Volume Title: A Significance Test for Time Series and Other Ordered Observations

Volume Author/Editor: W. Allen Wallis and Geoffrey H. Moore

Volume Publisher: NBER

Volume ISBN: 0-87014-446-4

Volume URL: http://www.nber.org/books/wall41-1

Publication Date: September 1941

Chapter Title: A Significance Test for Time Series and Other Ordered Observations

Chapter Author: W. Allen Wallis, Geoffrey H. Moore

Chapter URL: http://www.nber.org/chapters/c9924

Chapter pages in book: (p. 1 - 67)

A SIGNIFICANCE TEST FOR TIME SERIES

W. ALLEN WALLIS AND GEOFFREY H. MOORE

TECHNICAL PAPER

1

Technical Papers

With the publication of 'A Significance Test for Time Series and Other Ordered Observations', the National Bureau starts a series of *Technical Papers*. As the name indicates, the series will deal with procedures or with research findings that are technical in character and that have a somewhat more specialized appeal than our *Occasional Papers*. Individual issues of the technical series will not be without interest to informed readers at large, but in the main papers in this series will be addressed to specialists in the several fields of economic research.

Contributors of $25 or more to the National Bureau receive all our publications, including the *Technical Papers*. Associates and subscribers to our *Occasional Papers* will receive notice of each issue of the technical series. If they wish any issue of the *Technical Papers*, it will be sent upon request and counted as one of the series of *Occasional Papers* due them.

Single issues of the *Technical Papers* may be purchased from the National Bureau.

NATIONAL BUREAU OF ECONOMIC RESEARCH
1819 Broadway
New York, N. Y.

A SIGNIFICANCE TEST FOR TIME SERIES

and Other Ordered Observations

BY

W. Allen Wallis AND Geoffrey H. Moore
Stanford University *Rutgers University*

TECHNICAL PAPER 1: SEPTEMBER 1941

NATIONAL BUREAU OF ECONOMIC RESEARCH

1819 Broadway, New York

Contents

Tables

Charts

Preface

This paper presents a simple and economical test of the existence in time series, or other data in which sequence of appearance is an essential characteristic, of systematic tendencies related to sequence or order. Although it has seemed desirable to express the derivation of the test in mathematical terms, we have endeavored to explain the procedures in such a way that a reader to whom the mathematical expressions as such carry little meaning will nevertheless be able to grasp their general context. Use of the test involves no higher mathematics whatever.

The test is a byproduct of studies of the cyclical behavior of production carried out at the National Bureau in 1939–40 under Research Associateships provided by the Carnegie Corporation of New York. In the cyclical behavior of the production of major crops in the United States, Great Britain, France, and Germany we found no regular relation between business cycles and the specific cycles of quantity harvested, acreage planted or harvested, and average yield per acre. The striking contrast between this and our findings for most other production series necessitated a close examination of crop cycles. One plausible hypothesis is that crop production is dominated by a complex of factors whose resultant is essentially 'random'—weather, insect depredations, plant diseases, etc. Investigation of this hypothesis required a criterion of randomness of expansion and contraction in time series. Since the criterion which resulted is adaptable to a range of time series problems much wider than the one we originally faced, it is published independently of the analysis of the cyclical behavior of agricultural production.

We are deeply indebted to Milton Friedman for invaluable counsel and assistance on numerous aspects of this paper. Rollin F. Bennett, Arthur F. Burns, Louis Guttman, John H. Smith, Abraham Wald, Jacob Wolfowitz, and Holbrook Working have made especially careful criticisms of the manuscript, and Martha Anderson, Harold Barger, J. B. D. Derksen, Pavel Egoroff, Trygve Haavelmo, Roy W. Jastram, Milton Lifshitz, Jakob Marschak, Horst Mendershausen, Frederick C. Mills, Wesley C. Mitchell, Russell T. Nichols, Paul S. Olmstead, Julius Shiskin, Frederick F. Stephan, Vladimir P. Timoshenko, Gerhard Tintner, and C. Ashley Wright have also provided helpful comments at one stage or another of its preparation.

Milton Lifshitz calculated the fourth moments of the distribution of Section VII and made most of the computations for the sampling distribution and the examples, except those connected with the least squares polynomial, which we owe to Dorothy Karger Gottfried. The sampling distribution for twelve observations was computed chiefly by John D. McLean, for whose services we are grateful to the National Youth Administration at Stanford University. The charts are the work of H. Irving Forman.

A summary of this paper was read before the Nineteenth Annual Conference of the Pacific Coast Economic Association at Stanford University on December 28, 1940,[55] and appears in the *Journal of the American Statistical Association* for September 1941.[56]

Superscript numerals refer throughout the paper to entry numbers in the List of References.

I Introduction

Analyses of time series would be greatly facilitated by simple significance tests of general applicability. Simplicity is essential if tests are to be practicable, for time series usually contain many observations, and investigations using them often involve numerous series. The standard error of estimate, to cite one example, is too expensive a statistic for many investigations of time series, and is, besides, encumbered by assumptions and restrictions that narrowly circumscribe its applicability. Generality is especially desirable, for economic data seldom justify assumptions of normality, homoscedasticity, independence, etc., nor do they provide a basis for selecting any specific alternative to these assumptions. Furthermore, it is frequently advantageous to use significance tests with such devices as moving averages or even free-hand curves, whose very nature is abhorrent to modern tests of significance that have proved so potent with data (including those of economics) free from the peculiarities of time series.

A test of significance is, of course, a test of randomness, in that it shows whether the discrepancies between a set of data (a 'sample') and expectations based on some null hypothesis can reasonably be ascribed to chance. The simple question 'Can this sample be regarded as random?' is not, however, sufficiently exact to admit of an answer. For any sample either 'yes' or 'no' is justifiable if the statistician is allowed to frame his own definitions of the ambiguous elements in the question. These elements upon whose specification the answer usually hinges are of two types. First, the form of the population the inquirer has in mind must be specified; he must ask: 'Can this be regarded as a random sample from such and such a population?' Second, the characteristics with respect to which randomness is to be judged must be specified; the question should be amended still further to: 'With respect to this or that trait, can these data reasonably be regarded as a random sample from such and such a population?'

The characteristics with respect to which randomness is tested may be simply the values of certain parameters. If a normal population is assumed, for instance, the seventeenth or any pre-designated odd moment about the mean should not differ significantly from zero; or if a Poisson population is specified, the mean, variance, and third moment should not vary significantly from one another. The characteristic may also be a frequency distribution by certain stated intervals; thus, if a uniform distribution is assumed, the frequencies should be proportional to the lengths of the intervals. Another type of characteristic, particularly relevant to time series, and the subject of this paper, is the order of appearance of the observations in sampling.

It is essential that both the form of the population and the characteristic(s) by which randomness is to be judged be chosen entirely without reference to the sample. For any sample it is possible to find some population of which it can be regarded as a random representation; and an ingenious statistician can not only find a population, but can also justify theoretically its use with the subject matter under investigation. Similarly, it is always possible to select some characteristic of a sample with respect to which it does not appear to be a random sample of a specified population.

The necessity of specifying, *independently of the data under analysis,* what characteristic of the sample will be the test criterion and what hypothesis will be tested is often overlooked. A great many spurious findings in statistical investigations, especially in the field of correlation and regression analysis, are attributable to neglect of this fundamental tenet. Unless it is adhered to rigidly, the conclusions reached are at best suggestive hypotheses, perhaps worthy of further inquiry but in no sense substantiated. It is, of course, permissible to estimate parameters from the data, provided the *form* of the population is clearly specified without reference to the particular sample. This would be subject to the same limitations as is selecting the form, were it not usually possible to make an exact mathematical allowance for the extent to which the hypothesis is in this respect simply (in Fisher's felicitous phrase) a tautological reformulation of the observations; i.e., by deducting degrees of freedom.

In practice there is usually no difficulty in selecting the characteristic by which randomness is to be tested. Indeed, the decision is usually imposed by the nature of the problem or data, by the availability of established methods and tables, by considerations of economy in calculation, by the traditions of the particular field of study, etc. (It is not intended by this statement to ignore the great advances made in recent work, initiated and principally developed by J. Neyman and E. S. Pearson,[41] on the choice of test criteria. But this work is not yet generally practicable, partly for the reasons just suggested and partly because it usually presupposes specification of the form of the population—in this regard, however, see the second paragraph of the footnote at the end of Section VIII.)

On the other hand, the difficulties of specifying the population of which the data may be regarded as a random sample are, in the social sciences at least, usually considerable and frequently insuperable. And even when it is possible to specify the form of the population it may be difficult or impossible to obtain necessary estimates of parameters. In regression analyses, for example, the usual hypothesis is that the residuals are normally distributed about a mean of zero with a variance to be estimated from the data. But when there is only one observation for each value of the independent variate (which with economic time series is virtually always) there is no satisfactory way to estimate what variance the observations would have if the independent variate were constant, since the validity of the estimate depends upon the adequacy of the fitted regression and the test of its adequacy is the variance of the residuals (i.e., the standard error of estimate).

For these reasons there has been a great deal of interest recently in tests that are independent of the form of distribution.* A test of this nature, especially relevant to certain problems of time series analysis and to other problems involving ordered observations, is set forth in this paper. It is based upon sequences in direction of movement, that is, upon sequences of like sign in the differences between successive observations.

* See references 20, 23, 25, 26, 28, 30, 31, 32, 40, 42, 45, 48, 49, 52, 53, 54.

II General Method

Each point at which the series under analysis ceases to decline and starts to rise or ceases to rise and starts to decline is noted; these 'turning points' are thus relative maxima or minima, for the first differences change sign there. A turning point is a 'peak' when it is a relative maximum and a 'trough' when it is a relative minimum. The interval between consecutive turning points is a 'phase'. (The interval between consecutive troughs or peaks might be referred to as a 'cycle'.) When a phase starts from a trough and ends at a peak it is an 'expansion'; when it starts from a peak and ends at a trough, a 'contraction'. The 'length' or 'duration' of a phase is the number of intervals (hereafter referred to as 'years', though they may represent any system of denoting sequence) between the initial and terminal turning points of the phase.*

From these definitions several deductions may be drawn. The turning points in a series of N observations may be as few as zero or as many as $N - 2$; there will be none if the direction of movement is the same throughout the series, and $N - 2$ if it alternates regularly throughout the series. If the number of turning points is even, there will be the same number of peaks as troughs but a difference of one between the number of expansions and the number of contractions; if it is odd, there will be a difference of one between the number of peaks and the number of troughs but the same number of expansions as of contractions. The shortest possible phase, occurring when two consecutive observations are both turning points, is one year. The longest possible phase, occurring when the only turning points are at the second and penultimate observations, is $N - 3$ years. The sum of the phase lengths is the number of years between the first and last turning point; since neither the first nor last observation can be a turning point, it cannot exceed $N - 3$.

* The definition of a phase excludes the movement preceding the initial turning point and that following the final turning point in the series. This exclusion conforms with the definitions used in the National Bureau's technique of measuring cyclical behavior;[37] but for the present purpose it would be preferable, especially in short series, to include them and record two additional phases (albeit from a slightly different population). The advantage of their inclusion was not appreciated until most of our computations were complete, and then it did not seem sufficiently important to justify the extensive recalculations that would be required. The incomplete phases are not entirely ignored by the test of significance developed in this paper, for their duration affects the number and durations of the complete phases.

With these definitions and their corollaries in mind, the expected frequency distribution of phase lengths in a series of N observations drawn at random from a stable population can be calculated. It is apparent that, as such a series is being drawn, the greater the number of consecutive rises the less is the probability of an additional rise; for the higher the observation the smaller is the chance of drawing one that exceeds it. Perhaps surprisingly, the rapidity with which the basic distribution tapers off from its mode does not affect the expected frequency distribution of phase durations; in fact, it is shown below that this expected frequency distribution of phase durations is practically independent of the probability distribution of the original data.

The only restriction on the original probability distribution is that it be such (or else that the method of sampling be such) that the probability of two consecutive observations being identical is infinitesimal. This condition is fulfilled by all distributions for which the cumulative probability (i.e., the ogive) increases continuously; all continuous distributions, therefore, and hence virtually all metric data, meet the restriction.

Without specifying anything further about the form of the basic distribution, we may make a mathematical transformation of it that leads to a known distribution but leaves unaltered the pattern of rises and falls of the original observations. That is, if x represents the original variate we replace it by a new variate z, which is a mathematical function of x; the function we choose is such that z' is greater than, equal to, or less than z'' according as x' is greater than, equal to, or less than x'', where z' and z'' are the transformed values of two observations x' and x''.

A familiar transformation of this type is the rank transformation. If each observation is replaced by its rank according to magnitude within the entire series, the new variate has a simple and definite distribution; that is, z may be any integer from 1 to N (N being the number of observations in the sample) and the probability of each value is $1/N$. The ranks have exactly the same pattern of rises and falls as the original observations. The distribution of phase durations expected in a random arrangement of the digits 1 to N is, therefore, that to be expected in a sample of N from any population; that is to say, it is completely independent of assumptions about the original distribution, hence comparable with the observed distribution of phase durations in any set of data.

Another familiar example of such a transformation, one more easily handled analytically, is the probability transformation. Without knowing the original probability distribution, we may imagine each value of x to be replaced by its cumulative probability, i.e.,

$$z_i = \int_{-\infty}^{x_i} f(x)\,dx.$$

While this replacement cannot actually be performed, since we cannot know the cumulative probability when we do not know the basic distribution, it is obvious that whatever the original distribution, $f(x)$, may have been (provided, of course, that it meets the continuity condition indicated above), the distribution of z will be uniform—a straight line of unit height and length over the interval $0 \leq z \leq 1$. Indeed, this simply amounts to the tautology that any observation with a given probability is exactly as probable as any other observation with the same probability, for the probabilities are defined by what is in effect the condition of uniformity in their distribution. Whenever the original variate, x, increases, the transformed variate, z, increases also, and similarly for decreases; so the pattern of phase durations is precisely the same in the transformed values as in the original observations. We can, therefore, tabulate the actual distribution of phase lengths from the original observations and calculate the expected distribution of phase lengths for the transformed variate; each operation can be carried out without knowing the fundamental probability law and the results will be comparable in the strictest sense.

A completely general determination of the expected distribution of phase lengths in a random series can, then, be obtained by working with a uniform distribution of unit height and length.

III Derivation of Distribution

A preliminary step that illustrates the method of solution is to calculate the expected number of phases in a series of N observations. Three observations are required to define a turning point. The probability of any particular ordered set of three observations, z_1-z_2-z_3, is $dz_1 \cdot dz_2 \cdot dz_3$. The sum of the probabilities of all possible sets is the triple integral of this product over the entire possible range of z,

$$\int_0^1 \int_0^1 \int_0^1 dz_1\, dz_2\, dz_3 = 1.$$

It may be helpful to visualize this geometrically: z_1 may be plotted along the axis of abscissae, z_2 along the axis of ordinates, and z_3 along an axis perpendicular to the z_1-z_2 plane at its origin. Then, since each variate is uniformly distributed from zero to one, a cube of unit edge represents all possible sets of three observations, and the points within the cube are equally likely. To ask what the probability is that a set of three observations constitute a turning point is to ask in what portion of the volume of this cube z_2 is either greater than or less than both z_1 and z_3.

If the turning point is a trough, z_2 is less than z_1; hence only that part of the cube is acceptable which lies above the half of the z_1-z_2 base in which z_1 exceeds z_2; i.e.. above the triangle formed by the points (0, 0, 0), (1, 0, 0), and (1, 1, 0). And since z_2 must also be less than z_3, only that portion of the cube is acceptable which also lies within a projection of the triangle on the z_2-z_3 face defined by the points (0, 0, 0), (0, 1, 1), and (0, 0, 1). The square pyramid that has the z_1-z_3 face as its base and the point (1, 1, 1) as its apex thus represents all sequences z_1-z_2-z_3 that produce a trough; the volume of this portion of the cube is given by

$$\int_{z_3=0}^{1} \int_{z_2=0}^{z_3} \int_{z_1=z_2}^{1} dz_1\, dz_2\, dz_3,$$

which may be evaluated as follows:

$$\int_{z_2}^{1} dz_1 = 1 - z_2$$

$$\int_0^{z_3} (1 - z_2)\, dz_2 = z_3 - \frac{z_3^2}{2}$$

$$\int_0^1 \left(z_3 - \frac{z_3^2}{2}\right) dz_3 = \frac{1}{3}.$$

From the symmetry of the function it is apparent that the probability of a peak is also 1/3. The probability that any particular set of three observations constitute a turning point is, therefore, 2/3. Since in a series of N there are $N - 2$ sets of three consecutive items, the expected number of turning points is

$$\frac{2(N - 2)}{3}.$$

Since a phase is the uninterrupted expansion or contraction between two turning points, there are one fewer phases than turning points, and

$$\frac{2N - 7}{3}$$

is the expected number of phases. This is not absolutely accurate. since there may be no turning points but cannot be a negative number of phases; when there are no turning points, the number of phases is zero, or equal to the number of turning points. The expression obtained by subtracting one from the expected number of turning points is therefore too small by one times the probability that there will be no turning points. This occurs only if each item exceeds or is exceeded by its predecessor, the probability of which is

$$2\int_0^1\int_0^{z_N}\int_0^{z_{N-1}}\cdots\int_0^{z_2} dz_1\,dz_2\cdots dz_N = \frac{2}{N!}, \tag{1}$$

i.e., twice the reciprocal of the number of permutations of N different things. This amount should, therefore, be added to the expression for the expected number of phases. It is, however, so minute—less than 0.0000006 when N is only 10—and declines so rapidly as N increases, as to be utterly negligible.

An expansion of exactly one year is defined by four consecutive observations in which neither the first nor the third is as small as the second, and neither the second nor the fourth is as great as the third. The probability of any particular ordered set of four, z_1-z_2-z_3-z_4, is $dz_1 \cdot dz_2 \cdot dz_3 \cdot dz_4$; and its quadruple integral over all possible values is again 1. The probability that the four define a one-year expansion is the same integral over that portion of the four-dimensional hyper-cube in which z_1 is between z_2 and 1, z_2 between 0 and z_3 , z_3 between z_4 and 1, and z_4 between 0 and 1, i.e.,

$$\int_0^1\int_{z_4}^1\int_0^{z_3}\int_{z_2}^1 dz_1\,dz_2\,dz_3\,dz_4 = \frac{5}{24}.$$

Since the probability of a one-year contraction is exactly the same, 'the probability of a one-year phase is 5/12. And since there are $N - 3$ sets of four consecutive items in a sample of N,

$$\frac{5(N - 3)}{12}$$

is the expected number of phases of exactly one year's duration.

Similarly, the probability that a set of five consecutive observations define a phase of exactly two years' duration is

$$2 \int_0^1 \int_{z_5}^1 \int_0^{z_4} \int_0^{z_3} \int_{z_2}^1 dz_1\, dz_2\, dz_3\, dz_4\, dz_5 = \frac{11}{60}.$$

Since there are $N - 4$ sets of 5 consecutive items in a series of N, the expected number of two-year phases is

$$\frac{11(N - 4)}{60}.$$

Proceeding in the same way, the expected number of phases of exactly three years is

$$\frac{19(N - 5)}{360},$$

of exactly four years,

$$\frac{29(N - 6)}{2520},$$

of exactly five years,

$$\frac{41(N - 7)}{20160},$$

and of exactly six years,

$$\frac{55(N - 8)}{181440}.$$

These results are summarized in Table 1.

The general expression for the expected number of phases of exactly d years' duration is

$$\frac{2(d^2 + 3d + 1)(N - d - 2)}{(d + 3)!}. \tag{2}$$

This may be derived inductively. It may also be obtained by starting with the probability of a sequence of d rises or declines, the extremes of which may or may not be turning points. The probability of such a sequence, which represents a phase of d or more years, is shown by formula 1 to be

$$\frac{2}{(d + 1)!}.$$

By substituting $d + 1$ for d and differencing, then placing $d + 1$ for d in the result and differencing again, and finally multiplying by $N - d - 2$, expression 2 is obtained in the form

$$2(N - d - 2)\left[\left(\frac{1}{(d + 1)!} - \frac{1}{(d + 2)!}\right) - \left(\frac{1}{(d + 2)!} - \frac{1}{(d + 3)!}\right)\right].$$

Substituting $d + 1$ for d and differencing amounts to closing one end of the sequence by subtracting the probability of a rise or decline of $d + 1$ or more from the probability of a rise or decline of d or more; the process is performed twice to close both ends and thereby define

TABLE I

Expected Distribution of Phase Durations in Random Series

DURATION	FREQUENCY	PROBABILITY	PROBABILITY $N = 10$	$N = 20$	$N = 40$	$N = 70$	LIMITING PROBABILITY ($N = \infty$)
1	$\frac{5(N-3)}{12}$	$\frac{5(N-3)}{4(2N-7)}$	.6731	.6439	.6336	.6297	.6250
2	$\frac{11(N-4)}{60}$	$\frac{11(N-4)}{20(2N-7)}$	.2538	.2667	.2712	.2729	.2750
3	$\frac{19(N-5)}{360}$	$\frac{19(N-5)}{120(2N-7)}$	.0609	.0720	.0759	.0774	.0792
4	$\frac{29(N-6)}{2520}$	$\frac{29(N-6)}{840(2N-7)}$	.0106	.0146	.0161	.0166	.0173
5	$\frac{41(N-7)}{20160}$	$\frac{41(N-7)}{6720(2N-7)}$	.0014	.0024	.0028	.0029	.0031
6	$\frac{55(N-8)}{181440}$	$\frac{55(N-8)}{60480(2N-7)}$	.0001	.0003	.0004	.0004	.0005
...	...	...	...	...	...	...	...
d	$\frac{2(d^2+3d+1)(N-d-2)}{(d+3)!}$	$\frac{6(d^2+3d+1)(N-d-2)}{(d+3)!(2N-7)}$	...	...	...	...	$\frac{3(d^2+3d+1)}{(d+3)!}$
...	...	...	...	...	...	...	...
Total Frequency or Probability	$\frac{2N-7}{3}$	1	1	1	1	1	1
Mean Duration	$\frac{3(N+7-4e)}{2N-7}$	$\frac{3(N+7-4e)}{2N-7}$	1.4139	1.4661	1.4847	1.4916	1.500
Variance of Duration	$\frac{(24e-63)N^2+(12e-51)N-3(48e^2-140e+14)}{(2N-7)^2}$		0.4481	0.5175	0.5409	0.5494	0.5597

a phase. $N - d - 2$ is the number of sets of $d + 3$ consecutive items in a series of N, $d + 3$ being the number of points required to define a phase of d.

Division of each expectation by the expected total number of phases,

$$\frac{2N - 7}{3},$$

gives the probability that a phase selected at random will have a specified duration as

$$\frac{6(d^2 + 3d + 1)(N - d - 2)}{(d + 3)!(2N - 7)}. \tag{3}$$

It may not be apparent that this division is legitimate. Among the $N!$ equally likely possible arrangements of the digits 1 to N (referring now to a rank transformation instead of to a probability transformation) there are

$$N!\left(\frac{2N - 7}{3} + \frac{2}{N!}\right)$$

phases. Of these,

$$\frac{2N!(N - d - 2)(d^2 + 3d + 1)}{(d + 3)!}$$

are of duration d. The proportion of phases that are of duration d may therefore be obtained by dividing the second quantity by the first; and, neglecting the term

$$\frac{2}{N!}$$

this ratio is expression 3. Table 1 shows the values of expression 3 for d from 1 to 6, the numerical values of the probabilities for N = 10, 20, 40. and 70, and the limits approached by these probabilities as the sample size increases. The probabilities approach their limiting values rapidly; when N is as great as, say, 40, there is little advantage in allowing for the length of the series.

As noted in Section II, the same expectations can be calculated with a rank transformation. This brings out the fact that they are exactly those obtained by considering all possible arrangements of the actual observations in a particular sample, provided no two observations in the sample are equal. For example, the probability that three different observations define a turning point may be found by considering all permutations of three observations. Of the six permutations two produce a peak, two a trough, and two produce no turning point; hence the probability of a turning point is 2/3. Similarly, formula 1 is immediately apparent from the fact that two of the $N!$ permutations of N different numbers produce an uninterrupted movement; and from it the derivation of formula 2 proceeds as before.

Had the definition of a phase included the incomplete phase preceding the first turning point and that following the last turning point (see the footnote to the first paragraph of Section II), formula 2 would have been increased by

$$\frac{4(d + 1)}{(d + 2)!}$$

when $d < N - 1$ and by

$$\frac{2}{N!}$$

when $d = N - 1$. The expected total number of phases would have been

$$\frac{2N - 1}{3}$$

and formula 3 would have had a 1 in the denominator in place of the 7 and would have been greater by

$$\frac{12(d + 1)}{(d + 2)!(2N - 1)}.$$

IV Mean and Variance of Distribution

The mean, or expected, duration of a phase is derived by multiplying each duration by its probability and summing the products:

$$\text{Expected Duration} = \sum_{d=1}^{N-3} \frac{6d(d^2 + 3d + 1)(N - d - 2)}{(d + 3)!\,(2N - 7)} = \frac{3(N + 7 - 4e)}{2N - 7}$$

(where e is the natural logarithmic base, 2.7182818285). This summation is evaluated as follows: First place the factor

$$\frac{6}{2N - 7},$$

which is common to all terms of the sum, outside the summation sign and break the numerator of the remaining fraction into two terms: $(N - 2)(d^3 + 3d^2 + d) - (d^4 + 3d^3 + d^2)$. The first contains a factor $N - 2$ which may be placed outside the summation sign; the remaining factor is then written as $(d + 3)(d + 2)(d + 1) - 3(d + 3)(d + 2) + 5(d + 3) - 3$, to which it is equal. Upon dividing this by the denominator we have

$$\sum_{d=1}^{N-3} \frac{1}{d!} - 3\sum_{1}^{N-3} \frac{1}{(d + 1)!} + 5\sum_{1}^{N-3} \frac{1}{(d + 2)!} - 3\sum_{1}^{N-3} \frac{1}{(d + 3)!}.$$

These sums are easily evaluated from the fact that

$$\sum_{x=0}^{r} \frac{1}{x!}$$

approaches e so rapidly that unless r is very small the difference is negligible; hence they are essentially $e - 1$, $e - 2$, $e - 2\frac{1}{2}$, and $e - 2\frac{2}{3}$. The second term into which the numerator was broken is written as $(d+3)(d+2)(d+1)d - 3(d + 3)(d + 2)(d + 1) + 8(d + 3)(d + 2) - 13(d + 3) + 9$ and when this is divided by the denominator, it yields sums similar to those above.

The precise expression for the expected duration, allowing not only for the approximation involved in summing to infinity but also for that referred to in expression 1, is

$$\frac{3\left(N + 7 - 4\sum_{x=0}^{N} \frac{1}{x!}\right)}{2N - 7 + \frac{6}{N!}} + \frac{2(N - 1)}{N!\,(2N - 7) + 6},$$

which is exact for any value of N. When $N = 6$ it agrees with the simpler form to two decimal places, when $N = 7$ there is a difference of one in the third place, and when $N = 8$, there is a difference of one in the fourth place.

The variance of the distribution is simply the sum of the products of the probabilities by d^2, minus the square of the expected duration, and may be found by a similar process of summation to be essentially

$$\frac{3[(8e - 21)N^2 + (4e - 17)N - (48e^2 - 140e + 14)]}{(2N - 7)^2}$$
$$= \frac{2.238764N^2 - 18.380618N + 35.654290}{(2N - 7)^2}.$$

Calculation of the variance requires, after common factors are removed from the summation, evaluation of

$$\sum_{d=1}^{N-3} \frac{d^2(d^2 + 3d + 1)(N - d - 2)}{(d + 3)!}.$$

The numerator breaks into two terms: $(N - 2)(d^4 + 3d^3 + d^2) - (d^5 + 3d^4 + d^3)$. The first involves only a summation already evaluated in connection with the mean duration; the second may be written $(d + 3)(d + 2)(d + 1)\,d(d - 1) - 2(d + 3)(d + 2)\,(d + 1)d + 8(d + 3)(d + 2)(d + 1) - 21(d + 3)(d + 2) + 35(d + 3) - 27$, which (since the $(-1)!$ may here be regarded as infinite) reduces to summations of the form evaluated for the mean.

The variance has been evaluated for various values of N and entered in Table 1. As N increases, it rises toward a limiting value of 0.55969097. Because the phases in a single sample are not independent of one another, as is pointed out in Section VI, and because the number of phases is itself a stochastic variate related to the distribution of phase durations, the variance cannot be used in the usual way as a test of the observed mean duration (see Section VII).

V Empirical Verification

The foregoing mathematical deductions were checked by three empirical tests. The first involved 200 random series of 25 items each, the second, 300 random series of 50 items each, and the third, 200 random series of 75 items each. Each series was copied from a deck of N playing cards bearing the integers 1 to N, shuffled ten or a dozen times by the 'fan' method. Not all 700 series are completely independent, although all series of a given length are. One hundred of the series of fifty were gotten by omitting the integers above 50 in the first 100 series of 75. All 200 of the series of 25 were taken from the first 67 series of 75 by treating the integers 1–25, 26–50, and 51–75 as three independent series.

Deriving some series from longer ones does not involve as much duplication as it may seem to at first glance. The longer series are in no sense simple sums of their component series, for the manner in which the components are intermingled is an important characteristic of the full series. This becomes clear when we consider the problem of combining three independent series of 25 into a single series of 75. It would be necessary to determine by chance for each of the 75 positions which of the three series should fill it, this determination being such that each series would necessarily be selected exactly 25 times out of 75. A possible procedure would be to place in a bowl 25 chips of each of three colors, and draw these (without replacement) to determine which series should fill each position. While the frequency distribution of phase durations in the final series is not entirely unrelated to the distributions for the component series, the redundancy introduced by this economizing device was not deemed sufficient to offset the advantages of the increased number of series.

The turning points in the 700 series were marked and the lengths of the intervening phases tabulated. Table 2 gives the observed frequency distributions, the theoretical distributions, and the values of χ^2 for goodness of fit with the corresponding probabilities. Since in computing χ^2 the expected frequencies are adjusted to the observed in only one respect, the value of N, only one degree of freedom is lost. Had the expectations been calculated from column 3 of Table 1 instead of column 2, the total frequencies also would have been equal, and two degrees of freedom would have been lost; but the value of χ^2 would also have been smaller.

TABLE 2

Frequency Distributions of Phase Durations in 200 Random Series of 25 Items, 300 of 50 Items, and 200 of 75 Items

DURATION OF PHASE	FREQUENCY					
	Expected $N = 25$	Observed $N = 25$	Expected $N = 50$	Observed $N = 50$	Expected $N = 75$	Observed $N = 75$
1	1833.3333	1850	5875.0000	5895	6000.0000	6085
2	770.0000	776	2530.0000	2579	2603.3333	2624
3	211.1111	199	712.5000	663	738.8889	757
4	43.7302	49	151.9048	146	158.8095	118
5	8.4921	6	26.2351	35	27.6587	20
over 5			4.3601	2	4.6429	7
Total	2866.6667	2880	9300.0000	9320	9533.3333	9611
χ^2	2.2594		8.8914		15.6165	
n	4		5		5	
$P(\chi^2)$	.68		.11		.008	
$P(p)$[a]	.65		.70		.13	

[a] Probability for total number of phases; see Sec. VII.

Although the fit is adequate for the series of 25 and of 50, it is definitely bad for the series of 75, entirely because of a great deficiency of four-year phases—two-thirds of the value of χ^2 is contributed by this one class. The presumption that this result is fortuitous was confirmed by an additional 100 series of 75 integers gotten by using in order all two digit entries in Fisher and Yates' Table of Random Numbers,[24] the digits to the right being regarded as decimal places when two consecutive numbers were equal. A count of the four-year phases showed 80, surprisingly close to the expectation of 79.4.

For the significance test developed in this paper it is immaterial whether the expected frequencies are correct for phases longer than two years, provided the expected total number of such phases is accurate; for in applying the test to a single series, the expected frequencies of phases longer than two years are usually so small that it is necessary to combine all into a single group in order to meet the requirement that expected freuencies used in χ^2 be not too small. Most authorities state that the expected frequencies must be at least five and preferably ten, though recent investigations of the effect of small theoretical frequencies indicate that "except perhaps in the case when m [the theoretical frequency] $= 1$, the theoretical distribution of χ^2 is sufficiently closely realized".[27, 47]

The three-year phases cannot be made a separate group unless the expected number of phases of four or more years is sufficient to stand alone. The expression for this expectation is

$$\frac{5N - 31}{360},$$

which is almost exactly one when $N = 78$, two when $N = 150$, five when $N = 366$, and ten when $N = 726$. Since the expected number of phases in excess of two years is not large—it is

$$\frac{4N - 21}{60},$$

which is approximately one when $N = 20$, two when $N = 35$, five when $N = 80$, and ten when $N = 155$—not a great deal of information is lost by combining all into a single group. The test does not neglect entirely the lengths of phases in the last class, because these lengths influence the total number of phases and therefore the frequencies in all three classes. A test of the mean duration (see Section VII) salvages some of the information lost by grouping.

For each of the three empirical tests, therefore, χ^2 was recomputed by combining all durations except the first two into a single group. The resulting values of χ^2 are 0.5291, 3.6998, and 2.2112, which, being based on two degrees of freedom each, indicate probabilities of .77, .16, and .33, respectively.

VI The Test of Significance

After the frequency distribution of phase durations has been obtained from a time series, it may be compared with the expected distribution by the usual procedure for testing goodness of fit; that is, by squaring the differences between the observed and the expected frequencies, dividing by the expected frequencies, and summing the ratios.[22] This sum is essentially similar to χ^2, but since its sampling distribution is not quite that ordinarily associated with χ^2 it is advisable to distinguish it by the subscript p (denoting 'phase').

The reason χ^2_p is not distributed as χ^2 is that the phases within a single sample series drawn at random from a fixed population are not entirely independent of one another. When one long phase occurs, another long one is more probable than it would otherwise have been. A long rise, for example, tends to carry the series to unusually high values, thereby increasing the probability that the decline will be long. Short phases, on the other hand, tend to leave the series at central values, from which short phases are likely. (This positive serial correlation is offset to some extent by an inverse relation caused by the fact that a long phase reduces the number of observations available for other phases, which reduces not only the number of the other phases but also the relative frequency among them of long phases; except for very long phases or very short series, however, this counteracting effect is small.) Since the resultant positive correlation within samples makes very large and very small values of χ^2_p a little more likely than if the phase lengths were independent, it is to be expected that, except perhaps in short series, the variance of χ^2_p will somewhat exceed that of χ^2. In addition, since χ^2_p is virtually always based on two degrees of freedom, for which the χ^2 distribution is exceedingly skewed, this increased variance may be expected to raise the mean value of χ^2_p above that of χ^2. For two degrees of freedom, χ^2 has a mean of two and a variance of four.

In the preceding Section the χ^2 test was applied in disregard of the interdependence of phases within a series. In that Section, since phases from many series were thrown into a single frequency distribution, the independence of phases from different series tended to submerge the interdependence of phases from the same series. For a given number of series the importance of the interdependence increases with the series length, and this may have something to do with the series for $N = 25$ appearing to fit better than those for $N = 75$.

The problem of the exact sampling distribution of χ_p^2 for various values of N is not unlike (and, apparently, not simpler than) the similar problem for the rank correlation coefficient, for which no general solution has yet been found. Both Olds[42] and Kendall, Kendall and Smith[31] have devised what are essentially systematic methods of building up the distribution of the rank correlation coefficient for any value of N from the distribution for $N - 1$. They have also provided excellent approximations to bridge the gap between the point at which the patience necessary to evaluate the exact distributions is exhausted and that at which the limiting (normal) distribution becomes applicable. But no precise formula giving the probability as a function of the coefficient and the sample size has been discovered. Similarly, in tabulating the distribution of the rank correlation ratio Friedman[25] devised a method of building one exact distribution from another—later explained in detail by Kendall and Smith,[32] who added an approximating function to smooth the transition between exact and limiting distributions; but there is no general analytic expression for the probability. We have not been able to determine mathematically the sampling distribution of χ_p^2, but have found what seems a satisfactory working solution.

In the first place, we discovered a recursion formula for calculating the relative frequencies of the 2^{N-1} different arrangements of signs of first differences that occur in the $N!$ permutations of N different numbers. This formula states the number of permutations of N different numbers that produce the sequence of signs of differences shown in the r-th row of a matrix having 2^{N-1} rows and $N - 1$ columns formed as follows: In the first column fill in alternately plus and minus, starting with plus at the top of the column. In the second column enter two pluses, then two minuses, and alternate groups of two. In the third column enter four pluses, then four minuses, etc. In general, the j-th column starts with pluses in the first 2^{j-1} rows, then has minuses in the next 2^{j-1}, then another 2^{j-1} pluses, etc., alternating groups of 2^{j-1} to the bottom of the column. The last column has simply 2^{N-2} pluses followed by 2^{N-2} minuses. Then, denoting by $F_N(r)$ the number of permutations of N different integers that produce the sequences of signs given in the r-th row of this matrix,

$$F_N(r) = F_{i+1}(j)\binom{N}{i+1} - F_N(j) \tag{4}$$

where $r = 2^i + j$ $(0 \leq i \leq N - 2, 1 \leq j \leq 2^i)$, i.e., i is the largest power to which 2 can be raised without equaling or exceeding r, j is the difference between r and 2^i;

$$\binom{N}{i+1}$$

denotes the number of combinations of N things $i + 1$ at a time, i.e.,

$$\frac{N!}{(i+1)!(N-i-1)!};$$

and $F_N(1) = 1$ for all N. Division of $F_N(r)$ by $N!$ converts it from an absolute frequency to a probability.

The durations of the phases represented by the r-th row in the above rectangle may also be determined from the value of r: The length of the preliminary incomplete phase is the largest value of p_0 such that $r \equiv 0$ or 1, modulus 2^{p_0}; i.e., the largest value of p_0 such that r divided by 2^{p_0} leaves a remainder of 0 or 1. To find the length of the first complete phase write $r = k_1 \cdot 2^{p_0} + t_1$; i.e., let k_1 by the largest integer by which 2^{p_0} can be multiplied without exceeding r, and let t_1 by the remainder of 0 or 1. Take $r' = k_1 + t_1$ and determine the largest value of p_1 such that $r' \equiv 0$ or 1, mod 2^{p_1}; p_1 is then the length of the first complete phase. Similarly, the length of the second complete phase is found by starting with an r'' that has the same relation to r', through a k_2 and a t_2, that r' has to r through k_1 and t_1. The lengths of successive phases are found by repeating the process until eventually the congruence $1 \equiv 0$ or 1, mod 2^{p_m}, appears, whence $p_m = \infty$. This is to be replaced by

$$N - 1 - \sum_{i=0}^{m-1} p_i$$

as the length of the final incomplete phase.

This recursion expression enabled us to calculate the exact distribution of χ_p^2 for small values of N. The calculation requires not only a great many evaluations of the formula, but also computation of the corresponding values of χ_p^2 and subsequent cumulation of frequencies. Despite considerable shortcuts that can be introduced in actual calculation with the formulas, the procedure is laborious and has been carried only as far as $N = 12$. Table 3 gives the exact probability, P, of obtaining a χ_p^2 as large as or larger than each possible value, and also the mean and variance of χ_p^2, for $N = 6$ to 12, inclusive. These distributions are also shown in Chart 1, the probability scale being logarithmic. The χ^2 distribution for two degrees of freedom is a straight line on semi-logarithmic coordinates, since $\chi^2 = -2 \log P$ when $n = 2$ (denoting by n the number of degrees of freedom).

Because of the discontinuity of χ_p^2, the probability corresponding with a given χ_p^2 might properly be plotted at any abscissa from that value of χ_p^2 to the next lower value, and the value of χ_p^2 for a given P at any probability from that value to the next higher. For simplicity, the midpoints of both these intervals are plotted: at each

TABLE 3

Exact Distributions of χ_p^2, Six to Twelve Observations*

$N = 6$ χ_p^2	P	$N = 7$ χ_p^2	P	$N = 8$ χ_p^2	P	$N = 9$ χ_p^2	P	$N = 10$ χ_p^2	P	$N = 11$ χ_p^2	P	$N = 12$ χ_p^2	P
.4667	1.0000	.5515	1.0000	.2837	1.0000	.3576	1.0000	.3281	1.0000	.4792	1.0000	.6152	1.0000
.8667	.8694	.7333	.7893	.6837	.8432	1.1576	.7978	.6139	.9410	.5792	.9800	.6606	.9843
1.1939	.6750	.7515	.7028	.8437	.6654	1.2667	.6308	.7281	.9169	.8169	.9335	.7485	.8962
1.6667	.4528	.9333	.5361	.9200	.5897	1.6303	.6046	1.0554	.8129	.9169	.8436	.7939	.8907
2.3939	.3667	1.7333	.4933	1.3200	.5605	2.0667	.4893	1.3411	.6927	.9792	.7300	.8374	.8500
2.8667	.2222	2.1515	.3702	1.4800	.5062	2.4303	.4519	1.4190	.6059	1.0879	.7232	.9707	.7862
19.6667	.0528	2.3333	.3024	2.3637	.4946	2.7576	.3811	1.5853	.6010	1.2792	.6554	1.0152	.7204
		3.9333	.2774	2.6800	.4708	3.1576	.3743	1.7048	.5937	1.3169	.5763	1.0606	.6851
		5.6061	.1694	2.9347	.3924	3.2667	.3208	1.7717	.5917	1.5879	.5368	1.4152	.5846
		7.5045	.1171	3.0000	.2994	3.6667	.2150	1.8138	.5262	1.7000	.4728	1.4606	.5830
		8.9045	.0552	4.3746	.2931	4.0303	.1637	1.8190	.4186	1.8000	.4716	1.6374	.5695
				4.4546	.2346	4.0667	.1445	2.3126	.4070	2.0792	.4684	1.6828	.5332
				4.9346	.1936	4.7576	.1102	2.5769	.3745	2.2000	.4666	1.9333	.4870
				5.0000	.1332	5.6667	.0783	2.6762	.3274	2.3087	.4662	1.9485	.4863
				5.8193	.0645	6.0667	.0635	2.7431	.3269	2.4087	.4399	2.0667	.4277
				6.4546	.0332	7.4848	.0197	2.8626	.2739	2.4169	.4028	2.1556	.4275
						15.6667	.0049	2.9048	.2420	2.5000	.3916	2.2030	.4069
								2.9769	.2200	2.5792	.3842	2.2889	.3442
								3.2424	.1811	2.6879	.3044	2.3333	.3334
								3.8341	.1794	2.8087	.2742	2.5556	.3313
								3.9697	.1650	3.0256	.2608	2.6152	.3031
								4.3333	.1583	3.1087	.2302	2.6606	.3028
								4.4003	.1580	3.2130	.2006	2.7333	.2996
								4.6762	.1391	3.3000	.1472	2.8374	.2995
								4.8580	.1074	3.7792	.1471	2.8697	.2867
								5.1276	.0724	3.8000	.1468	2.8828	.2462
								5.4912	.0590	3.9087	.1327	2.9556	.2158
								6.5152	.0536	4.1169	.1278	3.2667	.2112
								7.1333	.0420	4.3130	.1259	3.4152	.2065
								11.3076	.0141	4.3879	.0991	3.4889	.1486
								12.9648	.0062	4.7256	.0909	3.9333	.1266
										5.0000	.0772	4.0697	.1266
										5.6087	.0772	4.1556	.1143
										5.7000	.0755	4.3485	.1128
										6.0130	.0552	4.3939	.1128
										8.2000	.0498	4.5707	.1123
										8.6348	.0320	4.6162	.1094
										9.4675	.0225	4.7333	.1013
										9.7348	.0183	5.6667	.0924
										10.2140	.0086	5.8030	.0924
										11.4348	.0045	5.8889	.0905
												6.0253	.0900
												6.7333	.0849
												6.8424	.0720
												6.9556	.0597
												7.5040	.0499
												7.6222	.0408
												8.5758	.0293
												8.8222	.0260
												9.2374	.0193
												9.2667	.0141
												10.5556	.0028
												19.6667	.0004

	$N = 6$	$N = 7$	$N = 8$	$N = 9$	$N = 10$	$N = 11$	$N = 12$
Mean	2.5078	2.4364	2.3802	2.3629	2.3544	2.3497	2.3478
Variance	17.0757	6.3017	3.7465	3.7080	3.8358	3.9657	4.4135

* Calculated from three frequency classes by combining all phases of more than two years into a single class and disregarding the correction of formula 1, which would affect slightly the frequency expected for this class.

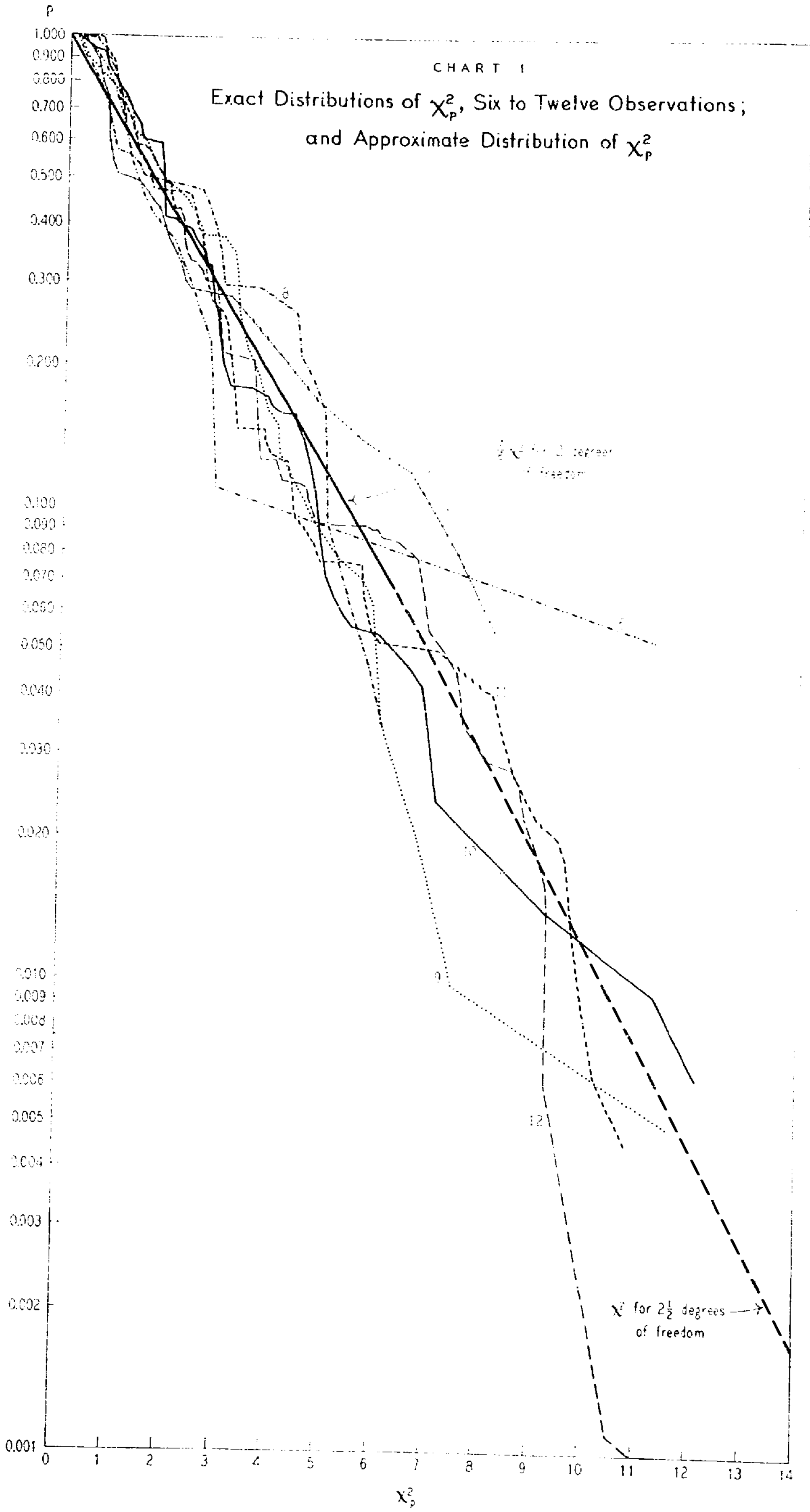


CHART 1
Exact Distributions of χ_P^2, Six to Twelve Observations; and Approximate Distribution of χ_P^2
P
χ^2 for $2\frac{1}{2}$ degrees of freedom
χ_p^2

TABLE 4

Distributions of χ_p^2 from 200 Random Series of 25 Items, 300 of 50 Items, and 200 of 75 Items; and Test of Homogeneity

χ_p^2	$N = 25$	$N = 50$	$N = 75$	Total	Contribution to χ^2
.025					
	7	15	7	29	0.9310
.115					
	10	7	10	27	3.1605
.215					
	6	10	15	31	5.9946
.395					
	15	8	10	33	5.9950
.455					
	16	12	4	32	8.2500
.585					
	8	24	4	36	9.1111
.635					
	2	19	9	30	7.9914
.780					
	10	11	7	28	0.7083
.955					
	4	17	6	27	4.7160
1.145					
	4	11	15	30	7.5278
1.310					
	8	14	8	30	0.1778
1.590					
	8	11	12	31	1.5914
1.845					
	13	11	5	29	4.1494
2.035					
	3	24	4	31	15.1774
2.215					
	6	8	13	27	5.1049
2.440					
	22	1	6	29	33.8391
2.595					
	4	11	10	25	2.5333
2.835					
	9	16	3	28	4.5833
3.195					
	9	11	7	27	0.3086
3.605					
	9	11	7	27	0.3086
4.195					
	8	8	12	28	3.3333
4.755					
	4	13	11	28	3.2083
5.895					
	8	14	6	28	0.8333
7.510					
	7	13	9	29	0.2874
13.975					
Total	200	300	200	700	$\chi^2 = 129.8252$

Since $n = 46$, $P(\chi^2)$ is less than .0000000001. Using unit class intervals, $\chi^2 = 18.604$ for $n = 18$, and $P(\chi^2) = .41$.

value of χ^2_p there is a point halfway logarithmically between the two bounding probabilities, and at each value of P a point midway between the two bounding values of χ^2_p. These midpoints are connected by the lines that appear in the chart. The same procedure is followed later in Charts 2, 3, and 6.

As a second step toward determining the sampling distribution, empirical distributions of χ^2_p were determined from the 700 series described in the preceding section; that is, 200 values of χ^2_p were computed for $N = 25$, 300 for $N = 50$, and 200 for $N = 75$, all from three frequency classes. The observed distributions are given in Table 4, the class intervals being so chosen as to make the sum of the frequencies in each class as near 30 as possible without taking account of more than two decimal places (though the values of χ^2_p were calculated to four places). Chart 2 shows the three distributions (in cumulative form on semi-logarithmic coordinates, as in Chart 1), each value of χ^2_p being plotted individually. Chart 3 shows the same thing for each of seven sets of 100 values, the division into sets according with the order of drawing the samples.

A χ^2 test of homogeneity was applied to the three distributions, and the last column of Table 4 shows the contribution to χ^2 from each frequency class. Since the total χ^2 is 129.83 for $n = 46$, there can be no doubt that the three distributions differ significantly. They do not differ, however, in respects important for the present test. In the first place, the differences among the tails—approximately the highest 30 per cent of the observed values of χ^2_p—are not significant, even statistically: for the range beyond $\chi^2_p = 2.6$, the sum of the contributions to χ^2 is 15.3963, indicating a probability of about .4. Charts 2 and 3 perhaps create an impression of divergence at the tails. This is because the curves necessarily converge at $P = 1.00$ for $\chi^2_p = 0$, and the high serial correlation resulting from cumulation tends to keep them together in that neighborhood. Furthermore, discrepancies are minimized at high and magnified at low probabilities by the logarithmic scale—which was chosen partly for this very reason (the low probabilities for a test of significance requiring the closest scrutiny) and partly because the relative rather than the absolute magnitude of errors is relevant to probability measurements. The similarity of the tails is better shown by Chart 4, depicting the three distributions of Chart 2 as histograms.

In the second place, the discrepancies in the lower range of χ^2_p largely reflect marked irregularities of the separate distributions. These irregularities are highlighted by the fineness of the class intervals in that range, and tend to disappear if the intervals are broadened. Thus, class intervals by units of χ^2_p, i.e., under 1, 1 to 2, 2 to 3, etc., with over 9 as the last class, result in a χ^2 of 18.604 based on 18 degrees of freedom, corresponding with a probability of .41.

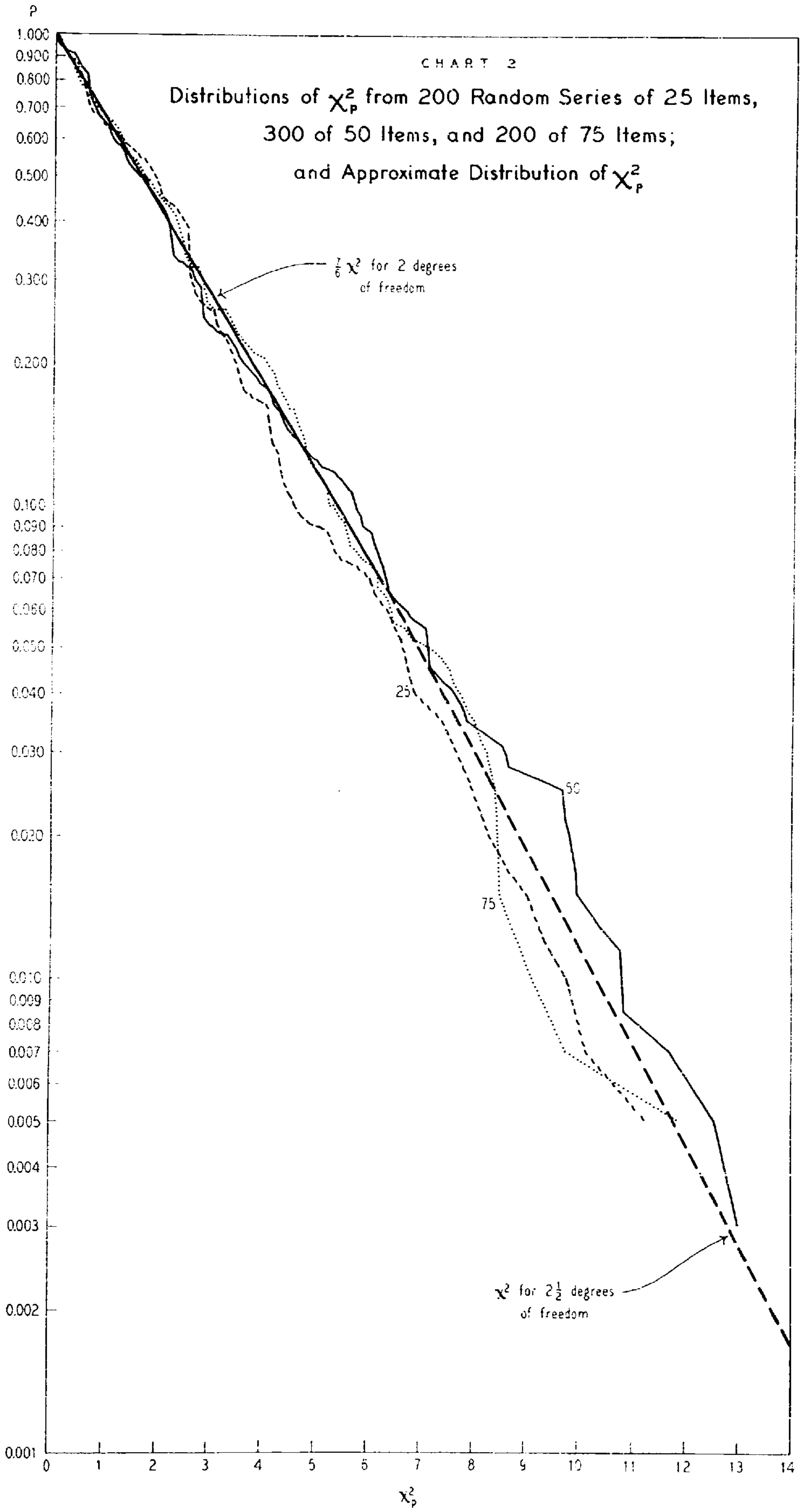
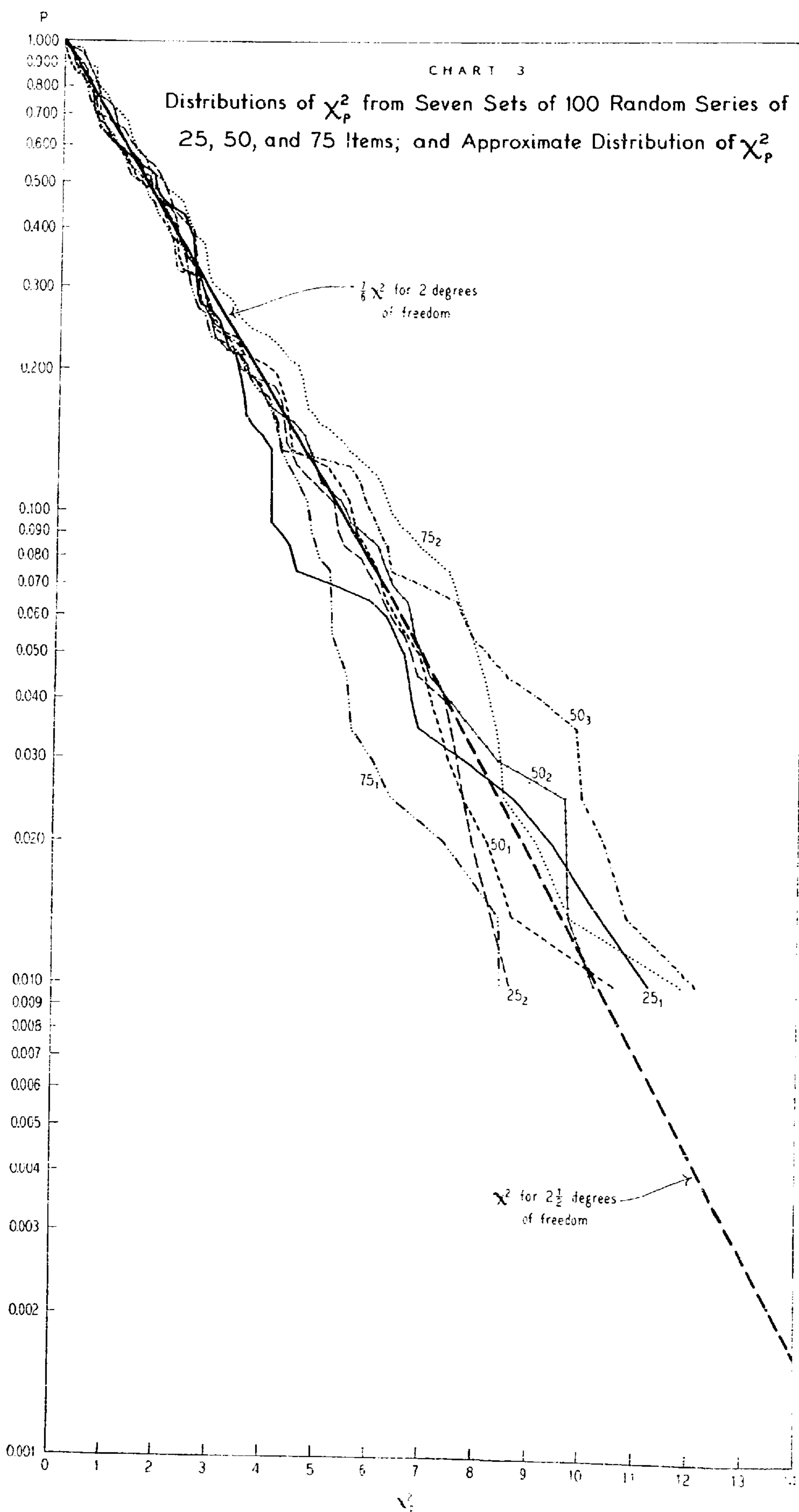


CHART 2

Distributions of χ^2_p from 200 Random Series of 25 Items, 300 of 50 Items, and 200 of 75 Items; and Approximate Distribution of χ^2_p



P
1.000
0.900
0.800
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.090
0.080
0.070
0.060
0.050
0.040
0.030
0.020
0.010
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
CHART 3
Distributions of χ^2_p from Seven Sets of 100 Random Series of 25, 50, and 75 Items; and Approximate Distribution of χ^2_p
$\frac{7}{6}\chi^2$ for 2 degrees of freedom
75_2
50_3
50_2
75_1
50_1
25_2
25_1
χ^2 for $2\frac{1}{2}$ degrees of freedom
0
1
2
3
4
5
6
7
8
9
10
11
12
13

The illustration in the penultimate paragraph of Section VIII further confirms this evidence that the non-homogeneity is a matter of erratic shifts in the numerous minor peaks of the distributions rather than of divergencies in fundamental form.

In any case, the tails of the distribution are of chief concern for a test of significance. When N is as large as 25 (and, as indicated by the exact distribution for $N = 12$, even when it is somewhat less) the distribution of χ_p^2 apparently is sufficiently near its limiting form for a single sampling distribution to be adequate.

The mean of the 700 values of χ_p^2 is 2.3049 and the variance is 5.0458. As an approach to the distribution of χ_p^2, it may simply be reduced by approximately one-seventh of its magnitude (more precisely by .3049/2.3049, but since these figures are merely estimates, and the ratio differs little from one-seventh, it seems sensible to use the more convenient figure) and compared with the χ^2 distribution for $n = 2$, which has a mean of two and tables for which are readily available. Such a comparison is shown in non-cumulative form in

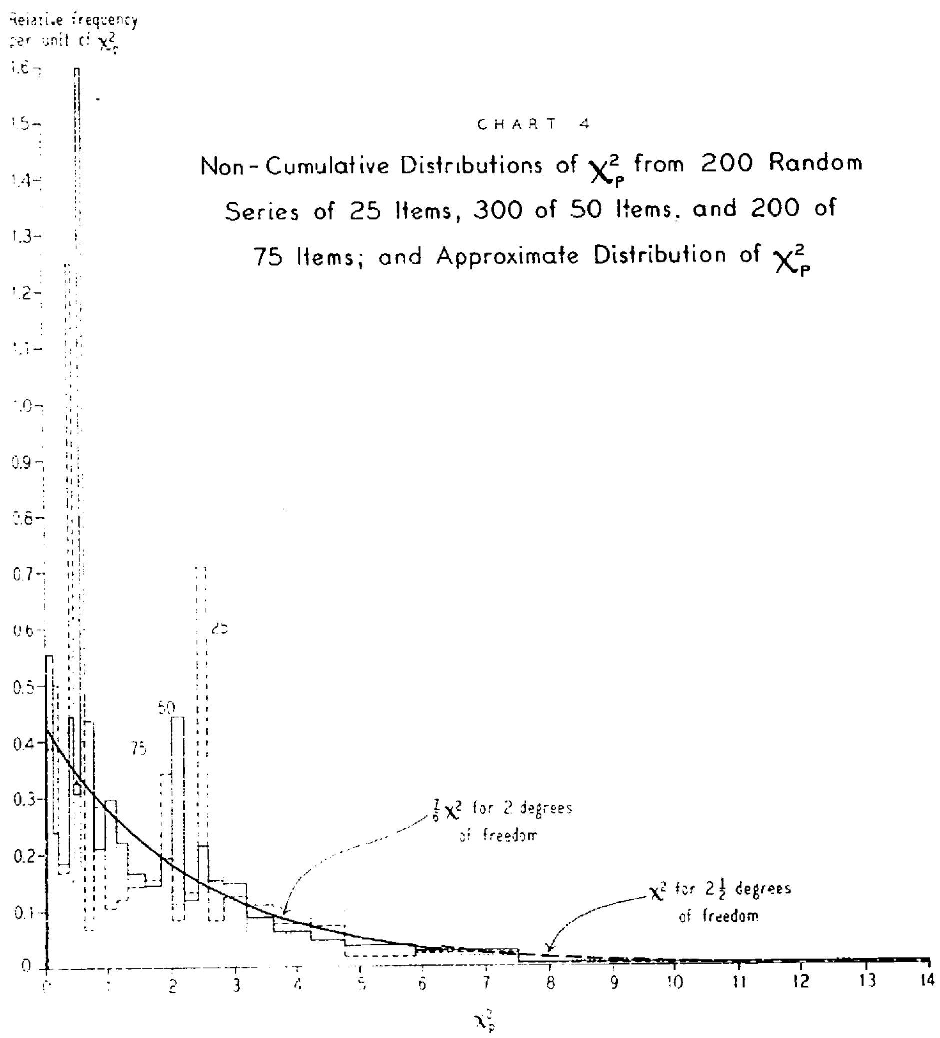
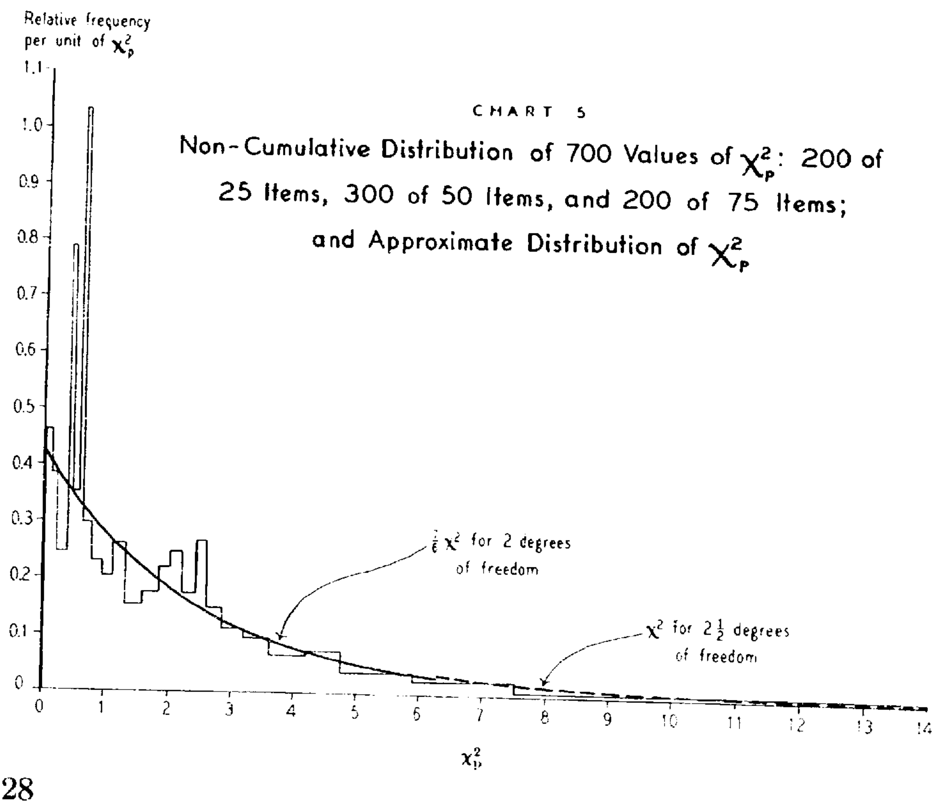
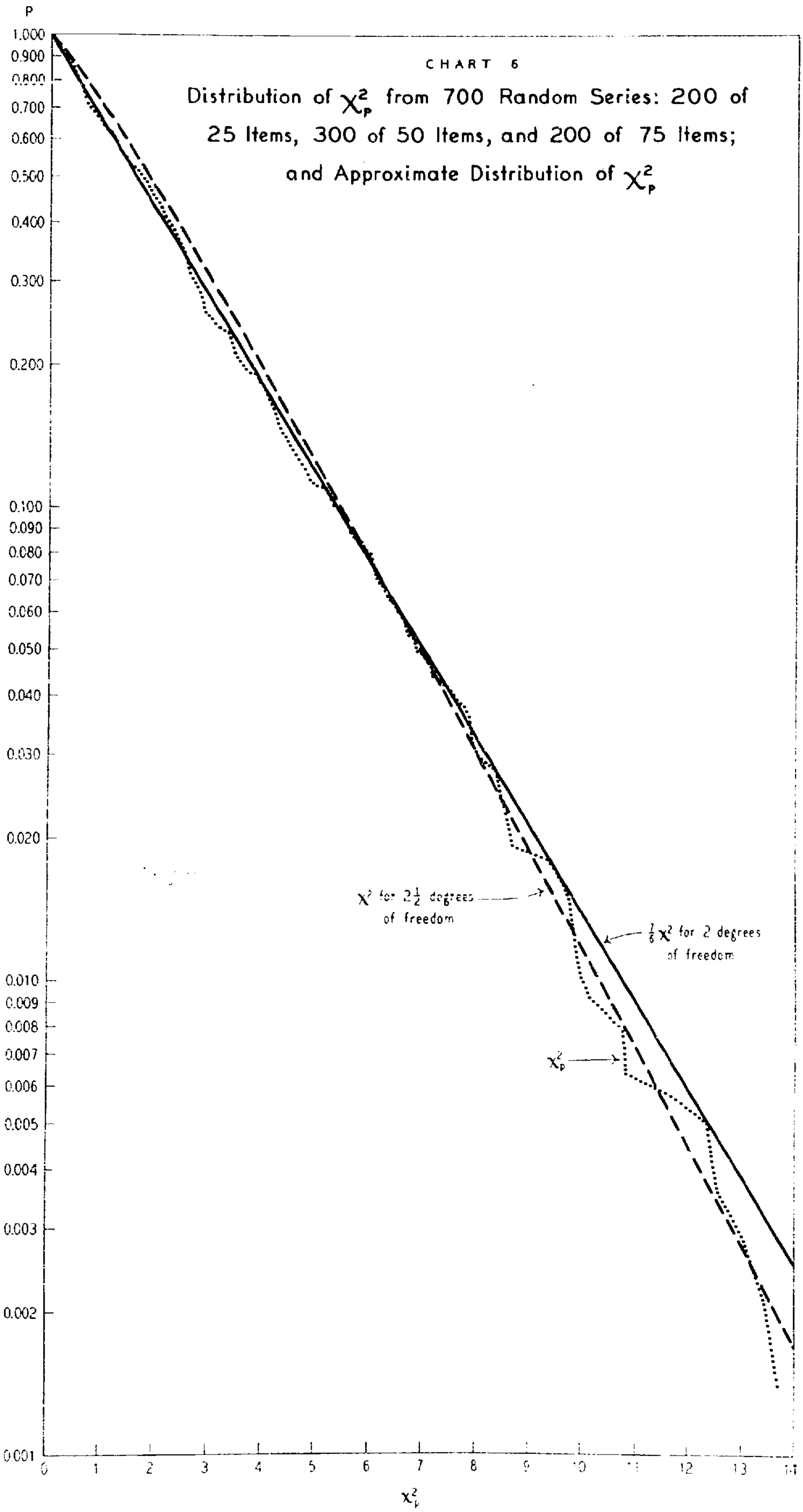


Chart 5, and in cumulative form in Chart 6; instead of reducing the values of χ_P^2 one-seventh, which would usually be the most convenient procedure, the values of χ^2 were increased one-sixth. In both charts the agreement between the line representing $\frac{7}{6}\chi^2$ and that representing the 700 observed values of χ_P^2 is quite good.

The fact that the variance of the observed values is less than that of $\frac{7}{6}\chi^2$ for $n = 2$ suggests, however, that a more satisfactory fit at the tails can be attained by using a χ^2 distribution having a variance of 5, e.g., χ^2 for $n = 2.5$. The values for χ^2 for $n = 2.5$ were obtained from the *Tables of the Incomplete Γ-Function*,[14] taking $p = .25$ and $u = \chi^2/\sqrt{5}$. Interpolations with respect to p were made linearly, and with respect to u, according to the logarithms of the probability (i.e., the logarithms of $1 - P$ in the notation of the *Tables* or of P in the present notation). This distribution is depicted by the dash line in Charts 1 to 6; its agreement with the observations at the tails—for χ_P^2 above about 5.5 and P below about .10—is very satisfactory indeed. In the main body of the distribution the curve whose mean value is equated to the sample mean gives a somewhat better fit.

It may occur to the reader that the mean values might have been equated by using χ^2 for $n = 2.3$. In the main body of the distribution this curve differs only slightly from $\frac{7}{6}$ of that for $n = 2$, and at the tails it lies as much below the observations as $\frac{7}{6}\chi^2$ for $n = 2$ lies above them. Hence the ready availability of tabulations for $n = 2$ is a decisive argument in its favor. Similarly, equating the variance by





P
1.000
0.900
0.800
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.090
0.080
0.070
0.060
0.050
0.040
0.030
0.020
0.010
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
CHART 6
Distribution of χ_p^2 from 700 Random Series: 200 of 25 Items, 300 of 50 Items, and 200 of 75 Items; and Approximate Distribution of χ_p^2
χ^2 for $2\frac{1}{2}$ degrees of freedom
$\frac{7}{6}\chi^2$ for 2 degrees of freedom
χ_p^2
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
χ_p^2

multiplication of the curve for $n = 2$ gives a result definitely poorer than that for $n = 2.5$. However, the use of χ^2 for $n = 2.5$ is based on only about 65 or 70 observations, the highest 10 per cent of the sample values of χ_p^2, and these, as pointed out in Section V, are not entirely independent.

If a curve of the χ^2 form is to be fitted to empirical observations, the maximum likelihood estimate of n proves to be the value that equates the digamma function[17, 43] of $\frac{n-2}{2}$ to the natural logarithm of one-half the geometric mean of the sample. For the 700 observations used here, the maximum likelihood estimate of n is 2.24. The values for $N = 25$, 50, and 75 are 2.24, 2.22, and 2.24, respectively. If both n and a coefficient of χ_p^2 are to be estimated by the criterion of maximum likelihood, n is the value that equates digamma of $\frac{n-2}{2}$ to the natural logarithm of $n/2$ times the ratio of the geometric mean to the arithmetic mean, and the coefficient is n divided by the arith-

TABLE 5

Tail of χ^2 Distribution for $2\frac{1}{2}$ Degrees of Freedom[a] for Use as Approximate Distribution of χ_p^2 when $N > 12$[b]

χ^2	P	χ^2	P	χ^2	P
5.448	*.10*	8.00	.0301	11.50	.0057
5.50	.0976	*8.009*	*.03*	11.75	.0050
5.674	*.09*	8.25	.0268	*11.755*	*.005*
5.75	.0869	8.50	.0238	12.00	.0044
5.927	*.08*	8.75	.0211	12.25	.0039
6.00	.0773	*8.836*	*.02*	12.50	.0035
6.163	*.07*	9.00	.0187	12.75	.0031
6.25	.0687	9.25	.0166	13.00	.0027
6.50	.0612	9.50	.0148	13.25	.0024
6.541	*.06*	9.75	.0131	13.50	.0022
6.75	.0543	10.00	.0116	13.75	.0019
6.898	*.05*	10.25	.0103	14.00	.0017
7.00	.0483	*10.312*	*.01*	14.25	.0015
7.25	.0429	10.50	.0091	14.50	.0013
7.401	*.04*	10.75	.0081	14.75	.0012
7.50	.0382	11.00	.0072	15.00	.0010
7.75	.0339	11.25	.0064	*15.085*	*.001*

[a] Calculated from *Tables of the Incomplete Γ-Function.*[44]

[b] P denotes the probability that χ^2 will equal or exceed the specified value. Interpolations may be made linearly with respect to log P. For values of χ_p^2 less than 6.3, $\frac{6}{7}\chi_p^2$ should be referred to the usual tables of χ^2 for two degrees of freedom; or P may be calculated as the reciprocal of the natural antilogarithm of $\frac{3}{7}\chi_p^2$ (or the reciprocal of the common antilogarithm of $.186126\chi_p^2$). When $N \leq 12$, see Table 3.

metic mean; for these data, n is 2.12 and the coefficient, .92. If n is fixed arbitrarily and a coefficient estimated by maximum likelihood, the estimate is the ratio of n to the arithmetic mean of the sample, in this case 2/2.3049 or approximately 6/7.

In practice, then, the procedure for interpreting χ_p^2, assumed always to be calculated from three frequency classes, is as follows: If χ^2_p is less than 6.3 (the point of intersection between the ogives of $\frac{7}{8}\chi^2$ for $n = 2$ and χ^2 for $n = 2.5$), reduce it one-seventh and refer to the usual χ^2 tables for two degrees of freedom. This procedure is satisfactory for all values of χ^2_p, but for values above 6.3 somewhat more accurate probabilities are apparently obtained by referring the whole value of χ^2_p to Table 5, which gives the distribution of χ^2 for $n = 2.5$. The curve, composed of two segments, corresponding with this procedure has been added to Charts 1–6. When $N \leq 12$ the exact distributions of Table 3 should, of course, be used.

A thorough mathematical investigation of the proper sampling distribution is much to be desired. It should determine the distribution of χ^2_p not merely for three frequency classes but also for more. More important, it should analyze the broader question of what form of test is most appropriate to phase durations.

VII An Auxiliary Test

A test of the mean phase duration would, as was mentioned in Section V, retrieve some of the information lost by throwing all phases of more than two years into a single frequency class; but the variance of the expected distribution of phase durations, it was pointed out in Section IV, cannot be used as a test of the observed mean duration. It is quite simple, however, to test the total number of phases; and, except for an unimportant discrepancy occasioned by excluding from the definition of a phase the incomplete phase before the first turning point and that after the last turning point, this is equivalent to a test of the mean duration. The mean phase duration is merely the total duration of all phases divided by the number of phases, and the total duration of all phases, plus the durations of the two incomplete phases, is a constant, $N - 1$. The mean duration, therefore, depends only upon the number of phases and the lengths of the two incomplete phases. The number of phases is simply the number of turning points reduced by one (except for the negligible qualification that the two are equal when there are no turning points).

Now the expected number of turning points is shown in Section III to be

$$\frac{2(N-2)}{3},$$

the variance of the number of turning points is

$$\frac{16N-29}{90},$$

the third moment about the mean is

$$-\frac{16(N+1)}{945},$$

and the fourth moment about the mean is

$$\frac{448N^2-1976N+2301}{4725}.$$

The variance may be found inductively by computing for a series of values of N the values of the second moment about the origin, utilizing the exact probabilities for all possible numbers of turning points, which are explained later in this Section. When $N > 3$, the second differences of this series have the constant value 8/9; since for $N = 4$ the value is 13/6 and the first difference is 12/5, the general expression for the second moment about the origin is

$$\frac{40N^2 - 144N + 131}{90}.$$

Deducting the square of the expected number of turning points leaves the required variance. A series of values of the third moment about the origin shows, when $N > 7$, constant third differences of 16/9; since for $N = 8$ the value is 1618/21, the first difference is 2665/63, and the second difference 592/45, the general expression for the third moment about the origin is

$$\frac{280N^3 - 1344N^2 + 2063N - 1038}{945}.$$

From this the third moment about the mean is easily computed by familiar formulas.[22] The fourth differences of a series of fourth moments about the origin are constantly 128/27 when $N > 7$; when $N = 8$ the fourth moment is 12683/35, the first difference is 52940/189, the second difference 23928/175, and the third difference 192/5. The fourth moment about the origin is, therefore,

$$\frac{2800N^4 - 15680N^3 + 28844N^2 - 19288N + 4263}{14175},$$

whence the fourth moment about the mean may be derived.

As N increases, the skewness, measured by the ratio of the squared third moment about the mean to the cubed variance, approaches zero, and the kurtosis, measured by the ratio of the fourth moment about the mean to the squared variance, approaches three. That the skewness and kurtosis approach their values for a normal distribution suggests that the distribution of the number of turning points approaches normality as the length of series increases. Furthermore, empirical comparisons indicate that normality is approached with such rapidity that the discrepancy can be ignored when $N > 12$. Hence the number of turning points can be regarded as normally distributed about

$$\frac{2(N - 2)}{3}$$

or the number of phases as normally distributed about

$$\frac{2N - 7}{3},$$

either with variance of

$$\frac{16N - 29}{90}.$$

In using the normal distribution, the discrepancy between the observed and expected numbers of turning points or of phases should be reduced in absolute value by one-half unit to allow for discontinuity, and the distribution should be truncated—i.e., the probability of values above $N - 2$ or below 0 deducted and the remaining probabilities raised proportionately. (With a single tail of the distribution, truncation is unimportant; with the two tails combined, as in Table 6, its chief effect is when the departure from expectation is negative and greater absolutely than is possible in the positive direction.) For the sake of comparison the approximate probabilities estimated from a truncated normal curve are given for $N = 12$ in Table 6, which shows for $N = 6$ to 12 the exact probability of obtaining a discrepancy from expectation as great as or greater (*in absolute value*) than that represented by each number of turning points.

The exact probabilities were calculated from a recursion formula derived inductively. Letting $f_N(t)$ be the number of permutations of N different numbers that have exactly t turning points, $f'_N(t)$ be the first cumulation of $f_N(t)$, and $f''_N(t)$ be the second cumulation—i.e.,

$$f'_N(t) = \sum_{i=0}^{t} f_N(i), \qquad \text{and} \qquad f''_N(t) = \sum_{j=0}^{t} f'_N(j)$$

the formula is

$$f''_N(t) = (t + 1)f''_{N-1}(t) + (N - t - 3)f''_{N-1}(t - 2) \tag{5}$$

where

$$f''_N(t) = 0 \quad \text{if} \quad t < 0 \quad \text{and} \quad f''_N(N - 2) = \frac{(N + 1)!}{3}.$$

Values of $f'_N(t)$ are obtained by differencing the series of $f''_N(t)$, and values of $f_N(t)$ by differencing $f'_N(t)$. (For a simpler recursion formula see Section IX.) An expression for $f''_N(t)$ directly in terms of N and t is

$$f''_N(t) = \frac{1}{2^{t-1}} \sum_{k=0}^{\frac{t}{2} \text{ or } \frac{t-1}{2}} (-1)^k (t - 2k + 1)^N A(k) \tag{5a}$$

where

$$A(k) = \sum_{i=0}^{\frac{k}{2} \text{ or } \frac{k-1}{2}} a_i b_{k-2i}$$

$$a_i = \frac{(N - t - 4 + j)!}{j!(N - t - 4)!}$$

$$b_h = \frac{(2N - t - 5)!}{h!(2N - t - 5 - h)!}.$$

The test of the number of turning points or phases, referred to briefly as the p-test, is, of course, related to the χ^2_p test. A frequency

TABLE 6

Exact Probability that the Discrepancy between Observed and Expected Numbers of Turning Points will Equal or Exceed (*in absolute value*) the Discrepancy corresponding with any Observed Number of Turning Points, Six to Twelve Observations; and Normal Approximation for Twelve Observations

							$N = 12$	
t[a]	$N = 6$	$N = 7$	$N = 8$	$N = 9$	$N = 10$	$N = 11$	Exact	Approximate
0	.0028	.0004	.0000	.0000	.0000	.0000	.0000	.0000
1	.0861	.0250	.0063	.0014	.0003	.0001	.0000	.0000
2	.5833	.2988	.1436	.0257	.0079	.0022	.0005	.0010
3	1.0000	1.0000	.6374	.1937	.0911	.0416	.0082	.0093
4	.2556	.6325	1.0000	.6694	.4098	.2373	.0642	.0693
5		.1329	.6374	1.0000	1.0000	.6978	.2739	.2780
6			.1436	.3847	.6950	1.0000	.7173	.7180
7				.0694	.2304	.6978	1.0000	1.0000
8					.0357	.2373	.4638	.4597
9						.0416	.1350	.1384
10							.0195	.0248

[a] t represents the number of turning points. 'Phases' may be read for 'turning points' if the entries of the first column are decreased by one and the first line is ignored.

For larger values of N, $\pm \dfrac{|3t - 2N + 4| - 1.5}{\sqrt{1.6N - 2.9}}$ may be regarded as a normal deviate. If t in this expression represents the number of phases, instead of the number of turning points, the 4 must be changed to 7.

distribution of phase durations may be thought of as compounded of two elements: first, the total number of phases occurring; and second, their proportionate distribution by duration. The p test is sensitive only to the first element, whereas the χ_p^2 test is sensitive chiefly (though by no means exclusively) to the second. χ_p^2 would be sensitive only to the second component if it were based on the relative frequencies of phase lengths expected with a specified total number of phases. It would, therefore, be independent of the p-test, and the two tests could be compounded by Fisher's method.[22] The relative frequencies expected with a specified total, however, are not those given by formula 3, which is valid only when the total number of phases is unrestricted; so until the proper expected relative frequencies and the sampling distribution of this kind of χ_p^2 are determined, such a combination cannot be made. Although the illustrations in the next section are concerned with the χ_p^2 test, the probabilities resulting from the p test have been recorded in the tables as $P(p)$.

By using only one tail of its distribution, the p test can be made to discriminate between series having too many (i.e., too short) cycles and those having too few (i.e., too long).

VIII Applications

To illustrate the application of the foregoing technique to an economic problem, we analyze sweetpotato production, yield per acre, and acreage harvested in the United States, 1868–1937[51] (Table 7 and Chart 7). This crop, selected from a number of crops to which we have applied the method in connection with the National Bureau's *Studies in Cyclical Behavior*, is presented here simply for its illustrative advantages; it is, however, fairly typical of the group in its cyclical behavior as judged by the present tests. The turning points are indicated by asterisks in Table 7.

Occasionally, as in this illustration, metric data include equal observations, presumably because of limitations on the accuracy of measurements. Only when these equal values are adjacent is there a point that neither continues nor reverses the direction of movement. In these instances perhaps the best procedure is to regard the ties not as truly equal but as a random sequence of unequal observations, tabulate the distribution of phase lengths for each possible arrangement of plus and minus signs between the ties, and average the resulting distributions, each weighted by the probability of the particular set of signs it represents, as computed by formula 4 (Section VI), using as N the number of observations in the tied sequence. This procedure, of course, may result in an observed distribution containing fractional frequencies. Thus, the sequence 0, 1, 1, 1, 2, (when the 0 and the 2 are known, in view of the preceding and succeeding values, to be turning points) represents four possibilities, which may be denoted by the signs of the first differences as $++++$, $+-++$, $++-+$, and $+--+$. The first possibility corresponds with one four year phase and each of the other three with two one-year and one two-year phases. Since the probabilities of the four cases are $\frac{1}{6}$, $\frac{1}{3}$, $\frac{1}{3}$, and $\frac{1}{6}$, respectively, the set of points represents $2\frac{2}{3}$ phases, of which $1\frac{2}{3}$ are of one year, $\frac{5}{6}$ of two years, and $\frac{1}{6}$ of four years. It would perhaps be preferable to carry the calculation through to the ultimate probability value for each possible frequency distribution and then obtain the weighted average of these probabilities, but ordinarily this result will not differ sufficiently from that based on the weighted average of the frequency distributions to justify the extra trouble. The assumption of randomness at the basis of this procedure is in conformity with the null hypothesis, so differences from the null hypothesis cannot be attributed to ties; but ties do reduce the sensitivity of the test to departures from randomness.

CHART 7

Sweetpotato Production, Yield per Acre, and Acreage Harvested
United States, 1868-1937

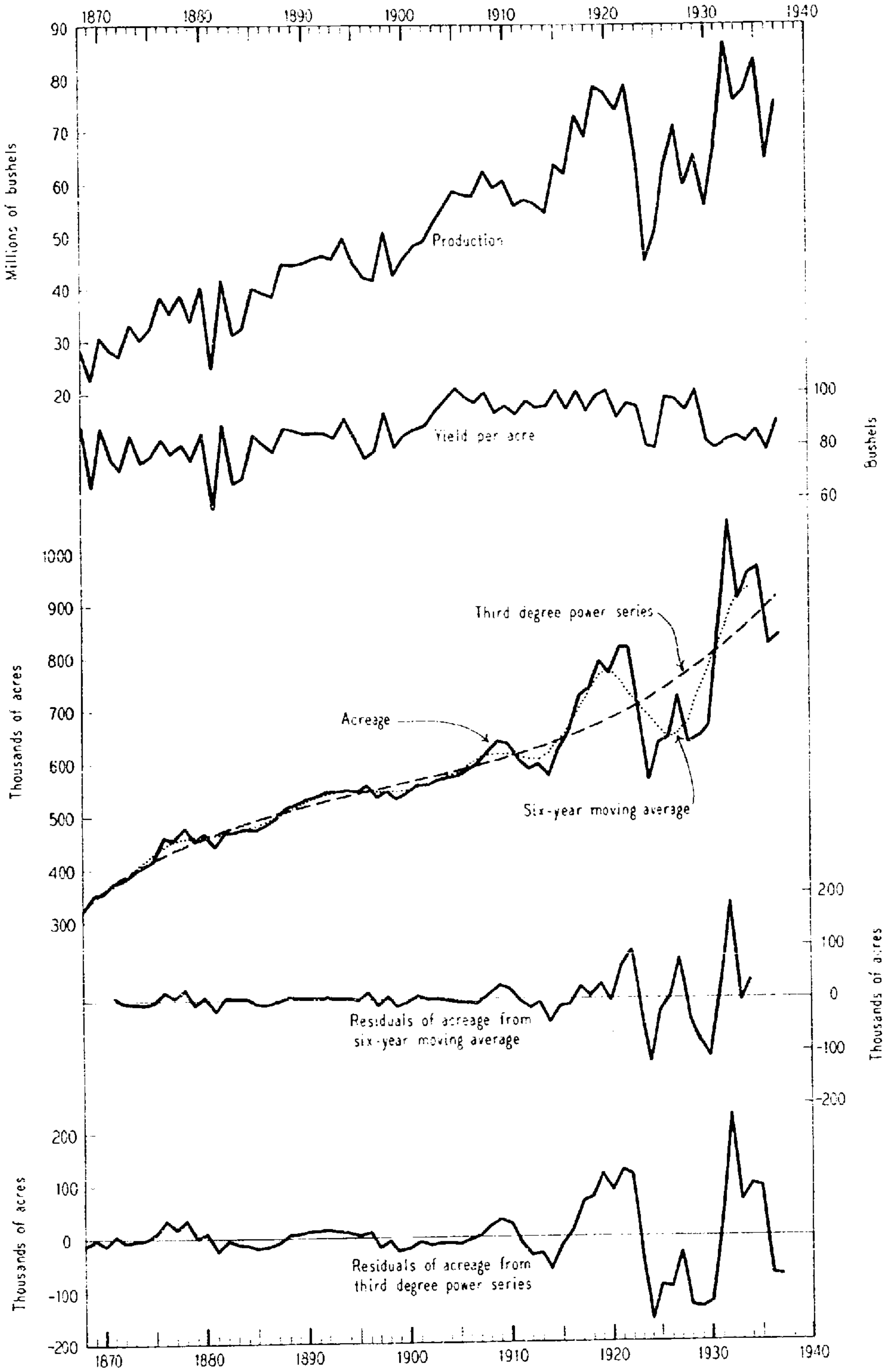


TABLE 7

Sweetpotato Production, Yield per Acre, and Acreage Harvested United States, 1868–1937[51]

	PRODUCTION (thousands of bushels)	YIELD PER ACRE (bushels)	ACREAGE HARVESTED (thousands) Actual	6-yr. moving average centered	Residual
1868	28,557	87.9	325		
1869	22,713*	64.7*	351		
1870	30,911*	87.8*	352		
1871	28,093	74.9	375	369	6
1872	27,148*	71.6*	379	382	−3
1873	33,266*	84.9*	392	397	−5
1874	30,150*	74.3*	406	413	−7*
1875	32,518	76.5	425	428	−3
1876	38,214*	83.1*	460*	441	19*
1877	35,196*	77.5*	454*	451	3*
1878	38,703*	80.8*	479*	458	21*
1879	33,851*	75.1*	451*	460	−9*
1880	40,128*	85.6*	469*	462	7*
1881	24,830*	56.3*	441*	463	−22*
1882	41,742*	89.0*	469	465	4*
1883	31,096*	66.2*	470	468	2*
1884	32,376	68.0	476*	473	3*
1885	40,111*	84.6*	474*	481	−7
1886	39,061	81.2	481	489	−8*
1887	38,528*	78.0*	494	498	−4
1888	44,838*	87.1*	515	508	7*
1889	44,779*	85.9	521	518	3*
1890	44,963	84.7*	531	528	3*†
1891	45,773	85.2*	537	535	2*
1892	46,364*	85.2*†	544	540	4*
1893	45,615*	83.7*	545	544	1*
1894	49,676*	90.6*	548*	546	2*
1895	44,886	82.4	545*	545	0*
1896	42,001	75.4*	557*	544	13*
1897	41,587*	78.3	531*	543	−12*
1898	50,743*	92.8*	547*	543	4*
1899	42,245*	79.6*	531*	544	−13*
1900	45,684	84.3	542	547	−5
1901	48,156	86.3	558*	552	6*
1902	48,975	87.8	558*†	558	0*
1903	52,871	93.6	565	565	0*†
1904	55,515	97.4	570	572	−2
1905	58,560*	102.0*	574	580	−6*
1906	57,750	98.7	585	591	−6*†
1907	57,332*	96.2*	596	603	−7*
1908	62,299*	100.3*	621	611	10
1909	58,994*	92.3*	639*	613	26*
1910	60,310*	95.1*	634	613	21
1911	55,285*	91.7*	603	609	−6
1912	56,644*	96.7*	586*	604	−18*

* Turning point.

† Ties. For method of handling see Sec. VIII, second paragraph.

TABLE 7—*Concluded*

	PRODUCTION (thousands of bushels)	YIELD PER ACRE (bushels)	ACREAGE HARVESTED (thousands) Actual	6-yr. moving average centered	Residual
1913	55,998	94.0*	596*	605	−9*
1914	54,145*	94.7	572*	617	−45*
1915	63,241*	100.9*	627	640	−13
1916	61,546*	93.5*	658	669	−11
1917	72,767*	100.4*	725	701	24*
1918	68,581*	92.9*	738	734	4*
1919	78,272*	99.0	791*	763	28*
1920	76,999	100.4*	767*	772	−5*
1921	73,708*	90.2*	817*†	753	64
1922	78,365*	95.9*	817*†	725	92*
1923	63,871	94.8	674	702	−28
1924	44,884*	79.6	564*	684	−120*
1925	50,139	78.8*	636	662	−26
1926	63,300	98.1*	645	644	1
1927	70,897*	97.9	724	651	73*
1928	59,178*	93.0*	636*	677	−41
1929	64,963*	100.6*	646	729	−83
1930	54,415*	81.3	669	779	−110*
1931	66,849	78.6*	850	821	29
1932	86,436*	81.9	1,056*	875	181*
1933	75,248*	82.9*	908*	914	−6*
1934	77,482	80.9*	958	926	32
1935	83,128*	85.8*	969*		
1936	64,144*	78.0*	822*		
1937	75,053	89.3	840		

* Turning point.
† Ties. For method of handling see Sec. VIII, second paragraph.

The phase durations for the sweetpotato series are tabulated in Table 8, together with the expected frequencies on the hypothesis that the observations are random and independent. From the values of χ^2_P and their corresponding probabilities (Table 8), it appears that the fluctuations in production conform well with chance; of the two components of total production, yield per acre conforms well and acreage harvested not at all, suggesting that fluctuations in production depend more upon fluctuations in yield than upon fluctuations in acreage.

It is, of course, apparent even from casual inspection of Table 7 or Chart 7 that total production does not constitute a random series, but has a marked upward trend. In general, the method here presented is not very sensitive to a primary trend. By 'primary' trend we mean an elementary function whose first and second derivatives have few, if any, changes in sign and only gradual changes in magnitude. It corresponds with the basic secular trend, as contrasted with long waves, trend-cycles, business cycles, seasonal variations, etc.

"A line of primary trend will trace out synoptically and elegantly the general secular movement without giving much heed to the details of the movement."[18] The removal or introduction of a trend can alter the order of magnitude between adjacent items (i.e., change the sign of their difference) only if the trend factor for a single year is greater than the difference between successive trend-adjusted items. Consider, for example, a set of observations 1, 2, 1, 2, 1, in which there is no trend. If an upward trend having a rate of rise of less than one unit per year is introduced, there will be no alteration in the pattern of expansions and contractions; for since none of the differences is less than one in absolute value it cannot have its sign changed by adding a quantity less than one.

Ordinarily, of course, there is no minimum to the absolute value of the random factor. When the difference between consecutive residuals from trend is less absolutely than the change due to trend between the two points, the sign of a difference may depend upon whether trend is included or eliminated. If, as frequently happens in economic time series, the sequence of residuals is such that differ-

TABLE 8

Frequency Distributions of Phase Durations in Sweetpotato Production, Acreage Harvested, and Yield per Acre United States, 1868–1937

DURATION OF PHASE	EXPECTED FREQUENCY[a]	OBSERVED FREQUENCY						
		PRODUCTION			ACREAGE[b]			YIELD
(years)		Total	Expansion	Contraction	Total	Expansion	Contraction	Total
1	27.917	32	16	16	18	5.5	12.5	30
2	12.100	10	4	6	2.5	2	0.5	14
3	3.431	3	2	1	3.5	2	1.5	1
4	0.737	0	0	0	1	1	0	0
5	0.128	0	0	0	1	1	0	0
6	0.019	1	1	0	0	0	0	1
7	0.002	0	0	0	0.5	0.5	0	0
8	0.000	0	0	0	0	0	0	0
9	0.000	0	0	0	1	1	0	0
10	0.000	0	0	0	0.5	0.5	0	0
Total	44.333	46	23	23	28	13.5	14.5	46
χ_p^2 [c]		1.363	1.323	0.920	13.487	14.676	5.444	1.696
$P(\chi_p^2)$		.56	.57	.67	.002	.001	.10	.48
Compound probability[22]			.75			.001		
$P(p)$ [d]		.73			.000			.73

[a] Expected frequencies for expansions or contractions are one-half those in this column.
[b] For explanation of fractional frequencies see Sec. VIII, second paragraph.
[c] Computed by combining all durations in excess of 2.
[d] Probability for total number of phases; see Sec. VII.

ences as small as the trend factor for a single year are rare, the distribution of phase lengths will not be much affected by the presence or absence of trend.

Still another factor minimizing the effect of a primary trend on the test is that a positive trend tends to lengthen expansions but to shorten contractions. In general it tends to make one-year contractions more numerous than one-year expansions, and expansions of more than one year more numerous than contractions of more than one year, without altering greatly the total number of phases of a given duration. Opposite effects are produced by negative trends. In such cases the existence of trend may be concealed by the frequency distribution of all phases, but be revealed by separate distributions of expansions and contractions.

Separate distributions of expansions and contractions are shown (Table 8) for production and acreage, but not for yield, which has little or no apparent trend. For production, both distributions conform well to the expected distribution and to each other. A test of homogeneity shows $\chi^2 = 1.4$, which, for $n = 2$, signifies a probability of .5. There is, therefore, no indication that the distribution of phases has been affected by the primary trend. For acreage the two distributions differ markedly in a manner attributable to trend, and the probability resulting from application of the χ^2 test for homogeneity is only .02. The non-randomness evidenced in the acreage series may, therefore, be at least partly attributable to a primary trend rather than to secondary movements.

In interpreting tests of the homogeneity of contractions and expansions, it should be remembered that the positive serial correlation in phase lengths pointed out in the second paragraph of Section VI has a tendency to produce homogeneity between the distributions of expansions and contractions. This affects also the probability compounded from the probabilities of the χ_p^2-s for expansions and contractions (Tables 8 and 10), since it means that the two probabilities are not entirely independent.

Lack of sensitivity to primary trend is a limitation of the technique from the viewpoint of detecting its existence. On the other hand, it is not difficult to determine by other methods whether a primary trend exists—the rank correlation between the variate and the date often affords a satisfactory test. And for determining whether the systematic variation contains secondary components, e.g., cyclical or seasonal variations, it is a decided advantage of the present method that it frequently—perhaps usually with economic data—gives satisfactory results regardless of the presence of trend, thus avoiding the complexities of trend elimination. It is, of course, possible for secondary components also to be concealed if their year to year changes are definitely smaller than the year to year random changes; this is not so likely as in the case of the primary trend, but it is a real possibility in the case of gradual movements, e.g., long waves.

A second example illustrates the usefulness of the technique as a criterion of the fit of moving averages and for selecting the proper period for a moving average. If a moving average or any other curve describes adequately the systematic variation in a series, the residuals should constitute a random series. If the period is too long, waves or cycles may appear in the residuals; if it is too short, the residuals will cluster too closely about the line.

To illustrate this application, ten moving averages having spans of from 2 to 11 years were fitted to the data on sweetpotato acreage and the residuals tested for randomness. Each moving average uses equal weights. Averages based on an even number of points, however, are centered half-way between observations so must be interpolated to find the smoothed value corresponding with a given observation, and this amounts to using a moving average based on one more year with the extremes receiving only half the weight of the intervening years. What is designated a six-year moving average with equal weights is thus really a seven-year moving average with equal weights for the central five years and weights of one-half for the first and seventh years. The moving averages were rounded to the number of figures appearing in the data.

Table 9 shows the values of χ_p^2, together with the corresponding probabilities, obtained by testing the residuals for randomness. Its chief feature is that the moving averages based on even numbers of years give better results than the corresponding averages based on odd numbers; the implicit tapering of the weight diagram involved in interpolating seems to improve the fit markedly. A second striking feature is that the probabilities first rise, then decline. Thus, the probabilities for the odd numbered moving averages reach a maximum of .24 at seven years while the even numbered give the best fit at six years, when the probability is .61. Had other weight diagrams been tested, they might have resulted in still better fits.

A 'better' fit, in the present sense, does not necessarily give a closer approximation to the data. It is one for which the residuals behave more like a series of independent, random observations, as judged by the sequences in the signs of the first differences. The closest fits to the original observations are given by the shortest moving averages; but these describe not only the systematic variation but also a portion of the random fluctuations. If the moving average is either too short or too long, χ_p^2 will be significantly large; but the source of its magnitude is not the same in the two cases. If the moving average is too short, there are too many short phases and too few long ones; if too

TABLE 9

Frequency Distributions of Phase Durations in Residuals from Moving Averages fitted to Sweetpotato Acreage Harvested United States, 1868–1937

PERIOD OF MOVING AVERAGE (years)		FREQUENCY OF PHASES: One Year	Two Years	Over two Years	Total	χ_p^2	$P(\chi_p^2)$	$P(p)$[a]
2	Expected	27.083	11.733	4.183	43	16.823	.0004	.0008
	Observed	46	8	1	55			
3	Expected	27.083	11.733	4.183	43	16.823	.0004	.0008
	Observed	46	8	1	55			
4	Expected	26.25	11.367	4.05	41.667	1.857	.45	.59
	Observed[b]	31.25	8.25	4.5	44			
5	Expected	26.25	11.367	4.05	41.667	4.626	.14	.31
	Observed[b]	19.417	9.917	7.333	36.667			
6	Expected	25.417	11	3.917	40.333	1.141	.61	.80
	Observed[b]	25.5	8.25	5.25	39			
7	Expected	25.417	11	3.917	40.335	3.287	.24	.96
	Observed[b]	28.5	6	5.5	40			
8	Expected	24.583	10.633	3.783	39	2.547	.34	.88
	Observed	25	7	6	38			
9	Expected	24.583	10.633	3.783	39	4.634	.14	.09
	Observed[b]	18.75	7.5	6.75	33			
10	Expected	23.75	10.267	3.65	37.667	3.175	.26	.33
	Observed[b]	23	5.5	5.5	34			
11	Expected	23.75	10.267	3.65	37.667	9.861	.01	.05
	Observed	22	2	7	31			

[a] Probability for total number of phases; see Sec. VII.
[b] For explanation of fractional frequencies see Sec. VIII, second paragraph.

long, there are too few short phases and too many long ones. This effect appears in the actual frequency distributions of residuals from the ten curves (Table 9).

If a series is conceived to have residuals that are in some sense random but not independent of one another, a moving average selected according to the present criterion will tend to include the part of the random element that is serially correlated with the preceding items. As is well known, serial correlation will produce 'cycles' in an otherwise random series.

Further examination of the residuals from the six-year moving average reveals that separate distributions for expansions and contractions do not differ significantly from expectation or from one another (Table 10). A test of homogeneity yields a χ^2 of 2.247 based on two degrees of freedom, which corresponds with a probability of .33 and therefore indicates no significant difference. According to the criterion of sequences in the direction of movement of the residuals, therefore, the six-year moving average seems to give a satisfactory fit. The values of this moving average and of the residuals from it (Table 7 and Chart 7) suggest that until about 1907 the systematic variation in sweetpotato acreage consisted principally of a simple trend and mild undulations of about 15 years, but that the undulatory movements became marked thereafter.

Were we presenting a detailed analysis of the sweetpotato data, instead of using them merely to illustrate a technique, it might be desirable to treat the two portions of the series separately, perhaps using different moving averages. If separate frequency distributions of phase durations in the residuals are made for the periods before and after 1907, there seems to be an excess of short phases in the earlier period and an excess of long phases in the later period, though in neither are the differences from expectation statistically significant. The values of χ_p^2 are 2.911 and 6.388, respectively, corresponding with probabilities of .29 and .06. Probably a longer moving average would be more satisfactory for the earlier period and a shorter one for the later period. A possible explanation (which we have not yet investigated) lies in the assertion of the U. S. Department of Agriculture[50] that "in 1909 there appears a marked concentration of production in certain states". Concomitant with the increase in specialization there may have been greater sensitivity of producers to price and cost factors. Inverse relationships between the price of

TABLE 10

Frequency Distributions of Durations of Expansions and Contractions in Residuals from Six-Year Moving Average of Sweetpotato Acreage Harvested

United States, 1868–1937

DURATION OF PHASE (years)	EXPECTED FREQUENCY	OBSERVED FREQUENCY[a]	
		Expansions	Contractions
1	12.708	11.5	14
2	5.500	6	2.25
3	1.557	2	2.5
4	0.333	0	0.5
5	0.058	0	0
6	0.008	0	0.25
Total	20.167	19.5	19.5
χ_p^2 [b]		0.161	2.903
$P(\chi_p^2)$		.93	.29
Compound probability[22]		.62	

[a] For explanation of fractional frequencies see Sec. VIII, second paragraph.

[b] Computed by combining all durations in excess of 2.

cotton and the acreage of sweetpotatoes harvested the following year, and between cotton acreage and sweetpotato acreage harvested the same year, are noticeable from the time of the 1914–18 War. The increased amplitude of fluctuations in sweetpotato acreage from the time of that War coincides with increased amplitudes in cotton prices and acreage.

In order to compare this new test with a more elaborate procedure frequently used in time series analysis, a power series $y = a + bx + cx^2 + dx^3 + \cdots$ was fitted by the method of least squares to the series on sweetpotato acreage harvested. The calculations were carried as far as the ninth degree term, using the technique of orthogonal polynomials,[22] but none beyond the third reduced the residual variance significantly. According to the usual criterion, therefore, a third degree curve would be regarded as giving an adequate fit (Chart 7). The residuals from the third degree curve were then subjected to the present test. There were 24 phases of one year, 3 of two years, and 9 of more than two years, producing a χ_p^2 of 12.47 and a probability of .004, from which it is clear that the fit of the third degree power series is quite inadequate. (The normal deviate for the total number of phases—see Section VII—is 2.26, indicating a probability of .02.). The power series required very much more time and labor for fitting and testing than did the moving averages; but the result seems considerably worse and the variance test completely misleading. The power series does give a good representation of the primary trend, but since in this particular case it is good only in a descriptive sense—even a fairly short projection being obviously absurd because the curve rises with increasing acceleration—it has little advantage over methods that avowedly produce mere descriptions. The shortcomings, of course, lie not in the method of least squares but in the fallacy of inferring that a third degree power series gives a thorough fit because no other power series effects a significant reduction in the standard error. It is doubtful that economic time series generally can be adequately represented by power series fitted in this way, though such functions may be useful for describing certain portions of the systematic variation (a use in which the present test is inappropriate).

An obvious limitation of the present test is that it by no means utilizes all the information in the data. This shortcoming, in fact, partly accounts for the usefulness of the technique; it can be applied to any distribution because it ignores characteristics not common to all distributions.

For particular problems additional tests, perhaps as general as this one, can usually be devised. (In applying them the caveats of Section I must be kept firmly in mind and 'cruel and unusual' tests avoided, for sufficient multiplication of tests is bound sooner or later to produce a 'significant' result.) It would be futile to attempt a detailed discussion of supplementary tests except in relation to some specific problem and body of data, but two suggestions will indicate the nature of the possibilities.

First, the magnitudes of the differences between successive observations may be taken into account, without introducing assumptions about the population form, by comparing the variance of the differences with the variance of the N observations, regarded as a finite population. If this is done in rank form, it is very nearly equivalent to testing the serial rank correlation coefficient, for none of the $N - 1$ differences that determine the variance of the differences can differ by more than one from the $N - 1$ differences by which the serial rank correlation is determined. (If the original data are used, rather than ranks, it is equivalent to testing a serial correlation coefficient from a finite population.) In the case of sweetpotato production, the serial rank correlation coefficient is $+.91$, a highly significant value. That the rank correlation coefficient suggests non-randomness whereas χ_p^2 suggests randomness is due to its greater sensitivity to the kind of trend present in these data: whether this is an advantage or disadvantage depends upon the problem (or, to express it differently, upon the nature of the alternative hypotheses).

Second, the signs of the observations (instead of the signs of their first differences, which enter the χ_p^2 test) can be tested for non-random sequences. When testing the six-year moving average, for example, we might determine whether there are non-random sequences in the signs of the residuals. If it is assumed that each residual is equally likely to be positive or negative, the expected number of completed sequences of like sign d years in duration is

$$\frac{N - d - 1}{2^{d+1}},$$

the expected total number of completed sequences is

$$\frac{N - 3}{2},$$

and the probability that a sequence selected at random will be of d years duration is

$$\frac{N - d - 1}{2^d(N - 2)}.$$

These results are easily obtained. A sequence terminal point (initial or final) occurs when a pair of signs is $+\,-$ or $-\,+$, and the probability of this is $1/2$. Since there are $N - 1$ pairs of consecutive signs in a series of N,

$$\frac{N - 1}{2}$$

is the expected number of terminal points. Since there are one fewer sequences than terminal points (note, however, an exception similar to that measured by expression 1, which in this case has a probability of 2^{1-N}) the expected number of sequences is

$$\frac{N-1}{2} - 1 = \frac{N-3}{2}.$$

A completed sequence of d involves d signs of one kind enclosed between two signs of the opposite kind, and the probability of this is

$$2(\tfrac{1}{2})^{d+2} = (\tfrac{1}{2})^{d+1};$$

multiplication by $N - d - 1$, the number of sequences of $d + 2$ in a series of N, yields the expected number of completed sequences of exactly d years. Such a test lacks generality because there is no reason to assume, in the absence of definite information, that positive and negative residuals are equally likely, even if the curve fits adequately; if the probabilities of positive and negative deviations are p and q, where $p + q = 1$, the expected number of completed sequences of d in the signs is $p^2q^2(p^{d-2} + q^{d-2})(N - d - 1)$ and the expected total number of completed sequences is $2pq(N - 1) - 1$.

TABLE 11

Frequency Distribution of Sequences of Like Sign in Residuals from Six-Year Moving Average of Sweetpotato Acreage Harvested United States, 1868–1937

DURATION OF SEQUENCE	FREQUENCY		CONTRIBUTION TO χ^2
	Expected	Observed	
(years)	($N = 64$)		
1	15.5000	7	4.6613
2	7.6250	4	1.7234
3	3.7500	8	4.8167
4	1.8438	2	0.0132
over 4	1.7812	2	0.0269
Total	30.5000	23	
χ^2			11.241
$P(\chi^2)$			.02

Table 11 shows the observed distribution of sequences in sign for the residuals from the six-year moving average of acreage harvested, and that expected if positive and negative deviations are equally probable, a reasonable assumption in this case. Since χ^2 is 11.241, based on 4 degrees of freedom, the probability of so great a divergence from expectation, were chance alone operating, is only .02. It is, of course, a well known characteristic of moving averages, particularly when the weighting is uniform or nearly so, that they tend to lie below the observations during certain types of movement and above during others. This condition can be improved by alterations in the weight diagram.

Still another consideration to bear in mind when interpreting the test is that a set of phase durations that appears random when viewed only as a frequency distribution may not have occurred in a random sequence. To test this, the theory of runs[38] could be applied to the series of phase durations, regarded as a series composed of three kinds of elements (though the serial correlation among phases may be sufficient to vitiate such a test).

An additional point, obvious yet none the less worthy of mention, is that the time unit may affect conclusions derived from the χ_p^2 test. For example, year to year movements in pig iron production, 1877–1936,[35] show a χ_p^2 of 3.60, corresponding with a probability of .2, whereas month to month movements show a χ_p^2 of 372.11, corresponding with an extraordinarily minute probability (Table 12). To put the point more generally: for certain types of continuous function whether the ordinates are correlated serially over a given interval depends upon the frequency with which they are recorded over that interval. Ordinates recorded frequently may be highly correlated serially, while those recorded infrequently may be entirely uncorrelated; but clearly many other types of result are possible.

TABLE 12

Frequency Distributions of Phase Durations in Pig Iron Production, Annually and Monthly[a] United States, 1877–1936[35]

DURATION OF PHASE	ANNUALLY Expected Frequency	ANNUALLY Observed Frequency	MONTHLY Expected Frequency	MONTHLY Observed Frequency
1	23.750000	16	298.750000	43
2	10.266667	7	131.266667	38
3	2.902778	2	37.736111	29
4	0.621429	0	8.216667	25
5	0.107788	0	1.450050	19
6	0.015763	1	0.215829	15
7	0.001996	1	0.027822	4
8	0.000223	0	0.003166	9
9	0.000022	0	0.000323	5
11	0.000000	0	0.000003	3
12	0.000000	0	0.000000	2
18	0.000000	0	0.000000	1
Total	37.666667	27	477.666667	193
χ_p^2 [b]	3.6019		372.1092	
$P(\chi_p^2)$	.21		.0000	
$P(p)$ [c]	.002		.0000	

[a] The annual figures are calendar year totals of the monthly figures.
[b] Computed by combining all durations in excess of 2.
[c] Probability for total number of phases; see Sec. VII.

An entirely different use for the χ_p^2 test, not pertaining especially to time series, may be illustrated by the homogeneity test of Table 4. If the three distributions are really homogeneous, the total value of χ^2 should be apportioned among the rows at random. Whether it is may be tested by determining whether the 24 contributions to χ^2 constitute a random series. χ_p^2 is found to be 1.0355, corresponding with a probability of .64 (Table 13). This test is entirely independent of the homogeneity test, since that test would be unaffected by any rearrangement of the rows while this one depends only on the arrangement of the rows. It therefore salvages information on order neglected in a χ^2 test of homogeneity or goodness of fit. The result of the χ_p^2 test could be fused with the result of the χ^2 test by R. A. Fisher's method[22] to obtain a single probability based upon both types of information, but there is no point in doing so in this particular case, since the homogeneity test of the 24 classes shows a probability so low that no other single test based on the same classes, however high its probability, could alter the inference of non-homogeneity. A test based on sequences of like sign in the differences from the mean contribution, similar to that shown in Table 11, is an alternative device for salvaging the order information that the χ^2 test disregards, provided the expectations are large enough to eliminate skewness.

TABLE 13

Frequency Distribution of Phase Durations in the Series of 24 Contributions for the χ^2 Test of Homogeneity in Table 4

DURATION OF PHASE	FREQUENCY	
	Expected	Observed
1	8.7500	6
2	3.6667	3
over 2	1.2500	1
Total	13.6667	10
χ_p^2	1.0355	
$P(\chi_p^2)$	.64	
$P(p)$[a]	.11	

[a] Probability for total number of phases; see Sec. VII.

χ_p^2 can also be used to test the independence of two variates, and in some circumstances is superior for this purpose to the rank correlation coefficient. The procedure is to arrange the pairs according to the order of magnitude of one variate and tabulate the distribution of phase durations in the other variate. If the two series are independent, the resulting value of χ_p^2 will not be significant. This test is likely to be more sensitive than the rank correlation coefficient when the relation between the two variates is not monotonic. Suppose, e.g., that arrangement of 15 pairs in ascending order according to one variate produces the following sequence in the second: 1, 3, 4, 8, 10, 11, 13, 15, 14, 12, 9, 7, 6, 5, 2. The value of χ_p^2 is 7.6667 (since there are no completed phases, it equals the expected total number of phases), indicating a significant relation between the two variates (Table 5). The rank correlation coefficient, on the other hand, is only .089, definitely not significant. A difficulty with the χ_p^2 test in this use, however, is that the conclusion occasionally depends upon which variate is chosen for arranging in order and which for counting the phase durations. If the numbers given above are arranged in order, and a tabulation made of the phase durations in the numbers of their positions as now listed, χ_p^2 is found to be only 0.1886, an unusually low value. While the two values of χ_p^2 that can be obtained from a single set of paired variates are not independent, a little experimentation will show that almost any pair of values is possible simultaneously. If the two variates are really independent, of course, neither value of χ_p^2 should be significant, but their interdependence makes it difficult to use both validly.*

* Jacob Wolfowitz points out that this ambiguity can be avoided by basing a test on sequences of consecutive ranks instead of on sequences in direction of movement. In the illustration above, for example, there are three runs of two consecutive numbers (3, 4; 10, 11; and 15, 14) and one run of three (7, 6, 5); this is also true when the listed numbers are arranged in order and the runs counted in the numbers indicating the positions as now listed (the runs then are 2, 3; 5, 6; 9, 8; and 14, 13, 12).

Mr. Wolfowitz derived this test of the independence of two series from a new criterion, somewhat analogous to the likelihood ratio, which he has devised for the choice of tests of significance when nothing is known about the form of the population. It is to be hoped that this important work will reach an early fruition and become generally available. In this connection, see the third from last paragraph of Sec. 1.

IX Historical Note

Investigations of the topic treated in this paper started at least as early as 1874. A brief mention of them may be of interest, since they have appeared in places not generally familiar to economic statisticians. Most of the writers seem to have been aware of few of their predecessors, and the present authors, unfortunately, were aware of none until their own work was in final form. They have no reason to suppose that the following account is complete—indeed, it would be surprising if the classical probability writers had entirely overlooked the problem.

Three statistical papers have suggested basing a test of significance on the frequency distribution of phase durations. The first was published by Louis Besson in 1920,[13] the second by R. A. Fisher in 1926,[21] and the third by W. O. Kermack and A. G. McKendrick in 1937.[33] Only the last seriously investigated the suggestion.

Besson is the only writer to obtain formula 2 (Section III), giving the distribution of phase durations as a function of N. He derived it for two special cases: (1) a discontinuous rectangular distribution in which the number of possible values is much larger than the number of items in the sample, and (2) a normal distribution; and he implied that the same formula had appeared in other instances. He did not, however, realize that it represents a completely general solution, although he apparently suspected as much. "Our formulae are exact," he writes, "in both the very different cases where all values are equally probable and where they follow the law of Gauss, as well as in still different cases. Hence it cannot be doubted that they possess a very great degree of generality, and no matter what law which the quantities occurring in meteorological applications might follow, the application of these formulae would not lead us into serious error; as a matter of fact such quantities usually follow the law of Gauss quite closely." Besson did not discuss the problem of determining whether differences between observed and expected frequencies are statistically significant. Although two papers[19, 29] have referred to Besson's formula, no use has been made of it, as far as we know.

R. A. Fisher[21] gave the limiting form of expression 3 (Section III) for large values of N in the form

$$3\left[\frac{1}{(d+1)!}-\frac{2}{(d+2)!}+\frac{1}{(d+3)!}\right]$$

and also the limiting value of the mean duration, variance, and third moment about the mean (Section IV). Since Fisher's object was to express the probability of a rise or fall at a given point as a function of its distance from the preceding turning point, he was interested only in infinite series and did not take account of the effect of N. Apparently he realized the generality of the formula, however, and he suggested basing a test of significance upon it: "The extreme rarity of runs of 5, 6, or 7 differences is of value in the use of such runs as evidence that a sequence is in parts not of a random character; such a test may be refined by counting all the runs of all lengths and comparing the frequency of each class observed with that predicted by the above distribution." He did not indicate how such a comparison might be made. Fisher's results seem to have been unnoticed, except for a single citation.[46] His contribution was elicited by a note by Bilham[16] showing the probability of three observations' defining a turning point to be 2/3.

The most important of the three statistical contributions is that of Kermack and McKendrick.[33] They derived formula 3 (Section III), but only in the limiting form

$$\frac{3(d^2 + 3d + 1)}{(d + 3)!}$$

and also the expected mean duration of a phase. In comparing the observed with the expected distributions, however, they ignored the interrelation among the phases in a single series and assumed that χ_p^2 is distributed as χ^2. They also gave the mean and variance of the total number of phases as functions of N. There are several striking similarities between their paper and ours: they pointed out the insensitivity to trend or slow periodic movements (they added a suggestion that testing a sequence made up of every k-th observation will increase the sensitivity in this respect); they wrote, "One obvious limitation of these criteria is that as they make use only of qualitative relationships and do not take into account the exact magnitude of the observations, they do not make full use of all the available information. It is to be noticed, however, that there is the compensating advantage that the criteria make no assumption whatever about the law of distribution of the observations, apart from the very general one that they are unequal"; and, as a final coincidence, their paper was partly financed by Carnegie funds. Their method of handling ties is one that we have recommended elsewhere[56] because of its simplicity, but since it assumes the true differences between tied observations (instead of the true values of the observations themselves) to be a random sequence, it is not as strictly correct as the one described in the second paragraph of Section VIII; the two procedures are identical, however, in the most common case, that where only two adjacent observations are equal, and they differ little unless there are many sequences of more than two ties or a few unusually long sequences.

There are also three principal mathematical treatments of the topic: a long paper by Kermack and McKendrick published in 1937[34] and two in treatises on combinatory analysis, Macmahon's of 1916[36] and Netto's of 1901.[39]

Kermack and McKendrick demonstrated the formulas utilized in their statistical paper,[33] and extended these in several directions not especially related to the present paper. At one point they used, in effect, the probability transformation which we explain in Section II. They simply stated that the observations can be regarded as all between 0 and 1 because the distribution of phase durations remains invariant under any one to one transformation; they did not, however, explicitly introduce a transformation to a uniform distribution, though this is implicit in the integrations they make.

Macmahon[36] does not consider phases in our sense, but he partitions a series into groups which are, as he points out, equivalent. He divides the observations in such a way that each group contains an ascending sequence. All observations in what we call an expansion of d, including both bounding turning points, thus constitute a single group of $d + 1$; each observation in what we call a contraction, excluding the bounding turning points, constitutes a group of 1. For example, he divides the sequence 8, 6, 7, 2, 9, 5, 4, 1, 3 as follows: 8; 6, 7; 2, 9; 5; 4; 1, 3 and treats it as a sequence of groups of 1, 2, 2, 1, 1, 2, in that order. (Given a sequence of groups as defined by Macmahon, it is easy to deduce the sequence of phase durations according to our definition.) Letting $a, b, c, \cdots$ represent the group sizes in the order of their appearance, where $a + b + c + \cdots = N$, Macmahon shows that the proportion of the permutations of N different observations which produce the sequence of groups a, b, $c, \cdots$ is given by the determinant

$$\begin{vmatrix} \frac{1}{a!} & \frac{1}{(a+b)!} & \frac{1}{(a+b+c)!} & \cdots \\ 1 & \frac{1}{b!} & \frac{1}{(b+c)!} & \cdots \\ 0 & 1 & \frac{1}{c!} & \cdots \\ \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdots \end{vmatrix}$$

He gives several theorems useful in manipulating this expression. (Compare equation 4, Section VI.)

Netto[39] summarizes a paper by Bienaymé[14] and a long series by André published between 1879 and 1896.[1–11] Bienaymé in 1874 stated that the number of phases, complete and incomplete, is normally distributed about a mean of

$$\frac{2N - 1}{3}$$

with variance of

$$\frac{16N - 29}{90}.$$

His paper actually reads: "le nombre des maxima et des minima sera probablement égal à

$$\frac{2N - 1}{3} \pm t\sqrt{\frac{16N - 29}{45}}$$

a probabilité correspondant à t etant donnée approximativement par l'integrale bien connue

$$\frac{2}{\sqrt{\pi}}\int_0^t e^{-x^2}\,dx."$$

A later paper[15] describes

$$\frac{2N - 1}{3}$$

as "le nombre des maxima et des minima, ou des séquences," meaning by a sequence what we call a phase, including incomplete phases; and in the second sense it is correct. The appearance of 45 instead by 90 is explained by the form in which Bienaymé writes the normal distribution. André's chief contribution was a recursion formula for the number of phases (including the incomplete phases before the first and after the last turning points):

$$f_N(p) = pf_{N-1}(p) + 2f_{N-1}(p - 1) + (N - p)f_{N-1}(p - 2)$$

where $f_N(p)$ represents the number of the permutations of N different numbers producing p phases. (Compare equations 5 and 5a, Section VII.)

None of the works we have seen has investigated the central problem of how to test the significance of the difference between the observed and expected distributions. Jones,[29] in referring to Besson's work,[13] cautions "against the use of the χ^2 test for testing the significance of these distributions since the *total* frequency of the observed and expected number of runs is not necessarily the same," but we are unable to see the point to this warning; the fact that the totals are free to vary seems simply to remove one linear constraint and so to allow one more degree of freedom for sampling fluctuations, though the χ^2 distribution is inapplicable for other reasons (see Section VI). Some of the investigations, particularly those by Macmahon[36] and by Kermack and McKendrick,[34] may prove valuable to a future researcher who carries out our suggestion of a thorough mathematical investigation into the question. None of the writers has considered any except direct applications of the technique to original data; that is, they have not considered its use with derived series which should be random according to the assumptions of the method of derivation.

X Summary

A simple and economical test of significance for time series (and other data in which the order of appearance is essential), which makes no assumption about the fundamental probability distribution, may be based on the frequency distribution of sequences of like sign in the first differences.

In a series of N independent random observations the expected number of completed runs of d in the signs of the first differences is

$$\frac{2(d^2 + 3d + 1)(N - d - 2)}{(d + 3)!}.$$

As the size of the sample increases, the proportion of runs of one approaches 5/8, the proportion of runs of two approaches 11/40, the proportion of runs longer than two approaches 1/10; and the average length of run approaches 1.5.

The expected number of runs of one is, then,

$$\frac{5(N - 3)}{12},$$

of runs of two,

$$\frac{11(N - 4)}{60},$$

and of runs longer than two,

$$\frac{4N - 21}{60}.$$

These three expectations may be compared with the observed frequencies by the usual method of summing the ratios of the squared deviations to the expectations. The sum is essentially similar to χ^2 for two degrees of freedom, but is denoted by χ_p^2 because its sampling distribution differs somewhat from that of χ^2. The tail of the distribution of χ_p^2, i.e., χ_p^2 above about 6.3 or P below about 1/15, is well described by the χ^2 distribution for $2\frac{1}{2}$ degrees of freedom, the .05 level of which falls at $\chi_p^2 = 6.898$ (Table 5); the main body of the distribution is covered by referring $\frac{6}{7}\chi_p^2$ to the usual χ^2 tables for 2 degrees of freedom. Although these empirical distributions seem adequate for practical work, a rigorous derivation of the true sampling distribution is much to be desired.

Since this test is, in general, not sensitive to the existence of a primary trend, it is especially useful in determining the presence of secondary components of the systematic variation, especially 'cyclical' fluctuations. It is useful also as an objective test of goodness of fit of smooth curves, particularly for curves that have not been fitted by mathematically efficient methods, e.g., freehand curves or moving averages. It also provides a criterion of the number of terms to be used in smoothing by moving averages. Still other uses, not pertaining especially to time series, are in salvaging the order information neglected in a χ^2 test of homogeneity and in detecting the existence of correlation.

A simpler test of the same nature may be based on the total number of completed runs in the signs of the first differences, since this is normally distributed with mean of

$$\frac{2N - 7}{3}$$

and variance of

$$\frac{16N - 29}{90}.$$

References

The following references (except the twelfth) are cited by number in superscripts to the text, footnotes, or tables. We have not been able to examine the eleventh entry. We are indebted to Herbert E. Jones for lending us the complete manuscript of the theoretical portions of his paper;[29] to Walter A. Shewhart for giving us a copy of his manuscript,[46] through which we learned of the work of Kermack and McKendrick[33,34] (and through that of the work of André,[5,10] Macmahon,[35] and Netto[39]); and to Frederick F. Stephan for directing our attention to Fisher's paper.[21]

1 André, Desiré, 'Développements de séc x et de tang x', *Comptes Rendus de l'Académie des Sciences* (Paris), 88 (1879), pp. 965–7

2 —— 'Sur les permutations alternées', *Journal de Mathématiques pures et appliquées*, Series 3, 7 (1881), pp. 167–84

3 —— 'Sur le nombre des permutations de n éléments qui présentent s séquences', *Comptes Rendus de l'Académie des Sciences* (Paris), 97 (1883), pp. 1356–8

4 —— 'Probabilité pour qu'une permutation donnée de n lettres soit une permutation alternée', *Comptes Rendus de l'Académie des Sciences* (Paris), 97 (1883), pp. 983–4

5 —— 'Étude sur les maxima, minima et séquences des permutations', *Annales scientifiques de l'École Normale supérieure* (Paris), Series 3, 1 (1884), pp. 121–34

6 —— 'Sur le partage en quatre groupes des permutations des n premiers nombres', *Comptes Rendus de l'Académie des Sciences* (Paris), 115 (1892), pp. 872–4

7 —— 'Note', *Bulletin de la Societé Mathématique de France*, 21 (1893), p. 131

8 —— 'Sur le triangle des séquences', *Comptes Rendus de l'Académie des Sciences* (Paris), 118 (1894), pp. 575–8, 726, 1026–8

9 —— 'Sur les permutations quasi alternées', *Comptes Rendus de l'Académie des Sciences* (Paris), 119 (1894), pp. 947–9

10 —— 'Sur les permutations quasi-alternées', *Journal de Mathématiques pures et appliquées*, Series 5, 1 (1895), pp. 315–50

11 —— in *Bulletin de la Société Philomathique de Paris*, Series 8, 5 (1893), p. 33, and 8 (1896), p. 5

12 Bertrand, J., 'Note relative au théorème de M. Bienaymé', *Comptes Rendus de l'Académie des Sciences* (Paris), 81 (1875), pp. 458, 491–2

13 Besson, Louis, (translated and abridged by Edgar W. Woolard), 'On the Comparison of Meteorological Data with Chance Results', *Monthly Weather Review*, 48 (1920), pp. 89–94

14 Bienaymé, J., 'Sur une question de probabilités', *Bulletin de la Société Mathématique de France*, 2 (1874), pp. 153–4

15 —— 'Application d'un théorème nouveau du Calcul des probabilités', *Comptes Rendus de l'Académie des Sciences* (Paris), 81 (1875), pp. 417–23

16 Bilham, E. G., 'Correlation Coefficients', *Quarterly Journal of the Royal Meteorological Society*, 52 (1926), p. 172

17 British Association for the Advancement of Science, *Mathematical Tables* (London, 1931), I, Table XI, pp. 42–6

18 Burns, Arthur F., *Production Trends in the United States since 1870* (National Bureau of Economic Research, 1934), pp. 44–5
19 Cowles, Alfred and Jones, Herbert E., 'Some *a Posteriori* Probabilities in Stock Market Action', *Econometrica*, 5 (1937), pp. 280–94
20 Dixon, W. J., 'A Criterion for Testing the Hypothesis that Two Samples are from the Same Population', *Annals of Mathematical Statistics*, 11 (1940), pp. 199–204
21 Fisher, R. A., 'On the Random Sequence', *Quarterly Journal of the Royal Meteorological Society*, 52 (1926), p. 250
22 —— *Statistical Methods for Research Workers* (Edinburgh and London, 1925–38), Sec. 14, 16, 18, 20, 21.1
23 —— *The Design of Experiments* (Edinburgh and London, 1935, 1937), Sec. 21
24 —— and Yates, F., *Statistical Tables for Biological, Agricultural, and Medical Research* (Edinburgh and London, 1938), Table XXXIII, pp. 82–7
25 Friedman, Milton, 'The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance', *Journal of the American Statistical Association*, 32 (1937), pp. 675–701
26 —— 'A Comparison of Alternative Tests of Significance for the Problem of *m* Rankings', *Annals of Mathematical Statistics*, 11 (1940), pp. 86–92
27 Hoel, Paul G., 'On the Chi-square Distribution for Small Samples', *Annals of Mathematical Statistics*, 9 (1938), pp. 158–65
28 Hotelling, Harold and Pabst, Margaret Richards, 'Rank Correlation and Tests of Significance involving no Assumption of Normality', *Annals of Mathematical Statistics*, 7 (1936), pp. 29–43
29 Jones, Herbert E., 'The Theory of Runs as Applied to Time Series', summarized in *Report of the Third Annual Research Conference on Economics and Statistics*, (Cowles Commission for Research in Economics, 1937) pp. 33–6
30 Kendall, M. G., 'A New Measure of Rank Correlation', *Biometrika*, 30 (1938), pp. 81–93
31 —— Kendall, Sheila F. H., and Smith, B. Babington, 'The Distribution of Spearman's Coefficient of Rank Correlation in a Universe in which all Rankings Occur an Equal Number of Times', *Biometrika*, 30 (1939), pp. 251–73
32 —— and Smith, B. Babington, 'The Problem of *m* Rankings', *Annals of Mathematical Statistics*, 10 (1939), pp. 275–87
33 Kermack, W. O., and McKendrick, A. G., 'Tests for Randomness in a Series of Observations', *Proceedings of the Royal Society of Edinburgh*, 57 (1937), pp. 228–40
34 —— 'Some Distributions associated with a Randomly Arranged Set of Numbers', *Proceedings of the Royal Society of Edinburgh*, 57 (1937), pp. 332–76
35 Macaulay, Frederick R., *The Movements of Interest Rates, Bond Yields and Stock Prices in the United States since 1856* (National Bureau of Economic Research, 1938), Table 27, col. 4
36 MacMahon, Percy A., *Combinatory Analysis*, (Cambridge, England, 1915), I, 187–216
37 Mitchell, Wesley C., and Burns, Arthur F., 'The National Bureau's Measures of Cyclical Behavior', National Bureau of Economic Research *Bulletin 57* (July 1, 1935)
38 Mood, A. M., 'The Distribution Theory of Runs', *Annals of Mathematical Statistics*, 11 (1940), pp. 367–92
39 Netto, Eugen, *Lehrbuch der Combinatorik*, (Leipzig, 1901), pp. 106–16

40 Neyman, J., 'On Statistics the Distribution of which is Independent of the Parameters Involved in the Original Probability Law of the Variables', *Statistical Research Memoirs*, 2 (1938), pp. 58–9
41 —— and Pearson, E. S., 'On the Problem of the Most Efficient Tests of Statistical Hypothesis', *Philosophical Transactions of the Royal Society of London*, Series A, 231 (1933), pp. 289–337
42 Olds, E. G., 'Distributions of Sums of Squares of Rank Differences for Small Numbers of Individuals', *Annals of Mathematical Statistics*, 9 (1938), pp. 133–48
43 Pairman, Eleanor, 'Tables of the Digamma and Trigamma Functions', *Tracts for Computers*, (Cambridge, England, 1919)
44 Pearson, Karl, ed., *Tables of the Incomplete Γ-Function* (London, 1922), Table I
45 Pitman, E. J. G., 'Significance Tests which may be Applied to Samples from any Population', *Supplement to the Journal of the Royal Statistical Society* 4 (1937), pp. 119–30, 225–32, and *Biometrika*, 29 (1937), pp. 322–35
46 Shewhart, Walter A., 'Contribution of Statistics to the Science of Engineering', *Proceedings of the Bicentennial Celebration of the University of Pennsylvania* (Philadelphia, in press).
47 Sukhatme, P. V., 'On the Distribution of χ^2 in Samples of the Poisson Series', *Supplement to the Journal of the Royal Statistical Society*, 5 (1938), pp. 75–9
48 Thompson, William R., 'On Confidence Ranges for the Median and Other Expectation Distributions from Populations of Unknown Distribution Form', *Annals of Mathematical Statistics*, 7 (1936), pp. 122–8
49 —— 'Biological Applications of Normal Range and Associated Significance Tests in Ignorance of Original Distribution Forms', *Annals of Mathematical Statistics*, 9 (1938), pp. 281–7
50 U. S. Department of Agriculture, *Agricultural Yearbook, 1925* (Washington, 1926), p. 365
51 —— *Agricultural Statistics, 1939* (Washington, 1939), p. 243
52 Wald, A., and Wolfowitz, J., 'Confidence Limits for Continuous Distribution Functions', *Annals of Mathematical Statistics*, 10 (1939), pp. 105–18
53 ———— 'On a Test Whether Two Samples are from the Same Population', *Annals of Mathematical Statistics*, 11 (1940), pp. 112, 147–62
54 Wallis, W. Allen, 'The Correlation Ratio for Ranked Data', *Journal of the American Statistical Association*, 34 (1939), pp. 533–38
55 —— and Moore, Geoffrey H., 'A Test of Significance for Time Series Analysis', *Proceedings of the Nineteenth Annual Conference of the Pacific Coast Economic Association* (1940), pp. 97–98
56 ———— 'A Significance Test for Time Series Analysis', *Journal of the American Statistical Association*, 36 (1941), pp. 401–9