# Envelope plots of ordered Mahalanobis distances: uses and efficient generation

Frank Critchley[1]*, Guobing Lu[2] and Richard A Atkinson[3]

[1] *Open University,* [2] *University of Bristol, and* [3] *University of Birmingham*

August 2005

*Address for correspondence*: Department of Statistics, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK. Email: F.Critchley@open.ac.uk

# ABSTRACT

Systematic departures from a reference sampling scheme, such as multivariate normality, can affect all the observed ordered Mahalanobis distances (OMDs) – not just a few extreme values – causing distinctive changes to their overall *pattern*. Envelope OMD plots provide a useful framework against which to 'read' such changes, often suggestive of further, confirmatory analyses. Illustrative examples include heavy-tailed and skew distributions and the presence of (clusters of) outliers. The exact distributions involved being intractable, the above framework is found by simulation. Fast, accurate approximations to it are also developed for use when such simulation is computationally prohibitive.

*Keywords*: DIAGNOSTIC PLOT; ENVELOPE PLOT; MULTIPLE OUTLIERS; MULTIVARIATE NORMAL DISTRIBUTION; ORDERED MAHALANOBIS DISTANCES.

# 1 INTRODUCTION

Let $x_1, \ldots, x_n$ be a sample from a $k$-variate distribution $F$, $\bar{x} := n^{-1} \sum x_i$ the sample mean and $\hat{\Omega} := n^{-1} \sum (x_i - \bar{x})(x_i - \bar{x})^T$ the empirical dispersion matrix, assumed nonsingular. Then, the Mahalanobis distance of each observation (from $\bar{x}$) is defined by

$$D_i := \sqrt{(x_i - \bar{x})^T \hat{\Omega}^{-1} (x_i - \bar{x})}. \tag{1.1}$$

A detailed description of the properties of the $\{D_i\}$ can be found in Mardia (1977).

The $\{D_i\}_{i=1}^n$ are widely suggested as basic tools for detecting outliers in multivariate data. For testing single outliers, they provide the most powerful test statistic under Ferguson's (1961) mean-slippage multivariate normal model. Indeed, in this case, $D_i$ is a strictly decreasing function of the likelihood ratio statistic (Wilks' Lambda) $\Lambda_i = \left|\hat{\Omega}(i)\right| / \left|\hat{\Omega}\right|$, where $\hat{\Omega}(i)$ denotes the sample dispersion matrix with $x_i$ deleted (Barnett and Lewis, 1998). If $D_{(1)} \leq D_{(2)} \leq \ldots \leq D_{(n)}$ denote the order statistics of the $\{D_i\}$, then $D_{(n)}$ can be used to identify the best candidate and, simultaneously, to test the mean-slippage hypothesis (Siotani, 1959; Press, 1972; Hawkins, 1980; Seber, 1984). However, when there are multiple outliers in the data, it is well-known that the $\{D_i\}$ suffer from the masking effect. That is, multiple outliers do not necessarily stand out as having the largest $D_i$ values, making their identification a challenging problem.

One way of tackling this problem is to form robust versions of $D_i$ by replacing $\bar{x}$ and $\hat{\Omega}$ with high-breakdown estimates. Various methods and algorithms have been proposed in the literature, including Rousseeuw (1985), Rousseeuw and Leroy (1987), Rousseeuw and van Zomeren (1990), Cook and Hawkins (1990), Hadi (1992), Atkinson and Mulira (1993), Atkinson (1994), Hawkins (1994), and Becker and Gather (1999, 2001).

Graphical techniques provide another approach. A variety of plots of, possibly transformed, observed ordered Mahalanobis distances (OMDs) against their typi-

cal (often, expected) values under multivariate normality have been proposed, as a check on both (I) this overall distributional assumption and (II) the possibility of outliers from it. Such plots are useful visual diagnostic tools that are quick to produce and provide a valuable initial diagnostic screen that can help guide subsequent data analysis. Under normality, $\{D_i^2\}_{i=1}^n$ are asymptotically independent $\chi_k^2$ random variables. Cox (1968) and Healy (1968) suggested either a $\chi_k^2$ probability plot for the $D_i^2$ or a normal probability plot for their cube roots. A slight improvement can be made if, instead, we base the probability plot on the Beta distribution, since

$$D_i^2 \sim (n-1)Beta(k/2, (n-k-1)/2) \tag{1.2}$$

(see Gnanadesikan and Kettenring (1972) and Small (1978), in both of which an unbiased estimate of $\Omega$ is used, changing the marginal distribution of $D_i^2$ accordingly). However, two drawbacks limit the application of such probability plots (Krzanowski and Marriott, 1994). The first is statistical dependence between the $\{D_i^2\}$, particularly marked for small sample sizes. The second is the masking problem described above.

Nevertheless, as we demonstrate, the overall *pattern* presented by the order statistics $\{D_{(i)}\}$ can reveal useful information about the structure of a multivariate data set. Either the existence of outliers, or systematic departure from a postulated distributional assumption, can affect all the $\{D_{(i)}\}$ – not just a few extreme values – causing a change in their overall pattern. In this sense, we might say that masking is a problem *because* attention is focused on just the largest $D_{(i)}$'s.

In this paper, we introduce enhanced, envelope plots of ordered Mahalanobis distances – *OMD envelope plots* for short – and illustrate their usefulness in a range of examples. These new plots add a framework against which to 'read' the pattern of the observed OMDs. Specifically, they add (accurate approximations to) the median, and upper and lower $\alpha/2$ points, of each $D_{(i)}$ under a reference sampling distribution. The distributions involved being intractable, this framework is ob-

tained by simulation. This framework respects, and reflects, the dependence among the $\{D^2_{(i)}\}$ inherent in, for example, the linear constraint:

$$\sum D^2_{(i)} = nk. \tag{1.3}$$

Section 2.1 describes the construction and use of OMD envelope plots, while Sections 2.2 and 2.3 give examples of (I) departures from underlying normality and (II) the presence of several forms of outlier, respectively, where these plots are helpful diagnostic indicators of appropriate further analyses. A well-known data set on Peruvian Indians is informatively re-analysed in Section 2.4. Section 3 provides fast, accurate approximations to the framework for use when simulation is computationally prohibitive.

# 2 THE OMD ENVELOPE PLOT

## 2.1 Construction and use

Consider the OMD statistics $\{D_{(i)}\}_{i=1}^n$ for a random sample of size $n$ from a specified $k$-variate distribution $F$. Let $m_i$ denote the median of $D_{(i)}$, and $l_i(\alpha)$ and $u_i(\alpha)$ its lower and upper $\alpha/2$ quantiles, respectively, for given $\alpha \in (0,1)$. The framework plot referred to above comprises the obvious linear interpolation of the three point objects

$$\{(m_i, l_i(\alpha)\}_{i=1}^n, \quad \{(m_i, m_i)\}_{i=1}^n \quad \text{and} \quad \{(m_i, u_i(\alpha)\}_{i=1}^n,$$

producing the straight line through the origin with unit slope, bounded above and below by strictly increasing, piecewise linear curves.

The $\{D_{(i)}\}$ being invariant under nonsingular affine transformation, it is natural to focus attention on distributions $F$ from families $\mathbb{F}$ closed under such transformations. In this case, by invariance, there is no loss in restricting attention to *standardised* distributions $F$, greatly reducing the burden of simulation. In par-

ticular, when $\mathbb{F} = \mathbb{N}_k$ comprises the nonsingular $k$-variate normal distributions, it suffices to consider the joint distribution of the $\{D_{(i)}\}$ induced by $F = N_k(0, I)$, the resulting framework depending, then, only on $n$, $k$ and $\alpha$. In the examples below, $\alpha = 0.05$ and each framework plot is generated from $10,000$ simulations, so that the estimated quantiles are essentially exact. For given data $\{x_i\}_{i=1}^n$, the values of $D_{(i)}$ are computed and the object $O_{data} := \{(m_i, D_{(i)})\}_{i=1}^n$ sent to the above framework plot. Throughout, the points of $O_{data}$ are denoted by a disk symbol '$\circ$'.

The graph obtained in this way is called the *OMD $F$-envelope plot* of the data. It plays a similar rôle to Atkinson's (1981) envelope plot. In the 'null' case where the $\{x_i\}$ are indeed a sample from $F$, we may expect – dependence issues aside – the $n$ points in $O_{data}$ to lie scattered about the central line in the framework plot, only a proportion $\alpha$ of them falling outside the region bounded by its lower and upper curves. In contrast, when the data depart systematically from the reference sampling scheme, we may expect corresponding departures from such a null overall pattern. This has been verified in many simulated data sets, examples of which are given below. In particular, the pattern of $O_{data}$ varies with both (I) the true underlying distribution and (II) the proportion of outliers, the number of outlier clusters, and the separation of outlier clusters from the bulk of the data. In common with many graphical methods, the power of the OMD envelope plot stems primarily from the human eye's ability to distinguish different point patterns. Possible interpretations of such patterns can then be checked by appropriate, follow-up analyses.

## 2.2 Checks on consistency between data and postulated distribution

### Example 1: Null Case

The OMD normal-envelope plot of a multivariate standard normal sample with $n = 60$ and $k = 2$ is shown at Figure 1. There is no evidence here to reject normality.

Note also the increase in both the spread $(u_i(\alpha) - l_i(\alpha))$ and the positive skewness $(u_i(\alpha) - m_i) - (m_i - l_i(\alpha))$ of $D_{(i)}$ as $i$ increases towards $n$.

When the data comprise a sample from a distribution with features importantly different from those of the postulated distribution $F$, the OMD $F$-envelope plot tends to have an overall different *shape*, a pattern which becomes clearer (less variable) as sample size increases. We illustrate this with examples when $F$ is normal but the data are either heavy-tailed or skew.

**Example 2: Data from a heavy-tailed distribution**

Here the data, shown in Figure 2(a), comprise a sample of 60 cases from a bivariate $t$-distribution with 4 degrees of freedom (Johnson and Kotz, 1972). Their OMD normal-envelope plot (Figure 2(b)) shows a markedly non-null pattern, the quadratic nature of $O_{data}$ reflecting the fact that the normal quantile attained by $D_{(i)}$ first decreases and then increases quite systematically with $i$, $D_{(i)}$ falling outside its $(1 - \alpha)$-central normal bounds a disproportionately large number of times. To aid interpretation, four symbols are linked between Figures 2(a) and 2(b), according as the corresponding OMD falls above or below the median, and inside or outside the normal-envelope.

It is therefore reasonable to reject the null hypothesis that the data are a normal sample, and to consider a heavier-tailed alternative. This tentative diagnosis is supported by Figure 3, where the sample size has been increased to 500 and the same ($\sim$ asymptotic) pattern of departure found.

Finally, it is confirmed by a follow-up analysis in which a bivariate $t$-distribution is fitted to the data, and the OMD envelope plot for *that* reference distribution produced. The result is displayed in Figure 2(c) which looks quite null, all but two points lying within the range between lower and upper curves, with no particular pattern discernable.

**Example 3: Data from a skew distribution**

Consider Gumbel's (1960) bivariate exponential distribution with distribution function $F_\theta(x, y) = 1 - e^{-x} - e^{-y} + e^{-(x+y+\theta xy)}$, $(x > 0, y > 0)$, where $\theta \in [0, 1]$ is a parameter. One attractive nature of this distribution is that the marginal distributions are exponential. Figure 4(a) shows 60 cases drawn from the distribution with $\theta = 1$, and Figure 4(b) its OMD normal-envelope plot.

Again, there is clear evidence to reject normality, the lower $O_{data}$ points lying about a line with slope less than 1, straying eventually below normal sampling limits, higher points then jumping suddenly, and roughly linearly, above these limits. Four symbols are linked, as before, between Figures 4(a) and 4(b) to aid interpretation. This same pattern is found asymptotically (Figure 5) and may indicate an underlying skew distribution, a suggestion again confirmed by fitting such a distribution to the data and using its OMD envelope plot (Figure 4(c)).

## 2.3 Checks on outliers

**Example 4: Data with a Light Outlier Cluster**

In practice, data sets frequently contain at least a small proportion of outliers from a majority distribution $F$, causing corresponding departures from null OMD $F$-envelope plots. For small proportions ('light' outlier clusters), a few $D_{(i)}$ with the largest indices become unduly large. In view of the overall constraint (**??**), this in turn depresses the $D_{(i)}$ values taken for smaller values of $i$. The pattern is rather different for 'heavy' outlier clusters (Example 5). We illustrate these patterns for simulated data sets with $F = N_2(0, I)$, the total sample size $n = 60$ containing $n_{out} = 1$, 5 or 25 outliers.

Figure 6 shows the scatter plot and OMD normal-envelope plot when $n_{out} = 1$, the single outlier being denoted by '$\times$'. In this envelope plot, $(m_n, D_{(n)})$ lies far above the upper curve. Although the other points are still inside the bounded region, they lie scattered – not around the median straight line – but around the

line with the same slope yet slightly lower intercept. This typically indicates the existence of a single outlier in the data, a diagnosis confirmed by deleting the case indicated and updating the OMD normal-envelope plot accordingly (not shown). Of course, in this situation, the usual test statistic $D^2_{(n)}$ also performs well.

Figures 7(a) and 7(b) display the corresponding plots when $n_{out} = 5$, the outliers (again denoted by '$\times$') being drawn from a mean-shift normal distribution. The masking effect of multiple outliers emerges clearly from these plots. The departure of the outlier cluster from the majority of the data is appreciable, but the value of the largest Mahalanobis distance lies within the normal range, so that the test statistic $D^2_{(n)}$ breaks down. Nevertheless, the OMD envelope plot contains useful information, clearly rejecting normality. Indeed, by extension of the single outlier case, five (or six) suspected outliers with the largest OMDs can be picked out, a diagnosis confirmed in the correspondingly updated OMD normal-envelope plot (Figure 7(c)).

**Example 5: Data with a Heavy Outlier Cluster**

A heavy outlier cluster changes the OMD pattern in a rather different way. The $O_{data}$ points tend to fluctuate around a line with slope less than 1 (this slope decreasing the higher the proportion of outliers), straying outside normal sampling limits for, especially, small values of $i$. Informally, the smallest distances are too big and the largest distances too small, reflecting the presence of a 'gap' between clusters of cases.

Figure 8 shows the same two plots as before, but now with $n_{out} = 25$. These cases are again highlighted across the linked plots, displaying a stronger masking effect: there is now no way to discern two subsamples in Figure 8(a). Nevertheless, this OMD envelope plot clearly rejects normality, its distinctive linear pattern (described above) suggesting the possibility of clustering in the data, a diagnosis confirmed by a hierarchical cluster analysis (not shown).

**Example 6: Data with two Outlier Clusters**

The 5-dimensional data set considered here consists of three clusters: the bulk of the data (size 43) are from the standard normal population, the other two outlier clusters, of sizes 15 and 2 respectively, are from mean-shift normal distributions. Figure 9(a) displays their scatterplot-matrix. Two points stand out in the OMD normal-envelope plot for the full data set (Figure 9(b)), suggesting that they are outlying from the rest. Moreover, the other points in this plot fluctuate about a line with slope less than 1, suggesting that further departures from normality may be present. Deleting the two cases indicated and considering the updated OMD normal-envelope plot (Figure 9(c)) does, indeed, suggest a heavy-outlier pattern for the remaining 58 cases – a possibility which, again, can be confirmed by cluster analysis.

## 2.4 Diagnosis in practice

**Example 7: Peruvian Indian Data**

This real data set (Ryan et al., 1976) comprises 39 bivariate observations, weight and height, collected in connection with the study of long-term environmental changes on the blood pressure of Peruvian Indians. It has been used by several authors to illustrate checks of normality. An anomalous observation (case 39) is clear both from a scatter plot of the data (Seber, 1984) and from a $\chi^2$ probability plot of OMDs (Krzanowski and Marriott, 1994). Removing this case, Krzanowski and Marriott (1994) reject normality of the remaining 38 cases due to curvature in the centre of the updated $\chi^2$ probability plot, while subsequent analysis leads Seber (1984, p. 144 and p. 154) to query their normality, attributing this to 'a cluster of taller Indians with lower than average weight' identified by him in the scatter plot.

Here, we display the OMD normal-envelope plot for the full data set in Figure 10(a). The point $(m_{39}, D_{(39)})$ lies far above the upper bound curve, separately from

other points, indicative of a very significant outlier. Moreover, the other points in this diagram also have clear non-null structure, all but the smallest fluctuating about a line with slope less than 1, a pattern which remains in the updated plot (Figure 10(b)) after deletion of the indicated outlier. Normality of the remaining 38 cases is clearly rejected, with the suggestion of a follow-up cluster analysis, in agreement with Seber.

# 3 APPROXIMATING THE FRAMEWORK PLOT

We focus attention here on the generation of framework plots when $F = N_k(0, I)$. A straightforward method to generate such plots is to directly simulate the desired quantiles, as above. However, especially in view of the matrix inversion involved, this may be prohibitive computationally when $n$ and/or $k$ is large, or when multiple envelope plots are required.

Accordingly, we develop here accurate, easily computed, approximations to the distribution of the $\{D_{(i)}\}$ under normality. We recall first some basic properties of the $\{D_i^2\}$ in this case, including (**??**) and (**??**) above, before analysing a variety of approximations.

## 3.1 Basic properties of the $\{D_i^2\}$ under normality

Under normality, the $\{D_i^2\}$ satisfy:

**P1** $\sum D_i^2 = nk$.

**P2** $\{D_i^2\}$ are (finitely) exchangeable random variables.

**P3** The marginal distribution of each $D_i^2$ is $(n-1)Beta(k/2, (n-k-1)/2)$.

**P4** Asymptotically, $\{D_i^2\}$ are independent $\chi_k^2$ random variables.

Note that properties P1 and P2 do not depend on normality of $F$ and that together they imply:

$$E(D_i^2) = k \text{ for each } i, \text{ while } \rho(D_i^2, D_j^2) = -(n-1)^{-1} \text{ for all } i \neq j,$$

so that the $\{D_i^2\}$ are asymptotically uncorrelated. Of course, the marginal distribution in P3 does depend on normality and, in turn, implies the asymptotic marginal $\chi_k^2$ distribution in P4.

For notational convenience, we denote $(D_1^2, \ldots, D_n^2)^T$ by $e = (e_1, \ldots, e_n)^T$. Our aim is to find a random vector $\hat{e} = (\hat{e}_1, \ldots, \hat{e}_n)^T$ whose distribution is close to that of $e$ – having as many of the properties P1 – P4 as possible – while $\hat{e}$ itself is quick to simulate from.

Results for four such approximations, described next, are displayed in Figures 11 and 12, for $(n, k) = (60, 2)$ and $(n, k) = (20, 5)$ respectively, (these values being chosen to represent, informally, 'just-asymptotic' and 'sub-asymptotic' situations). In each of these plots, $\alpha = 0.05$ and the solid curves are the essentially exact normal framework plot based on $10,000$ simulations, as above. The fast approximations, based on just 500 simulations, are plotted as $O_{data}$ whose points are, again, denoted by '$\circ$'.

## 3.2 Asymptotic approximations

Taking $\{\hat{e}_i\}$ to be i.i.d. as some $G$ on $(0, \infty)$ ensures P2, while taking (A1) $G = \chi_k^2$ or (A2) $G = (n-1)Beta(k/2, (n-k-1)/2)$ guarantees P4, the second of these delivering P3 too. However, in both cases, P1 is sacrificed: unlike $e$, whose distribution is degenerate (having support in the hyperplane $\overline{e} = k$), $\hat{e}$ is free to take values in a region of positive volume in $(0, \infty)^n$.

**A1: Chi-square Approximation**

As Figures 11(a) and 12(a) show, this approximation results in a relatively poor

framework plot. This is especially so in the second of these, both tail quantiles being then quite poorly approximated for all but the very smallest $\hat{e}_{(i)}$.

**A2: Beta Approximation**

The plots improve somewhat when the Beta distribution is used: see Figures 11(b) and 12(b). However, for $n$ small compared to $k$, ignoring the constraint P1 causes appreciable discrepancies in both tail quantiles for central values of $i$.

## 3.3 Dirichlet-type transformations

The following Dirichlet-type (for short, D-type) transformation is widely used in generating symmetric multivariate distributions (Fang, Kotz & Ng, 1990). Let the elements of $\xi = (\xi_1, \ldots, \xi_n)^T$ be i.i.d. as some $G$ on $(0, \infty)$ with mean $k$ and let

$$\eta = \eta(\xi) := \frac{k\xi}{\overline{\xi}}$$

with $\overline{\xi} := n^{-1} \sum \xi_i$. Then $\eta$ satisfies P1 and P2 while, under mild conditions, the $\{\eta_i\}$ are asymptotically independent, each having the same limiting distribution as $G$. Thus, taking (A3) $G = \chi^2_k$ or (A4) $G = (n-1)Beta(k/2, (n-k-1)/2)$ again guarantees P4. However, P3 is sacrificed in both cases.

The resulting D-type Chi-square and Beta approximations (A3) and (A4) show much improvement over their untransformed counterparts (A1) and (A2): see panels (c) and (d) in Figures 11 and 12. The D-type Beta approximation (panels (d)) performs particularly well, even when $(n, k) = (20, 5)$.

## 3.4 Numerical comparison

These approximations can be compared numerically using the following two measures of accuracy. Let $\hat{q}_i$ be the estimates of $q_i$ , where $q_i$ represents $l_i, m_i$ or $u_i$ and

put

$$S_q^{(1)} := n^{-1} \sum \left| \frac{\hat{q}_i - q_i}{q_i} \right| \text{ and } S_q^{(2)} := \sqrt{n^{-1} \sum \left( \frac{\hat{q}_i - q_i}{q_i} \right)^2}.$$

Clearly, both are measures of overall deviation, but on different scales.

Table 1 gives the results of 500 simulations for each approximation with $n = 60$, $k = 2$ and $\alpha = 0.05$. The accuracies of all four estimates of the median are broadly comparable, but those of the tail quantiles are much improved by Dirichlet transformation.

### 3.5 Recommended approximation

These numerical summaries and plots indicate that, given P4, accurate approximation is better served by requiring P1 and P2 rather than P3. For completeness, we note that neither Dirichlet transform method is improved by the variance correction procedure described in the Appendix. Overall, the D-type Beta method is our approximation of choice.


### Acknowledgements

We are grateful to Dong Qian Wang for helpful discussions, and to the EPSRC for support under research grant GR/K08246.


## APPENDIX: A note on D-type approximation

Let now $\widehat{e}_i$ and $\widetilde{e}_i$ denote the general member of the approximations (A3) and (A4) respectively, so that each has the same mean, $k$, as $e_i$. However, neither of their variances coincides with that of $e_i$, denoted $\sigma^2 = \sigma^2(n,k)$ and implicit in P3. For, using $\widehat{e}_i \sim (nk)Beta(k/2,(n-k-1)/2)$, $var(\widehat{e}_i) = (1+\widehat{r}^2)\sigma^2$ where

$$\widehat{r}^2 = \frac{n(k+1-2/k)+(1+2/k)}{(n+2/k)(n-k-1)}.$$

The exact distribution of $\widetilde{e}_i$ is unknown, but its variance can be approximated. An

asymptotic analysis gives $var(\widetilde{e}_i) = (1 - \widetilde{r}^2)\sigma^2 + O(n^{-2})$ where

$$\widetilde{r}^2 = (1 + 2/k)\frac{(n-1)(n^2-2n+9)}{n^2(n+1)(n+3)}.$$

Thus, we have variance inflation for (A3) and, asymptotically, variance deflation for (A4). Weighted methods can be applied to adjust the variance in both cases. For example, putting

$$\widehat{e}^* := \lambda\widehat{e} + (1-\lambda)\widehat{f}, \;\; \lambda \in (0,1),$$

where $\hat{e}$ and $\hat{f}$ are independent D-type Chi-square approximation vectors, retains P1 and P2, achieving $var(\widehat{e}_i^*) = \sigma^2$ by appropriate choice of $\lambda$. However, simulation results presented in Figure 13 show that this weighted approximation has a poorer performance for small $i$ than its unweighted counterpart reported in Figure 11(c).

## REFERENCES


Atkinson, A.C. (1981) Two graphical displays for outlying and influential observations in regression, *Biometrika*, **68**, 13-20.

Atkinson, A.C. and Mulira, H.M. (1993) The stalactite plot for the detection of multivariate outliers, *Statistics and Computing*, **3**, 27-35.

Atkinson, A.C. (1994) Fast very robust methods for the detection of multiple outliers, *J. A. Statist. Assoc.*, **89**,1329-39.

Barnett, V. and Lewis, T. (1998) *Outliers in Statistical Data*, 4th edition, Wiley, New York.

Becker, C. and Gather, U. (1999) The masking breakdown point of multivariate outlier identification rules, *J. A. Statist. Assoc.*, **94**, 947-955.

Becker, C. and Gather, U. (2001) The largest nonidentifiable outlier: a comparison of multivariate simultaneous outlier identification rules, *Computational Statistics and Data Analysis*, **36**, 119-127.

Cook, R.D. and Hawkins, D.M. (1990) Comment on Rousseeuw and van Zomeren (1990). *J. A. Statist. Assoc.*, **85**, 640-644.

Cox, D.R. (1968) Notes on some aspects of regression analysis, *J. R. Statist. Soc.*, A **131**, 265-79.

Fang, K.T., Kotz, S. and Ng, K.W. (1990) *Symmetric Multivariate and Related Distributions*, Chapman and Hall, London.

Ferguson, T.S. (1961) On the rejection of outliers, In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 253-87.

Gnanadesikan, R. and Kettenring, J.R. (1972) Robust estimates, residuals and outlier detection with multiresponse data, *Biometrics*, **28**, 81-124.

Gumbel,E.J. (1960) Bivariate exponential distributions, *J. A. Statist. Assoc.*, **55**, 698-707.

Hadi, A.S. (1992) Identifying multiple outliers in multivariate data, *J. R. Statist. Soc.* B **54**, 761-71.

Hawkins, D.M. (1980) *Identification of Outliers*, Chapman and Hall, London.

Hawkins, D.M. (1994) The feasible solution algorithm for the minimum covariance determinant estimator in multivariate data, *Computational Statistics and Data Analysis*, **17**, 197-210.

Healy, M.J.R. (1968) Multivariate normal plotting, *Appl. Statist.*, **17**, 157-61.

Johnson, N.J. and Kotz, S. (1972) *Distributions in Statistics: Continuous Multivariate Distributions*, Wiley, New York.

Krzanowski, W.J. and Marriott, F.H.C. (1994) *Multivariate Analysis*, Part I, Distributions, ordination and inference, Edward Arnold.

Mardia, K.V. (1977) Mahalanobis distances and angles, In *Multivariate Analysis (IV)*, ed. Krishnaiah, P.R., pp. 495-511.

Press, S.J. (1972) *Applied Multivariate Analysis*, Holt, Rinehart and Winston, New York.

Rousseeuw, P.J.(1985) Multivariate estimation with high breakdown points, in *Mathematical Statistics and Applications*, eds. W. Grossmann, G. Pflug, I. Vincze and W. Wertz, Dordrecht: Reidel, pp. 283-297.

Rousseeuw, P.J. and Leroy, A.M. (1987) *Robust Regression and Outlier Detection*, Wiley, New York.

Rousseeuw, P.J. and van Zomeren, B. (1990) Unmasking multivariate outliers and leverage points (with discussion), *J. A. Statist. Associ.*, **85**, 633-51.

Ryan, T. A., Jr., Joiner, B. L. and Ryan, B. F. (1976) Minitab: Student Handbook. Duxbury Press: North Scituate, Mass.

Small, N.J.H. (1978) Plotting squared radii, *Biometrika*, **65**, 657-8.

Seber, G.A.F. (1984) *Multivariate Observations*, Wiley, New York.

Siotani, M. (1959) The extreme value of the generalized distance of the individual points in the multivariate normal sample, *Ann. Institute Statist. Math.*, Tokyo, **10**, 183-208.

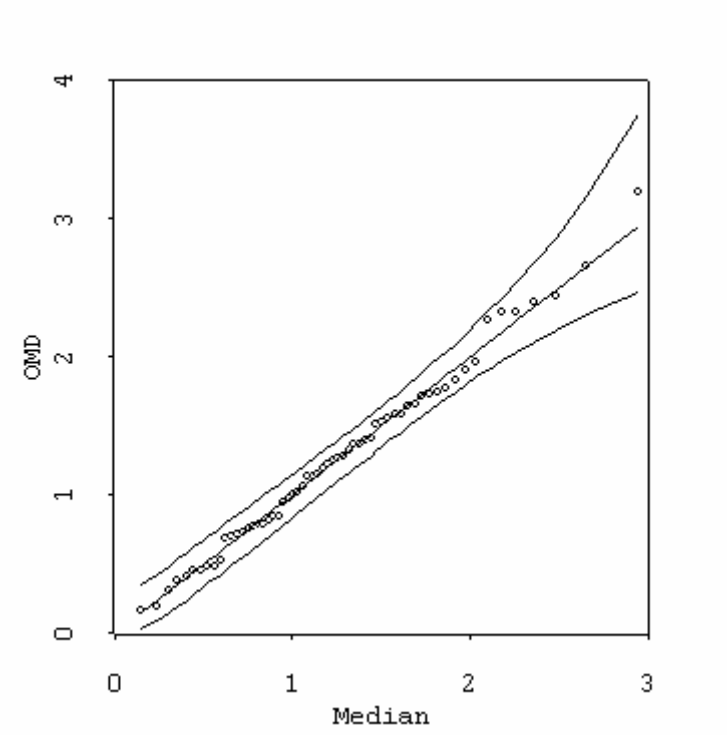



**Fig. 1:** OMD normal-envelope plot of 60 cases from $N_2(0, I)$, ($\alpha = 0.05$).

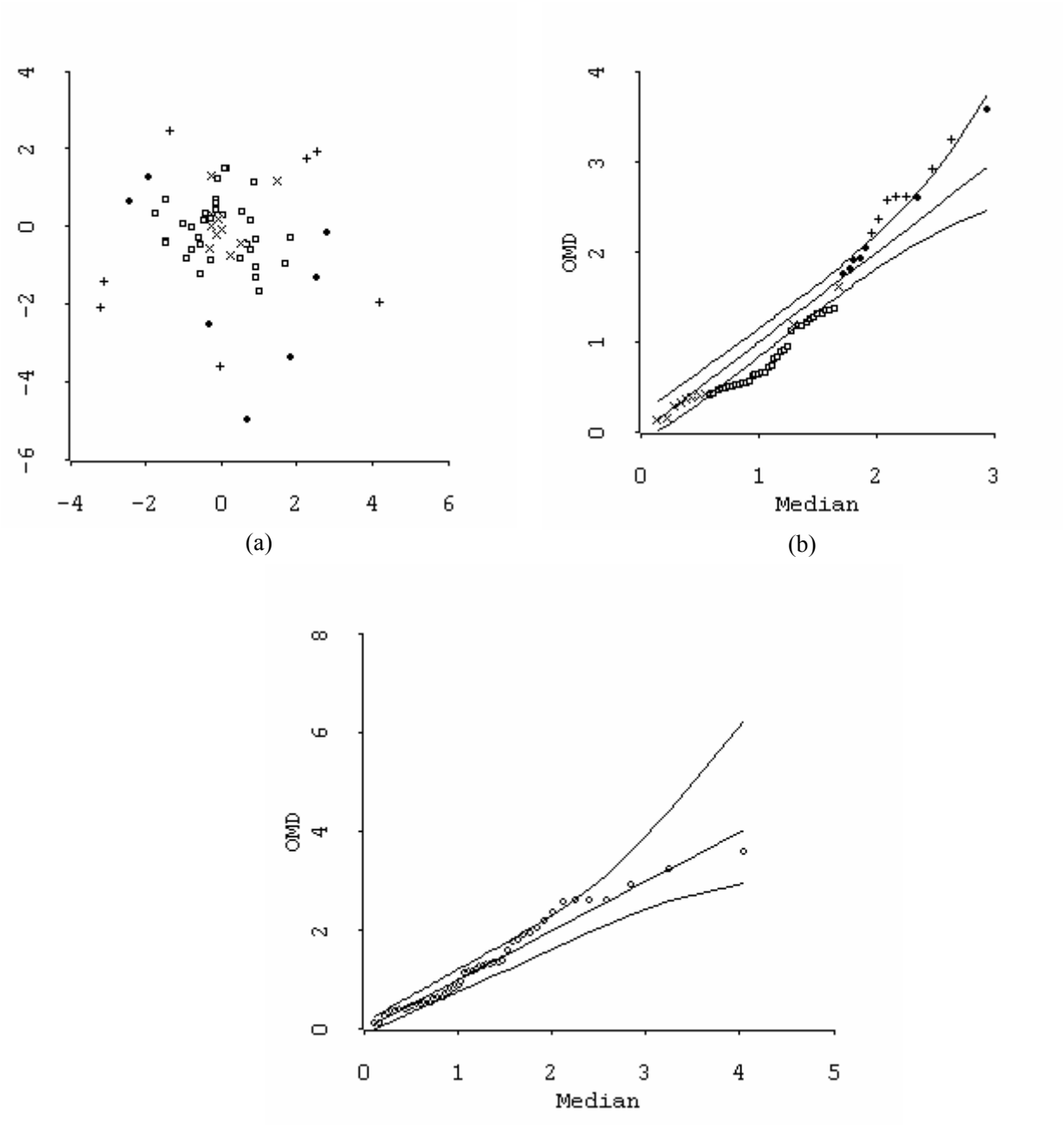
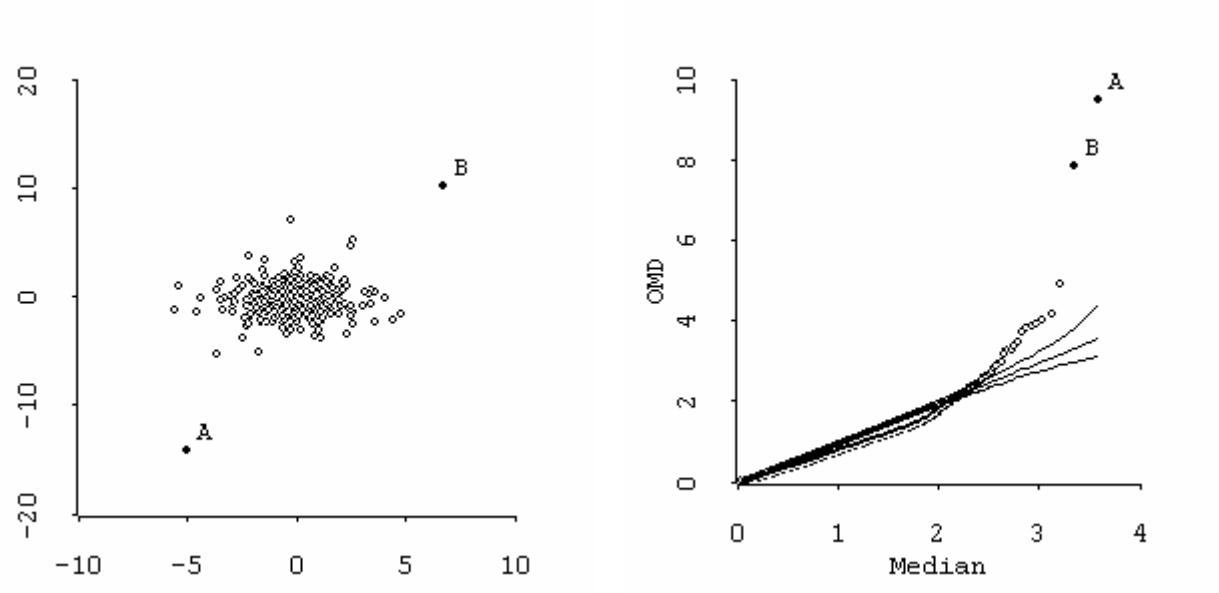



**Fig. 2:** Sample of size 60 from bivariate *t*-distribution with 4 degrees of freedom: (a) scatter plot; (b) OMD normal-envelope plot; (c) OMD fitted-*t*-envelope plot. In (a) and (b), four symbols are linked according as the corresponding OMD falls above or below the median, and inside or outside the normal-envelope.




(a) (b)

**Fig. 3:** Sample of size 500 from bivariate *t*-distribution with 4 degrees of freedom: (a) scatter plot; (b) OMD normal-envelope plot.

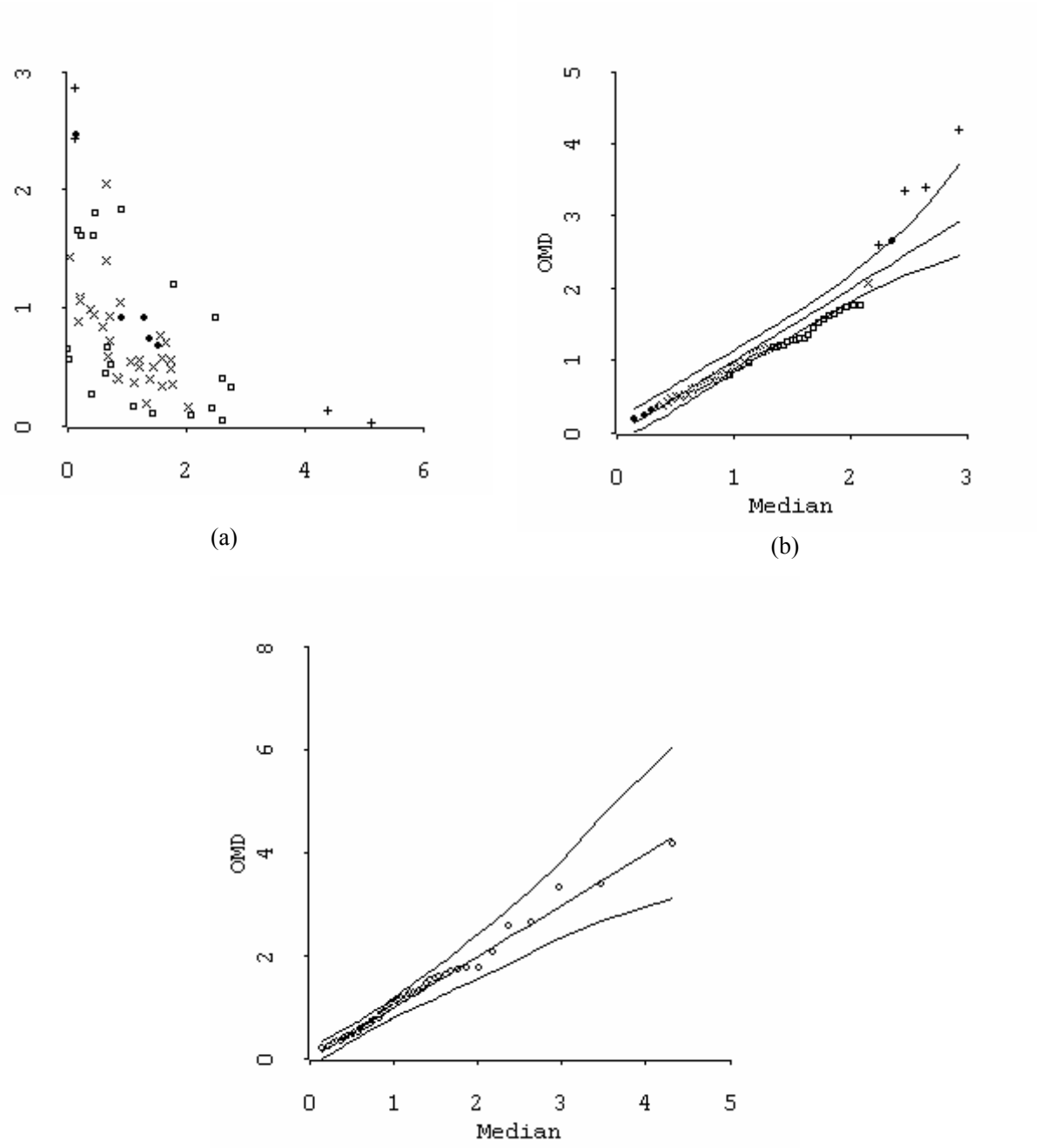



**Fig. 4:** Sample of size 60 from Gumbel's bivariate exponential distribution with $\theta = 1$: (a) scatter plot; (b) OMD normal-envelope plot; (c) OMD fitted-Gumbel-envelope plot. In (a) and (b), four symbols are linked according as the corresponding OMD falls above or below the median, and inside or outside the normal-envelope.

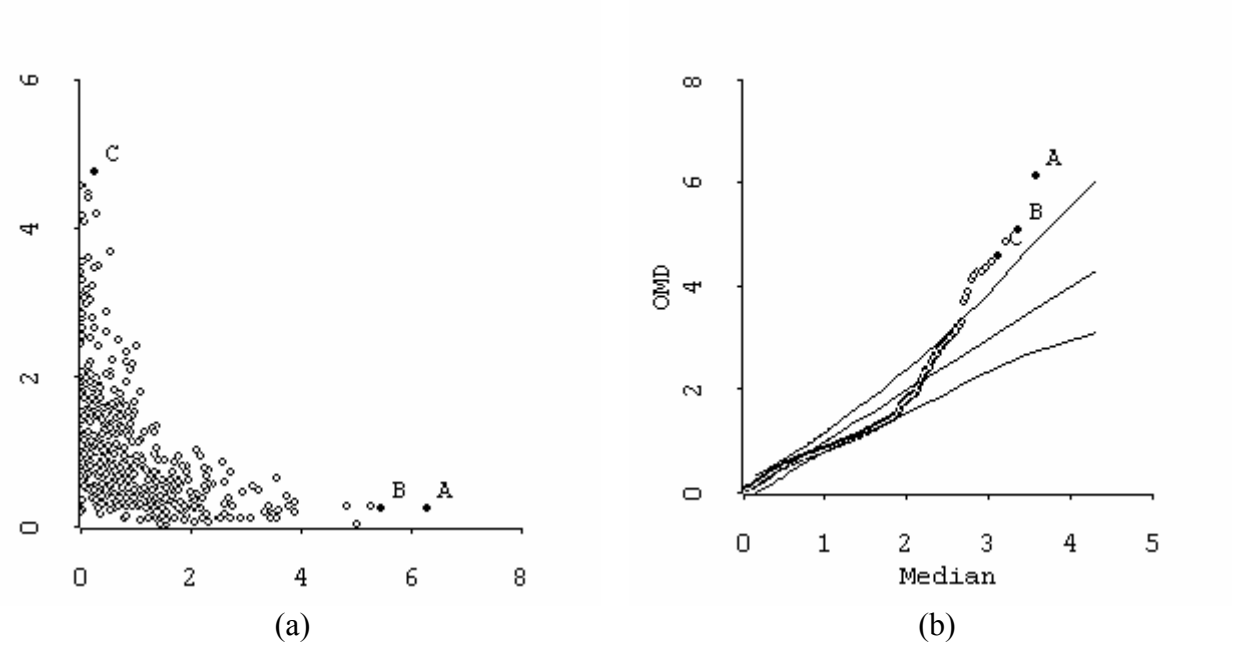
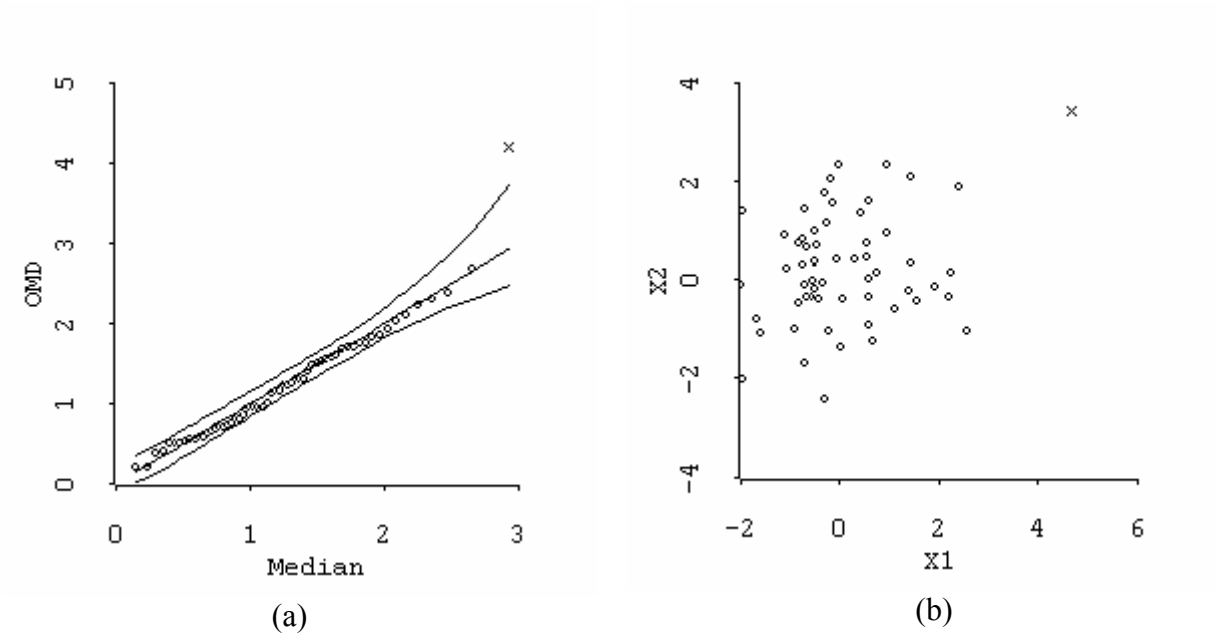



**Fig. 5:** Sample of size 500 from Gumbel's bivariate exponential distribution with $\theta = 1$: (a) scatter plot; (b) OMD normal-envelope plot.




(a) (b)

**Fig. 6:** Normal data ($n$ = 60) with one outlier (denoted "×"):
(a) OMD normal-envelope plot; (b) scatter plot.

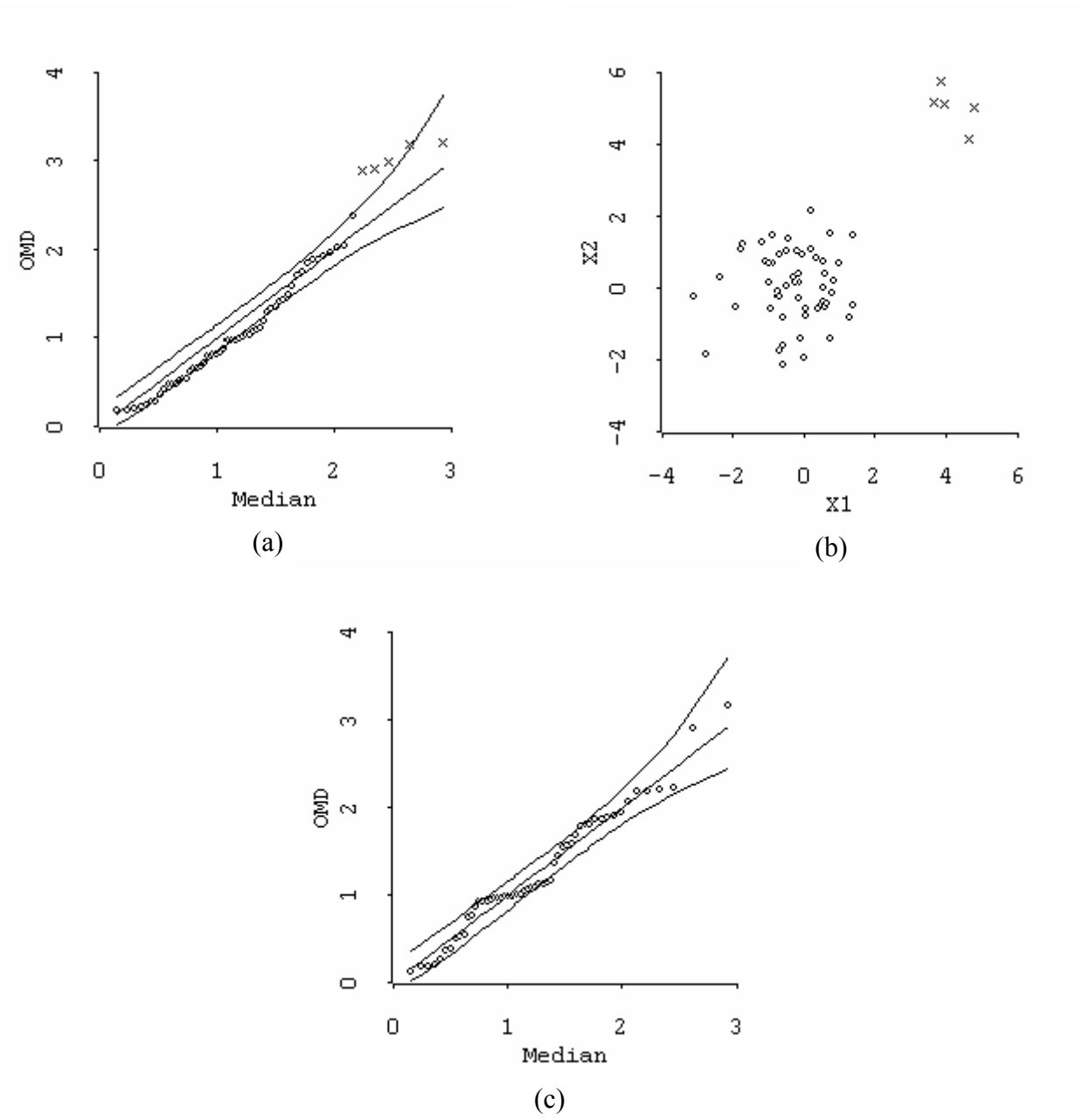
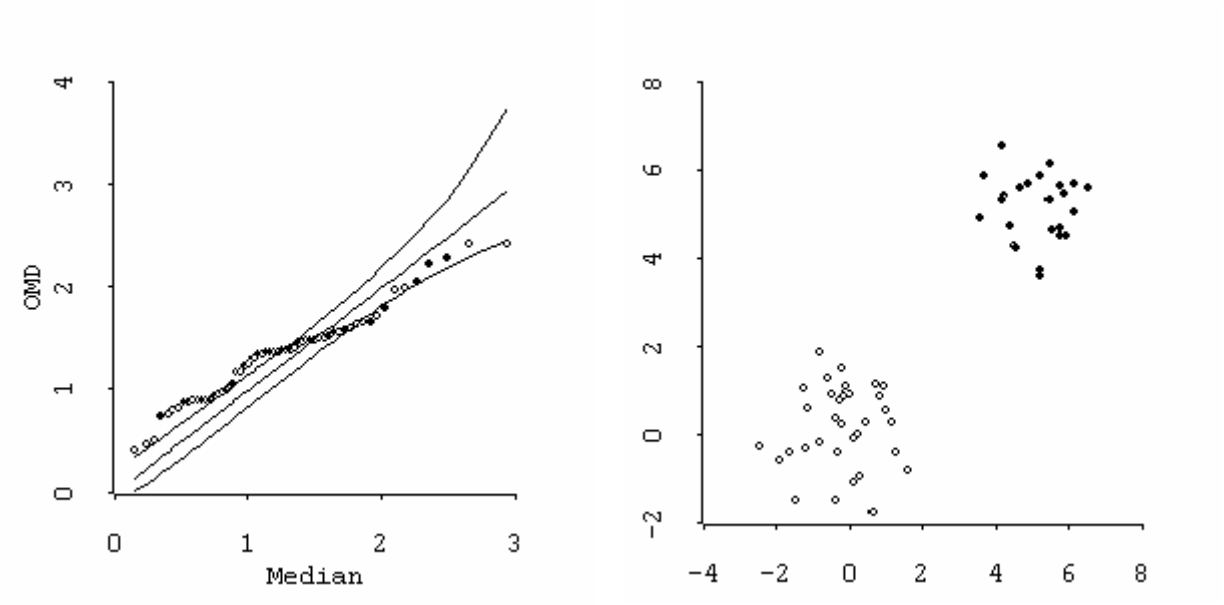
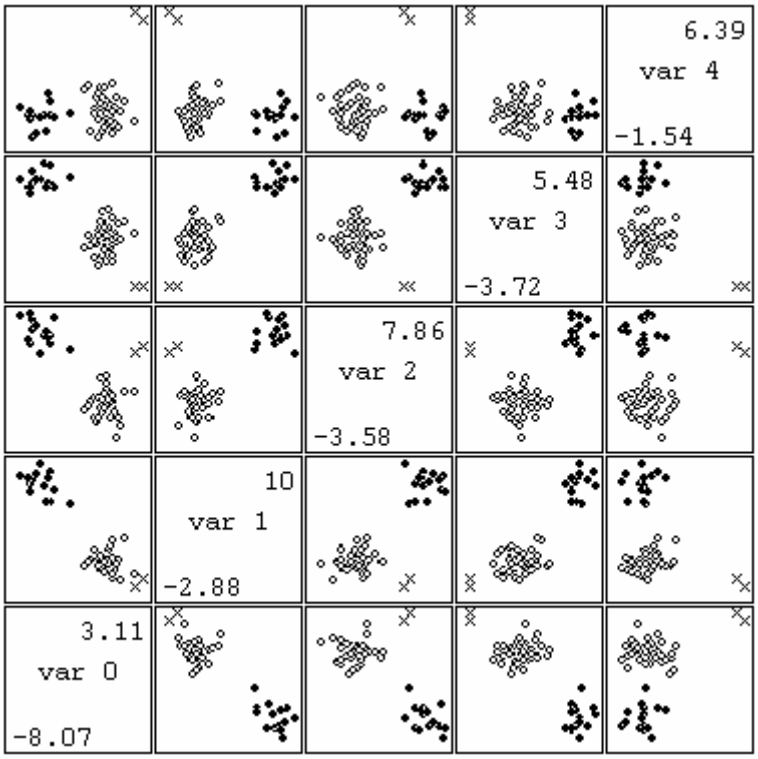



**Fig. 7:** Normal data ($n$ = 60) with 5 outliers (denoted "×"):
(a) OMD normal-envelope plot; (b) scatter plot;
(c) OMD normal-envelope plot ($n$ = 55) after deletion of 5 outliers.

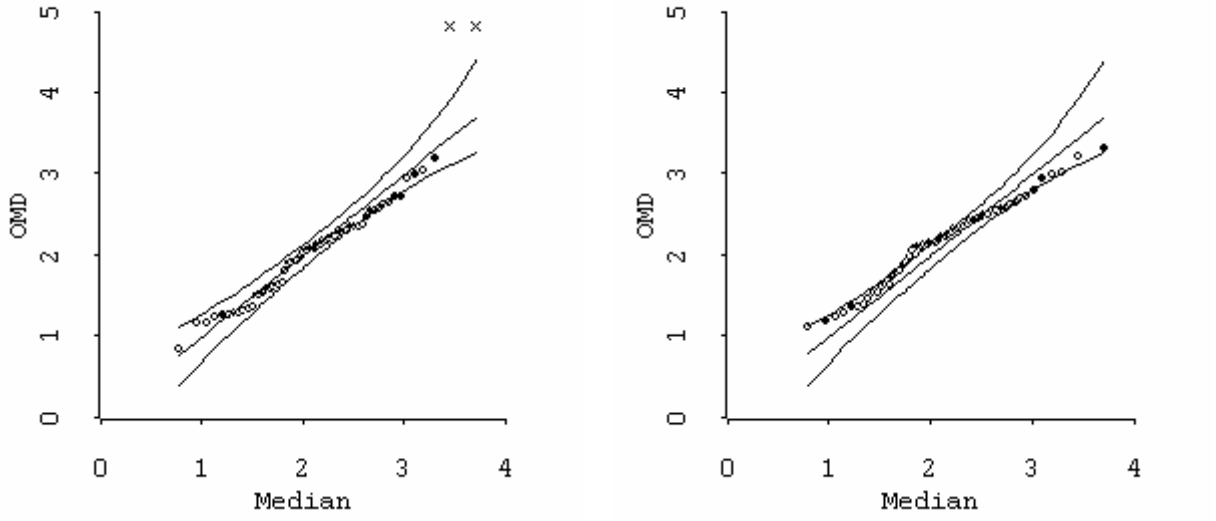



(a) (b)

**Fig. 8:** Normal data ($n$ = 60) with a heavy outlier cluster:
(a) OMD normal-envelope plot; (b) scatter plot.
The $n_{out}$ = 25 points in the outlier cluster are highlighted.

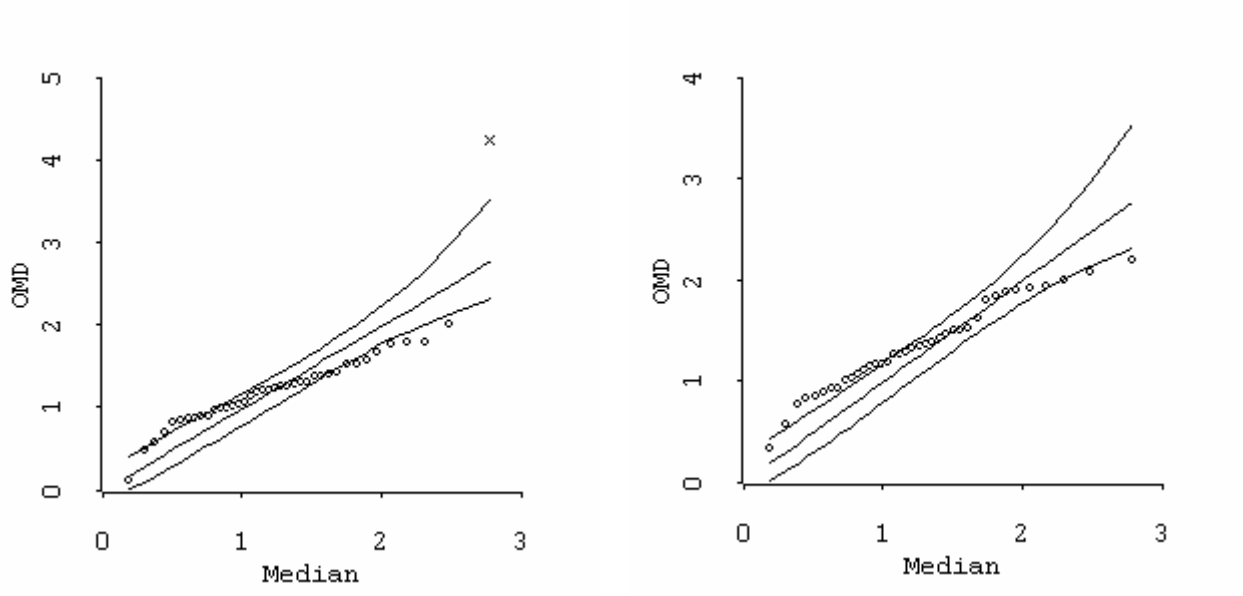



(a)



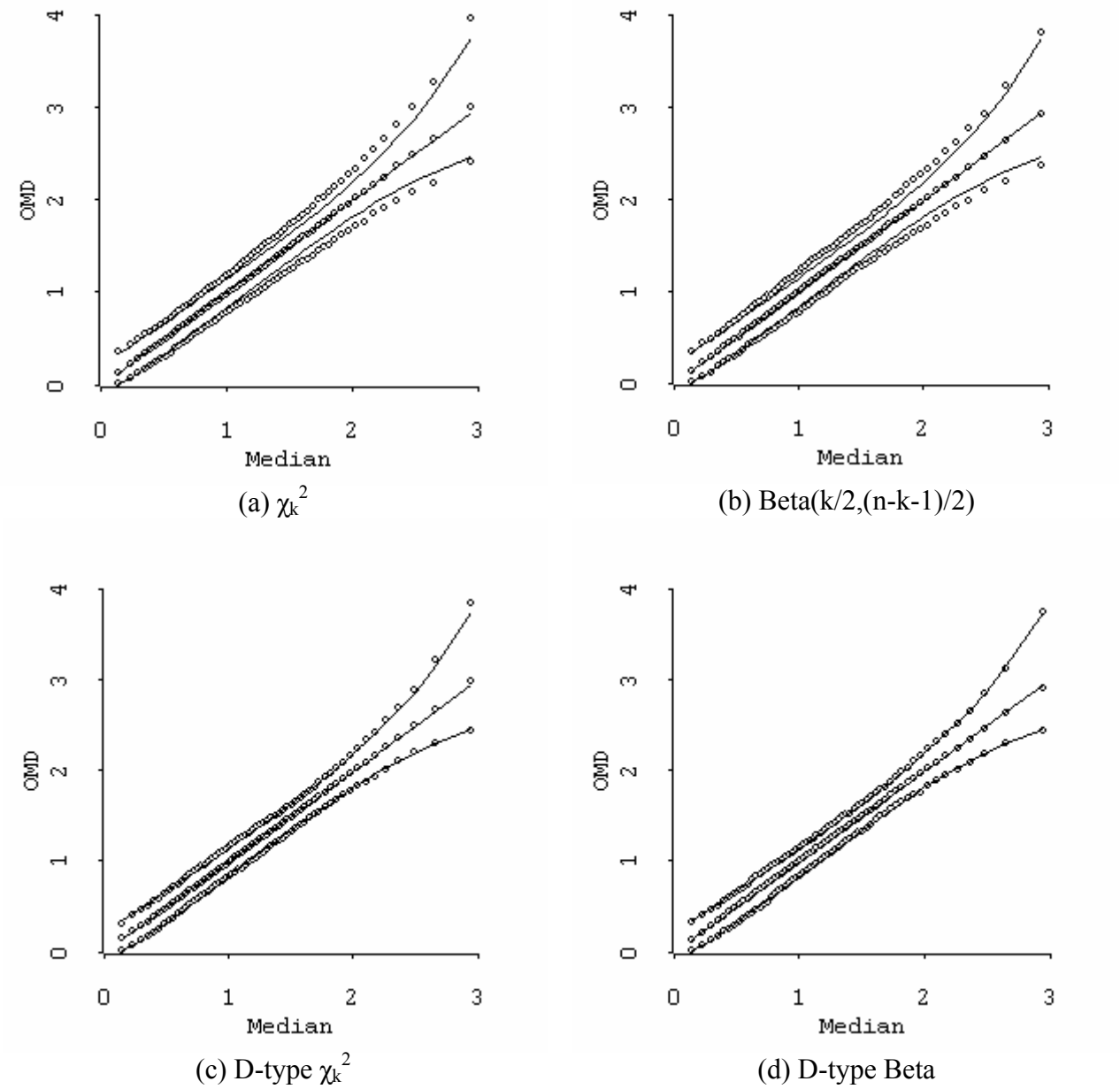

(b) (c)

**Fig. 9:** Normal data ($n$ = 60) with two outlier clusters:
(a) Scatterplot-matrix, (b) OMD normal-envelope plot;
(c) OMD normal-envelope plot ($n$ = 58) after deletion of 2 outliers.
The 2 cases in one cluster are denoted "×", the 15 in the other cluster "•",
and the bulk of data "○".




(a) (b)

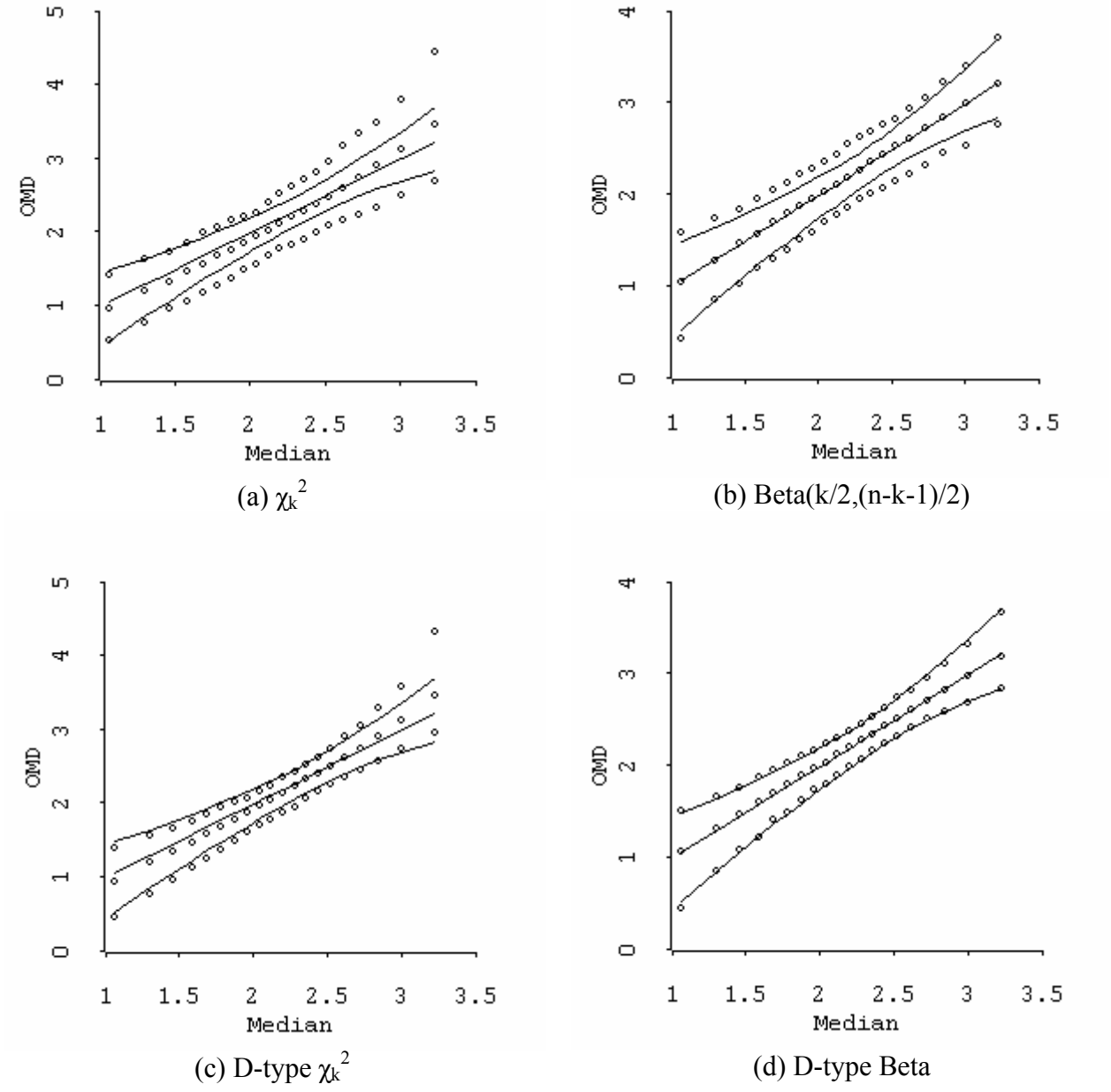
**Fig. 10:** Peruvian Indian weights and heights data: (a) OMD normal-envelope plot; (b) OMD normal-envelope plot after omitting the single outlier "×".




(a) $\chi_k^2$ (b) Beta(k/2,(n-k-1)/2)

(c) D-type $\chi_k^2$ (d) D-type Beta

**Fig. 11:** Approximations to OMD normal-envelope plot with $n = 60$, $k = 2$ and $\alpha = 0.05$.




(a) $\chi_k^2$ (b) Beta(k/2,(n-k-1)/2)

(c) D-type $\chi_k^2$ (d) D-type Beta

**Fig. 12:** Approximations to OMD normal-envelope plot with $n = 20$, $k = 5$ and $\alpha = 0.05$.

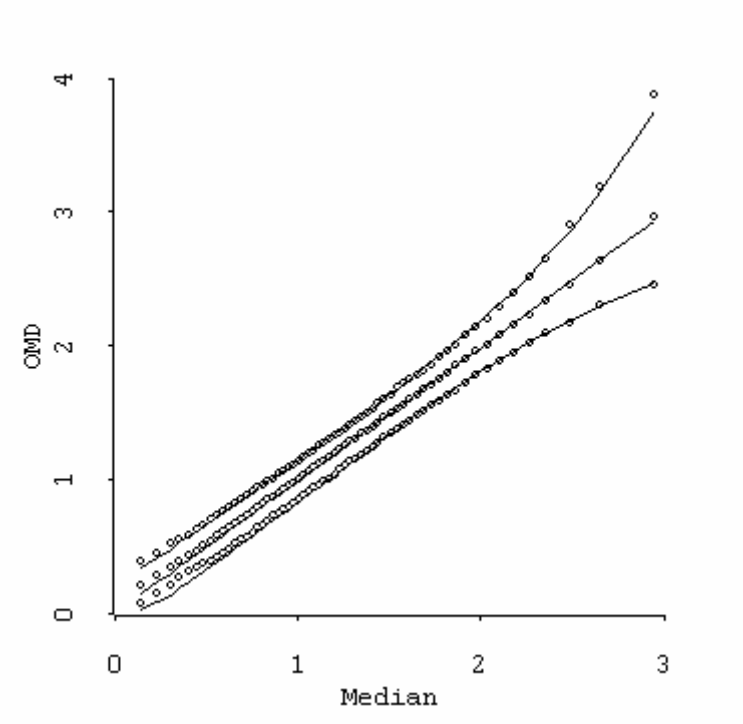



**Fig. 13:** Weighted D-type Chi-square approximation
to OMD normal-envelope plot with $n = 60$, $k = 2$ and $\alpha = 0.05$.

| Method | $S^{(1)}{}_l$ | $S^{(1)}{}_m$ | $S^{(1)}{}_u$ | $S^{(2)}{}_l$ | $S^{(2)}{}_m$ | $S^{(2)}{}_u$ |
|---|---|---|---|---|---|---|
| Chi-square | 56 | 12 | 38 | 61 | 13 | 42 |
| Beta | 44 | 6 | 46 | 53 | 8 | 48 |
| D-type Chi-square | 15 | 11 | 15 | 23 | 14 | 18 |
| D-type Beta | 13 | 4 | 8 | 24 | 8 | 13 |

**Table 1:** Accuracy measures $S^{(1)}$ and $S^{(2)}$ for four approximations to OMD normal-envelope plots, based on 500 simulations with $n = 60$, $k = 2$ and $\alpha = 0.05$.
Measures multiplied by 1000 and rounded to nearest integer.