# TESTING FOR NORMALITY

By R. C. GEARY, *Cambridge University Department of Applied Economics*

## 1. Introduction

The present communication, one of a series, has two main objectives:

(1) To show that probabilities derived from the well-known analyses of variance and other 'small sample' tables, which postulate universal normality, may differ seriously from the true probabilities when the universes are non-normal, even, in some cases, when the degree of non-normality is not considerable.

(2) To determine the most efficient tests of normality from a wide field of alternative symmetrical tests.

It may be useful to summarize very briefly previous work in so far as it is strictly relevant to this study.* The modern theory may be regarded as having been initiated by Karl Pearson who, in 1895, found the first approximation (i.e. to $n^{-1}$) to the variances and covariance of $\sqrt{b_1}$ and $b_2$ for samples drawn at random from any universe and, assuming that the $\sqrt{b_1}$ and $b_2$ were distributed jointly with normal probability, constructed 'probability ellipses' from which the probability of the same values occurring, had the universe, in fact, been normal, could be inferred very approximately. A considerable advance in moment determination was made by C. C. Craig (1928). In 1929, R. A. Fisher, in inventing cumulants, simple functions of the sample moments, and formulating rules for finding their semi-invariants, developed incidentally a technique for expanding to several terms in $1/n$ the moments of $\sqrt{b_1}$ and $b_2$ when the universe was normal. This paper was followed soon after by another (1930), fundamental for all succeeding work on this subject, in which R. A. Fisher ingeniously applied combinatorial technique to the finding of exact values of the moments of normal $\sqrt{b_1}$ and $b_2$, and gave *inter alia* the values of the second, fourth and sixth moments of $\sqrt{b_1}$ and of the first three moments of $b_2$. The fourth semi-invariant, together with many other normal semi-invariants of $b_2$, was determined by J. Wishart in 1930, and a further advance in R. A. Fisher's technique was made jointly by R. A. Fisher & J. Wishart in 1930. In 1932 Joseph Pepper gave the eighth normal moment of $\sqrt{b_1}$. Using R. A. Fisher's rules C. T. Hsu and D. N. Lawley in 1940 gave the exact values for normal random samples of the fifth and sixth moments of $b_2$. Using a method due to R. C. Geary (1933) (applying C. C. Craig's ideas (1928) to the normal problem), R. C. Geary & J. P. G. Worlledge have recently (1946) found the seventh moment of $b_2$.

So much for moment determination. In 1930, E. S. Pearson used appropriate Pearson-type curves, applied to R. A. Fisher's (1929) approximations of the semi-invariants, to find approximate frequency distributions of $\sqrt{b_1}$ and $b_2$. From the frequency distributions he computed a table of 1 % and 5 % probability points at intervals for $n$ from 50 to 5000 for $\sqrt{b_1}$ and for $n$ from 100 to 5000 for $b_2$.

Since at the time the prospect seemed remote of determining the frequency of normal $b_2$ on which reliance could be reposed for samples of moderate sizes, R. C. Geary (1935)† suggested that the ratio, $a$, of mean deviation to standard deviation computed from the origin

* An excellent account of the development of moment theory up to the year 1930 was given by J. Wishart (1930).

† The author was informed by M. Fréchet that this test was suggested by Bertrand, but has been unable to check the reference.

might be used as a test of normality, and gave the 1 and 5 % probability points for this test at intervals for normal samples of 6–100. E. S. Pearson compared experimentally Geary's test with $b_2$ and suggested, for samples so large that comparison could safely be made, that $b_2$ was probably somewhat more sensitive than $a$, a suggestion which will be examined theoretically in this communication. In 1935 also, R. C. Geary showed that there was a high (negative) correlation for normal samples between $a(1)$ (see 3·1) and $b_2$ for normal samples, and argued therefrom that the former should be nearly as efficient as $b_2$. In 1936, R. C. Geary gave a table of 1, 5 and 10 % probability points of $a(1)$ at intervals for samples of 11–1001. In 1938, a brochure by R. C. Geary & E. S. Pearson was published by the Biometrika Office entitled *Tests of Normality*, giving tables and diagrams of probability points of $a(1)\sqrt{b_1}$ and $b_2$. There is considerable literature dealing with the effect of universal non-normality on the normal tests, mostly by way of particular numerical examples: a selection of papers on this subject is included in the list of references at the end of the paper.

## 2. Effect of non-normality

### (a) The z-test

The effect of universal non-normality will first be considered in relation to the $z$-test. If $x_1, x_2, \ldots, x_{n'}$ and $y_1, y_2, \ldots, y_{n''}$ are two independent samples drawn at random from the same universe (normal or non-normal) it is easy to show that, if

$$z = \tfrac{1}{2}\log\frac{n''-1}{n'-1}\frac{\sum_{i=1}^{n'}(x_i-\bar{x})^2}{\sum_{i=1}^{n''}(y_i-\bar{y})^2} = \tfrac{1}{2}\log\frac{s'^2}{s''^2}, \tag{2·1}$$

then

$$\sigma_z^2 = \frac{(\beta_2-1)}{4}\left(\frac{1}{n'}+\frac{1}{n''}\right) = M_2, \tag{2·2}$$

when both $n'$ and $n''$ are so large that terms in $n'$ and $n''$ of degree less than $-1$ are regarded as negligible. This is an obvious generalization of the approximate formula given by R. A. Fisher* for normal samples, namely,

$$\sigma_z^2 = \frac{1}{2}\left(\frac{1}{n'}+\frac{1}{n''}\right) = M_2^0. \tag{2·3}$$

It may be useful also to give formulae for the first and second moments from zero for $z$ when the two random samples are drawn not necessarily from the same universes, though both universes have mean zero and the same variance $\lambda_2$:

$$\left.\begin{aligned}
2M_1' = {} & -\frac{1}{2\lambda_2^2}\left(\frac{\lambda_4'}{n'}+\frac{2\lambda_2^2}{n'-1}\right)+\frac{1}{2\lambda_2^2}\left(\frac{\lambda_4''}{n''}+\frac{2\lambda_2^2}{n''-1}\right)+\frac{1}{3\lambda_2^3}\left[\left(\frac{\lambda_6'}{n'^2}-\frac{\lambda_6''}{n''^2}\right)+12\lambda_2\left(\frac{\lambda_4'}{n'^2}-\frac{\lambda_4''}{n''^2}\right)\right.\\
& \left.+4\left(\frac{\lambda_3'^2}{n'^2}-\frac{\lambda_3''^2}{n''^2}\right)+8\lambda_2^3\left(\frac{1}{n'^2}-\frac{1}{n''^2}\right)\right]-\frac{3}{4\lambda_2^4}\left[\frac{(\lambda_4'+2\lambda_2^2)^2}{n'^2}-\frac{(\lambda_4''+2\lambda_2^2)^2}{n''^2}\right]+\ldots,\\
4M_2' = {} & \frac{1}{\lambda_2^2}\left[\left(\frac{\lambda_4'}{n'}+\frac{\lambda_4''}{n''}\right)+2\lambda_2^2\left(\frac{1}{n'-1}+\frac{1}{n''-1}\right)\right]\\
& -\frac{1}{\lambda_2^3}\left[\left(\frac{\lambda_6'}{n'^2}+\frac{\lambda_6''}{n''^2}\right)+12\lambda_2\left(\frac{\lambda_4'}{n'^2}+\frac{\lambda_4''}{n''^2}\right)+4\left(\frac{\lambda_3'^2}{n'^2}+\frac{\lambda_3''^2}{n''^2}\right)+8\lambda_2^3\left(\frac{1}{n'^2}+\frac{1}{n''^2}\right)\right]\\
& +\frac{1}{12\lambda_2^4}\left[\frac{33}{n'^2}(\lambda_4'+2\lambda_2^2)^2+\frac{33}{n''^2}(\lambda_4''+2\lambda_2^2)^2-\frac{6}{n'n''}(\lambda_4'+2\lambda_2^2)(\lambda_4''+2\lambda_2^2)\right]+\ldots
\end{aligned}\right\} \tag{2·4}$$

* *Statistical Methods for Research Workers*, 8th ed. p. 219.

where the $\lambda$'s indicate semi-invariants of the two universes of the orders indicated. In these formulae, in effect, terms to order $-2$ in $n'$, $n''$ are retained.

When both samples are large the frequency distribution of $z$ will approach normality provided that $\mu_4$ is finite. The effect of universal kurtosis can accordingly be assessed in a very rudimentary manner from (2·2) and (2·3). The $z$-deviate $\zeta$ corresponding to, say, the $2\frac{1}{2}$ % normal probability point is

$$\zeta = 1{\cdot}9600\sqrt{M_2^0}. \tag{2·5}$$

If, however, the universe were not normal and had, *in fact*, a variance $M_2$ with $\beta_2 \neq 3$, the *actual* probability of a deviation in excess of $\zeta$ in absolute value would be, not 0·05, but the normal probability appropriate to a unit variance deviate of $\zeta M_2^{-\frac{1}{2}}$. On this consideration the actual probabilities for different values of $\beta_2$, where the assumed probability is 0·05, are shown in the fifth column of Table 1.

Table 1. *Effect on probability of z of change in universal kurtosis, for large samples*

| $\beta_2$ | $M_2^0/M_2$ | $\sqrt{(M_2^0/M_2)}$ | $1{\cdot}9600\sqrt{(M_2^0/M_2)}$ | Actual probability |
|---|---|---|---|---|
| 1·5 | 4 | 2 | 3·9200 | 0·000089 |
| 2 | 2 | 1·4142 | 2·7718 | 0·0056 |
| 2·5 | 1·3333 | 1·1547 | 2·2632 | 0·024 |
| **3** | **1** | **1** | **1·9600** | **0·050** |
| 3·5 | 0·8000 | 0·8944 | 1·7530 | 0·080 |
| 4 | 0·6667 | 0·8165 | 1·6003 | 0·110 |
| 4·5 | 0·5714 | 0·7559 | 1·4816 | 0·138 |
| 5 | 0·5000 | 0·7071 | 1·3859 | 0·166 |
| 5·5 | 0·4444 | 0·6667 | 1·3065 | 0·191 |
| 6 | 0·4000 | 0·6325 | 1·2397 | 0·215 |

The table shows that, if the universe from which the samples are drawn has $\beta_2 = 6$, the true probability is about 1 in 5 instead of the assumed 1 in 20. It is, of course, true that universes with so large a kurtosis are unusual. This view cannot be held of the range 2·5–4 for $\beta_2$ in which the probability, assumed to be 0·05, can be anything, in fact, from 0·024 to 0·110. Accordingly, if universal kurtosis is markedly negative, use of the standard table masks significant differences; if kurtosis is positive the standard table exaggerates these differences. Unless systematic tests have established that kurtosis is negligible the standard table should not be used for testing significant differences in variance.

The foregoing analysis gives a theoretical explanation of the striking experimental results of E. S. Pearson (1931*b*) working, however, with a test function

$$x = \sum_{i=1}^{n'} (x_i - \bar{x})^2 \Big/ \left\{ \sum_{i=1}^{n'} (x_i - \bar{x})^2 + \sum_{i=1}^{n''} (y_i - \bar{y})^2 \right\}$$

and with sample sizes $n' = 5$ and $n'' = 20$, smaller than those contemplated in the present analysis. With 500 samples Pearson showed that when the frequency at the two tails together expected from normal theory was 15·4 (= probability 0·0308) the frequencies actually found in symmetrical universes with $\beta_2 = 2{\cdot}5$, 4·1 and 7·1 respectively were 7, 39 and 47, equivalent to probabilities of 0·014, 0·078 and 0·094.

If tests of normality indicate universal kurtosis, either of two courses might be adopted:

(i) Assume that $z$ is normally distributed with variance $M_2$ computed from (2·2) with $(\beta_2-3)$ estimated as $k_4/k_2^2$ from the sample, $k_2$ and $k_4$ being R. A. Fisher's (1929) cumulant functions.

(ii) Enter the standard table, not with $z$ computed from the samples but with $z\sqrt{(M_2^0/M_2)}$, estimating $M_2$ as in (i).

Both of these procedures are, of course, open to the objection that, unless the samples are extremely large the estimate of $\beta_2$ is unlikely to be accurate; the real $\beta_2$ might be larger or smaller than the estimate. Any probabilistic inferences should accordingly be accepted with reserve.

It is fortunate that the condition specified in the foregoing paragraphs, namely, that the numbers in the two samples are both large, rarely applies in practical applications. It more usually happens that the number of classes is small, whereas the number per class is relatively large. In this case E. S. Pearson (1931 *b*) has shown the first approximation to $\sigma_z^2$ is independent of $\beta_2$, from which he inferred that the actual probability when the total number of samples was large was inconsiderably influenced by kurtosis. In view of the foregoing analysis it seemed to the writer desirable to carry the inquiry a stage further.

Suppose, then, that $k$ samples are drawn at random from the same universe, $n_j$ in the $j$th sample, the total $\sum_j n_j = n$. It is assumed that $n$ is so large that terms in $n^{-2}$ are negligible, that the number of samples $k$ is small, and that all the $n_j$ are of the same order of magnitude as $n$, i.e. that if

$$n_j = \pi_j n, \quad \sum_{j=1}^{k} \pi_j = 1, \tag{2·6}$$

none of the $\pi_j$ is negligibly small.

Using R. A. Fisher's cumulant notation with subscript to indicate the sample from which the cumulants were computed, the mean for the $j$th sample is written $k_{1j}$ and its variance $k_{2j}$. Then

$$z = \tfrac{1}{2}\log\frac{X}{Y}, \tag{2·7}$$

where
$$(k-1)X = \sum_j n_j(k_{1j}-k_1)^2 = \sum n_j k_{1j}^2 - \frac{1}{n}\Sigma^2 n_j k_{1j},$$

so that
$$\frac{k-1}{n}X = \sum \pi_j(1-\pi_j)\,k_{1j}^2 - 2\sum_{j<j'}\sum \pi_j\pi_{j'}k_{1j}k_{1j'},$$

and
$$(n-k)\,Y = \sum_j (n_j-1)\,k_{2j},$$

so that
$$Y = \Sigma\phi_j k_{2j},$$

where
$$\phi_j = \frac{n_j-1}{n-k}.$$

Without loss of generality let the universal mean be zero and the variance unity. It may easily be shown that

$$EX = EY = 1.$$

Set
$$w = \frac{X}{Y} = \frac{\bar{X}+(X-\bar{X})}{\bar{Y}+(Y-\bar{Y})} = \{1+(X-1)\}\{1+(Y-1)\}^{-1}.$$

Then
$$\left.\begin{aligned} w &= \{1+(X-1)\}\{1-(Y-1)+(Y-1)^2-(Y-1)^3+\ldots\},\\ w^2 &= \{1+(X-1)\}^2\{1-2(Y-1)+3(Y-1)^2-4(Y-1)^3+\ldots\}.\end{aligned}\right\} \tag{2·8}$$

We shall compute the approximate values of $Ew$ and $Ew^2$, i.e. the values to order $n^{-1}$; the symbol $\doteq$ denotes 'equal to, to approximation required'. From values of the variances and covariances given by E. S. Pearson (1931*b*) in his equations (9)–(11), we have

$$\left.\begin{aligned} E(X-1)^2 &= \frac{2}{k-1}+(1-2k+\alpha_{-1}-1)\frac{\lambda_4}{n(k-1)^2},\\ E(X-1)(Y-1) &\doteq \frac{\lambda_4}{n},\\ E(Y-1)^2 &\doteq \frac{\lambda_4+2}{n}, \end{aligned}\right\} \tag{2·9}$$

with

$$\alpha_c = \sum_j \pi_j^c.$$

We require

$$\left(\frac{k-1}{n}\right)^2 X^2 \doteq \sum_j \pi_j^2(1-\pi_j)^2 k_{1j}^4 - 4\sum_j\sum_{j'} \pi_j^2(1-\pi_j)\pi_{j'}k_{1j}^3k_{1j'}$$
$$+2\sum_j\sum_{j'}\pi_j\pi_{j'}(1-\pi_j-\pi_{j'}+3\pi_j\pi_{j'})k_{1j}^2k_{2j'}^2 - 4\sum_j\sum_{j'}\sum_{j''}\pi_j\pi_{j'}\pi_{j''}(1-3\pi_j)k_{1j}^2k_{1j'}k_{1j''},$$

$$Y-1 = \Sigma\phi_j(k_{2j}-1) = \Sigma\phi_j k'_{2j}, \quad \text{say},$$

remembering that, by definition of cumulants,

$$Ek_{2j} = \lambda_2 = 1.$$

Also

$$(Y-1)^2 = \sum\phi_j^2 k'^2_{2j} + 2\sum_{j>j'}\sum \phi_j\phi_{j'}k'_{2j}k'_{2j'}.$$

It will be useful for what follows to note that

$$\phi_j \doteq \pi_j.$$

Using R. A. Fisher's formulae (1929) for formation of joint semi-invariants of $k_1$ and $k_2$, and noting that the $k$ samples are independent, we find from the foregoing

$$\left.\begin{aligned} n(k-1)\,EX(Y-1)^2 &\doteq (k-1)(\lambda_4+2),\\ n(k-1)^2\,EX^2(Y-1) &\doteq 2(k^2-1)\lambda_4,\\ n(k-1)^2\,EX^2(Y-1)^2 &\doteq (k^2-1)(\lambda_4+2). \end{aligned}\right\} \tag{2·10}$$

Then, from (2·8), (2·9), (2·10),

$$\left.\begin{aligned} Ew &\doteq 1+\frac{2}{n},\\ Ew^2 &\doteq \frac{k+1}{k-1}+\frac{1}{n(k-1)^2}\{6(k^2-1)-(k^2+2k-2-\alpha_{-1})\lambda_4\}. \end{aligned}\right\} \tag{2·11}$$

These are the formulae required. It will be noted

(i) that the terms free of $n^{-1}$ are independent of $\lambda_4$, which is equivalent to E. S. Pearson's result (1931*b*);

(ii) that the formulae (2·11) agree with the normal values

$$\left.\begin{aligned} E_0w &= \left(1-\frac{2}{n-k}\right)^{-1} \doteq 1+\frac{2}{n},\\ E_0w^2 &= \frac{k+1}{k-1}\left(1-\frac{2}{n-k}\right)^{-1}\left(1-\frac{4}{n-k}\right)^{-1} \doteq \frac{k+1}{k-1}\left(1+\frac{6}{n}\right), \end{aligned}\right\} \tag{2·12}$$

to $n^{-1}$ when $\lambda_4 = 0$;

(iii) the approximations at (2·11) are free of $\lambda_3$.

The approximations at (2·11) tend to confirm E. S. Pearson's result that, when $n$ is large compared with $k$, the effect of universal kurtosis is unimportant. It would be useful, however, to compute the approximate true probability for different values of $k$, $n$, $\lambda_4$ and $\alpha_{-1}$. For this and for subsequent work the following lemma* will be found useful:

*If $f(x)$ and $\phi(x)$ are two frequency densities with semi-invariants $L_m$ and $L'_m$ $(m = 1, 2, \ldots)$, respectively, then, formally,*

$$f(x) = \exp\left\{\sum_{m=1}^{\infty} \frac{(L_m - L'_m)}{m!}\left(-\frac{d}{dx}\right)^m\right\}\phi(x). \tag{2·13}$$

For the present application take as generating function $\phi$ the frequency distribution of $w$ in the normal case, i.e.

$$\phi(w) = \frac{\left(\frac{n-3}{2}\right)!\left(\frac{k-1}{n-k}\right)^{\frac{1}{2}(k-1)}}{\left(\frac{k-3}{2}\right)!\left(\frac{n-k-2}{2}\right)!} w^{\frac{1}{2}(k-3)}\left\{1 + \frac{(k-1)w}{(n-k)}\right\}^{-\frac{1}{2}(n-1)}, \tag{2·14}$$

and, from (2·11),
$$L_2 - L'_2 \simeq -\frac{(k^2 + 2k - 2 - \alpha_{-1})\lambda_4}{n(k-1)^2}. \tag{2·15}$$

Assume that
$$L_m - L'_m \simeq 0 \quad (m \neq 2).$$

Then if the 'normal theory' probability corresponding to the sample value $w$ be $p$, the approximate 'true' probability, subject to (2·15), will be about $(p + p')$, where $p'$ is given by

$$p' = \frac{(L_2 - L'_2)}{2}\int_w^{\infty} \phi''(w)\, dw = -\frac{(L_2 - L'_2)}{2}\phi'(w). \tag{2·16}$$

The term $p'$, of course, merely corrects for the non-normal term in $n^{-1}$ in the variance of $z$; it takes no account of corrections due to terms of higher (negative) orders in $n$ or even of non-normal terms in $n^{-1}$ in semi-invariants $L_m$ $(m > 2)$. The calculation is designed merely to show whether the standard table probability requires correction for universal kurtosis; this will appear if $p'$ is of the order of magnitude of $p$.

### (b) *The t-test*

In Geary's 1936 paper the expansion to terms in $n^{-2}$ of the first *four* moments of $t$, where

$$t = n^{\frac{1}{2}} k_1 / k_2^{\frac{1}{2}}, \tag{2·17}$$

were given. Following are the first *six* semi-invariants $L$ of $t$ to the same approximation as in the earlier paper:

$$\left.\begin{aligned}
L_1 &\simeq -\frac{1}{n^{\frac{1}{2}}}\left\{\frac{\lambda_3}{2} + \frac{3}{16n}(2\lambda_3 - 2\lambda_5 + 5\lambda_3\lambda_4)\right\} + \ldots,\\
L_2 &\simeq 1 + \tfrac{1}{4}(8 + 7\lambda_3^2)\,n^{-1} + (6 - 2\lambda_4 - \tfrac{3}{8}\lambda_3^2 - \tfrac{45}{8}\lambda_3\lambda_5 + \tfrac{177}{16}\lambda_3^2\lambda_4)\,n^{-2},\\
L_3 &\simeq -2\lambda_3 n^{-\frac{1}{2}} - (9\lambda_3 - 3\lambda_5 + \tfrac{15}{4}\lambda_3\lambda_4 + \tfrac{83}{8}\lambda_3^3)\,n^{-\frac{3}{2}},\\
{}^{\dagger}L_4 &\simeq (6 - 2\lambda_4 + 12\lambda_3^2)\,n^{-1} + (54 - 18\lambda_4 + 4\lambda_6 + 75\lambda_3^2 - 63\lambda_3\lambda_5 - 6\lambda_4^2 + 81\lambda_3^2\lambda_4 + \tfrac{699}{8}\lambda_3^4)\,n^{-2},\\
L_5 &\simeq -(60\lambda_3 - 6\lambda_5 - 20\lambda_3\lambda_4 + 105\lambda_3^3)\,n^{-\frac{3}{2}},\\
L_6 &\simeq (240 - 120\lambda_4 + 577\tfrac{1}{2}\lambda_3^2 + 16\lambda_6 - 210\lambda_3\lambda_5 - 150\lambda_3^2\lambda_4 + 1200\lambda_3^4)\,n^{-2}.
\end{aligned}\right\} \tag{2·18}$$

Throughout this subsection we take
$$\lambda_m = \lambda'_m / \lambda_2'^{\frac{1}{2}m},$$

* Due to Charlier and termed the "Differential Series" by the Scandinavian School.

† 1936 formula corrected.

where the $\lambda'_m$ are the semi-invariants of the parent universe. For these expressions terms in $n^{-\frac{5}{2}}$ are neglected. They were derived from the moments (from zero) $M'_i$ of $t$, which were obtained by the method described in the 1936 paper. It will be noted that, to the approximation used, the expressions involve only the first six semi-invariants of the parent universe. When the parent universe is normal all the $\lambda_i$ $(i > 2)$ are zero. The magnitude of the numerical coefficients in the foregoing approximate expressions for the $L_i$ indicate that, when the universal values of the $\lambda_i$, particularly those of uneven order, are not very small, the frequency distribution of $t$ may differ appreciably from the classical Gosset-Fisher (1908, 1925) distribution.

The formal Gram-Charlier expression for the frequency of $t$ could, of course, be written down at once from (2·18). It is doubtful, however, if the Gaussian can be regarded as the most appropriate generating function for the frequency of $t$ because, even when the parent universe is normal, the semi-invariants $T'_{2m}$ of the higher even orders are large for moderate values of $n$. For example,

$$L'_4/L'^2_2 = 6/(n-5); \quad L'_6/L'^3_2 = 240/(n-5)(n-7).$$

It is proposed to use (2·13) for finding the approximate frequency with

$$\phi(t) = T(t;n) = \left(\frac{n-2}{2}\right)! \left(1+\frac{t^2}{n-1}\right)^{-\frac{1}{2}n} \Big/ \left(\frac{n-3}{2}\right)! \, (\pi \overline{n-1})^{\frac{1}{2}}, \tag{2·19}$$

the Gosset-Fisher frequency. Let

$$T_1(t;n) = \left(1+\frac{t^2}{n-1}\right)^{-\frac{1}{2}n} \tag{2·20}$$

It can easily be shown that the $r$th derivative (in $t$) of $T_1$ is

$$T_1^{(r)}(t;n) = (-)^r \frac{(n+r-1)!}{(n-1)!\,(n-1)^r} \left\{ t^r - n_1 \frac{r(r-1)}{2} t^{r-2} + n_2 \frac{r(r-1)(r-2)(r-3)}{2.4} t^{r-4} \right.$$
$$\left. - n_3 \frac{r(r-1)(r-2)(r-3)(r-4)(r-5)}{2.4.6} t^{r-6} + \ldots \right\} \left(1+\frac{t^2}{n-1}\right)^{-\frac{1}{2}(n+2r)}, \tag{2·21}$$

with $$n_1 = \frac{n-1}{n+1}, \quad n_2 = \frac{(n-1)^2}{(n+1)(n+3)}, \quad n_3 = \frac{(n-1)^3}{(n+1)(n+3)(n+5)}, \quad \text{etc.}$$

Note that (2·21) assumes the Hermite form when $n = \infty$.

The theory will now be applied to particular examples using in all cases $n = 10$. The universes will be assumed to belong to the Karl Pearson system, so that (M. G. Kendall, 1941) the values of $\lambda_5$ and $\lambda_6$ can be derived (given $\lambda_3$ and $\lambda_4$) from the following equations:

$$\left.\begin{aligned} (1+4\eta)\lambda_3 + 2\xi &= 0, \\ (1+5\eta)\lambda_4 + 3\xi\lambda_3 + 6\eta &= 0, \\ (1+6\eta)\lambda_5 + 4\xi\lambda_4 + 24\eta\lambda_3 &= 0, \\ (1+7\eta)\lambda_6 + 5\xi\lambda_5 + 10\eta(4\lambda_4 + 3\lambda_3^2) &= 0. \end{aligned}\right\} \tag{2·22}$$

From the first two equations

$$\eta = (2\lambda_4 - 3\lambda_3^2)/(-10\lambda_4 + 12\lambda_3^2 - 12),$$

which, substituted in the first equation of (2·22), gives $\xi$. The values of $\xi$ and $\eta$, substituted in the third and fourth equations, give $\lambda_5$ and $\lambda_6$. From (2·18), the $L'_i$ being the semi-invariants when the parent universe is normal (i.e. the values found when all the $\lambda$'s are set equal to zero),

$$\left.\begin{aligned} L_1-L_1' &\simeq J_1 n^{-\frac{1}{2}}+K_1 n^{-\frac{3}{2}}, & L_4-L_4' &\simeq J_4 n^{-1}+K_4 n^{-2},\\ L_2-L_2' &\simeq J_2 n^{-1}+K_2 n^{-2}, & L_5-L_5' &\simeq K_5 n^{-\frac{3}{2}},\\ L_3-L_3' &\simeq J_3 n^{-\frac{1}{2}}+K_3 n^{-\frac{3}{2}}, & L_6-L_6' &\simeq K_6 n^{-2}. \end{aligned}\right\} \tag{2·23}$$

The $J$ and $K$ are the terms in the $\lambda$ in (2·18). To $n^{-2}$ (i.e. ignoring $n^{-\frac{5}{2}}$) the frequency generated from $T$ of (2·19) is as follows:

$$\begin{aligned} f(t) = T &+ n^{-\frac{1}{2}}\left\{J_1 D+\frac{J_3}{6}D^3\right\}+n^{-1}\left\{\frac{D^2}{2}(J_2+J_1^2)+\frac{D^4}{24}(J_4+4J_1J_3)+\frac{D^6}{72}J_3^2\right\}\\ &+n^{-\frac{3}{2}}\left\{K_1 D+\frac{D^3}{6}(K_3+3J_1J_2+J_1^3)+\frac{D^5}{120}(K_5+5J_1J_4+10J_2J_3+10J_1^2J_3)\right.\\ &\left.+\frac{D^7}{144}(J_3J_4+2J_1J_3^2)+\frac{J_3^3}{1296}D^9\right\}+n^{-2}\left\{\frac{D^2}{2}(K_2+2J_1K_1)+\frac{D^4}{24}(K_4+4J_1K_3\right.\\ &+4J_3K_1+3J_2^2+6J_1^2J_2+J_1^4)+\frac{D^6}{720}(K_6+6J_1K_5+20J_3K_3+15J_2J_4\\ &+60J_1J_2J_3+15J_1^2J_4+20J_1^3J_3)+\frac{D^8}{11{,}520}(16J_3K_5+10J_4^2+80J_2J_3^2\\ &\left.+80J_1J_3J_4+80J_1^2J_3^2)+\frac{D^{10}}{5184}(3J_3^2J_4+4J_1J_3^3)+\frac{J_3^4}{31{,}104}D^{12}\right\}, \end{aligned} \tag{2·24}$$

with
$$D^h=\left(-\frac{d}{dt}\right)^h T.$$

To $n^{-1}$, (2·24) agrees with the formula given by M. S. Bartlett (1935), in which, however, there is a small and obvious slip in a sign. The law of formation of the numerical coefficients of (2·24) is evident; for instance, the numerical coefficient of $D^8 J_2 J_3^2$ is $1/144 = 1/2!\,3!^2 2!$. The integrals $\int_t^{\infty}$ and $\int_{-\infty}^{-t}$ $(t>0)$ are found by reducing the exponent of $D$ by unity, as follows:

$$\int_{-\infty}^{-t} D\,dt = -T,\quad \int_t^{\infty} D^{2m}\,dt = \int_{-\infty}^{-t} D^{2m}\,dt = D^{2m-1},\quad \int_t^{\infty} D^{2m+1}\,dt = -\int_{-\infty}^{-t} D^{2m+1}\,dt = D^{2m}. \tag{2·25}$$

In normal theory the upper and lower $2\frac{1}{2}$ % points of $t$ are $\pm 2·262$ for $n = 10$. Table 2 shows the 'true' probabilities, i.e. the value of

$$\int_{-\infty}^{-2·262} f(t)\,dt \tag{2·26}$$

for parent universes specified by $\lambda_3$, $\lambda_4$, using (2·24).

There are two observations to be made on the results presented in this table. The first is that, despite the considerable number of terms (shown at (2·24)) included in the probability expansion, the values found in the successive terms cannot be regarded as satisfactorily convergent for so small a sample as 10, and, of course, the convergence disimproves with increasing $\sqrt{\beta_1}$. Taken all together, however, they seem consistent and significant. The second observation is that attention was confined to the negative 'tail' of the distribution. It may be assumed that, in all cases, the distortion would be very considerably less marked if regard were had to the probability for $|t| > 2·262$. Actually for universe 3 the probability is 0·056, not significantly different from the normal theory probability of 0·05. In justification of the attitude adopted above, the point might be put as follows:

We decide to accept the hypothesis that the universal mean is zero provided that the value of $t$ found from the particular sample satisfies $t_0 \leqslant t \leqslant t_1$, where

$$\text{Prob}\,(t < t_0) = \text{Prob}\,(t > t_1) = 0{\cdot}025.$$

The table is designed to show that if the parent universe is markedly asymmetrical the range $(t_0, t_1)$ may differ appreciably from $-t_0 = t_1 = 2{\cdot}262$.

Table 2. *Probabilities of t less than* $-2{\cdot}262$ *for samples of* 10 *for seven universes*

| Universe | $\lambda_3 = \sqrt{\beta_1}$ | $\lambda_4 = \beta_2 - 3$ | Probability |
|---|---|---|---|
| Normal | 0 | 0 | 0·025 |
| 2 | 0 | 1 | 0·024 |
| 3 | 1/2 | 0 | 0·041 |
| 4 | $1/\sqrt{2}$ | 1/2 | 0·047 |
| 5 | 1 | 0 | 0·072? |
| 6 | 1 | 1 | 0·086? |
| 7 | 1/2 | 1/2 | 0·043 |

As anticipated by earlier work (W. S. Gosset, 1908; R. C. Geary, 1936), the table shows that the distortion is slight for symmetrical universes; even when $\lambda_4 = 1$ (and $\lambda_3 = 0$) the probability (0·024) is practically identical with the normal value. There can be little doubt that the standard table probabilities can be seriously at variance with the true probabilities when the universes from which the samples are drawn are markedly asymmetrical.

### (c) *Difference of means*

R. A. Fisher's (1925) test of significance

$$t = \frac{(k_1' - k_1'')\sqrt{(n' + n'' - 2)}}{\{(n'-1)\,k_2' + (n''-1)\,k_2''\}^{\frac{1}{2}}}\sqrt{\frac{n'n''}{(n'+n'')}}, \tag{2·27}$$

for the difference of averages $k_1'$ and $k_1''$ in normal theory for random samples numbering $n'$ and $n''$ is, of course, a particular case of the analysis of variance considered in § ($a$) above. The second cumulants are $k_2'$ and $k_2''$. It is assumed that the unknown universal means and variances are equal. Suppose now that the random samples in reality have been derived from universes in which the means are equal but the other semi-invariants $\lambda_i'$ and $\lambda_i''$ are not necessarily zero for $i \geqslant 2$, or even necessarily equal. Since the universal means are assumed equal, without loss of generality we may take $\lambda_1' = \lambda_1'' = 0$. This general mathematical model seems to be the correct one; we are not trying to determine the probability of the samples being derived from the *same universe* but rather if they could conceivably have been drawn from universes with the *same arithmetic mean*, however much they may differ otherwise. The correctness or otherwise of the concept may be considered in relation to, say, the problem of deciding from two random samples which of two types of fertilizer is to be preferred from yield observations on a given crop on a given kind of land. Undoubtedly the prime problem will be that of ascertaining which is probably the better yielding (i.e. whether the arithmetic means are significantly different). Of considerably less importance is the question of which fertilizer is the more variable; of less importance still is the question of deciding, say, whether with approximately equal yields one universe is symmetrical and the other markedly asymmetrical. The point is that the question of the equality of universal means should be considered without assuming that the other semi-invariants in the universes from which the samples have been drawn are necessarily equal. This essentially is also the viewpoint in R. A. Fisher's randomization method.

Expanding the denominator of (2·27) in terms of $(k_2'-\lambda_2')$ and $(k_2''-\lambda_2'')$ and computing therefrom the first few terms of the first four moments of $t$, we find the following approximations to the first four semi-invariants:

$$\left.\begin{aligned}
AL_1 &\simeq -\frac{(\lambda_3'-\lambda_3'')}{2(n'\lambda_2'+n''\lambda_2'')},\\
A^2L_2 &\simeq \left(\frac{\lambda_2'}{n'}+\frac{\lambda_2''}{n''}\right)\left(1+\frac{2\overline{n'\lambda_2'^2+n''\lambda_2''^2}}{n'\lambda_2'+n''\lambda_2''}\right)\\
&\qquad +\frac{(n'^2-n''^2)(\lambda_4'\lambda_2''-\lambda_4''\lambda_2')}{n'n''(n'\lambda_2'+n''\lambda_2'')^2}+\frac{7(\lambda_3'-\lambda_3'')^2}{4(n'\lambda_2'+n''\lambda_2'')^2},\\
A^3L_3 &\simeq \frac{\lambda_3'}{n'^2}-\frac{\lambda_3''}{n''^2}-\frac{3(\lambda_3'-\lambda_3'')}{(n'\lambda_2'+n''\lambda_2'')}\left(\frac{\lambda_2'}{n'}+\frac{\lambda_2''}{n''}\right),\\
A^4L_4 &\simeq \frac{6(n'\lambda_2'^2+n''\lambda_2''^2)}{(n'\lambda_2'+n''\lambda_2'')^2}\left(\frac{\lambda_2'}{n'}+\frac{\lambda_2''}{n''}\right)^2-6\left(\frac{\lambda_3'}{n'^2}-\frac{\lambda_3''}{n''^2}\right)\frac{(\lambda_3'-\lambda_3'')}{(n'\lambda_2'+n''\lambda_2'')}\\
&\quad +\frac{18(\lambda_3'-\lambda_3'')^2}{(n'\lambda_2'+n''\lambda_2'')^2}\left(\frac{\lambda_2'}{n'}+\frac{\lambda_2''}{n''}\right)+\frac{\lambda_4'}{n'^3}+\frac{\lambda_4''}{n''^3}\\
&\quad -3\left(\frac{\lambda_2'}{n'}+\frac{\lambda_2''}{n''}\right)\frac{\{\lambda_4'(n'n''\lambda_2'+\overline{2n''^2-n'^2}\lambda_2'')+\lambda_4''(n'n''\lambda_2''+\overline{2n'^2-n''^2}\lambda_2')\}}{n'n''(n'\lambda_2'+n''\lambda_2'')^2},
\end{aligned}\right\}\qquad (2{\cdot}28)$$

with
$$A=\left\{\left(\frac{n'+n''}{n'n''}\right)\frac{(\lambda_2'\overline{n'-1}+\lambda_2''\overline{n''-1})}{(n'+n''-2)}\right\}^{\frac{1}{2}}.$$

Using formula (2·24) to the term in $n^{-1}$ with the Gosset-Fisher function again as generating function, Table 3 shows rough approximations, for four examples, to the 'true' probability of values of $t \leqslant \tau$, where $\tau$ is the (negative) value for probability 0·025 from the normal table, and $\lambda_2' = \lambda_2'' = 1$. When the two samples are drawn from different universes the distortion can accordingly be considerable. The third example suggests that if the universes are the same the distortion is small, a result to be anticipated from the fact (apparent from (2·28)) that, to the approximation used, the first two semi-invariants are equal to their normal theory values; this theory confirms the experimental results of E. S. Pearson & N. K. Adyanthaya (1929).

Table 3

| Example | $n'$ | $n''$ | $\lambda_3'$ | $\lambda_3''$ | $\lambda_4'$ | $\lambda_4''$ | Probability |
|---|---|---|---|---|---|---|---|
| 1 | 12 | 4 | 1 | −1 | 1 | −1 | 0·045 |
| 2 | 18 | 6 | 1 | −1 | 1 | −1 | 0·041 |
| 3 | 7 | 4 | $1/\sqrt{2}$ | $1/\sqrt{2}$ | 1/2 | 1/2 | 0·027 |
| 4 | 10 | 6 | 1 | 0 | 1 | 0 | 0·036 |

It should be remarked that the probabilities in Table 3 (as well as in Table 2) are merely rough approximations—the samples used are far too small for the results to have any pretension to accuracy. The object has been merely to show that the actual probability *could* be considerably at variance with that shown in the standard table, for small samples.

### 3. Sufficient conditions for approach to normality of $a(c)$ with increasing $n$

The remainder of the paper deals with the field of symmetrical tests of normality, homogeneous of degree zero, represented by (3·1). It is essential to establish the conditions of approach to normality of the frequency distribution of $a(c)$ as the sample number increases.

Let
$$a(c) = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|^c \bigg/ \left\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^{\frac{1}{2}c}, \tag{3·1}$$

where $\bar{x} = \Sigma x_i/n$ and $c$ is non-negative. It will be shown in succession that, subject to stated conditions, with increasing $n$,

(i) the frequency distribution of
$$a_1(c) = \frac{1}{n}\Sigma|x_i|^c \bigg/ \left\{\frac{1}{n}\Sigma x_i^2\right\}^{\frac{1}{2}c} \tag{3·2}$$
tends towards normality, and

(ii) the frequency distribution of $a_1(c)$ tends towards that of $a(c)$ and hence towards normality.

It is assumed, without loss of generality, that the universal mean of the universe from which the sample of $n$ is drawn is zero. Denote the $k$th absolute moment from zero by $\mu_{|k|}$, $k$ not being necessarily an integer. Given a positive quantity $\epsilon$ arbitrarily small, $\omega(\epsilon)$ can be found so that
$$\text{Prob}\left\{\left|\frac{1}{n}\Sigma(|x_i|^c - \mu_{|c|})\right| < \omega\sqrt{\frac{(\mu_{|2c|}-\mu_{|c|}^2)}{n}}\right\} > 1-\epsilon, \tag{3·3}$$
$$\text{Prob}\left\{\left|\frac{1}{n}\Sigma(x_i^2 - \mu_2)\right| < \omega\sqrt{\frac{(\mu_4-\mu_2^2)}{n}}\right\} > 1-\epsilon, \tag{3·4}$$

provided, of course, that $\mu_{|2c|}$ and $\mu_4$ exist. As $n$ increases $\omega$ may be envisaged as approaching the normal probability point appropriate to the probability $\epsilon$, since, in the conditions stated, $\Sigma|x_i|^c/n$ and $\Sigma x_i^2/n$ are normally distributed in the limit. For samples which satisfy the inequality in the brackets { } at (3·4) and if $n$ is so large that
$$\omega\sqrt{\left(\frac{\mu_4}{n}\right)} < \mu_2,$$

the denominator of (3·2) can be expanded to three terms (including the remainder) by Taylor's theorem, so that $a_1(c)$ may be written
$$a_1(c) = \mu_{|c|}\mu_2^{-\frac{1}{2}c}\left\{1+\frac{1}{n}\Sigma\left(y_i-\frac{c}{2}z_i\right)-\frac{c}{2n^2}\Sigma y_i\Sigma z_i+\frac{c(c+2)}{8n^2}(\Sigma z_i)^2\left(1+\frac{1}{n}\Sigma y_i\right)X\right\}, \tag{3·5}$$

with
$$y_i = (|x_i|^c - \mu_{|c|})/\mu_{|c|},$$
$$z_i = (x_i^2-\mu_2)/\mu_2,$$
$$X = \mu_2^{\frac{1}{2}c+2}\left\{\mu_2+\frac{\theta}{n}\Sigma(x_i^2-\mu_2)\right\}^{-\frac{1}{2}(c+4)} \quad (0<\theta<1).$$

With probability exceeding $(1-\epsilon)$ it is evident, from (3·4), that $X$ is maximized by

$$\left(1-\frac{\omega}{\mu_2}\sqrt{\frac{\mu_4}{n}}\right)^{-\frac{1}{2}(c+4)}$$

It will suffice, for the present purpose, to infer that

$$|X| < \kappa,$$

where $\kappa$ is a constant independent of $n$. We have now

$$E\frac{1}{n}\Sigma\left(y_i-\frac{c}{2}z_i\right) = 0.$$

Set

$$\sigma^2 = E\frac{1}{n^2}\left\{\Sigma\left(y_i-\frac{c}{2}z_i\right)\right\}^2$$

$$= \frac{1}{n}\left\{\frac{\mu_{|2c|}}{\mu_{|c|}^2}-\frac{c\mu_{|c+2|}}{\mu_{|c|}\mu_2}+\frac{c^2}{4}\frac{\mu_4}{\mu_2^2}-\left(\frac{c}{2}-1\right)^2\right\}, \tag{3·6}$$

and

$$\frac{1}{\sigma}\left(\frac{\mu_2^{\frac{1}{2}c}a_1(c)}{\mu_{|c|}}-1\right)-\frac{1}{n\sigma}\Sigma\left(y_i-\frac{c}{2}z_i\right) = u, \tag{3·7}$$

with

$$u = -\frac{c}{2n^2\sigma}\Sigma y_i\,\Sigma z_i+\frac{c(c+2)}{8n^2\sigma}(\Sigma z_i)^2\left(1+\frac{1}{n}\Sigma y_i\right)X. \tag{3·8}$$

For samples which satisfy the inequalities in { } at (3·3) and (3·4) and hence with a probability exceeding $(1-2\epsilon)$, we have

$$|u| < \frac{c\omega^2}{2\sigma}\frac{\sqrt{(\mu_{|2c|}\mu_4)}}{n\mu_{|c|}\mu_2}+\frac{c(c+2)\kappa\omega^2\mu_4}{8\sigma\quad n\mu_2^2}\left(1+\frac{\omega}{\mu_{|c|}}\sqrt{\frac{\mu_{|2c|}}{n}}\right) < \frac{\xi}{\sqrt{n}}, \tag{3·9}$$

where $\xi$ is independent of $n$. Or, briefly,

$$\text{Prob}\left\{|u| < \frac{\xi}{\sqrt{n}}\right\} > 1-2\epsilon, \tag{3·10}$$

so that $u$ tends in probability towards zero with $1/n$. Now (3·7) may be written in the form $u = Y'-Y$, where $Y'$ and $Y$ are the respective terms on the left side. If $A$ be any number and $F$ the total probability function, a well-known lemma (Fréchet, 1937, p. 164) shows that

$$|F_{Y'}(A)-F_Y(A)| \leqslant \left\{F_Y\left(A+\frac{\xi}{\sqrt{n}}\right)-F_Y\left(A-\frac{\xi}{\sqrt{n}}\right)\right\}+2\epsilon, \tag{3·11}$$

using (3·10). Hence the frequency distribution of

$$Y' = \frac{1}{\sigma}\left(\frac{\mu_2^{\frac{1}{2}c}a_1(c)}{\mu_{|c|}}-1\right) \tag{3·12}$$

tends towards that of

$$Y = \frac{1}{n\sigma}\Sigma\left(y_i-\frac{c}{2}z_i\right) \tag{3·13}$$

at every continuity point of the latter frequency, as $n$ tends towards infinity. But $Y$, from (3·13), is the simple average of $n$ random measures, and its frequency must tend towards normality provided that its standard deviation exists; from (3·6) it is evident that $\sigma$ is finite provided that $\mu_{|k|}$, where $k$ is the greater of $2c$ and 4, is finite. Here and in the remainder of this section it will be useful to remember that if $\mu_{|k|}$ exists so does $\mu_{|k'|}$ for $0 \leqslant k' \leqslant k$.

To prove that the frequency distribution of $a(c)$ tends towards that of $a_1(c)$ and hence towards normality with increasing $n$ it will be shown that $\Sigma\,|\,x_i-\bar{x}\,|^c/n$ tends in probability towards $\Sigma\,|\,x_i\,|^c/n$. Two cases will be considered separately: (1) $c \geqslant 1$, (2) $1 > c \geqslant 0$.

*Case* (1). $c \geqslant 1$

For values of $x_i$ for which $|\,x_i\,| \geqslant |\,\bar{x}\,|$,

$$|\,x_i-\bar{x}\,|^c-|\,x_i\,|^c = \pm c\bar{x}\,|\,x_i-\theta\bar{x}\,|^{c-1} \quad (0<\theta<1)$$

and when $|\,x_i\,| < |\,\bar{x}\,|$, $\quad ||\,x_i-\bar{x}\,|^c-|\,x_i\,|^c| \leqq (2^c+1)\,|\,\bar{x}\,|^c.$

Hence $$\frac{1}{n}\left|\sum_{i=1}^{n}(|\,x_i-\bar{x}\,|)^c-(|\,x_i\,|^c)\right| < |\,\bar{x}\,|\left(\frac{B}{n}\sum_{i=1}^{n}|\,x_i\,|^{c-1}+C\,|\,\bar{x}\,|^{c-1}\right), \tag{3·14}$$

$B$ and $C$ being independent of the $x_i$ and $n$ but depending on $c$. With $\epsilon$ arbitrarily small $\omega$ can be found so that

$$\left.\begin{aligned}&\text{Prob}\left\{|\,\bar{x}\,| < \omega\sqrt{\frac{\mu_2}{n}}\right\} > 1-\epsilon,\\ &\text{Prob}\left\{\left|\frac{1}{n}\Sigma(|\,x_i\,|^{c-1}-\mu_{|c-1|})\right| < \omega\sqrt{\frac{\mu_{|2c-2|}-\mu^2_{|c-1|}}{n}}\right\} > 1-\epsilon.\end{aligned}\right\} \tag{3·15}$$

Hence, from (3·14) and (3·15), if $\mu_2$ and $\mu_{|2c-2|}$ exist,

$$\text{Prob}\left\{\left|\frac{1}{n}\Sigma\,|\,x_i-\bar{x}\,|^c-\frac{1}{n}\Sigma\,|\,x_i\,|^c\right| < B'\frac{\omega\mu_{|c-1|}\mu_2^{\frac{1}{2}}}{\sqrt{n}}\right\} > 1-2\epsilon$$

for $n$ sufficiently large the constant $B'$ depending on $c$ but not on $n$. Hence for $c \geqq 1$, $\Sigma\,|\,x_i-\bar{x}\,|^c/n$ tends in probability towards $\Sigma\,|\,x_i\,|^c/n$. Incidentally, this proves that $\{\Sigma(x_i-\bar{x})^2/n\}^{\frac{1}{2}c}$ tends in probability towards $\{\Sigma x_i^2/n\}^{\frac{1}{2}c}$, the latter two expressions representing respectively the denominators of $a(c)$ and $a_1(c)$.

*Case* (2). $1 > c \geqslant 0$

Let $\bar{x}$ satisfy a probabilistic inequality identical in form with the first equation of (3·15) and let $\gamma$ be any positive quantity, fixed once for all. Let $n$ (presently to be defined further) be so large that

$$\gamma > \omega\sqrt{\frac{\mu_2}{n}}.$$

Then $$\frac{1}{n}\sum_{i=1}^{n}(|\,x_i-\bar{x}\,|^c-|\,x_i\,|^c) = \frac{1}{n}\left(\sum_{|x_i|\geqq\gamma}{}' + \sum_{|x_i|<\gamma}{}''\right)(|\,x_i-\bar{x}\,|^c-|\,x_i\,|^c). \tag{3·16}$$

When $|\,x_i\,| \geqslant \gamma$ (i.e. in $\Sigma'$),

$$|\,x_i-\bar{x}\,|^c-|\,x_i\,|^c = \pm c\bar{x}\,|\,x_i-\theta\bar{x}\,|^{c-1} \quad (0<\theta<1),$$

so that $$\text{Prob}\left\{|\,|x_i-\bar{x}\,|^c-|\,x_i\,|^c\,| < c\omega\left(\gamma-\omega\sqrt{\frac{\mu_2}{n}}\right)^{c-1}\sqrt{\frac{\mu_2}{n}}\right\} > 1-\epsilon. \tag{3·17}$$

When $|\,x_i\,| < \gamma$ (i.e. in $\Sigma''$), given $\eta$ arbitrarily small and positive, $n$ can be found so that

$$||\,x_i-\bar{x}\,|^c-|\,x_i\,|^c\,| < \eta, \tag{3·18}$$

when $$|\,\bar{x}\,| < \omega\sqrt{\frac{\mu_2}{n}},$$

since $|\,x\,|^c$ $(c>0)$ is uniformly continuous in $\Sigma''$. We then have

$$\text{Prob}\{||\,x_i-\bar{x}\,|^c-|\,x_i\,|^c\,| < \eta\} > 1-\epsilon. \tag{3·19}$$

Combining (3·17) and (3·19), it may be inferred that

$$\text{Prob}\left\{\left|\frac{1}{n}\Sigma\,|\,x_i-\bar{x}\,|^c-\frac{1}{n}\Sigma\,|\,x_i\,|^c\right|<c\omega\left(\gamma-\omega\sqrt{\frac{\mu_2}{n}}\right)^{c-1}\sqrt{\frac{\mu_2}{n}}+\eta\right\}>1-2\epsilon, \tag{3·20}$$

the first term of the upper limit in { } tending to zero as $n$ tends towards infinity, and $\epsilon$ and $\eta$ being arbitrarily small

We have accordingly shown that the numerator and denominator of $a(c)$ tends in probability towards those of $a_1(c)$. Hence $a(c)$ tends in probability towards $a_1(c)$. Hence, using the lemma cited at (3·11), the total frequency of $a(c)$ tends towards that of $a_1(c)$ which tends towards normality as $n$ tends towards infinity. Finally:

*If $c \geqslant 0$ the frequency distribution of $a(c)$, given by* (3·1), *tends towards normality as $n$ tends towards infinity provided that $\mu_{|k|}$, where $k$ is the greater of* 2*c and* 4, *is finite.*

It seems likely that an analogous theorem can be proved for $0 > c > -\frac{1}{2}$; we shall not, however, be concerned in this communication with negative values of $c$.

## 4. Moments of $a(c)$ for normal samples

While it will be shown in later sections that, with indefinitely large samples, $\sqrt{b_1}$ and $b_2$ are the most efficient tests of asymmetry and kurtosis, respectively, it by no means follows that other tests are inefficient or that they may not be useful supplements in cases in which the prime tests are indecisive as to the probable non-normality of a given sample. It is accordingly proposed to give here close approximations to the first four moments (from the origin) of $a(c)$ (given by (3·1)) for normal random samples of $n$.

For normal samples (R. A. Fisher, 1929; R. C. Geary, 1933)

$$M'_k\{a(c)\} = E\{a(c)\}^k = E\left\{\frac{1}{n}\sum_{i=1}^{n}|\,x_i-\bar{x}\,|^c\right\}^k \Big/ E\left\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^{\frac{1}{2}ck} \tag{4·1}$$

The exact value of the denominator is, of course, known, for

$$E\left\{\frac{1}{n}\Sigma(x_i-\bar{x})^2\right\}^{\frac{1}{2}k'} = \left(\frac{n-1}{n}\right)^{\frac{1}{2}k'} Es^{k'} = \left(\frac{2}{n}\right)^{\frac{1}{2}k'}\left(\frac{n+k'-3}{2}\right)!\Big/\left(\frac{n-3}{2}\right)!, \tag{4·2}$$

since, as usual, $(n-1)\,s^2 = \Sigma(x_i-\bar{x})^2$. It will be useful to expand $\log_e Es^{k'}$ with $k' = ck$ using Stirling's formula in (4·2):

$$\begin{aligned}\log_e Es^{k'} &= \frac{k'}{2}\log\frac{2}{n-1}+\log\left(\frac{n+k'-3}{2}\right)!-\log\left(\frac{n-3}{2}\right)!\\ &= \frac{(k'^2-2k')}{4(n-1)}-\frac{k'(k'-1)(k'-2)}{12(n-1)^2}+\frac{k'^2(k'-2)^2}{24(n-1)^3}-\frac{k'(k'-1)(k'-2)(3k'^2-6k'-4)}{120(n-1)^4}\\ &\quad+\frac{k'^2(k'-2)^2(k'^2-2k'-2)}{60(n-1)^5}-\frac{k'(k'-1)(k'-2)(3k'^4-12k'^3+24k'+16)}{252(n-1)^6}\\ &\quad+\frac{k'^2(k'-2)^2(3k'^4-12k'^3-4k'^2+32k'+32)}{336(n-1)^7},\end{aligned} \tag{4·3}$$

which checks for $k' = 1$ to $(n-1)^{-7}$ with Geary (1935, p. 354). Take

$$v(c) = \frac{1}{n}\sum_{i=1}^{n}|\,z_i\,|^c, \tag{4·4}$$

with

$$z_i = x_i-\bar{x}.$$

The moments of $v(c)$ will be found exactly as in the case of $c = 1$ (Geary, 1936) from the single or joint normal frequency distributions of $(z_1, z_2, \ldots)$. We find

$$M_1'\{v(c)\} = \frac{1}{\sqrt{\pi}}\left(\frac{2\overline{n-1}}{n}\right)^{\frac{1}{2}c}\left(\frac{c-1}{2}\right)!,$$

$$M_2'\{v(c)\} = \frac{1}{\sqrt{\pi}}\frac{(2\overline{n-1})^c}{n^{c+1}}\left(\frac{2c-1}{2}\right)! + \frac{2^c}{\pi}n^{-\frac{1}{2}}(n-1)^{-c}(n-2)^{\frac{1}{2}(2c+1)}\left[\left(\frac{c-1}{2}\right)!\right]^2 \tag{4·5}$$

$$\times\left\{1+\frac{1}{2!}\left(\frac{c+1}{2}\right)^2\left(\frac{2}{n-1}\right)^2+\frac{1}{4!}\left(\frac{c+1}{2}\right)^2\left(\frac{c+3}{2}\right)^2\left(\frac{2}{n-1}\right)^4+\frac{1}{6!}\left(\frac{c+1}{2}\right)^2\left(\frac{c+3}{2}\right)^2\left(\frac{c+5}{2}\right)^2\left(\frac{2}{n-1}\right)^6+\ldots\right\}. \tag{4·6}$$

For the third moment we write

$$M_3'\{v(c)\} = E\{v(c)\}^3 = \frac{n}{n^3}E\,|z_1|^{3c} + \frac{3n(n-1)}{n^3}E\,|z_1|^{2c}\,|z_2|^c + \frac{n(n-1)(n-2)}{n^3}E\,|z_1|^c\,|z_2|^c\,|z_3|^c$$

$$= A_1 + A_2 + A_3, \tag{4·7}$$

denoting the three terms on the right by $A_1$, $A_2$, $A_3$ respectively. Then

$$A_1 = \frac{1}{\sqrt{\pi}}\left(\frac{3c-1}{2}\right)!(2\overline{n-1})^{\frac{3}{2}c}n^{-\frac{1}{2}(4+3c)},$$

$$A_2 = \frac{3.2^{\frac{3}{2}c}}{\pi}\left(\frac{2c-1}{2}\right)!\left(\frac{c-1}{2}\right)!(n-2)^{\frac{1}{2}(3c+1)}(n-1)^{-\frac{3}{2}c}n^{-\frac{3}{2}}$$

$$\times\left\{1+\frac{(2c+1)(c+1)}{2!(n-1)^2}+\frac{(2c+3)(2c+1)(c+3)(c+1)}{4!(n-1)^4}+\ldots\right\},$$

$$A_3 = \left(\frac{2^c}{\pi}\right)^{\frac{3}{2}}(n-3)^{\frac{1}{2}(3c+2)}(n-2)^{-\frac{1}{2}(3c+1)}(n-1)\,n^{-\frac{3}{2}}\left[\left(\frac{c-1}{2}\right)!\right]^3\left\{1+\frac{3(c+1)^2}{2(n-2)^2}\right.$$

$$-\frac{(c+1)^3}{(n-2)^3}+\frac{(c+1)^2(c+3)(7c+9)}{8(n-2)^4}-\frac{(c+3)^2(c+1)^3}{2(n-2)^5}$$

$$\left.+\frac{(c+3)^2(c+1)^2(61c^2+310c+265)}{240(n-2)^6}+\ldots\right\}.$$

Similarly, for the fourth moment,

$$M_4'\{v(c)\} = E\{v(c)\}^4 = \frac{n}{n^4}E\,|z_1|^{4c} + \frac{4n(n-1)}{n^4}E\,|z_1|^{3c}\,|z_2|^c$$

$$+\frac{3n(n-1)}{n^4}E\,|z_1|^{2c}\,|z_2|^{2c}+\frac{6n(n-1)(n-2)}{n^4}E\,|z_1|^{2c}\,|z_2|^c\,|z_3|^c$$

$$+\frac{n(n-1)(n-2)(n-3)}{n^4}E\,|z_1|^c\,|z_2|^c\,|z_3|^c\,|z_4|^c$$

$$= C_1 + C_2 + C_3 + C_4 + C_5 \tag{4·8}$$

with

$$C_1 = \frac{2^{2c}}{\sqrt{\pi}}(n-1)^{2c}n^{-2c-3}\left(\frac{4c-1}{2}\right)!,$$

$$C_2 = \frac{2^{2(c+1)}}{\pi}(n-2)^{\frac{1}{2}(4c+1)}(n-1)^{-2c}n^{-\frac{5}{2}}\left(\frac{3c-1}{2}\right)!\left(\frac{c-1}{2}\right)!$$

$$\times\left\{1+\frac{(3c+1)(c+1)}{2!(n-1)^2}+\frac{(3c+3)(3c+1)(c+3)(c+1)}{4!(n-1)^4}+\ldots\right\},$$

$$C_3 = \frac{3.2^{2c}}{\pi}(n-2)^{\frac{1}{2}(4c+1)}(n-1)^{-2c}n^{-\frac{5}{2}}\left[\left(\frac{2c-1}{2}\right)!\right]^2\left\{1+\frac{(2c+1)^2}{2!(n-1)^2}+\frac{(2c+3)^2(2c+1)^2}{4!(n-1)^4}+\dots\right\},$$

$$C_4 = \frac{3.2^{2c+1}}{\pi^{\frac{3}{2}}}(n-3)^{2c+1}(n-2)^{-\frac{1}{2}(4c+1)}(n-1)\,n^{-\frac{5}{2}}\left(\frac{2c-1}{2}\right)!\left[\left(\frac{c-1}{2}\right)\right]!^2$$

$$\times\left\{1+\frac{(c+1)(5c+3)}{2(n-2)^2}-\frac{(c+1)^2(2c+1)}{(n-2)^3}+\frac{(c+1)(57c^3+227c^2+255c+81)}{24(n-2)^4}\right.$$

$$\left.-\frac{(2c+1)(c+1)^2(c+3)(5c+9)}{6(n-2)^5}+\dots\right\},$$

$$C_5 = \frac{2^{2c}}{\pi^2}(n-4)^{\frac{1}{2}(4c+5)}(n-3)^{-2c-1}(n-2)(n-1)\,n^{-\frac{5}{2}}\left[\left(\frac{c-1}{2}\right)!\right]^4$$

$$+\left\{1+\frac{3(c+1)^2}{(n-3)^2}-\frac{4(c+1)^3}{(n-3)^3}+\frac{(c+1)^2(7c^2+21c+15)}{(n-3)^4}-\frac{4(c+3)(c+1)^3(2c+3)}{(n-3)^5}\right.$$

$$\left.+\frac{(c+3)(c+1)^2(122c^3+671c^2+1070c+525)}{15(n-3)^6}-\dots\right\}.$$

Formulae (4·5), (4·6), (4·7) and (4·8) were checked from the corresponding formulae for $c = 1$ given in the author's 1936 paper.

From the following section it will be apparent that for indefinitely large samples the most sensitive test of kurtosis of the field $a(c)$ is found for $c = 4$. At the same time it is shown that there is really not much difference in efficiency for values of $c$ in the range $5 \geqslant c > 2$; moreover, the results in § 6 (in which the efficiency of the tests for $c = 4$ and $c = 1$ are compared from the power function viewpoint) suggest that, for samples of moderate size, the superiority, if any at all, of a test using $a(4) = b_2$ over other tests in the series may be even less marked. The disadvantage of $a(4)$ is that its frequency is not known for samples of all sizes; and if we could estimate, with any degree of confidence, the probability points of $a(c)$ for any value or values of $c > 2$ for medium-size samples we might, for practical purposes, dispense with $a(4)$ altogether, since, while we now know one way of solving the problem of determining the exact, or almost exact, frequency distribution of $a(4)$, it must be admitted that the method is extremely tedious. (From the theoretical point of view, however, the $a(4)$ problem must be solved since it remains a challenge to the mathematical skill of statisticians!) It will accordingly be of interest to study the order of magnitude of the semi-invariants of $a(c)$ for $c$ near 2.

Consider the case, for example, of $c = 2{\cdot}4$, not by any means, it is important to observe, the lowest value which would be used for tabulating. In Table 4 the first three moments are given for $n = 25$. The $L$'s represent, of course, the semi-invariants. The values of the functions for $a_1(c)$ (given by (3·2)) for $n = 24$ (i.e. the appropriate number of degrees of freedom for comparison with $a(c)$) are also given. These show that the moments of $a_1(c)$ are very close to those of $a(c)$, which suggests that, when $n$ is not less than, say, 20, the values of $B_1$, $B_2$ and corresponding functions of higher orders, if required, for $a_1(c)$ could be used for the determination of the probability points of $a(c)$. This is important from the computational point of view because the algebraic expressions for the normal moments of $a_1(c)$ are exceedingly simple whereas it must be conceded that (4·8) offers a grim prospect for the computer; furthermore, the principal term $C_5$ is rather slowly convergent unless $n > 50$ or so, whereas *exact* values for all values of $n$ can readily be found for the moments of $a_1(c)$ for normal samples.

Table 4. *Normal moments, etc., of $a(c)$ and $a_1(c)$ for $c = 2{\cdot}4$*

| | $a(2{\cdot}4)$ | $a_1(2{\cdot}4)$ |
|---|---|---|
| $n$ | 25 | 24 |
| $M_1'=L_1$ | 1·166252 | 1·1662524891 |
| $M_2'$ | 1·362004 | 1·362091186 |
| $M_3'$ | 1·592841 | 1·593151615 |
| $M_2=L_2$ | 0·001860 | 0·001946318 |
| $M_3=L_3$ | 0·000063 | 0·000069583 |
| $\sqrt{B_1}=L_3/L_2^{\frac{3}{2}}$ | 0·80 | 0·8104 |

As with (4·1) for $a(c)$, the moments (from the origin) of any order of $a_1(c)$ is the quotient of the moments of the same order for numerator and denominator, assuming that the universal mean is zero and the variance unity. Since the different members $x_i$ of the sample are independent—the difficulty with $a(c)$ is that the $(x_i-\bar{x})$ are *not* independent—for the moments of the numerator of (3·2) we require only

$$E\,|\,x\,|^{k'} = \frac{2}{\sqrt{(2\pi)}}\int_0^\infty dx\, x^{k'}\, e^{-\frac{1}{2}x^2} = \left(\frac{k'-1}{2}\right)!\,\frac{2^{\frac{1}{2}k'}}{\sqrt{\pi}}, \tag{4·9}$$

and for the denominator

$$Es^{k'} = E\left(\frac{1}{n}\sum_i x_i^2\right)^{\frac{1}{2}k} = \left(\frac{2}{n}\right)^{\frac{1}{2}k'}\left(\frac{n+k'-2}{2}\right)!\bigg/\left(\frac{n-2}{2}\right)!. \tag{4·10}$$

The case of $c = 4$ is particularly simple. The first four semi-invariants are as follows:

$$\left.\begin{aligned}
L_1 &= M_1' = \frac{3n}{(n+2)},\\
L_2 &= M_2 = \frac{24n^2(n-1)}{(n+2)^2(n+4)(n+6)},\\
L_3 &= M_3 = \frac{1728(n-1)(n-2)n^3}{(n+2)^3(n+4)(n+6)(n+8)(n+10)},\\
L_4 &= \frac{10{,}368n^4(n-1)(30n^4+168n^3-608n^2-2672n+3712)}{(n+2)^4(n+4)^2(n+6)^2(n+8)(n+10)(n+12)(n+14)}.
\end{aligned}\right\} \tag{4·11}$$

Moments, etc., for $a_1(c)$ for normal samples of 24 and 50 are contrasted for $c = 2{\cdot}4$ and $c = 4$ in Table 5. The contrast between the values of $\sqrt{B_1}$ and $(B_2-3)$ respectively for $a_1(2{\cdot}4)$ and $a_1(4)$ is striking in the extreme. Even for $n = 24$ $\sqrt{B_1}\,[a_1(2{\cdot}4)]$ and $B_2\,[a_1(2{\cdot}4)]$ are approaching the values at which a Gram-Charlier approximation to the frequency distribution may be reasonably convergent. Furthermore, the decline in the values of the $B$'s from $n = 24$ to $n = 50$ is marked for $a_1(2{\cdot}4)$, while the decline in the $B[a_1(4)]$ is very slow.

It is accordingly suggested that a table of probability points (perhaps 0·001, 0·01, 0·025, 0·05 and 0·10) of $a(c)$, for $c$ equal to, say, 2·2, be prepared for $n \geqslant 25$ on the assumption that Gram-Charlier applies throughout. For this purpose the values of the mean and variance for $n$ at intervals of, say, 10 should be computed from formulae (4·5) and (4·6); the $B_1$ and $(B_2-3)$ should, however, be computed as for $a_1(c)$. For lower sample sizes it might be well to use terms to order $n^{-2}$ which would render necessary the use of the fifth and sixth semi-invariants of $a_1(c)$. The formulae given by E. A. Cornish & R. A. Fisher (1937) (assuming Gram-Charlier) could be used to find the probability points. On account of the minuteness of the variance $L_2$ for $c$ near 2 it will be necessary to work to many places of decimals—at least 10. As stated at the outset, the test of kurtosis $a(2{\cdot}2)$ will be only slightly less efficient than $a(4)$ and it may be slightly more efficient than $a(1)$, the probability points of which are known approximately for samples of all sizes. In any case the $a(2{\cdot}2)$ table would be a useful adjunct to that of $a(1)$.

Table 5. *Normal moments, etc., of $a_1(c)$ for $c = 2{\cdot}4$ and $c = 4$*

| | $n=24$ | | $n=50$ | |
|---|---|---|---|---|
| | $c=2{\cdot}4$ | $c=4$ | $c=2{\cdot}4$ | $c=4$ |
| $M_1' = L_1$ | 1·1662524891 | 2·769231 | 1·1721603127 | 2·884615 |
| $M_2 = L_2$ | 0·001946318 | 0·559932 | 0·001058462 | 0·359550 |
| $M_3 = L_3$ | 0·000069583 | 0·752488 | 0·000022251 | 0·343337 |
| $L_4 = M_4 - 3L_2^2$ | 0·000004921 | 1·955999 | 0·000000919 | 0·711375 |
| $\sqrt{B_1} = L_3/L_2^{\frac{3}{2}}$ | 0·8104 | 1·7960 | 0·6462 | 1·5925 |
| $B_2 - 3 = L_4/L_2^2$ | 1·30 | 6·24 | 0·82 | 5·50 |

In an earlier paper (1935) the writer suggested that the correlation between $b_2$ and $a(1)$ for normal samples gave some indication of the relative efficiency of these two tests of normality. In this order of ideas it seems desirable to compute the approximate value of the correlation coefficient between $a(c)$ and $a(c')$, where $c$ and $c'$ are any two positive constants. In the first instance the universe from which the sample of $n$ was drawn was not necessarily normal. Since in the present application we will be concerned only with large samples we assume the universal mean known (and accordingly it may be taken as zero, i.e. $\lambda_1 = 0$), so that, instead of $a(c)$ we use, in reality, $a_1(c)$ given by (3·2). In the remainder of this section we write $a$ for $a_1(c)$ and $a'$ for $a_1(c')$:

$$a = \left(\frac{1}{n}\Sigma\,|\,x_i\,|^c\right)\bigg/\left(\frac{1}{n}\Sigma x_i^2\right)^{\frac{1}{2}c}, \qquad (4{\cdot}12)$$

$$a' = \left(\frac{1}{n}\Sigma\,|\,x_i\,|^{c'}\right)\bigg/\left(\frac{1}{n}\Sigma x_i^2\right)^{\frac{1}{2}c'}. \qquad (4{\cdot}13)$$

Set

$$\left.\begin{aligned} y_i &= (|\,x_i\,|^c - \mu_{|c|})/\mu_{|c|}, \\ y_i' &= (|\,x_i\,|^{c'} - \mu_{|c'|})/\mu_{|c'|}, \\ z_i &= (x_i^2 - \mu_2)/\mu_2, \\ \alpha &= \mu_{|c|}/\mu_2^{\frac{1}{2}c}, \quad \alpha' = \mu_{|c'|}/\mu_2^{\frac{1}{2}c'}, \\ C &= \frac{c+c'}{2}, \quad C_k = \frac{C(C+1)(C+2)\dots(C+k-1)}{k!}. \end{aligned}\right\} \qquad (4{\cdot}14)$$

Then

$$\frac{aa'}{\alpha\alpha'} = \left(1+\frac{1}{n}\Sigma y_i\right)\left(1+\frac{1}{n}\Sigma y_i'\right)\left(1+\frac{1}{n}\Sigma z_i\right)^{-\frac{1}{2}(c+c')} \qquad (4{\cdot}15)$$

The mean value of $aa'/\alpha\alpha'$ was found approximately (i.e. to terms in $n^{-3}$) by formally expanding the last factor in (4·15), multiplying by the first two factors, and setting down the mean value term by term, so that

$$M'_{cc'}/\alpha\alpha' = Eaa'/\alpha\alpha' \simeq \left\{1+\frac{1}{n^2}C_2 nEz^2 - \frac{1}{n^3}C_3 nEz^3\right.$$

$$+\frac{1}{n^4}C_4\left(nEz^4+\frac{6n\,\overline{n-1}}{2}E^2z^2\right)-\frac{1}{n^5}C_5(10n\,\overline{n-1}\,Ez^3Ez^2)$$

$$\left.+\frac{1}{n^6}C_6\,90\,\frac{n\,\overline{n-1}\,\overline{n-2}}{6}E^3z^2\right\}+\left\{-\frac{C_1}{n^2}(nEyz+nEy'z)+\frac{C_2}{n^3}(nEyz^2+nEy'z^2)\right.$$

$$-\frac{C_3}{n^4}[n(Eyz^3+Ey'z^3)+3n\,\overline{n-1}\,Ez^2(Eyz+Ey'z)]$$

$$+\frac{C_4}{n^5}[4n\,\overline{n-1}\,Ez^3(Eyz+Ey'z)+6n\,\overline{n-1}\,Ez^2(Eyz^2+Ey'z^2)]$$

$$\left.-\frac{30C_5}{n^6}\frac{n\,\overline{n-1}\,\overline{n-2}}{2}E^2z^2(Eyz+Ey'z)\right\}+\left\{\frac{1}{n^2}nEyy'-\frac{C_1}{n^3}nEyy'z\right.$$

$$+\frac{C_2}{n^4}[nEyy'z^2+2n\,\overline{n-1}\,Eyz\,Ey'z+n\,\overline{n-1}\,Eyy'\,Ez^2]$$

$$-\frac{C_3}{n^5}[n\,\overline{n-1}\,Eyy'\,Ez^3+3n\,\overline{n-1}\,(Eyz^2Ey'z+Ey'z^2Eyz+Eyy'zEz^2)]$$

$$\left.+\frac{C_4}{n^6}\left[\frac{6n\,\overline{n-1}\,\overline{n-2}}{2}Eyy'E^2z^2+12n\,\overline{n-1}\,\overline{n-2}\,EyzEy'zEz^2\right]\right\}. \quad (4·16)$$

The $E$'s in (4·16) are readily calculable from (4·14), e.g.

$$Eyy' = Ey_iy'_i = E(|x_i|^c-\mu_{|c|})(|x_i|^{c'}-\mu_{|c'|})/\mu_{|c|}\mu_{|c'|} = (\mu_{|c+c'|}/\mu_{|c|}\mu_{|c'|})-1.$$

It has been verified that when $c$ is substituted for $c'$ in (4·13) the formula agrees with that for the second moment of $a_1(c)$ given in § 6.

The coefficient of correlation is, of course,

$$R_{cc} = M_{cc'}/\sqrt{(M_{cc}M_{c'c'})}, \quad (4·17)$$

with

$$M_{cc'} = M'_{cc'}-M'_cM'_{c'}.$$

Formulae for the first and second moments, to the approximation required, for the computation of (4·17) are given in § 6.

As an application, the following are the values of the variances and the covariance for the test of normality $a(1)$ and $(b_2)$, i.e. in which $c$ and $c'$ have respectively the values 1 and 4, and where the universe belongs to the Pearson system with $\lambda_2 = 1$, $\lambda_3 = 0$ and $\lambda_4 = \frac{1}{2}$:

$$\left.\begin{aligned}\frac{M_{cc}}{\mu^2_{|c|}} &\simeq \frac{0·09313705}{n}-\frac{0·262961}{n^2}-\frac{0·196477}{n^3},\\ \frac{M_{c'c'}}{\mu^2_{|c'|}} &\simeq \frac{4·4286}{n}-\frac{92·25}{n^2}+\frac{831·2}{n^3},\\ \frac{M_{cc'}}{\mu_{|c|}\mu_{|c'|}} &\simeq -\frac{0·491}{n}+\frac{4·87}{n^2}-\frac{281·5}{n^3}.\end{aligned}\right\} \quad (4·18)$$

From (4·17) and (4·18), $R_{cc'}(n=100) \simeq -0{\cdot}826$ and $R_{cc'}(n=\infty) = -0{\cdot}764$. It is of great interest to find that, though the universe is markedly non-normal the correlation for indefinitely large samples is practically identical with the normal theory value of $-0{\cdot}767$ (Geary, 1935), another indication, no doubt, that normal theory inferences can usually be applied with confidence when the parent universe is not markedly unsymmetrical.

When samples are indefinitely large we find, from (4·16) and (4·17),

$$R_{cc'} = \frac{4\mu_{|c+c'|} - 2(c\mu_{|c|}\mu_{|c'+2|} + c'\mu_{|c'|}\mu_{|c+2|}) + (cc'\mu_4 - \overline{c-2}\,.\,\overline{c'-2})\,\mu_{|c|}\mu_{|c'|}}{\sqrt{(M_{cc}M_{c'c'})}}, \tag{4·19}$$

where, of course, the values to be taken here for $M_{cc}$ and $M_{c'c'}$ are found by substituting respectively $c'$ for $c$ and $c$ for $c'$ in the numerator. When, in addition, the parent universe is normal, we find

$$R^0_{cc'} = \frac{\left(\frac{c+c'-1}{2}\right)!\sqrt{\pi} - \left(\frac{c-1}{2}\right)!\left(\frac{c'-1}{2}\right)!\left(\frac{cc'+2}{2}\right)}{\sqrt{\left[\left\{\left(\frac{2c-1}{2}\right)!\sqrt{\pi} - \left(\frac{c-1}{2}!\right)^2\left(\frac{c^2+2}{2}\right)\right\}\left\{\left(\frac{2c'-1}{2}\right)!\sqrt{\pi} - \left(\frac{c'-1}{2}!\right)^2\left(\frac{c'^2+2}{2}\right)\right\}\right]}} \tag{4·20}$$

which reduces to $-1/\sqrt{\{12(\pi-3)\}}$ for $c = 1$, $c' = 4$, as it should (Geary, 1935). The following section will accord $b_2$ (i.e. $a(4)$) a decided primacy amongst tests of normality when the samples are indefinitely large. It may, therefore, be of interest to give the values of the correlation coefficients (for indefinitely large normal samples) between $b_2$ and $a(c)$ for selected values of $c$ (Table 6). The table suggests, in the high coefficients of correlation, except for $c$ very near 0 or 2, that all the $a(c)$ should be reliable tests of kurtosis, with no great difference between their efficiencies. The efficiency of any two tests would be identical, in the conditions stated, if the coefficient of correlation between them was $\pm 1$ because then, of course, they would be functionally, and not stochastically, related.

Table 6. *Correlation between $b_2$ and $a(c)$ for indefinitely large normal samples*

| Value of $c$ | Value of $R^0_{c4}$ | Value of $c$ | Value of $R^0_{c4}$ |
|---|---|---|---|
| 0 | 0 | 3 | 0·980 |
| 1 | −0·769 | 4 | 1 |
| 2 | 0 | 5 | 0·983 |
| 2·2 | 0·887 | 6 | 0·939 |
| 2·5 | 0·952 | ∞ | 0 |

## 5. The most efficient tests for indefinitely large samples

In this section we consider the efficiency of tests of kurtosis and asymmetry from the viewpoint of indefinitely large samples.

By definition a test will be regarded as *valid*, in relation to a field of continuous alternative universes including the normal, if its value for infinite samples drawn at random from the normal universe is different from its value for infinite samples from other universes of the field. As the sample number increases the test will become increasingly discriminatory of the normal as distinct from other universes of the field. This increased sensitivity might be given mathematical expression in some such terms as the following: given a probability $\alpha$ (say 0·01), the normal universe $W_0$ of the field and any other distribution $W_1$ of the field, a number $n_1$ can be found so that for $n \geqslant n_1$ the mean value of the test function for samples of $n$ from $W_1$ will lie at or beyond the $\alpha$ probability point of the test function for samples of $n$ from $W_0$: the smaller $n_1$ the more sensitive the test.

We consider, then, the infinite field of alternative tests of kurtosis represented by (3·1) when $c$ assumes all positive values, and the infinite field of alternative universes represented by the Gram-Charlier frequency

$$\frac{1}{\sqrt{(2\pi)}} \exp\left\{\sum_{i=3}^{\infty} \frac{\lambda_i}{i!}\left(-\frac{d}{dx}\right)^i\right\} e^{-\frac{1}{2}x^2}. \tag{5·1}$$

The universal variance is assumed to be unity, without loss of generality. The normal universe is a member of the field: it is found when all the $\lambda_i$ $(i > 2)$ are zero. We assume that the conditions of § 2 are satisfied so that for indefinitely large samples the frequency distribution of $a(c)$ for all parent universes is normal. Obviously the efficiency of any particular test (i.e. $a(c)$ for a particular value of $c$) in regard to the normal and a particular non-normal alternative (i.e. a Gram-Charlier frequency with particular values of the $\lambda_i$) will be adjudged by considering the ratio of

(i) the difference between the universal mean values of $a(c)$ for the normal and the particular non-normal parent universes; to

(ii) the standard deviation of $a(c)$ for indefinitely large normal samples.

The most efficient test will be $a(c)$ for $c$ a theoretically ascertainable function of the given $\lambda_i$ which makes the ratio a maximum.

For indefinitely large samples the mean value $\phi$ of $a(c)$ when the parent universe is given by (5·1) is

$$\phi = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty} dx\,|x|^c \exp\left\{\sum_i \frac{\lambda_i}{i!}\left(-\frac{d}{dx}\right)^i\right\} e^{-\frac{1}{2}x^2}. \tag{5·2}$$

Obviously $$\int_{-\infty}^{\infty} dx\,|x|^c\left(-\frac{d}{dx}\right)^{2m+1} e^{-\frac{1}{2}x^2} = 0.$$

Also, when $m \geqslant 1$,

$$\int_{-\infty}^{\infty} dx\,|x|^c\left(\frac{d}{dx}\right)^{2m} e^{-\frac{1}{2}x^2} = \left(\frac{c-1}{2}\right)!\,2^{\frac{1}{2}(c+1)}c(c-2)(c-4)\ldots(c-2m+2), \tag{5·3}$$

a result readily inferable from the obvious fact that the left side vanishes for $c = 0, 2, \ldots, 2m-2$. Accordingly

$$\phi = \left(\frac{c-1}{2}\right)!\frac{2^{\frac{1}{2}(c+1)}}{\sqrt{(2\pi)}}\left\{1 + \frac{\lambda_4}{24}c(c-2) + \left(\frac{\lambda_3^2}{72} + \frac{\lambda_6}{720}\right)c(c-2)(c-4) + \ldots\right\}. \tag{5·4}$$

The normal value is given by the first term.

From (4·3), (4·5) and (4·6) it is evident that the value of the standard deviation, for larger normal samples (retaining only $n^{-\frac{1}{2}}$) is

$$\sigma = \frac{2^{\frac{1}{2}c}}{\sqrt{(\pi n)}}\left(\left(\frac{2c-1}{2}\right)!\sqrt{\pi} - \left(\frac{c-1}{2}!\right)^2\frac{c^2+2}{2}\right)^{\frac{1}{2}}. \tag{5·5}$$

The principal term in the deviation $\phi - \phi^0$ (where $\phi^0$ is the normal value), from (5·4), is

$$\delta = \frac{\frac{1}{2}(c-1)!\,2^{\frac{1}{2}c}}{\sqrt{\pi}}\cdot\frac{\lambda_4 c(c-2)}{24}. \tag{5·6}$$

To a constant factor, the ratio $\delta/\sigma$ is given by the *first discriminant*

$$\rho(c) = c(c-2)\left\{\frac{\left(\frac{2c-1}{2}\right)!\sqrt{\pi}}{\left(\frac{c-1}{2}!\right)^2} - \frac{c^2+2}{2}\right\}^{-\frac{1}{2}} \tag{5·7}$$

It will now be shown that $\frac{d\rho(c)}{dc} = 0$ for $c = 4$.

The discriminant may be written in the form

$$\rho(c) = c(c-2)\left(\frac{2^c I_{2c}}{I_c} - \frac{c^2+2}{2}\right)^{-\frac{1}{2}}, \tag{5·8}$$

where

$$I_c = \int_0^{\frac{1}{2}\pi} \cos^c\theta\, d\theta, \tag{5·9}$$

and

$$\frac{\rho'(c)}{\rho(c)} = \frac{1}{c} + \frac{1}{c-2} - \frac{1}{2}\left\{2^c\left(\frac{I_{2c}\log 2}{I_c} + \frac{I'_{2c}}{I_c} - \frac{I_{2c}I'_c}{I_c^2}\right) - c\right\}\bigg/\left\{\frac{2^c I_{2c}}{I_c} - \frac{c^2+2}{2}\right\}. \tag{5·10}$$

From (5·9)

$$J_c = I'_c = \int_0^{\frac{1}{2}\pi} d\theta \log^c\theta \log\cos\theta. \tag{5·11}$$

From a fairly well-known property

$$J_0 = \int_0^{\frac{1}{2}\pi} d\theta \log\cos\theta = -\tfrac{1}{2}\pi\log 2. \tag{5·12}$$

In (5·10) we shall be concerned only with even positive integer values of $c$. We have at once

$$I_0 = \frac{\pi}{2},\quad I_2 = \frac{\pi}{4},\quad I_4 = \frac{3\pi}{16},\quad I_6 = \frac{5\pi}{32},\quad I_8 = \frac{35\pi}{256}. \tag{5·13}$$

From (5·11)

$$J_{2c} = \int_0^{\frac{1}{2}\pi} d\theta \cos^{2c}\theta \log\cos\theta = \int_0^{\frac{1}{2}\pi} d(\sin\theta)\cos^{2c-1}\theta\log\cos\theta,$$

which, by partial integration,

$$= \int_0^{\frac{1}{2}\pi} d\theta \sin\theta\left(\overline{2c-1}\sin\theta\cos^{2c-2}\theta\log\cos\theta + \frac{\cos^{2c-1}\theta\sin\theta}{\cos\theta}\right)$$

$$= (2c-1)(J_{2c-2} - J_{2c}) + I_{2c-2} - I_{2c}.$$

Hence

$$2cJ_{2c} = (2c-1)J_{2c-2} - I_{2c} + I_{2c-2}. \tag{5·14}$$

From (5·12), (5·13) and (5·14),

$$\left.\begin{aligned} &J_0 = -\tfrac{1}{2}\pi\log 2, & &J_6 \doteq (-60\pi\log 2 + 37\pi)/384,\\ &J_2 = (-2\pi\log 2 + \pi)/8, & &J_8 = (-840\pi\log 2 + 533\pi)/6144.\\ &J_4 = (-12\pi\log 2 + 7\pi)/64, & & \end{aligned}\right\} \tag{5·15}$$

Noting that $I'_{2c} = 2J_{2c}$ and substituting in the right side of (5·10) the values of $I$ and $J$ given by (5·13) and (5·15), we find $\rho'(4) = 0$. Table 7 gives the values of the discriminant for certain values of $c$.

The discriminant accordingly assumes a maximum value for $c = 4$, a result so remarkable that one might be inclined to suspect that it is a consequence of the form which was assumed for the alternative to the normal curve, a form which, in placing such emphasis on $\lambda_4$, high-lights, so to speak, $b_2$ ($= \lambda_4+3$ when $\lambda_2 = 1$ for indefinitely large samples) as a test of normality. From the algebraic point of view this is anything but obvious: the property emerges from quite a complicated piece of algebra. It may also be emphasized that the field of alternatives (5·1) is not arbitrary; it is a general form of frequency distribution when all the $\lambda_i$ are finite. Admittedly the discriminant takes account only of the term in $\lambda_4$ in the expansion; but this is certainly the most significant term for a wide class of frequency distributions, namely, those of homogeneous symmetrical functions of samples of $n$ as $n$ tends towards infinity under very general conditions for the parent universe, provided that the resulting frequency distribution can be assumed to have its third moment zero; for then the only term in $n^{-1}$ in the frequency distribution of the function will be the term in $\lambda_4$. The significance of the property demonstrated must not be overstressed since it is subject to many qualifications, but it gives strong grounds for holding that, for very large samples, $b_2$ is the most efficient test of normality of tests of type $a(c)$ in relation to a very extended class of alternative universes. At the same time Table 7 shows that there can be little difference in efficiency in the field $a(c)$ for $c$ ranging from close to 2 to about 5. There is but little doubt, on this showing, that $b_2$ is more sensitive than $a(1)$, a conclusion suggested on the basis of certain experimental results by E. S. Pearson (1935) and examined from the viewpoint of power function theory in § 6.

Table 7

| $0<c<2$ | Discriminant $\rho(c)$ | $2<c<\infty$ | Discriminant $\rho(c)$ |
|---|---|---|---|
| +0 | −2·334 | 2+0 | 4·460 |
| 0·1 | −2·541 | 2·1 | 4·508 |
| 0·2 | −2·725 | 2·5 | 4·666 |
| 0·5 | −3·188 | 3·0 | 4·801 |
| 0·7 | −3·441 | 3·9 | 4·898 |
| 1·0 | −3·758 | **4·0** | **4·900** |
| 1·1 | −3·851 | 4·1 | 4·898 |
| 1·5 | −4·166 | 5·0 | 4·818 |
| 1·9 | −4·405 | 6·0 | 4·602 |
| 2−0 | −4·460 | 7·0 | 4·288 |
| | | 8·0 | 3·906 |

Adverting to (5·4) in conjunction with (5·5), it might be asked if, on the analogy of the maximal property just demonstrated for the first discriminant, the function

$$\rho_2(c) = c(c-2)(c-4)\left\{\frac{\left(\frac{2c-1}{2}\right)!\sqrt{\pi}}{\left(\frac{c-1}{2}!\right)^2}-\frac{c^2+2}{2}\right\}^{-\frac{1}{2}}$$

has a turning point at $c = 6$. The answer is in the negative. The value of $\rho_2'(6)/\rho_2(6)$ is, in fact, 15/34. At the same time there must be a zero of $\rho_2'(c)$ very near $c = 6$ since

$$\rho_2(5{\cdot}9) = 8{\cdot}79, \quad \rho_2(6) = 9{\cdot}20, \quad \rho_2(6{\cdot}1) = 8{\cdot}56.$$

Analogous to the field on tests of kurtosis represented by (3·1) we may consider as a field of tests of asymmetry:

$$g(c) = \frac{1}{n}\{-\Sigma'\,|\,x_i-\bar{x}\,|^c+\Sigma''(x_i-\bar{x})^c\}\Big/\left\{\frac{1}{n}\Sigma(x_i-\bar{x})^2\right\}^{\frac{1}{2}c}, \qquad (5{\cdot}16)$$

where $\Sigma'$ extends to the observations $x_i$ less than the mean $\bar{x}$ and $\Sigma''$ to the rest of the sample. For $c = 3$ the test is, of course, $\sqrt{b_1}$. For normal samples

$$E\{g(c)\}^k = E\left\{-\frac{1}{n}\Sigma'\,|x_i-\bar{x}|^c+\frac{1}{n}\Sigma''(x_i-\bar{x})^c\right\}^k \Big/ E\left\{\frac{1}{n}\Sigma(x_i-\bar{x})^2\right\}^{\frac{1}{2}kc}, \qquad (5\cdot17)$$

the denominator of which is identical with the denominator of (4·1). Knowing the joint distribution (for normal samples) of $(x_1-\bar{x})$, $(x_2-\bar{x})$, ... (Geary, 1936), there is no theoretical difficulty in finding the mean values of the terms of the numerator for positive integer values of $k$. Here we shall be concerned only with the first and second moments, i.e. those for (5·17) for $k = 1$ and $k = 2$. We require the normal distribution of $z_1 = x_1-\bar{x}$ and the joint distribution of $z_1$ and $z_2 = x_2-\bar{x}$. These are

$$(z_1):\quad \left(\frac{n}{2\pi\,\overline{n-1}}\right)^{\frac{1}{2}} \exp\left\{-\frac{nz_1^2}{2(n-1)}\right\} dz_1,$$

$$(z_1, z_2):\quad \frac{1}{2\pi}\left(\frac{n}{n-2}\right)^{\frac{1}{2}} \exp\left\{-\frac{(n-1)(z_1^2+z_2^2)}{2(n-2)}-\frac{z_1z_2}{(n-2)}\right\} dz_1\,dz_2 = f(z_1, z_2)\,dz_1\,dz_2. \qquad (5\cdot18)$$

Clearly the odd normal moments of $g(c)$ are zero. Then

$$E\left\{-\frac{1}{n}\Sigma'\,|x_i-\bar{x}|^c+\frac{1}{n}\Sigma''(x_i-\bar{x})^c\right\}^2 = \frac{n}{n^2}E\,|z_1|^{2c}+\frac{n(n-1)}{n^2}E_1(z_1, z_2), \qquad (5\cdot19)$$

where $E_1(z_1, z_2)$ is the mean value of the two-dimensional terms. We then have

$$E_1(z_1, z_2) = \int_{-\infty}^{0}(-z_1)^c dz_1 \int_{-\infty}^{0}(-z_2)^c dz_2 f(z_1, z_2) - \int_{-\infty}^{0} dz_1(-z_1)^c \int_{0}^{\infty} dz_2 z_2^c f(z_1, z_2)$$
$$- \int_{0}^{\infty} dz_1 z_1^c \int_{-\infty}^{0} dz_2(-z_2)^c f(z_1, z_2) + \int_{0}^{\infty} dz_1 z_1^c \int_{0}^{\infty} dz_2 z_2^c f(z_1, z_2)$$
$$= \int_{0}^{\infty}\int_{0}^{\infty} z_1^c z_2^c dz_1 dz_2 \{f(-z_1, -z_2) - f(-z_1, z_2) - f(z_1, -z_2) + f(z_1, z_2)\}$$
$$= -\frac{2^{c+2}}{2\pi}\left(\frac{n}{n-2}\right)^{\frac{1}{2}} \frac{(n-2)^{c+1}}{(n-1)^{c+2}}\left(\frac{c}{2}!\right)^2 \left\{1+\frac{(c+2)^2}{3!(n-1)^2}+\frac{(c+2)^2(c+4)^2}{5!(n-1)^4}+\ldots\right\}, \qquad (5\cdot20)$$

$$Ez_1^{2c} = \frac{2c-1}{2}!\left(\frac{2\overline{n-1}}{n}\right)^c \frac{1}{\sqrt{\pi}}. \qquad (5\cdot21)$$

Also
$$E\left\{\frac{1}{n}\Sigma(x_i-\bar{x})^2\right\}^c = \left(\frac{2}{n}\right)^c \left(\frac{n+2c-3}{2}\right)! \Big/ \left(\frac{n-3}{2}\right)!. \qquad (5\cdot22)$$

We now have all the expressions required for the variance of normal $g(c)$. We require, for what follows, only the term in $n^{-1}$ which is

$$\sigma^2 = \frac{1}{n}\left\{\left(\frac{2c-1}{2}\right)!\frac{2^c}{\sqrt{\pi}} - \left(\frac{c}{2}!\right)^2 \frac{2^{c+1}}{\pi}\right\}. \qquad (5\cdot23)$$

Consider now a field of alternative universes represented by

$$\frac{1}{\sqrt{(2\pi)}}\left\{1+\frac{\lambda_3}{6}(x^3-3x)\right\} e^{-\frac{1}{2}x^2}, \qquad (5\cdot24)$$

the 'first approximation to the law of error' (for universal variance unity), obviously the most appropriate asymmetrical field, for different values of the parameter $\lambda_3$, and containing as a member of the field the normal distribution found for $\lambda_3 = 0$. For indefinitely large samples from (5·24) the mean value of $g(c)$ is

$$\delta = \frac{2\lambda_3}{6\sqrt{(2\pi)}}\int_0^\infty dx\, x^c(x^3-3x)\, e^{-\frac{1}{2}x^2} = \frac{c}{2}!\,(c-1)\,2^{\frac{1}{2}c}\frac{\lambda_3}{3}\sqrt{\frac{1}{2\pi}}. \tag{5·25}$$

From (5·23) and (5·25)

$$\frac{\delta}{\sigma} = \frac{\lambda_3 n^{\frac{1}{2}}}{6}\tau(c), \tag{5·26}$$

the *skew discriminant* $\tau(c)$ being given by

$$\tau(c) = (c-1)\left\{\left(\frac{2c-1}{2}\right)!\left(\frac{c}{2}!\right)^{-2}\frac{\sqrt{\pi}}{2}-1\right\}^{-\frac{1}{2}} = (c-1)\left(\frac{2^{c+1}}{2c+1}\frac{I_{2c+2}}{I_{c+1}}-1\right)^{-\frac{1}{2}}. \tag{5·27}$$

Log-differentiating,

$$\frac{\tau'(c)}{\tau(c)} = \frac{1}{c-1} - \frac{2^{c+1}}{2}\left\{\frac{1}{2c+1}\left(\frac{2J_{2c+2}}{I_{c+1}} - \frac{I_{2c+2}J_{c+1}}{I_{c+1}^2}\right) - \frac{I_{2c+2}}{I_{c+1}}\frac{2}{(2c+1)^2} + \frac{I_{2c+2}}{I_{c+1}}\frac{\log 2}{2c+1}\right\}\left\{\frac{2^{c+1}}{2c+1}\frac{I_{2c+2}}{I_{c+1}}-1\right\}^{-1}. \tag{5·28}$$

Setting $c = 3$ and using (5·13) and (5·15), we find that $\tau'(3) = 0$. Values of $\tau(c)$ for four values of $c$ are as follows:

| $c$ | $\tau(c)$ | $c$ | $\tau(c)$ |
|---|---|---|---|
| 2 | 2·370 | 4 | 2·389 |
| 3 | 2·450 | 5 | 2·236 |

Accordingly, for indefinitely large samples the test of asymmetry $g(c)$ is most efficient for $c = 3$, when the test becomes the familiar $\sqrt{b_1}$. The margin in favour of this value of $c$, as compared with others in the range $2 \leqslant c \leqslant 5$, is, however, quite small.

## 6. Tests of kurtosis from the power function viewpoint

It may be useful to open this section with an interpretation of the results of the previous section from the point of view of the type of error theory of J. Neyman & E. S. Pearson (1933, 1936). For this we consider two universes of the field, the normal $W_0$ and any non-normal universe $W_1$, and two tests of kurtosis $a(4) = b_2$ and $a(c_1)$ for a particular value $c_1$ of $c$. Suppose that samples are sufficiently large that $a(c)$, for samples from all universes of the field, may be regarded as normally distributed.

Given a probability $\alpha$, a sample number $n$ can be found so that the mean value of $a(c_1)$ from $W_1$ lies exactly at, say, the upper $\alpha$ probability point of the distribution of $a(c_1)$ from $W_0$. Then from the results established in the preceding section the value of $a(4)$ for the same sample of $n$ from $W_1$ could lie beyond the $\alpha$ probability point of $a(4)$ for normal samples of $n$. Suppose that the rule adopted was to regard as non-normal all samples for which $a(c)$ lies beyond the normal $\alpha$ probability point, and suppose that a very large number $N$ of samples were drawn, $N_0$ from universes not significantly different from normal (defining 'insignificance' in some manner) and $N_1$ from non-normal universes, so that $N = N_0 + N_1$, where $N_0$ and $N_1$ are not necessarily known in advance. Then using $a(c_1)$ the number of erroneous allocations will be approximately $\alpha N_0 + \frac{1}{2}N_1$, whereas using $a(4)$ the number will be $\alpha N_0 + (\frac{1}{2} - p)\, N_1$ $(\frac{1}{2} > p > 0)$, showing a definite advantage in favour of $a(4)$. The same conclusion emerges whatever value of $c \neq 4$ or whatever non-normal universe be taken for comparison.

The type of error approach reveals the theoretical weakness of using the method of § 5 for the assessment of relative efficiency of tests of normality; namely that the proportion of errors of judgment, even using $a(4)$, remains large, due fundamentally to concentrating on a single value (the mean) as typical or representative of samples from the non-normal universe; it is also a disadvantage that the sample number $n_1$ is necessarily a function of the particular value $c_1$ of $c$. The method has further disadvantages of which the principal are perhaps (i) a somewhat restricted field of alternative universes; (ii) the assumption that the samples were indefinitely large, essential to justify the normality of $a(c)$ for samples from any member of the universe field.

The Neyman-Pearson power function approach which will now be considered cannot be regarded as entirely free from these objections in its application to the material so far available from this research. It enables us, at any rate, to contemplate samples which, if not small, are within the range of experimental practicability.

The problem of the relative efficiency of the different members of a field of tests of kurtosis $a(c)$ will now be considered in its power function aspects. For the present purpose the *power* may be defined as follows:

Given a probability $\alpha$ (say 0·01), a sample number $n$, a particular value $c_1$ of $c$ and a non-normal parent universe $W_1$, the power, in relation to these data, represents the frequency of $a(c_1)$ for samples drawn at random from $W_1$ lying beyond the $\alpha$ probability point for $a(c_1)$ computed from samples drawn from a normal universe. The greater the power the more discriminatory the test. Accordingly, it is in theory necessary to know the frequency distribution of $a(c)$ for all sample sizes, for all values of $c$ and for all universes. Considering that the only frequency distribution of the field contemplated which can be regarded as determined for all sample sizes is $a(1)$ for normal samples (Geary, 1935, 1936), many compromises are necessary to give any kind of practical effect to the power concept. The compromises proposed are as follows:

(1) The form $a_1(c)$, given by (3·2), is used instead of the form $a(c)$ given by (3·1).

(2) Only large samples are dealt with.

(3) The field of alternative universes is restricted.

Using $a_1(c)$, the first four moments (from the origin) of $a_1(c)$ for samples from any universe can be expanded without real difficulty, and so approximate frequency distributions (using the Karl Pearson or Gram-Charlier systems) can be obtained. As to (1), from experiments in $a(1)$ and $a(4)$ the writer has verified that, for medium-sized normal samples, there is little difference between the probability points (e.g. 0·01, 0·05) of $a_1(c)$ and $a(c)$, though the higher semi-invariants (given $n$) are larger for the latter. In regard to (2) and (3) little confidence could be reposed in the values of the moments computed from expansions even to $n^{-3}$ unless the sample number was at least of the order of 100 when $c$ is greater than, say, 3; and, even if the moments were known exactly, the empirical frequencies would be more than doubtful for small samples. The approach finds its main justification in the consideration that any errors due to these necessary compromises may be presumed to apply more or less equally and in the same direction to the tests of kurtosis compared; generous, perhaps too generous, advantage is taken of this justification in the concluding part of this section.

Set, then,
$$a_1(c) = \left\{\frac{1}{n}\Sigma\,|\,x_i\,|^c\right\}\Big/\left\{\frac{1}{n}\Sigma x_i^2\right\}^{\frac{1}{2}c}, \tag{6·1}$$

so that
$$\frac{a_1(c)}{\alpha} = \left(1+\frac{1}{n}\Sigma y_i\right)\left(1+\frac{1}{n}\Sigma z_i\right)^{-\frac{1}{2}c}, \tag{6·2}$$

where
$$\alpha = \mu_{|c|}/\mu_2^{\frac{1}{2}c}, \quad y_i = (|\,x_i\,|^c-\mu_{|c|})/\mu_{|c|}, \quad z_i = (x_i^2-\mu_2)/\mu_2, \tag{6·3}$$

the universal mean being taken as zero, without loss of generality. Raising (6·2) to powers 1, 2, 3, 4, expanding to the required degree the final factor, multiplying by the first factor on the right, and setting down the mean value of each term we find, to $n^{-3}$,

$$\begin{aligned} M_1'/\alpha = 1 &- \frac{1}{n}\{k_1^{(1)}(11) - k_2^{(1)}(02)\} + \frac{1}{n^2}\{k_2^{(1)}(12) - k_3^{(1)}[(03)+3(11)(02)] + 3k_4^{(1)}(02)^2\} \\ &+ \frac{1}{n^3}\{k_3^{(1)}[3(11)(02) - (13)] + k_4^{(1)}[(04) - 3(02)^2 + 4(11)(03) + 6(12)(02)] \\ &- k_5^{(1)}[10(03)(02) + 15(11)(02)^2] + 15k_6^{(1)}(02)\}, \end{aligned} \qquad (6\cdot4)$$

$$\begin{aligned} M_2'/\alpha^2 = 1 &+ \frac{1}{n}\{k_2^{(2)}(02) - 2k_1^{(2)}(11) + (20)\} + \frac{1}{n^2}\{-k_3^{(2)}(03) + 3k_4^{(2)}(02)^2 \\ &+ 2k_2^{(2)}(12) - 6k_3^{(2)}(11)(02) - k_1^{(2)}(21) + k_2^{(2)}(20)(02) + 2k_2^{(2)}(11)^2\} \\ &+ \frac{1}{n^3}\{k_4^{(2)}[(04) - 3(02)^2] - 10k_5^{(2)}(03)(02) + 15k_6^{(2)}(02)^3 - 2k_3^{(2)}[(13) - 3(11)(02)] \\ &+ 4k_4^{(2)}[2(11)(03) + 3(12)(02)] - 30k_5^{(2)}(11)(02)^2 \\ &+ k_2^{(2)}[(22) - (20)(02)] - k_3^{(2)}[(20)(03) + 3(21)(02)] - 2k_2^{(2)}(11)^2 \\ &- 6k_3^{(2)}(12)(11) + 12k_4^{(2)}(11)^2(02) + 3k_4^{(2)}(20)(02)^2\}, \end{aligned} \qquad (6\cdot5)$$

$$\begin{aligned} M_3'/\alpha^3 = 1 &+ \frac{1}{n}\{k_2^{(3)}(02) - 3k_1^{(3)}(11) + 3(20)\} + \frac{1}{n^2}\{-k_3^{(3)}(03) + 3k_4^{(3)}(02)^2 \\ &+ 3k_2^{(3)}(12) - 9k_3^{(3)}(11)(02) - 3k_1^{(3)}(21) + 3k_2^{(3)}(20)(02) + 6k_2^{(3)}(11)^2 \\ &+ (30) - 3k_1^{(3)}(20)(11)\} + \frac{1}{n^3}\{k_4^{(3)}[(04) - 3(02)^2] - 10k_5^{(3)}(03)(02) + 15k_6^{(3)}(02)^3 \\ &- 3k_3^{(3)}[(13) - 3(11)(02)] + 6k_4^{(3)}[2(11)(03) + 3(12)(02)] - 45k_5^{(3)}(11)(02)^2 \\ &+ 3k_2^{(3)}[(22) - (20)(02)] - 3k_3^{(3)}[(20)(03) + 3(21)(02)] + 9k_4^{(3)}(20)(02)^2 \\ &- 6k_2^{(3)}(11)^2 - 18k_3^{(3)}(12)(11) + 36k_4^{(3)}(11)^2(02) - k_1^{(3)}(31) \\ &+ k_2^{(3)}(30)(02) + 3k_1^{(3)}(20)(11) \\ &+ 3k_2^{(3)}[(20)(12) + 2(21)(11)] - 9k_3^{(3)}(20)(11)(02) - 6k_3^{(3)}(11)^3\}, \end{aligned} \qquad (6\cdot6)$$

$$\begin{aligned} M_4'/\alpha^4 = 1 &+ \frac{1}{n}\{k_2^{(4)}(02) - 4k_1^{(4)}(11) + 6(20)\} + \frac{1}{n^2}\{-k_3^{(4)}(03) + 3k_4^{(4)}(02)^2 \\ &+ 4k_2^{(4)}(12) - 12k_3^{(4)}(11)(02) - 6k_1^{(4)}(21) + 6k_2^{(4)}(20)(02) + 12k_2^{(4)}(11)^2 \\ &+ 4(30) - 12k_1^{(4)}(20)(11) + 3(20)^2\} + \frac{1}{n^3}\{k_4^{(4)}[(04) - 3(02)^2] \\ &- 10k_5^{(4)}(03)(02) - 15k_6^{(4)}(02)^3 - 4k_3^{(4)}[(13) - 3(11)(02)] \\ &+ 8k_4^{(4)}[2(11)(03) + 3(12)(02)] - 60k_5^{(4)}(11)(02)^2 + 6k_2^{(4)}[(22) - (20)(02)] \\ &- 6k_3^{(4)}[(20)(03) + 3(21)(02)] + 18k_4^{(4)}(20)(02)^2 - 12k_2^{(4)}(11)^2 \\ &- 36k_3^{(4)}(12)(11) + 72k_4^{(4)}(11)^2(02) - 4k_1^{(4)}(31) + 4k_2^{(4)}(30)(02) \\ &+ 12k_1^{(4)}(20)(11) + 12k_2^{(4)}[(12)(20) + 2(21)(11)] \\ &- 36k_3^{(4)}(20)(11)(02) - 24k_3^{(4)}(11)^3 + (40) - 4k_1^{(4)}(30)(11) - 3(20)^2 \\ &- 6k_1^{(4)}(20)(21) + 3k_2^{(4)}(20)^2(02) + 12k_2^{(4)}(20)(11)^2\}, \end{aligned} \qquad (6\cdot7)$$

where $$k_r^{(p)} = \frac{\frac{1}{2}pc(\frac{1}{2}pc+1)(\frac{1}{2}pc+2)\dots(\frac{1}{2}pc+r-1)}{r!}, \quad (fg) = Ey_i^f z_i^g,$$

the latter, of course, the same for all $i$. The $(fg)$ required for the computation of (6·4)–(6·7) are

$$\left.\begin{aligned}
(11) &= (\mu_{|2+c|} - \mu_2\mu_{|c|})/\mu_2\mu_{|c|},\\
(02) &= (\mu_4 - \mu_2^2)/\mu_2^2,\\
(12) &= (\mu_{|4+c|} - 2\mu_{|2+c|}\mu_2 - \mu_{|c|}\mu_4 + 2\mu_{|c|}\mu_2^2)/\mu_{|c|}\mu_2^2,\\
(03) &= (\mu_6 - 3\mu_4\mu_2 + 2\mu_2^3)/\mu_2^3,\\
(04) &= (\mu_8 - 4\mu_6\mu_2 + 6\mu_4\mu_2^2 - 3\mu_2^4)/\mu_2^4,\\
(13) &= [\mu_{|6+c|} - 3\mu_{|4+c|}\mu_2 + 3\mu_{|2+c|}\mu_2^2 - \mu_{|c|}(\mu_6 - 3\mu_4\mu_2 + 3\mu_2^3)]/\mu_{|c|}\mu_2^3,\\
(21) &= [\mu_{|2c+2|} - 2\mu_{|c+2|}\mu_{|c|} - \mu_2(\mu_{|2c|} - 2\mu_{|c|}^2)]/\mu_{|c|}^2\mu_2,\\
(22) &= (\mu_{|2c+4|} - 2\mu_{|2c+2|}\mu_2 + \mu_{|2c|}\mu_2^2 - 2\mu_{|c+4|}\mu_{|c|} + 4\mu_{|c+2|}\mu_{|c|}\mu_2 - 3\mu_{|c|}^2\mu_2^2 + \mu_{|c|}^2\mu_4)/\mu_{|c|}^2\mu_2^2,\\
(20) &= (\mu_{|2c|} - \mu_{|c|}^2)/\mu_{|c|}^2,\\
(30) &= (\mu_{|3c|} - 3\mu_{|2c|}\mu_{|c|} + 2\mu_{|c|}^3)/\mu_{|c|}^3,\\
(31) &= (\mu_{|3c+2|} - 3\mu_{|2c+2|}\mu_{|c|} + 3\mu_{|c+2|}\mu_{|c|}^2 - \mu_{|3c|}\mu_2 + 3\mu_{|2c|}\mu_{|c|}\mu_2 - 3\mu_{|c|}^3\mu_2)/\mu_{|c|}^3\mu_2,\\
(40) &= (\mu_{|4c|} - 4\mu_{|3c|}\mu_{|c|} + 6\mu_{|2c|}\mu_{|c|}^2 - 3\mu_{|c|}^4)/\mu_{|c|}^4.
\end{aligned}\right\} \qquad (6{\cdot}8)$$

(6·8) is, of course, an immediate consequence of (6·3). The writer has checked the accuracy of formulae (6·4)–(6·7) by reference to the normal universe for $c = 1$.

The reader will have no illusions as to the magnitude of the task of applying the foregoing theory to particular cases. The formulae are set down, however, in the hope that other researchers will be sufficiently sensible of the importance of the theory to assist in building up a fairly extensive set of results. The writer has to be content, in the meantime, to consider the case of the symmetrical universe field given by

$$\frac{1}{\sqrt{(2\pi)}}\left\{1 + \frac{\lambda_4}{24}\left(\frac{d}{dx}\right)^4\right\} e^{-\frac{1}{2}x^2}, \qquad (6{\cdot}9)$$

when $\lambda_4 = \frac{1}{2}$, the normal being given, of course, for $\lambda_4 = 0$, and for $c = 4$ and $c = 1$. These values of $c$ are selected because the theory in § 5 has suggested that $a(4)$ is probably the most efficient of the test-field $a(c)$, while $a(1)$ is the only member of the field for which the normal

Table 8. *Moments from formulae* (6·8)

| $(fg)$ | $c=4$ | | $c=1$ | |
|---|---|---|---|---|
| | Normal | $\lambda_4=\frac{1}{2}$ | Normal | $\lambda_4=\frac{1}{2}$ |
| (11) | 4 | 5·428571 | 1 | 1·17021276 |
| (02) | 2 | 2·5 | 2 | 2·5 |
| (12) | 24 | 45·64286 | 3 | 4·88297871 |
| (03) | 8 | 14 | 8 | 14 |
| (04) | 60 | 138 | 60 | 138 |
| (13) | 216 | 544·2857 | 21 | 44·106383 |
| (21) | 256/3 | 177·71428 | 1·141593 | 1·75544898 |
| (22) | 2,720/3 | 2,481·92857 | 7·707963 | 14·766814 |
| (20) | 32/3 | 16·142857 | 0·570796 | 0·63834981 |
| (30) | 352 | 799·142857 | 0·429204 | 0·6405182 |
| (31) | 4,352 | 12,785·2653 | 3 | 5·236134 |
| (40) | 23,552 | 73,250·178 | — | 2·002492 |

distribution is known for samples of all sizes. The necessary moments ($fg$) given by (6·8) are shown in Table 8. Based on the values in this table, moments ($M'$) given by (6·4)–(6·7) of $a_1(c)$ and semi-invariants ($L$) derived therefrom are as follows. The normal values are, of course, known exactly but were computed for the purpose of checking the formulae:

$c = 4$; *normal universe*

$$\frac{L_1}{3} = \frac{M_1'}{3} \simeq 1 - \frac{2}{n} + \frac{4}{n^2} - \frac{8}{n^3},$$

$$\frac{M_2'}{9} \simeq 1 - \frac{4}{3n} - \frac{28}{n^2} + \frac{1040}{3n^3}, \quad \frac{L_2}{9} \simeq \frac{8}{3n} - \frac{40}{n^2} + \frac{1136}{3n^3},$$

$$\frac{M_3'}{27} \simeq 1 + \frac{2}{n} - \frac{48}{n^2} - \frac{1040}{n^3}, \quad \frac{L_3}{27} \simeq \frac{64}{n^2} - \frac{2368}{n^3},$$

$$\frac{M_4'}{81} \simeq 1 + \frac{8}{n} + \frac{40}{3n^2} - \frac{3520}{n^3}, \quad \frac{L_4}{81} \simeq \frac{3840}{n^3}.$$

$c = 4$; *universal* $\lambda_4 = \frac{1}{2}$

$$\frac{L_1}{3{\cdot}5} = \frac{M_1'}{3{\cdot}5} \simeq 1 - \frac{3{\cdot}357}{n} + \frac{11{\cdot}822}{n^2} + \frac{12{\cdot}1}{n^3},$$

$$\frac{M_2'}{(3{\cdot}5)^2} \simeq 1 - \frac{2{\cdot}286}{n} - \frac{57{\cdot}34}{n^2} + \frac{776{\cdot}03}{n^3}, \quad \frac{L_2}{(3{\cdot}5)^2} \simeq \frac{4{\cdot}4286}{n} - \frac{92{\cdot}25}{n^2} + \frac{831{\cdot}2}{n^3},$$

$$\frac{M_3'}{(3{\cdot}5)^3} \simeq 1 + \frac{3{\cdot}215}{n} - \frac{107{\cdot}47}{n^2} - \frac{2853{\cdot}89}{n^3}, \quad \frac{L_3}{(3{\cdot}5)^3} \simeq \frac{144{\cdot}61}{n^2} - \frac{6193{\cdot}95}{n^3},$$

$$\frac{M_4'}{(3{\cdot}5)^4} \simeq 1 + \frac{13{\cdot}143}{n} + \frac{20{\cdot}49}{n^2} - \frac{9529}{n^3}, \quad \frac{L_4}{(3{\cdot}5)^4} \simeq \frac{10{,}587}{n^3}.$$

$c = 1$; *normal universe*

$$L_1 = M_1' \simeq 0{\cdot}7978845608 + \frac{0{\cdot}19947114}{n} + \frac{0{\cdot}02493389}{n^2} - \frac{0{\cdot}03116737}{n^3},$$

$$L_2 \simeq \frac{0{\cdot}04507034}{n} - \frac{0{\cdot}07957747}{n^2} + \frac{0{\cdot}03978874}{n^3},$$

$$L_3 \simeq -\frac{0{\cdot}01685645}{n^2} + \frac{0{\cdot}07613597}{n^3}.$$

$c = 1$; *universal* $\lambda_4 = \frac{1}{2}$

$$\frac{L_1}{\mu_{|1|}} = \frac{M_1'}{\mu_{|1|}} \simeq 1 + \frac{0{\cdot}35239362}{n} - \frac{0{\cdot}159616}{n^2} - \frac{0{\cdot}745838}{n^3},$$

$$\frac{M_2'}{\mu_{|1|}^2} \simeq 1 + \frac{0{\cdot}79792429}{n} - \frac{0{\cdot}458012}{n^2} - \frac{1{\cdot}800648}{n^3}, \quad \frac{L_2}{\mu_{|1|}^2} \simeq \frac{0{\cdot}09313705}{n} - \frac{0{\cdot}262961}{n^2} - \frac{0{\cdot}196477}{n^3},$$

$$\frac{M_3'}{\mu_{|1|}^3} \simeq 1 + \frac{1{\cdot}336592}{n} - \frac{0{\cdot}850081}{n^2} - \frac{3{\cdot}239101}{n^3}, \quad \frac{L_3}{\mu_{|1|}^3} \simeq \frac{0{\cdot}053356}{n^2} + \frac{0{\cdot}204164}{n^3},$$

$$\mu_{|1|} = 0{\cdot}78126197.$$

Two sample sizes were considered: $n = 100$ and $n = 500$. For $n = 100$ and $c = 4$, the following are the Pearson Type IV frequencies of $a_1(4)$ when the parent universes are normal and have $\lambda_4 = \beta_2 - 3 = \frac{1}{2}$ respectively:

$$\text{Normal: } \lambda_4 = 0. \quad \left.\begin{aligned} &\kappa \cos^{11\cdot3350}\theta\, e^{13\cdot01543\theta}\, dx, \\ &\tan\theta = (x - 1\cdot873387)/0\cdot765849, \\ &\log_{10}\kappa = \bar{3}\cdot2644596. \end{aligned}\right\} \tag{6·10}$$

$$\lambda_4 = \tfrac{1}{2}: \quad \left.\begin{aligned} &\kappa \cos^{6\cdot0096}\theta\, e^{2\cdot31288\theta}\, dx, \\ &\tan\theta = (x - 2\cdot8522)/0\cdot9062, \\ &\log_{10}\kappa = \bar{1}\cdot7499974. \end{aligned}\right\} \tag{6·11}$$

The normal probability points shown in column (2) of Table 10 were derived from the foregoing normal frequency (6·10); the points in column (3) were derived from a Gram-Charlier formula (Geary, 1935). The 0·01 and 0·05 points given in column (2) are practically identical with those given by E. S. Pearson (1929) for $a(4)$, namely, 4·39 and 3·77. The powers given in column (4) are the aggregate frequencies lying beyond the values of the variate shown in column (2) on the assumption that the actual frequency was (6·11). The corresponding figures for $c = 1$ given in column (5) were based on a Gram-Charlier formula.

Table 9. *Power of* $a_1(c)$ *for* $c = 4$ *and* $c = 1$ *of discriminating* (6·9) *for* $\lambda_4 = \frac{1}{2}$ *from the normal* ($\lambda_4 = 0$) *at four normal theory probability levels. Samples of* 100

| Normal theory probability | Normal theory probability points | | Power for frequency (6·9) with $\lambda_4 = \frac{1}{2}$ | |
|---|---|---|---|---|
| (1) | $c=4$ (upper) (2) | $c=1$ (lower) (3) | $c=4$ (4) | $c=1$ (5) |
| 0·01 | 4·3836 | 0·7482 | 0·0648 | 0·0695 |
| 0·05 | 3·7744 | 0·7642 | 0·1995 | 0·1979 |
| 0·10 | 3·5195 | 0·7725 | 0·3163 | 0·3037 |
| 0·20 | 3·3110 | 0·7824 | 0·4525 | 0·4597 |

Before discussing the comparative powers in Table 9 it will be convenient to give a table, 11, on the same lines but for $n = 500$. On account of the larger sample size it has been necessary to change the reference-probabilities given in column (1). For the construction of this table Gram-Charlier formulae were used throughout—the probability points being determined from the E. A. Cornish & R. A. Fisher (1937) formulae—after verifying that for two of the probability levels, 0·01 and 0·05, the probability points for $c = 4$ (column (2) above) did not differ appreciably from those given by E. S. Pearson, namely, 3·60 and 3·37 (for $a(4)$), based on a Type IV curve.

The analysis in § 5 has enabled us to come fairly firmly to the conclusion that for indefinitely large samples $a(4)$ was to be preferred to $a(1)$ as a test of normality. We see from Tables 9 and 10 that this is subject to an important qualification. Table 9 shows that the discriminating power is definitely greater for samples of 500 for $a(4)$ than for $a(1)$, but the superiority is less emphatic than might have been anticipated from § 5. For medium-sized samples (Table 9) $a(4)$ exhibits no superiority. Of course, these conclusions are very tentative, as being based upon a single alternative and on particular sample sizes. The writer had proposed, in addition, to examine the universes (i) $\lambda_3 = 0$, $\lambda_4 = 1$ and (ii) $\lambda_3^2 = \lambda_4 = \frac{1}{2}$ as alternatives to the normal but time did not permit; he ventures to repeat the hope that other students will take the matter up.

Table 10. *Power of $a_1(c)$ for $c = 4$ and $c = 1$ of discriminating (6·9) for $\lambda_4 = \frac{1}{2}$ from the normal ($\lambda_4 = 0$) at four probability levels. Samples of 500*

| Normal probability | Normal probability points | | Power for frequency (6·9) with $\lambda_4 = \frac{1}{2}$ | |
|---|---|---|---|---|
| | $c=4$ (upper) | $c=1$ (lower) | $c=4$ | $c=1$ |
| (1) | (2) | (3) | (4) | (5) |
| 0·005 | 3·7062 | 0·773167 | 0·1934 | 0·2067 |
| 0·01 | 3·6094 | 0·775684 | 0·2920 | 0·2790 |
| 0·05 | 3·3766 | 0·782482 | 0·5955 | 0·5196 |
| 0·10 | 3·2695 | 0·786058 | 0·7392 | 0·6509 |

## 7. Conclusion and summary

In § 2 of the present paper it is shown that the actual probability of differences between means and variances derived from random samples on the nul-hypothesis may differ considerably from the probability derived from the standard tables (compiled on the assumption that the universal distribution is normal), when, in fact, the universal distribution is *not* normal. Accordingly, the standard tables cannot validly be used unless tests, based on the sample from which the inferences are to be drawn, or on a series of samples produced under similar conditions, have established the likelihood that the universal distribution is approximately normal. In certain cases—but these must be few—the nature of the material may, of itself, suffice to justify the assumption of universal normality. When universal normality cannot be assumed, the best course will be to correct the standard tables using, for this purpose, the moments (up to, say, the fourth) derived from the sample, in conjunction with the formulae given in § 2. This procedure is, of course, open to the objection that the moments derived from the sample may, in fact, differ substantially from the (in general unknown) universal moments, so that any probabilistic inference derived using sample moments must be accepted with reserve. If $b_2 = 3{\cdot}5$, say, it would be safer to assume that the universal value $\beta_2$ is 3·5, than to hope (without other evidence) that it is 3, the normal value; it might be 3·75 or even 4, when, usually, the standard table probabilities will be still further astray. It should not be difficult to construct supplementary tables giving very approximate corrections of the standard tables, using the moment expansions given in § 2, for different values of $\sqrt{\beta_1}$ and $\beta_2$. To compute unbiassed estimates of the latter, R. A. Fisher's $k$ statistics (1929) should, of course, be used.

It may be asked if testing for normality and, when necessary, correction for universal non-normality is worth the trouble. To answer this question it is desirable to have regard to the logical position of the statistician, concerned with drawing inferences from samples, whose characteristic approach may be defined as *reductio ad paene absurdum*: if an event is highly improbable it must be regarded for practical purposes as impossible. St Thomas Aquinas's* famous 'certitude of probability' is peculiarly apt as applied to the mental attitude of the statistician, from two quite different viewpoints. The first is that decision, and action based on that decision, for which there is not certainty, but merely probabilistic preference, is absolute. One does not say that one has a preference of 20 to 1 for Fertilizer A

* 'According to the Philosopher, certitude is not to be sought equally in every matter....Hence the certitude of probability suffices, such as may reach the truth in the greater number of cases, although it fails in the minority' (*Summa* IIa–IIae q. lxx, *a*. 2).

over Fertilizer B because the differences between the yields is at or near the 5 % probability point of some test functions: one necessarily decides without qualification that A is better than B.

The second aspect, which has the greater relevance in the present case, is that the statistician regards himself as endowed with 'certitude' when he knows that if he repeated an experiment, as to, say, significant differences in averages, a great number of times, he would be in error in attributing significant difference when, in fact, there was none, in a predetermined proportion of cases. He has certitude as to the probability though his decision in the individual case may be wrong. What is curious is that decisions (which, in effect, are absolute) can be based on probability levels which vary with the temperament of the statistician from perhaps a conservative 0·001 to a daring 0·1. For the particular statistician the probability level will vary with the case: for instance, the present writer would be inclined to suspect non-normality near the 10 % probability level of the $a(1)$ table, whereas he would not be disposed to attach significance in, say, analysis of variance, until about the $2\frac{1}{2}$ % level. Naturally the level will depend on the importance attaching to the decision.

Since all the statistician usually requires from the table of probability for a given measure of significance is whether, on the nul-hypothesis, the probability is 'small', absolute precision is not necessary in the probability. If the probability is thought to be minute, say 0·001, it does not matter if in actual fact it is 0·002 or 0·0005. If, on the contrary, the standard table value is approaching the statistician's level of decision it surely matters a great deal: if he thinks his judgment is likely to be erroneous in 1 out of 20 experiments it must be of importance if, in fact, the true probability is something like 1 in 10 or 1 in 5. These are the kinds of contrasts that appear from § 2, from comparison of standard table probabilities with 'actual' probabilities found when the samples were assumed to be randomly drawn from certain arbitrarily selected types of non-normal universes. The computed probabilities in § 2 admittedly make no claim to exactitude in most of the cases, since the formulae were strained by their application to small sample theory. The point is, however, that the estimates of the actual probabilities are unbiassed in regard to the 'normal theory' probabilities: if the former could be closer to the latter, they might also be further away.

There is one case which is in a quite exceptional category, namely that considered at the beginning of § 2. As far as the writer is aware, this case has never been examined *theoretically* before, despite the extreme simplicity of the algebra. It is shown that in the simplest case of analysis of variance, when the two sample numbers are of the same order of magnitude, the variance is proportional, approximately, to $(\beta_2 - 1)$, so that quite a small measure of universal kurtosis materially changes the probability. Statisticians must have been affected by a kind of hypnosis in favour of normal theory to have overlooked so trivial a point, a stricture from which the writer is not particularly concerned to exclude himself! An exception was E. S. Pearson (1931) who, on the basis of his results cited in § 2 (*a*), sounded a warning: 'The illustration should serve to emphasize the fact that certain of the "normal theory" tests can be used with greater confidence than others when dealing with samples from populations whose distribution laws are not known.'

An interesting chapter could be written on the fluctuations in the attitude of statisticians during the past century on the question of the occurrence of the normal frequency distribution in nature, a chapter, perhaps, in a large work on Fashions in the Sciences down the Ages. Amongst the following the historian may find the reasons for the prejudice in favour of the hypothesis of universal normality up to, say, the end of the last century:

(1) The fact that, to a close approximation, it applies in a wide range of *mathematical* conditions.

(2) The fact that the theory found practical applications predominantly in assessing the probability of errors in astronomical measurements and in games of chance where the mathematical model could reasonably be assumed to apply.

(3) The beauty of the mathematical theory and the facility of algebraic manipulation in the function involved.

(4) The general shape to the visual sense of such frequency distributions as were known, before $\chi^2$ imposed its discipline.

With the development, about the beginning of the century, of the theory of moments, statisticians became almost over-conscious of universal non-normality. The concomitant semi-invariant approach had quite a different background. The difference between the moment and Karl Pearson curve system on the one hand and semi-invariants and the Gram-Charlier system on the other is fundamentally that for the former normality is a particular case like any other, whereas for the latter normality is basic and generative. Each system has its advantages and disadvantages as applied to the determination of frequency distributions of which the lower moments are known. In fanciful terms one might say that in the ship Gram-Charlier one might sail in perfect safety but only within limited, and more or less ascertainable, range of Port Normality, whereas in the good craft Pearson one can sail the seven seas—at one's own risk.*

Our historian will find a significant change of attitude about a quarter-century ago following on the brilliant work of R. A. Fisher who showed that, when universal normality could be assumed, inferences of the widest practical usefulness could be drawn from samples of any size. Prejudice in favour of normality returned in full force and interest in non-normality receded to the background (though one of the finest contributions to non-normal theory was made during the period by R. A. Fisher himself), and the importance of the underlying assumptions was almost forgotten. Even the few workers in the field (amongst them the present writer) seemed concerned to show that 'universal non-normality doesn't matter': we so wanted to find the theory as good as it was beautiful. References (when there were any at all) in the text-books to the basic assumptions were perfunctory in the extreme. Amends might be made in the interest of the new generation of students by printing in leaded type in future editions of existing text-books and in all new text-books:

*Normality is a myth; there never was, and never will be, a normal distribution.*

This is an over-statement from the practical point of view, but it represents a safer initial mental attitude than any in fashion during the past two decades.

As already indicated, the present work is incomplete, especially on the experimental side. The writer hopes that he has created a *prima facie* case for the importance of testing for normality.

SUMMARY

(i) Inferences drawn from the standard (normal) tables of $z$ and $t$ may be seriously in error if the conditions in which the standard tables apply (the principal of which is that the universes from which the samples are drawn are normal) are ignored.

* This comment must not be taken as applying to the problem of curve-fitting, i.e. to fitting a smooth curve to given frequencies, but to the problem of estimating the frequency function given the first few semi-invariants.

(ii) Sufficient conditions are given for the approach to normality, with increasing sample size, of the field of tests of normality $a(c)$ (given by (3·1)) for $c > 0$.

(iii) Many term expansions of the first four moments of $a(c)$ for normal samples are given with practical applications designed to find the values of $c$ for which the moments could be used with confidence to find the frequency distributions for medium-size samples; semi-invariants of $a_1(2\cdot4)$ and $a_1(4)$ ($a_1(c)$ is given by (3·2)) are compared; correlations between $a_1(c)$ and $a_1(c')$ are examined.

(iv) For indefinitely large samples and a wide field of alternative universes $a(4)$ is found to be the most sensitive test of kurtosis and an analogous test of asymmetry $g(c)$ is found to be most sensitive for $c = 3$, $g(3)$ being the familiar $\sqrt{b_1}$.

(v) An examination of the relative efficiency of $a(1)$ and $a(4)$ from the Power Function point of view suggests that $a(4)$ is increasingly to be preferred as the sample size increases; for samples of moderate size $a(1)$ is probably as efficient as $a(4)$.

(vi) Throughout the paper a considerable range of formulae is given in case students may feel interested to carry the writer's researches a stage further so as to give a firmer basis to his conclusions or to modify them. It is suggested (§ 4) that the preparation of a table of probability points of $a(2\cdot2)$ for normal samples of different sizes be taken in hand.

REFERENCES


Baker, G. A. (1932). *Ann. Math. Statist.* **3**, 1.
Bartlett, M. S. (1935). *Proc. Camb. Phil. Soc.* **31**, 226.
Cornish, E. A. & Fisher, R. A. (1937). *Rev. Inst. Int. Statist.* **5**, 307.
Craig, C. C. (1928). *Metron*, **7**, 3.
Eden T. & Yates, F. (1933). *J. Agric. Sci.* **23**, 6.
Fisher, R. A. (1925). *Metron*, **5**, 90.
Fisher, R. A. (1929). *Proc. Lond. Math. Soc.* (2), **30**, 199.
Fisher, R. A. (1930). *Proc. Roy. Soc.* A, **130**, 16.
Fisher, R. A. & Wishart, J. (1931). *Proc. Lond. Math. Soc.* (2), **33**, 195.
Fréchet, M. (1937). *Géneralités sur les Probabilités. Variables aléatoires.*
Geary, R. C. (1933). *Biometrika*, **25**, 184.
Geary, R. C. (1935). *Biometrika*, **27**, 310, 353.
Geary, R. C. (1936). *J. Roy. Statist. Soc.* (*Supplement*), **3**, 178.
Geary, R. C. (1936). *Biometrika*, **28**, 295.
Geary, R. C. (1947). *Biometrika*, **34**, 68.
Geary, R. C. & Pearson, E. S. (1938). *Tests of Normality.*
Geary, R. C. & Worlledge, J. P. G. (1946). *Biometrika*, **34**, 98.
Gosset, W. S. (1908). *Biometrika*, **6**, 1.
Hsu, C. T. & Lawley, D. N. (1940). *Biometrika*, **31**, 238.
Kendall, M. G. (1941). *Biometrika*, **32**, 81.
Nair, A. N. K. (1942). *Sankhyā*, **5**, 393.
Neyman, J. & Pearson, E. S. (1933). *Philos. Trans.* A, **231**, 289.
Neyman, J. & Pearson, E. S. (1936). *Statist. Res. Mem.* **1**, 1.
Pearson, E. S. (1929). *Biometrika*, **21**, 337.
Pearson, E. S. (1930). *Biometrika*, **22**, 239.
Pearson, E. S. (1931*a*). *Biometrika*, **22**, 423.
Pearson, E. S. (1931*b*). *Biometrika*, **23**, 114.
Pearson, E. S. (1935). *Biometrika*, **27**, 333.
Pearson, E. S. & Adyanthaya, N. K. (1929). *Biometrika*, **21**, 259.
Pearson, Karl (1895). *Philos. Trans.* A, **186**, 343.
Pepper, Joseph (1932). *Biometrika*, **24**, 55.
Rider, P. R. (1931). *Ann. Math. Statist.* **2**, 48.
Rietz, H. L. (1939). *Ann. Math. Statist.* **10**, 265.
Shewhart, W. A. & Winters, F. W. (1928). *J. Amer. Statist. Ass.* **23**, 144.
Wishart, J. (1930). *Biometrika*, **22**, 224.
Yasukawa, K. (1934). *Tokohu Math. J.* **38**, 465.