TWO METHODS OF CONSTRUCTING EXACT TESTS, WITH APPLICATIONS TO TESTS OF GOODNESS OF FIT, NORMALITY, SERIAL INDEPENDENCE AND RANDOMNESS OF A SERIES OF EVENTS

By

J. DURBIN

TECHNICAL REPORT NO. 4
September 27, 1960

PREPARED UNDER THE AUSPICES OF
NATIONAL SCIENCE FOUNDATION GRANT G-9670
TWO METHODS OF CONSTRUCTING EXACT TESTS,
WITH APPLICATIONS TO TESTS OF GOODNESS OF FIT,
NORMALITY, SERIAL INDEPENDENCE AND RANDOMNESS OF A SERIES OF EVENTS

by

J. Durbin

TECHNICAL REPORT NO. 4

September 27, 1960

PREPARED UNDER THE AUSPICES

OF

NATIONAL SCIENCE FOUNDATION GRANT G-9670

APPLIED MATHEMATICS AND STATISTICS LABORATORIES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
TWO METHODS OF CONSTRUCTING EXACT TESTS, WITH APPLICATIONS TO
TESTS OF GOODNESS OF FIT, NORMALITY, SERIAL INDEPENDENCE
AND RANDOMNESS OF A SERIES OF EVENTS*

by

J. Durbin

Research Technique Division, London School of Economics

1. Introduction

For testing the hypothesis that a set of observations \( x_1, \ldots, x_n \)
is a random sample from a distribution having a specified continuous
distribution function \( F(x) \) a variety of procedures are available, the
most celebrated being Pearson's \( \chi^2 \) test. The \( \chi^2 \) test is flexible
and easy to use but it possesses an element of arbitrariness in the choice
of group boundaries; moreover it does not have the desirable property of
being an exact test in the sense of giving exact probabilities of rejection
when the null hypothesis is true. The Kolmogorov test, which has been
described and compared with related tests in an excellent article by
Darling [1957], is free from these objections and is known to have good
asymptotic power properties against alternatives specified in terms of
distance between distribution functions. Nevertheless, in the author's
experience its performance in practice with samples of moderate size has
been disappointing in that it has frequently failed to give a significant

* This paper is based mainly on research done at the University of North
Carolina, Chapel Hill with the support of the Office of Naval Research. The
principal results were presented in an invited address delivered to the
Institute of Mathematical Statistics and the Biometric Society at New York,
April 26, 1960. The work was completed and the paper written at Stanford
University with the support of the National Science Foundation.
result when the $X^2$ test has done so. Some empirical results which illustrate this tendency will be given later in the paper.

The reason for the disappointing performance of the Kolmogorov test is not far to seek. If the alternative distribution is such that the difference between the hypothetical distribution function and the alternative distribution function is relatively large for some value of the variate, one would anticipate that Kolmogorov's test would have high power. Such would be the case, for example, if the two distributions differed only in location. If, however, the difference between the two distribution functions is nowhere large, as is quite possible if their means and variances are the same even though the two frequency functions differ markedly in shape, one would not anticipate that Kolmogorov's statistic would be a powerful discriminator between the two hypotheses.

Putting $u_j = F(x_j)$ ($j = 1, \ldots, n$), the hypothesis that $x_1, \ldots, x_n$ have distribution function $F(x)$ is equivalent to the hypothesis that $u_1, \ldots, u_n$ are $U(0,1)$ variables i.e., are uniformly distributed on the interval $(0,1)$. Let $u_1, \ldots, u_n$ be ordered to give the ordered values $u(1) \leq u(2) \leq \ldots \leq u(n)$. The first part of this paper shows how to transform $u(1), \ldots, u(n)$ to a new set of values $v_1, \ldots, v_n$ which have the same joint distribution as $u(1), \ldots, u(n)$ when the null hypothesis is true but which give a set of points tending to be more heavily concentrated towards the left-hand end of the $(0,1)$ interval than towards the right-hand end under a wide class of alternatives. A variety of procedures can now be used for testing the hypothesis that $v_1, \ldots, v_n$ are ordered $U(0,1)$ variables, special interest being attached in this
paper to the one-sided Kolmogorov test. An indication is given in
section 6 of applications to some problems other than those of goodness
of fit, namely to tests based on the periodogram in time-series analysis
and to tests of the distribution of the intervals between successive events.

Because of wide applicability to problems in statistics and probability
as well as for purely mathematical interest, distributions related to
points scattered at random in an interval have been extensively studied
in many different guises. No attempt will be made here to review the
entire field; the following list of references is intended merely to
indicate work which seems to the author to be fairly closely connected
with the subject-matter of this paper. Among papers concerned specifically
with the random division of an interval we may mention those by Moran
[1947, 1951], Sherman [1950], Mauldon [1951], Darling [1953], Irwin [1955]
and Barton & David [1955, 1956]. The following papers are concerned with
goodness-of-fit aspects: K. Pearson [1933], Neyman [1937], E.S. Pearson
[1939] and Darling [1957]. Authors dealing with similar problems arising
in the study of exponential distributions are Sukhatme [1936, 1937],
Epstein & Tsao [1953], Epstein & Sobel [1954], Bartholemew [1956] and
Epstein [1960]. For applications to Poisson processes we mention the papers
by Greenwood [1946], Maguire et al [1952, 1953], Barnard [1953], Cox [1955],
and Bartholemew [1956]. Relevant papers concerned primarily with the order
statistics are those by Wilks [1948], Malmquist [1951], Renyi [1953] and
van Dantzig [1954].
A severe limitation to the use hitherto of the Kolmogorov test has been that it has only been available for tests of simple hypotheses. Thus it has not been usable even for the classical problem of testing for normality, except as an approximation, since the hypothesis is a composite one involving two nuisance parameters, the population mean and variance. In section 6 a general method is proposed for eliminating nuisance parameters thereby enabling us to convert a composite hypothesis into a simple hypothesis. To illustrate the method, suppose we wish to test whether \( x_1, \ldots, x_n \) are independent observations from a \( \text{N}(\mu, \sigma^2) \) distribution (i.e., a normal distribution with mean \( \mu \) and variance \( \sigma^2 \)), where \( \mu \) and \( \sigma^2 \) are unknown nuisance parameters. Let \( \bar{x}, s^2 \) denote the sample mean and variance and let \( \bar{x}', s'^2 \) denote observations of random variables independent of \( x_1, \ldots, x_n \) which have the same distribution as the mean and variance of a sample of \( n \) from a \( \text{N}(0,1) \) distribution. Define \( x_1', \ldots, x_n' \) by the relations

\[
\frac{x_j - \bar{x}'}{s'} = \frac{x_j - \bar{x}}{s} \quad (j = 1, \ldots, n)
\]

It is shown below that \( x_1', \ldots, x_n' \) are independent \( \text{N}(0,1) \) variables. Consequently the hypothesis of normality can be tested by applying Kolmogorov's test, or preferably the modified Kolmogorov test proposed in section 3, to \( x_1', \ldots, x_n' \). The price we have paid for the elimination of the nuisance parameters has been the introduction of an extraneous element of randomisation in the values of \( \bar{x}', s'^2 \).
2. Re-ordering of intervals

Suppose we have a sample of independently and identically distributed observations \( x_1, \ldots, x_n \) and wish to test the hypothesis that they come from a distribution with continuous distribution function \( F(x) \). When the null hypothesis is true the values \( u_j = F(x_j) \) \( (j = 1, \ldots, n) \) are independent \( U(0,1) \) variables. This implies that the \( n \) points determined by \( u_1, \ldots, u_n \) are randomly scattered on the \( (0,1) \) interval. One would expect a departure from the null hypothesis to be indicated by a tendency for some of the intervals between adjacent points to be shorter and for some of the intervals between adjacent points to be longer than is found on the hypothesis of random scatter. This suggests that we take as our starting point a study of the relative lengths of the intervals between adjacent points.

Denoting the ordered \( u \)'s by \( u(1), \ldots, u(n) \), these have the distribution

\[
dP = n! \, du(1) \ldots du(n) \quad (u(1) \leq u(2) \leq \cdots \leq u(n)). \tag{1}
\]

Let \( c_1 = u(1) \), \( c_j = u(j) - u(j-1) \) \( (j = 2, \ldots, n) \) and \( c_{n+1} = 1 - u(n) \). Since the Jacobian of the transformation from \( u(1), \ldots, u(n) \) to \( c_1, \ldots, c_n \) is unity, the \( c \)'s have the distribution

\[
dP = n! \, dc_1 \ldots dc_n \quad (c_1 \geq 0, \ldots, c_{n+1} \geq 0, \sum_{j=1}^{n+1} c_j = 1). \tag{2}
\]

A thorough discussion of this distribution has been given by Wilks (1948).
Note that we avoid difficulties arising from the fact that the distribution of $c_1,\ldots,c_{n+1}$ is singular by writing down the probability element for $c_1,\ldots,c_n$; the same device is used below for obtaining the distribution of $c(1),\ldots,c(n+1)$ and $g_1,\ldots,g_{n+1}$.

Since we are interested in the relative magnitudes of the $c$'s it is natural to consider the ordered $c$'s, $c(1) \leq c(2) \leq \cdots \leq c(n+1)$. Since there are $(n+1)!$ ways of permuting $n$ objects from $n+1$ the joint distribution of $c(1),\ldots,c(n)$ is, using (2),

$$dP = (n+1)! \, n! \, dc(1)\cdots dc(n) \left( 0 \leq c(1) \leq \cdots \leq c(n+1), \sum_{j=1}^{n+1} c(j) = 1 \right). \quad (3)$$

We now transform $c(1),\ldots,c(n+1)$ into a more manageable form by means of the transformation

$$g_j = (n+2-j)(c_j - c_{j-1}) \quad (c_0 = 0; \ j = 1,\ldots,n+1). \quad (4)$$

The Jacobian of the transformation is $(n+1)!$. Moreover each $g_j \geq 0$ and

$$\sum_{j=1}^{n+1} g_j = \sum_{j=1}^{n+1} c(j) = 1.$$ 

Consequently the distribution of $g_1,\ldots,g_n$ is

$$dP = n! \, dg_1 \cdots dg_n \left( g_1 \geq 0, \ldots, g_{n+1} \geq 0, \sum_{j=1}^{n+1} g_j = 1 \right). \quad (5)$$
Comparing (5) with (2) we see that the two distributions are identical. We therefore have the remarkable result that \( g_1, \ldots, g_{n+1} \), which depend on the ordered intervals, have the same distribution as the unordered intervals \( c_1, \ldots, c_{n+1} \). This result provides the key to the methods of test construction proposed in this and the following section. It depends essentially on the transformation (4) introduced by Sukhatme [1937] in the study of ordered exponentially-distributed variables and used by Mauldon [1951] and Dwass [1959] in the study of random intervals.

Putting

\[
  w_r = \sum_{j=1}^{r} g_j ,
\]

(6)

it follows that \( w_1, \ldots, w_n \) have the same distribution as the ordered \( U(0,1) \) variables \( u(1), \ldots, u(n) \). Consequently any test procedure depending on \( u(1), \ldots, u(n) \), such as the Kolmogorov test, has the same properties under the null hypothesis when it is based on \( w_1, \ldots, w_n \) as when it is based on \( u(1), \ldots, u(n) \).

It might be asked what has been gained by transforming to a set of values which have the same distribution as have the values we started with. The answer is that we hope to gain power. Some considerations which indicate why one would expect power to be increased in many situations will be discussed in the next two paragraphs.

The type of alternative we have in mind in this paper is one in which small intervals tend to be smaller and large intervals larger than on the null hypothesis. A way of making this idea precise is the following.
Let primes denote values obtained on the alternative hypothesis, i.e., \( c_j', c_j', g_j' \) and \( w_j' \) are the values on the alternative hypothesis corresponding to values \( c_j, c_j, g_j \) and \( w_j \) on the null hypothesis. Suppose that whenever \( c_i < c_j \) then \( c_i' / c_j' < c_i / c_j \), i.e., ratios of larger to smaller intervals tend to become magnified on the alternative hypothesis. We shall show that this implies that \( w_j' < w_j \) \((j = 1, \ldots, n)\). For \( c_j(j) < c_j(j+1) \), so that \( c_j(j) / c_j(j+1) < c_j(j) / c_j(j+1) \) \((j = 1, \ldots, n)\) (we ignore the event \( c_j(j) = c_j(j+1) \) since this has probability zero).

Thus

\[
\frac{c_{j+1} - c_j}{c_{j} - c_{j-1}} = \frac{c_{j+1}' / c_j' - 1}{1 - \frac{c_j'}{c_{j-1}' / c_j'}} > \frac{c_{j+1} / c_{j} - 1}{1 - \frac{c_{j-1}}{c_{j}' / c_{j}}}
\]

Consequently

\[
\frac{c_{j+1} - c_j}{c_{j+1} - c_j} > \frac{c_j - c_{j-1}}{c_j - c_{j-1}}, \text{ so that}
\]

\[
\frac{g_{j+1}'}{g_{j+1}} > \frac{g_j'}{g_j}, \quad (j = 1, \ldots, n)
\]

i.e., \( g_j' / g_j \) is an increasing function of \( j \). But \( \sum_{1}^{n+1} g_j' = \sum_{1}^{n+1} g_j = l \).
Consequently \( g'_1/g_1 < 1 \) and \( g'_{n+1}/g_{n+1} > 1 \). These results together imply that there is a value \( r \) such that \( g'_i \leq g_i \) for \( i = 2, \ldots, r \) and \( g'_i > g_i \) for \( i > r \) where \( 1 \leq r < n + 1 \). Since \( g'_i < g_i \) for \( j \leq r \) we have

\[
w'_j = \sum_{i=1}^{j} g'_i < \sum_{i=1}^{j} g_i = w_j ,\text{ while for } j > r \text{ we have } w'_j = 1 - \sum_{i=j+1}^{n+1} g'_i < 1 - \sum_{i=j+1}^{n+1} g_i = w_j .\text{ Thus } w'_j < w_j \text{ for all } j .
\]

It is remarkable that assuming only that ratios of larger to smaller
intervals become greater as we proceed from the null to the alternative
hypothesis we have been able to show that all values of the \( w' \)'s are
diminished. Consequently the sample distribution function calculated from
the \( w' \)'s is never less on the alternative hypothesis than on the null
hypothesis. It is unlikely that the sample distribution function of the
original \( u \)'s would give as clear an indication of departure from the null
hypothesis as that of the \( w' \)'s except perhaps for special types of departure
such as those arising from change of location, for the detection of which a
goodness-of-fit test would not normally be employed.

3. Some new tests of goodness-of-fit

From (4) and (6) we have

\[
w_j = c(1) + \cdots + c(j-1) + (n+2-j) c(j) \quad (j = 1, \ldots, n) \quad (7)
\]

where \( c(1) \leq \cdots \leq c(n+1) \) is the ordered set of intervals. The goodness-
of-fit tests proposed in this section are all tests of the hypothesis that
are distributed as ordered $U(0,1)$ variables. Large numbers of such tests could be devised; here we mention only three which have special features of interest. The first is included on account of its simplicity and because it requires only the widely-available tables of the $F$-distribution. We call it the modified median test since it is based on the median of the $w$'s. Consider the distribution of $w_r$. The probability element required is the probability that in a sample of $n$ uniform variables $r - 1$ are less than $w_r$, one is in the range $w_r, w_r + dw_r$ and $n - r$ are greater than $w_r$, i.e.

$$dP = \frac{n!}{(r-1)! (n-r)!} w_r^{r-1} (1 - w_r)^{n-r} dw_r \quad (0 \leq w_r \leq 1) \quad (8)$$

Thus $M_r = \frac{r}{n+1-r} \frac{1 - w_r}{w_r}$ has Fisher's variance-ratio distribution with $2(n+1-r)$, $2r$ degrees of freedom. The value of $w_r$ for any particular $r$ can be used as a test statistic, $r$ being at our disposal. The author recommends taking the median value, i.e. $r = \frac{1}{2} (n + 1)$ for $n$ odd and $r = \frac{1}{2} n$ (or $\frac{1}{2} (n + 1)$) for $n$ even. Since on the alternative we would expect $w$ to be smaller than on the null hypothesis a one-sided $F$-test is appropriate.

More important is the second of the three tests which we call the modified Kolmogorov test. This is obtained by considering the difference between the sample and population distribution functions corresponding to $w_1, \ldots, w_n$. 


The test statistic is

\[ K_m = \max_{r=1, \ldots, n} \left( \frac{r}{n} - w_r \right) \]  

(9)

Since we expect \( K_m \) to increase as we depart from the null hypothesis a one-sided test is appropriate. Consequently the test procedure is to reject when \( K_m \) is greater than the value tabulated for a one-sided Kolomogorov test. A suitable table has been given by Miller (1956).

The third test comes from the observation that \( \prod_{j=1}^{n} u_{(j)} = \prod_{j=1}^{n} u_j \).

Since \( \prod_{j=1}^{n} w_j \) is distributed like \( \prod_{j=1}^{n} u_{(j)} \) it therefore has the same distribution as \( \prod_{j=1}^{n} u_j \). Karl Pearson (1933) pointed out that \( \prod_{j=1}^{n} u_j \) is distributed like \( \exp\left(-\frac{1}{2} \chi^2\right) \), where \( \chi^2 \) has the chi-square distribution with \( 2n \) degrees of freedom, and hence can be used as the statistic for an exact test of goodness-of-fit. Applying Pearson's suggestion to \( \prod_{j=1}^{n} w_j \) we obtain a third exact test which we call the modified probability-product test. The test statistic is

\[ P_m = -2 \log \prod_{j=1}^{n} w_j \]  

(10)

which is tested as a chi-square variate with \( 2n \) degrees of freedom.
The power of the original probability-product test was studied by E.S. Pearson (1939) and in the light of his results it seems likely that the $P_m$ test will in theory have high power against a wide class of alternatives. However, the author does not recommend its use in practice in the form presented above since the value of the statistic $P_m$ seems to be too much affected by rounding errors in $u_1, \ldots, u_n$. It is possible that the test might be altered slightly so as to avoid this difficulty but the best way of doing so is not obvious to the author.

Some empirical results obtained by applying the first two tests to artificial data are given in the next section.

4. Some empirical results

The performance of the $M_r$ and $K_m$ tests proposed in section 3 has been compared with the performance of the $\chi^2$ and ordinary Kolmogorov tests on five samples of 50 observations from each of the following three distributions,

- **Exponential**: $\exp(-(x+1)) \, dx \quad (x \geq -1)$
- **Laplace**: $2^{-3/2} \exp(-2^{-1/2} |x|) \, dx \quad (-\infty \leq x \leq \infty)$
- **Normal**: $(2\pi)^{-1/2} \exp(-\frac{1}{2} x^2) \, dx \quad (-\infty \leq x \leq \infty)$

Each distribution has mean zero and variance unity and the null hypothesis is that the true distribution is the $\mathcal{N}(0,1)$ distribution defined by (13). Random deviates from these distributions have been tabulated by Quenouille (1959) and the five sets of three samples in this experiment are the values
given on the first five pages of Quenouille's tables. Deviates for distributions (11) and (12) were obtained by Quenouille by transformation of the corresponding deviates from distribution (13); consequently the samples from the three distributions correspond in the sense of being composed of values with the same probability integral transforms.

The following four tests were applied to each sample of 50 observations:

(a) The $\chi^2$ test, taking ten groups with an expected number of five in each group.

(b) The two-sided Kolmogorov test based on the statistic

$$K = \max_{j=1, \ldots, n} |S(x_j) - F(x_j)|,$$

where $S(x)$ is the sample distribution function and $F(x)$ is the hypothetical distribution function. Significance was assessed by referring to the tables of critical values of this statistic published by Miller (1956).

(c) The $M_r$ test proposed in section 3 taking $r = 25$.

(d) The modified Kolmogorov test based on the statistic $K_m$ defined by (9).

The results are given in Table 1. The entries are values of the appropriate test statistics. Significance at the 5% and 1% levels are denoted by single and double asterisks respectively.

With due reservations owing to the small number of samples considered, the results suggest the following tentative conclusions for the sample size and alternatives considered:
Table I: Comparison of performances of new and existing tests on artificial samples

<table>
<thead>
<tr>
<th>Population</th>
<th>Test</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Exponential</td>
<td>$\chi^2$</td>
<td>17.2*</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>$M_{25}$</td>
<td>2.20**</td>
</tr>
<tr>
<td></td>
<td>$K_m$</td>
<td>0.23**</td>
</tr>
<tr>
<td>Laplace</td>
<td>$\chi^2$</td>
<td>26.8**</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>0.19*</td>
</tr>
<tr>
<td></td>
<td>$M_{25}$</td>
<td>2.38**</td>
</tr>
<tr>
<td></td>
<td>$K_m$</td>
<td>0.25**</td>
</tr>
<tr>
<td>Normal</td>
<td>$\chi^2$</td>
<td>7.2</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>$M_{25}$</td>
<td>1.62*</td>
</tr>
<tr>
<td></td>
<td>$K_m$</td>
<td>0.16</td>
</tr>
</tbody>
</table>

The critical values of the four statistics are:

<table>
<thead>
<tr>
<th></th>
<th>$\chi^2$</th>
<th>K</th>
<th>$M_{25}$</th>
<th>$K_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5%</td>
<td>16.92</td>
<td>0.188</td>
<td>1.609</td>
<td>0.170</td>
</tr>
<tr>
<td>1%</td>
<td>21.67</td>
<td>0.226</td>
<td>1.967</td>
<td>0.211</td>
</tr>
</tbody>
</table>
(i) The unmodified Kolmogorov test is less powerful than the \( \chi^2 \) test.

(ii) The modified median test is more powerful than the unmodified Kolmogorov test but slightly less powerful than the \( \chi^2 \) test.

(iii) The modified Kolmogorov test is more powerful than the unmodified Kolmogorov test, slightly more powerful than the modified median test and about as powerful as the \( \chi^2 \) test.

In making these comparisons with the \( \chi^2 \) test it should be remembered that the \( \chi^2 \) test is not an exact test whereas the other three tests are exact. There is no suggestion that the \( \chi^2 \) test should be used as an absolute standard; it is merely a convenient yardstick which is used because of familiarity among statisticians. Moreover, the author wishes to make it clear that in presenting these results it has not been his intention to make any general statements regarding the relative powers of the various procedures; the intention has been merely to summarise the performance of the tests under the conditions of this experiment.

5. Applications to time-series and arrival-time distributions

A further application of the foregoing theory is to testing the hypothesis that a series of identically distributed normal variables have uniform spectral density, i.e. are serially independent. In the spectral approach to time series the basic statistic is the periodogram
\[ p_j = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} x_t \exp \left( \frac{(2\pi j t)/T}{2} \right) \right| \]

\[(j = 0, 1, \ldots, T/2 \text{ for } T \text{ even}) \]

\[(j = 0, 1, \ldots, (T-1)/2 \text{ for } T \text{ odd}) \]  

Let \( m = \frac{1}{2} T - 1 \) for \( T \) even and \( m = \frac{1}{2} (T - 1) \) for \( T \) odd. On the null hypothesis that \( x_1, \ldots, x_T \) are independent \( N(\mu, \sigma^2) \) variables the quantities \( 2\pi \rho_j/\sigma^2 \) \( (j = 1, \ldots, m) \) are independent exponential variables with density \( \exp(-x) \). Putting \( c_j = p_j/\sum_{i=1}^{m} p_i \) \( (j = 1, \ldots, m) \) we have the well-known result that \( c_1, \ldots, c_m \) are distributed like the intervals between successive ordered \( U(0, 1) \) variables, i.e. they have the distribution \( (2) \) with \( n + 1 = m \). Transforming to \( c(1), \ldots, c(m) \), then to \( e_1, \ldots, e_{m-1} \) and finally to \( w_1, \ldots, w_{m-1} \) as in section 2, we have reduced the problem to testing the hypothesis that \( w_1, \ldots, w_{m-1} \) are ordered \( U(0, 1) \) variables. For this purpose the modified Kolmogorov test recommended in section 3 is appropriate.

It might be asked what advantage this test is likely to have over existing exact tests such as the von Neumann test (1941) based on the statistic \( \frac{\sigma^2}{s^2} = \frac{T}{T-1} \sum_{t=2}^{T} (x_t - x_{t-1})^2 / \left( \sum_{t=1}^{T} (x_t - \bar{x})^2 \right) \), Fisher's test (1929) based on the largest of the \( c_j \)'s, or indeed the ordinary Kolmogorov test applied to the unordered \( c \)'s. The answer is similar to that given for the goodness-of-fit problem in section 3, namely that we expect it to have high power against a wider range of alternatives.
The method applies also in a rather obvious way to testing the hypothesis that events are occurring in a Poisson process. Suppose that one observes a Poisson process for a fixed length of time $T$ and that $n$ events occur at times $t_1, t_2, \ldots, t_n$ during the interval $(0, T)$. Let $u(j) = t_j/T (j = 1, \ldots, n)$. The $u(1), \ldots, u(n)$ are distributed in the form (1). Alternatively, suppose one fixes $n$ in advance and starting at time zero, records the times $t_1, \ldots, t_{n+1}$ at which $n + 1$ events occur. Let $u(j) = t_j/t_{n+1} (j = 1, \ldots, n)$. Then $u(1), \ldots, u(n)$ are distributed in the form (1). Simple proofs of these results are given by Epstein (1960, Appendix 1). In either case one can transform to the $c$'s, the $g$'s and the $w$'s as in sections 2 and 3 and employ the modified Kolmogorov test in the way described.

6. Tests based on systematic subsampling of the observations

Most of the work required to carry out the goodness-of-fit tests described in section 3 is likely to occur in making the probability integral transformation from the $x$'s to the $u$'s; indeed when the sample is large the amount of labour required can be prohibitively great. A way of reducing the work substantially is to confine the analysis to a systematic subsample of the observations. Suppose that the sample size is now denoted by $N$. Instead of transforming all the $x$'s we first arrange them in order of magnitude, giving $x(1) \leq x(2) \leq \cdots \leq x(N)$ say, and then transform every $k$th, i.e. we calculate $u(k) = F(x(k))$, $u(2k) = F(x(2k))$, $\ldots$, $u(nk) = F(x(nk))$, where $k$ is a suitable chosen integer and $n$ is the largest integer satisfying $nk + k - 1 \leq N$. 

Suppose for simplicity that \( N = nk + k - 1 \); otherwise reject
\( N - nk - k + 1 \) observations at random. The distribution of \( u_{(k)}, \ldots, u_{(nk)} \) is (see for example Wilks, (1948))

\[
dP = \frac{(nk + k - 1)!}{((k-1)!)^{n+1}} \left(\frac{k-1}{u_{(k)}} \prod_{j=1}^{n-1} (u_{(jk+k)} - u_{(jk)})^{k-1} (1 - u_{(nk)})^{k-1} \right) \, du_{(k)} \cdots du_{(nk)}
\]

\[
(u_{(k)} \leq u_{(2k)} \leq \cdots \leq u_{(nk)}) \tag{15}
\]

Let

\[
p_j = \frac{u_{(jk)}}{u_{(jk+k)}} \quad (j = 1, \ldots, n - 1)
\]

\[
p_n = u_{(nk)} \tag{16}
\]

The Jacobian is given by

\[
\frac{\partial(u_{(k)}, \ldots, u_{(nk)})}{\partial(p_1, \ldots, p_n)} = \frac{n}{\prod_{j=2}^{n} u_{(jk)}} \tag{17}
\]

The distribution of \( p_1, \ldots, p_n \) is, therefore,

\[
dP = \frac{(nk + k - 1)!}{((k-1)!)^{n+1}} \prod_{j=1}^{k-1} (1-p_1)^{k-1} \, dp_1 \prod_{j=2}^{2k-1} (1-p_2)^{k-1} \, dp_2 \cdots \prod_{j=nk}^{nk-1} (1-p_n)^{k-1} \, dp_n
\]

\[
(0 \leq p_j \leq 1, \text{ all } j) \tag{17}
\]

Thus \( p_1, \ldots, p_n \) are independently distributed, the distribution of \( p_j \) being of the Beta form.
\[ dP = (B(jk, k))^{-1} p_j^{jk-1} (1 - p_j)^{k-1} dp_j . \]  \hspace{1cm} (18)

Let \( z_j \) denote the probability integral transformation of \( p_j \), i.e.

\[ z_j = (B(jk, k))^{-1} \int_0^{p_j} x^{j-1} (1 - x)^{k-1} dx \quad (j = 1, \ldots, n) . \]  \hspace{1cm} (19)

Then \( z_1, \ldots, z_n \) are independent \( U(0, 1) \) variables. The actual values of the \( z \)'s determined by (19) may be obtained from the Tables of the Incomplete Beta Function (Pearson, 1948).

Let

\[ q_j = z_j^{1/j} \quad (j = 1, \ldots, n) . \]  \hspace{1cm} (20)

The joint distribution of \( q_1, \ldots, q_n \) is seen to be

\[ dP = n! \prod_{j=1}^{n} q_j^{2j-1} \prod_{i=1}^{n-1} dq_i \quad (0 \leq q_j \leq 1, \text{ all } j) . \]  \hspace{1cm} (21)

This is of the form (17) with \( k = 1 \). Let

\[ v_j = \prod_{i=j}^{n} q_i . \]  \hspace{1cm} (22)

We observe that \( q_j = v_j^{1/j} \) and \( q_n = v_n \). Thus (21) is equivalent to (17) with \( k = 1 \). Consequently the distribution of \( v_1, \ldots, v_n \) is given by putting \( k = 1 \) in (15), i.e.

\[ dP = n! \prod_{i=1}^{n} dv_i \quad (v_1 \leq v_2 \leq \cdots \leq v_n) . \]  \hspace{1cm} (23)

It follows that \( v_1, \ldots, v_n \) are distributed as ordered \( U(0, 1) \) variables.
We have transformed the systematic subsample obtained by taking every kth value into a sample of ordered \( U(0, 1) \) variables and may therefore now apply to \( v_1, \ldots, v_n \) the methods developed in section 2 and 3 for analysing \( u(1), \ldots, u(n) \). Of course a substantial amount of computation is required in order to proceed from \( u(k), \ldots, u(nk) \) to \( v_1, \ldots, v_n \). Whether or not this is worth while depends on how much computation is saved by analysing a systematic subsample instead of the whole sample. The procedure involves some loss of power but the author has not attempted to assess how great this is likely to be. Note that the same method can be employed to construct an exact test for equality of variances; however, in view of the amount of computation required the idea does not seem worth elaborating upon in detail.

7. The elimination of nuisance parameters by the method of random substitution

A serious limitation to the practical usefulness of the Kolmogorov test has been that there has not been available a simple method of allowing for the presence of nuisance parameters. For example, it has not been possible to use it to obtain an exact test of normality in the usual situation where the population mean and variance are unknown. The reader is referred to the paper by Kac et al (1955) for a thorough discussion of the difficulties involved. A similar limitation applies to the tests proposed in section 3. Consequently, if we are to bring these tests to the point of practical usefulness for the classical goodness-of-fit problem we require a practicable method of dealing with nuisance parameters.
A simple randomisation method for doing this will now be presented. Although the method is of wide application it is convenient to introduce it by applying it to the specific problem of testing for normality.

Suppose we wish to test the hypothesis that \( x_1, \ldots, x_n \) are independent observations from a normal distribution with unknown mean \( \mu \) and unknown variance \( \sigma^2 \). Let \( \bar{x} = n^{-1} \sum x_i \) and \( s^2 = (n - 1)^{-1} \sum (x_i - \bar{x})^2 ; \) furthermore let \( \bar{x}', s'^2 \) be observations of random variables independent of \( x_1, \ldots, x_n \) and distributed as the sample mean and variance calculated from a sample of \( n \) independent observations from a \( N(0, l) \) distribution, i.e., \( \bar{x}' \) has the distribution

\[
dP = \text{constant} \times \exp\left(-\frac{1}{2} n \bar{x}'^2\right) d\bar{x}'
\]  

and \( s'^2 \) has the distribution

\[
dP = \text{constant} \times (s'^2)^{\frac{1}{2} n - 3/2} \exp\left(-\frac{1}{2} (n - 1) s'^2\right) ds'^2.
\]

Define \( x'_1, \ldots, x'_n \) by the relations

\[
\frac{x'_i - \bar{x}'}{s'} = \frac{x_i - \bar{x}}{s}, \quad (i = 1, \ldots, n)
\]  

We shall show that \( x'_1, \ldots, x'_n \) are independent \( N(0, l) \) variables. Any of the exact tests previously discussed can now be applied to \( x'_1, \ldots, x'_n \). We have, in fact, transformed the composite hypothesis concerning \( x_1, \ldots, x_n \) into a simple hypothesis concerning \( x'_1, \ldots, x'_n \).
The proof is as follows. For the distribution of \( x_1, \ldots, x_n \) we have

\[
dP = \text{constant} \times \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\right) \, dx_1 \ldots dx_n.
\] (27)

Let

\[
\ell_i = \frac{x_i - \bar{x}}{s}, \quad (i = 1, \ldots, n) \quad (28)
\]

\[
m_r = \left(\frac{r-1}{r}\right)^{1/2} (\ell_r - \frac{1}{r-1} \sum_{i=1}^{r-1} \ell_i), \quad (r = 2, \ldots, n) \quad (29)
\]

\[
m_2 = (n - 1)^{1/2} \cos a_1 \cos a_2 \ldots \cos a_{n-2}
\]

\[
m_r = (n - 1)^{1/2} \cos a_1 \cos a_2 \ldots \cos a_{n-r} \sin a_{n-r+1}, \quad (r = 3, \ldots, n-1) \quad (30)
\]

\[
m_n = (n - 1)^{1/2} \sin a_1.
\]

(29) and (30) are the Helmert and polar transformations. Substituting in (27) we have the classical representation of the probability element of a normal sample in terms of \( \bar{x}, s^2 \) and the angular variables \( a_1, \ldots, a_{n-2} \) i.e.

\[
dP = \text{constant} \times \exp\left(-\frac{n}{2\sigma^2} (\bar{x} - \mu)^2\right) n^{-3} \exp\left(-\frac{1}{2} s^2 / \sigma^2\right) \times \cos^{n-3} a_1 \cos^{n-4} a_2 \ldots \cos a_{n-3} \, d\bar{x} \, ds^2 \, da_1 \ldots da_{n-2}. \] (31)

(Geary, 1933; see also Kendall & Stuart, 1958, pp 247, 250 for details of the derivation). Consequently \( \bar{x}, s^2 \) and the set of angles \( a_1, \ldots, a_{n-2} \)
are independently distributed with distributions

\[ dP = \text{constant} \times \exp \left\{ -\frac{n}{2\sigma^2} (\bar{x} - \mu)^2 \right\} d\bar{x} \quad (32) \]

\[ dP = \text{constant} \times s^{n-3} \exp\left\{ -\frac{1}{2} \frac{s^2}{\sigma^2} \right\} ds^2 \quad (33) \]

\[ dP = \text{constant} \times \cos^{n-3} a_1 \cos^{n-4} a_2 \cdots \cos a_{n-3} da_1 \cdots da_{n-2} \quad (34) \]

respectively.

More important from the present point of view is the converse result which for the sake of clarity we state in the form of the following lemma.

**Lemma** Suppose \( \bar{x}, s^2, a_1, \ldots, a_{n-2} \) are independent random variables with distributions (32), (33) and (34). Let \( x_1, \ldots, x_n \) be determined by (28), (29), and (30) together with the relation \( \sum_{i=1}^{n} l_i = 0 \). Then \( x_1, \ldots, x_n \) are independent \( \mathcal{N}(\mu, \sigma^2) \) variables.

To prove this, suppose that \( s^2 \) and \( a_1, \ldots, a_{n-2} \) are independent with distributions (33) and (34) respectively. Let

\[ y_2 = (n - 1)^{1/2} s \cos a_1 \cos a_2 \cdots \cos a_{n-2} \]

\[ y_r = (n - 1)^{1/2} s \cos a_1 \cos a_2 \cdots \cos a_{n-r} \sin a_{n-r+1} \quad (r = 3, \ldots, n-1) \]

\[ y_n = (n - 1)^{1/2} s \sin a_1 \]
The Jacobian of the transformation is

$$\frac{\partial(y_2, \ldots, y_n)}{\partial(s^2, a_1, \ldots, a_{n-2})} = \frac{1}{2} (n-1)^{n/2} s^{n-3} \cos^{n-3} a_1 \cos^{n-4} a_2 \cdots \cos a_{n-3}$$

(Kendall & Stuart, 1958, p. 247). On substitution we find for the distribution of $y_2, \ldots, y_n$

$$dp = \text{constant} \cdot \exp \left(-\frac{1}{2\sigma^2} \sum_{i=2}^{n} y_i^2 \right) dy_2 \cdots dy_n.$$

Suppose $\bar{x}$ independently has distribution (32). Let $x_1, \ldots, x_n$ be determined by

$$\sum_{i=1}^{n} x_i = nx$$

(36)

$$\left(\frac{r-1}{r}\right)^{1/2} \left( x_r - \frac{1}{r} \sum_{i=1}^{r-1} x_i \right) = y_r \quad (r = 2, \ldots, n). \quad (37)$$

On substituting in the joint distribution of $\bar{x}$ and $y_2, \ldots, y_n$ we find that $x_1, \ldots, x_n$ have the distribution (27). It remains to show that (35), (36) and (37) are equivalent to (28), (29), (30) and the relation

$$\sum_{i=1}^{n} \ell_i = 0;$$

the equivalence follows on summing both sides of (28) and on multiplying (30) through by $s$.

Suppose now that $x_1, \ldots, x_n$ have distribution (27) and that $a_1, \ldots, a_{n-2}$ are determined by (28), (29) and (30). Let $\bar{x}'$ and $s'^{2}$ have distributions (24) and (25) independently of $x_1, \ldots, x_n$. Suppose
that $\ell_1, \ldots, \ell_n$ are determined from $a_1, \ldots, a_{n-1}$ by (29), (30) and the relation \( \sum_{i=1}^n b_i = 0 \). Let $x'_1, \ldots, x'_n$ be determined by

\[
\frac{x'_i - \bar{x}'}{s} = \ell_i \quad (i = 1, \ldots, n) .
\]  

By the above Lemma it follows that $x'_1, \ldots, x'_n$ are independent $N(0, 1)$ variables. We have therefore established the claim that composite hypotheses concerning $x_1, \ldots, x_n$ can be tested as simple hypotheses concerning $x'_1, \ldots, x'_n$.

The method is clearly of wide application. Suppose in general that $x_1, \ldots, x_n$ have a distribution with distribution function $F(x_1, \ldots, x_n, \theta)$ depending on a set of parameters $\theta$ for which a sufficient set of statistics $t_1$ is available. Suppose that a transformation $T$ independent of $\theta$ carrying $x_1, \ldots, x_n$ into $(t_1, t_2)$ exists, $t_2$ being another set of statistics distributed independently of $t_1$, such that the inverse transformation $T^{-1}$ carries $(t_1, t_2)$ into $x_1, \ldots, x_n$ uniquely. Let $G(t_1, \theta)$ denote the distribution function of $t_1$ and let $t'_1$ be an observation of a random vector independent of $x_1, \ldots, x_n$ with distribution function $G(t'_1, \theta_0)$ where $\theta_0$ is a known value of $\theta$. Let $x'_1, \ldots, x'_n$ be the values obtained by applying $T^{-1}$ to $(t'_1, t_2)$. Then $x'_1, \ldots, x'_n$ are distributed with distribution function $F(x'_1, \ldots, x'_n, \theta_0)$. As for the lemma, the proof follows by merely reversing the transformation.
To illustrate the use of this general result consider the construction of an exact distribution-free test of serial correlation. The null hypothesis is that \( x_1', \ldots, x_n' \) are independently and identically distributed with an unknown continuous distribution function. Let \( x_{i_1}', x_{i_2}', \ldots, x_{i_n}' \) be the set of \( x \)'s arranged in increasing order of magnitude; \( i_1', i_2', \ldots, i_n' \) is then a permutation of the suffixes \( 1, 2, \ldots, n \). Let \( x_{i_1}', x_{i_2}', \ldots, x_{i_n}' \) be a sample of \( n \) independent \( N(0, 1) \) variables taken from a table of random normal deviates and arranged in increasing order of magnitude. The arrangement of the suffixes is taken to be the same as on \( x_{i_1}', x_{i_2}', \ldots, x_{i_n}' \). Calculate von Neumann's statistic

\[
\frac{s^2}{\hat{s}^2} = \frac{n \sum_{t=2}^{n} (x'_t - x'_{t-1})^2}{(n-1) \sum_{t=1}^{n} (x'_t - \bar{x}')^2},
\]

where

\[
\bar{x}' = \frac{1}{n} \sum_{t=1}^{n} x'_t
\]

and refer to tables of critical values of this statistic (Hart, 1942).

The key to the construction of this test is the observation that the order statistic \( x_{i_1}', x_{i_2}', \ldots, x_{i_n}' \) is a sufficient statistic for the unknown distribution function which, when the null hypothesis is true, is distributed independently of the permutation \( i_1', \ldots, i_n' \) of the suffixes \( 1, 2, \ldots, n \) (see Fraser, 1957). Other possible applications which spring to mind are the testing of bivariate and multivariate normality and the testing of serial correlation in least-squares regression.
The price that has to be paid for the elimination of nuisance parameters by this method is that an element of randomisation is introduced in the analysis of the data. The device can be objected to on the ground that it permits different investigators to draw different conclusions from the same set of data. Some statisticians object to the use of randomisation at any stage, design or analysis, as a matter of principle; others feel that randomisation is legitimate in the design of an experiment but not in the analysis of the results. The author does not wish to take up space on this occasion discussing his personal attitude to randomisation. He does, however, wish to point out an important operational distinction between randomisation in design and randomisation in analysis. In design an act of randomisation is performed once and for all, the experimental layout being arranged according to what turns up in the randomisation procedure. Since the experimental results achieved depend on the result of randomisation, it is hard for the investigator to see how randomisation has affected his conclusions without repeating the entire experiment. And even if he were to do this, any differences which emerged could just as well be due to changes in other conditions of the experiment as to a change in the result of the randomisation procedure.

At the analysis stage, however, the position is different. The experimental results are a fixed set of numbers and the investigator can repeat the randomisation procedure as often as he likes without affecting them.
Consequently if he wishes to know the extent to which his conclusions are affected by the result of randomisation he merely has to repeat the analysis after a fresh act of randomisation. This can be done as often as the investigator wishes. Of course, in order to preserve the exact probabilities of his test he must abide by the results indicated by his first act of randomisation.

As an illustration, the first sample of 50 for each of the three distributions considered in section 4 was tested for normality assuming the population mean and variance to be unknown. The sample means and variances were replaced by values $\bar{x}'$, $s_2'$ picked at random. For $\bar{x}'$ the quantity $1/\sqrt{50}$ times a random $N(0, 1)$ deviate was used and for $s_2'$ the quantity $(1/49) \left( 2 \sum_{j=1}^{24} e_j^2 + z^2 + 48 \right)$ was used, where $e_1$, ..., $e_{24}$ are random exponential deviates with distribution (11) while $z$ is a random $N(0, 1)$ deviate; all these random deviates were taken from Quenouille's tables. The hypothesis that the values $x_1'$, ..., $x_{50}'$ determined by (38) were independent $N(0, 1)$ variables was tested by each of the four tests considered in section 4. In order to explore the effect of randomisation the tests were repeated five times using a new pair of randomly chosen values $\bar{x}'$, $s_2'$ on each occasion. The results are given in Table 2. Entries in the table are values of the appropriate test statistics.
Table 2: Some results for random-substitution tests of normality

<table>
<thead>
<tr>
<th>Population</th>
<th>Test</th>
<th>Random pair $x', s'^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Exponential</td>
<td>$\chi^2$</td>
<td>21.2*</td>
</tr>
<tr>
<td></td>
<td>$K$</td>
<td>0.160</td>
</tr>
<tr>
<td></td>
<td>$M_{25}$</td>
<td>2.44**</td>
</tr>
<tr>
<td></td>
<td>$K_m$</td>
<td>0.263**</td>
</tr>
<tr>
<td>Laplace</td>
<td>$\chi^2$</td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>$K$</td>
<td>0.163</td>
</tr>
<tr>
<td></td>
<td>$M_{25}$</td>
<td>2.42**</td>
</tr>
<tr>
<td></td>
<td>$K_m$</td>
<td>0.222**</td>
</tr>
<tr>
<td>Normal</td>
<td>$\chi^2$</td>
<td>8.4</td>
</tr>
<tr>
<td></td>
<td>$K$</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td>$M_{25}$</td>
<td>1.75*</td>
</tr>
<tr>
<td></td>
<td>$K_m$</td>
<td>0.129</td>
</tr>
</tbody>
</table>
It has to be admitted that the amount of variation from random pair to random pair is disappointing for both the $X^2$ test and the ordinary Kolmogorov test. The reason is presumably the strong dependence of the test statistics on the sample mean and variance. On the other hand the amount of variation for the modified median and modified Kolmogorov tests is substantially less, suggesting that the effect of the re-ordering of the intervals has been to reduce the dependence of the test statistics on the randomly-substituted values.
References


Pearson, K., (1933). On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. *Biometrika*, 25, 379-410.


