PEARSON CURVES REVISITED

BY

H. SOLOMON and M. A. STEPHENS

TECHNICAL REPORT NO. 226

NOVEMBER 5, 1975

THIS RESEARCH WAS SPONSORED BY THE ARMY RESEARCH OFFICE
OFFICE OF NAVAL RESEARCH, AND AIR FORCE OFFICE OF
SCIENTIFIC RESEARCH BY CONTRACT NO.
ND0014-67-A-0112-0085 (NR-042-267)

Reproduction in Whole or in Part is Permitted
for any purpose of the United States Government
Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
PEARSON CURVES REVISITED

by

H. Solomon and M. A. Stephens

Technical Report No. 226
November 5, 1975

Prepared under Contract N00014-67-A-0112-0085
(NR-042-267)
Office of Naval Research
Herbert Solomon, Project Director

Reproduction in Whole or in Part is Permitted for
any Purpose of the United States Government

Approved for public release; distribution unlimited

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
1. INTRODUCTION

1.1 In this article we show, by means of a variety of examples, how useful Pearson curves can be in providing approximations to percentage points of theoretical distributions whose moments can be found, although they are otherwise analytically intractable; such a situation can arise in many different circumstances.

It seems to us timely to draw attention to the use of Pearson curves for two reasons. Firstly, and on the practical side, in recent years very extensive tables of percentage points of these curves have been produced which make their potential far greater than before. Secondly, and in a much more festive spirit, Professor Egon Pearson will celebrate his eightieth birthday on August 5, 1975, and we wish to dedicate this work to his honor. It was, of course, his father, Karl Pearson, who first introduced the system of Pearson curves, and Egon Pearson has justly earned his own renown for his contributions to both mathematical and applied statistics; but, amongst the latter, he has always taken an interest in the use of Pearson curves, and in fact his support and inspiration are largely behind the production of the newer tables referred to above.

If this report serves to interest a wider group of statisticians to the possibility of using Pearson curves to provide useful significance points, we know he would find this a fitting birthday tribute.
1.2 Pearson curves (P.C.) are probability densities \( y = f(x) \), which are solutions of

\[
\frac{1}{y} \frac{dy}{dx} = \frac{-(C_1 + x)}{C_0 + C_1 x + C_2 x^2}.
\]

As the constants are varied, curves of many different shapes can occur, and the range \( a < x < b \) of the variable \( x \) may be infinite or finite. Pearson distinguished several types of curve (Types 1 - 11), and as the years have passed many of the most important densities of statistics have been found to be members of the family. The classification and properties of the curves are described in Elderton (1953) and in Volumes 1 and 2 of Biometrika Tables for Statisticians (Pearson and Hartley, 1962, Vol. 1; 1972, Vol. 2).

The moments of a P.C. are of course determined by the values of the constants in (1), plus the constant of integration; conversely, if one were given four moments of a distribution \( D \), one could in principle solve for the constants and hence determine the P.C. with the same first four moments as \( D \), provided a P.C. exists with these moments; certain combinations do not lead to an acceptable solution of (1). The procedure is presented in detail in the references cited above.

In many cases the P.C. will then very closely approximate \( D \), and percentage points of the P.C., or tail probabilities, could be used to approximate those of \( D \) when these cannot easily be found analytically.

Alternatively, if density \( D \) has an end-point at, say, \( a \), the Pearson curve may be fitted by making its density also have an endpoint
at a, and thereafter matching the first three moments of the P.C. and D. If finite endpoints exist at both ends of the range for x, then only two moments will be needed and the two endpoints can be matched.

1.3 Percentage points of Pearson curves. Fortunately, there is a direct passage from the moments to the percentage points, and the tedious step of solving for \( f(x) \) itself can be eliminated as follows. Suppose a random variable \( x \) has a Pearson curve density, with mean \( \mu \) and central moments \( \mu_2, \mu_3, \mu_4 \). The percentage points of \( X \), the standardized \( x \) given by \( X = (x-\mu)/\sigma \), where \( \sigma = \sqrt{\mu_2} \), are obtained as functions of the shape parameters \( \beta_1 = \mu_3/\mu_2^2 \), and \( \beta_2 = \mu_4/\mu_2^2 \), and can be tabulated in double-entry tables against \( \sqrt{\beta_1} \) (or \( \beta_1 \)) and \( \beta_2 \). The percentage points of \( x \) are then easily obtained from those of \( X \).

Tables of standardized percentage points of Pearson curves were included in Vol. 1 of Biometrika Tables for Statisticians (B.T.S.); these were extended in a paper by Johnson, Nixon, Amos and Pearson (1963) and the new Tables reproduced in Vol. 2 of B.T.S. in 1972. The introductions in both volumes include summaries of the Pearson curve system, and, in Vol. 2, instructions and illustrations of how to use the Tables. Further extensions have recently been provided by Amos and Daniel (1971) and Bouwer and Bargmann (1974); they include FORTRAN programs from which the tables were prepared. Bouwer and Bargmann also give a program which may be used to calculate tail probabilities for a given value of \( X \).
Over many years Pearson curves have been used, in one way or another, to approximate continuous density functions. At one time they were quite often employed to fit a density to a sample of data, so that the sample moments were used to fit the curves; frequently \( f(x) \) was then explicitly calculated. This use has gradually fallen out of favor, partly, no doubt, because of the extensive algebra involved in solving for \( f(x) \) itself; also because statisticians were not attracted to use of a specific functional form which, it was known, was only an approximation of one curve to another, and apart from this did not represent in a meaningful way the true density (for example, the constants in (1) had no interpretation). Finally, and perhaps this is the severest criticism, the sampling variability of higher moments, derived from a sample, is so great that one could not be very confident of the accuracy of the approximation; it would be at least as good, probably, with a large sample, to use the histogram directly to obtain percentage points, provided, of course, this were drawn with small enough intervals.

The other major application, and the subject of this article, is to use the curves as approximations to densities with known moments. This technique has often been employed, but not often in the direct manner which we consider below; often only one of the special types (say Type III, the Gamma distribution, or Type V) has been used for approximating purposes, and this may involve using less than four moments to estimate the Pearson curve. There was often good reason for this; sometimes the algebra involved in calculating theoretical moments is difficult even to the third, and sometimes the user wanted, if the approximation proved
successful, to exploit the relative accessibility of, say, tables of the $\chi^2$ or Gamma distributions.

The accuracy of an approximation by a Pearson curve is difficult to assess; essentially, one curve is placed on top of another, and we know only that four moments (or perhaps an endpoint and three moments, or two end points and two moments) agree; the percentage points will match well for some shapes of curve, and not for others. This total absence of mathematical analysis of error has probably put off many potential users, though we might comment parenthetically that one of us (Herbert Solomon) attended a course many years ago, given by the eminent mathematical statistician Abraham Wald, devoted entirely to Pearson curves! (The other author (M. A. Stephens) gained his introduction from Egon Pearson.) It is worth remarking also that even when alternative systems of approximation are used (Gram-Charlier or similar series, for instance) such error analysis as can be made sometimes gives bounds which are not of much practical use; or alternatively, they require more than four moments for good results, or considerable expertise in deciding, say, which system of orthogonal functions to use.

We give below some examples where we have fitted Pearson curves by the different methods referred to above. The examples have been chosen because they cover a broad range of applications, and because the accuracy can be estimated. This can be done because exact points are available for certain parameter values, or because other excellent approximations are known to exist; nowadays it is also possible, of course, to compare with Monte Carlo studies in certain situations. In some examples the
approximations are not successful in one tail of the distribution. Our experience shows that this can occur when a very steep tail is to be approximated (of course, this is difficult to approximate by any curve-matching technique), but in general Pearson curves will be excellent in giving points in the long tail(s) of a distribution; happily these are often what are wanted in practical situations.
2. FITTING PEARSON CURVES

2.1 We start by giving the steps in fitting a Pearson curve to a theoretical distribution. Suppose the variable \( x \) to be approximated, has mean \( \mu \) and central moments \( \mu_2 = \sigma^2 \), \( \mu_3 \) and \( \mu_4 \).

(a) Calculate \( \beta_1 = \mu_3 / \mu_2^3 \), and \( \beta_2 = \mu_4 / \mu_2^2 \).

(b) Let \( X \) be the percentage point at significance level \( \alpha \), of a standardized P.C. variable, with these values of \( \beta_1 \) and \( \beta_2 \), found from the tables or programs referred to above. (Most tables use \( \sqrt{\beta_1} \), taken positively; Bouwer and Bargmann use \( \beta_1 \).)

(c) Calculate \( x'_{\alpha} = \sigma X_{\alpha} + \mu \) if \( \mu_3 \) is positive, or \( x'_{1-\alpha} = \mu - \sigma X_{\alpha} \) if \( \mu_3 \) is negative.

(d) Then \( x'_{\alpha} \) is the approximation for \( x_{\alpha} \), the \( \alpha \)-level percentage point of \( x \).

The published tables have been produced for standardized variables \( X \) with \( \mu_3 \) positive; for \( \mu_3 \) negative, consider the random variable \( Y = -X \). Variable \( Y \) will have the same moments as \( X \) except for a change of sign in \( \mu_3 \); clearly the upper \( \alpha \)-level point of \( Y \) is \((-1)\) times the lower \( \alpha \)-level point of \( X \). This leads to the formulas in (c) above.

(e) Use of three moments and an endpoint. Suppose only the first three moments are known, and the knowledge that the distribution of \( x \) begins at \( x = a \). Then \( X_0 = (x-a)/\sigma \) is the standardized lower \( 0\% \) point of \( X \). In the tables, for a given \( \beta_1 \) or \( \sqrt{\beta_1} \), find \( \beta_2 \) which gives the end point equal to \( X_0 \). Then use this \( \beta_2 \), with \( \beta_1 \) in
steps (b), (c), (d) above. This technique can be adapted in an obvious way when the upper endpoint is known.

If both endpoints are known, one must find the combination of $\beta_1$ and $\beta_2$ to give the standardized endpoints; this is a more tedious procedure and we have no examples of its use, so will not discuss it in detail.

In the next section we give examples of these methods in practice.
3. PEARSON CURVES FITTED TO QUADRATIC FORMS

3.1 Let \( x_i \), \( i = 1, 2, \ldots, k \), be independent standard normal variables, and let \( Q_k(c, a) \) be the quadratic form

\[
Q_k(c, a) = \sum_{i=1}^{k} c_i (x_i + a_i)^2
\]

where \( c, a \) are vectors \((c = c_1, c_2, \ldots, c_k)\) and \((a = a_1, a_2, \ldots, a_k)\) whose components are constants. When the \( a_i \) are all zero, the \( Q_k(c, a) \) will be shortened to \( Q_k(c) \); if possible, both will be shortened to \( Q_k \). \( Q_k(c) \) is a simple weighted sum of \( \chi^2_1 \) variables. The cumulants of \( Q_k(c, a) \) are very easily computed, from the formula for the r-th cumulant \( K_r \):

\[
K_r = 2^{r-1}(r-1)! \sum_{j=1}^{k} c_j^r (1 + r a_j^2).
\]

Applications of \( Q_k \), for finite \( k \), are very many; see, for example, references in Solomon (1960) or Jensen and Solomon (1974).

The quadratic form above, with \( k \) infinite, and \( a \) zero, arises in the asymptotic distribution theory of goodness-of-fit statistics of the Cramer-von Mises type. (cf, for example, Anderson and Darling (1952), Sukhatme (1972), Stephens (1970, 1971), Durbin and Knott (1972), De Wet and Venter (1973). The non-central form \( Q_\omega(c, a) \) arises in the distribution theory of these statistics when the null hypothesis is not true, i.e., in asymptotic power studies (Durbin and Knott (1972),
Many methods have been used to give probability or percentage points of $Q_k$, for various $k$ and $z$, combinations. Early tables, obtained by exact interpolation for $k = 2$ and $3$ were produced by Grad and Solomon (1955) and are reproduced in Owen (1962); note that $z$, as used here, is $a$ in Solomon (1960). More tables, for $k = 4$ and $5$, were given later by Kotz, Johnson and Boyd (1967a, 1967b). Quite a bit of analysis was used to obtain the exact points, and in the end extensive numerical computation is required. However, this can be and was made so accurate that all the points and probabilities in the tables referred to above will be regarded as exact.

An approximation which again can be made very accurate was given by Imhof (1961) for finite $k$; this also involves numerical integration (inversion of the characteristic function) and certain parameters can be varied to improve the accuracy. Imhof's method has been adapted for the case of infinite $k$ (with $z = 0$) by Durbin and Knott (1972); they curtail $Q_\infty(z)$ after $k$ terms and add one more term of the form $qX^2p$, the coefficient $q$ and degrees of freedom $p$ being decided by equating the first two moments of $Q_\infty(z)$ to those of the approximation. Durbin and Knott used this method to obtain percentage points of goodness-of-fit statistics.

Other approximations which can be used when moments to high order are available include expansions of the Gram-Charlier or Edgeworth type. Finally various authors, anxious to cut down the work involved in exact calculation of percentage points or probabilities, have suggested easier approximations, either using Pearson Type III curves, or transformations
to new variables which are approximated by normal variables. For a

general summary see Kotz, et al. (1967a, 1967b) and Jensen and Solomon
(1972). There is no doubt then that distributions of quadratic forms

have been of considerable interest for many years, and we shall use them

as our first illustration of using Pearson curves. The moments for any

$\zeta$ and $\xi$ are very easily obtained (to any order, incidentally) from (3);

four moments have been used to fit P.C. curves to the quadratic forms in

Tables la.1 and la.2.

3.2 Quadratic forms: $k$ finite. In Table la.1 percentage points are
given for central forms (i.e., $\zeta = \varrho$) and $k$ finite. The exact values
are taken from Solomon (1960) and Kotz, et al. (1967a, b); they are
compared with points given by the Pearson curve approximation and also
with values found by Imhof's method.

Table la.2 gives probabilities of the form $P(Q < t)$ for various $t$;
the results are presented in this way for comparison with tables in Jensen
and Solomon (1971). These authors examined a number of approximations,
including a transformation to a normal variable of their own, which
generally performed better than its competitors. We include therefore
only the exact, Pearson curve and Jensen-Solomon values. Among the
distributions considered is the non-central $\chi^2$ distribution, called $Q_1$,
with d.f. $k$ and non-centrality parameter $\lambda$; this can, of course, be
obtained from the general form $Q_k(\zeta, \varrho)$ by making all $c_i = 1$, and by
arranging the $a_i$ so that $\lambda = \sum a_i^2$. Also included are $Q$ distribu-
tions expressed as a sum of weighted $\chi^2_p$ distributions with $p$ not equal
to 1; these are obtained from the general form by repeating some of
the coefficients  \( c_1 \); in those illustrations it happens that \( a = \varphi \).

3.3 Quadratic forms: \( k \) infinite. In Table 1b results are given for
distributions of the form \( Q_\infty(x) \) which arise in connection with asymp-
totic distributions of goodness-of-fit statistics, specifically those
often called \( W^2 \), \( U^2 \), and \( D \) (see e.g. Stephens (1974)). The
illustrations given are for the situation where the distribution tested
is completely specified (called Case 0 in Stephens (1974)). Exact asymp-
totic distributions have long been known for this case, so that exact
values are available for comparison.

3.4 Comments on Table 1. In all comparisons of approximations to
percentage points, the meaningful comparison is between the \( \alpha \) for which
the point has been calculated, and the actual \( \alpha' \) achieved by the approxi-
mation, and not merely a comparison of the points themselves; with this
in mind, we examine the 3 parts of Table 1. For the upper tail, which
would generally be needed in practical uses of all these distributions,
the Pearson curves give excellent results, with negligible errors in \( \alpha \).
In Tables 1.a.1 and 1.a.2 they
compare well with Imhof's technique and also with the Jensen-Solomon
approximation; for other approximations, which will on the whole be worse,
see Jensen and Solomon (1972). For the steeper lower tail, the Pearson
curve approximation is not quite so effective, and becomes less so as the
coefficients \( c_1 \) become more unequal. Consider, for example, the Case
\( k = 3 \), \( c_1 = .2, .3, .5 \); for the lower tail 5% point, the P.C. is
in error by 9 units in the third decimal place; the difference between
the 5% and 10% points is 73 units, so that roughly the a-error could be estimated as about .6 of a percent. For \( k = 4 \), \( c_1 = 1.8, 1.2, 0.5, 0.5 \), lower 1% point, the true \( a' \) is probably nearer .025.

Imhof's technique gives excellent results in both tails, as one would expect for such a tailor-made approximation. Nevertheless it takes more computer time than fitting Pearson curves.

Several of the approximations have required the use of the recent more extensive curves (we used those of Amos and Daniel) since the \( (\beta_1, \beta_2) \) values are outside the range of the Biometrika tables. This applies also to \( W^2 \) in Table 1.3.

A considerable side benefit emerges from the accuracy of the Pearson curves in the upper tail of the quadratic form distribution, particularly for finite \( k \). Various authors have commented on the potential value of tables of percentage points of quadratic forms: Solomon (1960), for example, states that they would be useful for \( k \) up to 10 at least. Such sets of tables would need to be very lengthy to cover all the \( c_1 \) combinations of interest, and it would be difficult to see how one would interpolate for a set of \( c_1 \) for which the points were not tabulated. Now the problem has been in effect reduced, at least for the important upper tail, to calculating two principal parameters from the \( c_1 \), namely the \( \beta_1 \) and \( \beta_2 \) values; excellent approximations to the standardized percentage points depend only on these.
3.5 Goodness-of-fit statistics—further uses of Pearson curves. Pearson curves have recently been used to give the points for goodness-of-fit statistics in cases where parameters in the distribution tested (specifically, the normal and exponential distributions) were not known and had to be estimated from the data (Cases 1 to 4 in Stephens (1974); see also Table 54 in B.T.S., Vol. 2). Pearson curves have also been used to obtain significance points for the goodness-of-fit statistic $U^2$ for finite samples, since the first 4 moments can again be found; the exact distribution reduces to a problem in geometric probability, of finding the common volume between a simplex and a cylinder of expanding radius, and even for $N = 4$ this has great complexity. This particular application is of special interest, since Pearson has provided a thorough study of the different Pearson curve types which are used in the various interpolations necessary in $\sqrt{\beta_1}$ and in $\beta_2$ (B.T.S., Vol. 2, Pearson and Hartley, 1972).
4. RESULTANTS OF UNIT VECTORS

4.1 We turn now to an example in which we fit four moments or three moments and an endpoint. Suppose \( N \) unit vectors are uniformly distributed in two or three dimensions; i.e., if \( \mathbf{OP}_i, i = 1, 2, \ldots, N \), is a typical vector, the origin is fixed at \( O \) and \( P_i \) moves uniformly on the circumference of the circle with center \( O \) and radius 1 or on the surface of the sphere with center \( O \) and radius 1. Let \( \mathbf{R} \) be the resultant (i.e., vector sum) of the \( N \) vectors, with length \( R \). The distribution of \( R \) has been of interest in a great variety of problems; it is interesting to remark that one of the early formulations of the problem, for two dimensions, was given by Karl Pearson (1905). A general solution, for finite \( N \) and any number \( p \) of dimensions, is given by Watson (1922). For \( p = 2 \), the density takes a very difficult form,

\[
f_2(R) = R \int_0^\infty (J_0(t))^N J_0(Rt) t \, dt, \quad 0 \leq R \leq N,
\]

and the distribution function is

\[
F_2(R) = R \int_0^\infty (J_0(t))^N J_1(Rt) \, dt, \quad 0 \leq R \leq N.
\]

For \( p = 3 \), although a similar integral form exists, the solution may be put in closed form. Let

\[
P_r(R) = \frac{1}{2^{N-1} r!} \sum_{s=0}^\infty (-1)^s \binom{N}{s} \left( N - R - 2s \right)^r, \quad 0 < R < N.
\]
Then the density is $f_3(R) = R P_{N-2}(R)$, and the distribution function $F_3(R)$ is given by $1 - R P_{N-1}(R) - P_N(R)$. Here the symbol $\langle z \rangle$ means $\langle z \rangle = z$ for $z > 0$; $z = 0$ for $z < 0$. Thus the series in (6) terminates; nevertheless, as $N$ increases, it is difficult to use it accurately. Greenwood and Durand (1955) found tail probabilities and percentage points for $R$ in two dimensions by numerical evaluation of $F_2(R)$; Watson (1956) and Stephens (1964) have given points for three dimensions using $F_3(R)$. We call these the exact points. Greenwood and Durand also examined approximations of the Gram-Charlier types.

Cumulants of $Z = R^2$ can be calculated; $E(Z) = N$ for $p = 2$ or $3$; and further cumulants are

\[
\begin{align*}
p = 2 & \\
K_2(Z) & = N^2 - 1 \\
p = 3 & \\
K_3(Z) & = 2N(N-1)(N-2) \\
K_4(Z) & = 6N^4 - 36N^3 + 63N^2 - 33N \\
& 16(5N^4 - 30N^3 + 52N^2 - 27N)/45
\end{align*}
\]

These have been used, for $p = 2$, to give Pearson curve approximations to the distribution of $Z$, and hence to calculate percentage points of $Z$ and therefore of $R$ (Stephens (1964)).

4.2 Use of three moments and one end point. Since it is clear that $0 \leq Z \leq N^2$, one of the end points may also be used in the fit, as follows. The value of $Z$ is zero for $\alpha$ equal to zero. For a given $N$, Python

16
the standardized value $X_0$ corresponding to $Z = 0$ is, for two dimensions,

$$X_0 = (0 - N)/(N^2 - N)^{1/2} = -N/(N-1)^{1/2}.$$  

This is then used to find $\beta_2$ as described in Section 2(e). Similarly the upper endpoint can be used; this gives standardized $X_{100} = (N^2 - N)^{1/2}$.

A comparison of the exact values and these various approximations is given in Table 2, for two dimensions. It can be seen that for the points quoted, the 4-moment fit gives very slightly better results than the 3-moment-plus-fixed-point fit, though the practical difference in significance levels is negligible. Table 3 gives results, for three dimensions.

Again, all the three P.C. approximations give excellent results in the long upper tail, and the 4-moment method is good even in the lower tail. The 3-moment approximations, using the lower point as fixed point, also give remarkable accuracy in the long upper tail, but when the upper point is fixed, the difficulties of accurate interpolation for $\beta_2$ reveal themselves, and the lower tail points are generally not good. However, we should emphasize that these examples involve extremely long drawn out upper tails (for example, for $N = 10$, in the extreme upper tail, 1% of probability occupies nearly 41% of the entire range for $R$) and it is remarkable that fixing the upper tail and using only three moments can produce even the accuracy that it does. We have deliberately taken these illustrations to show how even very skew curves can be well-approximated, especially if all four moments are used.
The technique of fitting 4-moment Pearson curves has been used also for the distribution of $R/N$ when the typical unit vector, with spherical polar coordinates $(\theta, \phi)$, comes from the density on the sphere

$$f(\theta, \phi) = \frac{k}{4\pi \sin \theta} \sin \theta \exp(k \cos \theta), \quad 0 \leq \theta \leq \pi, \quad 0 \leq \phi \leq 2\pi.$$  

The above is called the Fisher distribution, and the distribution of $R/N$, although able to be written in closed form, is again difficult to compute, especially for large $N$. The Pearson curves give excellent approximations to the exact percentage points (Stephens, 1967). Tables were also constructed for $R/N$ when the vectors were from the corresponding density (the von Mises density) on the circle (Stephens, 1969). Here, the Pearson curves were invaluable, as exact points could not be found except by extensive numerical integration.
5. LINEAR COMBINATIONS OF UNIFORM ORDER STATISTICS

Suppose \( y_1 < y_2 < \ldots < y_n \) are the order statistics of a random sample of size \( n \) from a uniform distribution with limits \( A \) and \( B \), and consider the sample range

\[
\omega = y_n - y_1.
\]

This statistic can, of course, be used as an estimator of the population range \( R = B - A \). The distribution of \( \omega \) is well-known, and percentage points were given by Harter (1961), for the case when \( R = 12 \), i.e., the standard deviation of the uniform distribution is 1. By an obvious modification of Harter's tables we can get confidence bounds for \( R \). Thus, for \( n = 10 \),

\[
\Pr \left( \frac{2.0987}{\sqrt{12}} R < \omega < \frac{3.3367}{\sqrt{12}} R \right) = .90,
\]

giving a 90\% confidence interval for \( R \):

\[
1.038\omega < R < 1.651\omega.
\]

For \( n = 20 \) this becomes

\[
1.018\omega < R < 1.276\omega.
\]

The statistic \( \omega \) is a special case of a more general linear combination of the \( y_i \), \( V = \sum b_i y_i \), where \( b_i \) are constants. In discussing the distribution of \( V \) it is convenient to express it in terms of the
**spacings** \( u_i \) defined by

\[
u_i = y_i - y_{i-1}, \quad i = 2, 3, \ldots, n; \quad u_i = y_i; \quad u_{n+1} = 1 - y_n\]

and to write \( V = \sum_{i=1}^{n+1} a_i u_i \) where \( a_i \) are constants. For the range \( \omega \), \( a_1 = a_{n+1} = 0 \); all other \( a_i \) are \( 1 \). A statistic \( W \) of this type, sometimes called the thickened range (David, 1970), is

\[
W = (y_n - y_1) + (y_{n-1} - y_2) + \cdots + (y_{n-p+1} - y_p),
\]

where \( p \) is the greatest integer in \( n/2 \). In terms of spacings, the \( a_i \) in \( W = \sum_{i=1}^{n+1} a_i u_i \) are given by \( a_1 = a_{n+1} = 0 \); \( a_i = i-1 \), \( i = 2, \ldots, p+1 \); \( a_i = n+1-i \), \( i = p+2, \ldots, n \). The range \( \omega \) takes values between 0 and 1 (when \( R = 1 \)), and has a density symmetric about 0.5.

For \( R = 1 \), \( W \) always takes values between 0 and \( p \). Its density is symmetric around \( p/2 \) when \( n = 2p + 1 \); when \( n = 2p \), the mean is \( p^2/(2p+1) \) and the density is slightly skew. The variance is \( p(p+2)/(12(2p+3)) \) for \( n = 2p + 1 \), and is \( p(p^2+p+1)/(6(2p+1)^2) \) for \( n = 2p \). The exact distribution of \( V \), in general, can be found by a method due to Dwass (1961), but it has terms similar to those occurring in \( f_3(R) \) given in Section 2.2; namely terms of the type \( <z>^r \) where \( r \) increases with \( n \). Further the density takes an increasingly complicated form with repeated values of \( a_i \), according to the degree of repetition. Once again, then, for both accuracy and ease of computation, it will be useful to have excellent approximations, and these can be provided by
Pearson curves. The moments of $V$, for a uniform parent population, are obtained from (Stephens, 1972)

$$(n+1)\mu = \sum a_i$$

$$(n+1)(n+2)\sigma^2 = \sum (a_i - \bar{a})^2$$

$$(n+1)(n+2)(n+3)\mu_3 = 2 \sum (a_i - \bar{a})^3$$

$$(n+1)(n+2)(n+3)(n+4)\mu_4 = 6 \sum (a_i - \bar{a})^4 + 3(\sum (a_i - \bar{a})^2)^2$$.

These can be used to fit Pearson curves to $V$. Tables 4 and 5 give points for the range $\omega$, to compare with the exact points, and of the thickened range $W/n$. Also included with $W/n$ is the normal approximation with correct mean and variance; the density of $W/n$ approaches the normal with increasing $n$, while that of the range $\omega$, of course, does not.

Tables 4 and 5 show how very accurate Pearson curves can be expected to be when the distribution approximated is nearly normal, and particularly when it is symmetric. There would be many occasions where, as here, a limiting normal approximation exists, and where a Pearson curve will give a definite improvement in accuracy (as for $W/n$, $n = 5$, Table 5) if one is prepared to calculate the extra moments necessary.

The interest in $W$ is that it can be used to give confidence intervals for $R$ which will be much less sensitive to the presence outlying observations than those based on $\omega$, though of course they will have longer expected lengths. For $n = 10$ and 20 the 90% intervals are

$n = 10: \quad \frac{329W}{.329} < R < \frac{659W}{.659}$

$n = 20: \quad \frac{172W}{.172} < R < \frac{270W}{.270}$. 

21
6. THE RATIO OF RANGE TO STANDARD DEVIATION FOR A NORMAL SAMPLE

We conclude the examples with one where the appropriate Pearson curve is known to be of a different shape from that which it is desired to approximate, but which yet gives excellent results.

Pearson and Stephens (1964) give percentage points of the distribution of \( u = \omega/s \), where \( \omega \) is the range and \( s \) the standard deviation of a single sample of size \( n \) from a normal distribution. For values of \( n \) roughly below 15, exact points were available, but for higher values of \( \omega \) 4-moment P.C. fits were used. A table was included to compare P.C. and exact points for \( u < 15 \), and we reproduce some of the results in Table 6. The statistic is useful as a quick test for normality, particularly against the presence of outliers. The distribution is especially interesting, because it starts, for small \( n \), by being negatively skewed, is nearly symmetrical at \( n = 7 \) or 8, and thereafter becomes slowly more positively skewed.

The \( n = 3 \) situation is interesting. Pearson and Stephens give the P.C. density

\[
f(u) = 0.9573(u - 1.7324)^{-0.0101}(2.000 - u)^{-0.4970}, \quad 1.7324 \leq u \leq 2.0000,
\]

to compare with the known true density

\[
f(u) = \frac{3}{\pi} (1 - u^2/4)^{-1/2}, \quad \sqrt{3} \leq u \leq 2.
\]

It is remarkable how the P.C. curve (a U-shaped Beta distribution) gives
exactly the correct end points; even more interesting is the fact that the P.C. percentage points are remarkably accurate even though at the lower end point the true density is finite \( (\approx 6/\pi) \) while the P.C. approximation is infinite. Why this is so is further discussed in Pearson and Stephens (1964). As \( n \) increases, there continues to be an infinite density at the upper end point \( (n = 4, 5 \text{ for example}) \), and as one might expect, the P.C. gives poor results at the extreme values in this tail. However, for \( n > 6 \) the P.C. gives excellent results in both tails.
7. FINAL REMARKS

7.1 In this section we summarize the main points of this empirical demonstration of Pearson curves.

The main conclusion is that 4-moment fits give strikingly good results, measured in terms of the effective $\alpha$-level, whenever the curve is not very skew, and also in the long tail of a skew curve. For such a curve, the other tail might be short, but not necessarily steep (like, say, $\chi^2_n$); then Pearson curves would give good results in this tail also. (We have not used a simple $\chi^2_p$ as part of Illustration 1 because this is a Pearson curve (Type III) and the results would be overly convincing.) When the tail is steep, the points will be much less reliable.

7.2 Three-moments and an extreme point can also be used; they will probably improve the fit if the tail is steep at the extreme point fitted, but will then give poorer accuracy at the other end of the range.

The advantage of this type of fit seems more to be that it avoids the often considerable effort of working out the fourth moment; but when this fourth moment is available, it is better to use it.

7.3 For the results in this report, only quadratic interpolation in both $\sqrt{\beta_1}$ and $\beta_2$ has been used. This certainly appears adequate, and even linear interpolation, as suggested for the restricted tables in Vol. 1 of B.T.S., would often be enough. Further discussion of this point will be found in both volumes of these tables.

7.4 The accuracy of Pearson curves is extremely difficult to assess in any mathematical way, though it is hoped that a side product of this report might be to inspire further efforts in this direction. For the
moment, it seems as though experience is the best guide; in recent years reliance on experimental tests of techniques which do not lend themselves to mathematical analysis seems to have become more acceptable, perhaps because this situation arises so often in practical life. In any event, we have chosen a wide range of examples to try to show how effective the Pearson curve approximation technique can be. The method is ready for much greater exploitation, and further empirical examination, now that the extended tables and computer programs referred to above are available 7.5 We have not attempted to discuss the properties of Pearson curve approximations to sampled data, when the sample moments are of course very variable estimates of the true values. Some examples of such P.C. fits are in Vols. 1 & 2 of Biometrika Tables. We have sometimes tried to fit such curves to sampling distributions of statistics given by Monte Carlo techniques; in this case there seems to be little advantage over estimating the distribution function directly from the sample and interpolating for percentage points.

7.6 We have also not discussed the approximation of discrete distributions by Pearson curves; there are obvious difficulties of comparison, since, say, a 5% level will not usually be exactly attainable. Nevertheless, if a curve is fitted, it will usually give an \( \alpha \)-level percentage point which, if used with the discrete distribution, gives a true significance level which is very close to \( \alpha \) (often the closest attainable among the discrete set of possible levels). If the available \( \alpha \)'s are close together, such a curve fit could then be a very useful approximation, and it was used (Stephens, 1965) to give the significance points of the two-sample statistic \( U_{NM}^2 \). However, on these
lines, systems of discrete distributions, given by a difference equation analogous to Pearson's differential equation, have been investigated by Katz (1966) and Ord (1967); we hope to pursue a study of the possibilities of such systems.

7.7 Finally, we mention that one of us (M. A. Stephens) has computed Bayesian points for Pearson curves, i.e., significance points $X_L$ and $X_u$, for a given $\alpha$, such that the density function at $X_L$ and $X_u$ has the same values, and the sum of the probabilities in the tails beyond $X_L$ and $X_u$ adds to $\alpha$. Use of these for Bayesian inference might also be worth exploring.
Table 1.a.1

Quadratic Forms. Exact (E) and approximate percentage points for $Q_k(c)$

Approximations: P.C. = 4 moment Pearson Curve fits
I. = Imhof's technique

<table>
<thead>
<tr>
<th>$k$</th>
<th>$c$</th>
<th>Percentage Level $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>.01</td>
</tr>
<tr>
<td>3</td>
<td>1/3, 1/3, 1/3</td>
<td>E.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P.C.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I.</td>
</tr>
<tr>
<td></td>
<td>.3, .3, .4</td>
<td>E.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P.C.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I.</td>
</tr>
<tr>
<td></td>
<td>.2, .3, .5</td>
<td>E.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P.C.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I.</td>
</tr>
<tr>
<td></td>
<td>.1, .4, .5</td>
<td>E.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P.C.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I.</td>
</tr>
<tr>
<td></td>
<td>.1, .3, .6</td>
<td>E.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P.C.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I.</td>
</tr>
<tr>
<td>4</td>
<td>1.0, 1.0, 1.0, 1.0</td>
<td>E.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P.C.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I.</td>
</tr>
<tr>
<td></td>
<td>1.5, 1.5, .5, .5</td>
<td>E.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P.C.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I.</td>
</tr>
<tr>
<td></td>
<td>.2, 1.2, 1.2, .4</td>
<td>E.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P.C.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I.</td>
</tr>
</tbody>
</table>
### Table 1.a.1 (continued)

**Quadratic Forms.** Exact (E) and approximate percentage points for $Q_k(c)$

Approximations: P.C. = 4 moment Pearson Curve fits  
I. = Imhof's technique

<table>
<thead>
<tr>
<th>$k$</th>
<th>$c$</th>
<th>Percentage Level $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>.01</td>
</tr>
<tr>
<td>4</td>
<td>2.5, 0.5, 0.5, 0.5</td>
<td>E.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P.C.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I.</td>
</tr>
<tr>
<td>4</td>
<td>1.8, 1.2, 0.5, 0.5</td>
<td>E.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P.C.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I.</td>
</tr>
<tr>
<td>5</td>
<td>1.0, 1.0, 1.0, 1.0, 1.0</td>
<td>E.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P.C.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I.</td>
</tr>
<tr>
<td>5</td>
<td>1.8, 1.8, 0.6, 0.4, 0.4</td>
<td>E.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P.C.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I.</td>
</tr>
<tr>
<td>5</td>
<td>2.5, 1.0, 0.6, 0.5, 0.4</td>
<td>E.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P.C.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I.</td>
</tr>
<tr>
<td>5</td>
<td>2.8, 0.7, 0.5, 0.5, 0.5</td>
<td>E.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P.C.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P.C. (3,L)</td>
</tr>
<tr>
<td>5</td>
<td>3.0, 0.5, 0.5, 0.5, 0.5</td>
<td>E.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P.C. (3,L)</td>
</tr>
</tbody>
</table>
Table 1.a.2

Quadratic forms. Exact (E) and Approximate Values of $P(Q \leq t)$

$Q_1$ is a noncentral $\chi^2$ variable with d.f. $k$ and noncentrality parameter $\lambda$.

J-S = Jensen-Solomon approximation.
P.C. = 4 moment Pearson curve fit.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>t</th>
<th>E</th>
<th>J-S</th>
<th>P.C.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k=4, \lambda=4$</td>
<td>10</td>
<td>.7118</td>
<td>.7133</td>
<td>.7123</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>.9925</td>
<td>.9927</td>
<td>.9924</td>
</tr>
<tr>
<td>$Q_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k=16, \lambda=8$</td>
<td>20</td>
<td>.3369</td>
<td>.3356</td>
<td>.3368</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>.9632</td>
<td>.9635</td>
<td>.9632</td>
</tr>
<tr>
<td>$Q_k(c)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>($c_1=.4, c_2=.3, c_3=.3$)</td>
<td>0.2</td>
<td>.10472</td>
<td>.10650</td>
<td>.1047</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>.61020</td>
<td>.61048</td>
<td>.6102</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>.96977</td>
<td>.96971</td>
<td>.9698</td>
</tr>
<tr>
<td>$Q_k(c)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>($c_1=.5, c_2=.4, c_3=.1$)</td>
<td>0.2</td>
<td>.12818</td>
<td>.13950</td>
<td>.1293</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>.62822</td>
<td>.62894</td>
<td>.6338</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>.95895</td>
<td>.95855</td>
<td>.9595</td>
</tr>
<tr>
<td>$Q = 0.6x_2^2 + 0.3x_2^2 + 0.1x_2^2$</td>
<td>0.2</td>
<td>.0064</td>
<td>.0089</td>
<td>.0016</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>.6002</td>
<td>.6028</td>
<td>.5999</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>.9839</td>
<td>.9823</td>
<td>.9841</td>
</tr>
<tr>
<td>$Q = 0.6x_2^2 + 0.3x_4^2 + 0.1x_6^2$</td>
<td>1.0</td>
<td>.0334</td>
<td>.0367</td>
<td>.0320</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>.5804</td>
<td>.5837</td>
<td>.5799</td>
</tr>
<tr>
<td></td>
<td>8.0</td>
<td>.9913</td>
<td>.9899</td>
<td>.9914</td>
</tr>
</tbody>
</table>
Table 1.b

Quadratic Forms: Goodness-of-Fit Statistics

Comparison of Exact Points and P.C.

Approximations for Asymptotic Distributions

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Percentage Level $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.01</td>
</tr>
<tr>
<td>$U^2$</td>
<td></td>
</tr>
<tr>
<td>E.</td>
<td>.020</td>
</tr>
<tr>
<td>P.C.</td>
<td>.0241</td>
</tr>
<tr>
<td>$W^2$</td>
<td></td>
</tr>
<tr>
<td>E.</td>
<td>.025</td>
</tr>
<tr>
<td>P.C.</td>
<td>.0443</td>
</tr>
<tr>
<td>$D$</td>
<td></td>
</tr>
<tr>
<td>E.</td>
<td>.441</td>
</tr>
<tr>
<td>P.C.</td>
<td>.4877</td>
</tr>
</tbody>
</table>
Table 2
Vector Resultants

Exact (E) and Approximate Percentage Points for R/N

2 Dimensions

P.C. (4) means using 4 moments. (3,L) and (3,U) mean using three moments, lower or upper end point fixed.

<table>
<thead>
<tr>
<th>N</th>
<th>p = 2</th>
<th>Percentage Level α</th>
<th>.95</th>
<th>.99</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>E</td>
<td>.5402</td>
<td>.6550</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P.C. (4)</td>
<td>.5405</td>
<td>.6543</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P.C. (3,L)</td>
<td>.540</td>
<td>.654</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P.C. (3,U)</td>
<td>.544</td>
<td>.655</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>E</td>
<td>.3846</td>
<td>.4718</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P.C. (4)</td>
<td>.3846</td>
<td>.4717</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P.C. (3,L)</td>
<td>.385</td>
<td>.472</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P.C. (3,U)</td>
<td>.386</td>
<td>.472</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>E</td>
<td>.7542</td>
<td>.8793</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P.C. (4)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3
Vector Resultants

Exact (E) and Approximate Percentage Points for R/N

3 Dimensions

P.C. (4) means using 4 moments. (3,L) and (3,U) mean using three moments, lower or upper end point fixed.

<table>
<thead>
<tr>
<th>N</th>
<th>Percentage Level α</th>
<th>.01</th>
<th>.05</th>
<th>.95</th>
<th>.99</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>E</td>
<td>.0924</td>
<td>.1609</td>
<td>.7002</td>
<td>.8046</td>
</tr>
<tr>
<td></td>
<td>P.C. (4)</td>
<td>.1144</td>
<td>.1658</td>
<td>.6990</td>
<td>.8026</td>
</tr>
<tr>
<td></td>
<td>P.C. (3,L)</td>
<td>.0886</td>
<td>.1578</td>
<td>.6976</td>
<td>.8035</td>
</tr>
<tr>
<td></td>
<td>P.C. (3,U)</td>
<td>.1505</td>
<td>.1809</td>
<td>.7016</td>
<td>.8004</td>
</tr>
<tr>
<td>10</td>
<td>E</td>
<td>.0636</td>
<td>.1109</td>
<td>.5028</td>
<td>.5940</td>
</tr>
<tr>
<td></td>
<td>P.C. (4)</td>
<td>.0675</td>
<td>.1119</td>
<td>.5029</td>
<td>.5937</td>
</tr>
<tr>
<td></td>
<td>P.C. (3,L)</td>
<td>.0621</td>
<td>.1103</td>
<td>.5026</td>
<td>.5938</td>
</tr>
<tr>
<td></td>
<td>P.C. (3,U)</td>
<td>.0902</td>
<td>.1201</td>
<td>.5044</td>
<td>.5936</td>
</tr>
</tbody>
</table>
Table 4

Percentage Points of the Range \( w \) from a Uniform Distribution with Limits 0, \( \sqrt{12} \)

\( E = \) Exact; P.C. = Pearson curves

<table>
<thead>
<tr>
<th>( n )</th>
<th>.01</th>
<th>.05</th>
<th>.10</th>
<th>.90</th>
<th>.95</th>
<th>.99</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>.7693</td>
<td>1.1868</td>
<td>1.4414</td>
<td>3.0753</td>
<td>3.1993</td>
<td>3.3509</td>
</tr>
<tr>
<td></td>
<td>.7693</td>
<td>1.1868</td>
<td>1.4414</td>
<td>3.0753</td>
<td>3.1993</td>
<td>3.3508</td>
</tr>
<tr>
<td>10</td>
<td>1.7170</td>
<td>2.0987</td>
<td>2.2972</td>
<td>3.2752</td>
<td>3.3367</td>
<td>3.4103</td>
</tr>
<tr>
<td></td>
<td>1.7170</td>
<td>2.0987</td>
<td>2.2972</td>
<td>3.2752</td>
<td>3.3367</td>
<td>3.4103</td>
</tr>
<tr>
<td>20</td>
<td>2.4637</td>
<td>2.7154</td>
<td>2.8372</td>
<td>3.3709</td>
<td>3.4015</td>
<td>3.4378</td>
</tr>
<tr>
<td></td>
<td>2.4637</td>
<td>2.7155</td>
<td>2.8373</td>
<td>3.3709</td>
<td>3.4015</td>
<td>3.4378</td>
</tr>
</tbody>
</table>
Table 5
Percentage Points of W/n

E = exact; P.C. = Pearson curves; N = normal approximation

<table>
<thead>
<tr>
<th>n</th>
<th></th>
<th></th>
<th>Percentage Level α</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>.01</td>
<td>.05</td>
</tr>
<tr>
<td>5</td>
<td>E.</td>
<td>.0630</td>
<td>.0975</td>
</tr>
<tr>
<td></td>
<td>P.C.</td>
<td>.0632</td>
<td>.0982</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>.0564</td>
<td>.0985</td>
</tr>
<tr>
<td>8</td>
<td>E.</td>
<td>.1056</td>
<td>.1375</td>
</tr>
<tr>
<td></td>
<td>P.C.</td>
<td>.1057</td>
<td>.1377</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>.1013</td>
<td>.1367</td>
</tr>
</tbody>
</table>
Table 6
Ratio of Range to Standard Deviation
In a Normal Sample of Size n

*E* = exact; *P.C.* = Pearson curves

<table>
<thead>
<tr>
<th>n</th>
<th>0.01</th>
<th>0.05</th>
<th>0.95</th>
<th>0.99</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>E</td>
<td>1.7373</td>
<td>1.7576</td>
<td>1.9993</td>
</tr>
<tr>
<td></td>
<td>P.C.</td>
<td>1.7375</td>
<td>1.7575</td>
<td>1.9993</td>
</tr>
<tr>
<td>4</td>
<td>E</td>
<td>------</td>
<td>------</td>
<td>2.429</td>
</tr>
<tr>
<td></td>
<td>P.C.</td>
<td>------</td>
<td>------</td>
<td>2.431</td>
</tr>
<tr>
<td>5</td>
<td>E</td>
<td>------</td>
<td>------</td>
<td>2.753</td>
</tr>
<tr>
<td></td>
<td>P.C.</td>
<td>------</td>
<td>------</td>
<td>2.750</td>
</tr>
<tr>
<td>8</td>
<td>E</td>
<td>------</td>
<td>------</td>
<td>3.399</td>
</tr>
<tr>
<td></td>
<td>P.C.</td>
<td>------</td>
<td>------</td>
<td>3.398</td>
</tr>
</tbody>
</table>
References


JOHNSON, N. L. and KOTZ, S. (1968). Tables of Distributions of Positive
Definite Quadratic Forms in Central Normal Variables. Sankhyā, Series
B, 30, 303-314.

JOHNSON, N. L., NIXON, E., AMOS, D. E., and PEARSON, E. S. (1963). Table
of percentage points of Pearson curves, for given \( \beta_1 \) and \( \beta_2 \), expressed

KATZ, LEO. (1966). Unified treatment of a broad class of discrete proba-
bility distributions. Classical and Contagious Discrete Distributions

of distributions of quadratic forms in normal variables. I. Central

of distributions of quadratic forms in normal variables. II. Non-

and distributions of some related functionals on Brownian motion.

ORD, J. K. (1967). On a system of discrete distributions. Biometrika,
54, 649-656.

OWEN, D. B. (1962). Handbook of Statistical Tables. Addison-Wesley:
Reading, Massachusetts.

PEARSON, E. S. and HARTLEY, H. O. (1962). Biometrika Tables for Statisti-

PEARSON, E. S. and HARTLEY, H. O. (1972). Biometrika Tables for Statisti-

PEARSON, E. S. and STEPHENS, M. A. (1964). The ratio of range to standard
development in the same normal sample. Biometrika, 51, 484-487.


SOLOMON, H. (1960). Distribution of quadratic forms--tables and applica-
Laboratories, Stanford University, Stanford, California.


**REPORT DOCUMENTATION PAGE**

1. REPORT NUMBER  

   226  

2. GOVT ACCESSION NO.  

3. RECIPIENT'S CATALOG NUMBER  

4. TITLE (and Subtitle)  

   Pearson Curves Revisited  

5. TYPE OF REPORT & PERIOD COVERED  

   Technical Report  

6. PERFORMING ORG. REPORT NUMBER  

7. AUTHOR(s)  

   M. A. Stephens and H. Solomon  

8. CONTRACT OR GRANT NUMBER(s)  

   N00014-67-A-0112-0085  

9. PERFORMING ORGANIZATION NAME AND ADDRESS  

   Department of Statistics  
   Stanford University  
   Stanford, Calif. 94305  

10. PROGRAM ELEMENT, PROJECT, TASK  
    AREA & WORK UNIT NUMBERS  

    (N0-042-267)  

11. CONTROLLING OFFICE NAME AND ADDRESS  

   Office of Naval Research  
   Statistics & Probability Program  
   Code 436  
   Arlington, Virginia 22217  

12. REPORT DATE  

   November 5, 1975  

13. NUMBER OF PAGES  

   37  

14. MONITORING AGENCY NAME & ADDRESS (IF DIFFERENT FROM CONTROLLING OFFICE)  

15. SECURITY CLASS. (OF THIS REPORT)  

   Unclassified  

16. DISTRIBUTION STATEMENT (OF THIS REPORT)  

   Approved for Public Release; Distribution Unlimited  

17. DISTRIBUTION STATEMENT (OF THE ABSTRACT ONLY, IF DIFFERENT FROM REPORT)  

18. SUPPLEMENTARY NOTES  

19. KEYWORDS ( Continue on reverse side if necessary and identify by block number)  

   Pearson curves, approximating distributions, moments, curve-fitting techniques  

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  

   This report discusses the use of Pearson curves to give approximate percentage points to intractable distributions, when the first four moments (or three moments and one end-point) are available. It is shown how to fit the curves, and their effectiveness is illustrated with a wide range of examples; they are in general shown to give excellent approximations in the long tail(s) of a distribution.