The effect of a preliminary test of normality using $\sqrt{b_1}$ on Student’s t Distribution

This item was submitted to Loughborough University’s Institutional Repository by the/an author.

Additional Information:


Metadata Record: https://dspace.lboro.ac.uk/2134/32331

Publisher: © J.H. Dodgson

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the published version.
Page 118, line 
Replace the 
Page 178, the la 
"For distri 
distributio 
Page 178, the se 
"By co 
approximati 
some of our 
the margina 
the followi 
Page 183, line 11 
Add the word "being" to end of line. 
Page 186, line 17, and page 191, line 11: 
The reference should be to "Efron". 
Page 186, the sentence beginning on line 23 should read: 
"This would entail a rich population model incorporating the normal as 
a special case, like that of Box and Tiao (1973, Section 3.2.1), 
though this is restricted to the symmetric case, and indeed the Gram- 
Charlier."
ERRATA

Page 118, line 12:
Replace the word "kurtosis" with "skewness".

Page 174, the last sentence should read:
"For distribution [15] a mild linear relation is apparent, while for distribution [20] the very low correlation is barely discernible."

Page 178, the second paragraph before the formulae should read:
"By considering a limiting form of the bivariate Johnson approximation, on which the detailed analysis of Chapter 6 is based, some of our findings may be presented in a simple manner. As \( n \to \infty \), the marginal distributions of \( t \) and \( \sqrt{\theta_1} \) tend to normality, implying the following results:"

Page 183, line 18:
Add the word "being" to end of line.

Page 186, line 17, and page 191, line 1:
The reference should be to "Efron".

Page 186, the sentence beginning on line 23 should read:
"This would entail a rich population model incorporating the normal as a special case, like that of Box and Tiao (1973, Section 3.2.1), though this is restricted to the symmetric case, and indeed the Gram-Charlier."
The Effect of a Preliminary Test of Normality Using $\sqrt{b_1}$ on Student's $t$ Distribution

by

J.H. Dodgson

A Doctoral Thesis
Submitted in partial fulfilment of the requirements for the award of
Doctor of Philosophy
of the Loughborough University of Technology
January 1987

Contents

SYNOPSIS: iii

ACKNOWLEDGEMENTS: v

CHAPTER 1 - Defining the Problem. 1

CHAPTER 2 - The Moments and Correlation of $t$ and $\sqrt{b_1}$ for a Gram-Charlier Population. 22

CHAPTER 3 - A Package of Algorithms for the Gram-Charlier Distribution. 44

CHAPTER 4 - Assessing the Gram-Charlier Generators and Moments of $\sqrt{b_1}$. 70

CHAPTER 5 - The Marginal Distributions of $\sqrt{b_1}$ and $t$ and their Correlation. 102

CHAPTER 6 - The Conditional Distribution of $t$ given $\sqrt{b_1}$. 139

CHAPTER 7 - More Simulation, Conclusions and Extensions. 165

REFERENCES: 188

APPENDIX I - Auxiliary Formulae. 198

APPENDIX II - The Gram-Charlier Package GCPACK. 206

CONTENTS OF FLOPPY DISK. 213

FLOPPY DISK.
Synopsis

Student's t distribution is introduced with background including comments on its robustness properties. The ad hoc procedure of pretesting data for normality is discussed in the light of current advice, and previous work into its effectiveness reviewed. The approach to the problem is outlined: \( \sqrt{b_1} \) for test statistic, the Gram-Charlier distribution for population, approximations using the Johnson system.

Formulae for the moments of t are quoted; expansions for the moments of \( \sqrt{b_1} \), and its correlation with t, are derived using k-statistics. Results are illustrated with contour diagrams.

A Gram-Charlier pseudo-random number generator is developed to provide estimates of moments and assess the k-statistics approximations. A modified rejection method is described together with related algorithms including the inverse distribution function. A full specification is provided.

The Gram-Charlier generators based on rejection and the inverse distribution function are assessed on the basis of the sampling properties of k-statistics. Estimates of the moments of \( \sqrt{b_1} \) obtained from the derived expansions, simulation and another source are compared.

The estimated moments of \( \sqrt{b_1} \) are smoothed, the Johnson system fitted and approximations assessed using simulation. Power is determined. The system is fitted to t, similarly assessed, and robustness properties illustrated. Simulation estimates of the correlation between t and \( \sqrt{b_1} \) are compared with those from k-statistics, and smoothed values given. The question of correlation when fitting the Johnson bivariate system is investigated, and a simple approach adopted.

The effect of a preliminary test is explored by examining \( f(t/\sqrt{b_1}) \),
and extended to the distribution of $t$ given that normality is accepted/rejected. The effect of the preliminary test is characterised and the procedure shown to be ineffective; contributory factors are identified.

A small simulation study confirms previous findings and extends their generality. Final conclusions are drawn and put in context. Comments on the methods employed suggest areas for further work, in addition to those extending the study into preliminary testing.
Acknowledgements

This research was carried out while employed at Trent Polytechnic, Nottingham. I would like to express my appreciation to the Polytechnic (and Nottinghamshire County Council) for meeting most of the costs. I am grateful for the encouragement of colleagues in the Department of Mathematics, Statistics and Operational Research, particularly Peter Dixon, and practical help with computing provided by Mike Pate and members of the Department of Computing Services.

Discussion and correspondence with others in the statistical community has been both enjoyable and valuable. Any contributions not explicitly acknowledged have nevertheless been appreciated.

My supervisor Tony Pettitt is to be thanked for finding the problem, the patience to see it through, and further for the interest shown in my statistical career since the beginning of his own.

The manuscript has been professionally typed by Mrs. Wright at Loughborough; the Reprographic Section at Trent has helped in its production.

Finally I am indebted to my wife for providing the support which has enabled me to see the project through to completion.
Chapter One
Defining the Problem

1.1 **STUDENT’S t DISTRIBUTION AND THE ASSUMPTION OF NORMALITY.**

When dealing with univariate data, statistics as practised often takes the following form. Assume the observations represent a random sample from a population of known form but depending on one or more unknown parameters. The problem is then reduced to one of estimating the parameters. In simple cases summary statistics like the mean and variance provide intuitive estimates but generally the analysis will depend on the particular population model assumed, and certainly if one is concerned with optimality. Such point estimates are often quoted with their (estimated) standard errors though an indication of their precision is perhaps more conveniently expressed by giving interval estimates for the parameters — typically confidence intervals. In certain contexts, for example industrial quality control, the problem may fall within the framework of hypothesis testing, and the data tested for consistency with certain hypotheses specifying values or ranges of values for the parameters.

The most common model for continuous data is of course the normal distribution. Frequently justified by considerations of symmetry and unimodality, or the central limit theorem, the main appeal of this assumption is that problems involving location and scale find ready-made solutions in the literature. It is well-known that inference about the mean \( \mu \), when the population variance is also unknown, is then based on the statistic

\[
t = \frac{\bar{x} - \mu}{s / \sqrt{n}},
\]
having Student's t distribution with n-1 degrees of freedom. A related statistic was first proposed by Student (1908) and given in its present form by Fisher in the early twenties (discussed by Eisenhart (1979)). From the classical sampling theory viewpoint the use of Student's t distribution may be argued from Neyman-Pearson theory although there are interesting parallels in Bayes' and likelihood theory (Kendall and Stuart, 1979; Box and Tiao, 1973; Edwards, 1972).

In practice the assumptions underlying the use of Student's t distribution may be violated to a greater or lesser extent. Most obviously the population may be non-normal with samples exhibiting skewness or excess kurtosis for example; under certain circumstances the data may be prone to outliers or serial correlation. A statistical procedure that is relatively insensitive to departures from its underlying assumptions, particularly its population model, is termed "robust" (coined by Box in 1954) and robustness studies based on Monte Carlo simulation have become common with the advent of electronic computers. A sampling investigation into the robustness of the t statistic is reported in Pearson et al. (1929) and cited by Bartlett (1935) in an early mathematical approach in which the population is assumed to be Gram-Charlier. The literature on the subject is now extensive; for a recent discussion on the robustness of the common normal theory statistics using both real and simulated data see Pearson and Please (1975).

To sum up the position for the one-sample t statistic denote by $\sqrt{\beta_1}$ and $\beta_2$ the usual moment ratio measures of population skewness and kurtosis, their values for a normal population being of course 0 and 3 respectively. Quoting Pearson and Please (based on the
results of Geary (1947))

\[ E(t) = -\frac{1}{2} \sqrt{n} \beta_1 / n^{1/2} - O(n^{-3/2}) \]

\[ V(t) = 1 + \frac{1}{4} (8 + 7\beta_1) / n + O(n^{-2}) \]

\[ \sqrt{n} \beta_1 (t) = -2 \sqrt{n} (n^{1/2} - O(n^{-3/2}) \]

\[ \beta_2 (t) = 3 + 2(6 - \beta_2 + 6\beta_1) / n + O(n^{-2}) \]

One can understand the general conclusions that the statistic is quite insensitive to population kurtosis but for small to moderate sample sizes positive skewness in the population will shift the mean of \( t \) to the left and introduce negative skewness (changes being in the opposite direction for negative population skewness) with consequent effects on the true percentage points.

In practice we are unlikely to know the shape of the population though arguably we will have a clearer idea in the case of repeated routine analyses than when working with a unique sample. Gayen (1949) shows how to modify the percentage points of Student's \( t \) distribution if the population skewness and kurtosis may be assumed known. Various forms of modified \( t \) statistic which are generally more robust than the original have been suggested, for example Tukey and McLaughlin (1963) and Johnson (1978), and may be viewed as part of a general movement towards robust and adaptive procedures (Hogg, 1974).

Generations of practical statisticians have simply transformed extreme non-normal data, for example taking square roots or logarithms to remove skewness, before proceeding with the analysis (Bartlett 1935), though care is needed when interpreting the results in terms of the original units.
Concentrating on the t statistic, or its derivatives, may suggest a narrow view of statistical methodology excluding among others the simple to apply non-parametric procedures (at least for hypothesis testing) and the Bayesian approach with its scope for modelling uncertainty about the type of population as well about the population parameters (Box and Tiao, 1973). Yet Student's t distribution retains its central place among statistical methods because of its optimal properties for normal samples, the reassuring evidence of its robustness and not least its widespread acceptance and application in science and industry. The Weights and Measures Act 1979, introducing an "average" quality system to bring the UK into line with the EEC, gave the t test the authority of law being one half of the reference test (Bissell and Pridmore, 1981).

1.2 THE EFFECT OF A PRELIMINARY TEST OF NORMALITY: REVIEW.

Consider the situation where we have a unique sample and (for definiteness) require a confidence interval for the population mean. An apparently reasonable ad hoc procedure would be to carry out a preliminary test of normality using a standard test or probability plot and if normality seems a reasonable assumption construct a confidence interval using the percentage points of Student's t distribution in the usual way. If the assumption of normality is in doubt, and depending on the type and degree of non-normality, one could proceed as before and treat the solution as an approximation or use one of a number of alternative methods as indicated in the previous section.

This procedure, though perhaps less baldly stated, is implicitly
recommended in modern applied statistics texts which rightly stress the assumptions underlying the common normal theory statistics, outline their robustness properties and discuss ways of checking the consistency of data with the assumptions. Wetherill (1981) is a good example. Few authors, however, give definite advice on the routine use of preliminary tests and whether they should formally be part of the analysis.

Cox (1977) addresses this question in an examination of the role of significance tests. He makes the distinction between the primary and secondary aspects of the analysis. For our problem the primary aspect concerns the mean, the secondary aspect the assumption of normality. He writes "significance tests of hypotheses of simple secondary structure seem to be relatively unimportant; in particular so-called preliminary test procedures should, I think, be avoided if feasible". The question of the adjustment of method in the light of the data is mentioned in a later general address, where it is suggested, in the present context, that no adjustment is necessary (Cox, 1981).

Advice in the industrial context may be sought from the British Standards Institution. BS 2846 Part 4 (British Standards Institution, 1976) discusses the usual procedures based on normal theory statistics and states "as a start it is usually desirable to make a rough examination of this [normality] assumption, unless of course an adequate assurance of normality has been established from past examination of similar data". Probability plotting is suggested and there is mention of transformations and non-parametric methods. BS 5497 Part 1 (British Standards Institution, 1979) on the precision
of test methods (i.e. physical tests) does incorporate preliminary
tests for outliers, though not the usual tests of normality, as
part of the recommended procedure. Both standards have international
counterparts. One may also note that the EEC's average quality
system referred to earlier has no requirement for a formal test
of normality.

Bancroft and Han (1977) introduce the terminology "inference
based on conditional specification" and give a bibliography. They
are concerned with problems whose exact specification is unclear
and is decided using the results of preliminary test(s). The
question of normality when inference is about the population mean
would appear to be an important example but the emphasis is on whether
data may be pooled before subsequent analyses. These ideas are
introduced and extended in an otherwise conventional textbook
(Bancroft and Han, 1981). Though not directly helpful these sources
do acknowledge the problem of preliminary testing and insist that
the effect of any such tests on subsequent analyses should be taken
into account.

Box (1980) in discussing scientific modelling and robustness
has a broader perspective. He views modelling as a two-stage process:
criticism and estimation. The first stage is an iterative search for
the best model during which goodness-of-fit tests may formally play
a part. The estimation in the second stage is carried out conditional
on the model's truth using Bayes' theory. Robustness is considered
by allowing the introduction of a "discrepancy factor" - presumably
like $\sqrt{\lambda}$ - into the final stage. The question of a preliminary test
of normality before using Student's t distribution cannot fairly be
seen within the context of this paper. The approach to estimation
is particularly different and there is the distinction between our
concept of criterion robustness and the Bayesian concept of inference
robustness (Box and Tiao, 1973, p152). Nevertheless the recognition
of the role of significance tests during the search for a statistical
model is noteworthy. One might also cite Gilchrist's (1984)
introductory book on statistical modelling where the use of the
skewness statistic and probability plotting are suggested in the
identification stage and formal goodness-of-fit ($\chi^2$ and Kolmogorov-
Smirnov) in the validation stage.

Given the lack of generally accepted principles governing the
use of preliminary tests of goodness-of-fit we examine directly the
ad hoc procedure described at the beginning of this section. One
may hope that a fuller understanding of this common and typical
problem may help in the establishment of more general guidelines.

The principle underlying a preliminary test of normality
(whether used informally or formally with a specified significance
level) seems to be this:

The sampling distribution of the $t$ statistic for
samples which pass the preliminary test of normality
is well represented by Student's $t$ distribution, but
differs markedly (at least is less well represented)
for samples which fail the preliminary test.

It is this principle, or rather the question of its validity, that
is the main subject of this investigation.

There appears to have been only one paper specifically on this
topic though Pearson and Please (1975) do make brief mention of it in their review on robustness. The results in Table 1.1 (extracted from their Table 4) were obtained by generating 5000 samples of size 10 from three Johnson $S_U$ distributions with specified $\sqrt{\beta_1}$ and $\beta_2$. (The Johnson system is discussed in Section 1.4.) The heading "Total" gives the percentage of calculated $t$ values above (below) the upper (lower) 5% and 1% points of Student's $t$ distribution with 9 degrees of freedom. The samples were also screened for non-normality with the criteria reject if $|\sqrt{\beta_1}| < 1.14$ and/or $\beta_2 < 4.0$, the limit for $\sqrt{\beta_1}$ being the approximate 2.5% point and for $\beta_2$ the 5% point. The resulting percentages for $t$ are indicated by "Restricted". The percentage of samples restricted - the approximate power of the test of normality - ranged from 9% to 13%.

Table 1.1
Comparison of percentages beyond nominal values.
5000 samples (Total) of size 10 from $S_U$.
(From Pearson and Please (1975).)

<table>
<thead>
<tr>
<th>Population</th>
<th>100a</th>
<th>Above upper limit</th>
<th>Below lower limit</th>
<th>Percentage restriction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>Restricted</td>
<td>Total</td>
</tr>
<tr>
<td>$\sqrt{\beta_1} = 0.0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_2 = 3.6$</td>
<td>5</td>
<td>5.02</td>
<td>4.99</td>
<td>4.72</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.78</td>
<td>0.81</td>
<td>0.86</td>
</tr>
<tr>
<td>$\sqrt{\beta_1} = 0.4$</td>
<td>5</td>
<td>4.28</td>
<td>4.27</td>
<td>6.46</td>
</tr>
<tr>
<td>$\beta_2 = 3.6$</td>
<td>1</td>
<td>0.58</td>
<td>0.62</td>
<td>1.72</td>
</tr>
<tr>
<td>$\sqrt{\beta_1} = 0.8$</td>
<td>5</td>
<td>2.94</td>
<td>3.10</td>
<td>8.06</td>
</tr>
<tr>
<td>$\beta_2 = 3.6$</td>
<td>1</td>
<td>0.28</td>
<td>0.32</td>
<td>2.54</td>
</tr>
</tbody>
</table>
These results suggest a small net gain in using a preliminary test of normality though it is described by the authors as minimal. They speculate on whether the sample information $\sqrt{b_1}$ and $b_2$ could be put to better effect and in this context it is interesting to note that Johnson (1978) uses them in a modification of Gayen's (1949) procedure.

Easterling and Anderson (1978) report on a simulation study into the problem and argue along the following lines. Let $F$ denote a particular distribution and $G$ a statistic to test normality (without loss of generality suppose large values are significant) with $G_p$ denoting its $100p$-th percentile under the null hypothesis. Consider the conditional distributions

\begin{align}
    h(t|G \leq G_p, F) & \quad (1.1) \\
    h(t|G > G_p, F) & \quad (1.2)
\end{align}

for typical values of $p$ of 0.90 or 0.95. If the preliminary test is doing a useful task (1.1) should be adequately represented by Student's $t$ distribution and certainly more Student-like than (1.2) whatever the distribution $F$. To test normality they used the Anderson-Darling statistic $A^2$ (see for example Stephens (1974)) and the Shapiro-Wilk statistic $W$ (Shapiro and Wilk, 1965). The populations sampled were the normal itself, uniform, central and non-central Student's $t$ and exponential. Their method was to take repeated samples (of size 10 and 20) and classify them as significant or non-significant on the basis of $A^2$ and $W$ tests at the 10% level. Sampling continued until 1000 were obtained in each category. The resulting empirical distributions were compared.
with the t distribution using chi-square tests with twelve cells having boundaries at the nominal 5, 10, 20, ..., 90 and 95 percent points. Table 1.2 for W and n = 20 is typical of their results.

Table 1.2
Comparisons of chi-square values (ν = 11) for significant and non-significant samples.
1000 samples of size 20 for (1.1) and (1.2).
(From Easterling and Anderson (1978).)

<table>
<thead>
<tr>
<th>Population</th>
<th>$\sqrt{\beta_1}$</th>
<th>$\beta_2$</th>
<th>Preliminary W test of normality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Non-significant at 10% (1.1)</td>
</tr>
<tr>
<td>Normal</td>
<td>0.0</td>
<td>3.0</td>
<td>29.8</td>
</tr>
<tr>
<td>Uniform</td>
<td>0.0</td>
<td>1.8</td>
<td>14.9</td>
</tr>
<tr>
<td>t(7,0)*</td>
<td>0.0</td>
<td>5.0</td>
<td>9.4</td>
</tr>
<tr>
<td>t(3,0)</td>
<td>0.0</td>
<td>-</td>
<td>18.7</td>
</tr>
<tr>
<td>t(7,2)</td>
<td>1.2</td>
<td>7.5</td>
<td>75</td>
</tr>
<tr>
<td>Exponential</td>
<td>2.0</td>
<td>9.0</td>
<td>376</td>
</tr>
<tr>
<td>t(3,2)</td>
<td>-</td>
<td>-</td>
<td>1274</td>
</tr>
</tbody>
</table>

* $t(\nu,\delta)$ denotes a non-central t distribution with $\nu$ degrees of freedom and non-centrality parameter $\delta$.

If the preliminary test of normality is effective two features should be observed in Table 1.2. Column (1.1) should represent a chi-square variable with 11 degrees of freedom (expectation 11 with 95-th percentile 19.7) and for each population the chi-square value in column (1.1) should be smaller than that for column (1.2). What is especially disturbing is that for asymmetric populations the
preliminary test of normality has the opposite of the desired effect in making the representation by Student's t distribution worse. Commenting on the effect of a preliminary test of normality on the distribution of the t statistic based on these and other sets of results the authors conclude:

(1) For a normal population, as expected, there is no effect.

(2) For other symmetric populations the representation by Student's t distribution is improved but only for large population kurtosis is the effect pronounced.

(3) For asymmetric populations the representation by Student's t distribution in most cases is markedly worse.

Possible reasons for the anomalous behaviour of the t statistic suggested by the authors were:

(i) The numerator non-normal.

(ii) The expectation of the numerator non-zero.

(iii) The square of the denominator not chi-square.

Further work (related to (ii)) was reported as it was suspected that for asymmetric populations the effect of the preliminary test of normality was to shift the mean of t. To investigate this the population mean $\mu$ in the t statistic was replaced by $\bar{X}$, the mean of the 1000 sample means in the relevant category. The preliminary test of normality did become more effective in improving the representation by Student's t distribution in some cases though this gain appears of little relevance if we remember that in practice inference is about $\mu$. The importance of the shift in mean in understanding the effect of a preliminary test of normality will later be confirmed.
1.3 THE EFFECT OF A PRELIMINARY TEST OF NORMALITY: METHOD OF INVESTIGATION.

Like Easterling and Anderson we approach the problem by considering the distribution of the $t$ statistic conditional on the value (or range of values) of the statistic used in the preliminary test of normality, for varying degrees of non-normality in the population.

To test for normality we use the familiar sample measure of skewness

$$\sqrt{b_1} = \frac{m_3}{m_2^{3/2}},$$

where

$$m_r = \frac{1}{n} \sum (x_i - \bar{x})^r.$$

Percentage points are given by Pearson and Hartley (1976) though their history can be traced back to Pearson (1930). There are several reasons for this choice. It is the standard test against population skewness and as we have seen it is skewness (rather than kurtosis) that most effects the distribution of the $t$ statistic. Indeed it has been shown to be very powerful against skewed alternatives (Shapiro, Wilk and Chen, 1968). Unlike omnibus test statistics like $A^2$ and $W$ it is directly interpretable and for large samples may be used as an estimate of population skewness. Also being a moment statistic (like $t$ itself) distribution theory may be expected to be comparatively simple. We may note that recent work shows $\sqrt{b_1}$ to be closely related to other statistics: the Cramér-von Mises statistic (Pettitt, 1977); the efficient score statistic for a certain class of skew alternatives (Spiegelhalter, 1983).
Of particular interest is the effect of non-normality in the population on the conditional distribution of the t statistic given $\sqrt{\beta_1}$, and interpretation will be simpler if the amount of skewness and kurtosis may easily be controlled. In the first instance we assume a Gram - Charlier Type A distribution which has the general form

$$f(x) = \phi(x) \sum_{r=0}^{\infty} a_r H_r(x)$$

where $\phi(x)$ is the standard normal density, $a_r$ is a coefficient expressible in terms of moments up to order $r$ and $H_r(x)$ is the $r$th Hermite polynomial (Kendall and Stuart, 1977, Section 6.17).

Truncating after the fourth term and expressing in standard measure we obtain the usual form

$$f(x) = \phi(x) \left\{ 1 + \frac{\kappa_3}{3!} H_3(x) + \frac{\kappa_4}{4!} H_4(x) \right\}$$

$$-\infty < x < \infty .$$

Skewness and kurtosis are reflected in the population cumulants:

$$\kappa_3 = \sqrt{\beta_1}$$

$$\kappa_4 = \beta_2 - 3 .$$

The higher order cumulants are all zero, a distinct theoretical advantage when obtaining moment approximations using k-statistics.

The polynomials are

$$H_3(x) = x^3 - 3x$$

$$H_4(x) = x^4 - 6x^2 + 3 .$$

As previously noted Bartlett used this distribution in 1935.
to investigate the robustness of the t statistic; since then the
Gram–Charlier and the related Edgeworth distributions have been
regularly used in such studies.

The admissible regions in the $\beta_1 - \beta_2$ plane for positive
definite and unimodal distributions are given by Barton and Dennis
(1952) (also Draper and Tierney (1972) who correct one of the
Edgeworth regions) and reproduced in Figure 1.1. Ten points are
also identified which will be referred to in later chapters.

**Figure 1.1**

Positive definite and unimodal regions of the
Gram–Charlier distribution.
The restricted range of $\beta_1$ and $\beta_2$ is a disadvantage of this distribution to be weighed against its analytic simplicity. If the problem were to be approached using only Monte Carlo simulation the Johnson system (Section 1.4) or Tukey's lambda distribution (Ramberg, 1975) might provide useful population models because of their greater range of $\beta_1 - \beta_2$ values and their simplicity in generating pseudo-random variables. In the event a Gram - Charlier generator was developed to check and improve on moment approximations, and some further investigation carried out using simulated Johnson samples.

Supposing the Gram - Charlier population (1.3), we wish to investigate the relationship between the preliminary test of normality based on $\sqrt{b_1}$ and the distribution of the $t$ statistic. One direct formulation of the problem is to write

$$f(t|\sqrt{b_1}) = \frac{g(t, \sqrt{b_1})}{h(\sqrt{b_1})}. \quad (1.4)$$

Clearly the special case of a normal population ($\kappa_3 = 0, \kappa_4 = 0$ in (1.3)) does not interest us as $t$ and $\sqrt{b_1}$ will be independent. One argument for this is as follows. The sample mean and variance are joint sufficient statistics for the parameters of the normal distribution. The distribution of $\sqrt{b_1}$ does not depend on these parameters. Appealing to the theorem of Hogg and Craig (1956), $\sqrt{b_1}$ is independent of $\bar{x}$ and $s^2$ and therefore independent of the $t$ statistic.

Now an analytic solution to (1.4) might begin by finding the joint distribution of $E_{\bar{x}_1}$, $E_{\bar{x}_1}^2$ and $E_{\bar{x}_1}^3$ for a Gram - Charlier
population. Attempts along these lines were unsuccessful. Therefore approximations to the required marginal and joint distributions were sought, and subsequently obtained using the Johnson system.

1.4 THE JOHNSON SYSTEM.

The Johnson univariate system (Johnson, 1949a) is based on a transformation (or "translation") of the variate $X$ to the standard normal variate $Z$. This has the general form

$$ Z = \gamma + \delta f \left( \frac{X - \xi}{\lambda} \right) \tag{1.5} $$

where $f(*)$ is a monotonic function and $\gamma$, $\delta$, $\xi$ and $\lambda$ are parameters ($\delta$ and $\lambda$ are taken to be positive). A feature of this approach is that knowledge about the distribution of $Z$, for example percentage points, may be used in the study of the distribution of $X$.

The shape of the distribution of $X$ depends on $\gamma$ and $\delta$, while $\xi$ and $\lambda$ are location and scale parameters. A common alternative form of (1.5) is

$$ Z = \gamma + \delta f(Y) \tag{1.6} $$

where

$$ Y = \frac{X - \xi}{\lambda} $$

emphasis being on the distribution of $Y$. Figure 1.2 (taken from Johnson and actually illustrating the particular case $S_\lambda$) shows the general features of the transformation.
In fact three systems are needed to cover all possible points in the $\beta_1 - \beta_2$ plane:

1. The lognormal system $S_L$
   \[
   Z = \gamma + \delta \ln(X - \xi) \quad \xi < X.
   \]

2. The unbounded system $S_U$
   \[
   Z = \gamma + \delta \sinh^{-1}\left(\frac{X - \xi}{\lambda}\right).
   \]
(3) The bounded system $S_B$

$$Z = \gamma + \delta \ln \left( \frac{X - \xi}{\xi + \lambda - X} \right) \quad \xi < X < \xi + \lambda .$$

The lognormal distribution dates from the last century (Johnson and Kotz, 1970). It has only three parameters but it is convenient to let $\lambda = \pm 1$, in agreement with the sign of the third moment of $X$, and to write

$$Z = \lambda \left[ \gamma + \delta \ln \left( \frac{X - \xi}{\lambda} \right) \right]$$

for programming purposes.

Figure 1.3 shows the $S_U$ and $S_B$ regions divided by the lognormal line. The normal distribution itself is the limiting form of all three as $\delta \to \infty$.

**Figure 1.3**

Regions for the Johnson Systems $S_L$, $S_U$ and $S_B$. 

![Graph showing regions for Johnson Systems $S_L$, $S_U$, and $S_B$.]
Full details of the properties of all three systems, including their explicit density functions and moments, may be found in Johnson. Several methods have been used to fit these distributions including the use of

(i) percentiles,
(ii) moments,
(iii) maximum likelihood.

Since we proceed by finding approximations to the theoretical moments of the t statistic and $\sqrt{b_1}$, the method of moments is particularly suitable; Hill, Hill and Holder (1976) provide a convenient algorithm.

We note that the Johnson system may also be used as a population model allowing the introduction of skewness and kurtosis from the normal. It is not much used in theoretical work, there being three separate systems and only the lognormal is reasonably tractable. However it is convenient for simulation studies since one need only specify the required first four moments, determine the system and the associated parameters, and use the inverse of (1.5), viz

$$X = \xi + \lambda f^{-1}\left(\frac{Z - \gamma}{\delta}\right),$$

(1.7)

where $Z$ is a pseudo-random normal variable. This approach is used, in an extension of our study, in Chapter 7.

The Johnson bivariate system (Johnson, 1949b) is a natural extension of the univariate case and arguably provides the simplest method of approximating a bivariate distribution. Consider $X_1$ and $X_2$ having marginal transformations to normality
representing systems $S_I$ and $S_J$ respectively ($I, J = L, U, B$ and $N$ to include the normal distribution itself). The associated bivariate system $S_{IJ}$ for the joint distribution of $X_1$ and $X_2$ follows from the assumption that $Z_1$ and $Z_2$ have the standard bivariate normal distribution

\[
f(Z_1, Z_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left[ -\frac{1}{2(1-\rho^2)} \left( Z_1^2 - 2\rho Z_1 Z_2 + Z_2^2 \right) \right]. \tag{1.8}\]

Nine parameters are required: four for each marginal transformation plus $\rho$ the correlation coefficient of the transformed variates. There are ten distinct systems including $S_{NN}$, the bivariate normal itself. In fact the "logarithmic correlation surface" $S_{LL}$ was introduced by Wicksell (1917), and the "semi-logarithmic correlation surface" $S_{NL}$ by Yuan (1933).

It must be remembered that $Z_1$ and $Z_2$ being marginally normally distributed does not guarantee that their joint distribution is given by (1.8). This assumption is a further aspect of the approximation. Problems encountered with fitting $S_{IJ}$ prompt further discussion on the underlying assumptions in Chapter 5.

Conditional (array) distributions take a particularly simple form in this system. Assuming $Z_1$ and $Z_2$ are standard bivariate normal variates then it is well known that the conditional distribution
of $Z_2$ given $Z_1$ is $N(\rho Z_1, 1-\rho^2)$. This also represents the
conditional distribution of $\gamma_2 + \delta_2 f_J(Y_2)$ given $Y_1$ (arguing in
terms of the simpler representation of (1.6)). Therefore

$$\left[ \gamma_2 + \delta_2 f_J(Y_2) - \rho(\gamma_1 + \delta_1 f_I(Y_1)) \right](1-\rho^2)^{-\frac{1}{2}}$$

is $N(0,1)$. We conclude that given $Y_1$ the conditional distribution
of $Y_2$ follows the same system as its marginal distribution with

$\gamma_2$ replaced by $\left[ \gamma_2 + \rho(\gamma_1 + \delta_1 f_I(Y_1)) \right](1-\rho^2)^{-\frac{1}{2}}$

and $\delta_2$ replaced by $\delta_2(1-\rho^2)^{-\frac{1}{2}}$.

Johnson gives further details and discusses particular examples.

Finally we note that although $E(Y_2|Y_1)$, the regression line
of $Y_2$ on $Y_1$, is generally complicated, the median of the conditional
distribution of $Y_2$ given $Y_1$, the so-called median regression, is
quite simple. This median $\tilde{Y}_2$ has a corresponding $Z_2$ value of
zero. Therefore

$$\gamma_2 + \delta_2 f_J(\tilde{Y}_2) - \rho(\gamma_1 + \delta_1 f_I(Y_1)) = 0$$

giving

$$f_J(\tilde{Y}_2) = (\rho \gamma_1 - \gamma_2)\delta_2^{-1} + \rho \delta_1 \delta_2^{-1} f_I(Y_1).$$

Explicit functions and graphs are given by Johnson. It is worth
remarking that for the median regression to be linear we require

$$\rho \delta_1 = \delta_2, \quad \rho \gamma_1 = \gamma_2 \quad \text{and} \quad f_I \equiv f_J,$$

or that the system is $S_{NN^*}$ the bivariate normal case.

The next chapter is devoted to the problems of obtaining the
required moments to fit the Johnson approximations.
Chapter Two

The Moments and Correlation of $t$ and $\sqrt{b_1}$ for a Gram–Charlier Population

0 moments big as years!  

John Keats.

2.1 THE MOMENTS OF THE $t$ STATISTIC.

In our search for a bivariate Johnson representation of the joint distribution of $t$ and $\sqrt{b_1}$ in sampling from a Gram–Charlier population we require the first four moments of $t$. These may be obtained from expansions given by Geary (1936, 1947). The second paper gives expressions for the first six cumulants of $t$ to $O(n^{-2})$ in terms of the population cumulants. The first four are

$$L_1 = -\frac{1}{n^4} \left( \frac{\lambda^3}{2} + \frac{3}{16n} \left( 2\lambda_3 - 2\lambda_5 + 5\lambda_3 \lambda_4 \right) \right) + \ldots$$

$$L_2 = 1 + \frac{1}{4n} \left( 8 + 7\lambda_3^2 \right) + \frac{1}{n^2} \left( 6 - 2\lambda_4 - \frac{3}{8} \lambda_3^2 - \frac{45}{8} \lambda_3 \lambda_5 \right. + \left. \frac{177}{16} \lambda_3^2 \lambda_4 \right) + \ldots$$

$$L_3 = -\frac{2}{n^3} \lambda_3 - \frac{1}{n^{3/2}} \left( 9\lambda_3 - 3\lambda_5 + \frac{15}{4} \lambda_3 \lambda_4 + \frac{83}{8} \lambda_3^3 \right) + \ldots$$

$$L_4 = \frac{1}{n} \left( 6 - 2\lambda_4 + 12\lambda_3^2 \right) + \frac{1}{n^2} \left( 54 - 18\lambda_4 + 41\lambda_6 + 75\lambda_3^2 \right. - \left. 63\lambda_3 \lambda_5 - 6\lambda_4^2 + 81\lambda_3^2 \lambda_4 + \frac{699}{8} \lambda_3^4 \right) + \ldots$$

where $\lambda = \frac{\kappa_m}{\kappa_2^{1/2}}$
and $\kappa_m$ represents the $m^{th}$ population cumulant. In more familiar notation

$$\lambda_3 = \sqrt{\beta_1}$$

$$\lambda_4 = \beta_2 - 3.$$

The above formulae simplify for a Gram–Charlier population since the cumulants (and therefore the $\lambda$'s) of order five and above are zero. We may then write

$$E(t) = L_1$$

$$V(t) = L_2$$

$$\sqrt{\beta_1}(t) = \frac{L_3}{L_2^{3/2}}$$

$$\beta_2(t) = 3 + \frac{L_4}{L_2^2}.$$

In Chapter 1 simplified expressions for these moments were given and we noted that for small to moderate sample sizes population skewness will particularly affect the expected value and skewness of $t$. One convenient way to appreciate the effect of skewness and kurtosis on the moments of a statistic is to plot contours in the $\beta_1 - \beta_2$ plane where the moments have constant value. Contours of $E(t)$ and $\sqrt{\beta_1}(t)$ for sample size 25 using the above expressions, and obtained with the GHOST-80 graphical output system (Prior, 1985) using cubic splines, are shown in Figures 2.1 and 2.2. They illustrate the position for positive population skewness; if negative the sign of all the contour values would be positive.
Figure 2.1

Contours of $E(t)$ for a Gram-Charlier population, $n = 25$, using Geary (1947).
--- indicates the positive definite region.
The contours labelled 1 to 7 have values $-0.02(-0.02)-0.14$; that for zero coincident with the $\beta_2$ axis.
Figure 2.2
Contours of $\sqrt{B_1(t)}$ for a Gram-Charlier population, $n = 25$, using Geary (1947). 
--- indicates the positive definite region.
The contours labelled 1 to 5 have values $-0.1(-0.1)-0.5$; that for value zero coincident with 
the $\beta_2$ axis.
The diagrams clearly show the effect of increasing population skewness in making both $E(t)$ and $\sqrt{\beta_1}(t)$ more negative. Increasing population kurtosis has a similar effect, though it is much less marked except where the population skewness is large.

2.2 THE MOMENTS OF $\sqrt{b_1}$.

Consider now the moments of $\sqrt{b_1}$. For a normal population one may cite Pearson (1930) who obtained expansions for the standard deviation and kurtosis using the then recently introduced $k$-statistics (Fisher, 1929). In the same year Fisher (1930) obtained moments of the related cumulant ratio in closed form. A clear account of the former may be found in Kendall and Stuart (1977, Section 12.18) and this approach is adopted here since its extension to the non-normal case is relatively straightforward. One should also mention the computerised approach to obtaining the moments of moment statistics of Shenton, Bowman and Sheehan (1971). Comparison of the methods is postponed until Chapter 4 where further estimates based on simulation will also be presented.

The $k$-statistics are simply the unbiased estimators of the corresponding population cumulants, i.e.

$$E(k_r) = \kappa_r.$$  

The first four are

$$k_1 = \frac{1}{n} s_1$$

$$k_2 = \frac{1}{n(2)} (ns_2 - s_2^2)$$
\( k_3 = \frac{1}{n^{3/3}} \left( n^2 s_3^2 - 3 n s_2 s_1 + 2 s_1^3 \right) \)

\( k_4 = \frac{1}{n^{4/4}} \left\{ (n^3 + n^2) s_4 - 4(n^2 + n) s_3 s_1 - 3(n^2 - n) s_2^2 + 12 n s_2 s_1^2 - 6 s_1^4 \right\} \),

where \( s_r = \sum x^r \)

and \( n[p] = n(n-1)\ldots(n-p+1) \).

In terms of the sample moments these are

\[ k_1 = m_1 \]

\[ k_2 = \frac{n}{n-1} m_2 \]

\[ k_3 = \frac{n^2}{(n-1)(n-2)} m_3 \]

\[ k_4 = \frac{n^2}{(n-1)(n-2)(n-3)} \left\{ (n+1)m_4 - 3(n-1)m_2^2 \right\} \).

Fisher introduced the \( k \)-statistics as a means of obtaining the sampling moments of moment statistics like \( \sqrt{b_1} \), although their use seems largely confined to the normal case.

We first express \( \sqrt{b_1} \) in terms of the \( k \)-statistics:

\[ \sqrt{b_1} = C g \]

where \( C = \frac{n - 2}{\sqrt{n(n-1)}} \)

and \( g = \frac{k_3}{k_2^{3/2}} \).
Now write
\[ g = \frac{k_3}{\kappa_2^{3/2}} \left( 1 + \frac{k_2 - \kappa_2}{\kappa_2} \right)^{-3/2} \]

where \( \kappa_r \) denotes the \( r \)th population cumulant. Expanding, and writing
\[ w = \frac{k_2 - \kappa_2}{\kappa_2} \]
for brevity, we have
\[ g = \frac{k_3}{\kappa_2^{3/2}} \left\{ 1 - \frac{3}{2} w + \frac{15}{8} w^2 - \frac{35}{16} w^3 + \frac{315}{128} w^4 - \frac{693}{256} w^5 \\
+ \frac{3003}{1024} w^6 - \frac{6435}{2048} w^7 + \ldots \right\} . \]

Taking expectations, and denoting \( E\{k_3(k_2 - \kappa_2)^r\} \) by \( \mu^c(32^r) \), we have
\[ E(g) = \frac{-1}{\kappa_2^{3/2}} \left\{ \mu^c(3) - \frac{3}{2\kappa_2} \mu^c(32) + \frac{15}{8\kappa_2} \mu^c(32^2) - \frac{35}{16\kappa_2} \mu^c(32^3) + \frac{315}{128\kappa_2} \mu^c(32^4) - \frac{693}{256\kappa_2} \mu^c(32^5) + \frac{3003}{1024\kappa_2} \mu^c(32^6) - \frac{6435}{2048\kappa_2} \mu^c(32^7) + \ldots \right\} . \]

To express the moments in terms of cumulants we use the expressions of Cook (1951) and their extension in Kratky, Reinfelds, Hutcheson and Shenton (1972). For example, consider \( \mu^c(32^2) \). The relation expressing the corresponding crude moment in terms of cumulants is
\[ \mu''_{21} = \kappa_{21} + \kappa_{20}\kappa_{01} + 2\kappa_{11}\kappa_{10} + \kappa_{10}\kappa_{01} \]
or equivalently
\[ \mu_{12} = \kappa_{12} + \kappa_{02} \kappa_{10} + 2\kappa_{11} \kappa_{01} + \kappa_{01}^2 \kappa_{10} . \]

Taking the moment of the second variable about its mean we have

\[ \kappa_{12} = \kappa_{12} + \kappa_{02} \kappa_{10} \]

since \[ \kappa_{01} = 0 . \]

Hence \[ \mu^c_{(32^2)} = \kappa(32^2) + \kappa(2^2) \kappa_3 . \]

The other \[ \mu^c_{(32^r)} \] terms are treated in a similar fashion though for clarity all these auxiliary formulae are relegated to Appendix I. Note that \[ \kappa(3^s2^r) \] is \( O(n^{-(s+r-1)}) \) and that the auxiliary formulae are taken as far as \( O(n^{-3}) \) where the published expressions allow.

Substituting in the appropriate formulae from Appendix I we have to \( O(n^{-3}) \)

\[
\mathbb{E}(g) = \frac{1}{\kappa_2^{3/2}} \left\{ \kappa_3 - \frac{3}{2\kappa_2} \kappa(32) + \frac{15}{8\kappa_2^2} \left[ \kappa(32^2) + \kappa(2^2) \kappa_3 \right] 
- \frac{35}{16\kappa_2^3} \left[ \kappa(32^3) + \kappa(2^3) \kappa_3 + 3\kappa(2^2) \kappa(32) \right] 
+ \frac{315}{128\kappa_2^4} \left[ \kappa(2^4) \kappa_3 + 4\kappa(2^3) \kappa(32) + 6\kappa(32^2) \kappa(2^2) + 3\kappa(2^2) \kappa(32) \kappa_3 \right] 
- \frac{693}{256\kappa_2^5} \left[ 10\kappa(2^3) \kappa(2^2) \kappa_3 + 15\kappa^2(2^2) \kappa(32) \right] 
+ \frac{45045}{1024\kappa_2^6} \kappa^3(2^2) \kappa_3 \right\}. \tag{2.1}
\]
We may now express $E(g)$ in terms of the population cumulants using the formulae in Kendall and Stuart (1977, Section 12.16). For example

$$
\kappa(32^2) = \frac{1}{n} \kappa_7 + \frac{16}{n(n-1)} \kappa_5 \kappa_2 + \frac{12(2n-3)}{n(n-1)^2} \kappa_4 \kappa_3 + \frac{48}{(n-1)^2} \kappa_2^2 \kappa_3.
$$

We may now obtain $E(g)$ for any population with known cumulants, though this approach is particularly appropriate for the Gram-Charlier distribution since $\kappa_r = 0$ for $r > 5$. The relevant cumulant formulae are also collected in Appendix I for completeness.

The higher moments may be obtained in a similar way.

$$
E(g^2) = \frac{1}{\kappa_2^2} \left\{ \frac{1}{3} - \frac{3}{\kappa_2^2} \mu c(3^2) + \frac{6}{\kappa_2^2} \mu c(3^2 2^2)ight. \\
- \frac{10}{3} \frac{\mu c(3^2 3)}{\kappa_2} + \frac{15}{4} \frac{\mu c(3^2 4)}{\kappa_2} - \frac{21}{5} \frac{\mu c(3^2 5)}{\kappa_2} \\
+ \frac{28}{6} \frac{\mu c(3^2 6)}{\kappa_2} - \frac{36}{7} \frac{\mu c(3^2 7)}{\kappa_2} + \ldots \right\}.
$$
Using the auxiliary formulae in Appendix I, and again to $O(n^{-3})$,

\[
E(g^2) = \frac{1}{3} \left\{ \kappa(3^2) + \frac{1}{3} \frac{\kappa^2}{3} \left[ \kappa(3^2 2) + 2 \kappa(32) \kappa_3 \right] \right. \\
+ \frac{6}{\kappa_2^2} \left[ \kappa(3^2 2^2) + 2 \kappa(32^2) \kappa_3 + 2 \kappa^2(32) \right. \\
\left. + \kappa(3^2) \kappa(2^2) + \kappa(2^2) \kappa_3^2 \right] \\
- \frac{10}{\kappa_2^3} \left[ 2 \kappa(32^3) \kappa_3 + 3 \kappa(3^2 2) \kappa(2^2) \right. \\
\left. + 6 \kappa(32^2) \kappa(32) + \kappa(2^3) \kappa(3^2) \right. \\
\left. + \kappa(2^3) \kappa_3^2 + 6 \kappa(2^2) \kappa(32) \kappa_3 \right] \\
+ \frac{15}{\kappa_2^4} \left[ \kappa(2^4) \kappa_3^2 + 12 \kappa(32^2) \kappa(2^2) \kappa_3 \right. \\
\left. + 8 \kappa(2^3) \kappa(32) \kappa_3 + 12 \kappa(2^2) \kappa(2^2) \kappa_3^2 \right. \\
\left. + 3 \kappa^2(2^2) \kappa(3^2) + 3 \kappa^2(2^2) \kappa_3^2 \right] \\
- \frac{21}{\kappa_2^5} \left[ 10 \kappa(2^3) \kappa(2^2) \kappa_3^2 + 30 \kappa^2(2^2) \kappa(32) \kappa_3 \right. \\
\left. + \frac{420}{\kappa_2^6} \kappa^3(2^2) \kappa_3^2 \right]. \tag{2.2}
\]

As before we may now express $E(g^2)$ in terms of the population cumulants using the formulae of Kendall and Stuart.
$$g^3 = \frac{k_3^3}{\kappa_2^{9/2}} \left\{ 1 + \frac{k_2 - \kappa_2}{\kappa_2} \right\}^{-9/2}$$

$$= \frac{k_3^3}{\kappa_2^{9/2}} (1 + w)^{-9/2} \quad \text{(say)}$$

$$= \frac{k_3^3}{\kappa_2^{9/2}} \left\{ 1 - \frac{9}{2} w + \frac{99}{8} w^2 - \frac{429}{16} w^3 + \frac{6435}{128} w^4 - \frac{21879}{256} w^5 + \frac{138567}{1024} w^6 - \frac{415701}{2048} w^7 + \ldots \right\}.$$ 

$$E(g^3) = \frac{1}{\kappa_2^{9/2}} \left\{ \mu' (3^3) - \frac{9}{2\kappa_2} \mu^{3c} (3^3_2) + \frac{99}{8\kappa_2} \mu^{3c} (3^3_2^2) - \frac{429}{16\kappa_2} \mu^{3c} (3^3_2^3) + \frac{6435}{128\kappa_2} \mu^{3c} (3^3_2^4) - \frac{21879}{256\kappa_2} \mu^{3c} (3^3_2^5) + \frac{138567}{1024\kappa_2} \mu^{3c} (3^3_2^6) - \frac{415701}{2048\kappa_2} \mu^{3c} (3^3_2^7) + \ldots \right\}.$$ 

We now use the auxiliary formulae as before, their pattern suggesting the following to be correct to $O(n^{-3})$.

$$E(g^3) = \frac{1}{\kappa_2^{9/2}} \left\{ \kappa(3^3) + 3\kappa(3^2) \kappa_3 + \kappa^3_3 - \frac{9}{2\kappa_2} \left[ \kappa(3^3_2) + 3\kappa(3^2_2) \kappa_3 + 3\kappa(3^2) \kappa(32) + 3\kappa(32) \kappa_3^2 \right] + \frac{99}{8\kappa_2} \left[ 3\kappa(3^2_2) \kappa_3 + 3\kappa(32^2) \kappa(3^2) + 6\kappa(3^2_2) \kappa(32) + \kappa(3^3) \kappa(2^3) + 3\kappa(32^2) \kappa_3^2 + 6\kappa^2(32) \kappa_3 + 3\kappa(3^2) \kappa(2^3) \kappa_3 + \kappa(2^3) \kappa_3^3 \right] \right\}.$$
\[
- \frac{429}{16 \kappa^3} \left[ 3 \kappa (2 \kappa)^3 \right] \kappa_3^2 + 3 \kappa (2 \kappa)^3 \kappa (3 \kappa)^2 \kappa_3^2 + 18 \kappa (2 \kappa)^2 \kappa (3 \kappa)^2 \kappa_3^2 \\
+ 9 \kappa (2 \kappa)^2 \kappa (2 \kappa)^3 \kappa_3^2 + \kappa (2 \kappa)^3 \kappa_3^3 + 6 \kappa (3 \kappa)^2 \kappa_3^2 \\
+ 9 \kappa (2 \kappa)^2 \kappa (3 \kappa)^2 \kappa_3^3 + 9 \kappa (3 \kappa)^2 \kappa (2 \kappa)^2 \kappa_3^3 \\
+ 36 \kappa (2 \kappa)^2 \kappa (3 \kappa)^2 \kappa_3 + 9 \kappa (2 \kappa)^2 \kappa (2 \kappa)^3 \kappa_3^2 \\
+ 643 \kappa^3 \kappa_3^3 + 18 \kappa (3 \kappa)^2 \kappa (2 \kappa)^3 \kappa_3^2 + 12 \kappa (2 \kappa)^3 \kappa (3 \kappa)^2 \kappa_3^2 \\
+ 36 \kappa (2 \kappa)^2 \kappa (3 \kappa)^2 \kappa_3 + 9 \kappa (2 \kappa)^2 \kappa (2 \kappa)^3 \kappa_3^2 \\
+ 3 \kappa (2 \kappa)^3 \kappa_3^3 \right] \\
- \frac{21879}{256 \kappa^2} \left[ 10 \kappa (2 \kappa)^3 \kappa (2 \kappa)^3 \kappa_3^3 + 45 \kappa (2 \kappa)^2 \kappa (2 \kappa)^2 \kappa (3 \kappa)^2 \kappa_3^2 \right] \\
+ \frac{2078505}{1024 \kappa^6} \kappa (2 \kappa)^3 \kappa_3^3 \right) . 
\]

(2.3)

In fact this expression is only used to $O(n^{-2})$ as all the necessary cumulant formulae are not available.

\[
G^4 = \frac{k_3^4}{\kappa^3} \left[ 1 + \frac{k_2^2 - k_2}{\kappa^2} \right]^{-6} \\
= \frac{k_3^4}{\kappa^3} (1 + w)^{-3} \quad \text{(say)} \\
= \frac{k_3^4}{\kappa^3} \left[ 1 - 6w + 21w^2 - 56w^3 + 126w^4 - 252w^5 + 462w^6 - 792w^7 + \ldots \right] .
\]
\[ E(g^4) = \frac{1}{\kappa_2^6} \left\{ \mu^4 (3^4) - \frac{6}{\kappa_2} \mu^2 (3^4_2) + \frac{21}{\kappa_2^2} \mu c(3^4_2) + \frac{56}{\kappa_2^3} \mu c(3^4_2^3) + \frac{126}{\kappa_2^4} \mu c(3^4_2^4) - \frac{252}{\kappa_2^5} \mu c(3^4_2^5) + \frac{462}{\kappa_2^6} \mu c(3^4_2^6) - \frac{792}{\kappa_2^7} \mu c(3^4_2^7) + \ldots \right\} . \]

Even with the comprehensive tables of Kratky et al., we can only express \( E(g^4) \) to \( o(n^{-2}) \).

\[ E(g^4) = \frac{1}{\kappa_2^6} \left\{ 4\kappa (3^3) \kappa_3 + 3\kappa^2 (3^2) + 6\kappa (3^2) \kappa_3^2 + \kappa_3^4 \right\} \\
- \frac{6}{\kappa_2} \left\{ 6\kappa (3^2) \kappa_3^2 + 12\kappa (3^2) \kappa (32) \kappa_3 + 4\kappa (32) \kappa_3^3 \right\} \\
+ \frac{21}{\kappa_2^2} \left\{ 4\kappa (32^2) \kappa_3^3 + 12\kappa^2 (32) \kappa_3^2 + 6\kappa (3^2) \kappa (2^2) \kappa_3^2 + \kappa (2^2) \kappa_3^4 \right\} \]

Finally this may be expressed in terms of the population cumulants as before.

A partial check on equations (2.1) to (2.4) comes from comparing them with Fisher's (1930) results for the moments of \( g \) in sampling from a normal distribution. Assuming normality we have

\[ \kappa (32^R) = 0 \]

and more generally

\[ \kappa (p^q_2^R) = 0 \]
if $pq$ is odd or if $q = 1$ and $p > 2$ (Kendall and Stuart, 1977, p. 316-7).

We find that (2.1) and (2.3) reduce, as expected, to

$$E(g) = E(g^3) = 0.$$ 

Equation (2.2) becomes (after yet more algebra)

$$E(g^2) = \frac{6n(n^2-8n+35)}{(n-1)^3(n-2)}$$

or, expanding in powers of $n^{-1}$,

$$E(g^2) = 6\left(\frac{1}{n} - \frac{3}{n^2} + \frac{11}{3n^3} + \frac{89}{n^4} + \ldots\right).$$

Fisher gives

$$E(g^2) = \frac{6n(n-1)}{(n-2)(n+1)(n+3)}$$

$$= 6\left(\frac{1}{n} - \frac{3}{n^2} + \frac{11}{3n^3} - \frac{31}{4n^4} + \ldots\right),$$

agreeing to $O(n^{-3})$. Similarly (2.4) becomes

$$E(g^4) = \frac{108n^2}{(n-1)^2(n-2)^2}$$

while Fisher gives

$$E(g^4) = \frac{108n^2(n-1)^2(n^2+27n-70)}{(n-2)^3(n+1)(n+3)(n+5)(n+7)(n+9)},$$

clearly agreeing to $O(n^{-2})$. 
Glancing ahead to Chapter 4 we find that the expression for $E(g)$ to be quite useful, while that for $E(g^2)$ only acceptable for large samples and small departures from normality. Regretably the expressions for $E(g^3)$ and $E(g^4)$, at least to this order, do not repay the algebraic effort expended, though they might possibly find a role in asymptotic arguments.

Figures 2.3 and 2.4 illustrate the contours of

$$E(\sqrt{b_1}) = CE(g)$$

and

$$SD(\sqrt{b_1}) = CE\{E(g^2) - E(g)^2\}$$

for sample size 100 using (2.1) and (2.2).

We note that $E(\sqrt{b_1})$ increases with population skewness as expected, and the tendency for $SD(\sqrt{b_1})$ to increase with population kurtosis, though Figure 2.4 should not be considered very accurate for extreme non-normality (particularly in kurtosis).

2.3 The correlation of $t$ and $\sqrt{b_1}$.

We now derive an expansion for the correlation coefficient of $t$ and $\sqrt{b_1}$ using the techniques of the previous section. The question of the correlation of transformed variates (needed for the $s_{1j}$ representation of $t$ and $\sqrt{b_1}$) is discussed in Chapter 5.

The correlation coefficient of $t$ and $\sqrt{b_1}$ is given by

$$\rho_{t\sqrt{b_1}} = \frac{E(t\sqrt{b_1}) - E(t)E(\sqrt{b_1})}{SD(t)SD(\sqrt{b_1})}.$$
Figure 2.3

Contours of $E(\nu_{b_1})$ for a Gram-Charlier population, $n = 100$, using (2.1).

---- indicates the positive definite region.

The contours labelled 1 to 5 have values $0.2(0.2)1.0$; that for value zero coincident with the $b_2$ axis.
Figure 2.4

Contours of $\text{SD}(\sqrt{b_1})$ for a Gram-Charlier population, $n = 100$, using (2.2).

--- indicates the positive definite region.

The contours labelled 1 to 10 have values 0.2(0.05)0.65.
We therefore require the expected value of the product $t\sqrt{b_1}$. Now

$$t = \frac{(x-n)\sqrt{n}}{s}$$

which may be expressed

$$t = \sqrt{\frac{n}{k_2}} \frac{k_3}{k_2^{3/2}}$$

if the population mean is zero. (Recall that we consider the Gram-Charlier population in standard measure.) From the previous section we have

$$\sqrt{b_1} = C_g$$

where

$$C = \frac{n - 2}{\sqrt{n(n-1)}}$$

and

$$g = \frac{k_3}{k_2^{3/2}}.$$

We now derive an expansion for $E(tg)$.

$$tg = \sqrt{\frac{n}{k_2}} \frac{k_1}{k_2^{1/2}} \cdot \frac{k_3}{k_2^{3/2}}$$

$$= \sqrt{\frac{n}{k_2}} \frac{k_1 k_3}{k_2^{3/2}}$$

$$= \sqrt{\frac{n}{k_2}} \left(1 + \frac{k_3}{k_2} \right)^{-2}$$

$$= \sqrt{\frac{n}{k_2}} \left(1 + w \right)^{-2}\quad\text{(say)}$$

$$= \sqrt{\frac{n}{k_2}} \frac{k_1 k_3}{k_2^{3/2}} \left\{1 - 2w + 3w^2 - 4w^3 + 5w^4 - 6w^5 + 7w^6 - 8w^7 + \ldots \right\}.$$
Taking expectations, and denoting \( E(k_1 k_3 (k_2 - \kappa_2)^\tau) \) by \( \mu^{''c}(132\tau) \), we have

\[
E(tg) = \sqrt{\frac{\kappa_2}{\kappa}} \left\{ \mu^{''}(13) - \frac{2}{\kappa_2^2} \mu^{''c}(132) + \frac{3}{\kappa_2^2} \mu^{''c}(132^2) - \frac{4}{\kappa_2^3} \mu^{''c}(132^3) + \frac{5}{\kappa_2^4} \mu^{''c}(132^4) - \frac{6}{\kappa_2^5} \mu^{''c}(132^5) + \frac{7}{\kappa_2^6} \mu^{''c}(132^6) - \ldots \right\}.
\]

It proves convenient to reorder the components in the moment terms at this stage as follows.

\[
E(tg) = \sqrt{\frac{\kappa_2}{\kappa}} \left\{ \mu^{''}(13) - \frac{2}{\kappa_2^2} \mu^{''c}(213) + \frac{3}{\kappa_2^2} \mu^{''c}(2^213) - \frac{4}{\kappa_2^3} \mu^{''c}(2^313) + \frac{5}{\kappa_2^4} \mu^{''c}(2^413) - \frac{6}{\kappa_2^5} \mu^{''c}(2^513) + \frac{7}{\kappa_2^6} \mu^{''c}(2^613) - \ldots \right\}.
\]

As for the moments of \( \sqrt{\kappa} \) in the previous section we now use expressions relating the moments in terms of cumulants. For example

\[
\mu^{'''}_{211} = \kappa_{211} + 2\kappa_{111} \kappa_{100} + \kappa_{201} \kappa_{010} + \kappa_{210} \kappa_{001} + 2\kappa_{110} \kappa_{101} + \kappa_{200} \kappa_{011} + \kappa_{211} \kappa_{100} + 2\kappa_{101} \kappa_{100} \kappa_{010} + 2\kappa_{110} \kappa_{100} \kappa_{001} + \kappa_{100} \kappa_{010} \kappa_{001}.
\]

Taking the moment of the first variable about its mean we have
\[ \mu_{211}'' = \kappa_{211} + \kappa_{201} \kappa_{010} + \kappa_{210} \kappa_{001} + 2 \kappa_{110} \kappa_{101} + \kappa_{200} \kappa_{011} + \kappa_{200} \kappa_{010} \kappa_{001} \]

since \( \kappa_{100} = 0 \).

Hence

\[ \mu_{211}''(213) = \kappa(213) + \kappa(2^23) \kappa_1 + \kappa(2^13) \kappa_3 + 2 \kappa(21) \kappa(23) + \kappa(2^2) \kappa(13) + \kappa(2^2) \kappa_1 \kappa_3. \]

Again these auxiliary formulae have been collected in Appendix I.

Substituting in the appropriate formulae we have

\[ E(tg) = \frac{\sqrt{\pi}}{\kappa_2} \left\{ \kappa(13) + \kappa_3 \kappa_1 - \frac{2}{\kappa_2} \left[ \kappa(213) + \kappa(21) \kappa_3 + \kappa(23) \kappa_1 \right] \right. \]

\[ + \frac{3}{\kappa_2} \left[ \kappa(2^213) + \kappa(2^23) \kappa_1 + \kappa(2^13) \kappa_3 + 2 \kappa(21) \kappa(23) + \kappa(2^2) \kappa(13) + \kappa(2^2) \kappa_3 \kappa_1 \right]. \]

\[ - \frac{4}{\kappa_2} \left[ \kappa(2^31) \kappa_3 + \kappa(2^33) \kappa_1 + 3 \kappa(213) \kappa(2^2) + 3 \kappa(2^21) \kappa(23) + 3 \kappa(2^23) \kappa(21) + \kappa(2^3) \kappa(13) + \kappa(2^3) \kappa_3 \kappa_1 \right] \]

\[ + 3 \kappa(2^2) \kappa(21) \kappa_3 + 3 \kappa(2^2) \kappa(23) \kappa_1 \]

\[ + \frac{5}{\kappa_2} \left[ \kappa(2^4) \kappa_3 \kappa_1 + 6 \kappa(2^33) \kappa(2^2) \kappa_1 + 6 \kappa(2^21) \kappa(2^2) \kappa_3 \right] \]
\[ + 4\kappa(2^3) \kappa(23) \kappa_1 + 4\kappa(2^3) \kappa(21) \kappa_3 \\
+ 12\kappa(2^2) \kappa(21) \kappa(23) + 3\kappa^2(2^2) \kappa(13) \\
+ 3\kappa^2(2^2) \kappa_3 \kappa_1 \right\} \\
- \frac{6}{\kappa_2^5 \left\{ 10\kappa(2^3) \kappa(2^2) \kappa_3 \kappa_1 + 15\kappa^2(2^2) \kappa(21) \kappa_3 \\
+ 15\kappa^2(2^2) \kappa(23) \kappa_1 \right\} + \ldots \right\}. \tag{2.5} \]

The above expression is used to \( O(n^{-2}) \) as it was not possible to complete the \( O(n^{-3}) \) terms. The formulae allowing us to express (2.5) in terms of the population cumulants may be found in Appendix I.

Simulation results to be presented in Chapter 5 show (2.5) (like (2.2) for \( E(g^2) \)) to be rather inaccurate except for large samples and small departures from normality. Nevertheless the contours in Figure 2.5 have been obtained using this expression for sample size 100 since they do give a fair indication of the way \( \rho_{t \sqrt{\beta_1}} \) varies over the \( \beta_1 - \beta_2 \) plane for a Gram-Charlier population.

We see that in essence \( \rho_{t \sqrt{\beta_1}} \) increases with kurtosis and decreases with skewness. Remembering that our main aim is to investigate the use of \( \sqrt{\beta_1} \) in a preliminary test of normality before using Student's \( t \) distribution this correlation may be expected to be of fundamental importance.
Figure 2.5

Contours of the correlation of $t$ and $\sqrt{b_1}$ for a Gram-Charlier population, $n = 100$, using (2.5).

----- indicates the positive definite region.
The contours labelled 1 to 6 have values $0.1(0.1)0.6$. 
Chapter Three

A Package of Algorithms for the Gram-Charlier Distribution

3.1 INTRODUCTION.

To investigate the accuracy of the moment approximations in the previous chapter a Gram-Charlier pseudo-random number generator was sought. No such algorithm could be traced (confirmed by Sowey, 1986) and work began developing one, resulting in the "package" of Gram-Charlier algorithms written in Fortran described here. The problems and methods associated with some of the algorithms are described and a final section gives a brief specification of them all. The results of tests of the generators are presented in Chapter 4 and the complete listing forms Appendix II.

For background we first sketch two well-known methods for generating variates having a specified distribution: the inverse distribution method and the rejection method.

Suppose we wish to generate a variate X (assumed continuous) having probability density function \( f(x) \) and distribution function \( F(x) \). The method based on the inverse distribution function \( F^{-1}(.) \) is as follows.

1. Generate \( U \sim U(0,1) \).
2. Determine \( X = F^{-1}(U) \).

It is easy to show that \( X \) has the required distribution.

\[
P(X \leq x) = P(F^{-1}(U) \leq x)
= P(U \leq F(x))
= F(x)
\]
The obvious limitation of the method is in obtaining \( F^{-1}(.) \) which for many common distributions is not available in closed form.

The rejection method (von Neumann, 1951) requires \( X \) to be defined over a finite interval. Let this interval be \((A,B)\) and further suppose the density is bounded above by \( M \), i.e. \( f(x) \leq M \). See Figure 3.1.

**Figure 3.1**
Illustration of the rejection method.

![Figure 3.1 Illustration of the rejection method.](image)

1. Generate \( U_1, U_2 \sim U(0,1) \).
2. Determine \( X = A + (B-A)U_2 \) (i.e. \( X \sim U(A,B) \)).
3. If \( MU_1 \leq f(X) \) accept \( X \), otherwise repeat (1)-(3).

One can show that \( X \) has the required distribution as follows.
\[ P(X \leq x/\text{accept}) = \frac{P(X \leq x \cap \text{accept})}{P(\text{accept})} \]

\[ = \int_{A}^{x} \frac{f(x)dx/M(B-A)}{1/M(B-A)} \]

\[ = F(x). \]

Evidently the efficiency of the procedure depends on \(1/M(B-A)\).

For further discussion of these and other methods see, for example, Morgan (1984).

3.2 A MODIFIED REJECTION METHOD FOR THE GRAM-CHARLIER DISTRIBUTION.

Considering the Gram-Charlier distribution

\[ f(x) = \phi(x) \left(1 + \frac{k_3}{3!} H_3(x) + \frac{k_4}{4!} H_4(x)\right) \]

we note that the rejection method cannot be used without modification because of its infinite range of support. The approach eventually adopted (following a suggestion of Professor A.C. Atkinson and much experimentation) combines rejection in the centre with an inversion technique applied to tails fitted to the distribution function. See Figure 3.2.

**Figure 3.2**

Illustration of the Modified Rejection Method.
The procedure is "initialised" by determining the values of A, B, H and four parameters used to fit the tails $T_L(x)$ and $T_U(x)$.

A and B, corresponding to small tail areas $P_A$ and $1 - P_B$ and defining the central rejection region, are obtained by calling the inverse distribution function. This is described in the next section.

In the algorithms listed in Appendix II we have set $P_A = 0.005$ and $P_B = 0.995$, giving 99% of the distribution generated by the "true" rejection method, though the relevant DATA statements may be easily altered.

For greatest rejection efficiency we obtain the maximum value of the pdf H (corresponding to smallest M in Figure 3.1), and for simplicity restrict our attention to unimodal curves. The mode is obtained by solving $f'(x) = 0$ using the Newton-Raphson method. (The numerical methods used in this chapter are discussed, for example, in Phillips and Taylor (1973).) Writing for convenience the Gram-Charlier density in its general form

$$f(x) = \phi(x) \sum a_r H_r(x)$$

we have

$$f'(x) = \phi(x) \sum a_r H'_r(x) - x\phi(x) \sum a_r H_r(x)$$

$$= -\phi(x) \sum a_r (xH_r(x) - H'_r(x)) .$$

Now

$$H'_r(x) = r H_{r-1}(x)$$

(Kendall and Stuart, 1977, Section 6.14), though the above expression may be simplified by noting
This may be shown as follows. By definition

\[
D^r \phi(x) = (-1)^r H_r(x) \phi(x) \quad (3.1)
\]

hence

\[
D^{r+1} \phi(x) = (-1)^r \left\{ H'_r(x) \phi(x) - x H_r(x) \phi(x) \right\}
\]

\[
= (-1)^{r+1} \left\{ x H_{r+1}(x) - H'_r(x) \right\} \phi(x)
\]

but

\[
D^{r+1} \phi(x) = (-1)^{r+1} H_{r+1}(x) \phi(x) .
\]

Writing therefore

\[
f'(x) = -\phi(x) \sum a_r H_{r+1}(x)
\]

\[= -\phi(x) q(x) \]

(and noting the form of \(\phi(x)\)) we solve

\[q(x) = 0\]

using the iterative procedure

\[
x_{r+1} = x_r - \frac{q(x_r)}{q'_1(x_r)}
\]

where

\[q(x) = x + \frac{\kappa_3}{6} H_4(x) + \frac{\kappa_4}{24} H_5(x),\]

\[q'(x) = 1 + \frac{2}{3} \kappa_3 H_3(x) + \frac{5}{24} \kappa_4 H_4(x) .\]
and with starting value $x_0 = 0$.

Regarding the tails we fit

$$T_L(x) = \exp\left\{-\left(\frac{x - \theta_L}{\sigma_L}\right)^2\right\}$$

(3.3)

to the lower tail (i.e. $-\infty$ to $A$) of the distribution function, determining $\theta_L$ and $\sigma_L$ so as to match the ordinate and first derivative of $F(x)$ at $A$. Similarly

$$T_U(x) = 1 - \exp\left\{-\left(\frac{x - \theta_U}{\sigma_U}\right)^2\right\}$$

(3.4)

is fitted to the upper tail ($B$ to $\infty$) with $\theta_U$ and $\sigma_U$ determined for agreement at $B$.

Considering $T_L(x)$ we require

$$T_L(A) = F(A) = P_A$$
$$T_L'(A) = F'(A) = f(A).$$

The first criterion gives

$$\exp\left\{-\left(\frac{x - \theta_L}{\sigma_L}\right)^2\right\} = P_A$$

$$\left(\frac{x - \theta_L}{\sigma_L}\right)^2 = -\ln P_A$$

$$\theta_L = A - \sigma_L (-\ln P_A)^{\frac{1}{2}}.$$
Noting that
\[ T_L(x) = - \frac{2}{\sigma_L} \left( \frac{x - \theta_L}{\sigma_L} \right) \exp \left\{ - \left( \frac{x - \theta_L}{\sigma_L} \right)^2 \right\} \]

the second criterion gives
\[ - \frac{2}{\sigma_L^2} (A - \theta_L) \exp \left\{ - \left( \frac{A - \theta_L}{\sigma_L} \right)^2 \right\} = f(A) . \]

Consideration of the expressions above allow this to be written
\[ - \frac{2}{\sigma_L^2} \sigma_L (-\bar{n}_A \bar{P})^{\frac{1}{2}} \bar{P}_A = f(A) \]
giving
\[ \sigma_L = - \frac{2 \bar{P}_A (-\bar{n}_A \bar{P})^{\frac{1}{2}}}{f(A)} . \]

In a similar manner the values of \( \theta_U \) and \( \sigma_U \) in \( T_U(x) \) may be shown to be given by
\[ \theta_U = B - \sigma_U (-\bar{n}_U \theta_B)^{\frac{1}{2}} \]
\[ \sigma_U = \frac{2 \bar{P}_B (-\bar{n}_U \theta_B)^{\frac{1}{2}}}{f(B)} \]

where \( \theta_B = 1 - \bar{P}_B \).

The justification for choosing this particular form of approximation for the tails of the Gram-Charlier distribution function is entirely empirical: other approximations were tried but did not perform nearly so well in terms of the sampling moments of the k-statistics (discussed
in Chapter 4). However there is some theoretical support if we note that the Gram-Charlier distribution will be dominated by its standard normal term in the extreme tails. Considering therefore the standard normal distribution function \( \Phi(x) \) and writing

\[
1 - \Phi(x) = q = z^{-1}
\]

we have (Kendall and Stuart, 1977, p. 158)

\[
x \sim (2z_n z)^{1/2} \left( 1 - \frac{\xi_n \lambda + \xi_n \xi_n z}{2z_n n} \right)
\]

i.e.

\[
x \sim (2z_n z)^{1/2} \left( 1 - \frac{K_1}{(z_n z)^{1/2}} - \frac{K_2}{(z_n z)^{1/2}} \right)
\]

\((K_1, K_2 \text{ constants})\). As \( z \to \infty \) (i.e. in the extreme upper tail) the second and third terms tend to zero, the latter being apparent from l'Hospital's rule:

\[
\lim_{z \to \infty} \frac{\xi_n \xi_n z}{(z_n z)^{1/2}} = \lim_{z \to \infty} \frac{1}{\frac{2}{z_n n} z} \frac{1}{\frac{1}{z}} = \lim_{z \to \infty} \frac{2}{(z_n z)^{1/2}}.
\]

Therefore a very rough approximation to an upper percentage point is given by

\[
x = (2z_n z)^{1/2}
\]

\[
= \sqrt{2}(-z_n q)^{1/2}.
\]
The inverse of (3.4) is a linear transformation of this expression.

Having called the initialisation procedure to determine $A$, $B$, $H$, $\theta_L$, $\sigma_L$, $\theta_U$ and $\sigma_U$ a sequence of pseudo-random Gram-Charlier variates may be generated. The method consists of generating a $U(0,1)$ variate which determines the region of the Gram-Charlier variate: lower tail, centre or upper tail. If in a tail the inverse of (3.3) or (3.4) is used with the uniform variate as argument; if in the centre the Gram-Charlier variate is obtained by rejection. The algorithm is summarised as follows:

1. Generate $U_1 \sim U(0,1)$

2. If $U_1 < P_A$ lower tail: $X = \theta_L + \sigma_L (-\frac{1}{n} U_1)^{\frac{1}{n}}$,  
   if $U_1 > P_B$ upper tail: $X = \theta_U + \sigma_U (-\frac{1}{n} (1-U_1))^{\frac{1}{n}}$,  
   otherwise centre: Rejection:
   
   (A) Generate $U_2, U_3 \sim U(0,1)$.
   (B) $X = A + (B-A)U_3$.
   (C) If $H U_2 \leq f(X)$ accept $X$, otherwise repeat (A)-(C).

3.3 THE INVERSE GRAM-CHARLIER DISTRIBUTION FUNCTION.

Writing again the Gram-Charlier density

$$f(x) = \phi(x) \sum_{x} \alpha_H (x)$$

we obtain the distribution function $F(x)$. 
\[ F(x) = \int_{-\infty}^{\infty} \phi(x) \sum a_r H_r(x) \]
\[ = \int_{-\infty}^{\infty} \phi(x) \, dx + \int_{-\infty}^{\infty} \sum_{r=1}^{\infty} a_r H_r(x) \phi(x) \, dx , \]

it being assumed that \( \kappa_3, \kappa_4 \) lie within the positive definite region.

Now from (3.1)
\[ \int H_r(x) \phi(x) \, dx = -H_{r-1}(x) \phi(x) \]

hence
\[ F(x) = \phi(x) - \phi(x) \sum_{r=1}^{r-1} a_r H_r(x) \]

i.e.
\[ F(x) = \phi(x) - \phi(x) \left\{ \frac{\kappa_3}{6} H_2(x) + \frac{\kappa_4}{24} H_3(x) \right\} . \]

The inverse \( F^{-1}(.) \) is not obviously amenable to a closed form solution. However the availability of approximations to \( \phi^{-1} \) does suggest one iterative approach. Writing

\[ F(x) = \phi(x) - D(x) = p \]

we have

\[ x_{r+1} = \phi^{-1}(p + D(x_r)) \]

with starting value

\[ x_0 = \phi^{-1}(p) . \]
Though such procedures have only first order convergence this approach was found to be satisfactory in the centre of the distribution; however it was prone to failure in the tails caused by the argument of $\phi^{-1}(.)$ falling outside the range 0 to 1. Rather than attempting to get around this difficulty (e.g. introducing a multiplier into the argument) the more direct (and second order) Newton-Raphson method was again used.

Writing

$$G(x) = F(x) - p$$

we solve $G(x) = 0$.

Noting $G'(x) = f(x)$

we have

$$x_{r+1} = x_r - \frac{(F(x_r) - p)}{f(x_r)}$$

with starting value

$$x_0 = \phi^{-1}(p)$$

In the extreme tails ($p < 10^{-6}$ or $p > 1 - 10^{-6}$ in the algorithm listed in Appendix II) $\phi^{-1}(p)$ provides an approximate solution. The algorithm also contains a "back-up" regular falsi routine to be used in the (very rare) event of Newton-Raphson failing to converge.

3.4 DETERMINATION OF THE GRAM-CHARLIER REGION.

We first confirm that the Gram-Charlier distribution satisfies the general expression

$$\beta_2 > 1 + \beta_1$$
(Kendall and Stuart, 1977, p. 95). In terms of the cumulants of the standardised distribution this is

\[ \kappa_4 > \kappa_3^2 - 2. \]

Assuming this condition is satisfied we check whether the distribution lies within the positive-definite and unimodal regions illustrated earlier in Figure 1.1. For a positive-definite (i.e. valid) distribution we require \( f(x) > 0 \) for all \( x \). Writing

\[
\begin{align*}
f(x) &= \phi(x) \left\{ 1 + \frac{\kappa_3}{3} H_3(x) + \frac{\kappa_4}{4} H_4(x) \right\} \\
&= \phi(x) \left\{ 1 + a_3 H_3(x) + a_4 H_4(x) \right\} \\
&= \phi(x) p(x)
\end{align*}
\]

we require

\[ p(x) > 0 \text{ for all } x. \]

The boundary, in terms of \( a_3 \) and \( a_4 \), is given by the solution of

\[
\begin{align*}
p(x) &= 0 \\
p'(x) &= 0
\end{align*}
\]

(see for example Gow, 1960, Section 19.13). We have
\begin{align*}
p(x) &= 1 + a_3 H_3(x) + a_4 H_4(x) \\
&= 1 + a_3(x^3 - 3x) + a_4(x^4 - 6x^2 + 3) \\
p'(x) &= 3a_3 H_2(x) + 4a_4 H_3(x) \\
&= 3a_3(x^2 - 1) + 4a_4(x^3 - 3x) \\
\end{align*}

giving

\[
\begin{align*}
4a_4 x^4 + a_3 x^3 - 6a_4 x^2 - 3a_3 x + 3a_4 + 1 &= 0 \\
-4a_4 x^3 + 3a_3 x^2 - 12a_4 x - 3a_3 &= 0
\end{align*}
\]

Using the dialytic method of Sylvester (Muir, 1960, Section 266) the solution is given by

\[
\Delta_p = 0
\]

where

\[
\Delta_p = \begin{vmatrix}
a_4 & a_3 & -6a_4 & -3a_3 & 3a_4 + 1 & 0 & 0 \\
0 & a_4 & a_3 & -6a_4 & -3a_3 & 3a_4 + 1 & 0 \\
0 & 0 & a_4 & a_3 & -6a_4 & -3a_3 & 3a_4 + 1 \\
4a_4 & 3a_3 & -12a_4 & -3a_3 & 0 & 0 & 0 \\
0 & 4a_4 & 3a_3 & -12a_4 & -3a_3 & 0 & 0 \\
0 & 0 & 4a_3 & 3a_3 & -12a_4 & -3a_3 & 0 \\
0 & 0 & 0 & 4a_4 & 3a_3 & -12a_4 & -3a_3
\end{vmatrix}
\]

The positive-definite region illustrated in Figure 1.1 was obtained by Barton and Dennis (1952) in this way. However Shenton (1951) using
another argument gives the positive definite criteria

\[(8\kappa_4 + 2\kappa_4^2 + 4\kappa_3^2)^3 > (24\kappa_4^2 + 2\kappa_4^3 + 24\kappa_3^2 + 6\kappa_3^2\kappa_4)^2\]

and

\[4\kappa_4^3 - \kappa_4^2 + 4\kappa_4\kappa_3^2 > 3\kappa_3^2\kappa_4 + 4\kappa_3^4.\]

In fact \(\Delta_p\) is positive within the positive-definite region and negative without, but for simplicity the Shenton criteria were used.

For the unimodal problem consider the solution of

\[f'(x) = 0.\]

Again following Barton and Dennis the unimodal region will be bounded by the curve where \(f(x)\) passes from one to three modes, i.e., \(f(x)\) has a point of inflexion. Now (equation (3.2))

\[f'(x) = -\phi(x) \sum a_r H_{r+1}(x)\]

and therefore

\[f''(x) = \phi(x) \sum a_r H_{r+2}(x)\]

giving the conditions

\[
\begin{align*}
H_2(x) + a_3 H_5(x) + a_4 H_6(x) &= 0 \\
H_1(x) + a_3 H_4(x) + a_4 H_5(x) &= 0
\end{align*}
\]

Explicitly these are
\[
\left\{ \begin{array}{l}
ax^6 + a_3x^5 - 15a_4x^4 - 10a_3x^3 + (45a_4 + 1)x^2 + 15a_3x - 15a_4 - 1 = 0 \\
ax^5 + a_3x^4 - 10a_4x^3 - 6a_3x^2 + (15a_4 + 1)x + 3a_3 = 0
\end{array} \right.
\]

with corresponding determinant
\[ \Delta_U = \begin{bmatrix} a_4 & a_3 & -15a_4 & -10a_3 & (45a_4 + 1) & 15a_3 & (-15a_4 - 1) & 0 & 0 & 0 & 0 \\ 0 & a_4 & a_3 & -15a_4 & -10a_3 & (45a_4 + 1) & 15a_3 & (-15a_4 - 1) & 0 & 0 & 0 \\ 0 & 0 & a_4 & a_3 & -15a_4 & -10a_3 & (45a_4 + 1) & 15a_3 & (-15a_4 - 1) & 0 & 0 \\ 0 & 0 & 0 & a_4 & a_3 & -15a_4 & -10a_3 & (45a_4 + 1) & 15a_3 & (-15a_4 - 1) & 0 \\ 0 & 0 & 0 & 0 & a_4 & a_3 & -15a_4 & -10a_3 & (45a_4 + 1) & 15a_3 & (-15a_4 - 1) \\ 0 & a_4 & a_3 & -10a_4 & -6a_3 & (15a_4 + 1) & 3a_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_4 & a_3 & -10a_4 & -6a_3 & (15a_4 + 1) & 3a_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_4 & a_3 & -10a_4 & -6a_3 & (15a_4 + 1) & 3a_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_4 & a_3 & -10a_4 & -6a_3 & (15a_4 + 1) & 3a_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & a_4 & a_3 & -10a_4 & -6a_3 & (15a_4 + 1) & 3a_3 \end{bmatrix} \]
This formidable determinant is used to check for unimodality being negative inside the region, zero on the boundary and generally positive outside. However there is a small area above the "node" - determined to be at (4.71, 0.688) - where $\Delta_U$ passes from zero to negative. To avoid any problems in this area, and the unnecessary evaluation of the determinant, unimodality is rejected if the point corresponding to $\kappa_3, \kappa_4$ lies above the line $\beta_1 = 0.668$ or to the right of the line joining (5.4, 0) to (4.71, 0.688). It may of course be possible to derive simpler criteria.
3.5 **SPECIFICATION OF THE ALGORITHMS IN GCPACK.FOR.**

**MAIN ALGORITHMS.**

SUBROUTINE INRVGC(C3,C4,H,A,THL,SIGL,B,THU,SIGU,IFault)

DESCRIPTION:
Intialises the parameters required by RVGC.

**INPUT PARAMETERS:**
- C3 Real : \( \kappa_3 \), 3rd population cumulant
- C4 Real : \( \kappa_4 \), 4th population cumulant

**OUTPUT PARAMETERS:**
- H Real : Value of \( f(x) \) at mode
- A Real : Percentile \( P_{0.5} \)
- THL Real : \( \theta_L \)
- SIGL Real : \( \sigma_L \)
- B Real : Percentile \( P_{99.5} \)
- THU Real : \( \theta_U \)
- SIGU Real : \( \sigma_U \)

IFault: Integer: Failure indicator:
- 0 success
- 1,2,3 \( \kappa_3 \) and \( \kappa_4 \) outside unimodal region (see ITYPGC)
- 4 XMOLDGC failure (for H)
- 5 PPGC failure (for A or B)

**INTERNAL CALLS:**
- ITYPGC,XMODGC,FGC,PPGC

**EXTERNAL CALLS:**
- ALNORM (implicit),PPND (implicit)

**NOTES:**
Called once before generating a sequence of pseudo-random variates using RVGC.
The areas corresponding to the percentiles A and B are denoted PA and PB in the DATA statement and must be consistent with their values in RVGC.

REAL FUNCTION RVGC(C3,C4,H,A,THL,SIGL,B,THU,SIGU)

DESCRIPTION:
Generates a pseudo-random variate from the Gram-Charlier distribution (restricted to unimodal case).

INPUT PARAMETERS

\begin{align*}
C3 & \quad \text{Real: 3rd and 4th population cumulants} \\
C4 & \\
H & \\
A & \\
THL & \\
SIGL & \quad \text{Real: Parameters determined by INRVGC} \\
B & \\
THU & \\
SIGU &
\end{align*}

OUTPUT FUNCTION VALUE:

\begin{align*}
\text{Real: The pseudo-random variate}
\end{align*}

INTERNAL CALLS:

FGC

EXTERNAL CALLS:

G85CAF

NOTES:

Takes approximately 5.9 seconds (CPU time on the Trent Polytechnic DEC-20) to generate $10^4$ observations when $\kappa_3 = \sqrt{0.3}$, $\kappa_4 = 2.0$. 
REAL FUNCTION PPGC(C3,C4,P,IFAULT)

DESCRIPTION:
The quantile (percentage point) of the Gram-Charlier distribution corresponding to the lower tail area.

INPUT PARAMETERS:

\[
\begin{align*}
C3 & \quad \text{Real: 3rd and 4th population cumulants} \\
C4 & \\
P & \quad \text{Real: Lower tail area}
\end{align*}
\]

OUTPUT FUNCTION VALUE:

Real: The quantile

OUTPUT PARAMETER:

IFAULT. Integer: Failure indicator:

0 Success
1 $P$ outside range $(0,1)$
2 normal approximation in extreme tails
3 failure to converge

INTERNAL CALLS:

FFGC,DFGC

EXTERNAL CALLS:

ALNORM (implicit),PPND

NOTES:
The procedure is regulated by constants in the DATA statement.
In particular PPACC controls the accuracy.
If $P$ is generated from U$(0,1)$ PPGC provides a pseudo-random Gram-Charlier variate.
Using $G5CAF$ for $P$, PPGC takes approximately 13.5 seconds to generate $10^4$ observations when $\kappa_3 = \sqrt{0.3}$, $\kappa_4 = 2.0$. 
FUNCTION ITYPGC(C3,C4)

DESCRIPTION:

Determines whether C3 and C4 define a positive-definite or unimodal Gram-Charlier distribution.

INPUT PARAMETERS:

C3  }  Real: 3rd and 4th population cumulants
C4  }

OUTPUT FUNCTION VALUE:

Integer: 0 unimodal
1 positive-definite (but not unimodal)
2 outside positive-definite region
3 impossible region ($s_2 < 1 + s_1$)

INTERNAL CALLS:

DETLU

REAL FUNCTION FGC(C3,C4,X)

DESCRIPTION:

The probability density function of the Gram-Charlier distribution.

INPUT PARAMETERS:

C3 }  Real: 3rd and 4th population cumulants
C4
X  Real: Gram-Charlier variate

OUTPUT FUNCTION VALUE:

Real: The function evaluated at X

REAL FUNCTION DFGC(C3,C4,X)

DESCRIPTION:

The distribution function of the Gram-Charlier distribution.
INPUT PARAMETERS:

\[
\begin{align*}
C3 & \quad \text{Real: 3rd and 4th population cumulants} \\
C4 & \\
X & \quad \text{Real: Gram-Charlier variate}
\end{align*}
\]

OUTPUT FUNCTION VALUE:

- Real: The function evaluated at \( X \)

EXTERNAL CALL:

ALNORM

REAL FUNCTION XMODGC(C3,C4,IFAULT)

DESCRIPTION:

The mode of the Gram-Charlier distribution (assumed unimodal).

INPUT PARAMETERS:

\[
\begin{align*}
C3 & \quad \text{Real: 3rd and 4th population cumulants} \\
C4 & \\
X & \quad \text{Real: Gram-Charlier variate}
\end{align*}
\]

OUTPUT FUNCTION VALUE:

- Real: The mode of the distribution

OUTPUT PARAMETER:

IFAULT Integer: Failure indicator:

- 0 success
- 1 not converged

INTERNAL CALLS:

QGC,3DGC
INTERNAL AUXILIARY ALGORITHMS.

REAL FUNCTION QGC(C3,C4,X)

DESCRIPTION:
Function to obtain the derivative of the Gram-Charlier pdf
(for XMODGC).

INPUT PARAMETERS:

\[
\begin{align*}
C3 & \quad \text{Real: 3rd and 4th population cumulants} \\
C4 & \\
X & \quad \text{Real: Gram-Charlier variate}
\end{align*}
\]

OUTPUT FUNCTION VALUE:
Real: The function evaluated at X

REAL FUNCTION QDGC(C3,C4,X)

DESCRIPTION:
Derivative of QGC (for XMODGC).

INPUT PARAMETERS:

\[
\begin{align*}
C3 & \quad \text{Real: 3rd and 4th population cumulants} \\
C4 & \\
X & \quad \text{Real: Gram-Charlier variate}
\end{align*}
\]

OUTPUT FUNCTION VALUE:
Real: The function evaluated at X
REAL FUNCTION DETLU(A,N)

DESCRIPTION:

The determinant of a square matrix.

INPUT PARAMETERS:

A Real: Square matrix

N×N array

N Integer: Dimension of matrix

OUTPUT FUNCTION VALUE:

Real: The determinant

EXTERNAL AUXILIARY ALGORITHMS.

REAL FUNCTION ALNORM(X,.UPPER.)

DESCRIPTION:

Tail area of the standard normal distribution.

INPUT PARAMETERS:

X Real: Standard normal variate

.UPPER. Logical: Tail required

.True. from X to infinity

.False. from minus infinity to X

OUTPUT FUNCTION VALUE:

Real: Tail area

SOURCE:


NOTES:

ALNORM is called with .UPPER. set to .FALSE..

May be replaced by any standard normal lower tail area (distribution function) algorithm.
REAL FUNCTION PPND(P, IFAULT)

DESCRIPTION:

The quantile (percentage point) of the standard normal distribution

   corresponding to the lower tail area.

INPUT PARAMETER:

P    Real: Lower tail area

OUTPUT FUNCTION VALUE:

Real: The quantile

OUTPUT PARAMETER:

IFAULT Integer: Failure indicator:

   0  success

   1  P outside range (0,1)

SOURCE:


NOTES:

May be replaced by any standard normal quantile (percentage point)

   algorithm.

REAL FUNCTION G05CAF(DUM)

DESCRIPTION:

Generates a pseudo-random variate uniformly distributed on the

   interval (0,1)

INPUT PARAMETER:

DUM    Real: Dummy argument (required by FORTRAN syntax)

OUTPUT FUNCTION VALUE:

Real: The pseudo-random variate

SOURCE:

The NAG FORTRAN Library Mark 10 (Numerical Algorithms Group, 1984).
NOTES:

A repeatable sequence of variates, corresponding to integer I, may be obtained by first calling subroutine GØ5CBF(I); for a non-repeatable sequence first call GØ5CCF.

May be replaced by any $U(0,1)$ generator.
Chapter Four

Assessing the Gram-Charlier Generators

and the Moments of $\sqrt{b_1}$

4.1 THE k-STATISTIC PROPERTIES OF THE GRAM-CHARLIER GENERATORS.

We consider the two algorithms for generating Gram-Charlier random variables

(1) RVGC - the modified rejection method

(2) PPGC - the inverse distribution function

described in the previous chapter. Clearly both depend critically on the auxiliary $U(0,1)$ generator. We use G05CAF in the NAG library (Numerical Algorithms Group, 1984), a multiplicative congruential generator having the form

$$N = 13^{13} \times N \mod 2^{59}$$

$$G05CAF = N/2^{59}.$$ 

This was designed on the basis of the results of spectral tests up to dimension 8 and has been subjected to extensive empirical tests (Knuth, 1981; Maclaren, personal communication). Apart from the fact that the inverse distribution function is obtained by an iterative approximation method (2) is "exact" (with any non-uniformity in the auxiliary generator being reflected in an obvious way). Method (1) however is an approximation and needs testing, indeed it was developed with the aid of the criteria about to be discussed.

Now we saw in Chapter 1 that the Gram-Charlier distribution is completely defined by its third and fourth cumulants. Remembering the
distribution is in standard measure we have

\[ \kappa_1 = \mu = 0 \]
\[ \kappa_2 = \sigma^2 = 1 \]
\[ \kappa_3 = \sqrt{\beta_1} \]
\[ \kappa_4 = \beta_2 - 3 \]
\[ \kappa_r = 0 \quad r \geq 5. \]

Fisher's k-statistics were introduced in Chapter 2 and used to obtain moment approximations. Being unbiased estimators of their corresponding population cumulants their sampling properties should provide a useful check on the Gram-Charlier generators. The first four were defined in Section 2.2. Additionally we use \( k_5 \) and \( k_6 \):

\[ k_5 = \frac{1}{n^5} \left\{ (n^4 + 5n^3)s_5 - 5(n^3 + 5n^2)s_4s_1 - 10(n^3 - n^2)s_3s_2 \right. \\
+ 20(n^2 + 2n)s_3s_1^2 + 30(n^2 - n)s_2^2s_1 \\
- 60n s_2^3 + 24s_1^5 \} \]

\[ k_6 = \frac{1}{n^6} \left\{ (n^5 + 16n^4 + 11n^3 - 4n^2)s_6 - 6(n^4 + 16n^3 + 11n^2 - 4n)s_5s_1 \\
- 15n(n - 1)^2(n + 4)s_4s_2^2 - 10(n^4 - 2n^3 + 5n^2 - 4n)s_3^2 \\
+ 30(n^3 + 9n^2 + 2n)s_4s_1^2 + 120(n^3 - n)s_3s_2s_1 \\
+ 30(n^3 - 3n^2 + 2n)s_2^3 - 120(n^2 + 3n)s_3^3s_1 \\
- 270(n^2 - n)s_2^2s_1^2 + 360n s_2^4s_1^2 - 120 s_1^6 \} . \]
We wish to verify that the means of the k-statistics in simulated sampling are close to their corresponding theoretical cumulants. To this end we also obtain their theoretical variances using the formulae in Section 12.16 Kendall and Stuart (1977).

\[
V(k_1) = \frac{2}{n} \, (k_1 = \overline{x})
\]
\[
= \frac{1}{n}
\]

\[
V(k_2) = E(k_2 - \kappa_2)^2
\]
\[
= \mu(2^2)
\]
\[
= \kappa(2^2)
\]
\[
= \frac{\kappa_4}{n} + \frac{2\kappa_2}{n-1}
\]
\[
= \frac{\kappa_4}{n} + \frac{2}{n-1}
\]

\[
V(k_3) = \kappa(3^2)
\]
\[
= \frac{1}{n} \kappa_6 + \frac{9}{n-1} \kappa_4 \kappa_2 + \frac{9}{n-1} \kappa_3^2 + \frac{6n}{(n-1)(n-2)} \kappa_2^3
\]
\[
= \frac{9}{n-1} \kappa_4 + \frac{9}{n-1} \kappa_3^2 + \frac{6n}{(n-1)(n-2)}
\]

\[
V(k_4) = \kappa(4^2)
\]
\[
= \frac{1}{n} \kappa_8 + \frac{16}{n-1} \kappa_6 \kappa_2 + \frac{48}{n-1} \kappa_5 \kappa_3 + \frac{34}{n-1} \kappa_4^2
\]
\[
+ \frac{72n}{(n-1)(n-2)} \kappa_4 \kappa_2^2 + \frac{144n}{(n-1)(n-2)} \kappa_3 \kappa_2^2 + \frac{24n(n+1)}{(n-1)(n-2)(n-3)} \kappa_2^4
\]
\[
= \frac{34}{n-1} \kappa_4^2 + \frac{72n}{(n-1)(n-2)} \kappa_4 + \frac{144n}{(n-1)(n-2)} \kappa_3^2 + \frac{24n(n+1)}{(n-1)(n-2)(n-3)} \kappa_2^4.
\]
Now writing directly in the Gram-Charlier form for brevity

\[ V(k_5) = \kappa(5^2) \]
\[ = \frac{850n}{(n-1)(n-2)} \kappa_4^2 + \frac{1500n}{(n-1)(n-2)} \kappa_4 \kappa_3 + \frac{6000(n+1)}{(n-1)(n-2)(n-3)} \kappa_4 \]
\[ + \frac{1800n(n+1)}{(n-1)(n-2)(n-3)} \kappa_3^2 + \frac{120n^2(n+5)}{(n-1)(n-2)(n-3)(n-4)} \]

\[ V(k_6) = \kappa(6^2) \]
\[ = \frac{4950n}{(n-1)(n-2)} \kappa_4^3 + \frac{n(n+1)}{(n-1)(n-2)(n-3)} \left[ \frac{15300 \kappa_4^2 + 54000 \kappa_4 \kappa_3}{(n-1)(n-2)(n-3)} \right]
\[ + \frac{8100 \kappa_4 \kappa_3^2}{(n-1)(n-2)(n-3)(n-4)} \left( 5400 \kappa_4 \right)
\[ + \frac{21600 \kappa_3^3}{(n-1)(n-2)(n-3)(n-4)(n-5)} \right]. \]

The estimated means of the first six \( k \)-statistics for samples of size 50 at ten points in the \( \beta_1 - \beta_2 \) plane (identified on Figure 1.1) are shown in Table 4.1. Within the unimodal region the results from both methods are shown; outside (but within the positive definite region) only the inverse distribution function is used. Both are based on ten thousand replications. The table also shows the theoretical and estimated standard errors of \( k_r \), and differences between \( \hat{E}(k_r) \) and \( k_r \) in units of the standard error of estimate (i.e. \( (\hat{E}(k_r) - k_r)100/SE(k_r) \)) greater in magnitude than 2 and 3 are indicated by dashed and solid underlining respectively.
from Gram-Charlier distributions.

(a) Theoretical, where for all distributions

\[ E(k_1) = 0, \ E(k_2) = 1, \ E(k_3) = \kappa_3, \ E(k_4) = \kappa_4, \ E(k_5) = E(k_6) = 0. \]

(b) Simulation using the modified rejection method, 10000 samples.

(c) Simulation using the inverse distribution function, 10000 samples.

Distribution (1) \( \kappa_3 = 0.0 \ \kappa_4 = 1.0 \)

<table>
<thead>
<tr>
<th>( r )</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1414E0</td>
<td>-0.1833E-2</td>
<td>0.1416E0</td>
</tr>
<tr>
<td>2</td>
<td>0.2466E0</td>
<td>0.1004E1</td>
<td>0.2499E0</td>
</tr>
<tr>
<td>3</td>
<td>0.5579E0</td>
<td>-0.2306E-2</td>
<td>0.5946E0</td>
</tr>
<tr>
<td>4</td>
<td>0.1667E1</td>
<td>\textcolor{red}{0.1086E1}</td>
<td>0.1736E1</td>
</tr>
<tr>
<td>5</td>
<td>0.5929E1</td>
<td>0.2452E-1</td>
<td>0.5884E1</td>
</tr>
<tr>
<td>6</td>
<td>0.2510E2</td>
<td>\textcolor{red}{0.1468E1}</td>
<td>0.2231E2</td>
</tr>
</tbody>
</table>
Table 4.1 (continued)

Distribution (2) $\kappa_3 = 0.0$ $\kappa_4 = 2.0$

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>SE($k_r$)</td>
<td>E($k_r$)</td>
<td>SE($k_r$)</td>
</tr>
<tr>
<td>1</td>
<td>0.1414E0</td>
<td>-0.8005E-3</td>
<td>0.1415E0</td>
</tr>
<tr>
<td>2</td>
<td>0.2843E0</td>
<td>0.1007E1</td>
<td>0.2912E0</td>
</tr>
<tr>
<td>3</td>
<td>0.7035E0</td>
<td>-0.4579E-2</td>
<td>0.7455E0</td>
</tr>
<tr>
<td>4</td>
<td>0.2528E1</td>
<td>0.2121E1</td>
<td>0.2198E1</td>
</tr>
<tr>
<td>5</td>
<td>0.1016E2</td>
<td>-0.8106E-1</td>
<td>0.7525E1</td>
</tr>
<tr>
<td>6</td>
<td>0.5071E2</td>
<td>0.1702E1</td>
<td>0.2978E2</td>
</tr>
</tbody>
</table>

Distribution (4) $\kappa_3 = \sqrt{0.3} = 0.5477$ $\kappa_4 = 1.0$

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>SE($k_r$)</td>
<td>E($k_r$)</td>
<td>SE($k_r$)</td>
</tr>
<tr>
<td>1</td>
<td>0.1414E0</td>
<td>0.1012E-2</td>
<td>0.1399E0</td>
</tr>
<tr>
<td>2</td>
<td>0.2466E0</td>
<td>0.9994E0</td>
<td>0.2464E0</td>
</tr>
<tr>
<td>3</td>
<td>0.6052E0</td>
<td>0.5695E0</td>
<td>0.5526E0</td>
</tr>
<tr>
<td>4</td>
<td>0.1923E1</td>
<td>0.9948E0</td>
<td>0.1435E1</td>
</tr>
<tr>
<td>5</td>
<td>0.7562E1</td>
<td>0.2526E0</td>
<td>0.4351E1</td>
</tr>
<tr>
<td>6</td>
<td>0.3458E2</td>
<td>-0.3060E1</td>
<td>0.1588E2</td>
</tr>
</tbody>
</table>
Table 4.1 (continued)

Distribution (5) $\kappa_3 = \sqrt{0.3} = 0.5477$ $\kappa_4 = 2.0$

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>$SE(k_r)$</td>
<td>$E(\hat{k}_r)$</td>
<td>$SE(\hat{k}_r)$</td>
</tr>
<tr>
<td>1</td>
<td>0.1414EO</td>
<td>0.6951E-3</td>
<td>0.1414EO</td>
</tr>
<tr>
<td>2</td>
<td>0.2843EO</td>
<td>0.1006E1</td>
<td>0.2874EO</td>
</tr>
<tr>
<td>3</td>
<td>0.7416EO</td>
<td>0.5350EO</td>
<td>0.7526EO</td>
</tr>
<tr>
<td>4</td>
<td>0.2703E1</td>
<td>0.2152E1</td>
<td>0.234CE1</td>
</tr>
<tr>
<td>5</td>
<td>0.1161E2</td>
<td>-0.4546EO</td>
<td>0.8468E1</td>
</tr>
<tr>
<td>6</td>
<td>0.5925E2</td>
<td>-0.1439EO</td>
<td>0.3677E2</td>
</tr>
</tbody>
</table>

Distribution (3) $\kappa_3 = 0.0$ $\kappa_4 = 3.5$

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>$SE(k_r)$</td>
<td>$E(\hat{k}_r)$</td>
</tr>
<tr>
<td>1</td>
<td>0.1414EO</td>
<td>-0.6761E-3</td>
</tr>
<tr>
<td>2</td>
<td>0.3329EO</td>
<td>0.1001E1</td>
</tr>
<tr>
<td>3</td>
<td>0.8777EO</td>
<td>-0.7361E-2</td>
</tr>
<tr>
<td>4</td>
<td>0.3796E1</td>
<td>0.3523E1</td>
</tr>
<tr>
<td>5</td>
<td>0.1652E2</td>
<td>-0.6883E-1</td>
</tr>
<tr>
<td>6</td>
<td>0.9681E2</td>
<td>0.7676E-1</td>
</tr>
</tbody>
</table>
Table 4.1 (continued)

Distribution (6) $\kappa_3 = \sqrt{0.3} = 0.5477$ $\kappa_4 = 3.5$

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>$SE(k_r)$</td>
<td>$E(k_r)$</td>
</tr>
<tr>
<td>1</td>
<td>0.1414E0</td>
<td>0.3655E-2</td>
</tr>
<tr>
<td>2</td>
<td>0.3329E0</td>
<td>0.1006E1</td>
</tr>
<tr>
<td>3</td>
<td>0.9086E0</td>
<td>0.5651E0</td>
</tr>
<tr>
<td>4</td>
<td>0.3915E1</td>
<td>0.3538E1</td>
</tr>
<tr>
<td>5</td>
<td>0.1786E2</td>
<td>0.627OE-2</td>
</tr>
<tr>
<td>6</td>
<td>0.1043E3</td>
<td>-0.3182E1</td>
</tr>
</tbody>
</table>

Distribution (7) $\kappa_3 = \sqrt{0.7} = 0.8367$ $\kappa_4 = 2.0$

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>$SE(k_r)$</td>
<td>$E(k_r)$</td>
</tr>
<tr>
<td>1</td>
<td>0.1414E0</td>
<td>0.2625E-2</td>
</tr>
<tr>
<td>2</td>
<td>0.2843E0</td>
<td>0.9992E0</td>
</tr>
<tr>
<td>3</td>
<td>0.7896E0</td>
<td>0.8415E0</td>
</tr>
<tr>
<td>4</td>
<td>0.2921E1</td>
<td>0.1941E1</td>
</tr>
<tr>
<td>5</td>
<td>0.1330E2</td>
<td>0.5985E-2</td>
</tr>
<tr>
<td>6</td>
<td>0.6940E2</td>
<td>-0.7457E1</td>
</tr>
</tbody>
</table>
Table 4.1 (continued)

Distribution (8) $\kappa_3 = \sqrt{0.7} = 0.8367$ $\kappa_4 = 3.0$

<table>
<thead>
<tr>
<th>r</th>
<th>$SE(k_r)$</th>
<th>$E(\hat{k}_r)$</th>
<th>$SE(\hat{k}_r)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1414E0</td>
<td>0.3406E-2</td>
<td>0.1419E0</td>
</tr>
<tr>
<td>2</td>
<td>0.3175E0</td>
<td>0.9986E0</td>
<td>0.3191E0</td>
</tr>
<tr>
<td>3</td>
<td>0.8984E0</td>
<td>0.8524E0</td>
<td>0.8158E0</td>
</tr>
<tr>
<td>4</td>
<td>0.3679E1</td>
<td>0.2968E1</td>
<td>0.2186E1</td>
</tr>
<tr>
<td>5</td>
<td>0.1742E2</td>
<td>0.2150E-1</td>
<td>0.7214E1</td>
</tr>
<tr>
<td>6</td>
<td>0.9797E2</td>
<td>-0.7616E1</td>
<td>0.3112E2</td>
</tr>
</tbody>
</table>

Distribution (9) $\kappa_3 = 1.0$ $\kappa_4 = 2.0$

<table>
<thead>
<tr>
<th>r</th>
<th>$SE(k_r)$</th>
<th>$E(\hat{k}_r)$</th>
<th>$SE(\hat{k}_r)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1414E0</td>
<td>-0.2266E-3</td>
<td>0.1414E0</td>
</tr>
<tr>
<td>2</td>
<td>0.2843E0</td>
<td>0.1001E1</td>
<td>0.2881E0</td>
</tr>
<tr>
<td>3</td>
<td>0.8238E0</td>
<td>0.1001E1</td>
<td>0.6933E0</td>
</tr>
<tr>
<td>4</td>
<td>0.3074E1</td>
<td>0.1992E1</td>
<td>0.1834E1</td>
</tr>
<tr>
<td>5</td>
<td>0.1444E2</td>
<td>-0.2069E-1</td>
<td>0.6129E1</td>
</tr>
<tr>
<td>6</td>
<td>0.7638E2</td>
<td>-0.1030E2</td>
<td>0.2601E2</td>
</tr>
</tbody>
</table>
Table 4.1 (continued)

Distribution (10) $\kappa_3 = 1.0 \kappa_4 = 2.8$

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(c)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>$SE(k_r)$</td>
<td>$E(k_r)$</td>
<td>$SE(k_r)$</td>
</tr>
<tr>
<td>---</td>
<td>-----------</td>
<td>-----------</td>
<td>-----</td>
</tr>
<tr>
<td>1</td>
<td>0.1414EO</td>
<td>-0.8041E-3</td>
<td>0.1405EO</td>
</tr>
<tr>
<td>2</td>
<td>0.3112EO</td>
<td>0.9992EO</td>
<td>0.3114EO</td>
</tr>
<tr>
<td>3</td>
<td>0.9086EO</td>
<td>0.9924EO</td>
<td>0.7910EO</td>
</tr>
<tr>
<td>4</td>
<td>0.3652E1</td>
<td>0.2820E1</td>
<td>0.2164E1</td>
</tr>
<tr>
<td>5</td>
<td>0.1773E2</td>
<td>0.1591E-2</td>
<td>0.7292E1</td>
</tr>
<tr>
<td>6</td>
<td>0.9885E2</td>
<td>-0.9669E1</td>
<td>0.3163E2</td>
</tr>
</tbody>
</table>

Comparing the $E(k_r)$ with $\kappa_r$ we find that in general there is good agreement when using the inverse distribution function with only $E(k_6)$ tending to be below expectation in the more extreme non-normal cases. The modified rejection method does not perform as well with generally larger discrepancies between $E(k_r)$ and $\kappa_r$, though without any obvious pattern among the distributions considered. Thinking in terms of the standard error of estimate may however be too severe from a practical point of view - equivalent to a significant test for the mean based on a sample of 10000. The absolute errors in $E(k_r)$ in terms of $SE(k_r)$ are less than 6% for $k_4$, 4% for $k_5$ and 9% for $k_6$.

Comparing the theoretical and estimated standard errors for both methods one may note for the higher order $k$-statistics a suggestion of the ordering
\( SE(k_r) > SE(k_r^*) > SE(\hat{k}_r^*) \).

Evidently the sampling distribution of the higher order \( k \)-statistics is not sufficiently well represented in the extreme tails, and this is particularly true for the inverse distribution method. (Noting that we are discussing skew distributions this partly explains the earlier remark about \( E(\hat{k}_6) \).

For guidance on the interpretation of these results the same exercise was performed on the NAG normal generator G05DDF (which uses the method of Brent, 1974) and the results, again for ten thousand replications, are shown in Table 4.2.

**Table 4.2**

Mean and standard error of \( k \)-statistics for samples of size 50 from the standard normal distribution.

(a) Theoretical, where \( E(k_1) = 0, E(k_2) = 1, E(k_r) = 0 \ r \geq 3 \).

(b) Simulation using the NAG N(0,1) generator, 10000 samples.

<table>
<thead>
<tr>
<th>( r )</th>
<th>( SE(k_r) )</th>
<th>( E(\hat{k}_r) )</th>
<th>( SE(\hat{k}_r) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1414E0</td>
<td>0.7996E-3</td>
<td>0.1403E0</td>
</tr>
<tr>
<td>2</td>
<td>0.2020E0</td>
<td>0.1001E1</td>
<td>0.2036E0</td>
</tr>
<tr>
<td>3</td>
<td>0.3571E0</td>
<td>-0.5048E-2</td>
<td>0.3557E0</td>
</tr>
<tr>
<td>4</td>
<td>0.7441E0</td>
<td>-0.1165E-2</td>
<td>0.7351E0</td>
</tr>
<tr>
<td>5</td>
<td>0.1801E1</td>
<td>-0.8509E-2</td>
<td>0.1761E1</td>
</tr>
<tr>
<td>6</td>
<td>0.5103E1</td>
<td>0.1180E-1</td>
<td>0.4811E1</td>
</tr>
</tbody>
</table>
The estimated means are seen to be very satisfactory and further there is good agreement between the theoretical and estimated standard errors. Comparison with Table 4.1 shows the standard errors in general to be much smaller than those for a Gram-Charlier distribution. These observations may be explained by noting that the moments of the $k$-statistics, and in turn their moments, will depend on the population cumulants which have value zero for order three and above for a normal distribution. Inverting the normal distribution function (algorithm AS111 Beasley and Springer, 1977) gives substantially the same results, and in particular does not exhibit smaller standard errors for the higher order $k$-statistics.

The examination of the sampling properties of the $k$-statistics seems appropriate when generating distributions like the normal and Gram-Charlier which are defined in terms of their low order cumulants, especially when concerned with the sampling moments of moment statistics whose value depends on those cumulants. One would like to know more, however, about the general effects on a distribution of the high order cumulants.

Some information is provided by the study of Pearson, Johnson and Burr (1979) who investigate the percentage points of eight systems of distribution having common first four moments at points within the $\beta_1 - \beta_2$ plane. Their vast array of results is not easy to summarise, indeed no hard conclusions are drawn. However it may be said that there can be very good agreement apart from in the extreme tails (meaning beyond the 5% to 1% points), that agreement is better in "long" tails than "steep" tails, and generally more agreement may be expected with $\beta_1, \beta_2$ not too far from their normal values (which in their terms
probably includes the Gram-Charlier region. A further study would be needed, however, to assess the extent to which these conclusions apply to the general sampling properties of a distribution.

We now consider in detail the sampling moments of \( \sqrt{b_1} \) for a Gram-Charlier population.

4.2 THE MOMENTS OF \( \sqrt{b_1} \).

We proceed by comparing the moments of \( \sqrt{b_1} \) obtained by the expansions of Chapter 2 with those obtained by simulation. There is however a third source of information which we now briefly review.

Shenton, Bowman and Sheehan (1971) describe a technique for obtaining the moments of moment statistics. A statistic \( t \) (their notation) is expressed as a function of crude moments which in turn are expressed in terms of orthogonal polynomials associated with the particular population. Representing \( t \) by a multivariate Taylor expansion the expectations of powers of \( t \) are obtained using recursive schemes and other numerical techniques made feasible with the availability of powerful computing facilities and sophisticated software. For the moments of the skewness and kurtosis statistics in sampling from a Gram-Charlier population Bowman and Shenton (1971, 1973) are relevant. The tables in Bowman and Shenton (1975) give their most comprehensive published results for the moments of \( \sqrt{b_1} \) and \( b_2 \) providing for a variety of non-normal distributions (Pearson, normal mixture, Gram-Charlier etc.) at certain points in the \( \beta_1 - \beta_2 \) plane terms in asymptotic expansions together with minimum sample sizes for their safe use.
In Table 4.3 are presented estimates of $E(\sqrt{b_1})$ for $n = 25, 50, 100, 200$ at ten points in the $\beta_1-\beta_2$ plane (see Figure 1.1) using

1. formula (2.1),
2. simulation using the inverse distribution function,
3. the coefficients in Bowman and Shenton.

The inverse distribution function was used because of its better k-statistic properties discussed earlier and because it allows us to use the same method over the whole positive definite region; 10000 replications were used for $n = 25, 50, 100$ and 5000 for $n = 200$; estimated standard errors are in parenthesis. Bowman and Shenton do not provide coefficients at distribution (5) (for $\beta_1 = 0.3$ between (4) and (6) they have $\beta_2 = 4.5$ which is rather close to (4)); results for $n$ below the recommended minimum are in parenthesis. It may be recalled that contours of $E(\sqrt{b_1})$ obtained by formula for $n = 100$ were presented earlier in Figure 2.3.
Table 4.3

Estimates of $E(\sqrt{b_1})$ based on the following methods:

Formula (2.1).

Simulation using the inverse distribution function;
10000 samples for $n = 25, 50, 100$; 5000 samples for $n = 200$;
(estimated SE's in units of 0.0001).

Bowman and Shenton (1975);
- not available; (below minimum $n$).

Symmetric Distributions $\beta_1 = 0$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$S_2 = 4.0$</th>
<th>Simulation</th>
<th>$S_2 = 5.0$</th>
<th>Simulation</th>
<th>$S_2 = 6.5$</th>
<th>Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.0094 (62)</td>
<td>0.0069 (78)</td>
<td>0.0328 (106)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>-0.0054 (49)</td>
<td>-0.0063 (64)</td>
<td>-0.0013 (87)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>-0.0014 (36)</td>
<td>-0.0013 (48)</td>
<td>0.0020 (63)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.0036 (38)</td>
<td>-0.0058 (48)</td>
<td>0.0121 (63)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$E(\sqrt{b_1}) = 0$ with
Formula and Bowman-Shenton

Distribution (4) $\beta_1 = 0.3 \ (\sqrt{\beta_1} = 0.5477)\ S_2 = 4.0$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formula</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.4589</td>
<td>0.4409 (55)</td>
<td>0.4320</td>
</tr>
<tr>
<td>50</td>
<td>0.5005</td>
<td>0.4934 (44)</td>
<td>0.4907</td>
</tr>
<tr>
<td>100</td>
<td>0.5228</td>
<td>0.5141 (32)</td>
<td>0.5199</td>
</tr>
<tr>
<td>200</td>
<td>0.5348</td>
<td>0.5336 (33)</td>
<td>0.5341</td>
</tr>
<tr>
<td>$n$</td>
<td>Formula</td>
<td>Simulation</td>
<td>Bowman-Shenton</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>-------------</td>
<td>----------------</td>
</tr>
<tr>
<td>25</td>
<td>0.5300</td>
<td>0.4259 (76)</td>
<td>-</td>
</tr>
<tr>
<td>50</td>
<td>0.5280</td>
<td>0.5062 (59)</td>
<td>-</td>
</tr>
<tr>
<td>100</td>
<td>0.5347</td>
<td>0.5209 (45)</td>
<td>-</td>
</tr>
<tr>
<td>200</td>
<td>0.5404</td>
<td>0.5418 (46)</td>
<td>-</td>
</tr>
</tbody>
</table>

Distribution (6) $\beta_1 = 0.3$ ($\sqrt{\beta_1} = 0.5477$) $\beta_2 = 6.5$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formula</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.7104</td>
<td>0.4723 (105)</td>
<td>(0.4701)</td>
</tr>
<tr>
<td>50</td>
<td>0.5817</td>
<td>0.5123 (82)</td>
<td>0.5239</td>
</tr>
<tr>
<td>100</td>
<td>0.5549</td>
<td>0.5502 (60)</td>
<td>0.5414</td>
</tr>
<tr>
<td>200</td>
<td>0.5492</td>
<td>0.5406 (61)</td>
<td>0.5459</td>
</tr>
</tbody>
</table>

Distribution (7) $\beta_1 = 0.7$ ($\sqrt{\beta_1} = 0.8367$) $\beta_2 = 5.0$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formula</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.8009</td>
<td>0.6730 (69)</td>
<td>0.6692</td>
</tr>
<tr>
<td>50</td>
<td>0.8048</td>
<td>0.7665 (54)</td>
<td>0.7693</td>
</tr>
<tr>
<td>100</td>
<td>0.8164</td>
<td>0.8075 (40)</td>
<td>0.8070</td>
</tr>
<tr>
<td>200</td>
<td>0.8254</td>
<td>0.8211 (40)</td>
<td>0.8230</td>
</tr>
</tbody>
</table>
Table 4.3 (continued)

Distribution (8) $\beta_1 = 0.7$ ($\sqrt{\beta_1} = 0.8367$) $\beta_2 = 6.0$

<table>
<thead>
<tr>
<th>n</th>
<th>Formula</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.9651</td>
<td>0.6991 (89)</td>
<td>(0.6996)</td>
</tr>
<tr>
<td>50</td>
<td>0.8564</td>
<td>0.7828 (70)</td>
<td>0.7902</td>
</tr>
<tr>
<td>100</td>
<td>0.8364</td>
<td>0.8208 (51)</td>
<td>0.8204</td>
</tr>
<tr>
<td>200</td>
<td>0.8342</td>
<td>0.8285 (52)</td>
<td>0.8303</td>
</tr>
</tbody>
</table>

Distribution (9) $\beta_1 = 1.0$ $\beta_2 = 5.0$

<table>
<thead>
<tr>
<th>n</th>
<th>Formula</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.9493</td>
<td>0.8071 (65)</td>
<td>(0.8029)</td>
</tr>
<tr>
<td>50</td>
<td>0.9604</td>
<td>0.9207 (49)</td>
<td>0.9204</td>
</tr>
<tr>
<td>100</td>
<td>0.9755</td>
<td>0.9625 (36)</td>
<td>0.9649</td>
</tr>
<tr>
<td>200</td>
<td>0.9864</td>
<td>0.9858 (37)</td>
<td>0.9837</td>
</tr>
</tbody>
</table>

Distribution (10) $\beta_1 = 1.0$ $\beta_2 = 5.8$

<table>
<thead>
<tr>
<th>n</th>
<th>Formula</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>1.0986</td>
<td>0.8352 (81)</td>
<td>(0.8388)</td>
</tr>
<tr>
<td>50</td>
<td>1.0085</td>
<td>0.9430 (64)</td>
<td>0.9403</td>
</tr>
<tr>
<td>100</td>
<td>0.9944</td>
<td>0.9771 (47)</td>
<td>0.9776</td>
</tr>
<tr>
<td>200</td>
<td>0.9948</td>
<td>0.9992 (47)</td>
<td>0.9907</td>
</tr>
</tbody>
</table>
Examination of Table 4.3 shows close agreement between the Bowman-Shenton and simulation results, indeed there is a suggestion that the minimum sample sizes are rather conservative. Formula (1.1) somewhat overestimates $E(\sqrt{b_1})$, the inaccuracy decreasing with $n$ and increasing with degree of non-normality (particularly kurtosis). As a rough indication it appears useful for $n$ around 25 in the region of distribution (4), increasing to 50 near (5) and requiring rather large $n$ (up to 100) for further non-normality within the positive definite region.

Table 4.4 presents similar comparisons for $V(\sqrt{b_1})$. Contours of the standard deviation by formulae for $n = 100$ were presented in Figure 2.4.
Table 4.4
Estimates of $\text{V}(\nu_b_1)$ based on the following methods:
Formulae (2.1) and (2.2).
Simulation using the inverse distribution function;
10000 samples for $n = 25, 50, 100$; 5000 samples for $n = 200$;
(estimated SE's in units of 0.0001).
Bowman and Shenton (1975);
- not available; (below minimum $n$).

Distribution (1) $\beta_1 = 0$ $\beta_2 = 4.0$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.3002</td>
<td>0.3835 (60)</td>
<td>0.3831</td>
</tr>
<tr>
<td>50</td>
<td>0.2107</td>
<td>0.2381 (35)</td>
<td>0.2380</td>
</tr>
<tr>
<td>100</td>
<td>0.1258</td>
<td>0.1297 (19)</td>
<td>0.1342</td>
</tr>
<tr>
<td>200</td>
<td>0.0687</td>
<td>0.0725 (14)</td>
<td>0.0711</td>
</tr>
</tbody>
</table>

Distribution (2) $\beta_2 = 0$ $\beta_2 = 5.0$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.4591</td>
<td>0.6030 (83)</td>
<td>(0.4530)</td>
</tr>
<tr>
<td>50</td>
<td>0.3106</td>
<td>0.4052 (56)</td>
<td>0.4028</td>
</tr>
<tr>
<td>100</td>
<td>0.1919</td>
<td>0.2279 (31)</td>
<td>0.2239</td>
</tr>
<tr>
<td>200</td>
<td>0.1072</td>
<td>0.1170 (7)</td>
<td>0.1165</td>
</tr>
</tbody>
</table>
Table 4.4 (continued)

Distribution (3) $\beta_1 = 0 \quad \beta_2 = 6.5$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.4962</td>
<td>1.1280 (139)</td>
<td>(2.9163)</td>
</tr>
<tr>
<td>50</td>
<td>0.4129</td>
<td>0.7516 (98)</td>
<td>0.7469</td>
</tr>
<tr>
<td>100</td>
<td>0.2799</td>
<td>0.3920 (56)</td>
<td>0.3867</td>
</tr>
<tr>
<td>200</td>
<td>0.1624</td>
<td>0.1965 (40)</td>
<td>0.1920</td>
</tr>
</tbody>
</table>

Distribution (4) $\beta_1 = 0.3 \quad \beta_2 = 4.0$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.3197</td>
<td>0.3035 (49)</td>
<td>0.3057</td>
</tr>
<tr>
<td>50</td>
<td>0.2037</td>
<td>0.1911 (31)</td>
<td>0.1909</td>
</tr>
<tr>
<td>100</td>
<td>0.1199</td>
<td>0.1030 (16)</td>
<td>0.1042</td>
</tr>
<tr>
<td>200</td>
<td>0.0655</td>
<td>0.0548 (12)</td>
<td>0.0543</td>
</tr>
</tbody>
</table>

Distribution (5) $\beta_1 = 0.3 \quad \beta_2 = 5.0$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.6282</td>
<td>0.5701 (86)</td>
<td>-</td>
</tr>
<tr>
<td>50</td>
<td>0.3404</td>
<td>0.3537 (54)</td>
<td>-</td>
</tr>
<tr>
<td>100</td>
<td>0.1975</td>
<td>0.1985 (29)</td>
<td>-</td>
</tr>
<tr>
<td>200</td>
<td>0.1084</td>
<td>0.1038 (21)</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 4.4 (continued)

### Distribution (6) $\beta_1 = 0.3 \quad \beta_2 = 6.5$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>1.1045</td>
<td>1.0935 (146)</td>
<td>(2.2406)</td>
</tr>
<tr>
<td>50</td>
<td>0.5347</td>
<td>0.6782 (96)</td>
<td>0.6941</td>
</tr>
<tr>
<td>100</td>
<td>0.3090</td>
<td>0.3631 (53)</td>
<td>0.3614</td>
</tr>
<tr>
<td>200</td>
<td>0.1713</td>
<td>0.1863 (37)</td>
<td>0.1803</td>
</tr>
</tbody>
</table>

### Distribution (7) $\beta_1 = 0.7 \quad \beta_2 = 5.0$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.8244</td>
<td>0.4723 (80)</td>
<td>(0.6211)</td>
</tr>
<tr>
<td>50</td>
<td>0.3748</td>
<td>0.2950 (52)</td>
<td>0.2915</td>
</tr>
<tr>
<td>100</td>
<td>0.2040</td>
<td>0.1588 (27)</td>
<td>0.1578</td>
</tr>
<tr>
<td>200</td>
<td>0.1097</td>
<td>0.0799 (18)</td>
<td>0.0819</td>
</tr>
</tbody>
</table>

### Distribution (8) $\beta_1 = 0.7 \quad \beta_2 = 6.0$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>1.5060</td>
<td>0.7877 (125)</td>
<td>(0.9477)</td>
</tr>
<tr>
<td>50</td>
<td>0.5817</td>
<td>0.4965 (79)</td>
<td>0.4994</td>
</tr>
<tr>
<td>100</td>
<td>0.2988</td>
<td>0.2615 (40)</td>
<td>0.2667</td>
</tr>
<tr>
<td>200</td>
<td>0.1586</td>
<td>0.1344 (28)</td>
<td>0.1360</td>
</tr>
</tbody>
</table>
Table 4.4 (continued)

Distribution (9) $\beta_1 = 1.0$, $\beta_2 = 5.0$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.9493</td>
<td>0.4275 (77)</td>
<td>(0.4934)</td>
</tr>
<tr>
<td>50</td>
<td>0.3967</td>
<td>0.2377 (46)</td>
<td>0.2433</td>
</tr>
<tr>
<td>100</td>
<td>0.2081</td>
<td>0.1298 (24)</td>
<td>0.1299</td>
</tr>
<tr>
<td>200</td>
<td>0.1106</td>
<td>0.0667 (16)</td>
<td>0.0673</td>
</tr>
</tbody>
</table>

Distribution (10) $\beta_1 = 1.0$, $\beta_2 = 5.8$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>1.6785</td>
<td>0.6619 (119)</td>
<td>0.4283</td>
</tr>
<tr>
<td>50</td>
<td>0.6027</td>
<td>0.4071 (76)</td>
<td>0.4039</td>
</tr>
<tr>
<td>100</td>
<td>0.2952</td>
<td>0.2176 (37)</td>
<td>0.2172</td>
</tr>
<tr>
<td>200</td>
<td>0.1536</td>
<td>0.1105 (23)</td>
<td>0.1118</td>
</tr>
</tbody>
</table>

Table 4.4 shows again that the Bowman-Shenton results are generally accurate. There is some discrepancy however for the most non-normal distribution (10) and smallest sample size 25, the simulation estimate being some 20 standard errors higher. The minimum sample size in this case is stated as 24 which is apparently too low. The estimates by formula are reasonable for distributions in the region of (1) and (4) for $n = 50$, and useful generally for $n = 100$ within the unimodal region (i.e. including (2) and (5)). Regretably for more extreme non-normality,
particularly in kurtosis, they require very large sample sizes, though may be of use in asymptotic studies.

Tables 4.5 and 4.6 compare the estimates of the third and fourth moments of $\sqrt{b_1}$, respectively.
Table 4.5

Estimates of $\mu_3(\sqrt{b_1})$ based on the following methods:

- Formulae (2.1), (2.2) and (2.3).
- Simulation using the inverse distribution function;
  - 10000 samples for $n = 25, 50, 100$; 5000 samples for $n = 200$.
- Bowman and Shenton (1975);
  - not available; (below minimum $n$).

### Symmetric Distribution $\beta_1 = 0$

<table>
<thead>
<tr>
<th>n</th>
<th>Formulae (1) $\beta_2 = 4.0$</th>
<th>Simulation (2) $\beta_2 = 5.0$</th>
<th>Simulation (3) $\beta_2 = 6.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.0049</td>
<td>0.0159</td>
<td>0.0112</td>
</tr>
<tr>
<td>50</td>
<td>-0.0039</td>
<td>0.0081</td>
<td>-0.0077</td>
</tr>
<tr>
<td>100</td>
<td>0.0014</td>
<td>0.0027</td>
<td>-0.0055</td>
</tr>
<tr>
<td>200</td>
<td>0.0008</td>
<td>-0.0015</td>
<td>-0.0011</td>
</tr>
</tbody>
</table>

$\mu_3(\sqrt{b_1}) = 0$ with Formulae and Bowman-Shenton

### Distribution (4) $\beta_1 = 0.3$ $\beta_2 = 4.0$

<table>
<thead>
<tr>
<th>n</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.0025</td>
<td>-0.0242</td>
<td>(0.0218)</td>
</tr>
<tr>
<td>50</td>
<td>0.0345</td>
<td>-0.0182</td>
<td>-0.0196</td>
</tr>
<tr>
<td>100</td>
<td>0.0139</td>
<td>-0.0106</td>
<td>-0.0083</td>
</tr>
<tr>
<td>200</td>
<td>0.0042</td>
<td>-0.0034</td>
<td>-0.0026</td>
</tr>
</tbody>
</table>
Table 4.5 (continued)

Distribution (5) $\beta_1 = 0.3 \; \beta_2 = 5.0$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>-0.3640</td>
<td>-0.1522</td>
<td>-</td>
</tr>
<tr>
<td>50</td>
<td>0.0037</td>
<td>-0.0925</td>
<td>-</td>
</tr>
<tr>
<td>100</td>
<td>0.0148</td>
<td>-0.0296</td>
<td>-</td>
</tr>
<tr>
<td>200</td>
<td>0.0056</td>
<td>-0.0077</td>
<td>-</td>
</tr>
</tbody>
</table>

Distribution (6) $\beta_1 = 0.3 \; \beta_2 = 6.5$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>-1.4954</td>
<td>-0.5090</td>
<td>(-12.90)</td>
</tr>
<tr>
<td>50</td>
<td>-0.0889</td>
<td>-0.2161</td>
<td>(-0.2408)</td>
</tr>
<tr>
<td>100</td>
<td>0.0118</td>
<td>-0.0683</td>
<td>-0.0589</td>
</tr>
<tr>
<td>200</td>
<td>0.0074</td>
<td>-0.0103</td>
<td>-0.0120</td>
</tr>
</tbody>
</table>

Distribution (7) $\beta_1 = 0.7 \; \beta_2 = 5.0$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>-0.8404</td>
<td>-0.1837</td>
<td>(-1.3396)</td>
</tr>
<tr>
<td>50</td>
<td>-0.0301</td>
<td>-0.1099</td>
<td>-0.1100</td>
</tr>
<tr>
<td>100</td>
<td>0.0187</td>
<td>-0.0376</td>
<td>-0.0361</td>
</tr>
<tr>
<td>200</td>
<td>0.0082</td>
<td>-0.0093</td>
<td>-0.0100</td>
</tr>
</tbody>
</table>
Table 4.5 (continued)

Distribution (8) $\beta_1 = 0.7 \quad \beta_2 = 6.0$

<table>
<thead>
<tr>
<th>n</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>-2.6881</td>
<td>-0.5058</td>
<td>(-5.3913)</td>
</tr>
<tr>
<td>50</td>
<td>-0.1983</td>
<td>-0.2182</td>
<td>(-0.2482)</td>
</tr>
<tr>
<td>100</td>
<td>0.0087</td>
<td>-0.0678</td>
<td>-0.0687</td>
</tr>
<tr>
<td>200</td>
<td>0.0097</td>
<td>-0.0154</td>
<td>-0.0161</td>
</tr>
</tbody>
</table>

Distribution (9) $\beta_1 = 1.0 \quad \beta_2 = 5.0$

<table>
<thead>
<tr>
<th>n</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>-1.2029</td>
<td>-0.1727</td>
<td>(-0.6510)</td>
</tr>
<tr>
<td>50</td>
<td>-0.0617</td>
<td>-0.0855</td>
<td>-0.0964</td>
</tr>
<tr>
<td>100</td>
<td>0.0195</td>
<td>-0.0328</td>
<td>-0.0330</td>
</tr>
<tr>
<td>200</td>
<td>0.0096</td>
<td>-0.0093</td>
<td>-0.0093</td>
</tr>
</tbody>
</table>

Distribution (10) $\beta_1 = 1.0 \quad \beta_2 = 5.8$

<table>
<thead>
<tr>
<th>n</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>-3.3002</td>
<td>-0.4625</td>
<td>(-1.3916)</td>
</tr>
<tr>
<td>50</td>
<td>-0.2595</td>
<td>-0.2421</td>
<td>-0.2208</td>
</tr>
<tr>
<td>100</td>
<td>0.0067</td>
<td>-0.0731</td>
<td>-0.0664</td>
</tr>
<tr>
<td>200</td>
<td>0.0110</td>
<td>-0.0173</td>
<td>-0.0165</td>
</tr>
</tbody>
</table>
Table 4.6

Estimates of $\mu_4(\sqrt{b_1})$ based on the following methods:

- Formulae (2.1), (2.2), (2.3) and (2.4).
- Simulation using the inverse distribution function;
  - 10000 samples for $n = 25, 50, 100$; 5000 samples for $n = 200$.
- Bowman and Shenton (1975);
  - not available; (below minimum $n$).

Distribution (1) $\beta_1 = 0$ $\beta_2 = 4.0$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.9754</td>
<td>0.5096</td>
<td>(1.5675)</td>
</tr>
<tr>
<td>50</td>
<td>0.2570</td>
<td>0.1825</td>
<td>0.1888</td>
</tr>
<tr>
<td>100</td>
<td>0.0659</td>
<td>0.0512</td>
<td>0.0566</td>
</tr>
<tr>
<td>200</td>
<td>0.0167</td>
<td>0.0156</td>
<td>0.0155</td>
</tr>
</tbody>
</table>

Distribution (2) $\beta_1 = 0$ $\beta_2 = 5.0$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>2.4345</td>
<td>1.0484</td>
<td>(-12.89)</td>
</tr>
<tr>
<td>50</td>
<td>0.6498</td>
<td>0.4821</td>
<td>0.4266</td>
</tr>
<tr>
<td>100</td>
<td>0.1676</td>
<td>0.1499</td>
<td>0.1474</td>
</tr>
<tr>
<td>200</td>
<td>0.0426</td>
<td>0.0411</td>
<td>0.0404</td>
</tr>
</tbody>
</table>
Table 4.6 (continued)

Distribution (3) \( \beta_1 = 0 \) \( \beta_2 = 6.5 \)

<table>
<thead>
<tr>
<th>n</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>5.8529</td>
<td>3.2133</td>
<td>(175.7)</td>
</tr>
<tr>
<td>50</td>
<td>1.5747</td>
<td>1.5336</td>
<td>(2.1801)</td>
</tr>
<tr>
<td>100</td>
<td>0.4077</td>
<td>0.4652</td>
<td>0.4466</td>
</tr>
<tr>
<td>200</td>
<td>0.1037</td>
<td>0.1206</td>
<td>0.1123</td>
</tr>
</tbody>
</table>

Distribution (4) \( \beta_1 = 0.3 \) \( \beta_2 = 4.0 \)

<table>
<thead>
<tr>
<th>n</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>1.1032</td>
<td>0.3363</td>
<td>(-0.3547)</td>
</tr>
<tr>
<td>50</td>
<td>0.2680</td>
<td>0.1353</td>
<td>0.1337</td>
</tr>
<tr>
<td>100</td>
<td>0.0648</td>
<td>0.0377</td>
<td>0.0385</td>
</tr>
<tr>
<td>200</td>
<td>0.0158</td>
<td>0.0104</td>
<td>0.0098</td>
</tr>
</tbody>
</table>

Distribution (5) \( \beta_1 = 0.3 \) \( \beta_2 = 5.0 \)

<table>
<thead>
<tr>
<th>n</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>3.1334</td>
<td>1.0654</td>
<td>-</td>
</tr>
<tr>
<td>50</td>
<td>0.7446</td>
<td>0.4196</td>
<td>-</td>
</tr>
<tr>
<td>100</td>
<td>0.1804</td>
<td>0.1260</td>
<td>-</td>
</tr>
<tr>
<td>200</td>
<td>0.0442</td>
<td>0.0321</td>
<td>-</td>
</tr>
</tbody>
</table>
### Table 4.6 (continued)

#### Distribution (6) \( \beta_1 = 0.3 \ \beta_2 = 6.5 \)

<table>
<thead>
<tr>
<th>n</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>8.7297</td>
<td>3.3152</td>
<td>(-40.38)</td>
</tr>
<tr>
<td>50</td>
<td>1.8985</td>
<td>1.3822</td>
<td>1.2092</td>
</tr>
<tr>
<td>100</td>
<td>0.4588</td>
<td>0.4077</td>
<td>0.4023</td>
</tr>
<tr>
<td>200</td>
<td>0.1130</td>
<td>0.1042</td>
<td>0.0995</td>
</tr>
</tbody>
</table>

#### Distribution (7) \( \beta_1 = 0.7 \ \beta_2 = 5.0 \)

<table>
<thead>
<tr>
<th>n</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>4.5439</td>
<td>0.8676</td>
<td>(9.8768)</td>
</tr>
<tr>
<td>50</td>
<td>0.9400</td>
<td>0.3547</td>
<td>0.3784</td>
</tr>
<tr>
<td>100</td>
<td>0.2067</td>
<td>0.0972</td>
<td>0.0947</td>
</tr>
<tr>
<td>200</td>
<td>0.0477</td>
<td>0.0217</td>
<td>0.0229</td>
</tr>
</tbody>
</table>

#### Distribution (8) \( \beta_1 = 0.7 \ \beta_2 = 6.0 \)

<table>
<thead>
<tr>
<th>n</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>11.13</td>
<td>2.1915</td>
<td>(-26.86)</td>
</tr>
<tr>
<td>50</td>
<td>2.0039</td>
<td>0.8688</td>
<td>0.6845</td>
</tr>
<tr>
<td>100</td>
<td>0.4302</td>
<td>0.2313</td>
<td>0.2386</td>
</tr>
<tr>
<td>200</td>
<td>0.0988</td>
<td>0.0563</td>
<td>0.0581</td>
</tr>
</tbody>
</table>
Table 4.6 (continued)

Distribution (9) $\beta_1 = 1.0$ $\beta_2 = 5.0$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>5.8385</td>
<td>0.7805</td>
<td>0.9775</td>
</tr>
<tr>
<td>50</td>
<td>1.1235</td>
<td>0.2714</td>
<td>0.2871</td>
</tr>
<tr>
<td>100</td>
<td>0.2315</td>
<td>0.0735</td>
<td>0.0744</td>
</tr>
<tr>
<td>200</td>
<td>0.0509</td>
<td>0.0175</td>
<td>0.0170</td>
</tr>
</tbody>
</table>

Distribution (10) $\beta_1 = 1.0$ $\beta_2 = 5.8$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Formulae</th>
<th>Simulation</th>
<th>Bowman-Shenton</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>13.16</td>
<td>1.8638</td>
<td>(-17.35)</td>
</tr>
<tr>
<td>50</td>
<td>2.2160</td>
<td>0.7439</td>
<td>0.5595</td>
</tr>
<tr>
<td>100</td>
<td>0.4431</td>
<td>0.1836</td>
<td>0.1761</td>
</tr>
<tr>
<td>200</td>
<td>0.0966</td>
<td>0.0389</td>
<td>0.0413</td>
</tr>
</tbody>
</table>

Table 4.5 shows that there is good agreement between the simulation and Bowman-Shenton estimates of $\mu_3(\sqrt{\beta_1})$. There is some disagreement (around 9%) at distribution (10) for $n = 50$ but without an "internal" estimate of standard error such discrepancies are difficult to interpret. This cannot be said for the formulae estimates. The method (at least to $O(n^{-2})$) is not useful, though strangely rather accurate for the most extreme non-normal distribution at $n = 50$. 
A similar picture is revealed in Table 4.6 for $\mu_4'(q_{b_1})$. Only for extreme non-normality and smaller sample sizes is there any major disagreement between the simulation and Bowman-Shenton estimates (up to 25% for distributions (9) and (10)), with some suggestion that the Bowman-Shenton results are slightly low for extreme kurtosis. (The possibility that we are witnessing a characteristic of the Gram-Charlier generator cannot, of course, be dismissed.) The formulae estimates are of little use, though do provide asymptotic results for slight non-normality.

A small selection of moment comparisons is given in Dodgson (1983a), while Dodgson (1983b) compares the Bowman-Shenton and simulation estimates and discusses Johnson approximations to the distribution of $v_{b_1}$ (described in the next chapter).

It is clear (in retrospect) that Fisher's k-statistic approach is only of limited use in deriving the sampling moments of moment statistics for non-normal populations. This contrasts with typical examples in the literature where normality is assumed; the simplification afforded by the vanishing cumulants of order 3 and above allow, for example, the derivation of a useful approximation to the fourth moment of $v_{b_1}$ (Kendall and Stuart, 1977). When dealing with the Gram-Charlier distribution (characterised by its vanishing cumulants of order 5 and above) we saw in Chapter 2 that it is apparently only possible, with the available published formulae, to derive $E(v_{b_1})$ and $V(v_{b_1})$ to $O(n^{-3})$, and $\mu_3'(v_{b_1})$ and $\mu_4'(v_{b_1})$ to $O(n^{-2})$. The limited success of these expressions is disappointing, particularly since one of the reasons for choosing the Gram-Charlier population was its
relative simplicity when obtaining sampling moments via k-statistics.

The Bowman-Shenton results are generally very satisfactory but are not easily portable and in published form only available for a fixed grid in $\beta_1-\beta_2$ for certain populations. The authors comment: "the course grid included can be justified since (1) the digital program involved is quite complicated and requires some two-three minutes per parameter point entry CPU time, and (2) it is unlikely that the results can be derived on anything but large computers" (Bowman and Shenton, 1975).

One alternative approach would be to computerise the k-statistics using an algebraic manipulation language. This might be expected to give more flexible and portable results in the present case as well as being of general interest. An example of such a language is MACSYMA (MIT, 1983) which can perform symbolic and numerical manipulations including the direct evaluation of integrals. However MACSYMA only became available at Trent Polytechnic while drafting this section and its value has yet to be explored.

Johnson approximations to the marginal distributions of $\sqrt{\beta_1}$ and $t$ based on their respective first four moments are obtained in the next chapter. Their correlation is also discussed so as to complete the representation of their joint distribution.
Chapter Five

The Marginal Distributions of $\sqrt{b_1}$ and t

and their Correlation

5.1 SMOOTHING THE ESTIMATED MOMENTS OF $\sqrt{b_1}$.

In Chapter 3 expansions were derived for the moments of $\sqrt{b_1}$ assuming a Gram-Charlier population. These were compared with estimates based on simulation and the expansions of Bowman and Shenton in Chapter 4. Noting that of the derived expansions only that for $E(\sqrt{b_1})$ appears of practical use we concentrate on the other two approaches and consider combining or smoothing the tables of estimates.

Since the moments of $\sqrt{b_1}$ vary smoothly for changing population skewness and kurtosis a simple regression approach is used, the four sample sizes (chosen simply to be representative) treated separately. However it is worth noting the suggestion of Lewis and Orav (1985) that the theoretical form of asymptotic expansions for the moments of a moment statistic provide starting points for two-stage simulation experiments regressing on sample size.

For each moment in turn the standard methods of variable selection (Draper and Smith, 1981, Chapter 6) were used with "explanatory variables" $\sqrt{\beta_1}$, $\beta_2$, $\beta_1$, $\beta_2^2$ and $\sqrt{\beta_1} \cdot \beta_2$. Originally rounded "agreed" values from the simulation and Bowman-Shenton results were used as independent variable (the latter not always being available) but it was found more satisfactory, and less arbitrary, to use both sets of results considered as "replicates" as necessary (Shenton and Bowman's point $\beta_1 = 0.3, \beta_2 = 4.5$ included).
The calculations were performed on the statistical package MINITAB (Ryan, Joiner and Ryan, 1985) taking guidance from the STEPWISE facility, though using personal judgement in selecting variables for reasonable consistency over n and the amount of acceptable residual variation. For the third and fourth moments the use of "centred" values of $\sqrt{\beta_1}$ and $\beta_2$ (i.e. with their mean subtracted) was found helpful (Marquardt, 1980).

The final results, in the form of a data file convenient for later use, are given in Table 5.1. The details of the smoothing may be found in Table 5.2. It will be apparent that most uncertainty is with $\mu_4(\sqrt{b_1})$ (and therefore $\beta_2(\sqrt{b_1})$) for n = 25 and 50.

It will be noted that Table 5.1 also includes values for the correlation of $t$ and $\sqrt{b_1}$. These will be discussed in Section 5.4.
Table 5.1
Smoothed estimates of moments of $\nu_1$ and correlation with $t$.
(a) record no., (b) distribution no., (c) $\beta_1$, (d) $\beta_2$, (e) $n$,
(f) $E(\nu_1)$, (g) $V(\nu_1)$, (h) $\hat{\nu}_1(\nu_1)$, (i) $\beta_2(\nu_1)$, (j) $\rho(t,\nu_1)$.

<table>
<thead>
<tr>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
<th>(f)</th>
<th>(g)</th>
<th>(h)</th>
<th>(i)</th>
<th>(j)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>4.0</td>
<td>25</td>
<td>0.000</td>
<td>0.375</td>
<td>0.000</td>
<td>3.42</td>
<td>0.236</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
<td>4.0</td>
<td>50</td>
<td>0.000</td>
<td>0.238</td>
<td>0.000</td>
<td>3.00</td>
<td>0.240</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>4.0</td>
<td>100</td>
<td>0.000</td>
<td>0.134</td>
<td>0.000</td>
<td>3.17</td>
<td>0.262</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.0</td>
<td>4.0</td>
<td>200</td>
<td>0.000</td>
<td>0.071</td>
<td>0.000</td>
<td>3.37</td>
<td>0.227</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.0</td>
<td>5.0</td>
<td>25</td>
<td>0.000</td>
<td>0.616</td>
<td>0.000</td>
<td>2.80</td>
<td>0.390</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.0</td>
<td>5.0</td>
<td>50</td>
<td>0.000</td>
<td>0.411</td>
<td>0.000</td>
<td>3.04</td>
<td>0.413</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.0</td>
<td>5.0</td>
<td>100</td>
<td>0.000</td>
<td>0.144</td>
<td>0.000</td>
<td>3.15</td>
<td>0.408</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.0</td>
<td>6.5</td>
<td>25</td>
<td>0.000</td>
<td>1.116</td>
<td>0.000</td>
<td>2.56</td>
<td>0.545</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.0</td>
<td>6.5</td>
<td>50</td>
<td>0.000</td>
<td>0.741</td>
<td>0.000</td>
<td>2.78</td>
<td>0.580</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.0</td>
<td>6.5</td>
<td>100</td>
<td>0.000</td>
<td>0.391</td>
<td>0.000</td>
<td>3.04</td>
<td>0.582</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.0</td>
<td>6.5</td>
<td>200</td>
<td>0.000</td>
<td>0.197</td>
<td>0.000</td>
<td>3.07</td>
<td>0.575</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.0</td>
<td>6.5</td>
<td>100</td>
<td>0.000</td>
<td>0.391</td>
<td>0.000</td>
<td>3.04</td>
<td>0.582</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0.0</td>
<td>6.5</td>
<td>200</td>
<td>0.000</td>
<td>0.197</td>
<td>0.000</td>
<td>3.07</td>
<td>0.575</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0.3</td>
<td>4.0</td>
<td>25</td>
<td>0.430</td>
<td>0.315</td>
<td>-0.136</td>
<td>3.89</td>
<td>0.203</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.3</td>
<td>4.0</td>
<td>50</td>
<td>0.491</td>
<td>0.187</td>
<td>-0.235</td>
<td>4.69</td>
<td>0.178</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.3</td>
<td>4.0</td>
<td>100</td>
<td>0.517</td>
<td>0.106</td>
<td>-0.261</td>
<td>3.74</td>
<td>0.196</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>0.3</td>
<td>4.0</td>
<td>200</td>
<td>0.534</td>
<td>0.057</td>
<td>-0.279</td>
<td>3.39</td>
<td>0.155</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>0.3</td>
<td>5.0</td>
<td>25</td>
<td>0.445</td>
<td>0.555</td>
<td>-0.385</td>
<td>3.14</td>
<td>0.357</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>0.3</td>
<td>5.0</td>
<td>50</td>
<td>0.502</td>
<td>0.359</td>
<td>-0.451</td>
<td>3.00</td>
<td>0.349</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.3</td>
<td>5.0</td>
<td>100</td>
<td>0.527</td>
<td>0.194</td>
<td>-0.363</td>
<td>2.98</td>
<td>0.347</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>0.3</td>
<td>5.0</td>
<td>200</td>
<td>0.538</td>
<td>0.100</td>
<td>-0.218</td>
<td>2.80</td>
<td>0.335</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>0.3</td>
<td>6.5</td>
<td>25</td>
<td>0.468</td>
<td>1.076</td>
<td>-0.459</td>
<td>2.77</td>
<td>0.512</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>0.3</td>
<td>6.5</td>
<td>50</td>
<td>0.520</td>
<td>0.690</td>
<td>-0.373</td>
<td>2.70</td>
<td>0.518</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>0.3</td>
<td>6.5</td>
<td>100</td>
<td>0.543</td>
<td>0.363</td>
<td>-0.293</td>
<td>2.96</td>
<td>0.516</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.3</td>
<td>6.5</td>
<td>200</td>
<td>0.545</td>
<td>0.183</td>
<td>-0.147</td>
<td>2.90</td>
<td>0.503</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>0.3</td>
<td>6.5</td>
<td>50</td>
<td>0.476</td>
<td>0.475</td>
<td>-0.522</td>
<td>4.06</td>
<td>0.313</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>0.3</td>
<td>6.5</td>
<td>100</td>
<td>0.770</td>
<td>0.291</td>
<td>-0.624</td>
<td>3.94</td>
<td>0.266</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>0.3</td>
<td>7.0</td>
<td>25</td>
<td>0.806</td>
<td>0.158</td>
<td>-0.541</td>
<td>3.69</td>
<td>0.258</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>0.3</td>
<td>7.0</td>
<td>50</td>
<td>0.824</td>
<td>0.082</td>
<td>-0.375</td>
<td>3.27</td>
<td>0.239</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.3</td>
<td>7.0</td>
<td>100</td>
<td>0.699</td>
<td>0.801</td>
<td>-0.685</td>
<td>3.47</td>
<td>0.426</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>0.3</td>
<td>7.0</td>
<td>200</td>
<td>0.831</td>
<td>0.135</td>
<td>-0.315</td>
<td>3.18</td>
<td>0.364</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.3</td>
<td>7.0</td>
<td>50</td>
<td>0.788</td>
<td>0.502</td>
<td>-0.644</td>
<td>3.21</td>
<td>0.390</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>0.3</td>
<td>7.0</td>
<td>100</td>
<td>0.822</td>
<td>0.266</td>
<td>-0.503</td>
<td>3.42</td>
<td>0.379</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>0.3</td>
<td>7.0</td>
<td>200</td>
<td>0.831</td>
<td>0.135</td>
<td>-0.315</td>
<td>3.18</td>
<td>0.364</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>0.8</td>
<td>5.0</td>
<td>25</td>
<td>0.807</td>
<td>0.414</td>
<td>-0.668</td>
<td>5.19</td>
<td>0.280</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>0.8</td>
<td>5.0</td>
<td>50</td>
<td>0.922</td>
<td>0.240</td>
<td>-0.842</td>
<td>5.30</td>
<td>0.204</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>0.8</td>
<td>5.0</td>
<td>100</td>
<td>0.963</td>
<td>0.131</td>
<td>-0.738</td>
<td>4.66</td>
<td>0.192</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>0.8</td>
<td>5.0</td>
<td>200</td>
<td>0.986</td>
<td>0.068</td>
<td>-0.553</td>
<td>3.89</td>
<td>0.166</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>0.8</td>
<td>5.0</td>
<td>50</td>
<td>0.938</td>
<td>0.405</td>
<td>-0.881</td>
<td>3.90</td>
<td>0.307</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.8</td>
<td>5.0</td>
<td>100</td>
<td>0.978</td>
<td>0.215</td>
<td>-0.692</td>
<td>4.05</td>
<td>0.291</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>0.8</td>
<td>5.0</td>
<td>200</td>
<td>0.992</td>
<td>0.109</td>
<td>-0.472</td>
<td>3.70</td>
<td>0.271</td>
<td></td>
</tr>
</tbody>
</table>
Table 5.2
Selected variables ("centred" for μ₃ and μ₄) and residual root mean square for smoothing the moments of \( \sqrt{b_1} \).

<table>
<thead>
<tr>
<th>Selected variables</th>
<th>Residual rms</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E(\sqrt{b_1}) )</td>
<td></td>
</tr>
<tr>
<td>( n = 25 )</td>
<td>( \sqrt{\beta_1}, \sqrt{\beta_1 \beta_2} )</td>
</tr>
<tr>
<td>50</td>
<td>( \sqrt{\beta_1}, \sqrt{\beta_1 \beta_2} )</td>
</tr>
<tr>
<td>100</td>
<td>( \sqrt{\beta_1}, \sqrt{\beta_1 \beta_2} )</td>
</tr>
<tr>
<td>200</td>
<td>( \sqrt{\beta_1}, \sqrt{\beta_1 \beta_2} )</td>
</tr>
<tr>
<td>( V(\sqrt{b_1}) )</td>
<td></td>
</tr>
<tr>
<td>( n = 25 )</td>
<td>( \beta_2, \beta_1, \beta_2^2 )</td>
</tr>
<tr>
<td>50</td>
<td>( \beta_1, \beta_2^2 )</td>
</tr>
<tr>
<td>100</td>
<td>( \beta_1, \beta_2^2 )</td>
</tr>
<tr>
<td>200</td>
<td>( \beta_1, \beta_2^2 )</td>
</tr>
<tr>
<td>( \mu_3(\sqrt{b_1}) )</td>
<td></td>
</tr>
<tr>
<td>( n = 25 )</td>
<td>( \sqrt{\beta_1}, \beta_2, \beta_2^2, \sqrt{\beta_1 \beta_2} )</td>
</tr>
<tr>
<td>50</td>
<td>( \sqrt{\beta_1}, \beta_2, \sqrt{\beta_1 \beta_2} )</td>
</tr>
<tr>
<td>100</td>
<td>( \sqrt{\beta_1}, \beta_2, \sqrt{\beta_1 \beta_2} )</td>
</tr>
<tr>
<td>200</td>
<td>( \sqrt{\beta_1}, \beta_2, \sqrt{\beta_1 \beta_2} )</td>
</tr>
<tr>
<td>( \mu_4(\sqrt{b_1}) )</td>
<td></td>
</tr>
<tr>
<td>( n = 25 )</td>
<td>( \sqrt{\beta_1}, \beta_2, \beta_2^2 )</td>
</tr>
<tr>
<td>50</td>
<td>( \sqrt{\beta_1}, \beta_2, \beta_2^2, \sqrt{\beta_1 \beta_2} )</td>
</tr>
<tr>
<td>100</td>
<td>( \sqrt{\beta_1}, \beta_2, \beta_2^2, \sqrt{\beta_1 \beta_2} )</td>
</tr>
<tr>
<td>200</td>
<td>( \sqrt{\beta_1}, \beta_2, \beta_2^2, \sqrt{\beta_1 \beta_2} )</td>
</tr>
</tbody>
</table>
5.2 FITTING A MARGINAL DISTRIBUTION TO $\sqrt{\beta_1}$.

In Section 1.4 the Johnson system was briefly described and notice given of its use to approximate the marginal and joint distributions of $t$ and $\sqrt{\beta_1}$. Here we consider fitting the system to $\sqrt{\beta_1}$.

For a normal population Johnson's $Su$ was first fitted to the distribution of $\sqrt{\beta_1}$ by Pearson (1963), using the values of the moments provided by Fisher (1930). D'Agostino and Pearson (1973) give a more thorough treatment which is cited by Biometrika tables (Pearson and Hartley, 1976) in giving corrected percentage points using this method. The non-normal case is investigated by Shenton and Bowman (1975) who use their moment expressions to fit $Su$ and Cornish-Fisher expansions with an $Su$ kernel to both $\sqrt{\beta_1}$ and $\beta_2$ in sampling from a normal mixture. For $\sqrt{\beta_1}$ they conclude that the $Su$ fit is "amazingly good".

The algorithm AS 99 (Hill, Hill and Holder, 1976; corrected in Hill and Wheeler, 1981) is used to fit the Johnson distribution. The related AS 100 (Hill, 1976; corrected in Hill and Wheeler, 1981 and Dodgson and Hill, 1983) allows the transformation of a standard normal variate to a Johnson variate, and vice versa, thereby giving Johnson quantiles and tail areas from the corresponding normal values. These in turn are obtained from AS 111 (Beasley and Springer, 1977) and AS 66 (Hill, 1973). Recently Griffiths and Hill (1985) have collected some of the more important Applied Statistics algorithms, and include those used in this work.

For the normal case, and using the moments provided by Fisher, it is found that the percentage points of $\sqrt{\beta_1}$ (i.e. 5%, 2.5%, 1%
and 0.5%) agree with those in Biometrika tables.

The Johnson fit for the non-normal case based on the smoothed moments of the previous section is investigated as follows. For certain of the Gram-Charlier populations and sample sizes listed previously the Johnson distribution is fitted and the quantiles $X_p$ obtained for $P = 0.05, 0.10, 0.30, 0.50, 0.70, 0.90$ and 0.95. A sampling experiment (using the inverse distribution function in GCPACK) is performed to obtain the empirical probabilities of falling below these quantiles, and the results for three of the more non-normal distributions displayed in Table 5.3. (Empirical probabilities of exceeding the nominal 5% and 1% points are also available.)

Notice first that the fitted Johnson is not necessarily $S_U$, as in the normal case, but varies depending on the values of the third and fourth central moments of $\nu_1$. However the supported ranges of the $S_L$ and $S_B$ systems will only be a factor if they prevent the satisfactory representation of $\nu_1$ over the observed range. The instance of the substitution of $S_L$ when $S_B$ failed to converge (near the normal point and just above the $S_L$ line in Figure 1.3) is a feature of the algorithm and this fit, like the others, must be judged on its merits. The standard errors corresponding to the varying values of $P$ are approximate since the quantiles $X_p$ are derived from the $S_I$ fit; the maximum possible standard errors, of course, are the values for $P = 0.5$.

The results suggest that the Johnson system provides a satisfactory approximation to the non-null distribution of $\nu_1$, generally improving with increasing $n$ and proximity to the normal
Table 5.3

Marginal distribution of $\hat{\nu}_1$.

Quantiles $X_p$ from Johnson $S_1$ fit (smoothed moments).

Simulation estimates $\hat{P}$ of corresponding $P$;

10000 samples for $n = 25, 50, 100; 5000$ for $n = 200$.

Distribution (6) $\beta_1 = 0.3 \quad \beta_2 = 6.5$

<table>
<thead>
<tr>
<th>$P$</th>
<th>$n = 25$</th>
<th>$\hat{P}$</th>
<th>$n = 50$</th>
<th>$\hat{P}$</th>
<th>$n = 100$</th>
<th>$\hat{P}$</th>
<th>$n = 200$</th>
<th>$\hat{P}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>-1.4104</td>
<td>0.0559</td>
<td>-0.9589</td>
<td>0.0534</td>
<td>-0.5020</td>
<td>0.0474</td>
<td>-0.1760</td>
<td>0.0474</td>
</tr>
<tr>
<td>0.10</td>
<td>-0.9705</td>
<td>0.1017</td>
<td>-0.6213</td>
<td>0.0974</td>
<td>-0.2555</td>
<td>0.0968</td>
<td>-0.0093</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.30</td>
<td>-0.0361</td>
<td>0.2972</td>
<td>0.1041</td>
<td>0.2927</td>
<td>0.2451</td>
<td>0.2990</td>
<td>0.3286</td>
<td>0.2962</td>
</tr>
<tr>
<td>0.50</td>
<td>0.5737</td>
<td>0.4902</td>
<td>0.5879</td>
<td>0.5000</td>
<td>0.5750</td>
<td>0.5030</td>
<td>0.5555</td>
<td>0.5022</td>
</tr>
<tr>
<td>0.70</td>
<td>1.1156</td>
<td>0.6878</td>
<td>1.0286</td>
<td>0.7002</td>
<td>0.8866</td>
<td>0.7038</td>
<td>0.7765</td>
<td>0.6954</td>
</tr>
<tr>
<td>0.90</td>
<td>1.7467</td>
<td>0.9127</td>
<td>1.5595</td>
<td>0.9060</td>
<td>1.2981</td>
<td>0.8980</td>
<td>1.0859</td>
<td>0.8958</td>
</tr>
<tr>
<td>0.95</td>
<td>1.9843</td>
<td>0.9542</td>
<td>1.7661</td>
<td>0.9552</td>
<td>1.4781</td>
<td>0.9510</td>
<td>1.2303</td>
<td>0.9486</td>
</tr>
</tbody>
</table>

* $S_L$ fitted as $S_B$ failed to converge.
Table 5.3 (continued)

Distribution (8) $\beta_1 = 0.7$  $\beta_2 = 6.0$

<table>
<thead>
<tr>
<th>P</th>
<th>$n = 25$</th>
<th>$n = 50$</th>
<th>$n = 100$</th>
<th>$n = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$X_p(S_B)$</td>
<td>$\hat{p}$</td>
<td>$X_p(S_B)$</td>
<td>$\hat{p}$</td>
</tr>
<tr>
<td>0.05</td>
<td>-0.9490</td>
<td>0.0465</td>
<td>-0.5233</td>
<td>0.0460</td>
</tr>
<tr>
<td>0.10</td>
<td>-0.5154</td>
<td>0.0930</td>
<td>-0.1882</td>
<td>0.0942</td>
</tr>
<tr>
<td>0.30</td>
<td>0.3122</td>
<td>0.3051</td>
<td>0.4738</td>
<td>0.2932</td>
</tr>
<tr>
<td>0.50</td>
<td>0.8117</td>
<td>0.4797</td>
<td>0.8797</td>
<td>0.4985</td>
</tr>
<tr>
<td>0.70</td>
<td>1.2453</td>
<td>0.7122</td>
<td>1.2297</td>
<td>0.7183</td>
</tr>
<tr>
<td>0.90</td>
<td>1.7580</td>
<td>0.9160</td>
<td>1.6317</td>
<td>0.9208</td>
</tr>
<tr>
<td>0.95</td>
<td>1.9596</td>
<td>0.9562</td>
<td>1.7838</td>
<td>0.9576</td>
</tr>
</tbody>
</table>
Table 5.3 (continued)

Distribution (10) $\beta_1 = 1.0 \quad \beta_2 = 5.8$

<table>
<thead>
<tr>
<th>$P$</th>
<th>$n = 25$</th>
<th></th>
<th>$n = 50$</th>
<th></th>
<th>$n = 100$</th>
<th></th>
<th>$n = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$X_P(S_B)$</td>
<td>$\hat{P}$</td>
<td>$X_P(S_B)$</td>
<td>$\hat{P}$</td>
<td>$X_P(S_U)$</td>
<td>$\hat{P}$</td>
<td>$X_P(S_U)$</td>
</tr>
<tr>
<td>0.05</td>
<td>-0.6812</td>
<td>0.0481</td>
<td>-0.2633</td>
<td>0.0514</td>
<td>0.1433</td>
<td>0.0509</td>
<td>0.4139</td>
</tr>
<tr>
<td>0.10</td>
<td>-0.2588</td>
<td>0.1012</td>
<td>0.0699</td>
<td>0.1018</td>
<td>0.3715</td>
<td>0.1050</td>
<td>0.5647</td>
</tr>
<tr>
<td>0.30</td>
<td>0.5026</td>
<td>0.2950</td>
<td>0.6857</td>
<td>0.2845</td>
<td>0.7840</td>
<td>0.3100</td>
<td>0.8441</td>
</tr>
<tr>
<td>0.50</td>
<td>0.9434</td>
<td>0.4832</td>
<td>1.0403</td>
<td>0.4946</td>
<td>1.0282</td>
<td>0.4906</td>
<td>1.0160</td>
</tr>
<tr>
<td>0.70</td>
<td>1.3210</td>
<td>0.7245</td>
<td>1.3352</td>
<td>0.7268</td>
<td>1.2454</td>
<td>0.6933</td>
<td>1.1749</td>
</tr>
<tr>
<td>0.90</td>
<td>1.7691</td>
<td>0.9174</td>
<td>1.6643</td>
<td>0.9129</td>
<td>1.5215</td>
<td>0.9047</td>
<td>1.3892</td>
</tr>
<tr>
<td>0.95</td>
<td>1.9486</td>
<td>0.9528</td>
<td>1.7871</td>
<td>0.9497</td>
<td>1.6420</td>
<td>0.9514</td>
<td>1.4884</td>
</tr>
</tbody>
</table>
point in the $\beta_1 - \beta_2$ plane. For moderate sample sizes and extreme non-normality the empirical distribution exhibits an excess between the 50th and 70th percentiles (as given by the Johnson fit) and a lighter upper tail.

One might suggest the use of the Pearson statistic to check the Johnson fit. However the fact, as shown in Bishop, Fienberg and Holland (1975), that $X^2$ increases with sample size (in this context number of samples) whenever the fitted model is incorrect indicates that it cannot be interpreted in the traditional sense when used to assess an approximation in a simulation experiment. Following the above authors, dividing $X^2$ by the number of samples gives a comparative measure whose values confirm the comments in the previous paragraph.

The power of $\sqrt{b_1}$ in detecting non-normality is clearly of interest in our investigation. Table 5.4 present the power of one-sided tests at the nominal 5% and 1% levels (percentage points from Biometrika tables) obtained from the Johnson approximations using the smoothed moments.

Table 5.4
Power of $\sqrt{b_1}$ in tests of normality from Johnson $S_1$ fit (smoothed moments).

<table>
<thead>
<tr>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>n</th>
<th>5%</th>
<th>1%</th>
<th>Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4.0</td>
<td>25</td>
<td>0.1170</td>
<td>0.0417</td>
<td>$S_U$</td>
</tr>
<tr>
<td>(1)</td>
<td></td>
<td>50</td>
<td>0.1368</td>
<td>0.0531</td>
<td>$S_N$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.1404</td>
<td>0.0599</td>
<td>$S_U$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.1405</td>
<td>0.0633</td>
<td>$S_U$</td>
</tr>
</tbody>
</table>
### Table 5.4 (continued)

<table>
<thead>
<tr>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>n</th>
<th>5%</th>
<th>1%</th>
<th>Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5.0</td>
<td>25</td>
<td>0.1873</td>
<td>0.0909</td>
<td>SB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>0.2024</td>
<td>0.1095</td>
<td>$S_N^*$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.2039</td>
<td>0.1144</td>
<td>$S_N$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.2003</td>
<td>0.1142</td>
<td>$S_U$</td>
</tr>
<tr>
<td>0</td>
<td>6.5</td>
<td>25</td>
<td>0.2649</td>
<td>0.1715</td>
<td>$S_B$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>0.2730</td>
<td>0.1853</td>
<td>$S_B$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.2664</td>
<td>0.1823</td>
<td>$S_N$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.2641</td>
<td>0.1819</td>
<td>$S_N$</td>
</tr>
<tr>
<td>0.3</td>
<td>4.0</td>
<td>25</td>
<td>0.2992</td>
<td>0.1186</td>
<td>$S_U$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>0.4672</td>
<td>0.2253</td>
<td>$S_U$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.6729</td>
<td>0.4502</td>
<td>$S_U$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.8601</td>
<td>0.7242</td>
<td>$S_U$</td>
</tr>
<tr>
<td>0.3</td>
<td>5.0</td>
<td>25</td>
<td>0.3808</td>
<td>0.2087</td>
<td>$S_L^*$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>0.5131</td>
<td>0.3465</td>
<td>$S_B$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.6414</td>
<td>0.4909</td>
<td>$S_B$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.7901</td>
<td>0.6719</td>
<td>$S_B$</td>
</tr>
<tr>
<td>0.3</td>
<td>6.5</td>
<td>25</td>
<td>0.4499</td>
<td>0.3208</td>
<td>$S_B$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>0.5241</td>
<td>0.4089</td>
<td>$S_B$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.6169</td>
<td>0.5052</td>
<td>$S_B$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.7373</td>
<td>0.6383</td>
<td>$S_L^*$</td>
</tr>
</tbody>
</table>
Table 5.4 (continued)

<table>
<thead>
<tr>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>n</th>
<th>5%</th>
<th>1%</th>
<th>Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>5.0</td>
<td>25</td>
<td>0.5102</td>
<td>0.2951</td>
<td>$S_U$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>0.7033</td>
<td>0.5265</td>
<td>$S_U$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.8568</td>
<td>0.7497</td>
<td>$S_U$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.9620</td>
<td>0.9218</td>
<td>$S_L$</td>
</tr>
<tr>
<td>0.7</td>
<td>6.0</td>
<td>25</td>
<td>0.5444</td>
<td>0.3863</td>
<td>$S_B$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>0.6739</td>
<td>0.5496</td>
<td>$S_B$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.8084</td>
<td>0.7126</td>
<td>$S_L$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.9263</td>
<td>0.8758</td>
<td>$S_L$</td>
</tr>
<tr>
<td>0.7</td>
<td>5.0</td>
<td>25</td>
<td>0.6016</td>
<td>0.3563</td>
<td>$S_U$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>0.8162</td>
<td>0.6592</td>
<td>$S_U$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.9344</td>
<td>0.8723</td>
<td>$S_U$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.9888</td>
<td>0.9760</td>
<td>$S_U$</td>
</tr>
<tr>
<td>0.7</td>
<td>5.8</td>
<td>25</td>
<td>0.6127</td>
<td>0.4396</td>
<td>$S_B$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>0.7658</td>
<td>0.6488</td>
<td>$S_B$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.8945</td>
<td>0.8266</td>
<td>$S_U$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.9741</td>
<td>0.9526</td>
<td>$S_U$</td>
</tr>
</tbody>
</table>

* $S_N$ or $S_L$ fitted as $S_B$ failed to converge.
Examining Table 5.4 we notice first that for symmetric alternatives the power of \( \sqrt{b_1} \) is approximately constant for increasing \( n \). The values in Table 5.1 suggest an explanation: the first and third moments of \( \sqrt{b_1} \) are of course zero, \( \text{VAR}(\sqrt{b_1}) \) decreases like \( n^{-1} \) while \( \beta_2(\sqrt{b_1}) \) increases with \( n \) - the decreasing variability appears to be "balanced" by the increasing tails. Generally, as expected, power increases with \( n \) and population skewness. The sample size needed for power to exceed 50% is approximately 50 for \( \beta_1 = 0.3 \) and 25 for \( \beta_1 = 0.7 \). A further interesting feature is the decrease in power with increasing population kurtosis associated with high population skewness and large \( n \). This cannot be attributable to the fitting process as the same phenomenon occurs with simulation estimates of power; examination of Table 5.1 this time reveals no simple explanation.

One may finally note the simplicity of using the Johnson system with the associated Applied Statistics algorithms to represent the sampling distribution of a statistic whenever the first four moments (exact or approximate) are available. In particular, power calculations to a practical accuracy are quite straightforward.

5.3 FITTING A MARGINAL DISTRIBUTION TO \( t \).

It was shown in Chapter 2 that the moments of the \( t \) statistic, expressed in terms of population cumulants, are given by Geary (1947). Using the algorithms discussed in the previous section it is therefore a simple matter to fit the Johnson system. To assess the fit an identical sampling experiment to that described in the previous section was carried out and the results given in Table 5.5.
Table 5.5
Marginal distribution of t.

Quantiles $X_p$ from Johnson $S_I$ fit (moments from Geary, 1947).

Simulation estimates $\hat{P}$ of corresponding $P$;
10000 samples for $n = 25, 50, 100$; 5000 for $n = 200$.

Distribution (6) $\beta_1 = 0.3$ $\beta_2 = 6.5$

<table>
<thead>
<tr>
<th></th>
<th>$n = 25$</th>
<th></th>
<th>$n = 50$</th>
<th></th>
<th>$n = 100$</th>
<th></th>
<th>$n = 200$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$X_p(S_L)$</td>
<td>$\hat{P}$</td>
<td>$X_p(S_L)$</td>
<td>$\hat{P}$</td>
<td>$X_p(S_L)$</td>
<td>$\hat{P}$</td>
<td>$X_p(S_N)$</td>
<td>$\hat{P}$</td>
</tr>
<tr>
<td>0.05</td>
<td>-1.8884</td>
<td>0.0497</td>
<td>-1.7839</td>
<td>0.0495</td>
<td>-1.7291</td>
<td>0.0500</td>
<td>-1.6755</td>
<td>0.0532</td>
</tr>
<tr>
<td>0.10</td>
<td>-1.4514</td>
<td>0.0514</td>
<td>-1.3779</td>
<td>0.0977</td>
<td>-1.3396</td>
<td>0.1024</td>
<td>-1.3099</td>
<td>0.1088</td>
</tr>
<tr>
<td>0.30</td>
<td>-0.5866</td>
<td>0.3062</td>
<td>-0.5595</td>
<td>0.2978</td>
<td>-0.5456</td>
<td>0.3053</td>
<td>-0.5478</td>
<td>0.3044</td>
</tr>
<tr>
<td>0.50</td>
<td>-0.0221</td>
<td>0.5011</td>
<td>-0.0139</td>
<td>0.4978</td>
<td>-0.0094</td>
<td>0.5020</td>
<td>-0.0201</td>
<td>0.4888</td>
</tr>
<tr>
<td>0.70</td>
<td>0.5156</td>
<td>0.7025</td>
<td>0.5150</td>
<td>0.6947</td>
<td>0.5158</td>
<td>0.7042</td>
<td>0.5077</td>
<td>0.6904</td>
</tr>
<tr>
<td>0.90</td>
<td>1.2474</td>
<td>0.8948</td>
<td>1.2499</td>
<td>0.9041</td>
<td>1.2551</td>
<td>0.9059</td>
<td>1.2697</td>
<td>0.9092</td>
</tr>
<tr>
<td>0.95</td>
<td>1.5808</td>
<td>0.9458</td>
<td>1.5910</td>
<td>0.9511</td>
<td>1.6021</td>
<td>0.9513</td>
<td>1.6353</td>
<td>0.9558</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$n = 25, 50, 100$</th>
<th></th>
<th>$n = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.0022</td>
<td></td>
<td>0.0031</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0030</td>
<td></td>
<td>0.0042</td>
</tr>
<tr>
<td>0.30</td>
<td>0.0046</td>
<td></td>
<td>0.0065</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0050</td>
<td></td>
<td>0.0071</td>
</tr>
</tbody>
</table>
### Table 5.5 (continued)

Distribution (8) \( \beta_1 = 0.7 \) \( \beta_2 = 6.0 \)

<table>
<thead>
<tr>
<th>( p )</th>
<th>( n = 25 )</th>
<th>( n = 50 )</th>
<th>( n = 100 )</th>
<th>( n = 200 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( X_p(S_U) )</td>
<td>( \hat{p} )</td>
<td>( X_p(S_L) )</td>
<td>( \hat{p} )</td>
</tr>
<tr>
<td>0.05</td>
<td>-1.9882</td>
<td>0.0518</td>
<td>-1.8461</td>
<td>0.0493</td>
</tr>
<tr>
<td>0.10</td>
<td>-1.5082</td>
<td>0.1093</td>
<td>-1.4193</td>
<td>0.0978</td>
</tr>
<tr>
<td>0.30</td>
<td>-0.6011</td>
<td>0.3123</td>
<td>-0.5733</td>
<td>0.2985</td>
</tr>
<tr>
<td>0.50</td>
<td>-0.0337</td>
<td>0.5038</td>
<td>-0.0203</td>
<td>0.4981</td>
</tr>
<tr>
<td>0.70</td>
<td>0.4951</td>
<td>0.7052</td>
<td>0.5073</td>
<td>0.7090</td>
</tr>
<tr>
<td>0.90</td>
<td>1.2081</td>
<td>0.9023</td>
<td>1.2264</td>
<td>0.8988</td>
</tr>
<tr>
<td>0.95</td>
<td>1.5355</td>
<td>0.9519</td>
<td>1.5544</td>
<td>0.9490</td>
</tr>
</tbody>
</table>
### Table 5.5 (continued)

Distribution (10) $\beta_1 = 1.0$ $\beta_2 = 5.8$

<table>
<thead>
<tr>
<th>$P$</th>
<th>$n = 25$</th>
<th>$\hat{p}$</th>
<th>$n = 50$</th>
<th>$\hat{p}$</th>
<th>$n = 100$</th>
<th>$\hat{p}$</th>
<th>$n = 200$</th>
<th>$\hat{p}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>-2.0446</td>
<td>0.0516</td>
<td>-1.8797</td>
<td>0.0494</td>
<td>-1.7907</td>
<td>0.0481</td>
<td>-1.7383</td>
<td>0.0530</td>
</tr>
<tr>
<td>0.10</td>
<td>-1.5418</td>
<td>0.1016</td>
<td>-1.4336</td>
<td>0.0996</td>
<td>-1.3806</td>
<td>0.0966</td>
<td>-1.3448</td>
<td>0.1004</td>
</tr>
<tr>
<td>0.30</td>
<td>-0.6122</td>
<td>0.2942</td>
<td>-0.5734</td>
<td>0.3047</td>
<td>-0.5594</td>
<td>0.2971</td>
<td>-0.5470</td>
<td>0.3022</td>
</tr>
<tr>
<td>0.50</td>
<td>-0.0421</td>
<td>0.4943</td>
<td>-0.0238</td>
<td>0.5044</td>
<td>-0.0164</td>
<td>0.4978</td>
<td>-0.0116</td>
<td>0.4982</td>
</tr>
<tr>
<td>0.70</td>
<td>0.4839</td>
<td>0.6950</td>
<td>0.4956</td>
<td>0.7100</td>
<td>0.5065</td>
<td>0.6997</td>
<td>0.5101</td>
<td>0.6946</td>
</tr>
<tr>
<td>0.90</td>
<td>1.1913</td>
<td>0.9003</td>
<td>1.2049</td>
<td>0.9088</td>
<td>1.2276</td>
<td>0.8977</td>
<td>1.2402</td>
<td>0.9014</td>
</tr>
<tr>
<td>0.95</td>
<td>1.5183</td>
<td>0.9506</td>
<td>1.5328</td>
<td>0.9537</td>
<td>1.5599</td>
<td>0.9481</td>
<td>1.5810</td>
<td>0.9528</td>
</tr>
</tbody>
</table>
The evidence from Table 5.5 is that the Johnson system provides a very good fit to the non-normal distribution of the t statistic. The fit is certainly better than that for $\nu b_1$, presumably because of the accuracy of the moments, but possibly also because of the nature of the true distributions of t and $\nu b_1$.

Now it was pointed out in Chapter 1 and illustrated with contour diagrams in Chapter 2 that the effect of population non-normality on the t statistic is essentially that of population skewness on the first and third sampling moments. Using the procedure outlined above we are in a position to investigate the effect of population kurtosis on the percentage points of t. A convenient way of illustrating this is with a modified form of box plot. Originally conceived to illustrate the range, interquantile range and median (Spear, 1952) it has more recently become the ubiquitous tool of exploratory data analysis (Tukey, 1977). We use a thick line to represent the region between the 25% and 75% points broken at the median, with thin lines extending to the 5% and 95% points (acknowledging Tufte, 1983).

In Figure 5.1 we illustrate the distribution of the t statistic for five populations: normal (i.e. from tabulated percentage points) and Gram-Charlier with $\beta_1 = 0, 0.3, 0.7$ and 1 while holding $\beta_2 = 5$ (distributions (2), (5), (7) and (9) in Figure 1.1). Sample sizes are $n = 5, 10, 25$ and 50, the smaller sample sizes introduced to illustrate the general effect more clearly.
Figure 5.1
The effect of population skewness on the t statistic.
Box plots illustrating $t_p$ for $P = 0.05, 0.25, 0.5, 0.75$ and 0.95.
We confirm the suggestion in Chapter 1 that for small to moderate sample sizes the effect of increasing population (positive) skewness is to shift the "location" of the t statistic to the left and introduce negative skewness. We note also that for very small samples population kurtosis will shorten the tails, while for large samples population skewness has little effect on the percentage points, evidently due to the central limit theorem.

We remark finally that the box plots provide an effective means of illustrating the robustness properties of the t statistic. Their use need not be confined to empirical distributions; a one-dimensional representation of a univariate distribution can be an effective aid in many areas of statistics, particular when comparisons are being made (Dodgson, 1986).

5.4 SMOOTHED SIMULATION ESTIMATES OF THE CORRELATION OF t AND $\sqrt{b_1}$.

To complete our Johnson bivariate representation of the joint distribution of t and $\sqrt{b_1}$ we need to consider their correlation. This was partially investigated in Section 2.3 using the k-statistic technique and Figure 2.5 gives approximate contours over the $\beta_1$-$\beta_2$ plane for $n = 100$. We now present simulation estimates, allowing us to assess the accuracy of the previous method, and give smoothed estimates which will be used later in the investigation. It may be noted that the Johnson method requires the correlation of variables after their marginal transformation to normality. This question is the subject of the final section.
Table 5.6 compares the correlation of $t$ and $\sqrt{b_1}$ given by

(1) formula (2.5), essentially $E(t/\sqrt{b_1})$ to $O(n^{-2})$, together with expressions for the means, and

(2) simulation.

The details regarding the simulation and points in the $\beta_1-\beta_2$ plane are given in the table and are identical to those used for the comparisons in Chapter 4.

To obtain approximate standard errors we use the expression valid for a bivariate normal population (Kendall and Stuart, 1977, equation (16.74))

$$V(r) = \frac{(1-\rho^2)^2}{n-1} \left[ 1 + \frac{\rho^2}{2n} \right] + O(n^{-3}) , \quad (5.1)$$

the sample estimate simply replacing $\rho$. While bivariate normality cannot be assumed (though will certainly provide a good approximation for large $n$) this expression may be expected to give some indication of the precision of the estimates.
Table 5.6
Estimates of $\rho_{\varepsilon \varepsilon}$ based on the following methods:

- Formulae (2.5), (2.1) and Geary (1947).
- Simulation using the inverse distribution function;
- 10000 samples for $n = 25, 50, 100$; 5000 samples for $n = 200$;
- (approximate SE's in units of 0.0001).

<p>| Distribution (1) $\beta_1 = 0$ $\beta_2 = 4.0$ | Distribution (2) $\beta_1 = 0$ $\beta_2 = 5.0$ |
|-----|-----|-----|-----|-----|</p>
<table>
<thead>
<tr>
<th>n</th>
<th>Formulae</th>
<th>Simulation</th>
<th>n</th>
<th>Formulae</th>
<th>Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.2861</td>
<td>0.2414 (94)</td>
<td>25</td>
<td>0.5273</td>
<td>0.3874 (85)</td>
</tr>
<tr>
<td>50</td>
<td>0.2744</td>
<td>0.2487 (94)</td>
<td>50</td>
<td>0.4812</td>
<td>0.4181 (83)</td>
</tr>
<tr>
<td>100</td>
<td>0.2665</td>
<td>0.2589 (93)</td>
<td>100</td>
<td>0.4449</td>
<td>0.4134 (83)</td>
</tr>
<tr>
<td>200</td>
<td>0.2623</td>
<td>0.2328 (134)</td>
<td>200</td>
<td>0.4264</td>
<td>0.4027 (119)</td>
</tr>
</tbody>
</table>

<p>| Distribution (3) $\beta_1 = 0$ $\beta_2 = 6.5$ | Distribution (4) $\beta_1 = 0.3$ $\beta_2 = 4.0$ |
|-----|-----|-----|-----|-----|</p>
<table>
<thead>
<tr>
<th>n</th>
<th>Formulae</th>
<th>Simulation</th>
<th>n</th>
<th>Formulae</th>
<th>Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>1.0513</td>
<td>0.5420 (71)</td>
<td>25</td>
<td>0.1335</td>
<td>0.1982 (96)</td>
</tr>
<tr>
<td>50</td>
<td>0.7967</td>
<td>0.5693 (68)</td>
<td>50</td>
<td>0.1444</td>
<td>0.1733 (97)</td>
</tr>
<tr>
<td>100</td>
<td>0.6739</td>
<td>0.5780 (67)</td>
<td>100</td>
<td>0.1457</td>
<td>0.2022 (96)</td>
</tr>
<tr>
<td>200</td>
<td>0.6201</td>
<td>0.5632 (97)</td>
<td>200</td>
<td>0.1457</td>
<td>0.1516 (138)</td>
</tr>
</tbody>
</table>
Table 5.6 (continued)

<table>
<thead>
<tr>
<th>Distribution (5) $\beta_1 = 0.3$ $\beta_2 = 5.0$</th>
<th>Distribution (6) $\beta_1 = 0.3$ $\beta_2 = 6.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>Formulae</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
</tr>
<tr>
<td>25</td>
<td>0.3197</td>
</tr>
<tr>
<td>50</td>
<td>0.3405</td>
</tr>
<tr>
<td>100</td>
<td>0.3250</td>
</tr>
<tr>
<td>200</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Distribution (7) $\beta_1 = 0.7$ $\beta_2 = 5.0$</th>
<th>Distribution (8) $\beta_1 = 0.7$ $\beta_2 = 6.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>Formulae</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
</tr>
<tr>
<td>25</td>
<td>0.1306</td>
</tr>
<tr>
<td>50</td>
<td>0.1751</td>
</tr>
<tr>
<td>100</td>
<td>0.1883</td>
</tr>
<tr>
<td>200</td>
<td>0.1920</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Distribution (9) $\beta_1 = 1.0$ $\beta_2 = 5.0$</th>
<th>Distribution (10) $\beta_1 = 1.0$ $\beta_2 = 5.8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>Formulae</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
</tr>
<tr>
<td>25</td>
<td>0.0206</td>
</tr>
<tr>
<td>50</td>
<td>0.0627</td>
</tr>
<tr>
<td>100</td>
<td>0.0842</td>
</tr>
<tr>
<td>200</td>
<td>0.0937</td>
</tr>
</tbody>
</table>
The results in Table 5.6 indicate that the k-statistic approach to the correlation of t and \( v_{b_1} \) in non-normal sampling is of little value. Confining our attention to large samples we may note that for symmetric populations their accuracy decreases with increasing kurtosis while for moderately skewed distributions the worst results tend to be associated with kurtosis near the normal value. We conclude, as in Chapter 4, that k-statistics are of little practical use for non-normal populations.

Unlike the case of the moments of \( v_{b_1} \), there does not appear to have been any previous investigation into the correlation of t and \( v_{b_1} \) (presumably this would only be prompted by an investigation of the type currently being undertaken). We therefore rely on the simulation estimates and smooth them across the \( \beta_1-\beta_2 \) plane in the manner previously described for the moments of \( v_{b_1} \).

The results provide the final column of the data file exhibited as Table 5.1. The smoothing was carried out with the variables \( \beta_1, \beta_2 \) and \( \beta_2^2 \) for all sample sizes, the largest rms being 0.00063 for \( n = 25 \). With distribution (9) and \( n = 100 \) and 200 there is a little concern, with the smoothed value approximately 0.02 greater than the original in both cases. Taken with a previous observation on the k-statistics results there is perhaps a suggestion that the correlation between t and \( v_{b_1} \) is most "problematic" for skew populations having the lowest possible kurtosis.

5.5 CORRELATION FOR FITTING JOHNSON'S S_{ij}.

It was explained in Section 1.4 that having fitted marginal
transformations to variates $X_1$ and $X_2$

$$Z_1 = \gamma_1 + \delta_1 f_1\left(\frac{X_1 - \xi_1}{\lambda_1}\right)$$

$$Z_2 = \gamma_2 + \delta_2 f_2\left(\frac{X_2 - \xi_2}{\lambda_2}\right)$$

the $S_{IJ}$ representation is completed by determining the correlation of $Z_1$ and $Z_2$.

When fitting to an empirical distribution this is readily achieved by calculating the correlation of the data pairs following their respective transformations. Two examples that could be cited are Johnson (1949b) and Schreuder and Hafley (1977) who fit $S_{UU}$ and $S_{BB}$ respectively. However if, as in our case, we have only marginal moments and the correlation of the original variables another approach is needed.

This problem, on closer examination, turns out to be far from straightforward and we will not attempt to cover all aspects here. Our priority will be on describing a simple method to be used in the remainder of the investigation. Some other possible approaches are outlined in the final section of the thesis.

We avoid double subscripts by writing $X$ and $Y$ for the original variables having correlation $\rho_{XY}$ before transformation and $\rho$ after. Two special cases of $S_{IJ}$ have been discussed by Kowalski (1972) in an investigation into the effect of non-normality on the correlation coefficient. He states

$$\rho_{XY} = \frac{\rho}{(e-1)^{\frac{1}{2}}} \quad \text{for } S_{NL} \quad (5.2)$$
and \[ \rho_{XY} = \frac{\exp(\rho-1)}{e-1} \quad \text{for } S_{LL}. \tag{5.3} \]

These results are most suspect. Neither contain \( \delta \) terms (one for \( S_{NL} \), two for \( S_{LL} \)) which would permit

\[ \rho_{XY} \rightarrow \rho \]
as the distribution tends to normality. (The equivalent expression to (5.2) given by Yuan (1933) does allow the correct limiting value, though for a different parameterisation of the marginal lognormal.)

Transposing (5.2) we obtain

\[ \rho = 1.3108\rho_{XY}. \]

Allowing \( \rho \) its full range this implies

\[ -0.7629 \leq \rho_{XY} \leq 0.7629. \]

Similarly from (5.3) we obtain

\[ \rho = 1.5413 + \ln \rho_{XY}. \]

implying

\[ 0.0788 \leq \rho_{XY} \leq 0.5820. \]

It is worth noting that both expressions give the correlation after transformation exceeding that before, and consequently a restricted range on \( \rho_{XY} \). In particular (5.3) excludes independent \( X \) and \( Y \), as this implies \( \rho = -\infty \). There is an interesting analogue in Fréchet (1959) who states that following marginal transformations to normality the resulting correlation will lie in a restricted interval contained within \([-1, 1]\).

We proceed by deriving a general expression valid for all \( S_{IJ} \).
using "statistical differentials". (I am grateful to Professor
N.L. Johnson for this suggestion.)

Consider the following Taylor expansion of the function

\[ h(X,Y) = h + h_x \delta X + h_y \delta Y + \frac{1}{2} (h_{XX} \delta X^2 + 2h_{XY} \delta X \delta Y + h_{YY} \delta Y^2) + \ldots \]

where

\[ \delta X = X - \mu_X \]
\[ \delta Y = Y - \mu_Y \]

and where \( h \) and its partial derivatives (denoted \( h_x, h_y \) etc.)
are evaluated at \((\mu_X, \mu_Y)\). Taking expectations

\[ E[h(X,Y)] = h + \frac{1}{2} h_{XX} \sigma_X^2 + h_{XY} \text{Cov}(X,Y) + \frac{1}{2} h_{YY} \sigma_Y^2. \]

Consider

\[ h(X,Y) = \left[ \gamma_X + \delta_X f_X \left( \frac{X - \xi_X}{\lambda_X} \right) \right] \left[ \gamma_Y + \delta_Y f_Y \left( \frac{Y - \xi_Y}{\lambda_Y} \right) \right] \]
\[ = Z_X(X) Z_Y(Y) \quad \text{(say).} \]

Then

\[ h_x(X,Y) = \delta_X f'_X \left( \frac{X - \xi_X}{\lambda_X} \right) Z_Y(Y) \]
\[ h_{XX}(X,Y) = \delta_X f''_X \left( \frac{X - \xi_X}{\lambda_X} \right) Z_Y(Y) \]
\[ h_{YY}(X,Y) = Z_X(X) \delta_Y f''_Y \left( \frac{Y - \xi_Y}{\lambda_Y} \right) \]
\[ h_{XY}(X,Y) = \delta_X f'_X \left( \frac{X - \xi_X}{\lambda_X} \right) \delta_Y f'_Y \left( \frac{Y - \xi_Y}{\lambda_Y} \right) \]
where prime denotes differentiation with respect to \( X \) or \( Y \) according to the function. Noting that \( Z_X(X) \) and \( Z_Y(Y) \) are standard normal variables their correlation is simply \( E[h(X,Y)] \). Therefore

\[
\rho \propto Z_X(\mu_X) Z_Y(\mu_Y) + \frac{1}{2} \delta_X f''_X \left( \frac{\mu_X - \xi_X}{\lambda_X} \right) Z_Y(\mu_Y) \sigma_X^2 \\
+ \delta_X f'_X \left( \frac{\mu_X - \xi_X}{\lambda_X} \right) \delta_Y f'_Y \left( \frac{\mu_Y - \xi_Y}{\lambda_Y} \right) \text{Cov}(X,Y) \\
+ \frac{1}{2} Z_X(\mu_X) \delta_Y f''_Y \left( \frac{\mu_Y - \xi_Y}{\lambda_Y} \right)^2 . \tag{5.4}
\]

This expression simplifies considerably if

\[
Z_X(\mu_X) = Z_Y(\mu_Y) = 0 .
\]

In fact for Johnson transformations it is the median of the original variable which is mapped to zero. Therefore for \( X \) and \( Y \) having distributions not too asymmetric we have

\[
\rho \propto \delta_X \delta_Y f'_X \left( \frac{\mu_X - \xi_X}{\lambda_X} \right) f'_Y \left( \frac{\mu_Y - \xi_Y}{\lambda_Y} \right) \text{Cov}(X,Y) . \tag{5.5}
\]

A general theorem by Lancaster (1957) is relevant in this context (also Kendall and Stuart, 1979, Section 33.44). If a bivariate distribution of \((X,Y)\) can be obtained from the bivariate normal by separate transformations on \( X \) and \( Y \), the correlation in the transformed distribution cannot exceed in absolute value the correlation in the bivariate normal distribution, i.e. \( |\rho_{XY}| \leq |\rho| \).
This result is intuitively reasonable when one imagines the effect on the elliptical contours of the bivariate normal surface of transformations on the marginal distributions. Unless the transformations are identical (apart from location and scale) the linear association between the variables is bound to be decreased.

We illustrate the discussion up to this point with an example. Johnson (1949a,b) demonstrates his univariate and bivariate systems with the joint distribution of the length and breadth of a large sample of beans. The same data, attributed to Johanssen, were used by Pretorius (1930) to illustrate the fitting of a bivariate Gram-Charlier (Type AA) distribution. The two approaches are compared in Mardia (1970).

Details of Johnson's original $S_{UU}$ fit together with that obtained using algorithm AS 99 (see Section 5.2) are included in Table 5.7. The moments quoted by Johnson differ slightly from those given by Mardia, and determined with Sheppard's correction, and these latter have been used with the algorithm. The main cause of the discrepancies between the parameters, however, is certainly the fact that Johnson, not having the algorithm, had to use an abac to obtain $\gamma$ and $\delta$, and subsequently $\lambda$ and $\xi$. 
Table 5.7

Fitting Johnson's $S_{UU}$ to the length and breadth of 9440 beans.

(a) Johnson's original fit.
(b) Mardia's moments AS 99 fit.

<table>
<thead>
<tr>
<th>Moment</th>
<th>Length</th>
<th>Breadth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a)</td>
<td>(b)</td>
</tr>
<tr>
<td>mean</td>
<td>14.399</td>
<td>14.4046</td>
</tr>
<tr>
<td>sd</td>
<td>0.904</td>
<td>0.8998</td>
</tr>
<tr>
<td>$\sqrt{\beta_1}$</td>
<td>-0.910</td>
<td>-0.9106</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>4.863</td>
<td>4.8629</td>
</tr>
<tr>
<td>Parameter</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(a)</td>
<td>(b)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>2.38</td>
<td>2.4498</td>
</tr>
<tr>
<td>$\delta$</td>
<td>2.64</td>
<td>2.6394</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>1.5192</td>
<td>1.4808</td>
</tr>
<tr>
<td>$\xi$</td>
<td>16.0745</td>
<td>16.1026</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original data (a)</th>
<th>Transformed data (a)</th>
<th>Formulae (5.4) and (5.5) (a)</th>
<th>Formulae (5.4) and (5.5) (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td>0.7811</td>
<td>0.746</td>
<td>0.840 (5.4) 0.843</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.853 (5.5) 0.855</td>
</tr>
</tbody>
</table>

Before commenting on the figures for correlation we develop formulae (5.4) and (5.5) for the system $S_{UU}$. For both marginal transformations
\[ Z = \gamma + \delta f \left( \frac{x - \xi}{\lambda} \right) \]

where \[ f(x) = \sinh^{-1} \left( \frac{x - \xi}{\lambda} \right). \]

We have \[ f'(x) = \frac{1}{\lambda \sqrt{y^2 + 1}} \]

and \[ f''(x) = \frac{-y}{\lambda^2 (y^2 + 1)^{3/2}} \]

where (in Johnson's original notation)

\[ y = \frac{x - \xi}{\lambda}. \]

Table 5.7 allows us to compare

(1) the original correlation (from Mardia),
(2) the correlation following Johnson's data transformations,
(3) the transformed correlation using formulae (5.4) and (5.5)

with both fits.

One immediate remark is that the correlation after the marginal data transformations is less than the original correlation. This contravenes Lancaster's theorem. Thus we cannot have achieved a transformation to bivariate normality, and yet by restricting our attention to marginal transformations there is no reason to expect that we could. In fact the transformation provides a reasonable approximation to bivariate normality because, as will be seen shortly, the \( S_{uu} \) representation of the distribution, at least to the extent of its regression properties, is very good.

In contrast the formulae do give an increase in correlation following transformation of approximately 8%. We also note that
the simple expression (5.5) provides a reasonable approximation to (5.4). This is expected for marginals not too asymmetric, the table showing moderate skewness in the variables.

Despite the fact that the correlation increases following transformation in line with Lancaster's theorem the use of the formulae without further work seems premature. Some factors that would need investigation are the effects of marginal skewness and kurtosis, and consequent system, and the importance of the neglected terms. Apparently there is no guidance in the literature where Johnson's approach is accepted. Again given that the marginal transformations are used it is difficult to argue against the resulting correlation, especially when based on a very large sample.

These and related issues are returned to in Section 7.3. At present we still need a method of estimating the correlation for an $S_{IJ}$ representation of a joint distribution given not an empirical distribution but approximations to the marginal moments and the correlation of the original variables.

The method suggested is simple: assume the correlation after transformation is equal to the correlation of the original variables.

We take the precaution of comparing the method with Johnson's using the bean data. Now for $(X,Y)$ represented by $S_{IJ}$ the conditional (array) distribution of $Y$ given $X$ follows the same system as $Y$, but with modified $\gamma$ and $\delta$ (see Section 1.4). We are concerned with $S_{UU}$ having marginal moments (Johnson, 1949a)

$$\mu_j' = -\omega^j \sinh \Omega$$
\[ \mu_2 = \frac{1}{2} (\omega-1)(\omega \cosh 2\Omega + 1) \]

\[ \mu_3 = -\frac{1}{4} \omega^4 (\omega-1) \left\{ \omega (\omega+2) \sinh 3\Omega + 3 \sinh \Omega \right\} \]

\[ \mu_4 = \frac{1}{8} (\omega-1)^2 \left\{ \omega^2 (\omega^4 + 2\omega^3 + 3\omega^2 - 3) \cosh 4\Omega + 4\omega^2 (\omega+2) \cosh 2\Omega + 3(2\omega + 1) \right\} \]

where \[ \omega = e^{\delta-2} \]

and \[ \Omega = \gamma/\delta \]

Subroutines JNARR and SUMOM (on the floppy disk) determine respectively the modified \( \gamma \) and \( \delta \) values for a conditional distribution, and the first four moments of the SU system. These are used to compare the moments of the distribution of breadth on length, and length on breadth, using three approaches:

(a) Johnson's (1949a,b) fit with correlation determined from the transformed data.

(b) Johnson's fit with correlation determined using the original data.

(c) Algorithm AS 99 with Mardia's (1970) moments and correlation from the original data.

Numerical results are presented in Table 5.8 while Figure 5.2 compares Johnson's fit (a) and the proposed method (c) with the empirical moments. (A to H correspond to Johnson's (1949b) Figures 10 to 17, themselves derived from Pretorius (1930).)

A general comment on Table 5.8 is that results (b) and (c) are similar, indicating that it is the value of correlation, rather than the numerical accuracy of the marginal fits, that
Table 5.8

Moments of conditional (array) distributions for bean data.

(a) Johnson's fit with correlation from transformed data.

(b) Johnson's fit with correlation from original data.

(c) AS 99 with Mardia's moments and correlation from original data.

<table>
<thead>
<tr>
<th>L</th>
<th>(E(B/L)) (a)</th>
<th>(E(B/L)) (b)</th>
<th>(E(B/L)) (c)</th>
<th>(SD(B/L)) (a)</th>
<th>(SD(B/L)) (b)</th>
<th>(SD(B/L)) (c)</th>
<th>(\sqrt{\beta_1(B/L)}) (a)</th>
<th>(\sqrt{\beta_1(B/L)}) (b)</th>
<th>(\sqrt{\beta_1(B/L)}) (c)</th>
<th>(\beta_2(B/L)) (a)</th>
<th>(\beta_2(B/L)) (b)</th>
<th>(\beta_2(B/L)) (c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>7.013</td>
<td>6.957</td>
<td>6.960</td>
<td>0.356</td>
<td>0.342</td>
<td>0.341</td>
<td>-0.491</td>
<td>-0.464</td>
<td>-0.474</td>
<td>3.473</td>
<td>3.418</td>
<td>3.443</td>
</tr>
<tr>
<td>11</td>
<td>7.187</td>
<td>7.143</td>
<td>7.145</td>
<td>0.329</td>
<td>0.314</td>
<td>0.312</td>
<td>-0.475</td>
<td>-0.448</td>
<td>-0.457</td>
<td>3.452</td>
<td>3.400</td>
<td>3.420</td>
</tr>
<tr>
<td>12</td>
<td>7.379</td>
<td>7.348</td>
<td>7.347</td>
<td>0.299</td>
<td>0.285</td>
<td>0.283</td>
<td>-0.452</td>
<td>-0.427</td>
<td>-0.431</td>
<td>3.423</td>
<td>3.374</td>
<td>3.389</td>
</tr>
<tr>
<td>13</td>
<td>7.595</td>
<td>7.576</td>
<td>7.574</td>
<td>0.268</td>
<td>0.254</td>
<td>0.252</td>
<td>-0.416</td>
<td>-0.393</td>
<td>-0.391</td>
<td>3.381</td>
<td>3.337</td>
<td>3.345</td>
</tr>
<tr>
<td>14</td>
<td>7.844</td>
<td>7.839</td>
<td>7.834</td>
<td>0.237</td>
<td>0.222</td>
<td>0.221</td>
<td>-0.358</td>
<td>-0.336</td>
<td>-0.324</td>
<td>3.320</td>
<td>3.281</td>
<td>3.281</td>
</tr>
<tr>
<td>15</td>
<td>8.142</td>
<td>8.150</td>
<td>8.147</td>
<td>0.206</td>
<td>0.193</td>
<td>0.194</td>
<td>-0.253</td>
<td>-0.233</td>
<td>-0.203</td>
<td>3.234</td>
<td>3.202</td>
<td>3.195</td>
</tr>
<tr>
<td>16</td>
<td>8.493</td>
<td>8.516</td>
<td>8.526</td>
<td>0.187</td>
<td>0.175</td>
<td>0.181</td>
<td>-0.074</td>
<td>-0.057</td>
<td>-0.001</td>
<td>3.156</td>
<td>3.134</td>
<td>3.140</td>
</tr>
<tr>
<td>17</td>
<td>8.839</td>
<td>8.878</td>
<td>8.919</td>
<td>0.190</td>
<td>0.180</td>
<td>0.195</td>
<td>-0.126</td>
<td>0.138</td>
<td>0.209</td>
<td>3.170</td>
<td>3.155</td>
<td>3.199</td>
</tr>
</tbody>
</table>
Table 5.8 (continued)

Length on breadth

<table>
<thead>
<tr>
<th>B</th>
<th>E(L/B)</th>
<th>SD(L/B)</th>
<th>$\sqrt{\beta_1(B/L)}$</th>
<th>$\beta_2(B/L)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a)</td>
<td>(b)</td>
<td>(c)</td>
<td>(a)</td>
</tr>
<tr>
<td>6.5</td>
<td>11.332</td>
<td>11.129</td>
<td>11.124</td>
<td>-0.748</td>
</tr>
<tr>
<td></td>
<td>1.278</td>
<td>1.242</td>
<td>1.247</td>
<td>4.039</td>
</tr>
<tr>
<td>7.0</td>
<td>12.401</td>
<td>12.284</td>
<td>12.285</td>
<td>-0.726</td>
</tr>
<tr>
<td></td>
<td>1.021</td>
<td>0.981</td>
<td>0.984</td>
<td>3.994</td>
</tr>
<tr>
<td>7.5</td>
<td>13.450</td>
<td>13.404</td>
<td>13.414</td>
<td>-0.679</td>
</tr>
<tr>
<td></td>
<td>0.781</td>
<td>0.739</td>
<td>0.739</td>
<td>3.905</td>
</tr>
<tr>
<td>8.0</td>
<td>14.460</td>
<td>14.466</td>
<td>14.480</td>
<td>-0.571</td>
</tr>
<tr>
<td></td>
<td>0.573</td>
<td>0.534</td>
<td>0.530</td>
<td>3.721</td>
</tr>
<tr>
<td>8.5</td>
<td>15.392</td>
<td>15.433</td>
<td>15.426</td>
<td>-0.321</td>
</tr>
<tr>
<td></td>
<td>0.433</td>
<td>0.400</td>
<td>0.395</td>
<td>3.419</td>
</tr>
<tr>
<td>9.0</td>
<td>16.206</td>
<td>16.278</td>
<td>16.212</td>
<td>0.067</td>
</tr>
<tr>
<td></td>
<td>0.397</td>
<td>0.373</td>
<td>0.361</td>
<td>3.287</td>
</tr>
</tbody>
</table>
Figure 5.2
Johnson $S_{UU}$ fit to bean data.
- Johnson's fit.
- - - proposed method.
- - - empirical values.

A

Regression
Breadth on length

B

Skewness
Breadth on length

C

Skewness
Breadth on length

D

Kurtosis
Breadth on length

E

Regression
Length on breadth

F

Kurtosis
Length on breadth

G

Skewness
Length on breadth

H

Kurtosis
Length on breadth
is most influential. There is close agreement for the regressions $E(B/L)$ and $E(L/B)$ among the three methods, particular when the regressor is in the central region. Figure 5.2, A and E, indicates that both the original and proposed methods give satisfactory fits, the latter giving slightly better representation for small values of the regressor.

It may be argued that in practical terms the higher order moments become progressively less important. Certainly the Johnson representation becomes less effective. The shift in the scedasticity curves with the higher correlation of the suggested method gives a better fit for all but the largest values of the regressor. This shift is also evident and generally effective for skewness and kurtosis, though sampling variability (presumably) makes the empirical values appear rather erratic. For these four curves Johnson comments on his own fit: "these lines would be improved if they were displaced to the right". The proposed method achieves this, though to less than the ideal degree. The new scedasticity curves are still disappointing, and further work would be needed to decide whether this is a feature of the data or the method.

The general impression, at least for this example, is that the Johnson approach tends to underestimate the association between the variables following their assumed transformation to bivariate normality. The proposed method does slightly better. Ideally an improved method should be found to achieve the transformation to bivariate normality, and there is some discussion on this possibility later. For the present we argue that the proposed
method gives a fit that is at least as good as the original. Moreover it provides a simple solution to the problem of fitting $S_{ij}$ in a theoretical investigation where only marginal moments and the original correlation are available. It will be used in the sequel to represent the joint distribution of $t$ and $\sqrt{b_1}$. 
Chapter Six

The Conditional Distribution of $t$ given $\sqrt{b_1}$

6.1 THE CONDITIONAL (ARRAY) DISTRIBUTION.

The purpose of our investigation is to study the effect of a preliminary test of normality based on $\sqrt{b_1}$ on the distribution of the $t$ statistic. As a first target consider the conditional (array) density $f(t/\sqrt{b_1} = k)$, where the population is assumed Gram-Charlier and the skewness and kurtosis are allowed to vary.

The previous chapter concerned the fitting of the Johnson system to the marginal and joint distributions of $t$ and $\sqrt{b_1}$. One advantage of the Johnson bivariate system is the simple form of the array distributions, and these were used in the discussion on the methods of estimating correlation. (The method is described in Section 1.4, with related algorithm JNABR on the disk.)

The diagramatic presentation of results may require some explanation. Against the vertical axis are represented with box plots, illustrating the 5%, 25%, 50%, 75% and 95% points (as in Figure 5.1), the following distributions:

(1) The nominal (i.e. tabulated) distribution of $t$.

(2) The marginal distribution of $t$, using a Johnson approximation based on the smoothed moments in Table 5.1.

(3) The conditional (array) distribution of $t$, using the Johnson method with the smoothed moments and correlation in Table 5.1, and the given value of $\sqrt{b_1}$.

The horizontal axis gives the values of $\sqrt{b_1}$. A partial box plot,
typically showing only the central or upper half of the distribution, together with the nominal 5% and 1% points, indicates the extent to which $\sqrt{b_1}$ may be considered "extreme" and/or "significant", and further gives a rough indication of power in detecting non-normality.

We begin by considering symmetric populations and choose distribution (3) (see Figure 1.1) having (for the Gram-Charlier population) the most extreme kurtosis $\beta_2 = 7.0$. Figure 6.1 illustrates the position for $n = 25$.

Examining the diagram we see immediately the close agreement between the nominal and marginal distributions of $t$. This robustness is expected for a symmetric population, as discussed in Sections 1.1 and 5.3. Turning to the conditional distributions the dependence of $t$ on $\sqrt{b_1}$ is reflected in $f(t/b_1)$ shifting linearly towards the upper tail with increasing $\sqrt{b_1}$, with spread and symmetry effectively unaltered. (The approximation in question is $S_B$ whose moments cannot be expressed in closed form, though examination of Table 36, Biometrika Tables Volume 2 (Pearson and Hartley, 1976), suggests the second, third and fourth moments are not exactly constant.)

At $\sqrt{b_1} = 0$ the conditional distribution is centred on zero, like the nominal and marginal distributions, though with less spread. With increasing $\sqrt{b_1}$, i.e. the test of normality becoming more significant, the conditional distribution becomes less well represented by the nominal (and marginal) distributions. However because the conditional distribution has less spread, it can be the case that its upper percentage points approximate more closely to the nominal values for "significant" $\sqrt{b_1}$ than for $\sqrt{b_1} = 0$. (Similarly for lower percentiles and large negative $\sqrt{b_1}$.)
Figure 6.1

The following distributions are illustrated for distribution (3), \( \beta_1 = 0 \) and \( \beta_2 = 7.0 \), \( n = 25 \):

- Nominal and marginal distributions of \( t \).
- Conditional (array) distribution of \( t \) given \( \nu_{b_1} \).
- Marginal distribution of \( \nu_{b_1} \) (part), with nominal percentage points.
Figure 6.2

The following distributions are illustrated for distribution (3), $B_1 = 0$ and $B_2 = 7.0$, $n = 100$:

Nominal and marginal distributions of $t$.

Conditional (array) distribution of $t$ given $\sqrt{b_1}$.

Marginal distribution of $\sqrt{b_1}$ (part), with nominal percentage points.
Figure 6.2 is the comparative diagram for \( n = 100 \). Note that the nominal and marginal distributions are virtually identical by virtue of the robustness of \( t \), and virtually normal by the central limit theorem. We see that the implied line of median regression is steeper than its counterpart in Figure 6.1 (drawn to the same scale), a fact which results more from the variance ratio of \( t \) to \( \sqrt{b_1 (0(n^{-1}))} \) than the increase in estimated correlation (see Table 5.1). Thus the shift in the conditional distribution of \( t \) relative to its nominal distribution for a given change in \( \sqrt{b_1} \) is greater.

It is this shift in the conditional distribution from being centred on zero for \( \sqrt{b_1} = 0 \) to the "right" for increasing \( \sqrt{b_1} \) which characterises the behaviour of \( t \) for symmetric populations. The shift is dependent on the correlation which (again from Table 5.1) increases with population kurtosis and, to an extent, sample size. Thus as \( \sqrt{b_1} \) becomes more significant the distribution of \( t \) becomes less well represented (over the whole range) by the nominal distribution.

The power of \( \sqrt{b_1} \) in detecting non-normality is also relevant. Since we are dealing with symmetric populations this is not great, and it can be seen from the diagrams to be approximately 27% for a test at the 5% level for both \( n = 25 \) and 100. The power of \( \sqrt{b_1} \) was discussed in Section 5.2 and figures for all ten distributions and four sample sizes are presented in Table 5.4.

It is interesting to compare our findings with one of the conclusions of Eastering and Anderson (1978), outlined in Section 1.2, that for symmetric non-normal populations a preliminary test of normality improves the representation of the distribution of
the $t$ statistic by Student's $t$ distribution, though only for large population kurtosis is the effect pronounced. The above suggests that such an effect will occur using $\sqrt{b_1}$ for the preliminary test, and the relevance of kurtosis to the correlation between $t$ and $\sqrt{b_1}$. However we resist the temptation to draw any further conclusions at this stage and consider asymmetric populations.

Distribution (10) represents the most extreme non-normal distribution in this part of the study, having $\beta_1 = 1$ and $\beta_2 = 5.8$. The nominal, marginal and conditional distributions for $n = 25$ and 100 are illustrated, respectively, in Figures 6.3 and 6.4.

Figure 6.3, for $n = 25$, shows how the skewness in the population affects the marginal distribution of $t$ in relation to its nominal distribution - characterised previously as a shift to the "left" with introduced negative skewness. The population skewness is also, of course, reflected in the position of the marginal distribution of $\sqrt{b_1}$, and it may be confirmed that the power of a test of normality at the 5% level is approximately 61%.

As expected the conditional distribution shifts with $\sqrt{b_1}$, though the correlation is less than for the symmetric population (3). The best match between the nominal and conditional distributions occurs for $\sqrt{b_1}$ taking values near 1.0, the population skewness. Thus we see that the representation by the nominal distribution improves as $\sqrt{b_1}$ increases from zero to 1.0, i.e. as $\sqrt{b_1}$ becomes more significant. Thus we are finally focusing on our original problem, identified by Easterling and Anderson (1978) and outlined in Section 1.2: a preliminary test of normality can make worse the
Figure 6.3

The following distributions are illustrated for distribution (10), $\beta_1 = 1$ and $\beta_2 = 5.8$, $n = 25$:

- Nominal and marginal distributions of $t$.
- Conditional (array) distribution of $t$ given $\nu b_1$.
- Marginal distribution of $\nu b_1$ (part), with nominal percentage points.
The following distributions are illustrated for distribution (10), $\beta_1 = 1$ and $\beta_2 = 5.8$, $n = 100$:

Nominal and marginal distributions of $t$.

Conditional (array) distribution of $t$ given $\sqrt{b_1}$.

Marginal distribution of $\sqrt{b_1}$ (part), with nominal percentage points.
representation of the distribution of $t$ by its nominal distribution.

A comparison of the nominal and marginal distributions of $t$ in Figure 6.4 shows that with the larger sample size of 100 the robustness of $t$ essentially "overcomes" the non-normality in the population. For skew populations the power of $\sqrt{b_1}$ in a test of normality naturally increases with sample size and is approximately 89% for a test at the 5% level. Again we see the shift in the conditional distribution with $\sqrt{b_1}$, and the best representation of the conditional by the marginal around $\sqrt{b_1} = 1.0$, a value which would be judged very highly significant.

Having examined two non-normal populations (and two sample sizes) let us try to understand the general picture. Several factors need to be taken into consideration, of which the correlation between $t$ and $\sqrt{b_1}$ is crucial.

The contours in Figure 2.5 based on the k-statistic approximation, the values in Table 5.1 giving smoothed simulation estimates, and further work to be presented in the final chapter all show that correlation increases with population kurtosis and decreases with population skewness. It may be remembered, however, that for our population model we cannot have skewness without kurtosis.

Consider now the effect of increasing population skewness. (It may be assumed that kurtosis is large enough to allow this, say $5 \leq \beta_2 \leq 7$.) The marginal distribution of $t$ becomes distorted slightly in the manner previously illustrated and discussed. The distribution of $\sqrt{b_1}$ shifts with $\sqrt{b_1}$. Its power in detecting non-normality increases and "ideal" values of $\sqrt{b_1}$ (i.e. near zero) become further into the lower tail.
The correlation between $t$ and $\sqrt{b_1}$, though decreasing slightly with skewness, will be sufficient for the location of the conditional distribution of $t$ to depend on $\sqrt{b_1}$. A small value of $\sqrt{b_1}$, being in the lower tail of its marginal, will therefore result in the conditional distribution of $t$ being shifted to the left of its nominal.

The effect of increasing sample size is to exaggerate the shift effect because of the reduction in the variability of $\sqrt{b_1}$. Examination of Table 5.1 indicates that for the more non-normal distributions the correlation between $t$ and $\sqrt{b_1}$ decreases as $n$ increases from 25, and this is to some extent compensatory. A smaller sample size is included in a later supplementary investigation, which also extends the $\beta_1$ and $\beta_2$ values beyond the Gram-Charlier region.

These points may be illustrated by displaying the conditional distributions $f(t/\sqrt{b_1})$ with $\sqrt{b_1}$ held fixed for all ten distributions. We choose $\sqrt{b_1} = 0$, of special interest being the "ideal" value. The conditional distributions, together with the nominal, are illustrated for sample sizes 25 and 100 in Figures 6.5 and 6.6 respectively. Given with the $\beta_1$ and $\beta_2$ values is the power of rejecting normality using $\sqrt{b_1}$ in a test at the 5% level.

What is most noticeable is that given a symmetric sample the location of the conditional distribution can vary so much with the actual amount of skewness and kurtosis in the population. Generally the shift in the conditional distribution increases with non-normality, population skewness being responsible for making $\sqrt{b_1} = 0$ "extreme", and kurtosis providing the correlation
Figure 6.5

The nominal distribution of $t$, and the conditional distribution $f(t/\sqrt{b_1} = 0)$ for distributions (1) to (10), $n = 25$. 

<table>
<thead>
<tr>
<th>Distribution</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_1$</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.7</td>
<td>0.7</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>$b_2$</td>
<td>4.0</td>
<td>5.0</td>
<td>6.5</td>
<td>4.0</td>
<td>5.0</td>
<td>6.5</td>
<td>5.0</td>
<td>6.0</td>
<td>5.0</td>
<td>5.8</td>
</tr>
<tr>
<td>Power (5%)</td>
<td>0.12</td>
<td>0.19</td>
<td>0.26</td>
<td>0.30</td>
<td>0.38</td>
<td>0.45</td>
<td>0.51</td>
<td>0.54</td>
<td>0.60</td>
<td>0.61</td>
</tr>
</tbody>
</table>
Figure 6.6

The nominal distribution of $t$, and the conditional distributions $f(t/\sqrt{\beta_1} = 0)$ for distributions (1) to (10), $n = 100$. 

<table>
<thead>
<tr>
<th>Distribution</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.7</td>
<td>0.7</td>
<td>1.0</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>4.0</td>
<td>5.0</td>
<td>6.5</td>
<td>4.0</td>
<td>5.0</td>
<td>6.5</td>
<td>5.0</td>
<td>6.0</td>
<td>5.0</td>
<td>5.8</td>
</tr>
<tr>
<td>Power (5%)</td>
<td>0.14</td>
<td>0.20</td>
<td>0.27</td>
<td>0.67</td>
<td>0.64</td>
<td>0.62</td>
<td>0.86</td>
<td>0.81</td>
<td>0.93</td>
<td>0.89</td>
</tr>
</tbody>
</table>
between \( t \) and \( \sqrt{b_1} \). The effect of increasing skewness, however, is attenuated by the corresponding decrease in correlation.

For moderate non-normality the shift effect increases with sample size. Of course by \( n = 100 \) the marginal and nominal distributions of \( t \) will be virtually identical for all the distributions considered.

These arguments must be tempered to some extent by considerations of power, and the probability of \( \sqrt{b_1} \) taking such a relatively extreme value. Such questions are best considered in the context of the distribution of \( t \) conditional on \( \sqrt{b_1} \) lying within an interval - the subject of the next section. However on the present evidence we must question the value of a preliminary test of normality. Consider, for example, distribution (5), exhibiting fairly mild non-normality. For \( n = 25 \) the power of rejecting normality for a test at the 5% level is approximately 38%; \( \sqrt{b_1} = 0 \) lies just above the 25% point of the marginal distribution of \( \sqrt{b_1} \). Yet the difference between the nominal and conditional distributions of \( t \) is quite marked.

A final point concerns conditioning on the value of \( \sqrt{b_1} \) near the population value \( \sqrt{\beta_1} \). For symmetric populations the spread of the conditional distribution, which itself is symmetric, is reduced (with kurtosis), suggesting inference based on the nominal distribution will be conservative. For asymmetric populations the effect of such conditioning is to correct for the location of the conditional distribution, with little reduction in spread and with the skewness in the marginal distribution persisting in the conditional.
These remarks suggest the possibility of using $\sqrt{b_1}$ in some ancillary way. The definition of ancillarity given by Cox and Hinkley (1974, p35) appears to exclude this possibility in a formal sense, in particular because the conditional distribution of $t$ given $\sqrt{b_1}$ does depend on $\sqrt{b_1}$ (at least in approximation).

Another approach, suggested by Johnson (1978) and referred to in Chapter 1 in the context of robust and adaptive procedures, modifies the value of the $t$ statistic using the observed value of $\sqrt{b_1}$. Further speculation on these points will be found in the final chapter.

6.2 THE CONDITIONAL DISTRIBUTION GIVEN $\sqrt{b_1}$ LIES WITHIN AN INTERVAL.

By examining the conditional distribution $f(t/\sqrt{b_1})$ we have begun to understand the effect of a preliminary test of normality using $\sqrt{b_1}$ on the distribution of the $t$ statistic. However a formal decision on whether to accept or reject normality requires an arbitrary significance level and corresponding rejection region. In this section, therefore, we consider the distribution of $t$ given that $\sqrt{b_1}$ lies within an interval.

A direct approach is via distribution functions. Let $F(\sqrt{b_1})$ denote the marginal distribution function of $\sqrt{b_1}$, and $F(t, \sqrt{b_1})$ denote the joint distribution function of $t$ and $\sqrt{b_1}$. We have

$$P(t \leq t_0/\sqrt{b_1} \leq k) = \frac{P[(t \leq t_0) \cap (\sqrt{b_1} \leq k)]}{P(\sqrt{b_1} \leq k)}$$

$$= \frac{F(t_0, k)}{F(k)}.$$
We write
\[ F(t_{out}/b_1 \leq k) = \frac{F(\theta, k)}{F(k)} . \] (6.1)

Following Wilks (1962) the function \( F(t_{out}/b_1 \leq k) \) is termed the conditional distribution function of \( t \) given \( b_1 \leq k \). (Note that conventionally \( F(t/k) \) represents the conditional distribution function of \( t \) given \( b_1 = k \).)

By the nature of the Johnson approximation to the joint distribution function of \( t \) and \( b_1 \), attention now turns to the bivariate normal distribution function.

Let \( Z_1 \) and \( Z_2 \) denote variables following the standard bivariate normal distribution with correlation \( \rho \). Equation (1.8) gives their joint pdf; we denote their joint distribution function \( \Phi(z_1, z_2; \rho) \). Methods of evaluating \( \Phi(z_1, z_2; \rho) \) are discussed in Johnson and Kotz (1972). We follow Daley (1974) and write
\[ \Phi(z_1, z_2; \rho) = \frac{1}{2} \left[ \Phi(z_1) + \Phi(z_2) - \delta(z_1, z_2) \right] - T(z_1, a_1) - T(z_2, a_2) \] (6.2)

where
\[ \Phi(z) \] is the standard normal distribution function,
\[ T(z, a) \] is Owen's (1956) \( T \)-function,
\[ \delta(z_1, z_2) = \begin{cases} 0 & \text{if } z_1 z_2 > 0 \text{ or } z_1 z_2 = 0 \text{ and } z_1 + z_2 > 0 \\ 1 & \text{otherwise,} \end{cases} \]
\[ a_i = \frac{(z_i + 1/z_i - \rho)}{(1-\rho^2)^{1/2}}, \quad i = 1, 2, \text{ where } z_3 = z_1. \]

The normal integral \( \Phi(z) \) may be evaluated with algorithm AS 66 (references in Section 5.2). The \( T \)-function, with conventional
arguments, is defined

\[ T(h, a) = \frac{1}{2\pi} \int_0^\infty \frac{\exp((-h^2/2)(1+x^2))}{1 + x^2} \, dx \quad \infty < h, a < \infty. \]

Algorithm AS 76 (Young and Minder, 1974; corrected in Hill, 1978 and Thomas, 1979, noting also pp 113 and 336 of Applied Statistics, 28 (1979)) evaluates this integral.

In fact difficulties were encountered with AS 76 (and indeed the earlier AS 4) because of the nature of its arguments when evaluating (6.2) for \((Z_1, Z_2)\) varying over its region. In order to guarantee a failsafe routine the following results, given with others by Owen (1956), were used in preference to AS 76 if the conditions (in computational terms) were met:

\[ T(h, 0) = 0; \]
\[ T(0, a) = \frac{1}{2\pi} \tan^{-1} a; \]
\[ T(h, \infty) = \begin{cases} 0.5[1 - \phi(h)] & h > 0 \\ 0.5 \phi(h) & h < 0. \end{cases} \]

Also for large but not extreme arguments we used

\[ T(h, a) = \frac{1}{2\pi} \tan^{-1} a - 0.5[\phi(h) - 0.5] + \frac{1}{2\pi} \tan^{-1} \left( \frac{1}{a} \right) \text{ for } |ha| > 4 \]

and

\[ T(h, a) = 0.5 \phi(h) + 0.5 \phi(ah) - \phi(h) \phi(ah) - T(ah, \frac{1}{a}) \text{ for } a > 1. \]

The first is based on an approximation to a function discussed by Nicholson (1943), shown by Owen to be related to the T-function; the second is an identity given by Owen. The conditions for these five cases are tested in the order presented.
Subsequent to this development Chou (1985) published AS R55 to overcome the same difficulties. At present this has not been evaluated.

We now return to our task of approximating the conditional distribution $F(t/v_{b_1} \leq k)$. Let $Z_t(t)$ and $Z_{v_{b_1}}(v_{b_1})$ represent the (marginal) Johnson transformations of $t$ and $v_{b_1}$, the transformed variates assumed to follow the standard bivariate normal distribution with correlation $\rho$. From (6.1) our approximation is

$$F(t/v_{b_1} \leq k) = \frac{\phi(Z_t(t), Z_{v_{b_1}}(k); \rho)}{\phi(Z_{v_{b_1}}(k))}.$$

To continue with the presentation of results using box plots we require the quantiles of $F(t/v_{b_1} \leq k)$. We write

$$\frac{\phi(z_1, z_2; \rho)}{\phi(z_2)} = p$$

i.e.

$$\phi(z_1, z_2; \rho) - p\phi(z_2) = 0.$$

For given $z_2$, $\rho$ and $p$ this is solved for $z_1$ using regular falsi.

To investigate the effect of the decision to accept/reject normality on the distribution of the $t$ statistic we also require $F(t/v_{b_1} > k)$. Now

$$P(t \leq t_o/v_{b_1} > k) = \frac{P((t \leq t_o) \cap (v_{b_1} > k])}{P(v_{b_1} > k)}.$$

For the numerator we have

$$P(t \leq t_o) = P((t \leq t_o) \cap (v_{b_1} \leq k)] + P((t \leq t_o) \cap (v_{b_1} > k)].$$
Therefore

\[
F(t/\sqrt{b_1} > k) = \frac{F_t(t) - F(t, k)}{1 - F_{\sqrt{b_1}}(k)}.
\]

To obtain the quantiles of \( F(t/\sqrt{b_1} > k) \) an argument similar to that above requires us to solve for \( z_1 \)

\[
\phi(z_1) - \phi(z_1, z_2; \rho) - p(1 - \phi(z_2)) = 0.
\]

Again regular falsi is used.

We now compare the nominal distribution of \( t \) with the conditional distributions \( f(t/\sqrt{b_1} \leq k) \) and \( f(t/\sqrt{b_1} > k) \), where \( k \) is a conventional critical value, i.e. we compare the nominal with the "accept" and "reject" conditional distributions. (We are considering a one-sided test of normality; the two-sided case is considered in the final section.) Choosing sample size 25 and test of normality at the 5% level we have \( k = 0.711 \) (Pearson and Hartley, 1976). Figure 6.7 illustrates the position for distributions (1), (2) and (3), i.e. symmetric populations with increasing kurtosis.

The comparison continues in Figure 6.8, for distributions (4), (5) and (6) \( (\beta_1 = 0.3) \), and finally Figure 6.9, illustrating distributions (7) and (8) \( (\beta_1 = 0.7) \), and also distributions (9) and (10) \( (\beta_1 = 1.0) \).

The approximate power of the corresponding test of normality is also given.

Figures 6.7 to 6.9 effectively integrate the information provided by the array distributions discussed in the previous section, indeed literally, being the integral of \( f(t/\sqrt{b_1}) \) over the accept and reject regions.
Figure 6.7

The nominal distribution of $t$, and the conditional distributions $f(t/\text{Accept})$ and $f(t/\text{Reject})$ for
distributions (1), (2) and (3), test at 5% (one-sided), $n = 25$. 

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0</td>
<td>4.0</td>
<td>0.12</td>
</tr>
<tr>
<td>(2)</td>
<td>0</td>
<td>5.0</td>
<td>0.19</td>
</tr>
<tr>
<td>(3)</td>
<td>0</td>
<td>6.5</td>
<td>0.26</td>
</tr>
</tbody>
</table>
Figure 6.8

The nominal distribution of $t$, and the conditional distributions $f(t/\text{Accept})$ and $f(t/\text{Reject})$ for distributions (4), (5) and (6), test at 5% (one-sided), $n = 25$.

<table>
<thead>
<tr>
<th>Accept/Reject</th>
<th>A</th>
<th>R</th>
<th>A</th>
<th>R</th>
<th>A</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_1$</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_2$</td>
<td>4.0</td>
<td>5.0</td>
<td>6.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td>0.30</td>
<td>0.38</td>
<td>0.45</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 6.9

The nominal distribution of t, and the conditional distributions f(t/Accept) and f(t/Reject) for distributions (7) to (10), test at 5% (one-sided), n = 25.

<table>
<thead>
<tr>
<th>Accept/Reject</th>
<th>A</th>
<th>R</th>
<th></th>
<th>Accept/Reject</th>
<th>A</th>
<th>R</th>
<th></th>
<th>Accept/Reject</th>
<th>A</th>
<th>R</th>
<th></th>
<th>Accept/Reject</th>
<th>A</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution</td>
<td>(7)</td>
<td>(8)</td>
<td>(9)</td>
<td>(10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.7</td>
<td>0.7</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5.0</td>
<td>6.0</td>
<td>5.0</td>
<td>5.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td>0.51</td>
<td>0.5%</td>
<td>0.60</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 6.7 for symmetric populations confirms the effectiveness of a preliminary test of normality in the sense that the conditional distribution of \( t \) for accept is closer to the nominal distribution than that for reject. The effect of increasing kurtosis, i.e. increasing the correlation between \( t \) and \( \sqrt{b_1} \) and therefore the shift, is again apparent. For distribution (3) as well as considerable positive shift for reject there is moderate negative shift for accept. The power of \( \sqrt{b_1} \) in a test of normality, though increasing with kurtosis, is severely limited for the reasons detailed earlier.

Figure 6.8, illustrating the position for populations with mild skewness, shows that the locations of the accept and reject distributions are shifted either side of the nominal distribution. The positive shift for reject is slightly greater; the preliminary test, it may be argued, is (just) effective. For this sample size the power of the preliminary test of normality is less than 50%.

The effect of the preliminary test of normality when greater skewness is present in the population is illustrated in Figure 6.9. We see that the reject distribution is better represented by the nominal than the accept. The power of the preliminary test of normality exceeds 50% in all cases. For the greatest skewness (distributions (9) and (10)) the "correction" in the conditional distribution resulting from \( \sqrt{b_1} \) being in the critical region, with probability approximately 60%, is particularly marked.

To examine the effect of increasing the sample size we consider the case when \( n = 100 \). In fact the pattern of distributions is very similar to that exhibited in Figures 6.7 to 6.9, and it will be sufficient to comment without further diagrams.
For symmetric populations ((1), (2) and (3)) the picture is virtually identical to Figure 6.7; so too is the power. For mild population skewness ((4), (5) and (6)) the left shift for accept exceeds the right shift for reject. Thus preliminary goodness-of-fit is undesirable for less severe non-normality. The power of the preliminary test exceeds 50%, and for this sample size falls slightly with kurtosis (67% for (4) to 62% for (6)). For more extreme non-normality ((7) to (10)) the preliminary test is most undesirable with the accept distribution shifted considerably to the left and the reject only very slightly overcompensated to the right. Power, which again decreases with kurtosis, exceeds 80% in all cases.

We briefly remark on the effect of performing the preliminary test at the 1% instead of 5% level, comments based on the equivalent diagrams to Figures 6.7 to 6.9 for the case n = 25. Essentially there is greater right shift for reject and less left shift for accept than for the 5% test. Only for distribution (10) is the preliminary test having an adverse effect (in our restricted sense). The price paid, of course, is loss of power which for n = 25 only reaches 44% for (10).

6.3 THE CONDITIONAL DISTRIBUTION FOR THE TWO-SIDED CASE.

Finally we consider briefly the question of a two-sided test of normality. For a test at the 100α% level we accept normality if \(|\sqrt{b_1}|\) does not exceed the upper 100α/2% point of the null distribution of \(\sqrt{b_1}\). Consider therefore \(P(t/|\sqrt{b_1|} \leq k)\). Now
\[ P(t \leq t_0/|b_1| \leq k) = \frac{P[(t \leq t_0) \cap (|b_1| \leq k)]}{P(|b_1| \leq k)}. \]

Therefore
\[ F(t/|b_1| \leq k) = \frac{F(t, k) - F(t, -k)}{F_{b_1}(k) - F_{b_1}(-k)}. \]

On transforming to normality the percentage points of the conditional distribution are obtained by solving for \( z_t \), using regular falsi,
\[ \frac{\Phi(z_t, z_\theta; \rho) - \Phi(z_t, z_\lambda; \rho)}{\Phi(z_\theta) - \Phi(z_\lambda)} = p \]

where \( z_\theta \) and \( z_\lambda \) correspond to transformed \( k \) and \( -k \) respectively.

Also, in a manner similar to that used for \( F(t/|b_1| > k) \), we obtain
\[ F(t/|b_1| > k) = \frac{F_{\theta}^{-1}(t) - [F(t, k) - F(t, -k)]}{1 - [F_{b_1}(k) - F_{b_1}(-k)]}. \]

Percentage points are obtained in the usual way by solving
\[ \frac{\Phi(z_t) - \Phi(z_t, z_\theta; \rho) + \Phi(z_t, z_\lambda; \rho)}{1 - \Phi(z_\theta) + \Phi(z_\lambda)} = p. \]

A comparison between the one- and two-sided cases is made in Figure 6.10, using as illustration distributions (3) and (10) for \( n = 25 \). As in the previous diagrams the nominal distribution of \( t \) may be compared with the accept and reject conditional distributions. The significance level for both one-sided and two-sided tests is 5%, and the power for the two-sided tests has been calculated for further comparison.
Figure 6.10

The nominal distribution of t, and the conditional distributions f(t/Accept) and f(t/Reject) for distributions (3) and (10), one- and two-sided tests at 5%, n = 25.

<table>
<thead>
<tr>
<th>Accept/Reject</th>
<th>Distribution</th>
<th>(3)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>one-sided</td>
<td>0</td>
<td>1.0</td>
</tr>
<tr>
<td>B_1</td>
<td>two-sided</td>
<td>6.5</td>
<td>5.8</td>
</tr>
<tr>
<td>Power</td>
<td>0.26</td>
<td>0.44</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.58</td>
</tr>
</tbody>
</table>
For symmetric distribution (3) we note again the relative effectiveness of the one-sided test of normality. In the two-sided case however the symmetric conditional distribution has less/greater variability corresponding to accept/reject compared with the nominal distribution, roughly in equal measure (accept barely preferred). This may be explained by thinking of the joint distribution of $t$ and $\sqrt{b_1}$, positively correlated with both marginals symmetrically distributed about zero, the tails for $\sqrt{b_1}$ being longer than for its nominal; "averaging" $t$ over the inner and outer regions of $\sqrt{b_1}$ will produce this effect. There is a marked improvement in power, though only to 44%.

By contrast we see that for distribution (10), having $\sqrt{\beta_1} = 1.0$, there is little difference between the accept and reject distributions corresponding to the one- and two-sided cases. The effect of the non-normality on the marginal distribution of $\sqrt{b_1}$ is dominated by the shift resulting from the population skewness. Clearly with increasing skewness (positive or negative) the opposing rejection region becomes irrelevant. There will naturally be some loss in power resulting from the use of the 2.5% point.

We conclude that the distinction between a one-sided and two-sided preliminary test of normality only becomes relevant when discussing symmetric populations. In this case the improvement in the representation of the distribution of the $t$ statistic by its nominal distribution achieved by the one-sided test is all but lost when using the corresponding two-sided procedure.
Chapter Seven

More Simulation, Conclusions and Extensions

7.1 SUPPLEMENTARY SIMULATION INVESTIGATION USING JOHNSON'S $S_U$

Up to this point we have assumed a Gram-Charlier population. This was chosen so that population skewness and kurtosis could be varied, and also because of the supposed simplicity of obtaining approximations to sampling moments. Choice of sample size was also more in keeping with moment expansions than simulation, though this technique was eventually found to be necessary.

Our objective now is to use simulation to extend the discussion beyond the Gram-Charlier distribution (and the constraints of its admissible $\beta_1 - \beta_2$ region), taking the opportunity to introduce a smaller sample size, and to investigate the possibility of presenting our arguments without the present software (and methodological) overheads.

The Johnson system has been used extensively to obtain approximations to marginal, joint and conditional distributions. It now serves as our population model since, as pointed out in Section 1.4, it is particularly convenient for simulation, being a transformation from the standard normal distribution. We use system $S_U$ whose admissible region, illustrated in Figure 1.3, is similar to, but extends beyond, that of the Gram-Charlier distribution. From (1.7) we may write

$$X = \xi + \lambda \sinh\left(\frac{Z - \chi}{\delta}\right).$$

Following our previous experience we propose 20 points which
typify the admissible region. Table 7.1 gives the details and includes, in the form of a data file, the values of the parameters of the corresponding $S_U$ distribution (in standard measure) obtained using AS 99.

Note that point 2, $(0.5, 4)$, lies on the lognormal boundary and has been "nudged" into the $S_U$ region by assigning values $(0.499, 4.001)$.

For a given point, and corresponding parameters, we need only generate a standard normal variate for inclusion in (7.1). We use G05DDF (Numerical Algorithms Group, 1984).

We first re-examine the crucial question of the correlation between $t$ and $\sqrt{b_1}$. We consider sample sizes 10 and 25, and estimate the correlation at the twenty points using 10000 samples. It may be noted that the correlation is always less than 0.4 and from (5.1) the maximum standard error is therefore approximately 0.008.

For each sample size the estimates were "smoothed" over the $\beta_1 - \beta_2$ plane in the manner described in Section 5.4 for a Gram-Charlier population. The selected "variables" were again $\beta_1$, $\beta_2$, and $\beta_2^2$, the resulting expressions providing a simple means for determining a finer grid of points to produce contour plots. (It was found helpful to introduce zero at $(0, 3)$ in the "data".) Figures 7.1 and 7.2 illustrate the results for sample sizes 10 and 25 respectively. Like the contour plots in Chapter 2 they were obtained with GHOST-80 (Prior, 1985).
Table 7.1
Simulation from Johnson's $S_4$, 20 points in the admissible region and corresponding values of parameters:
(a) distribution/record no., (b) $\beta_1$, (c) $\beta_2$, (d) $\alpha$, 
(e) $\delta$, (f) $\lambda$, (g) $\xi$.

<table>
<thead>
<tr>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
<th>(f)</th>
<th>(g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0000</td>
<td>4.0000</td>
<td>0.000000</td>
<td>2.321155</td>
<td>2.109381</td>
<td>0.000000</td>
</tr>
<tr>
<td>2</td>
<td>0.4990</td>
<td>4.0010</td>
<td>-5.021475</td>
<td>3.711922</td>
<td>1.710513</td>
<td>-3.201318</td>
</tr>
<tr>
<td>3</td>
<td>0.0000</td>
<td>5.0000</td>
<td>0.000000</td>
<td>1.820506</td>
<td>1.553774</td>
<td>0.000000</td>
</tr>
<tr>
<td>4</td>
<td>0.5000</td>
<td>5.0000</td>
<td>-0.895084</td>
<td>2.107190</td>
<td>1.702280</td>
<td>-0.833831</td>
</tr>
<tr>
<td>5</td>
<td>1.0000</td>
<td>5.0000</td>
<td>-3.824605</td>
<td>2.797672</td>
<td>1.224697</td>
<td>-2.394895</td>
</tr>
<tr>
<td>6</td>
<td>0.0000</td>
<td>6.0000</td>
<td>0.000000</td>
<td>1.610431</td>
<td>1.311777</td>
<td>0.000000</td>
</tr>
<tr>
<td>7</td>
<td>0.5000</td>
<td>6.0000</td>
<td>-0.526586</td>
<td>1.746755</td>
<td>1.395554</td>
<td>-0.503166</td>
</tr>
<tr>
<td>8</td>
<td>1.0000</td>
<td>6.0000</td>
<td>-1.114211</td>
<td>1.966771</td>
<td>1.451209</td>
<td>-0.986433</td>
</tr>
<tr>
<td>9</td>
<td>1.5000</td>
<td>6.0000</td>
<td>-3.286683</td>
<td>2.407185</td>
<td>1.022491</td>
<td>-2.040824</td>
</tr>
<tr>
<td>10</td>
<td>0.0000</td>
<td>7.0000</td>
<td>0.000000</td>
<td>1.489201</td>
<td>1.168771</td>
<td>0.000000</td>
</tr>
<tr>
<td>11</td>
<td>0.5000</td>
<td>7.0000</td>
<td>-0.384170</td>
<td>1.572646</td>
<td>1.223614</td>
<td>-0.369528</td>
</tr>
<tr>
<td>12</td>
<td>1.0000</td>
<td>7.0000</td>
<td>-0.697790</td>
<td>1.688982</td>
<td>1.274473</td>
<td>-0.645411</td>
</tr>
<tr>
<td>13</td>
<td>1.5000</td>
<td>7.0000</td>
<td>-1.231830</td>
<td>1.866430</td>
<td>1.279589</td>
<td>-1.047417</td>
</tr>
<tr>
<td>14</td>
<td>2.0000</td>
<td>7.0000</td>
<td>-2.983059</td>
<td>2.183448</td>
<td>0.902619</td>
<td>-1.837151</td>
</tr>
<tr>
<td>15</td>
<td>0.0000</td>
<td>8.0000</td>
<td>0.000000</td>
<td>1.408197</td>
<td>1.071602</td>
<td>0.000000</td>
</tr>
<tr>
<td>16</td>
<td>0.5000</td>
<td>8.0000</td>
<td>-0.307720</td>
<td>1.466126</td>
<td>1.111128</td>
<td>-0.296452</td>
</tr>
<tr>
<td>17</td>
<td>1.0000</td>
<td>8.0000</td>
<td>-0.521470</td>
<td>1.540887</td>
<td>1.150229</td>
<td>-0.489731</td>
</tr>
<tr>
<td>18</td>
<td>1.5000</td>
<td>8.0000</td>
<td>-0.807913</td>
<td>1.642330</td>
<td>1.178505</td>
<td>-0.726307</td>
</tr>
<tr>
<td>19</td>
<td>2.0000</td>
<td>8.0000</td>
<td>-1.306473</td>
<td>1.790613</td>
<td>1.154824</td>
<td>-1.074511</td>
</tr>
<tr>
<td>20</td>
<td>2.5000</td>
<td>8.0000</td>
<td>-2.786701</td>
<td>2.035791</td>
<td>0.820929</td>
<td>-1.702544</td>
</tr>
</tbody>
</table>
Figure 7.1

Contours of the correlation of $t$ and $\sqrt{b_1}$ for a Johnson $S_u$ population, $n = 10$.

--- indicates the lognormal boundary.

The contours labelled 1 to 6 have values 0.05 (0.05) 0.3.
Figure 7.2
Contours of the correlation of $t$ and $\sqrt{b_1}$ for a
Johnson $S_U$ population, $n = 25$.
------ indicates the lognormal boundary.
The contours labelled 1 to 8 have values 0.05 (0.05) 0.4.
We see further evidence, this time over a larger $\beta_1 - \beta_2$ region, that the correlation between $t$ and $\sqrt{b_1}$ increases with population kurtosis and decreases with population skewness. There is every reason to suppose that this is quite a general result. We may also comment that the rate of decrease of correlation with skewness is reduced in the case of extreme kurtosis.

While thinking more generally we comment on the requirement of both the Gram-Charlier distribution and Johnson's $S_U$ that skewness must be accompanied by excess kurtosis. For Johnson's system, Figure 1.2 shows that if this is not the case a finite range of support is implied (with the lognormal as boundary). Of course the assumption of normality made when using Student's $t$ distribution implies an infinite range. The evidence provided by empirical distributions is likely to be selective and, like the histograms in Pearson and Please (1975) where "practical boundaries" may be involved, inconclusive.

We now investigate directly the question of the effect of a preliminary test of normality using $\sqrt{b_1}$ on the distribution of the $t$ statistic. Samples are generated from one of the $S_U$ distributions and the following distributions are investigated:

1. The marginal (and joint) distributions of $t$ and $\sqrt{b_1}$.
2. The conditional distribution of $t$ given $\sqrt{b_1} \leq$ critical value.
3. The conditional distribution of $t$ given $\sqrt{b_1} >$ critical value.

To aid comparison an equal number of samples were used for each category. One thousand were taken, those for (1) being independent of (2) and (3), with sampling continuing for the conditional
distributions as necessary.

We illustrate the position for distributions [15], \( \beta_1 = 0 \) and \( \beta_2 = 8 \), and [20], \( \beta_1 = 2.5 \) and \( \beta_2 = 8 \), i.e. a symmetric and skew distribution having common kurtosis. Choosing sample size 10 we need to derive the percentage points of \( \sqrt{\beta_1} \) (assuming normal sampling) as these are not available in Biometrika Tables for \( n \) below 20. Using an \( s_u \) approximation (discussed in Section 5.2) we take the upper 5% point to be 0.949.

Table 7.2 gives the estimated moments of the marginal (unconditional) and conditional accept and reject distributions of the t statistic, sampling from distributions [15] and [20], and allows comparison with those of the nominal distribution, i.e. Student's t with \( v = 9 \) (\( V(t) = v/(v-2) \) and \( \beta_2^2(t) = 3 + 6/(v-4) \), Johnson and Kotz, 1970). The three empirical distribution for each case are also compared with box plots in Figure 7.3. (The statistics and plots were obtained with the statistical package MINITAB (Ryan, Joiner and Ryan, 1985); the box plot algorithm is further described in Velleman and Hoaglin (1981).)

We take together the information in Table 7.2 and Figure 7.3 and consider first symmetric distribution [15]. Allowing for sampling variation inherent in such a small scale experiment we find our previous conclusions are reinforced: The distribution of the t statistic is robust for symmetric populations, though for small samples and high kurtosis the tails are shortened (c.f. Section 5.3). The effect of the preliminary test is to introduce slight negative shift for accept (in relation to the nominal and marginal) and considerable positive shift for reject.
Table 7.2
Simulation estimates of the moments of the marginal and conditional accept and reject distributions of the $t$ statistic for Johnson $S_U$ distribution, test of normality using $\sqrt{b_1}$ at $5\%$, $n = 10$, compared with the nominal ($\text{Student's t, } \nu = 9$). (1000 samples, approximate SE's in units of 0.001.)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>$\sqrt{b_1}$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nominal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1.134</td>
<td>0</td>
<td>4.2</td>
</tr>
<tr>
<td><strong>Distribution [15] $b_1 = 0$ $b_2 = 8$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marginal</td>
<td>-0.046(35)</td>
<td>1.108</td>
<td>0.058</td>
<td>3.20</td>
</tr>
<tr>
<td>Accept</td>
<td>-0.109(36)</td>
<td>1.151</td>
<td>0.005</td>
<td>3.20</td>
</tr>
<tr>
<td>Reject</td>
<td>0.586(27)</td>
<td>0.859</td>
<td>-0.269</td>
<td>4.05</td>
</tr>
<tr>
<td><strong>Distribution [20] $b_1 = 2.5$ $b_2 = 8$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marginal</td>
<td>-0.261(38)</td>
<td>1.212</td>
<td>-0.722</td>
<td>4.14</td>
</tr>
<tr>
<td>Accept</td>
<td>-0.335(43)</td>
<td>1.374</td>
<td>-0.653</td>
<td>4.09</td>
</tr>
<tr>
<td>Reject</td>
<td>-0.043(34)</td>
<td>1.082</td>
<td>-1.296</td>
<td>7.39</td>
</tr>
</tbody>
</table>
The marginal and conditional accept and reject distributions of the t statistic for Johnson $S_U$ distribution, test of normality using $\sqrt{b_1}$ at 5%, $n = 10$. Box plots of empirical distributions (1000 samples).

Distribution [15] $\beta_1 = 0$ $\beta_2 = 8$

1 **---------------------** I + I-------------------*** * Marginal

2 * **---------------------** I + I-------------------***

Accept

3 ******---------------------** I + I----------**** **

Reject

Distribution [20] $\beta_1 = 2.5$ $\beta_2 = 8$

1 OO ******---------------------** I + I---------- * Marginal

2 * *** ******---------------------** I + I----------****

Accept

3 0 00000 ******---------------------** I + I----------**

Reject
Thus the preliminary test may be deemed effective. Thus we confirm the findings of the previous chapter, though without the consideration of the power of the test which clearly will be small.

For asymmetric distribution [20] the marginal distribution of $t$ exhibits negative shift and negative skewness, typical of small $n$ as we know from previous discussion and particularly Figure 5.1. The preliminary test of normality results in accept having a similar offset to the marginal, while reject has a very definite correction in location. Thus we have a very simple demonstration of the shift effect discussed in detail in the previous chapter.

A further feature brought out here is the skewness and particularly kurtosis evident in the "corrected" reject distribution for small $n$ and extreme non-normality. One might compare with Figure 6.9, Gram-Charlier distribution (10) for $n = 25$, where some skewness is evident. In fact extreme tail behaviour was not previously considered as it was felt that the Johnson approximation might be inaccurate.

We also take the opportunity to examine the relation between $t$ and $\sqrt{b_1}$, shown to be so crucial in understanding the effect of the preliminary test, to eliminate the possibility of the correlation coefficient masking a non-linear relation. Figure 7.4 gives scatter plots of $t$ against $\sqrt{b_1}$ for the two simulation experiments described above. It may be noted that the plot for distribution [15] is approximately bivariate normal, while for distribution [20] the skewness in the marginals is evident. Both exhibit mild (linear) correlation.
Figure 7.4

Relation between $t$ and $\sqrt{b_1}$ for Johnson $S_U$ distribution, $n = 10$, 1000 samples. (Numbers represent concurrent points at this resolution, + if over 9.)

Distribution [15] $\beta_1 = 0$ $\beta_2 = 8$

\[
\begin{array}{ccccccc}
5.5+ & - & - & - & - & - & - \\
2.5+ & - & - & * & * & 2 & *35**4*223242222** & * \\
-0.5+ & * & 25323*5*75957+9+89++787623*22 & * \\
-3.5+ & - & - & - & - & - & - \\
\end{array}
\]

Distribution [20] $\beta_1 = 2.5$ $\beta_2 = 8$

\[
\begin{array}{ccccccc}
4.0+ & - & - & - & - & - & - \\
0.0+ & * & 867948++8++++++7+75837*33* \\
-4.0+ & - & - & - & - & - & - \\
-8.0+ & - & - & - & - & - & - \\
\end{array}
\]
A few general points arise from the work in this section. The use of EDA techniques, e.g. box plots, in simulation has been advocated by Lewis and Orav (1985). It is the effectiveness of the box plot for comparing distributions that leads to the modified form being used so extensively in the main part of this study. However their use with empirical distributions having several thousand observations, typical in simulation, may give rise to problems in the tails where values falling outside certain multiples of the interquartile range are plotted individually. There is the danger that a technique designed to identify possible outliers in a small sample may mislead the eye when used with a very large sample containing no outliers by giving undue weight to the tails. In the latter case a histogram arguably gives a better representation.

Finally it may be noted that the simulation program was written in Fortran and the empirical distributions read into MINITAB. However for very small scale simulation, e.g. for illustration in teaching, a MINITAB macro is available which will generate samples from the twenty Johnson $S_U$ distributions listed in Table 7.1 (Dodgson, 1985).

7.2 CONCLUSIONS AND DISCUSSION.

We now summarise the main conclusions of our investigation and discuss their implications. Some wider issues are raised if not resolved. Some technical points arising from the work are discussed in the final section, which goes on to speculate on how the study might be extended.
Our objective was to investigate the effect of a preliminary test of normality on the subsequent use of Student's t distribution for inference about the population mean. The available advice seems to be that a preliminary assessment of normality, while not a formal part of the analysis, is a sensible precaution. The only previous investigation into this question suggested entirely the opposite by demonstrating that counter-intuitive results could occur. However there was only speculation on possible causes and these were not followed up.

In our investigation we have used the skewness statistic to test for normality, as non-normality characterised by skewness most affects the distribution of the t statistic, and assumed a Gram-Charlier population to allow the introduction of non-normality in the shape of skewness and kurtosis. The essence of the problem is the conditional distribution of t given $\nu b_1$. We have proceeded to an approximation of both the array distribution and the conditional distribution given that $\nu b_1$ lies within/without the rejection region, and thus been able to demonstrate graphically the effect of the preliminary test on the distribution of t.

It is now clear that the effect of the preliminary test may be explained in terms of the correlation between t and $\nu b_1$ (for the particular population in question) and the observed value of $\nu b_1$ relative to its marginal distribution. What we are witnessing is essentially a first order phenomenon. This has also been demonstrated, in a more simple way, in the previous section.

We have found that for a symmetric population the conditional distribution of t is centred on zero for $\nu b_1 = 0$ and shifts to the
right with increasing $\sqrt{b_1}$. For asymmetric populations the location of the marginal distribution of $\sqrt{b_1}$ reflects the population skewness. Small values of $\sqrt{b_1}$, being in the lower tail, result in the conditional distribution of $t$ being shifted to the left, while increasing $\sqrt{b_1}$ up to the population value $\sqrt{b_1}^*$ results in a correction in the location of the conditional distribution. We have also examined directly the distribution of $t$ conditional on $\sqrt{b_1}$ passing and failing the preliminary test. Our doubt about the value of a preliminary test includes the case of a symmetric population if a two-sided test is used.

With a little asymptotic licence we may present some of the findings of our study in a simple manner. The joint distribution of $t$ and $\sqrt{b_1}$ may roughly be represented by a bivariate normal distribution. Using standard results we may therefore write

$$E(t/\sqrt{b_1}) \propto \mu_c + \rho_{t/\sqrt{b_1}} \frac{\sigma_t}{\sigma_{\sqrt{b_1}}} (\sqrt{b_1} - \mu_{\sqrt{b_1}}) \quad (7.2)$$

$$V(t/\sqrt{b_1}) \propto \sigma_t^2 (1 - \rho^2_{t/\sqrt{b_1}}). \quad (7.3)$$

Consider (7.2) for the symmetric case. We have $\mu_c = \mu_{\sqrt{b_1}} = 0$ and it is clear, because of the positive correlation, that the conditional mean moves to the right as $\sqrt{b_1}$ increases from zero. In the asymmetric case the population skewness (assumed positive) results in $\mu_c$ becoming slightly negative while $\mu_{\sqrt{b_1}}$ increases to reflect the population skewness. It is clear that a small value of $\sqrt{b_1}$ will lead to a negative shift in $E(t/\sqrt{b_1})$, and that this will be corrected for $\sqrt{b_1}$ near (possibly just greater than) the population skewness.
This approximate formulation also allows us to comment on
the effect of some other factors. We have seen that the correlation
between \( t \) and \( \sqrt{b_1} \) increases with kurtosis and decreases with
skewness. For a symmetric population, where of course the marginal
and nominal distributions of \( t \) are very close, the "effectiveness"
of the preliminary test results from the kurtosis in the population,
by providing the correlation and hence the shift with large \( \sqrt{b_1} \).
Of course the effect is not entirely first order. Under these
circumstances we see from (7.3) that the variability in the
conditional distribution will decrease, as previously observed in
Figure 6.5. However we are principally concerned with asymmetric
populations, since these result in the distortion of the distribution
of \( t \), and we have detailed the unforeseen results that can occur.
Examining (7.2) we see that although (for fixed \( \sqrt{b_1} \)) the shift
increases with \( \sqrt{b_1} \) and hence \( \mu_{\sqrt{b_1}} \), the effect is lessened by the
corresponding decrease in the correlation.

We are also able to comment on the effect of sample size.
Increasing \( n \), while having little effect on the correlation, will
result in an increase in the standard deviation ratio in (7.2)
and thereby in the shift effect. At the same time the marginal
distribution of \( t \) will be tending to its nominal (and both to
normality).

This leads us to comment on the relevance of the robustness
of \( t \), which while not the focus of the study has provided the
option of "do not test", to help put our "accept" and "reject"
distributions in context. The fact that as \( n \) increases the power
to detect non-normality increases while the sensitivity of \( t \) to
that non-normality decreases is an argument in itself against using a preliminary test for moderate to large samples, further strengthened by the comments on the shift effect made above. The possibility of finding a critical value that is effective when used in a preliminary test (based on robustness considerations rather than an arbitrary significance level) has been raised by Please (personal communication). Our work suggests that such a value cannot be found.

Overall we are lead to only one conclusion, echoing Cox (Section 1.2), though we are in no doubt: a preliminary test of normality should be avoided.

This advice may be unremarkable to those whose approach to statistics rarely involves such a test. But we have reached this conclusion working from the viewpoint of classical sampling theory where the value of a preliminary test might appear self-evident.

It should be added that we are discussing the analysis of a unique sample. If we are concerned with the routine analysis of a series of similar samples we would be in a better position to model the variation. Probability plotting and goodness-of-fit would be helpful in choosing a distribution (or transformation), and appropriate inferential methods would follow. With a unique sample it may be argued that an assumption of normality (or any other distribution) made so that standard methods of inference may be used, is just that - taken to be true but not verifiable. Statistical inference may then be seen as a generalisation of the familiar IF-THEN kind: IF common and reasonable a priori, THEN a useful step forward while not over sensitive to the requirements of IF.
Another view is that the results of any methods used to establish the model are "invisible" to the subsequent inference based on that model. This seems implicit in much recent discussion advocating the integration of exploratory methods, used in the initial stages of an analysis, with more formal probabilistic methods whose precise specification depends on the preliminary stage (e.g. Chatfield, 1985). To insist that inference should be carried out conditionally on the value of any statistic(s) used in the specification of the model seems both impractical and counter to the reductionist aims of model fitting. In particular there is no limit to the number of statistics that could be used, whose mutual dependence would have to be accounted for in terms of the finally accepted rather than "true" model.

In fact one distinction between statistical modelling and statistical inference is that in the former there is generally no concern about problems of selection or conditionality, just a search for a parsimonious model. Statistical inference on the other hand, at least as understood in terms of Neyman-Pearson theory, takes us from assumptions to conclusion in a series of steps resembling the logic of predicate calculus. Indeed the attraction of such definite reasoning applied to real world problems with all their attendant uncertainty has lead some to believe that statistical inference provides the foundation for scientific inference. Guttman (1985) provides an interesting commentary.

In fact statistics is more restrictive in scope, and more immediately useful, providing a basis for decision making when
practical considerations require one, as in quality control, and more generally a means of assessing evidence in the course of scientific enquiry. However standard methods support, and themselves are supported by, the belief that there is to be a single analysis of a unique set of data; beyond that they are not directly helpful and can even be contradictory.

While these matters can and have been discussed at length there is now a more pressing need to focus our thoughts on what might be termed the theory of statistical practice. The reason is the ubiquitous computer. Nelder (1986) has written that for the next stage in statistical computing - the expert system - "we are now faced with the challenge of formalizing what we do". While we might debate both the "we" and "do" (Wynn, 1980; Boyd, 1980) the relevant point concerns the formalization: given agreed optimum methods how do we devise a strategy for analysis?

We have investigated the relatively simple problem of inference about a population mean given that the use of Student's t distribution is optimal for a normal sample. However the pretesting of data for normality has been shown to be an inappropriate strategy. This suggests that there can be no simplistic approach to the formalization of statistics, with successful developments most likely in relatively circumscribed fields having agreed goals and methods.

So where are we left? One can hardly deny the advantages of making population assumptions in both simplifying and enriching an analysis. But in practice the search for a model is not all-embracing, beginning (and usually ending) with the simplest for which standard results are available. Preliminary goodness-of-fit, which we have
shown to be ineffective, is arguably not the best use of the sample information beyond that sufficient for the intended model.

No simple way forward arises from this work, though there are some pointers which are expanded on in the final section. Yet in understanding the effect of a preliminary test of normality on Student's t distribution we have a new perspective which may influence our approach to statistical problems in general.

7.3 LOOKING BACK AND FORWARD.

Looking back our first comment concerns the use of k-statistics. These were used, in the first instance, to obtain approximations to the moments of \( \nu b \_1 \) and its correlation with \( t \), approximations which were subsequently found to be inadequate. We conclude that at present k-statistics are really only appropriate for normal theory, though this might change if the algebra were to be computerised.

The Gram-Charlier distribution was found to be a convenient population for the introduction of skewness and kurtosis into the normal, its advantage of having vanishing high order cumulants not fully realised because of the inadequacy of the k-statistics approximations. The characterisation of non-normality with a grid of points in the \( \beta_1 - \beta_2 \) plane, while not new, allowed more detailed analysis than is found in many robustness studies which typically use an arbitrary collection of standard distributions.

The Johnson system gave satisfactory approximations which were easy to fit and apply using the associated algorithms. The corresponding bivariate system, while conceptual difficulties remain (see below), provided the route to the conditional distributions
on which our investigation hinged. The representation of related distributions with box plots proved valuable both for interpretation and communication.

The question of the correlation when fitting the Johnson bivariate system is the one technical matter of sufficient general interest to justify further work. The standard approach of using the correlation of the marginally transformed data gives rise to difficulties with the assumption that bivariate normality results, if this correlation is less than in the original data. It is also not applicable if, as in our study, only moments are available. Instead we simply used the original correlation, which on present evidence gives results at least good as the standard method while avoiding any conflict with theory.

Another approach, which we explored briefly, used a Taylor expansion to derive a relation between the correlation before and after transformation (equation (5.4)). This appears promising, with the correlation in the example showing the appropriate increase after transformation; but the discrepancy with the standard method (based on the same transformations) only exemplifies the problem.

Ideally we require a direct transformation to bivariate normality, and we cannot expect any method based on marginal transformations to univariate normality to be the best way of achieving this. There has been some discussion on this point in the literature, e.g. Kendall and Stuart (1979) on the canonical analysis of contingency tables, and Gnanadesikan (1977) on Box-Cox transformations.

Multivariate theory, and particularly that of multivariate
transformations, is certainly relevant though the choice of non-normal models is rather limited. At a more practical level there is much to be learned about the relationship between the joint distribution of marginally transformed variables and the bivariate normal distribution. Simulation studies based on the transformation of bivariate normal samples might be a first step.

Another, and more speculative, line of enquiry concerns the use of normal scores to construct univariate and by extension bivariate pseudo-samples (Andrews et al., 1972). Some preliminary (though inconclusive) work has been carried out on the effect on correlation of transforming a pseudo-sample. The value of replacing data by normal scores in the likelihood equation for $\rho$ assuming bivariate normal sampling (prompted by the EM algorithm, Dempster et al., 1977) might also be explored.

Returning to our main enquiry an obvious extension would be to investigate the use of other test statistics, such as sample kurtosis $b_2$ and the omnibus statistics $A^2$ and $W$ (referred to in Chapter 1). The moment statistics $\sqrt{b_1}$ and $b_2$, separately or combined (D'Agostino and Pearson, 1973; correction 1974), seem to offer the most promise given the relevance of both population skewness and kurtosis to the effect of a preliminary test. We might speculate that (1) the test statistics will be highly correlated for non-normal sampling, and (2) similar undesirable features of preliminary goodness-of-fit will be observed.

Investigating the effect of pretesting on other normal theory statistics such as Student's $t$ for the two-sample problem, chi-square for variance and $F$ for the variance ratio would allow
further generalisation, particular into the relationship between the preliminary test and robustness properties. Other examples could be cited, almost without limit if one widens the scope of the preliminary test.

The characterisation of non-normality could also include that of a normal population contaminated with outliers; the normal mixture provides one such model. Possibly the concept of a sensitivity curve or surface (see e.g. Prescott, 1976) could be adapted to investigate the effect of one or two outliers on a preliminary test.

Another line of enquiry is suggested by the key importance we have demonstrated of the correlation between the preliminary test statistic and the following inferential procedure: the two stages of the analysis could be rendered independent by sample splitting. Using part of the sample for model selection and the remainder for estimation may be seen as an extension to current ideas on sample re-use (Efron, 1982). This may be of value for moderate to large samples, and would not be difficult to investigate.

We end by drawing attention to the need for a more effective integration of the preliminary and so-called confirmatory parts of an analysis. At the theoretical level progress would be achieved by providing statistics like $\sqrt{b_1}$ with a more effective ancillary role. This would entail a rich population model incorporating the normal as a special case, like that of Box and Tiao (1973) and indeed the Gram-Charlier. However the problem of a preliminary test is likely to be subsumed under the wider issues which are regularly rehearsed in any debate on the relative
merits of the various schools of statistical inference (Barnett, 1982).

The typical practitioner has more pressing problems of inference, and would welcome most advice on how best to apply standard methods. And while this has not been our objective, we do hope to have made a contribution.
References


PEARSON, E.S. and ADYANTHAYA, N.K. (1929). The distribution of frequency constants in small samples from non-normal symmetrical and skew populations, (Part II). Biometrika, 21, 259-286.


Appendix I

Auxiliary Formulae

AI.1 MOMENTS IN TERMS OF CUMULANTS.

Formulae are quoted from Cook (1951), or the more comprehensive Kratky et al. (1972), which are modified, and expressed generally to $O(n^{-3})$, for substitution into the stated formulae. Having illustrated the method, the typing of extremely long expressions is avoided by referring to the page number in Kratky et al. thus KRHS p.N.

$E(g)$ - equation (2.1).

\[ \mu''_{11} = \kappa_{11} + \kappa_{10} \kappa_{01} \quad (\mu'' \equiv \mu' \text{ in KRHS}) \]

\[ \mu''_{11} = \kappa_{11} \]

\[ \mu_{c}(32) = \kappa(32) \]

\[ \mu''_{21} = \kappa_{21} + \kappa_{20} \kappa_{01} + 2\kappa_{11} \kappa_{10} + \kappa_{10} \kappa_{01} \]

\[ \mu''_{12} = \kappa_{12} + \kappa_{02} \kappa_{10} + 2\kappa_{11} \kappa_{01} + \kappa_{01} \kappa_{10} \]

\[ \mu''_{12} = \kappa_{12} + \kappa_{02} \kappa_{10} \]

\[ \mu_{C}(32^2) = \kappa(32^2) + \kappa(2^2) \kappa_3 \]

\[ \mu''_{31} = \kappa_{31} + \kappa_{30} \kappa_{01} + 3\kappa_{21} \kappa_{10} + 3\kappa_{20} \kappa_{11} + 3\kappa_{20} \kappa_{10} \kappa_{01} + 3\kappa_{11} \kappa_{10}^2 + \kappa_{10}^2 \kappa_{01} \]

\[ \mu_{c}(32^3) = \kappa(32^3) + \kappa(2^3) \kappa_3 + 3\kappa(2^2) \kappa(32) \]
\[
\mu_{41}^n = \kappa_{41} + \kappa_{40} \kappa_{01} + 4\kappa_{31} \kappa_{10} + 4\kappa_{30} \kappa_{11} + 4\kappa_{30} \kappa_{10} \kappa_{01} + 6\kappa_{21} \kappa_{20} \\
+ 6\kappa_{21} \kappa_{10}^2 + 3\kappa_{20} \kappa_{01} + 12\kappa_{20} \kappa_{11} \kappa_{10} + 6\kappa_{20} \kappa_{10} \kappa_{01} \\
+ 4\kappa_{11} \kappa_{10}^3 + \kappa_{10}^4 \kappa_{01}
\]

\[
\mu^c(32^4) = \kappa(2^3) \kappa_{3} + 4\kappa(2^3) \kappa(32) + 6\kappa(32^2) \kappa(2^2) + 3\kappa^2(2^2) \kappa_{3}
\]

\[
\mu_{51}^n = \kappa_{51} + \kappa_{50} \kappa_{01} + 5\kappa_{41} \kappa_{10} + 5\kappa_{40} \kappa_{11} + 5\kappa_{40} \kappa_{10} \kappa_{01} + 10\kappa_{31} \kappa_{20} \\
+ 10\kappa_{31} \kappa_{10}^2 + 10\kappa_{30} \kappa_{21} + 10\kappa_{30} \kappa_{20} \kappa_{01} \\
+ 20\kappa_{30} \kappa_{11} \kappa_{10} + 10\kappa_{30} \kappa_{10} \kappa_{01} + 30\kappa_{21} \kappa_{20} \kappa_{10} + 10\kappa_{21} \kappa_{10}^3 \\
+ 15\kappa_{20} \kappa_{11} + 15\kappa_{20} \kappa_{10} \kappa_{01} + 30\kappa_{20} \kappa_{11} \kappa_{10} + 10\kappa_{20} \kappa_{10}^3 \kappa_{01} \\
+ 15\kappa_{11} \kappa_{10}^4 + \kappa_{10}^5 \kappa_{01}
\]

\[
\mu^c(32^5) = 10\kappa(2^3) \kappa(2^2) \kappa_{3} + 15\kappa^2(2^2) \kappa(32)
\]

\[
\mu_{61}^n = \ldots \text{ KRHS p.} \, 15
\]

\[
\mu^c(32^6) = 15\kappa^3(2^2) \kappa_{3}
\]

\(\mathbf{E}(g^2) - \text{equation (2.2).}\)

\[
\mu_2^l = \kappa_2 + \kappa_1^2
\]

\[
\mu^l(3^2) = \kappa(3^2) + \kappa_3^2
\]

\[
\mu_{21}^n = \kappa_{21} + 2\kappa_{11} \kappa_{10} + \kappa_{10}^2 \kappa_{01}
\]

\[
\mu_{21}^c = \kappa_{21} + 2\kappa_{11} \kappa_{10}
\]

\[
\mu^c(3^22) = \kappa(3^22) + 2\kappa(32) \kappa_{3}
\]
\[ \mu''_{22} = \kappa_{22} + 2\kappa_{12}\kappa_{10} + 2\kappa_{21}\kappa_{01} + 2\kappa_{11} + \kappa_{20}\kappa_{02} + \kappa_{02}\kappa_{10} \]
\[ + 4\kappa_{11}\kappa_{10}\kappa_{01} + \kappa_{20}\kappa_{01} + \kappa_{02}\kappa_{10} \]
\[ u^c(3^22^2) = \kappa(3^22^2) + 2\kappa(32^2)\kappa_3 + 2\kappa^2(32) + \kappa(3^2)\kappa(2^2) + \kappa(2^2) + \kappa_3^2 \]
\[ u''_{32} = \ldots \text{KRHS p.3} \]
\[ \mu''_{1c}(3^22^3) = 2\kappa(32^3)\kappa_3 + 3\kappa(3^22)\kappa(2^2) + 6\kappa(32^2)\kappa(32) + \kappa(2^3)\kappa(3^2) \]
\[ + \kappa(2^3)\kappa_3^2 + 6\kappa(2^2)\kappa(32)\kappa_3 \]
\[ \mu''_{42} = \ldots \text{KRHS p.5} \]
\[ \mu''_{1c}(3^22^4) = \kappa(2^4)\kappa_3^2 + 12\kappa(32^2)\kappa(2^2)\kappa_3 + 8\kappa(2^3)\kappa(32)\kappa_3 + 12\kappa(2^2)\kappa^2(32) \]
\[ + 3\kappa^2(2^2)\kappa(3^2) + 3\kappa^2(2^2)\kappa_3^2 \]
\[ \mu''_{52} = \ldots \text{KRHS p.15} \]
\[ \mu''_{1c}(3^22^5) = 10\kappa(2^3)\kappa(2^2)\kappa_3^2 + 30\kappa^2(2^2)\kappa(32)\kappa_3 \]
\[ \mu''_{62} = \ldots \text{KRHS p.42} \]
\[ \mu''_{1c}(3^22^6) = 15\kappa^3(2^2)\kappa_3^2 \]

\[ \text{E}(g^3) \text{ - equation (2.3).} \]
\[ \mu'_{3} = \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3 \]
\[ \mu'(3^3) = \kappa(3^3) + 3\kappa(3^2)\kappa_3 + \kappa_3^3 \]
\[ \mu''_{31} = \kappa_{31} + 3\kappa_{21}\kappa_{10} + \kappa_{30}\kappa_{01} + 3\kappa_{20}\kappa_{11} + 3\kappa_{11}\kappa_{10} + 3\kappa_{20}\kappa_{10}\kappa_{01} + \kappa_{30}\kappa_{01} \]

\[ \mu''_{32} = \ldots \text{ KRHS p.3} \]

\[ \mu''_{32}^c(3^2) = \kappa(3^2) + 3\kappa(3^2)\kappa_3 + 3\kappa(3^2)\kappa(32) + 3\kappa(32)\kappa_3^2 \]

\[ \mu''_{33} = \ldots \text{ KRHS p.5} \]

\[ \mu''_{33}^c(3^2) = 3\kappa(3^2)\kappa_3 + 3\kappa(3^2)\kappa(32) + 6\kappa(3^2)\kappa(32) + \kappa(3^2)\kappa(2^2) \]

\[ + 3\kappa(3^2)\kappa_3^2 + 6\kappa(3^2)\kappa_3 + 3\kappa(3^2)\kappa(2^2)\kappa_3 + \kappa(2^2)\kappa_3^3 \]

\[ \mu''_{43} = \ldots \text{ KRHS p.16} \]

\[ \mu''_{43}^c(3^2) = \kappa(3^2)\kappa_3^3 + 18\kappa(3^2)\kappa(32)\kappa_3 + 9\kappa(3^2)\kappa(32)\kappa_3^2 \]

\[ + 9\kappa(3^2)\kappa(2^2)\kappa_3 + \kappa(2^2)\kappa_3^3 + 6\kappa(3^2) + 9\kappa(3^2)\kappa(32)\kappa(2^2) \]

\[ + 9\kappa(32)\kappa(2^2)\kappa_3^2 \]

\[ \mu''_{53} = \ldots \text{ KRHS pp 42-43} \]

\[ \mu''_{53}^c(3^2) = 10\kappa(3^2)\kappa(2^2)\kappa_3 + 45\kappa(2^2)\kappa(32)\kappa_3 \]

\[ \mu''_{63} = \ldots \text{ KRHS pp 50-51} \]

\[ \mu''_{63}^c(3^2) = 15\kappa(2^2)\kappa_3^3 \]
(2.3) subsequently expressed to $O(n^{-2})$.

$E(g) - \text{equation (2.4)}.$

The modified expressions are to $O(n^{-2})$.

\[ \mu_4' = \kappa_4 + 4\kappa_3 \kappa_1 + 3\kappa_2^2 + 6\kappa_2 \kappa_1^2 + \kappa_1^4 \]

\[ \mu_4'' = ... \text{KRHS p.3} \]

\[ \mu_4^c(3^42) = 6\kappa(3^22)\kappa_3^2 + 12\kappa(3^2)\kappa(32)\kappa_3 + 4\kappa(32)\kappa_3^3 \]

\[ \mu_4'' = ... \text{KRHS p.5} \]

\[ \mu_4^c(3^42^2) = 4\kappa(3^2)\kappa_3^3 + 12\kappa^2(32)\kappa_3^2 + 6\kappa(3^2)\kappa(2^2)\kappa_3^2 + \kappa(2^2)\kappa_3^4 \]

\[ \mu_4'' = ... \text{KRHS p.16} \]

\[ \mu_4^c(3^42^3) = \kappa(3^2)\kappa_3^4 + 12\kappa(32)\kappa(2^2)\kappa_3^3 \]

\[ \mu_4'' = ... \text{KRHS p.44} \]

\[ \mu_4^c(3^42^4) = 3\kappa^2(2^2)\kappa_3^4 \]

$E(tg) - \text{equation (2.5)}.$

\[ \mu_1'' = \kappa_1 + \kappa_{10}^2 \]

\[ \mu''(13) = \kappa(13) + \kappa_1^2 \kappa_3 \]
\[ \mu_{111}^{'''} = \kappa_{111} + \kappa_{011}\kappa_{100} + \kappa_{110}\kappa_{001} + \kappa_{101}\kappa_{010} + \kappa_{100}\kappa_{001}\kappa_{010} \]

\[ \mu^{'''}_{211} = \ldots \text{ KRHS p.2} \]

\[ \mu^{'''}_{311} = \ldots \text{ KRHS p.3} \]

\[ \mu^{'''}_{411} = \ldots \text{ KRHS p.6} \]

\[ \mu^{'''}_{511} = \ldots \text{ KRHS p.17} \]

\[ \mu^{'''}_{511} = 10\kappa_{213}\kappa_{113} \kappa_{221} + 15\kappa_{213}\kappa_{211}\kappa_{113} + 15\kappa_{213}\kappa_{211}\kappa_{221} \]

\[ (2.5) \text{ subsequently expressed to } O(n^{-2}). \]
AI.2 CUMULANT EXPRESSIONS IN TERMS OF POPULATION CUMULANTS.

The following formulae, taken from Kendall and Stuart (1977, Section 12.16) and simplified for the Gram-Charlier distribution by setting $\kappa_r = 0$ for $r \geq 5$, are sufficient to express equations (2.1) to (2.5) to their stated order. The last four, which are given with the higher cumulants, follow from "Rule 10" (loc. cit. p.305).

\[
\kappa(2^2) = \frac{\kappa_4}{n} + \frac{2\kappa_2^2}{n-1}
\]

\[
\kappa(2^3) = \frac{12\kappa_4\kappa_2}{n(n-1)} + \frac{4(n-2)}{n(n-1)^2} \kappa_3^2 + \frac{8}{(n-1)^2} \kappa_2^3
\]

\[
\kappa(2^4) = \frac{8(4n^2 - 9n + 6)}{n^2(n-1)^3} \kappa_4^2 + \frac{144}{n(n-1)^2} \kappa_4 \kappa_2^2 \\
+ \frac{96(n-2)}{n(n-1)^3} \kappa_3^2 \kappa_2 + \frac{48}{(n-1)^3} \kappa_2^4
\]

\[
\kappa(3^2) = \frac{9}{n-1} \kappa_4 \kappa_2 + \frac{9}{n-1} \kappa_3^2 + \frac{6n}{(n-1)(n-2)} \kappa_2^3
\]

\[
\kappa(3^3) = \frac{162(5n-12)}{(n-1)^2(n-2)} \kappa_4 \kappa_3 \kappa_2 + \frac{36(7n^2 - 30n + 34)}{(n-1)^2(n-2)^2} \kappa_3^3 \\
+ \frac{108n(5n-12)}{(n-1)^2(n-2)^2} \kappa_3 \kappa_2^3
\]

\[
\kappa(32) = \frac{6}{n-1} \kappa_3 \kappa_2
\]

\[
\kappa(32^2) = \frac{12(2n-3)}{n(n-1)} \kappa_4 \kappa_3 + \frac{48}{(n-1)^2} \kappa_2 \kappa_3^2
\]
\[ \kappa(3^2_2) = \frac{9(3n-5)}{n(n-1)^2} \kappa_4^2 + \frac{18(6n-11)}{(n-1)^2(n-2)} \kappa_4 \kappa_2^2 \]
\[ + \frac{18(9n-20)}{(n-1)^2(n-2)} \kappa_3 \kappa_2 + \frac{36n}{(n-1)^2(n-2)} \kappa_2^4 \]
\[ \kappa(3^2_2) = \frac{360(2n-3)}{n(n-1)^3} \kappa_3 \kappa_2^3 + \frac{24(5n-12)}{n(n-1)^3} \kappa_3^3 \]
\[ + \frac{480}{(n-1)^3} \kappa_3 \kappa_2^3 \]
\[ \kappa(3^2_2) = \frac{36(29n^2 - 103n + 93)}{n(n-1)^3(n-2)} \kappa_4 \kappa_2^2 + \frac{36(38n^2 - 155n + 160)}{n(n-1)^3(n-2)} \kappa_4 \kappa_3 \]
\[ + \frac{72(14n - 23)}{(n-1)^3(n-2)} \kappa_4 \kappa_2^3 + \frac{144(19n - 44)}{(n-1)^3(n-2)} \kappa_3 \kappa_2^2 \]
\[ + \frac{288n}{(n-1)^3(n-2)} \kappa_2^5 \]
\[ \kappa(21) = \kappa_3 \frac{n}{n} \]
\[ \kappa(31) = \kappa_4 \frac{n}{n} \]
\[ \kappa(2^2_1) = \kappa_5 \frac{5}{n^2} + \frac{4 \kappa_3 \kappa_2}{n(n-1)} \]
\[ \kappa(3^2_1) = \kappa_6 \frac{6}{n^2} + \frac{6 \kappa_4 \kappa_2}{n(n-1)} + \frac{6 \kappa_3^2}{n(n-1)} \]
Appendix II

The Gram-Charlier Package GCPACK

FILE: GCPACK.FOR

DESC: PACKAGE OF ALGORITHMS FOR THE GRAM-CHARLIER DISTRIBUTION

P(X) = PHI(X)((1+(C3/6)*H3(X)+(C4/24)*H4(X))

CEXT: NAG: G05CAF (OR ANY U(0,1) GENERATOR)

; ALNORM (AS66) (OR ANY N(0,1) DF)

; PPND (AS111) (OR ANY N(0,1) PP)

AUTH: J H DODGSON, DEPT OF MATHS, STATS AND OR,

TRENT POLYTECHNIC, NOTTINGHAM, UK

HIST: >07-82, 06-84 MODS, 12-86 COS MODS

SUBROUTINE INRVGC(C3, C4, H, A, THL, SIGL, B, THU, SIGU, IFAULT)

DESC: INITIALISES H TO SIGU FOR RVGC

INP: C3, C4

OUTP: H...IFault

CINT: ITYPGC, XMODGC, FGC, PPGC

CEXT: (ALNORM), (PPND)

HIST: >07-82, 06-84 COS MODS

REAL C3, C4, XMODE, H, A, FACL, THL, SIGL, B, QB, FACU, THU, SIGU,

1 0, 0, 1, 0, 2, 0, 0.005, 0.995/

H=ZERO
A=ZERO
THL=ZERO
SIGL=ZERO
B=ZERO
THU=ZERO
SIGU=ZERO

CHECK CUMULANTS

IFault = ITYPGC(C3, C4)

IF(IFault .NE. 0)RETURN ! C3, C4 OUTSIDE UNIMODAL REGION

DETERMINE H

IFault = 4
XMODE = XMODGC(C3, C4, IFXMOD)

IF(IFXMOD .NE. 0)RETURN ! NO SOLUTION

H = FGC(C3, C4, XMODE)

DETERMINE A, THL AND SIGL

IFault = 5
A = PPGC(C3, C4, PA, IFPPGC)

IF(IFPPGC .NE. 0)RETURN ! NO SOLUTION

FACL = SQRT(-ALOG(PA))
SIGL = -TWO*PA*FACL/FGC(C3, C4, A)
THL = A-SIGL*FACL

DETERMINE B, THU AND SIGU

B = PPGC(C3, C4, PB, IFPPGC)

IF(IFPPGC .NE. 0)RETURN ! NO SOLUTION

QB = ONE-PB
FACU = SQRT(-ALOG(QB))
SIGU = TWO*QB*FACU/FGC(C3, C4, B)

THU = B-SIGU*FACU

IFault = 0
REAL FUNCTION RVGC(C3,C4,H,A,THL,SIGL,B,THU,SIGU)

DESC: GENERATES (UNIMODAL) GRAM-CHARLIER RANDOM VARIABLE

INP: C3, C4; H ... SIGU (FROM INRVGC)

OUTP: RVGC

CINT: FGC

CEXT: NAG: G05CAF

HIST: >07-82, 06-84 MODS

REAL C3, C4, H, A, THL, SIGL, R, THU, SIGU, BMA, U, DUM,
1 ZERO, ONE, PA, PB
DATA ZERO, ONE, PA, PB/
1 0.0, 1.0, 0.005, 0.995/
U=G05CAF(DUM)
IF(U.LT.PA)GO TO 2
IF(U.GT.PB)GO TO 3

CENTRE

BMA=B-A
1 RVGC=A+BMA*G05CAF(DUM)
IF(H*G05CAF(DUM).LE.FGC(C3, C4, RVGC))RETURN GO TO 1

LOWER TAIL
2 RVGC=THL+SIGL*SQR(-ALOG(U))
RETURN

UPPER TAIL
3 RVGC=THU+SIGU*SQR(-ALOG(ONE-U))
RETURN

END

REAL FUNCTION PPGC(C3,C4,P,IFAULT)

DESC: DETERMINES QUANTILE (PERCENTAGE POINT) OF GRAM-CHARLIER DISTRIBUTION

INP: C3, C4, P

OUTP: PPGC, IFAULT

CINT: FGC, DPGC

CEXT: (ALNORM), PPND

HIST: >07-82, 06-84 MODS

REAL C3, C4, X0, X1, X2, DFMP0, DFMP1, DFMP2, DIV, CORR, D1STEP,
1 ZERO, ONE, THREE, SIX, SMALL1, SMALL2, PPACC, CORLIM, STEP, SXLIM
1 ITLIM1, ITLIM2/
2 0.0, 1.0, 3.0, 6.0, 1.0E-6, 1.0E-12, 1.0E-5, 6.0, 0.1, 6.0, 8, 20

DATA ZERO, ONE, THREE, SIX, SMALL1, SMALL2, PPACC, CORLIM, STEP, SXLIM,
1 ITLIM1, ITLIM2/
2 0.0, 1.0, 3.0, 6.0, 1.0E-6, 1.0E-12, 1.0E-5, 6.0, 0.1, 6.0, 8, 20

IF(IFAULT=1)
PPGC=ZERO
IF((P.LE.ZERO).OR.(P.GE.ONE))RETURN ! P OUTSIDE RANGE
IF(IFAULT=2)
PPGC=PPND(P, IPPPND)
IF((P.LT.SMALL1).OR.(P.GT.ONE-SMALL1))RETURN ! NORMAL APPROX
IF(IFAULT=3)
NEWTON-RAPHSON
X0=PPGC
IT=1
DIV=FGC(C3,C4,X0)
IF(ABS(DIV).LT.SMALL2)GO TO 4
CORR=(DFGC(C3,C4,XO)-P)/DIV
IF(ABS(CORR).GT.CORLIM)GO TO 4
X1=X0-CORR
IF(ABS(CORR).LT.PPACC)GO TO 2
IT=IT+1
IF(IT.GT.ITLIM1)GO TO 3
X0=X1
GO TO 1
C N-R SOLUTION
2 IFAULT=0
PPGC=X1
RETURN
C CHECK IF LAST N-R X0,X1 OK FOR REGULAR FALSI
3 DFMPO=DFGC(C3,C4,X0)-P
DFMP1=DFGC(C3,C4,X1)-P
IF((DFMPO*DFMP1.LT.ZERO).AND.(ABS(CORR).LT.STEP))GO TO 6
C NEW X0,X1 FOR REGULAR FALSI
4 X0=PPGC
DFMPO=DFGC(C3,C4,X0)-P
DISTEP=SIGN(STEP,DFMPO)
5 X1=X0-DISTEP
IF(ABS(X1).GT.SXLIM)RETURN ! NO SOLUTION
DFMP1=DFGC(C3,C4,X1)-P
DFMPO=DFMP1*DFMP1.LT.ZERO)GO TO 6
X0=X1
DFMPO=DFMP1
GO TO 5
C REGULAR FALSI
6 IT=1
7 DIV=DFMP1-DFMPO
IF(ABS(DIV).LT.SMALL2)RETURN ! NO SOLUTION
CORR=(X1-X0)*DFMP1/DIV
IF(ABS(CORR).GT.CORLIM)RETURN ! NO SOLUTION
X2=X1-CORR
IF(ABS(CORR).LT.PPACC)GO TO 9
IT=IT+1
IF(IT.GT.ITLIM2)RETURN ! NO SOLUTION
DFMP2=DFGC(C3,C4,X2)-P
IF(DFMPO*DFMP2.LT.ZERO)GO TO 8
X0=X1
DFMPO=DFMP1
8 X1=X2
DFMP1=DFMP2
GO TO 7
C R F SOLUTION
9 IFAULT=0
PPGC=X2
RETURN
END

FUNCTION ITYPGC(C3,C4)
C DESC: TESTS C3,C4 FOR POSITIVE-DEFINITE AND UNIMODAL
C GRAM-CHARLIER DISTRIBUTION
C INP : C3,C4
C OUTP:ITYPGC
C CINT:DETLU (OR ANY DETERMINANT ALGORITHM)
C HIST: 07-82, 06-84 COS MODS
C
REAL A(11,11),C3,C4,B1,A3,A4,SH1L,SH1R,SH2L,SH2R,
ZERO,ONE,TWO, TWOPFO,THREE,FOUR, SIX,EIGHT, TEN, FIFTN, TWTYFO,
1
FOTYFI,CONU1, CONU2, CONU3, CONC3, CONC4, SMALL
DATA
ZERO,ONE, TWO, TWOPFO,THREE,FOUR, SIX, EIGHT, TEN, FIFTN, TWTYFO,
1
FOTYFI,CONU1,CONU2, CONU3, CONC3, CONC4, SMALL/
2
0.0, 1.0, 2.0, 2.4, 3.0, 4.0, 6.0, 8.0, 10.0, 15.0, 24.0, 45.0, 0.688,
3
0.9971, 2.393, 0.166666667, 0.416666667E-1, 1.0E-12/1
B1=C3*C3
C TEST IF BETA2 .GT. 1+BETA1
ITYPGC=3
IF(C4.LE.B1-TWO)RETURN ! IMPOSSIBLE
ITYPGC=2
IF(ABS(C3).GE.SMALL)GO TO 1
C SYMMETRIC TESTS FOR POSITIVE-DEFINITE AND UNIMODAL
IF((C4.LE.ZERO).OR.(C4.GE.FOUR))RETURN ! NOT POS-DEF
ITYPGC=1
IF(C4.GE.TWOPFO)RETURN ! POS-DEF
ITYPGC=0
RETURN ! UNI
C ASYMMETRIC TESTS FOR POSITIVE-DEFINITE AND UNIMODAL
C TEST FOR POSITIVE-DEFINITE
1 SH1L=(EIGHT*C4+TWO*C4*C4+FOUR*Bl)**3
SH1R=(TWTYFO*C4*C4+TWO*C4**3+TWTYFO*B1+SIX*B1*C4)**2
SH2L=FOUR*C4**3-C4*"4+FOUR*C4*Bl
SH2R=THREE*B1*C4**4+FOUR*B1*B1
IF((SH1L.LE.SH1R).OR.(SH2L.LE.SH2R))RETURN ! NOT POS-DEF
ITYPGC=1
C SCREEN FOR UNIMODAL
IF((B1.GT.CONU1).OR.(CONU2*C4+B1.GT.CONU3))RETURN ! POS-DEF
C TEST FOR UNIMODAL
A3=C3*CONC3
A4=C4*CONC4
A(1,1)=A4
A(1,2)=A3
A(1,3)=-FIFTN*A4
A(1,4)=-TEN*A3
A(1,5)=FOTYFI*A4+ONE
A(1,6)=FIFTN*A3
A(1,7)=-FIFTN*A4-ONE
DO 2 J=8,11
2 A(1,J)=ZERO
DO 3 I=2,5
A(I,1)=ZERO
DO 3 J=2,11
3 A(I,J)=A(I-1,J-1)
A(6,1)=A4
A(6,2)=A3
A(6,3)=-TEN*A4
A(6,4)=-SIX*A3
A(6,5)=FIFTN*A4+ONE
A(6,6)=THREE*A3
DO 4 J=7,11
REAL FUNCTION FGC(C3,C4,X)

DESC: PDF OF GRAM-CHARLIER DISTRIBUTION
INP : C3, C4, X
OUTP: FGC
HIST: >07-82, 06-84 COS MODS

REAL C3,C4,A3,A4,X,X2,X3,X4,

1   HALF,ONE,THREE,SIX,CONC3,CONC4,CONPHI
DATA HALF,ONE,THREE,SIX,CONC3,CONC4,CONPHI/
1   0.5,1.0,3.0,6.0,0.166666667,0.416666667E-1,0.398942281/
A3=C3*CONC3
A4=C4*CONC4
X2=X*X
X3=X2*X
X4=X3*X
FGC=CONPHI*EXP(-HALF*X2)*(ONE+A3*(X3-THREE*X)+A4*(X4-
1   SIX*X2+THREE))
RETURN
END

REAL FUNCTION DFGC(C3,C4,X)

DESC: DF OF GRAM-CHARLIER DISTRIBUTION
INP : C3, C4, X
OUTP: DFGC
CEXIT: ALNORM
HIST: >07-82, 06-84 COS MODS

REAL C3,C4,A3,A4,X,X2,X3,

1   HALF,ONE,THREE,CONC3,CONC4,CONPHI
DATA HALF,ONE,THREE,CONC3,CONC4,CONPHI/
1   0.5,1.0,3.0,6.0,0.166666667,0.416666667E-1,0.398942281/
A3=C3*CONC3
A4=C4*CONC4
X2=X*X
X3=X2*X
DFGC=ALNORM(X,.FALSE.)-CONPHI*EXP(-HALF*X2)*(A3*(X2-ONE)+
1   A4*(X3-THREE*X))
RETURN
END

REAL FUNCTION XMODGC(C3,C4,IFault)

DESC: MODE OF GRAM-CHARLIER DISTRIBUTION (ASSUMED UNIMODAL)
INP : C3, C4
REAL FUNCTION QGC(C3,C4,X)

DESC: FUNCTION TO OBTAIN DERIVATIVE OF GRAM-CHARLIER PDF (FOR XMODGC)

INP : C3, C4, X

OUTP: QGC

HIST: >07-82, 06-84 COS MODS

REAL C3, C4, A3, A4, X, X2, X3, X4, X5,
1 THREE, SIX, TEN, FIFTN, CONC3, CONC4

DATA THREE, SIX, TEN, FIFTN, CONC3, CONC4 /
1 3.0, 6.0, 10.0, 15.0, 0.166666667, 0.416666667E-1 /
A3 = C3*CONC3
A4 = C4*CONC4
X2 = X*X
X3 = X2*X
X4 = X3*X
X5 = X4*X
QGC = X + A3*(X4 - SIX*X2 + THREE) + A4*(X5 - TEN*X3 + FIFTN*X)
RETURN
END

REAL FUNCTION QDGC(C3,C4,X)

DESC: DERIVATIVE OF QGC (FOR XMODGC)

INP : C3, C4, X

OUTP: QDGC

HIST: >07-82, 06-84 COS MODS
REAL C3,C4,AA3,AA4,X,X2,X3,X4,
1    ONE,THREE,SIX,CONNC3,CONNC4
DATA: ONE,THREE,SIX,CONNC3,CONNC4/
1    1.0,3.0,6.0,0.666666667,0.208333333/
AA3=C3*CONNC3
AA4=C4*CONNC4
X2=X*X
X3=X2*X
X4=X3*X
QDGC=ONE+AA3*(X3-THREE*X)+AA4*(X4-SIX*X2+THREE)
RETURN
END

REAL FUNCTION DETLU(A,N)
C DESC:DETERMINANT OF SQUARE MATRIX
C INP:A (NXN),N
C OUTP:DETLU
C AUTH:M B PATE
C
REAL A(N,N),DIPROD,S,
1    ZERO,ONE,SMALL
DATA ZERO,ONE,SMALL/0.0,1.0,1.0E-12/
DETLU=ZERO
DIPROD=ONE
DO 4 J=1,N
DO 3 I=1,N
IF(ABS(A(J,J)).LT.SMALL)RETURN
IP=MINO(I,J)
S=ZERO
IF(IP.EQ.1)GO TO 2
1 S=S+A(I,K)*A(K,J)
2 S=A(I,J)-S
IF(I.GT.J)S=S/A(J,J)
3 A(I,J)=S
4 DIPROD=DIPROD*A(J,J)
DETLU=DIPROD
RETURN
END
Contents of Floppy Disk

Note:
We include GCPACK and files containing certain of the Fortran subroutines and functions developed during this investigation which might be of interest to others. Further details of these, and indeed all the algorithms used in the work, may be obtained from the author.

GCPACK
TMOM47 - moments of $t$
RB1M85 - moments of $\sqrt{b_1}$
TFUN - Owen's $T$-function (including AS 76)
BIPHI - bivariate normal integral
JFPDF - Johnson percentage points and distribution function
JFARUM - JNARR and SUMDM (Section 5.5)