and Samorodnitsky (1987) for $P = \text{Uniform}([0, 1]^d)$ remains true when P is any continuous distribution on R^d with proper conditions. In the latter, the lower bounds are of the form

$$C\lambda^{2(2d-1)} \exp\{-2\lambda^2\} \quad (\lambda > 0).$$

7. Applications

7.1. The life of a steel roller

The life of a steel roller is a very interesting practical problem raised by a steel plant in 1958, which was advertised for solution in a Chinese journal. More than thirty years had passed before Cheng, Zhu, Wei and Shi (1991) provided an answer quite recently through the PP theory. The problem is as follows. While making rolled steel, a very big metal ball is used. Each time of steel rolling, the

ball is put in the device randomly, and after rolling it shows a sign of wear and tear, which is approximately a big ring with a great circle of the sphere in the middle. If any point of the sphere falls in the ring m times, then the metal ball has to be scrapped. The question is how many times the ball can be used on average.

Without loss of generality, we assume the radius of the ball is one. Consider one rolling first. Let $2h$ be the thickness of the ring of wear and tear, and x a unit vector perpendicular to the great circle. A point a is in the ring if and only if $|a^\tau x| \leq h$. For n times of rolling , let $X_1, \ldots, X_n$ be the n unit vectors perpendicular to the n rings of wear and tear. The number of rings where a particular point a falls in is $\sum_{i=1}^{n} I(|a^\tau X_i| \leq h)$. Note that, without loss of generality, $X_1, \ldots, X_n$ are iid with a uniform distribution on the half ball $\{\alpha = (\alpha_1, \alpha_2, \alpha_3) : |\alpha| = 1, \alpha_3 \geq 0\}$. Denote the corresponding empirical measure by P_n. Then

$$\sum_{i=1}^{n} I\Big(|a^\tau X_i| \leq h\Big) = nP_n I\Big(|a^\tau x| \leq h\Big)$$

is actually a process indexed by the unit sphere. Evidently, the ball is scrapped if

$$\sup_{|a|=1} \Big(nP_n I(|a^\tau x| \leq h)\Big) \geq m. \tag{7.1}$$

The number of times that a ball can be used is

$$\tau = \tau(m, h) = \inf\left\{n : \sup_{|a|=1} \Big(nP_n I(|a^\tau x| \leq h)\Big) \geq m\right\}.$$

The left hand side of (7.1) is a typical PP statistic whose asymptotic distribution can be derived. Then the asymptotic distribution of τ follows.

Both theoretical results and numerical simulations are presented in Cheng et al. (1991). For example, one theorem says that as $m \to \infty$,

$$\lim(\tau(m, h)/m - 1/h) = 0 \quad \text{a.s.}$$

An empirical formula based on the simulation is

$$E\tau(m, h) \cong m/h - 2.6388\frac{\sqrt{m}}{h} + 13.7116.$$

7.2. Other applications

The principal component analysis (PCA) method is frequently applied to meteorology data sets, which usually contain a number of outliers. In this case,

classical PCA often fails to give reasonable results. Chang, Shi and Chen (1990) tried the robust PP PCA discussed in this paper (cf. Section 5) and analyzed a data set from the records of monthly mean precipitation for 160 stations in 35 years (1951-1985) in China. In order to compare this method with the classical one, they chose a relatively "clean" data set of 500 hPa monthly mean height from Jan. 1951 to Dec. 1960 and 29 spots in Asia, and created a "dirty" data set by deleting the data of year 1960 and adding a few outliers to it. The numerical results showed that for the clean data set, two methods gave basically the same results; but for the dirty data set, the results given by the robust PP method remained almost the same, while the classical method gave entirely different results.

Instead of the usual principal components in latent root regression, Yan (1990) used the robust PP principal components to construct the robust latent root regression (RLRR), and the robust principal component regression (RPCR). He applied these two regression methods to analyze a set of medical data. Additionally, he compared these two methods with the LS regression and the classical principal component regression (CPCR). His numerical experiments showed that the RLRR and RPCR were not only robust but also able to cope with colinearality while CPCR could only stand colinearality but was sensitive to outliers, and LS regression performed poorly when either colinearality or outliers occurred.

When an attack plane launches missiles to a target, it needs to choose a distance for good launching according to its state relative to the target. The relative state can be described by six variables (predictors). To build a model for the launching distance (response) based on observations, quite a few regression methods have been applied. However, none could achieve the required accuracy. Tian and Rong (1993) adopted the PP regression technique. They used the sum of ridge polynomials, $\sum_{i=1}^{m} g_i(a_i^T x)$, to approximate the regression function, where g_i $(i = l, \ldots, m)$ are polynomials. The numerical results showed that this method achieved the required accuracy.

Acknowledgments

This work was supported by the National Natural Science Foundation of China. The authors would like to thank the referees for their helpful comments and suggestions.

References

Adler, R. J. and Brown, L. D. (1986). Tail behaviour for suprema of empirical processes. *Ann. Probab.* **14**, 1–30.

Adler, R. J. and Samorodnitsky, G. (1987). Tail behaviour for the suprema of Gaussian processes with applications to empirical processes. *Ann. Probab.* **15**, 1339–1351.

Alexander, K. S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.* **12**, 1041–1067.

Beran, R. and Millar, P. W. (1986). Confidence sets for a multivariate distribution. *Ann. Statist.* **14**, 431–443.

Cai, Y. H. (1991). The goodness-of-fit test for a multivariate distribution by using PP and bootstrap method. *J. Sys. Sci. & Math. Sci.* **11**, 51–62. (in Chinese)

Chang, H., Shi, J. and Chen, Z. (1990). Projection pursuit principal component analysis and its application to meteorology. *Acta meteorologica Sinica* **4**, 254–263.

Cheng, P. and Li, G. (1986). Projection Pursuit — a kind of new statistical methods. *Chinese J. Appl. Probab. Statist.* **2**, 267–276. (in Chinese)

Cheng, P., Li, G., et al. (1986). *Lecture Notes on Projection Pursuit.* Institute of Systems Science, Academia Sinica, Beijing. (in Chinese)

Cheng, P. and Zhu, L. X. (1992). The tail probability inequalities for the PP type Cramér-Von Mises statistics. In: *Probability and Statistics, Proc. of Special Program at NanKai Institute, Tianjin, China, Aug. 1988 - May 1989* (Edited by Jiang, Yan, Cheng and Wu), 46–55. World Scientific, Singapore.

Cheng, P., Zhu, L. X., Wei, G. and Shi, P. D. (1991). On the life of sphere roller. *Acta Math. Sci.* **11**, 308–316.

Cui, H. J. (1990). Is p-dimensional population normal when distributions on its any N projection directions are normal? *Math. in Practice and Theory* **4**, 19–23. (in Chinese)

Cui, H. J. (1992). The sufficient and necessary condition for weakly consistent PP dispersion matrix being robust. *Chinese J. Appl. Probab. Statist.* **8**, 113–121. (in Chinese)

Donoho, D. L. (1982). Breakdown properties of multivariate location estimators. Ph.D. Qualifying Paper, Dept. of Statist., Harvard University.

Friedman, J. H. (1987). Exploratory projection pursuit. *J. Amer. Statist. Assoc.* **82**, 249–266.

Friedman, J. H. and Tukey, J. W. (1974). A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. Comput. C* **23**, 881–889.

Hall, P. (1989a). On projection pursuit regression. *Ann. Statist.* **17**, 573–585.

Hall, P. (1989b). On polynomial-based projection indices for exploratory projection pursuit. *Ann. Statist.* **17**, 589–605.

Huber, P. J. (1985). Projection pursuit. *Ann. Statist.* **13**, 435–475.

Huber, P. J. (1988). Spurious structure? Technical Report, Center for Intelligent Control Systems, MIT.

Jones, M. C. and Sibson, R. (1987). What is projection pursuit? *J. Roy. Statist. Soc. Ser.A* **150**, 1–36.

Kruskal, J. B. (1969). Toward a practical method which helps uncover the structure of a set of multivariate observations by finding the linear transformation which optimizes a new "index of condensation". In *Statistical Computation* (Edited by R. C. Milton and J. A. Nelder). Academic, New York.

Kruskal, J. B. (1972). Linear transformation of multivariate data to reveal clustering. In *Multidimensional Scaling: Theory and Application in the Behavioral Sciences, I, Theory.* Seminar Press, New York and London.

Li, G. (1984). Convergence of robust PP estimators for dispersion matrices and principal components. *J. Sys. Sci. & Math. Sci.* **4**, 1–14. (in Chinese)

Li, G. (1986). Convergence rate of projection pursuit estimators for dispersion matrices. *Acta Math. Appl. Sinica* **9**, 42–49. (in Chinese)

Li, G. (1987). Some properties of PP-type estimators for multivariate location and dispersion. *J. Sys. Sci. & Math. Sci.* **7**, 220–228.

Li, G. and Chen, Z. (1985). Projection-pursuit approach to robust dispersion matrices and principal components: Primary theory and Monte Carlo. *J. Amer. Statist. Assoc.* **80**, 759–766.

Li, G. and Zha, W. (1991). PP Neyman test for multivariate Goodness of fit. In: *Statistics in China* (Edited by X. Chen et al.), 261–274. Longman Academic, Scientific & Technical, London.

Li, Y. F. (1989). Exploring projection pursuit methods and their applications. Master Thesis, Dept. Appl. Math., Tianjin University. (in Chinese)

Malkovich, J. F. and Afifi, A. A. (1973). On tests for multivariate normality. *J. Amer. Statist. Assoc.* **68**, 176–179.

Miller, J. R. G. (1985). Discussion of "Projection pursuit" by P. J. Huber. *Ann. Statist.* **13**, 510–513.

Neyman, J. (1937). "Smooth test" for goodness of fit. *Skand. Aktuartidskr.* **20**, 149–155.

Öhvrik, J. (1987). Structure in noise. Technique Report, Department of Statistics, University of Stockholm.

Öhvrik, J. (1988). On the distribution of the Kolmogorov distance in the worst direction. Technique Report, Department of Statistics, University of Stockholm.

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer-Verlag, New York.

Shi, P. D. (1991). The asymptotic behavior and applications of a class of PP R-statistics. *J. Sys. Sci. & Math. Sci.* **4**, 84–96.

Shi, P. and Li, G. (1991). Bootstrapping PP L-statistics for tests. *J. Sys. Sci. & Math. Sci.* **4**, 158–172.

Shi, P. and Li, G. (1992). A class of PP L-statistics for tests. *Acta Math. Appl. Sinica* (English series) **8**, 27–44.

Sun, J. Y. (1989). The P-values of projection pursuit. Technical Report, No.104, Department of Statistics, Stanford University.

Switzer, P. (1970). Numerical classification. In *Geostatistics* (Edited by D. F. Merriam), 31–47. Plenum, New York.

Switzer, P. and Wright, R. M. (1971). Numerical classification applied to certain Jamaican eocene nummulitids. *Math. Geol.* **3**, 297–311.

Tang, X. and Li, G. (1992). PP U-statistics and their applications. In: *Probability and Statistics, Proc. of Special Program at NanKai Institute, Tianjin, China, Aug. 1988 - May 1989* (Edited by Jiang, Yan, Cheng and Wu), 209–227. World Scientific, Singapore.

Tian, Z. and Rong, H. W. (1993). Analyzing data of attack plane by projection pursuit regression. *Chinese J. Appl. Probab. Statist.* **9** (to appear, in Chinese).

Yan, G. Y. (1990). Robust latent root regression and principal component regression. Master Thesis, The Fourth Military Medical University, Xian, China. (in Chinese)

Zhang, H. (1987). The asymptotic normality of some multivariate estimators. Sino-American Statistical Meeting, contributed papers, 577–579, Aug. 31 – Sept. 2, Beijing, China.

Zhang, H. (1988). A PP goodness-of-fit test and its asymptotic properties. *J. Sys. Sci. & Math. Sci.* **8**, 234–242. (in Chinese)

Zhang, H. (1989). Some PP test statistics and their properties. *Acta Math. Appl. Sinica* **12**, 82–95. (in Chinese)

Zhang, J. (1990a). On the theory of projection pursuit. Ph.D. Thesis, Institute of Systems Science, Academia Sinica, Beijing, China. (in Chinese)

Zhang, J. (1990b). The lower bounds for the distributions of PP Kolmogorov-Smirnov statistics. *Chinese Bull. Sci.* **35**, 1444–1447. (in Chinese)

Zhang, J. (1991). Asymptotic theory of robust PP estimators for principal components and dispersion matrix (III): Bootstrap confidence sets, bootstrap tests. *J. Sys. Sci. & Math. Sci.* **4**, 290–301.

Zhang, J. (1992). The tail estimation for the distribution of the suprema of a P-bridge. *J. Sys. Sci. & Math. Sci.* **5**, 274–286.

Zhang, J. and Cheng, P. (1989). Asymptotic powers of some PP tests. *J. Sys. Sci. & Math. Sci.* **9**, 370–382. (in Chinese)

Zhang, J. and Cheng, P. (1991). The upper bounds for the distributions of PP Kolmogorov-Smirnov statistics. *Acta Mathematica Sinica* **34**, 388–401. (in Chinese)

Zhang, J., Zhu, L. X. and Cheng, P. (1989). Asymptotic theory of robust PP estimators for principal components and dispersion matrix (I), (II), submitted to *J. Multivariate Anal.*

Zhang, J., Zhu, L. X. and Cheng, P. (1993). Exponential bounds for the uniform deviation of a class of empirical processes. *J. Multivariate Anal.*, to appear.

Zhu, L. X. (1990). Probability estimation for the distribution of Kolmogorov distance in the worst direction. *Chinese Bull. Sci.* **35**, 1625–1627. (in Chinese)

Zhu, L. X. (1991a). Tail probability for the supreme of P-bridge with applications to empirical processes. *Chinese Bull. Sci.* **36**, 332–333. (in Chinese)

Zhu, L. X. (1991b). The tail probability for the distribution of Kolmogorov distance in the worst direction. In: *Chinese Math. into 21st Century* (Edited by W. Wu and M. Cheng), 284–296, Peking University Press.

Zhu, L. X. and Fang, K. T. (1992). Projection pursuit approximation for nonparametric regression. In: *Proc. of Order Statistics and Nonparametrics: Theory and Methods* (Edited by P. K. Sen and I. A. Salama), North Holland.

Institute of Systems Science, Academia Sinica, Beijing 100080.

(Received April 1991; accepted September 1992)