Department Copy

χ - PLOTS FOR INDEPENDENCE

by
Paul Switzer
Stanford University
and
N. I. Fisher
CSIRO, Australia

TECHNICAL REPORT NO. 8
OCTOBER 1984

PREPARED UNDER THE AUSPICES OF
OF
NATIONAL SCIENCE FOUNDATION
GRANT MCS 81-09584

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

χ - PLOTS FOR INDEPENDENCE

by
Paul Switzer
Stanford University
and
N. I. Fisher
CSIRO, Australia

TECHNICAL REPORT NO. 8
OCTOBER 1984

PREPARED UNDER THE AUSPICES OF
OF
NATIONAL SCIENCE FOUNDATION
GRANT MCS 81-09584

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

χ - PLOTS FOR INDEPENDENCE

N.I. FISHER
CSIRO
Division of Mathematics & Statistics
Lindfield, NSW, Australia 2070

and

P. SWITZER
Department of Statistics
Stanford University
Stanford, CA94305, USA

SUMMARY

This paper presents a graphical method to be used, in conjunction with a scatterplot, to investigate possible association of two variates as manifested in a sample of bivariate measurements. The method is designed so that the plot is approximately horizontal under independence, and under various forms of association it produces corresponding characteristic patterns. Examples given include application to study of regression residuals, dependence of two spatial point processes, serial association of time series, and comparison of two time series.

Some key words: association, correlation, graphical method, nonparametric method, probability plotting, regression residuals, tests of independence, time series

1. INTRODUCTION

The dependence between a pair of variates can be various with potentially surprising aspects. Therefore, a global summary statistic such as a correlation coefficient will not usually convey the quality of so complex a property as dependence. In this paper, a graphical method is proposed, which attempts to convey explicit dependence information in detail, empirically without prior model structure.

Bivariate numeric data $(X_1,Y_1),\ldots,(X_n,Y_n)$ are commonly and usefully presented graphically in the form of a scatterplot. In particular, if obvious regression effects are conceivably absent or have been removed, then the (X,Y) scatterplot may be used to reveal residual association patterns. In this paper we propose another plotting tool; the n data pairs (X_i,Y_i) $i = 1,\ldots,n$ are transformed into n pairs (λ_{ni},χ_{ni}), $i = 1,\ldots,n$ providing a second scatterplot, the χ-plot or chi-plot, intended to reveal more detailed and explicit information regarding the nature of the association between X and Y.

Roughly, χ_{ni} measures the failure of the bivariate distribution function to factor into a product of marginal distribution functions at the sample argument (X_i,Y_i), and λ_{ni} measures the distance from (X_i,Y_i) to the bivariate median. Each of the transforms χ and λ thus operates on the full set of 2n numbers. The χ transform provides an indication of X - Y association separately for each data point, and the λ transform positions that data point with respect to the X and Y marginal distributions.

The proposed data transforms are invariant under monotonic re-scaling of the X and Y measurements and the χ-plot is scaled so that all points fall in the rectangle $[-1,1] \times [-1,1]$. Positive and negative departure from independence appear as corresponding deviations from the horizontal line $\chi = 0$, with calculated allowances for scatter due to sampling variability. Increasing the sample size n tends to decrease the amount of scatter in the χ-plot.

In section 2 the χ and λ transforms are defined and described. Some additional properties of these transforms under random sampling are discussed in section 4. Section 3 presents a series of examples based on real and simulated bivariate data sets for which the (X,Y) scatterplot and the χ-plot are presented together. Included are examples of regression residuals versus predictor values, lagged time series, bivariate time series, and comparison of spatial point patterns. The range of these examples goes beyond the context of the available sampling distribution theory of section 2 and section 4. On the other hand the simulated data examples are specifically included to illustrate the reasonableness of

this approximate theory and to show the effects of both mild and moderate correlation in bivariate normal situations.

2. DESCRIPTION OF THE BIVARIATE DATA TRANSFORMS

2.1 The data transform χ

For each of the bivariate sample points (X_i,Y_i), $i = 1,\ldots,n$ partition the (X,Y) plane into quadrants, which are formed by intersecting the regions $X \leqslant X_i$ and $Y \leqslant Y_i$. The remaining $n - 1$ bivariate sample points will distribute themselves among the four quadrants with relative frequencies denoted by A_{ni}, B_{ni}, C_{ni}, D_{ni}, summing to unity. Under random sampling, some multinomial distribution will give the variability of these quadrant frequencies, conditional on the sample cut-point (X_i,Y_i), for each of the n 2×2 tables.

These quadrant counts are now used to define a sample bivariate distribution function H_n and sample marginal distribution functions F_n, G_n evaluated at the data points, viz.,

$$H_n(X_i,Y_i) \equiv H_{ni} = A_{ni} = \sum_{j \neq i} I[X_j \leqslant X_i,\ Y_j \leqslant Y_i]/(n - 1)$$

$$F_n(X_i) \equiv F_{ni} = A_{ni} + B_{ni} = \sum_{j \neq i} I[X_j \leqslant X_i]/(n - 1)$$

$$G_n(Y_i) \equiv G_{ni} = A_{ni} + D_{ni} = \sum_{j \neq i} I[Y_j \leqslant Y_i]/(n - 1)$$

where $I[A] = 0,1$ according as A is false or true. If X and Y are statistically independent then one would expect the sample bivariate distribution function to be the same as the product of the marginal distribution functions, except for sampling variability, at each of the sample cut-points.

With appropriate scaling, the differences $H_{ni} - F_{ni}G_{ni}$, $i = 1,\ldots,n$ will each behave asymptotically like a standard normal variable under random sampling from independent marginals, given the cut-point (X_i,Y_i). The appropriate scaling factor, which depends on the marginal frequencies, is $(n^{\frac{1}{2}}/S_{ni})^{-1}$ where

$$S_{ni}^2 = F_{ni}(1 - F_{ni})G_{ni}(1 - G_{ni}).$$

Therefore, a signed and scaled measure of the departure from bivariate statistical independence, as evidenced at the sample cut-point (X_i,Y_i), is the χ-transform defined as

$$\chi_{ni} = (H_{ni} - F_{ni}G_{ni})/S_{ni}, \qquad i = 1,\ldots,n.$$

The χ-transform remains undefined for those two, three, or four sample points where $S_{ni} = 0$, assuming no ties at the extreme data values. It may be useful to note that the numerator $H_{ni} - F_{ni}G_{ni}$ is the same as $A_{ni}C_{ni} - B_{ni}D_{ni}$, the cross-product difference in the corresponding 2×2 table.

At each sample point the transform χ_{ni} is actually a correlation coefficient between dichotomized X values and dichotomized Y values. Therefore, all values of χ_{ni} lie in the interval $[-1,1]$. Furthermore, it can be checked that $\sqrt{n}\ \chi_{ni}$ is exactly the signed square root of the usual χ^2 statistic for testing independence in the 2×2 table generated by the cut-point (X_i,Y_i). In the case that the variable Y is a strictly increasing function of X, it turns out that $\chi_{ni} = 1$ for all sample cut-points; similarly, when Y is a strictly decreasing function of X then $\chi_{ni} = -1$ for all sample cut-points. Finally, correlations between χ values for different sample cut-points are discussed later in section 4.1.

2.2 The data transform λ

Under independence of the measured variables X and Y, the approximate normal distribution of each χ_{ni}, i = 1,...,n is unaffected by conditioning on corresponding marginal frequencies F_{ni}, G_{ni} that are not too close to zero or one. Therefore, let $\lambda(F_{ni},G_{ni}) \equiv \lambda_{ni}$ be a real-valued function of the marginal frequencies. Any plot of χ_{ni} vs. λ_{ni}, i = 1,...,n would asymptotically show normal vertical scatter with variance 1/n around the horizontal $\chi = 0$, under independence of X and Y. Association between X and Y should be revealed by departures from such zero-centred vertical scatter on a (λ,χ) plot, called briefly a <u>χ-plot</u>. The graphical aspect of this plot will then depend on the nature of the X, Y association and the choice of function λ.

While it is sometimes customary in graphical procedures to represent non-null behaviour by a straight line with either non-zero slope or non-zero intercept, this would be akin to trying to summarize dependence by a single global correlation coefficient.

The examples of the following section use

$$\lambda_{ni} = 4(\mathrm{sgn}_{ni})\max\{(F_{ni} - \tfrac{1}{2})^2,(G_{ni} - \tfrac{1}{2})^2\}$$

where

$$\mathrm{sgn}_{ni} = \mathrm{sgn}\{(F_{ni} - \tfrac{1}{2})(G_{ni} - \tfrac{1}{2})\}.$$

All values of this λ must lie in the interval [-1,1]. When the data are a random bivariate sample from independent continuous marginals, then the values of λ_{ni}, i = 1,...,n are individually uniformly distributed. However, when X and Y are associated then the values of λ_{ni} may show clustering. In particular, if X and Y are positively correlated, the λ_{ni} values will tend to positive values, and conversely for negative correlation.

Thus $|\lambda_{ni}|$ may be viewed as a measure of the distance from the sample point (X_i, Y_i) to the bivariate median of the distribution. Hence, when a departure from independence is exhibited by some sample points according to the χ criterion, then the χ-plot should show where such points lie with respect to the centre of the bivariate distribution.

An undesirable aspect of this signed λ functional is its discontinuity at $F = \frac{1}{2}$ and at $G = \frac{1}{2}$. One might have used $|\lambda|$ instead; however the signs of the λ_{ni} values do carry dependence information and it seems a little wasteful not to use them. Other λ functionals we considered were

(i) $|\lambda| = 4(F_{ni} - \frac{1}{2})(G_{ni} - \frac{1}{2})$, related to Spearman's rank correlation.

(ii) $|\lambda| = 4\{|F_{ni} - \frac{1}{2}| - |G_{ni} - \frac{1}{2}| - \frac{1}{2}\}^2$, also uniformly distributed under independence.

3. ILLUSTRATIONS

3.1 Modifications of the χ-plot

The sampling behaviour of the χ transform of the data will be erratic for those sample points at the edges of the data distribution, and the asymptotic normal theory will be an inappropriate approximation. Therefore, those extreme sample points for which $|\lambda_{ni}| \geqslant 4\{1/(n-1) - \frac{1}{2}\}^2$ have not been plotted in the illustrations of the next section. This censoring criterion will eliminate at most eight points.

Also, the scale of the χ transform, in the illustrations following, has been monotonically modified to

$$\tilde{\chi} = \sin(\tfrac{1}{2}\pi\chi)$$

in order to clarify behaviour of χ-plots under sampling from bivariate normal distributions with correlation ρ. This modification retains the correlation range [-1,1] and has the additional property that $\tilde{\chi}_{ni}$ is asymptotically equal to ρ as λ_{ni} approaches zero, i.e., for sample cut-points near the bivariate median. The modified $\tilde{\chi}$ has asymptotic standard deviation $\frac{1}{2}\pi n^{-\frac{1}{2}}$ under independence. If there is no degree of monotone dependence between X and Y, but there is some dependence of a more complex nature, this will be manifested in the $\tilde{\chi}$-values in terms of increased scatter, and possibly non-uniform increase in scatter, along the λ-axis.

As a final important comment before turning to the examples, we note that the χ_{ni}- and λ_{ni}-values depend on the data only through the ranks of X_i among $X_1,\ldots,X_n$ and $Y_1,\ldots,Y_n$, so that the plot is resistant to the effect of gross outliers. Both Pearson and Spearman sample correlations, $\hat{\rho}$ and $\hat{\rho}_s$ respectively, are quoted with each scatterplot.

3.2 Examples

We begin with some examples of artificial data, to illustrate the behaviour of the χ-plot under independence and under different forms of dependence.

Example 1. Figure 1a is a scatterplot of 100 pseudo-random pairs of independent variables X, Y drawn from the standard bivariate normal distribution and Figure 1b shows the related χ-plot. The points in the χ-plot are scattered reasonably uniformly along the λ-axis, with approximately half of them on each side of the line $\chi = 0$; a couple of points are lying outside the approximate 95% pointwise probability region.

Example 2.

Figure 2a is a scatterplot of 100 pseudo-random pairs (X,Y) drawn from the standard bivariate normal distribution with correlation $\rho = 0.1$. The χ-plot for these data, in Figure 2b, shows differences from the χ-plot in Figure 1b characteristic of monotone positive association: the χ-values tend to increase as $\lambda \uparrow 0$ and to fall away again as λ continues to increase, the majority of points lie above $\chi = 0$, rather more than 2½% lie above the upper 97½% line, and the majority of points lie to the right of the axis $\lambda = 0$.

Example 3.

The data in Figure 3a are as for Example 2, except with $\rho = 0.5$. The various departures from null behaviour described in Example 2 are much more marked in Figure 3b and additionally, scatter increases for λ near 1. Note that near $\lambda = 0$, the χ-values approximate the sample correlation of 0.65.

In the limit, when $\rho = 1$, all χ-values are equal to 1, and the λ-values lie in the range [0,1], concentrated near zero and more and more sparse as $\lambda \to 1$.

Example 4.

The data shown in Figure 4a are 100 observations drawn from the standard bivariate normal distribution with zero correlation and satisfying the restrictions $Y \geqslant 0$ and $|X^2 + Y^2 - 1| \leqslant 0.5$. Thus, there is essentially no monotone association, but rather moderate non-monotone association. The effect of this on the χ-plot is shown in Figure 4b, with far too many points lying outside the 95% probability region.

The next three examples are applications to artificial data sets for which the assumptions upon the χ-plot is based are in some way violated. The point is that the plot still behaves reasonably under the null hypothesis of independence of X and Y. Later real-life examples of the same type behave rather differently.

Example 5.

Figure 5a was obtained by generating a sample of 100 independent observations Z_i, $1 \leqslant i \leqslant 100$ from the standard normal distribution and forming the dependent sequence of pairs $(X_i,Y_i) = (Z_i,Z_{i+1})$, $1 \leqslant i \leqslant 99$. The χ-plot to check on serial association in the Z-sequence is shown in Figure 5b. There is little difference to be seen between this plot and the χ-plot in Figure 1b, apart from a possibly reduced scatter along the χ-axis, there being no points outside the bounds.

Example 6.

Figure 6a is a scatter plot of 100 observations Z_i as in Example 5, each Z_i being plotted against its observation number, i. Again, comparison of the corresponding χ-plot in Figure 6b with that in Figure 1b reveals little to disturb one's tranquility about the hypothesis of independence of Z_i and i.

Example 7.

Figure 7a shows 100 points (X_i,Y_i), with $X_1,\ldots,X_{100}$ being independently sampled from the standard normal distribution, and independently, $Y_1,\ldots,Y_{100}$ being drawn from the same distribution but serially correlated with lag 1 correlation 0.9. Thus the hypothesis being tested by the χ-plot in Figure 7b is of independence of the values measured at the same time point of two time series. The only obvious departure from the behaviour of the χ-plot in Figure 1b can be explained

by the presence of a small amount of monotone association between X_i and Y_i as manifested by the Spearman's rho-value of 0.06, arising from sampling variability.

The feature we have observed in simulated data sets for the situations in Examples 5, 6 and 7 is a slight tendency for the χ-values to be less dispersed than under the hypothesis that $(X_1,Y_1),\ldots,(X_n,Y_n)$ constitute a random sample, X and Y being independent. It would appear that the denominator of χ is overestimating the standard deviation of the numerator, a point which we shall pursue elsewhere. The effect of this on the procedure is to make it slightly conservative in detecting dependence.

We turn now to some applications to real data sets.

Example 8.

A test for independence of two spatial point patterns: Diggle (1983) describes procedures for sparse sampling of an area containing two species of trees, for example, to test for independence of their respective patterns. One method is based on measuring the distances X, Y from a given point on a randomly located regular $p \times p$ grid to the nearest member of each species, and testing the resulting p^2 data pairs for independence. Diggle (1983, pp129-139) gives the locations of 703 hickories and 929 oaks in Lansing Woods, and maps of these in Figure 17 (op.cit.). Figure 8a is a scatterplot of (X,Y) data obtained in this way from an 8×8 grid. The χ-plot in Figure 8b reveals departure from the null model. In fact, the magnitudes of the sample correlations for these data are essentially the same as those for Example 2, yet comparison of Figure 8b with Figure 2b shows interesting differences. For example, χ-values in Figure 8b for $\lambda \to -1$ seem

reasonably well-behaved, yet the other part of the plot, particularly for $\lambda \geqslant 0$, clearly points to negative association. Now λ-values near 1 correspond to data pairs X_i, Y_i which are both small or both large, so a sensible conjecture is that the negative association derives from pairs of small distances. One way of enhancing the χ-plot to investigate this further would be to label the $\chi - \lambda$ values according to the corresponding (X_i, Y_i) pair in the scatterplot.

Example 9.

The 117 points in Figure 9a were digitised from a scatterplot appearing in a paper by Olmstead and Tukey (1947). The data set was used to motivate a "corner" test of association, for situations in which the association may be manifested in peripheral points. The associated χ-plot in Figure 9b indicates some form of association, some of which is positive dependence. However, it is useful to compare this Figure with Figure 2b, as we did in the previous example. In this case, it appears that λ-values near -1 lead to unusually high χ-values, as judged by Figure 2b and the sample correlations. These λ-values correspond to peripheral points X, Y which are large and of opposite signs, and suggest further investigation along the lines described in Example 8.

Examples 8 and 9 emphasize the fact that there is a certain amount of information in the χ-plot about the nature of the dependence between X and Y which is not obvious, or not available, using a scatterplot and a formal test.

Example 10.

Gnanadesikan (1977) describes a procedure for testing multivariate normality which relies on computing "scaled residuals" of the data from the mean value, and converting each residual to a squared

radius and direction (r^2,θ). For bivariate data, r^2 should be approximately χ^2_2 and θ uniform, and r^2 and θ should be approximately independent. Gnanadesikan (1977,Fig.28h) gives a scatterplot of 100 $(X = \exp(-\frac{1}{2}r^2),\ Y = \theta)$ values derived from a sample simulated from a bivariate χ^2_{10} distribution; our Figure 10a was obtained by digitising his scatterplot. The corresponding χ-plot in Figure 10b indicates a small degree of negative association, and also rather more scatter in the χ-values than would be expected under independence. Comparison with Example 4, and our studies of other data sets, suggest that the increased scatter may be attributable to "holes" in the data mass. In fact, these are evident in the scatterplot, and are confirmed by a cell count test used by Gnanadesikan.

It is not uncommon in practice to encounter a situation in which one is interested in the relationship between a random variable Y and an ordinary variable X, particularly in regression analysis where Y and X may be a residual and a design variable, respectively. The χ-plot may be formally applied to investigate this, as suggested in Example 6, and as is done in the next example.

Example 11.

Figure 11a is a scatterplot of 100 data points kindly supplied by Dr. R.L. Sandland, CSIRO, from a regression analysis. $X_i = i$, the observation number, and Y_i is the i^{th} residual from a fitted model. There is essentially no monotone association between the variables, but some form of non-linear association is evident in the χ-plot in Figure 11b; the increased dispersion of the χ-values is at least partly attributable to a "hole" in the data mass, as discussed in the previous example.

Example 12.

Corresponding to Example 5, we consider a real-life application of the χ-plot to testing a time series for serial association. Ali and Giacotto (1982) give several series of stock price movements, specifically, monthly log-price changes over the period January 1, 1966 to December 31, 1976; we shall examine the Alcoa series $\{Z_i\}$. The authors contend that the observations may be regarded as independent, based on their analysis of sample lag correlation coefficients. Figure 12a is a scatterplot of the 130 pairs (Z_i, Z_{i+1}) and Figure 12b the χ-plot. Comparison with Figure 5b, the χ-plot for the synthetic time series data, suggests a slight but definite negative association. One reason for the differing conclusions could be that data of this type tend to have long-tailed distributions, which may render invalid the formal test for serial correlation.

Example 13.

Corresponding to Example 7, we consider the comparison of corresponding time values in two parallel time series. The two series we shall use are derived from those given by Whittle (1954) on sunspot observations in the Northern and Southern solar hemispheres. Denoting Whittle's series by $\{U_t\}$ and $\{V_t\}$, define $X_t = U_t - V_t$, $Y_t = U_t + V_t$, $t = 1,2,\ldots,132$. Under a certain hypothesis, the series $\{X_t\}$ and $\{Y_t\}$ are supposed to be independent. Figure 13a is a scatterplot of $(0.001\ X_t,\ 0.001\ Y_t)$. The obvious lack of independence is reflected in the bizarre pattern of the χ-plot in Figure 13b. Study of a plot of X_t against t suggests that this may be due in part to scale change of the X_t-series corresponding to location and scale change of the Y-series.

4. ADDITIONAL PROPERTIES OF THE χ-PLOT AND POSSIBLE EXTENSIONS

4.1 Induced correlations between χ-values

Under random sampling from a bivariate distribution with independent marginal distribution functions F and G, the asymptotic joint distribution of χ_{ni} and χ_{nj} is bivariate normal with positive correlation, conditional on the sample cut-points (X_i,Y_i) and (X_j,Y_j). This asymptotic correlation is given by

$$-2 \log \operatorname{corr}(\chi_{ni},\chi_{nj}) =$$

$$= \left| \log \frac{F(X_i)\{1-F(X_j)\}}{F(X_j)\{1-F(X_i)\}} \right| + \left| \log \frac{G(Y_i)\{1-G(Y_j)\}}{G(Y_j)\{1-G(Y_i)\}} \right| .$$

So, while the asymptotic distributions of $\chi_{n1},\chi_{n2},\ldots$ are all the same and do not depend on the cut-points, the joint asymptotic distributions do indeed depend on the cut-points. It follows that the correlation between χ_{ni} and χ_{nj} is least when one sample cut-point is near an extremity of its marginal distribution. The correlation is greatest when both sample cut-points are near the bivariate median and, in this case, approximately,

$$\operatorname{corr}(\chi_{ni},\chi_{nj}) = 1 - 2\{|F(X_i) - F(X_j)| + |G(Y_i) - G(Y_j)|\}$$

which may itself be estimated by substituting the sample marginal distribution functions.

The average correlation between two values of χ, still under independence of X and Y, may be calculated by averaging over the distribution of the two cut-points. We may regard $F(X_i)$, $F(X_j)$, $G(Y_i)$, $G(Y_j)$ as four independent uniform random variables and the expected value of the asymptotic correlation between χ_{ni} and χ_{nj} is then

$$\mathrm{corr}(\chi_{ni},\chi_{nj}) = (\pi^2 - 8)^2/16 \simeq 0.22 \ .$$

4.2 Relation to tests of the independence hypothesis

The asymptotic distribution theory may be used directly to construct tests of independence. For any single fixed cut-point, the value of χ_{ni} may be compared to a standard normal distribution, one-tailed or two-tailed as appropriate. Of course, this is just a slight extension of the usual χ^2 test of independence formed from the corresponding 2×2 table. Tests with power against broader alternatives to independence may be constructed from the joint asymptotic distribution of χ_{ni}, χ_{nj}, etc., using several sample cut-points. Since the asymptotic correlation matrix of these χ's can be estimated consistently as in section 4.1 one could, for example, define a test statistic as the appropriate quadratic form with an asymptotic chi-squared distribution under the independence hypothesis.

The unscaled numerator of χ_{ni} for a sample point (X_i,Y_i) may be written as in section 2.1, viz., $H_{ni} - F_{ni}G_{ni}$, in terms of the sample cdf's. It may be noted that $\Sigma_i H_{ni}$ is a linear function of Kendall's $\hat{\tau}$ statistic and that $\Sigma_i F_{ni}G_{ni}$ is a linear function of Spearman's $\hat{\rho}_s$ statistic, from which the following relationship is deduced:

$$\Sigma_i(H_{ni} - F_{ni}G_{ni}) = \frac{n}{12}\left(3\hat{\tau} - \frac{n+1}{n-1}\,\hat{\rho}_s\right) \ .$$

A discussion of such omnibus statistics for assessing bivariate association may be found in Schweizer and Wolff (1981) and in Blum, Kiefer, and Rosenblatt (1961).

However, the approach we take is not to reduce the n sample points to a single statistic. Rather we let each sample point speak separately, in the context of the full sample, regarding the degree of association

between X and Y through its new co-ordinates $(\lambda_{ni},\tilde{\chi}_{ni})$ on the χ-plot. The approach is not unlike the examination of a P-P transformation of individual sample points as opposed to calculation of an omnibus goodness-of-fit statistic.

4.3 Behaviour of the χ-plot under non-independence

If X and Y are not independent then, for large n, the sequence of values of the data transform χ_{ni}, corresponding to the fixed cut-point (X_i,Y_i), will tend to some limiting value $\chi(F,G,H)$ in the interval $[-1,1]$, generally different from zero. For a given bivariate distribution this limiting value is not, in general, a function of $\lambda(F,G)$, the horizontal plotting axis. Rather, the limiting value of χ_{ni} will depend on $F(X_i)$ and $G(Y_i)$ separately so that, even for large n, the vertical scatter of the χ-plot at a fixed value of λ may not shrink to zero. Also, the vertical scatter of the χ-plot due to sampling variability will no longer have constant asymptotic standard deviation $1/\sqrt{n}$ when X and Y are not independent. In fact, this standard deviation will depend on the limiting value of χ_{ni} as well as the limiting marginal frequencies, viz.

$$n^{\frac{1}{2}}.SD(\chi_{ni}) = (1-\chi^2+v\chi)^{\frac{1}{2}} \quad \text{where}$$

$$v = \frac{4\{\frac{1}{2}-F(X_i)\}\{\frac{1}{2}-G(Y_i)\}}{[F(X_i)\{1-F(X_i)\}G(Y_i)\{1-G(Y_i)\}]^{\frac{1}{2}}} \quad .$$

For sample cut-points near the bivariate median, the vertical scatter of the χ-plot is multiplied by the approximate factor $(1-\chi^2)^{\frac{1}{2}}$.

One could define a three-dimensional version of the χ-plot in which the value of χ_{ni} is plotted against the pair of sample marginal frequencies (F_{ni},G_{ni}) rather than against $\lambda_{ni}(F_{ni},G_{ni})$. For example, the magnitudes of χ_{ni} could be represented by circles of varying diameter

with centres located at (F_{ni},G_{ni}); solid circles could be used for positive values and outline circles for negative values. Of course, one could simply plot the points (F_{ni},G_{ni}) on the unit square, but this loses the dependence information contained in the χ_{ni} values.

Alternatively, each data point (X_i,Y_i) could be represented by its observation number i both in the scatterplot and in the χ-plot, or more crudely, by one of the integers $1,\ldots,9$, depending on which third of the ordered X-values and of the ordered Y-values it belonged to. These suggestions relate to the proposals made in Example 8 and 9 for further investigation.

We are also considering extensions to the following situations: X and Y are each multivariate; pairwise plotting of association for three or more jointly distributed variables; X and Y have specified marginal distributions a priori as in Example 10. None of these extensions seems straightforward.

Acknowledgements

We are most grateful to Murray Cameron, Peter Diggle, William Dunsmuir and Ron Sandland for assistance in finding suitable data sets, and to Paul Tukey for some early discussions about the problem.

References

Ali, M.M. & Giaccotto, C. (1982). The identical distribution hypothesis for stock market prices - Location-and Scale-shift alternatives. *J.Am.Statist. Assoc.* 77, 19-28.

Blum, J.R., Kiefer J., and Rosenblatt, M., (1961). Distribution free tests of independence based on the sample distribution. *Ann.Math. Statist.* 32, 485-498.

Diggle, P.J. (1983). *Statistical Analysis of Spatial Point Patterns.* London: Academic Press.

Gnanadesikan, R. (1977). *Methods for Statistical Analysis of Multivariate Data.* New York: Wiley.

Olmstead, P.S. & Tukey, J.W. (1947). A corner test for association. *Ann.Math. Statist.* 18, 495-513.

Schweizer, B. & Wolff, E.F. (1981). On nonparametric measures of dependence for random variables. *Ann.Statist.* 9, 879-885.

Whittle, P. (1954). A statistical investigation of sunspot observations with special reference to H. Alfven's sunspot model. *Astrophysical J.* 120, 251-260.

Figure 1 (a) Data for Example 1; $\hat{\rho} = 0.0432$, $\hat{\rho}_s = 0.0124$
(b) χ-plot for data in Fig. 1a

Figure 2 (a) Data for Example 2; $\hat{\rho} = 0.1276$, $\hat{\rho}_s = 0.1650$
(b) χ-plot for data in Fig. 2a

Figure 3 (a) Data for Example 3; $\hat{\rho} = 0.6482$, $\hat{\rho}_s = 0.6434$
(b) χ-plot for data in Fig. 3a

Figure 4 (a) Data for Example 4; $\hat{\rho} = -0.0120$, $\hat{\rho}_s = 0.0050$
(b) χ-plot for data in Fig. 4a

Figure 5 (a) Data for Example 5; $\hat{\rho} = -0.0087$, $\hat{\rho}_s = 0.0038$
(b) χ-plot for data in Fig. 5a

Figure 6 (a) Data for Example 6; $\hat{\rho} = 0.0647$, $\hat{\rho}_s = 0.0383$
(b) χ-plot for data in Fig. 6a

Figure 7 (a) Data for Example 7; $\hat{\rho} = 0.0595$, $\hat{\rho}_s = 0.0596$
(b) χ-plot for data in Fig. 7a

Figure 8 (a) Data for Example 8; $\hat{\rho} = -0.1687$, $\hat{\rho}_s = -0.1864$
(b) χ-plot for data in Fig. 8a

Figure 9 (a) Data for Example 9; $\hat{\rho} = 0.1879$, $\hat{\rho}_s = 0.2188$
(b) χ-plot for data in Fig. 9a

Figure 10 (a) Data for Example 10; $\hat{\rho} = 0.0673$, $\hat{\rho}_s = -0.0600$
(b) χ-plot for data in Fig. 10a

Figure 11 (a) Data for Example 11; $\hat{\rho} = 0.0060$, $\hat{\rho}_s = 0.0229$
(b) χ-plot for data in Fig. 11a

Figure 12 (a) Data for Example 12; $\hat{\rho} = -0.0663$, $\hat{\rho}_s = -0.1022$
(b) χ-plot for data in Fig. 12a

Figure 13 (a) Data for Example 13; $\hat{\rho} = 0.00022$, $\hat{\rho}_s = -0.1002$
(b) χ-plot for data in Fig. 13a

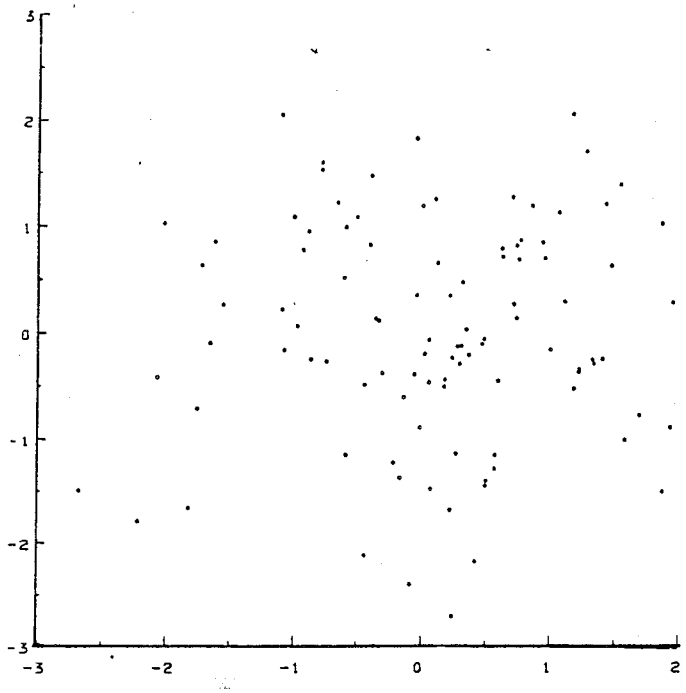

FIGURE 1a

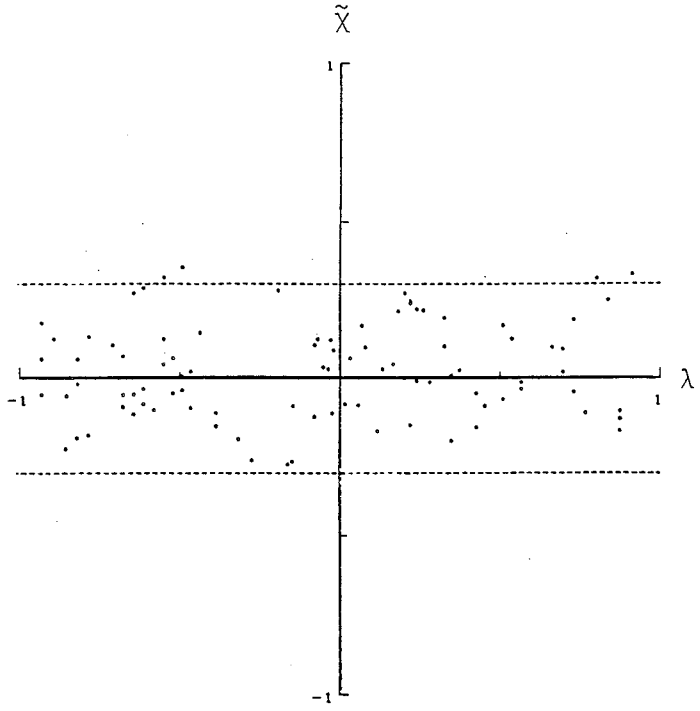

FIGURE 1b

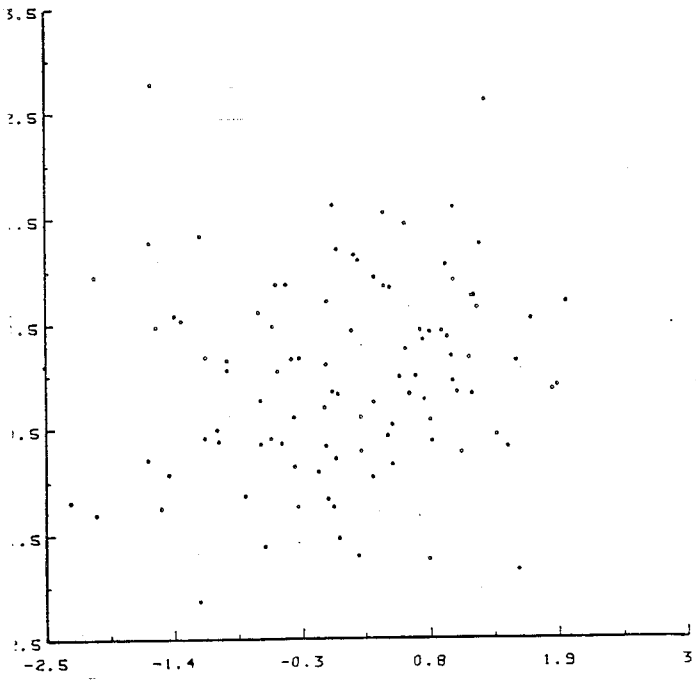

FIGURE 2a

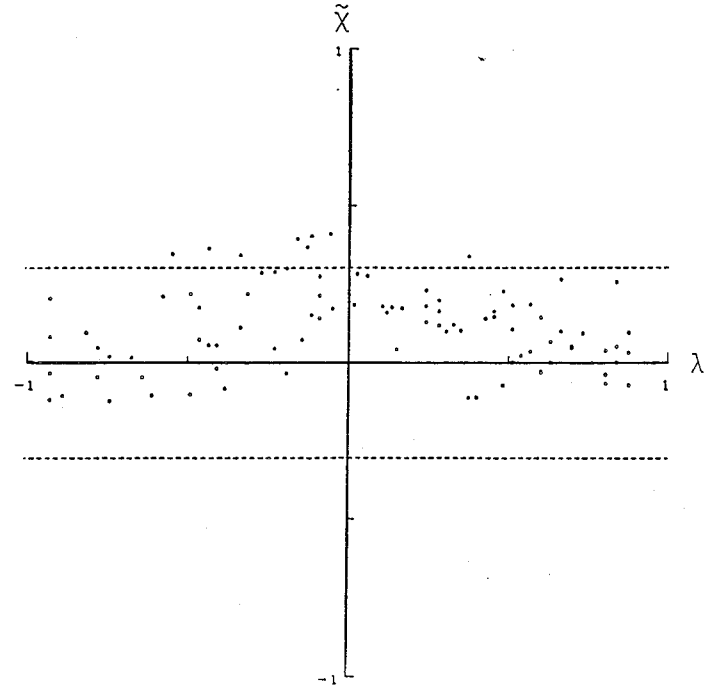

FIGURE 2b

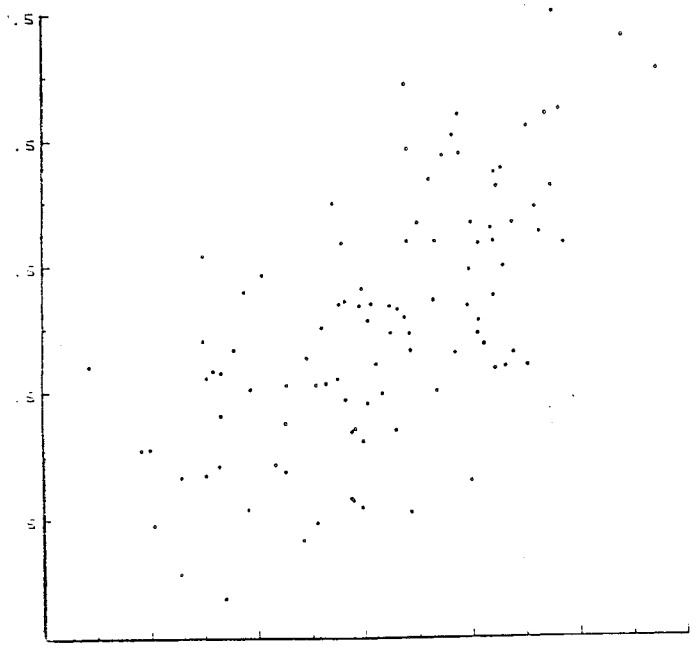

FIGURE 3a

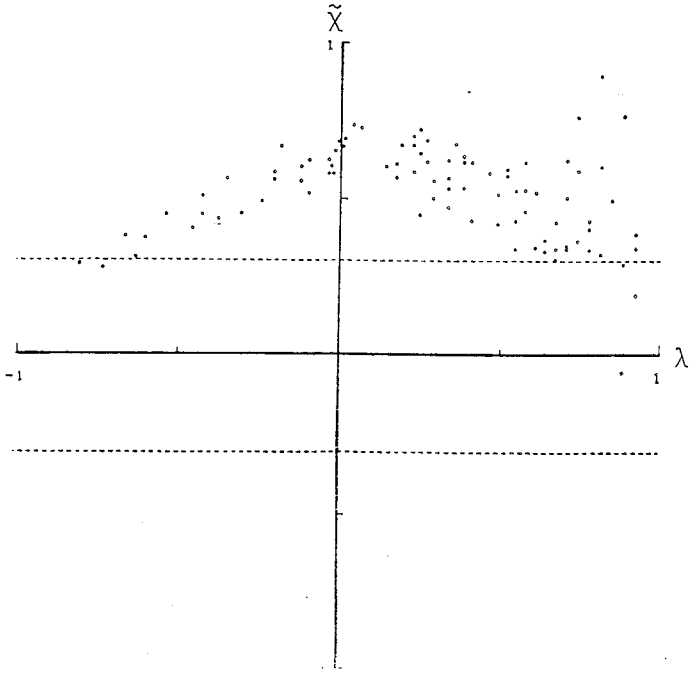

FIGURE 3b

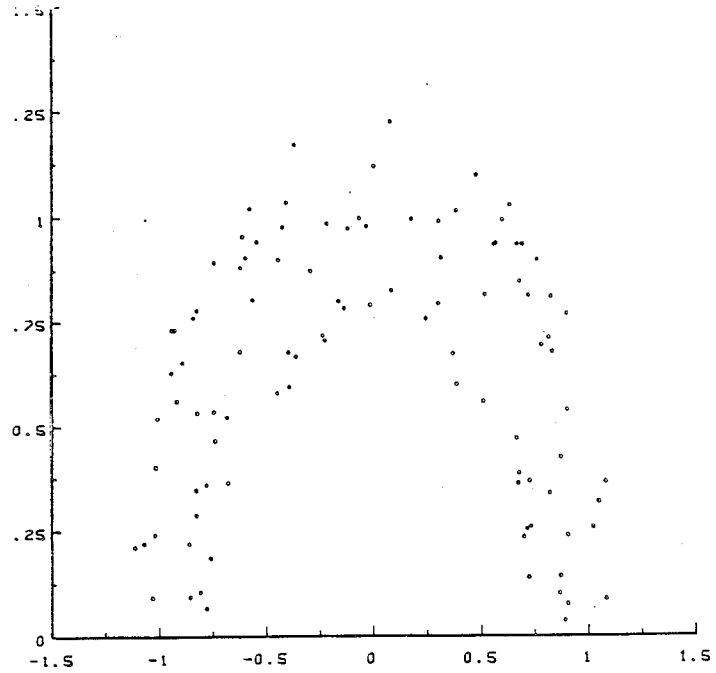

FIGURE 4a

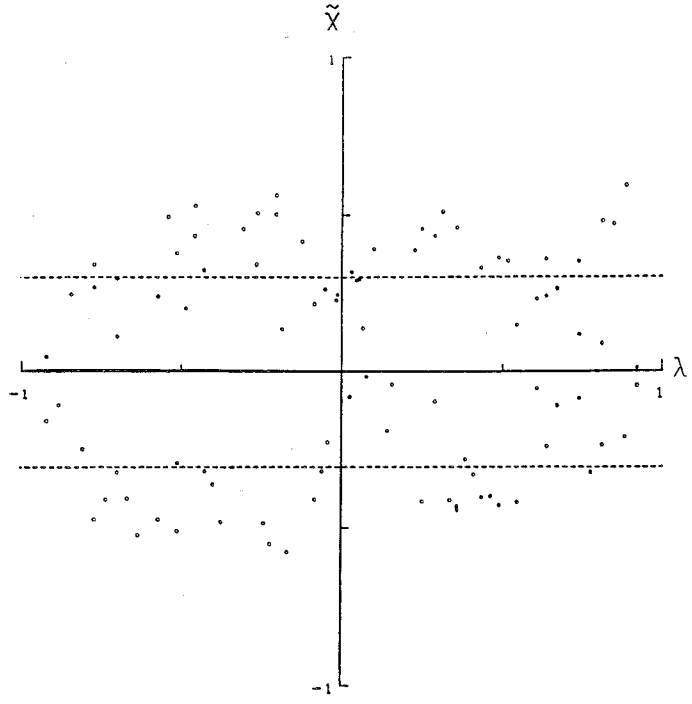

FIGURE 4b

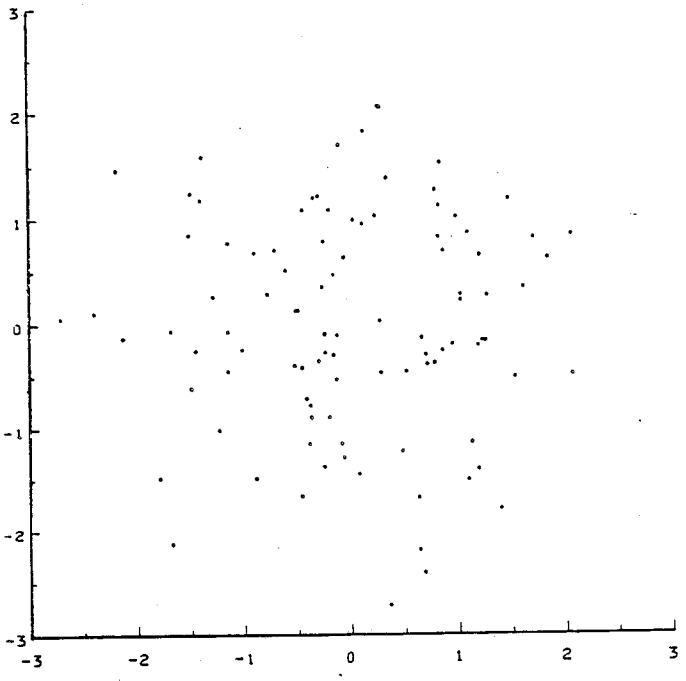

FIGURE 5a

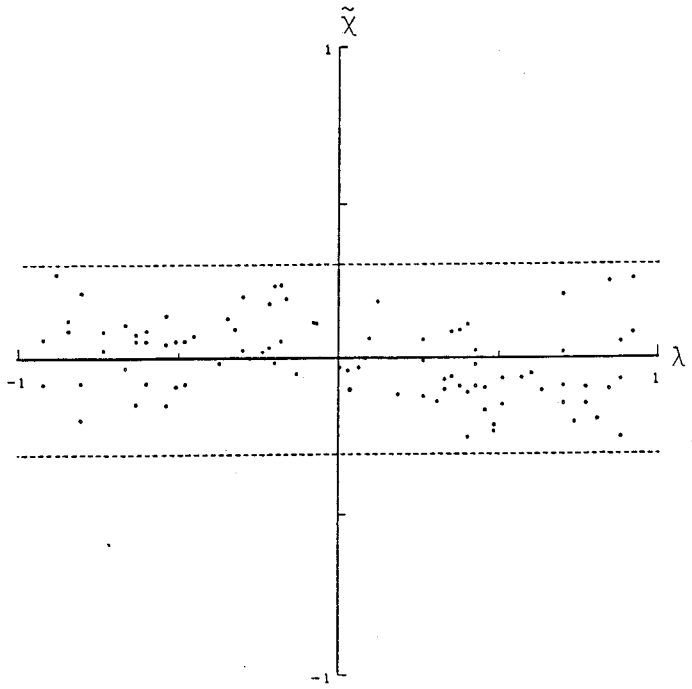

FIGURE 5b

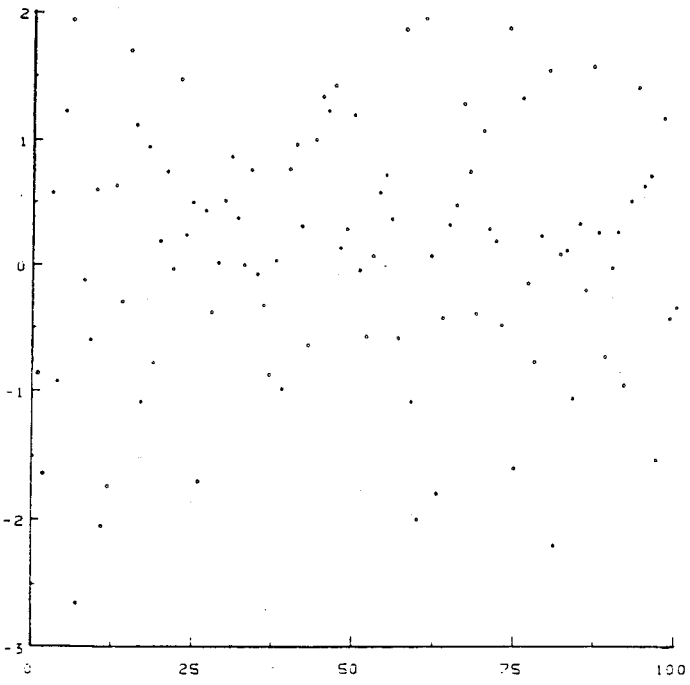

FIGURE 6a

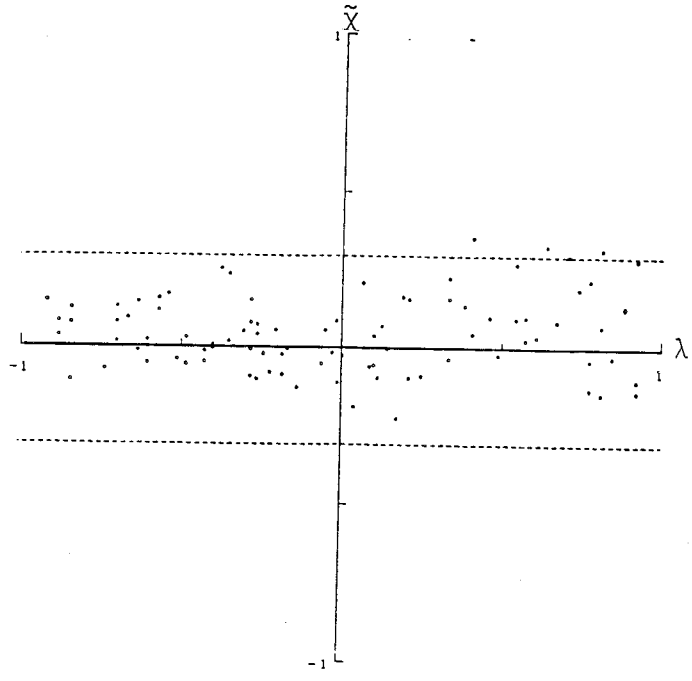

FIGURE 6b

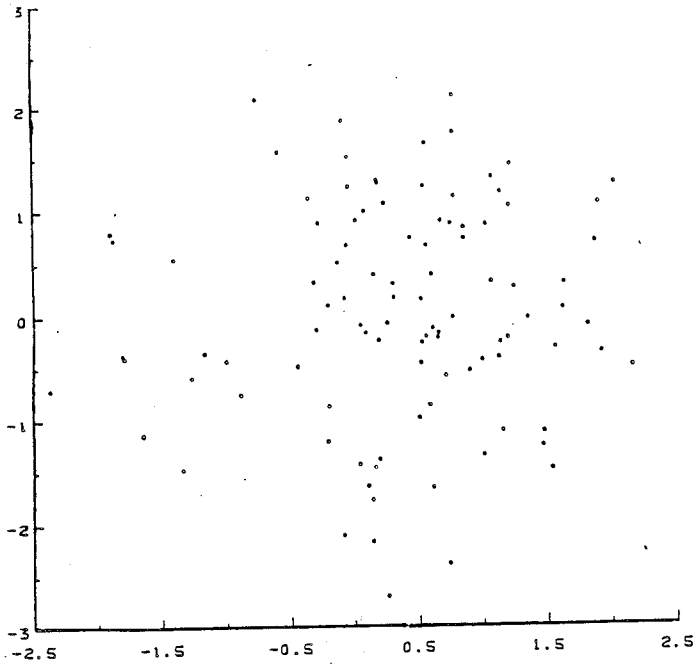

FIGURE 7a

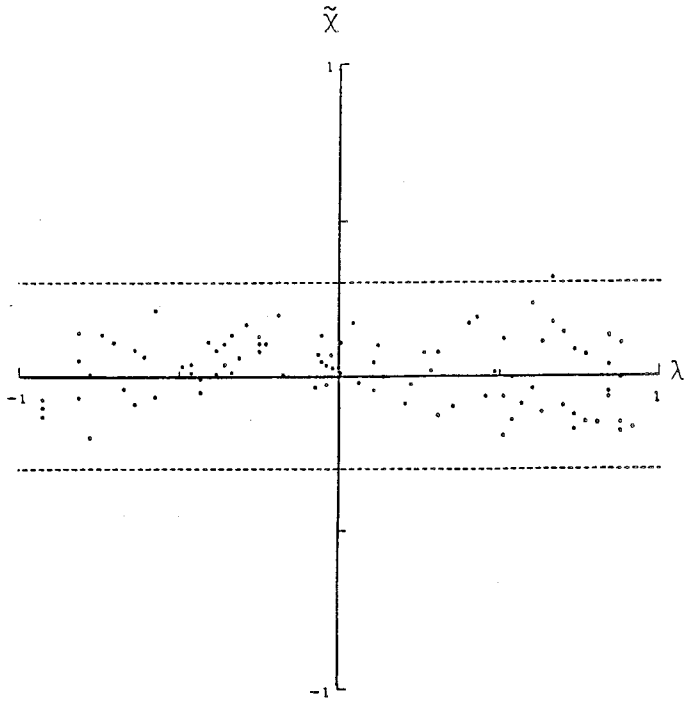

FIGURE 7b

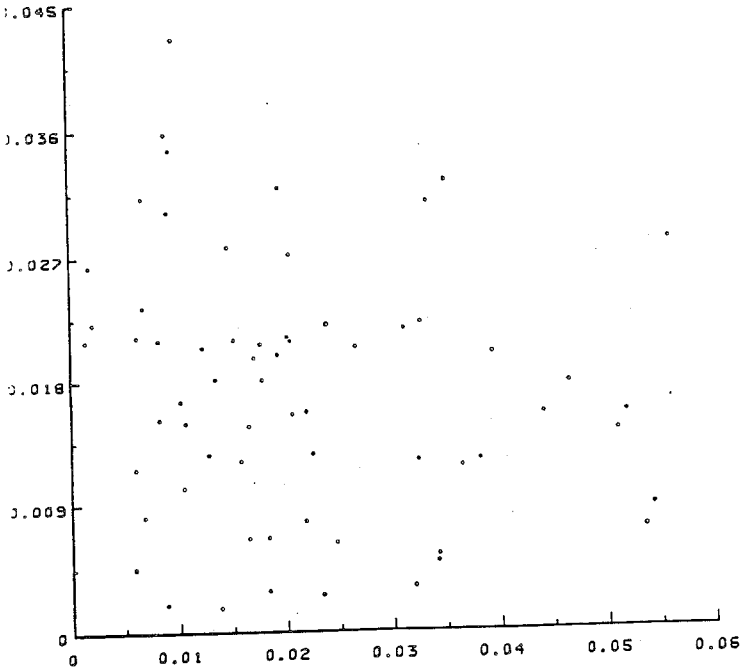

FIGURE 8a

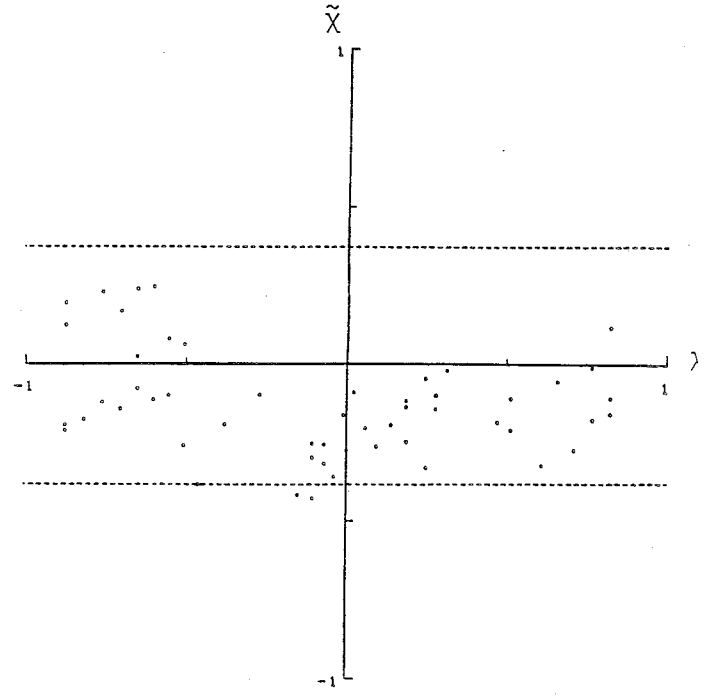

FIGURE 8b

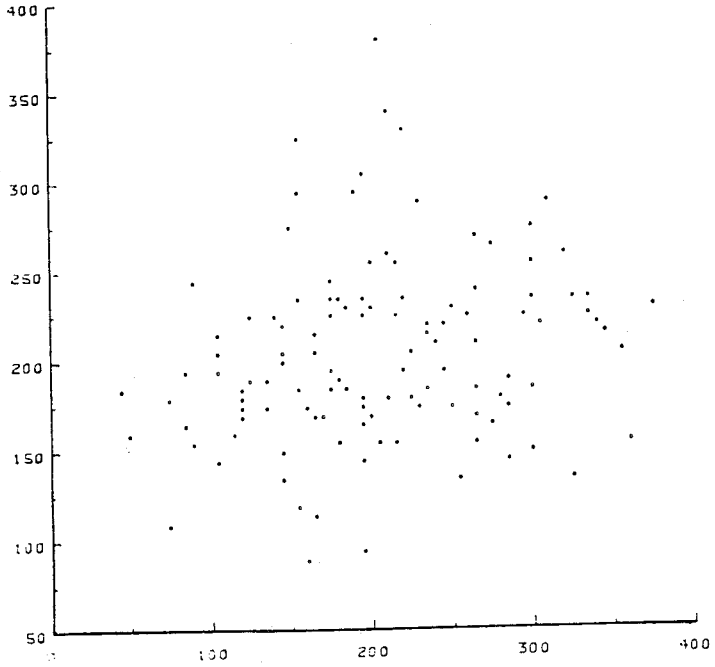

FIGURE 9a

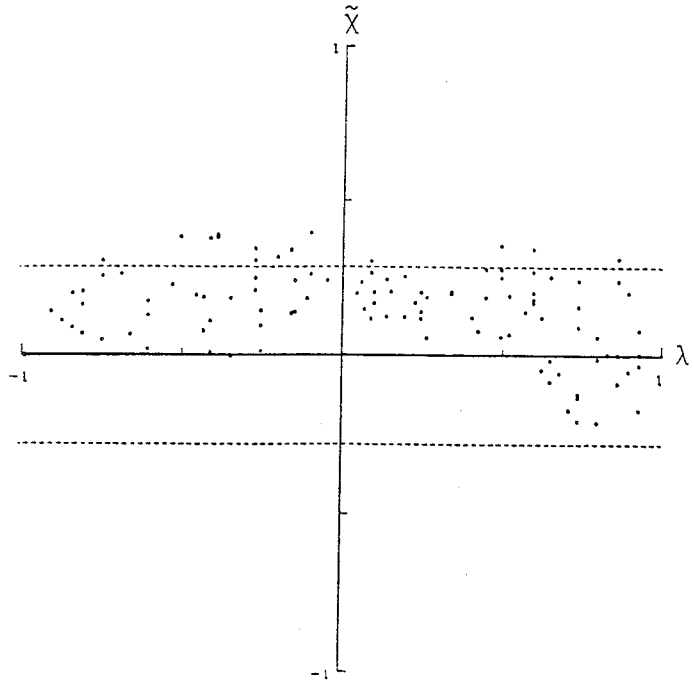

FIGURE 9b

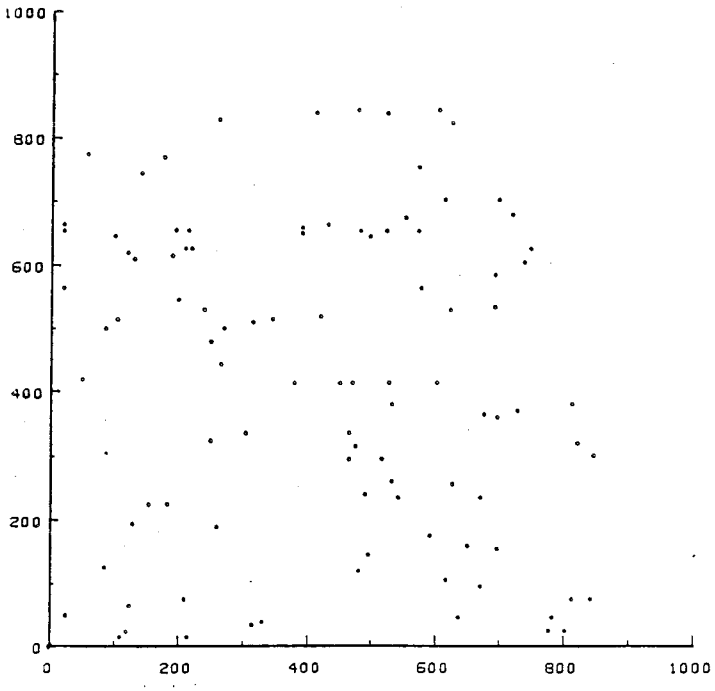

FIGURE 10a

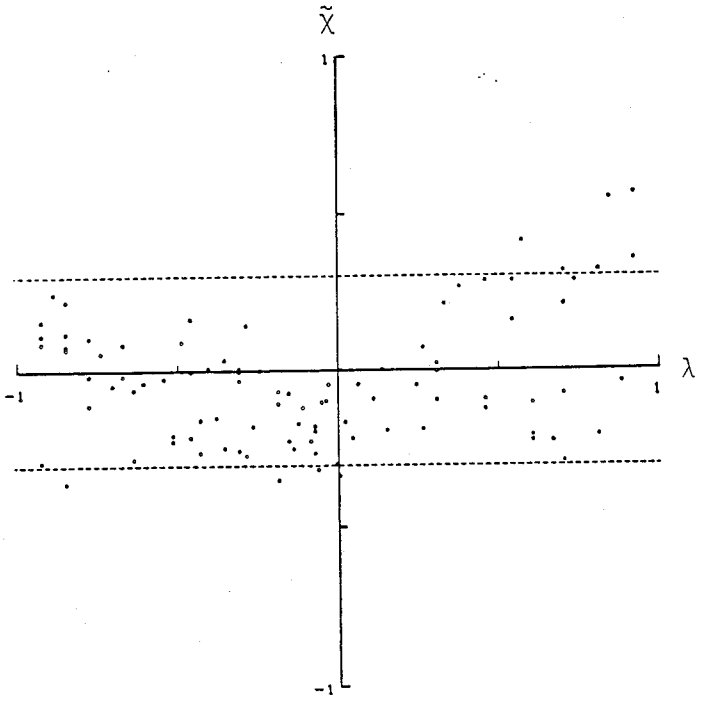

FIGURE 10b

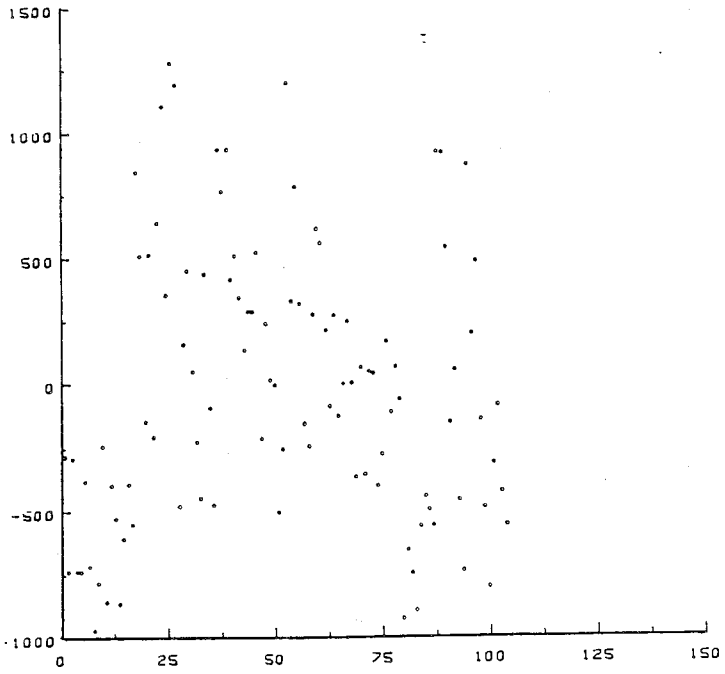

FIGURE 11a

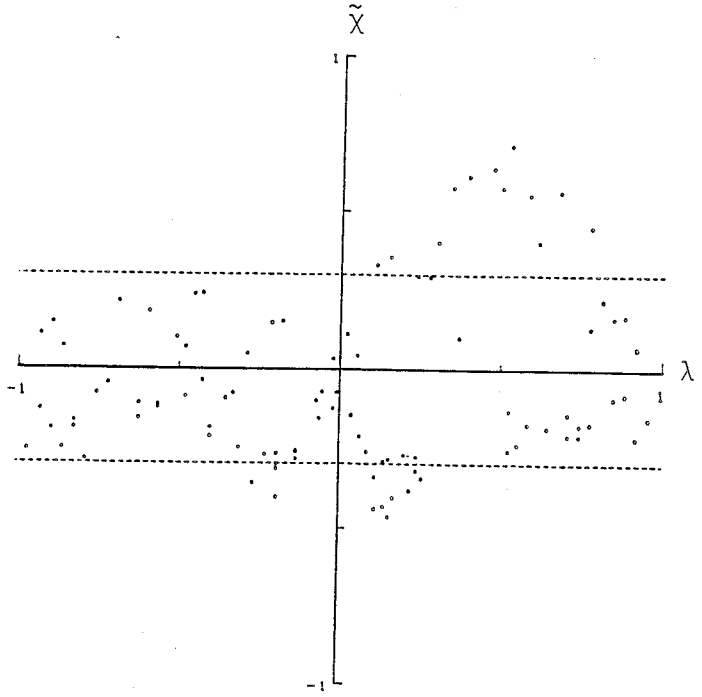

FIGURE 11b

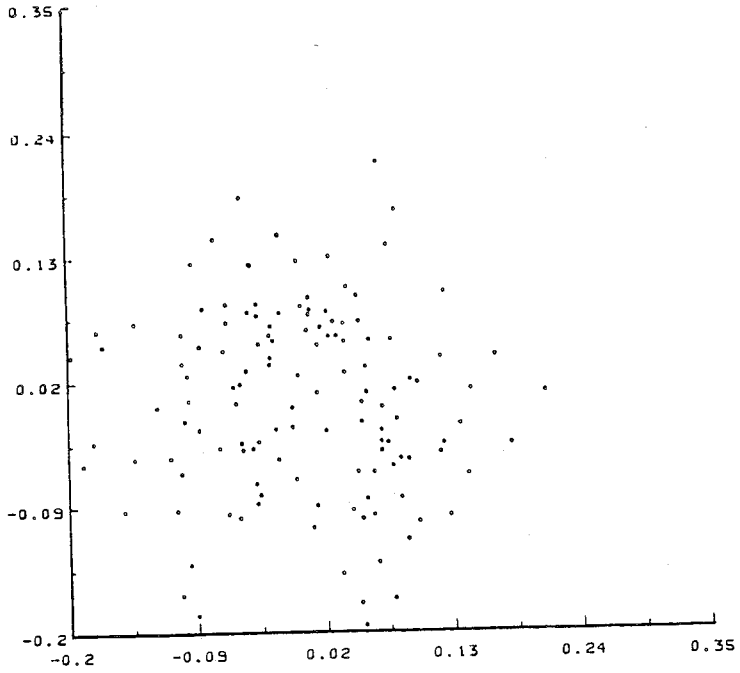

FIGURE 12a

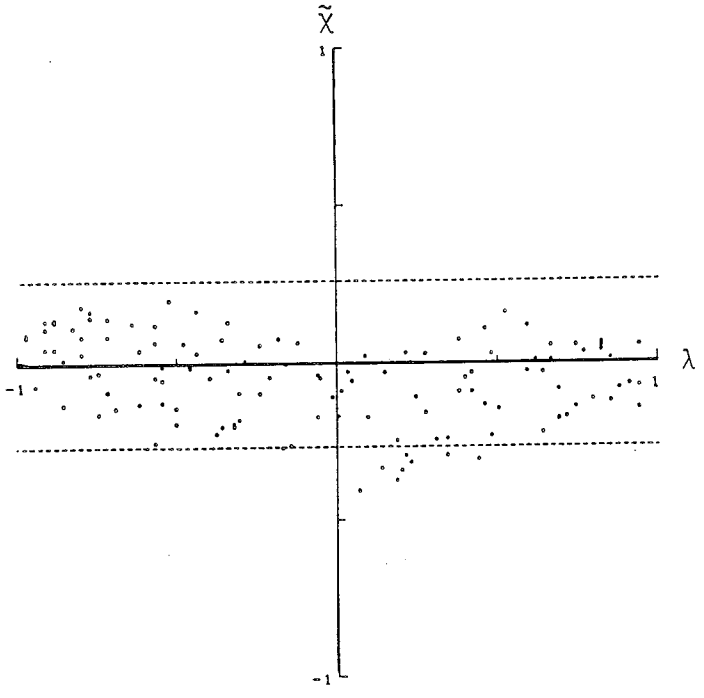

FIGURE 12b

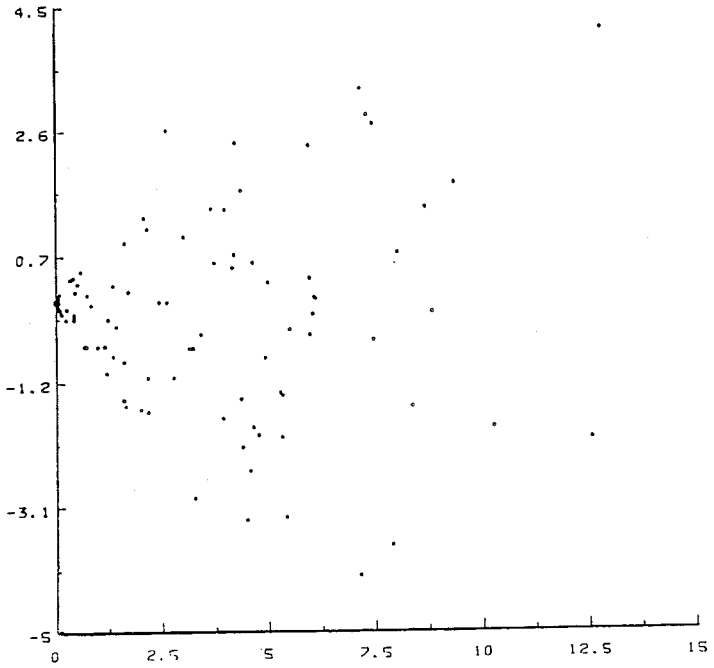

FIGURE 13a

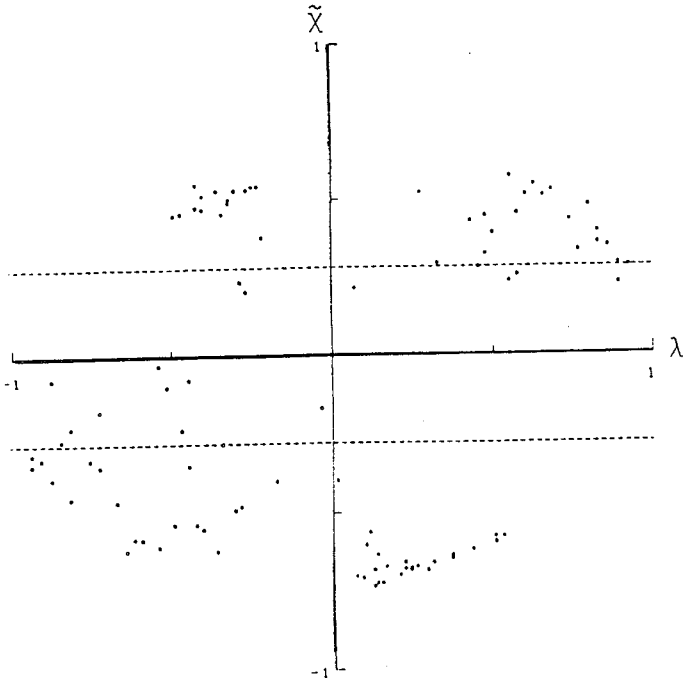

FIGURE 13b