Ann. Inst. Statist. Math.
Vol. 41, No. 3, 461-475 (1989)

MULTIVARIATE SYMMETRY VIA PROJECTION PURSUIT

DAVID K. BLOUGH

Department of Agricultural Economics, University of Arizona, Tucson, AZ 85721, U.S.A.

(Received June 29, 1988; revised February 1, 1989)

Abstract. Blough (1985, *Ann. Inst. Statist. Math.*, **37**, 545-555) developed a multivariate location region for a random p-vector X. The dimension of this region provides information on the degree of symmetry possessed by the distribution of X. By considering all one-dimensional projections of X, it is possible to ascertain the dimension of the location region. Projection pursuit techniques can therefore be used to study symmetry in multivariate data sets. An example from an Entomology investigation is presented illustrating these methods.

Key words and phrases: Multivariate symmetry, location region, projection pursuit.

1. Introduction

Establishing the location of an asymmetric probability distribution is difficult due to the fact that there are many reasonable measures of location for such distributions. Doksum (1975) addresses this issue in the univariate case, and Blough (1985) extends those results to the multivariate case. In the latter, for a given p-variate distribution function F, a closed, convex location region in $\mathscr{R}^p$ (or some proper lower-dimensional subspace) is constructed, any point of which is a reasonable location parameter for F. Reasonable measure of location here refers to parameters which satisfy certain axioms of location (see Doksum (1975)).

As Doksum, Blough and more recently MacGillivray (1986) show, the size of the location region can be used to characterize a distribution in terms of symmetry. It is the purpose of this paper to develop a definition of multivariate symmetry and show how this relates to the structure of the location region. This is done in Section 2. Section 3 deals with the detection of symmetry in an underlying p-variate distribution based on a random sample from that distribution. This is accomplished by the technique of projection pursuit. An example from an Entomology investigation is given in Section 4.

2. Multivariate symmetry

We begin by defining order of symmetry for p-variate probability distributions. Let $\boldsymbol{X}$ be a random p-vector, F its corresponding distribution function. Also, let K_l be the $p \times p$ matrix given by

$$K_l = [k_{ij}]_l = \begin{cases} 0 & \text{if} \quad i \neq j, \\ 1 & \text{if} \quad i = j \neq l, \\ -1 & \text{if} \quad i = j = l. \end{cases}$$

DEFINITION 2.1. A random p-vector $\boldsymbol{X}$ (or equivalently F) is said to be symmetric of degree d if there exists a vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d, 0, \dots, 0)' \in \mathscr{R}^p$ and an orthogonal transformation T such that

$$T(\boldsymbol{X} - \boldsymbol{\theta}) \text{ has the same distribution as } K_1 K_2 \cdots K_d(T(\boldsymbol{X} - \boldsymbol{\theta}))\,.$$

Thus geometrically, symmetry of degree 1 means $\boldsymbol{X}$ is symmetric about a $(p-1)$-dimensional hyperplane. More generally, $\boldsymbol{X}$ symmetric of degree d means $\boldsymbol{X}$ is symmetric about d mutually orthogonal $(p-1)$-dimensional hyperplanes, and thus $\boldsymbol{X}$ is symmetric about their $(p-d)$-dimensional intersection. We note also that if $\boldsymbol{X}$ is symmetric of degree p, then $\boldsymbol{X}$ is symmetric about a point (namely, $\boldsymbol{\theta}$). In this case, $\boldsymbol{X}$ is termed centro-symmetric.

A brief review of the construction of the location region is necessary. Consider first the univariate case. Let F be the univariate distribution function of the random variable X. Doksum (1975) defines the following quantities: the location functional

$$m_F(u) = \frac{1}{2}[F^{-1}(u) + F^{-1}(1-u)]\,, \qquad 0 \le u \le \frac{1}{2}\,,$$

$$\underline{\theta}_F = \inf\left\{ m_F(u) \colon 0 \le u \le \frac{1}{2} \right\},$$

$$\bar{\theta}_F = \sup\left\{ m_F(u) \colon 0 \le u \le \frac{1}{2} \right\}.$$

The univariate location interval for F is then

$$I_F = [\underline{\theta}_F, \bar{\theta}_F]\,.$$

Notice that in the case that F is symmetric about a point θ, I_F reduces to just the single point θ. Now consider a bivariate random vector $\boldsymbol{X}$ with distribution function F. Let B_F denote the location region for $\boldsymbol{X}$. Blough (1985) constructs B_F in the bivariate case in the following manner. Let R_α denote a counterclockwise rotation of the plane through an angle α, and let F_α denote the distribution function of $R_\alpha \boldsymbol{X}$. Then B_{F_α} is a location rectangle obtained by taking the Cartesian product of the two univariate location intervals of the marginal distributions of F_α. That is,

$$B_{F_\alpha} = I_{F_{1\alpha}} \times I_{F_{2\alpha}} ,$$

with $F_{1\alpha}$ and $F_{2\alpha}$ being the first and second univariate marginal distribution functions of F_α, respectively. By rotating back to the original coordinate system and intersecting all such location rectangles, the location region results:

$$B_F = \bigcap_{\alpha \in (0, 2\pi]} R_{-\alpha}(B_{F_\alpha}) .$$

Thus, B_F is a closed convex set in the plane.

In higher dimensions, B_F can be constructed similarly; that is, by considering all rotated univariate marginal distributions. Let T be an orthogonal transformation of $\mathscr{R}^p$. Let F_T be the distribution function of $T\boldsymbol{X}$, where $\boldsymbol{X}$ is a p-variate random vector. B_{F_T} is a location hyperrectangle obtained by taking the Cartesian product of the p univariate location intervals of the marginal distributions of F_T. Thus,

$$B_{F_\alpha} = I_{F_{1T}} \times I_{F_{2T}} \times \cdots \times I_{F_{pT}} ,$$

with F_{iT} being the i-th univariate marginal distribution function of F_T. Then

$$B_F = \bigcap T^{-1}(B_{F_T}) ,$$

the intersection being over all orthogonal transformations T of $\mathscr{R}^p$. (That the construction of B_F in dimensions higher than two parallels the bivariate case is discussed in the Appendix.) Hence, B_F is a closed convex set in $\mathscr{R}^p$.

How does the structure of B_F relate to the degree of symmetry in $\boldsymbol{X}$? All information regarding the structure of B_F is found in its width function $W(\boldsymbol{a})$, where $\boldsymbol{a} \in \mathscr{R}^p$ is a unit vector (Blough (1985)). In particular, if $\boldsymbol{a}_0$ is a direction of symmetry, $W(\boldsymbol{a}_0) = 0$. This implies B_F is at most $(p-1)$-dimensional. Hence, if $\boldsymbol{X}$ is symmetric of degree d, it possesses d mutually orthogonal directions of symmetry, and so B_F is at most $(p-d)$-dimensional. Also then, if $\boldsymbol{X}$ is centrosymmetric, B_F is 0-dimensional; that is, B_F consists of only the point of symmetry ($\boldsymbol{\theta}$).

Theoretical and empirical determinations of the structure of B_F in relation to symmetry will now be developed.

3. Detection of symmetry in multivariate samples

The foregoing development enables one to characterize completely the degree of symmetry in a p-variate distribution. Theoretically, all such information is contained in the width function of the location region. Thus, to obtain information from a random sample on the degree of symmetry of the underlying distribution, the techniques of projection pursuit are readily employed.

3.1 *Symmetry via projection pursuit*

Methods of projection pursuit were first developed by Kruskal (1969, 1972), and implemented by Friedman and Tukey (1974). An excellent review of the literature on projection pursuit (PP) is given by Huber (1985). Application of PP techniques with numerous examples can be found in Friedman (1987).

Determination of symmetry can be addressed by the application of PP methods. We will employ one-dimensional projections of each observed p-variate data point. Initially, the data will be "sphered" (Friedman (1987)) to remove location, scale, and correlation dependencies. It remains to define the projection index. To this end, consider the linear combination

$$\boldsymbol{a}'\boldsymbol{X} \quad \text{with} \quad \boldsymbol{a}'\boldsymbol{a} = 1, \quad \boldsymbol{a} \in \mathscr{R}^p .$$

Let F_a be the (univariate) distribution function of $\boldsymbol{a}'\boldsymbol{X}$. All information regarding the shape of B_F is contained in all such projections. This results from the following:

(1) By construction, $B_F = \bigcap T^{-1}(B_{F_T})$, where the intersection is taken over all orthogonal transformations T of $\mathscr{R}^p$, and where B_{F_T} is the location hyperrectangle of $T\boldsymbol{X}$.

(2) The boundaries of B_{F_T} depend on the distribution function F_T only through its univariate marginals. For details, see Blough (1985).

Hence, since each row of $T\boldsymbol{X}$ is a projection as well as a univariate marginal distribution of $T\boldsymbol{X}$, the result follows. Let $m_{F_a}(u)$ be the univariate location functional for F_a:

$$m_{F_a}(u) = \frac{1}{2}\,[F_a^{-1}(u) + F_a^{-1}(1-u)]\,,$$

for $0 \leq u \leq 1/2$. If F is symmetric of degree 1, then B_F is $(p-1)$-dimensional and thus there exists $\boldsymbol{a}_0 \in \mathscr{R}^p$ and $\theta_1 \in \mathscr{R}^1$ such that

(1) $m_{F_{a_0}}(u) = \theta_1$ for all u in $[0, 1/2]$, and

(2) the width function $W(\boldsymbol{a}_0) = 0$.

In fact, $\boldsymbol{a}_0$ will be orthogonal to the hyperplane of symmetry. Hence, PP will seek to find the direction in which the width of the location region is minimized. This motivates the use of the following (affine invariant) projection index:

$$Q(\boldsymbol{a}) = \sup_{0 \leq u \leq 1/2} m_{F_a}(u) - \inf_{0 \leq u \leq 1/2} m_{F_a}(u) .$$

In a direction of symmetry $\boldsymbol{a}_0$, $Q(\boldsymbol{a}_0) = 0$. The larger $Q(\boldsymbol{a})$ is, the more F deviates from symmetry in direction $\boldsymbol{a}$.

Given a random sample $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$, the sample analog of the projection index is

$$\hat{Q}(\boldsymbol{a}) = \max_{1 \leq i \leq n} \frac{1}{2} [(\boldsymbol{a}'\boldsymbol{x})_{(i)} + (\boldsymbol{a}'\boldsymbol{x})_{(n-i+1)}]$$

$$- \min_{1 \leq i \leq n} \frac{1}{2} [(\boldsymbol{a}'\boldsymbol{x})_{(i)} + (\boldsymbol{a}'\boldsymbol{x})_{(n-i+1)}] ,$$

where $(\boldsymbol{a}'\boldsymbol{x})_{(i)}$ is the i-th order statistic from the projected sample $(\boldsymbol{a}'\boldsymbol{x}_1, \boldsymbol{a}'\boldsymbol{x}_2, \ldots, \boldsymbol{a}'\boldsymbol{x}_n)$. Once a direction $\boldsymbol{a}_0$ is found which minimizes $\hat{Q}(\boldsymbol{a})$, a test for symmetry (see Doksum *et al.* (1977)) can be applied to test the hypothesis that F_{a_0} is symmetric. If accepted, further projection pursuit can be undertaken to determine a possibly higher degree of symmetry (see below).

3.2 *Structure removal*

If the hypothesis that F_{a_0} is symmetric is accepted, higher orders of symmetry may be investigated by structure removal. That is, since B_F has been found to be at most $(p-1)$-dimensional, we will seek to transform B_F orthogonally so that the first coordinate of all points in B_F is the same. This will make the $(p-1)$-dimensional hyperplane of symmetry perpendicular to the first coordinate axes. Projection pursuit will then continue on the last $(p-1)$ coordinates.

Thus, we seek an orthogonal transformation H such that for some $\lambda \in \mathscr{R}$,

$$H\boldsymbol{a}_0 = \pm \lambda \boldsymbol{e}_1 ,$$

where $e_1 = (1, 0, 0, \ldots, 0)' \in \mathscr{R}^p$, $a_0 = (a_{01}, a_{02}, \ldots, a_{0p})' \in \mathscr{R}^p$. This can be accomplished with the Householder matrix H given by

$$H = I - 2hh',$$

where

$$\lambda = -(\text{sign of } a_{01}),$$

$$h_1 = \sqrt{\frac{1}{2}\left(1 - \frac{a_{01}}{\lambda}\right)},$$

$$h_i = \frac{a_{0i}}{2\lambda h_1}, \qquad i = 2, 3, \ldots, p$$

(see, for example, Kennedy and Gentle (1980)).

Now transform the data to $(Hx_1, Hx_2, \ldots, Hx_n)$ and consider only the last $(p-1)$ coordinates of each transformed data point. Apply projection pursuit to this data. This procedure can be repeated until the hypothesis of symmetry in a direction which minimizes the current $\hat{Q}$ is rejected, or until the hypothesis of a centrosymmetric distribution is not rejected.

At each step, necessarily orthogonal candidate directions of symmetry will be obtained. More specifically, let $v_1 = a_1^* \in \mathscr{R}^p$ be the first candidate direction of symmetry. That is, initially, $\hat{Q}(a)$ is minimized at a_1^*. If the Doksum-Fenstad-Auberge test for symmetry of $F_{a_1^*}$ fails to reject the hypothesis of symmetry, we proceed to step 2; otherwise, declare F asymmetric and stop. In step 2, let $H_1^* = H_1$ be the $p \times p$ Householder matrix such that

$$H_1^* a_1^* = H_1 a_1^* = \pm \lambda e_1,$$

where e_j is the j-th unit vector in $\mathscr{R}^p$. Transform the data via H_1^* to $(H_1^* x_1, H_1^* x_2, \ldots, H_1^* x_n)$ and consider the last $(p-1)$ coordinates of each such point. Let $a_2^* \in \mathscr{R}^{p-1}$ be the direction in which $\hat{Q}(a)$ is minimized. If the Doksum-Fenstad-Auberge test in this direction fails to reject the hypothesis of symmetry, proceed to step 3; otherwise, declare F symmetric of degree and stop.

Let $v_2 = H_1^* \begin{pmatrix} 0 \\ a_2^* \end{pmatrix}$. Then v_1 and v_2 are orthogonal:

$$v_1' v_2 = a_1^{*\prime} H_1^* \begin{pmatrix} 0 \\ a_2^* \end{pmatrix} = e_1' \begin{pmatrix} 0 \\ a_2^* \end{pmatrix} = 0.$$

In step 3, let H_2 be the $(p-1) \times (p-1)$ Householder matrix such that

$$H_2^* \begin{pmatrix} 0 \\ a_2^* \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & H_2 & \\ 0 & & & \end{pmatrix} \begin{pmatrix} 0 \\ a_2^* \end{pmatrix} = \pm \lambda e_2 .$$

Transform the data to $(H_2^* H_1^* x_1, H_2^* H_1^* x_2, \ldots, H_2^* H_1^* x_n)$ and consider only the last $(p-2)$ coordinates of each such data point. Let $a_3^* \in \mathscr{R}^{p-2}$ be the direction in which $\hat{Q}(a)$ is minimized for this data. Test for symmetry. If rejected, proceed to step 4; otherwise, declare x symmetric of degree 2 and stop.

Let $v_3 = H_1^* H_2^* \begin{pmatrix} 0 \\ 0 \\ a_3^* \end{pmatrix}$. Then v_1, v_2 and v_3 are mutually orthogonal:

$$v_1' v_3 = a_1^{*\prime} H_1^* H_2^* \begin{pmatrix} 0 \\ 0 \\ a_3^* \end{pmatrix} = e_1' H_2^* \begin{pmatrix} 0 \\ 0 \\ a_3^* \end{pmatrix} = e_1' \begin{pmatrix} 0 \\ 0 \\ a_3^* \end{pmatrix} = 0 .$$

Also,

$$\begin{aligned} v_2' v_3 &= (0 \quad a_2^{*\prime}) H_1^* H_1^* H_2^* \begin{pmatrix} 0 \\ 0 \\ a_3^* \end{pmatrix} \\ &= (0 \quad a_2^{*\prime}) H_2^* \begin{pmatrix} 0 \\ 0 \\ a_3^* \end{pmatrix} = e_2' \begin{pmatrix} 0 \\ 0 \\ a_3^* \end{pmatrix} = 0 . \end{aligned}$$

Thus, at stage k, we have $v_1, v_2, \ldots, v_k$ mutually orthogonal directions of symmetry. The procedure continues by transforming each observation x_i to the vector $H_{k-1}^* H_{k-2}^* \cdots H_2^* H_1^* x_i$ and considering the last $(p-k+1)$ coordinates. Here

$$H_j^* = \begin{pmatrix} I_{j-1} & 0 \\ 0 & H_{j-1} \end{pmatrix}, \qquad \text{with} \qquad H_j^* \begin{pmatrix} \mathbf{0}_{j-1} \\ a_j^* \end{pmatrix} = \pm \lambda_j e_j .$$

In the original coordinate system,

$$v_j = H_1^* H_2^* \cdots H_{j-1}^* \begin{pmatrix} \mathbf{0}_{j-1} \\ a_j^* \end{pmatrix},$$

for $j = 1, 2, \ldots, p$ with $H_0^* = I_{p \times p}$.

This procedure is designed to determine the degree of symmetry in a p-variate distribution. If we progress through p steps, failing to reject the hypothesis of symmetry each time, we can conclude that the distribution is centrosymmetric. Thus, one application of this procedure is to test for centrosymmetry in a distribution. An example of this is presented in the next section.

4. Example

An investigation conducted by an entomologist at the University of Arizona consisted of collecting data from two adjacent cotton fields, each of which was irrigated by a different method. Field 1 received above ground drip irrigation, and Field 2 received underground drip irrigation. The average daily vapor pressure deficit (VPD) was measured on each field at three elevations: ground level (low), top of crop canopy (high), and midway between the two (midheight). Observations were obtained from the center of each field for 41 days during the summer of 1986. The primary goal in this investigation was to study the effects of irrigation method on the field environment, since the environment is known to affect insect infestation in cotton fields. To stabilize the variance, base 10 logarithms of the original VPD values were taken.

A spatial time series model was fit to this 6-variate time series model by way of the Kalman filter. An assumption needed for the filter is that the innovations (residuals) are 6-variate normally distributed. In particular, this implies the distribution of innovations should be centrosymmetric. We use the techniques of Section 3 to assess this assumption. It should be noted that the variance-stabilizing transformation mentioned above might also remove skewness in the data. Hence, by testing for centrosymmetry, we are in effect checking the appropriateness of taking logarithms.

The one-step ahead residuals from the model fit are given in Table 1. Prior to PP, the data were "sphered". Optimal projections were obtained by using the programming language GAUSS on a personal computer with a quasi-Newton optimization routine. The "coarse stepping" minimization algorithm suggested by Friedman (1987) was used prior to quasi-Newton minimization.

The results are as follows:

$$\upsilon_1' = (-0.017, \ -0.019, \ -0.007, \ 0.000, \ 0.999, \ 0.000)\,,$$

$$\upsilon_2' = (0.133, \ -0.052, \ 0.157, \ -0.082, \ 0.002, \ 0.974)\,,$$

$$\upsilon_3' = (0.188, \ -0.395, \ 0.868, \ 0.157, \ 0.001, \ -0.173)\,,$$

$$\upsilon_4' = (0.129, \ 0.447, \ 0.031, \ 0.881, \ 0.011, \ 0.075)\,,$$

Table 1. Vapor pressure deficit data: time series residuals.

OBS	Above ground drip irrigation			Underground drip irrigation		
	low	midheight	high	low	midheight	high
1	0.331	0.323	0.157	− 0.174	− 0.233	− 0.257
2	0.084	0.138	− 0.357	− 0.330	− 0.358	− 0.340
3	0.255	− 0.200	− 0.377	0.439	0.389	0.289
4	− 0.824	0.325	0.046	0.005	0.033	0.105
5	0.365	− 0.584	− 0.155	0.090	0.161	0.153
6	− 0.196	− 0.082	− 0.151	− 0.147	− 0.137	− 0.116
7	0.209	− 1.156	0.007	− 0.932	− 0.983	− 1.091
8	0.428	− 0.071	0.035	− 0.092	− 0.091	− 0.150
9	0.201	− 0.281	0.056	− 1.144	− 1.595	− 1.542
10	0.029	0.671	− 0.144	0.508	0.905	0.354
11	0.237	− 0.289	0.011	0.728	0.623	0.813
12	0.091	0.552	0.145	0.176	0.197	0.328
13	0.017	0.242	0.006	0.086	0.105	0.214
14	0.189	0.218	− 0.030	0.058	0.088	0.168
15	0.098	0.102	0.049	0.120	0.083	0.082
16	0.034	0.164	− 0.073	− 0.124	− 0.107	− 0.064
17	0.049	− 0.068	0.050	− 0.004	− 0.048	− 0.090
18	− 0.118	− 0.091	0.196	0.378	0.363	0.369
19	0.106	− 0.455	− 0.104	− 0.433	− 0.399	− 0.387
20	− 0.026	− 0.386	− 0.345	0.057	0.044	− 0.016
21	− 0.204	− 0.316	− 0.359	− 0.684	− 0.771	− 0.988
22	0.046	− 0.137	− 0.080	0.062	0.059	0.017
23	0.072	− 0.174	0.326	0.243	0.351	0.463
24	0.227	0.306	0.306	0.639	0.678	0.800
25	− 0.009	0.092	0.149	− 0.243	− 0.228	− 0.213
26	0.119	0.600	− 0.010	0.146	0.136	0.133
27	− 0.057	− 0.207	0.077	− 0.026	− 0.028	− 0.019
28	− 0.010	0.198	− 0.025	− 0.383	− 0.392	− 0.433
29	− 0.002	0.043	0.044	0.149	0.126	0.102
30	0.106	− 0.206	0.022	0.172	0.175	0.190
31	− 0.139	0.146	− 0.115	− 0.099	− 0.118	− 0.134
32	− 0.053	− 0.779	− 0.091	− 0.309	− 0.363	− 0.464
33	− 0.027	− 0.269	− 0.109	− 0.001	− 0.076	− 0.063
34	0.035	0.025	0.141	0.215	0.246	0.279
35	0.057	0.312	0.175	0.071	0.065	0.084
36	0.124	1.033	0.293	− 0.011	0.005	0.012
37	− 0.025	0.140	− 0.267	0.171	0.138	0.092
38	− 0.081	0.057	− 0.056	− 0.509	− 0.548	− 0.618
39	− 0.034	− 0.472	− 0.195	0.002	− 0.007	− 0.032
40	− 0.339	0.006	− 0.492	− 0.141	− 0.053	− 0.091
41	0.033	0.251	0.149	− 0.518	− 0.584	− 0.864

$$v_5' = (-0.544,\ 0.647,\ 0.462,\ -0.266,\ 0.006,\ 0.012)\,,$$

$$v_6' = (0.796,\ 0.471,\ 0.080,\ -0.348,\ 0.022,\ -0.125)\,,$$

$$H_1^* = \begin{pmatrix} -0.017 & -0.019 & -0.007 & 0.000 & 0.999 & 0.000 \\ & 0.999 & 0.000 & 0.000 & 0.018 & 0.000 \\ & & 0.999 & 0.000 & 0.006 & 0.000 \\ & & & 0.999 & 0.000 & 0.000 \\ & & & & 0.017 & 0.000 \\ \text{(sym)} & & & & & 0.999 \end{pmatrix},$$

$$H_2^* = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ & -0.055 & 0.156 & -0.082 & 0.133 & 0.974 \\ & & 0.977 & 0.012 & -0.020 & -0.144 \\ & & & 0.994 & 0.010 & 0.075 \\ & & & & 0.983 & -0.123 \\ \text{(sym)} & & & & & 0.101 \end{pmatrix},$$

$$H_3^* = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ & 1 & 0 & 0 & 0 & 0 \\ & & -0.808 & -0.188 & -0.136 & 0.541 \\ & & & 0.980 & -0.014 & 0.056 \\ & & & & 0.990 & 0.041 \\ \text{(sym)} & & & & & 0.838 \end{pmatrix},$$

$$H_4^* = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ & 1 & 0 & 0 & 0 & 0 \\ & & 1 & 0 & 0 & 0 \\ & & & -0.837 & -0.187 & -0.515 \\ & & & & 0.981 & -0.052 \\ \text{(sym)} & & & & & 0.856 \end{pmatrix},$$

$$H_5^* = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ & 1 & 0 & 0 & 0 & 0 \\ & & 1 & 0 & 0 & 0 \\ & & & 1 & 0 & 0 \\ & & & & -0.451 & 0.893 \\ \text{(sym)} & & & & & 0.451 \end{pmatrix},$$

$$H_6^* = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ & 1 & 0 & 0 & 0 & 0 \\ & & 1 & 0 & 0 & 0 \\ & & & 1 & 0 & 0 \\ & & & & 1 & 0 \\ \text{(sym)} & & & & & -1.000 \end{pmatrix}.$$

Figure 1 shows the plots of the function of symmetry $\hat{Q}$ in each of the six candidate directions obtained from PP. We select an overall $\alpha = 12\%$, so that by Bonferroni, each test is conducted at the 2% level; thus the 98% confidence bands in each graph (these are the "B-Bands" of Doksum *et al.* (1977)). Since a horizontal line fits between the confidence bands in all six directions, we accept the centrosymmetric hypothesis.

Although the Bonferroni method guarantees an overall level of 12% for these 6 tests, it does not provide protection against the "selection effect" of testing for symmetry in directions suggested by the data. In as much as projection pursuit is primarily a data exploration technique, the resultant possible bias in a formal test procedure cannot be avoided.

If interest lies in testing directly for centrosymmetry, a modification of the above procedure allows for this. In this case, consider maximizing the projection index $\hat{Q}(\boldsymbol{a})$. In the case of centrosymmetry, $Q(\boldsymbol{a})$ is identically 0. An appropriate test would then reject the null hypothesis of centrosymmetry if the maximum $\hat{Q}$ were large. In order to assess the significance of this test statistic, it is necessary to derive a reference distribution of maximum $\hat{Q}$ values obtained under the null hypothesis of centrosymmetry. In lieu of analytical derivation of this distribution, it can be obtained via a Monte Carlo study of the results of maximizing $\hat{Q}$ when applied to a known centrosymmetric distribution. This approach is discussed in general by Friedman (1987).

As an example, we will apply this modified procedure to the above Entomology data. Since these represent time series residuals obtained by fitting a time series model under the assumption of normally distributed error terms, it is natural to take the 6-variate normal distribution as the null distribution. This was done and the following Monte Carlo study was conducted: 100 6-variate normal random samples, each of size 41 were generated. More specifically, since location, scale, and dependency structure of the null distribution do not affect the projection index, each observation consisted of 6 independent, randomly generated univariate standard normal variables. A histogram of the 100 resulting values of the maximum $\hat{Q}$ is presented in Fig. 2. When the maximum $\hat{Q}$ was computed for the Entomology data, the value 1.036 was obtained. With respect to the histogram of the null distribution, this is not significant. Hence, we again find no evidence of asymmetry in the underlying distribution.

5. Conclusion

A non-parametric technique for determining the degree of symmetry possessed by a multivariate distribution has been developed. The method is based on projection pursuit with a projection index obtained from a test for symmetry developed by Doksum (1975). The intimate links between this index, the dimension of the location region developed by Blough

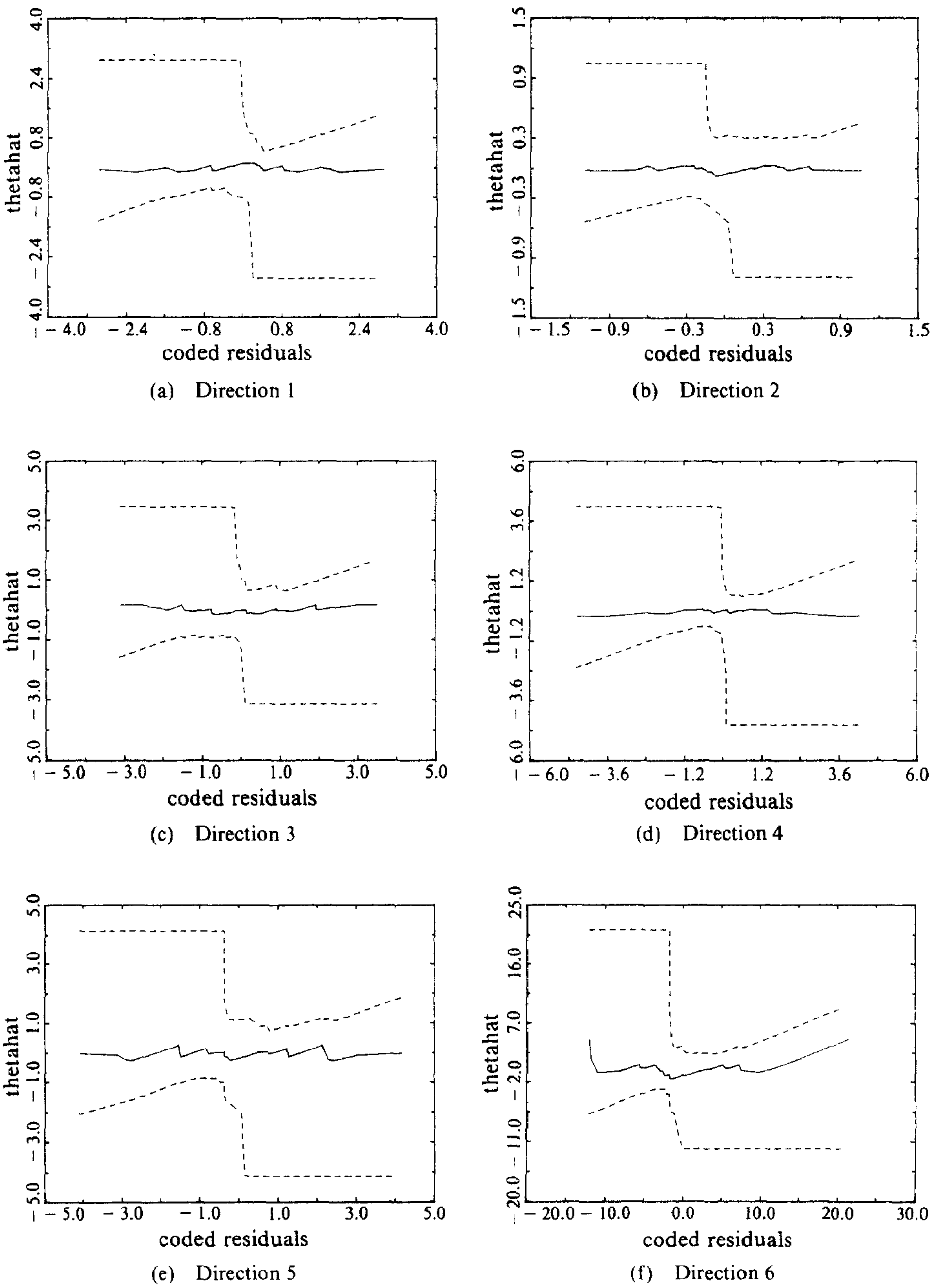

Fig. 1. Empirical symmetry functions (solid lines) with 98% confidence bands (dashed lines).

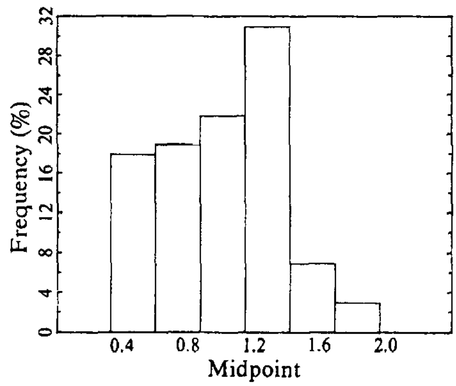

Fig. 2. Histogram of 100 maximum $\hat{Q}$ values obtained by sampling from the 6-variate normal distribution 100 times.

(1985), and the degree of symmetry in the distribution are what make the projection pursuit techniques viable.

Acknowledgements

The author wishes to thank the referees for their suggestions, in particular for suggesting the modified procedure for testing directly for centrosymmetry. The author also wishes to thank Professor Irene Terry of the Department of Entomology at the University of Arizona for the use of her data in the example.

Appendix

The location region for a p-variate random vector X with distribution function F can be constructed in a manner completely analogous to that used in the bivariate case developed by Blough (1985). The details of one point in the construction do warrant further discussion, however. We need to show that a location functional $\boldsymbol{\theta}_X$ is equivariant under orthogonal transformations. To this end, let R be a $p \times p$ orthogonal matrix. We wish to prove that if $\boldsymbol{\theta}_X$ is in the location region A constructed via Method I in Blough's paper, then

$$R\boldsymbol{\theta}_X = \boldsymbol{\theta}_{RX} .$$

Now there exists an orthogonal matrix P such that $R = P'QP$ where Q is of the block-diagonal form

$$\begin{pmatrix} \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix} & & & & \\ & \begin{pmatrix} -1 & & \\ & \ddots & \\ & & -1 \end{pmatrix} & & & \\ & & \begin{pmatrix} \cos\alpha_1 & \sin\alpha_1 \\ -\sin\alpha_1 & \cos\alpha_1 \end{pmatrix} & & \\ & & & \ddots & \\ & & & & \begin{pmatrix} \cos\alpha_r & \sin\alpha_r \\ -\sin\alpha_r & \cos\alpha_r \end{pmatrix} \end{pmatrix}$$

for some r (see, for example, Herstein (1964), pp. 306–307). Hence $Q = PRP'$. Now geometrically, PX represents a change of basis. Let γ_{PX} be the location functional for $\boldsymbol{X}$ in this new basis corresponding to $\boldsymbol{\theta}_X$ in the old basis; that is,

$$\boldsymbol{\theta}_X = P^{-1}\gamma_{PX} = P'\gamma_{PX} .$$

Then

$$R\boldsymbol{\theta}_X = P'QP\boldsymbol{\theta}_X = P'Q\gamma_{PX} = P'\gamma_{QPX}$$

(applying the results of Doksum and Blough, since Q represents, geometrically, one-dimensional reflections and two-dimensional rotations)

$$= P'\gamma_{PRP'PX} = P'\gamma_{PRX} = \boldsymbol{\theta}_{RX} .$$

References

Blough, D. K. (1985). Measures of location in the plane, *Ann. Inst. Statist. Math.*, **37**, 545–555.

Doksum, K. A. (1975). Measures of location and asymmetry, *Scand. J. Statist.*, **2**, 11–22.

Doksum, K. A., Fenstad, G. and Auberge, R. (1977). Plots and tests for symmetry, *Biometrika*, **64**, 473–487.

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. Ser. C*, **23**, 881–889.

Herstein, I. N. (1964). *Topics in Algebra*, Xerox College Publishing, Waltham, Massachusetts.

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

Kennedy, W. J. and Gentle, J. E. (1980). *Statistical Computing*, Marcel Dekker, New York.

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, *Statistical Computation*, (eds. R. C. Milton and J. A. Nelder), Academic Press, New York.

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

MacGillivray, H. L. (1986). Skewness and asymmetry: Measures and orderings, *Ann. Statist.*, **14**, 994–1011.