MISCELLANEA

CONTENTS

RANDOMNESS AND RANDOM SAMPLING NUMBERS

By M. G. KENDALL and B. BABINGTON SMITH

(Development of certain sections of an address on "Randomness" given by M. G. Kendall to the Study Group of the Royal Statistical Society on May 11th, 1937.)

Introduction

Randomness and Probability.

1. IN colloquial speech the word "random" is applied to any method of choice which lacks aim or purpose; and this usage is also found in certain sciences. In statistics, however, the word has a somewhat different and more definite significance, closely related to probability. It appears, in fact, that for statistical purposes the ideas of randomness and probability are inseparable, whether one belongs to the "intuitive" school which regards probability as an undefinable, or to the opposing "frequency" school which seeks to explain it in terms of statistical frequencies.

2. Many writers are content to define randomness in terms of probability, and this appears to be the general practice of followers of the "intuitive" school. J. M. Keynes, for example, states explicitly that randomness "must be defined with reference to probability, not to what will happen 'in the long run'; though these may be two senses of it, corresponding to subjective and objective probability respectively." * In a similar sense Jeffreys writes: "(By random) we mean that every selection of m objects

* *A Treatise on Probability*, 1921, p. 291. Keynes' definition (taking probability as subjective) is "a is taken at random from the class S for the purposes of the propositional function $S(x) . \phi(x)$, relative to evidence h if 'x is a' is irrelevant to the probability $\phi(x)/S(x) . h$."

from the original n is equally probable." * Adherents of the "frequency" theory have an additional course open to them —namely, to define both randomness and probability in terms of frequency. Thus J. A. Venn † at the beginning of a chapter devoted to the concept of randomness says, "The scientific conception involved in the correct use of this term is, I apprehend, nothing more than that of aggregate order and individual irregularity (or apparent irregularity) which has already been described in the preceding chapters"—the chapters in question dealing with statistical series and forming the basis of Venn's theory of probability.

3. From the point of view of a formal development of the subject, the definition of randomness in terms of probability has much to recommend it, inasmuch as attention can then be centred on the latter. But the logical convenience of this course should not be allowed to lead to the neglect of a scrutiny of the concept of randomness itself, particularly in view of the practical importance of being able to judge whether any given method of sampling is random or not.

4. Whatever may be the proper relationship between the two concepts in logic, in a discussion of the technique of random sampling randomness must be considered in its own right. For the purposes of this paper the logical aspect has been relegated to the background. We take randomness and probability to be undefined ideas obeying certain intuitively formulated principles, which will be found sufficient to give results of practical application. At the same time we have every reason to suppose that the approach outlined in this paper can be formalized and put on a logical basis within the framework of existing notions on the subject and we hope to carry out this work on another occasion.

Randomness and Purpose not Antithetical.

5. It may be desirable to emphasize at the outset that statistical randomness cannot be defined in terms of absence of design. For instance, it is not legitimate to define a method of sampling as random if it follows no law of choice. So far as is known, the irrational number π gives a series of digits which are random in a certain sense. Nevertheless this series is produced by a purposive process, and is, in fact, given by a comparatively simple law of formation. The number is not random, but it *may* generate a random series of digits. The suggestion has been made that to require a series of objects to be chosen without following any law is itself to lay down a law of choice. This, we think, is a logical confusion, but the fundamental

* *Scientific Inference*, 1931, p. 24.
† *The Logic of Chance*, 1888, p. 96.

thought behind the remark is sound—namely, that a definition of randomness must be sought elsewhere than in the negation of purpose.

Types of Universe.

6. In discussing the various problems associated with randomness and random sampling, we find it useful to distinguish between the following types of a universe of discourse.

(i) By a finite universe we mean a universe containing a finite number of members. Similarly, by an infinite universe we mean one containing an infinite number of members. In certain sampling problems it is possible to adopt modes of procedure which virtually turn a finite universe into an infinite one. If, for example, on sampling, we constantly replace the members drawn from the universe, the universe is in effect infinite, being continually replenished and never approaching exhaustion.

(ii) A continuous universe is one whose members, considered according to some variable, form a continuous set. Similarly for a discontinuous universe. Any continuous universe must be infinite, but the converse is not true. For example, the universe of positive integers is discontinuous and infinite.

It is generally assumed that continuous universes exist and, in fact, a large part of the theory of statistics depends, if not on the existence of continuous universes, at least on the supposition that many universes can, to a satisfactory degree of approximation, be represented by continuous universes.

A universe may be continuous with regard to one variate and discontinuous with regard to another. For example, the universe of rational numbers between -1 and $+1$ is continuous in regard to magnitude and discontinuous in regard to sign.

(iii) A universe is said to be existent if each member has an actual existence. For example, the universe of inhabitants of the United Kingdom is existent. We may also regard as existent a universe whose members, though not given by enumeration, are specified by a law of formation. The universe of positive integers is a case in point. But the universe of irrational numbers does not necessarily exist, and there are grounds for supposing that in fact it cannot properly be regarded as existent.

In distinction to an existent universe, a universe is said to be *hypothetical* when all its members do not exist. For example, the universe of the throws of a given die is hypothetical, and so is the universe of possible yields of wheat in a given field. We regard the distinction drawn in this paragraph as extremely important in the theory of random selection.

Random Sampling

Random Aggregates.

7. It is in accordance with intuitive ideas on the subject that a set of entities obtained in a random way may be very unrandom in appearance. For example, a hand dealt at random from a pack of cards may consist of thirteen of a suit. It appears, therefore, that we cannot hope to define a random sample in terms of the properties of the sample itself, but only as a member of a class of samples —that is to say, by considering it as a member of a universe of higher type. What is required is a definition of a random *method* of sampling. When this is obtained a random sample can be easily defined as a sample obtained by a random method.

Random Selection from Hypothetical Universes.

8. It appears, in fact, that no aggregate of things can be said to be random unless it be considered as one of a set of such aggregates. This is one case of a general proposition which is fundamental to the frequency theory of probability. The expression " a random set of objects " may mean one of two things: either that the set was obtained by a random method, or that the objects are arranged randomly—that is to say, that their *arrangement* was chosen at random from the set of possible arrangements. In speaking of a random sample, we always mean the former.

9. At this point we rule out of the present discussion the drawing of samples from hypothetical universes. For existent universes we can clearly speak with meaning of selecting a number of members. For hypothetical universes, however, it is more doubtful whether the process of selection is really exercised at all. In sampling the throws of a die, for instance, our only method of obtaining members of the universe is actually to throw the die. To talk of the aggregate so obtained as a random selection of throws—or a selection of any kind—clearly involves certain assumptions about the nature of the event. It is a selection without a choice.

10. It appears that what actually happens in such cases is that we take the event and construct about it by a mental fiction a hypothetical universe consisting of other conceivable ways in which the event could have happened. We then assume that when the event happened it was equally possible for all the other members of the hypothetical universe to have happened too, and that for some reason or other one of these members was chosen to assume reality. This is an assumption of a different character from that met in ordinary sampling from existent universes, although the distinction has not usually been drawn. The problems which arise in this connection are, in fact, of an abstract metaphysical character

verging at times on the theological, and seem to require a good deal of further discussion. In this paper we leave these special problems on one side and confine our attention to the problems of drawing samples from existent universes only. We may, however, *utilize* material drawn from a hypothetical universe, such material forming an existent universe.

Human Bias.

11. It is becoming increasingly evident that sampling left to the discretion of a human individual is not random, although he may be completely unconscious of the existence of bias, or indeed actively endeavouring to avoid it. House-to-house sampling, the sampling of crop yields, even ticket drawing have all been found to give results widely divergent from expectation. Apart from theoretical considerations, there is thus practical evidence to show that it is insufficient to define a random method as one free from purposive selection. The criterion of randomness in selection must be of a more objective kind.

Definition of a Random Method of Selection in Terms of Independence.

12. For the purpose of the discussion we require, at this point, a notion of independence. For the present we take this concept to be undefined, merely noting that it may be expressed in terms of probability. With its aid we may define a random method of selection, applied to the characteristic C of a Universe U, as a method which is independent of C in U.

13. It is important to notice that this definition of random selection relates to a particular characteristic which is under consideration. There is no such thing as a random method of selection *per se*, considered apart from the universe whose members are being selected. A method which would be random for one universe is not necessarily random for another, and even within the same universe a method which is random in respect of one characteristic is not necessarily random in respect of another.

14. This accords with general ideas on the subject. For example, a possible method of sampling inhabitants of a street is to take, say, every tenth house. This may give a random sample, but if every tenth house is a corner house, the sample may, or may not, lose its randomness. To decide this point, we shall have to consider the properties of the universe which are under investigation. If we were inquiring into the proportion of inhabitants with blue eyes, it might be sufficient to take the corner houses, on the assumption that the colour of eyes was independent of geographical location. On the

other hand, if we were sampling for income, the method might fail, since corner houses have, in general, higher rents and rates than others, and we should therefore expect to find them inhabited by people with larger incomes.

15. A practical question of great importance which arises in this connection is: How are we to determine whether a given method is independent of a given characteristic? The answer is that we cannot determine it without doubt, for to do so would require a full knowledge of the universe; and this is almost always in practice denied us, for otherwise there would be no point in a sampling inquiry. The assumption of independence must therefore be made with more or less confidence on *a priori* grounds. It is part of the hypothesis on which our ultimate expression of opinion is based.

The Use of Digits in Random Sampling.

16. Although a random method is necessarily related to a characteristic of the universe, one frequently requires a sample which may be regarded as random in respect of all the characteristics of the universe. It might be thought, at first sight, that it would be impossible to draw a random sample for this purpose, because, in the very nature of selection, it is necessary to sample according to at least one characteristic of the universe; and this particular characteristic can hardly be independent of the method of selection when, in fact, that method is based upon it.

17. In practice this difficulty can most conveniently be overcome by superposing on the universe a new characteristic and sampling according to that characteristic. By far the most serviceable characteristic which can be superposed on an enumerable universe is that of ordinal number. If we can number our universe, the problem of drawing a random sample reduces to that of finding a series of random numbers.

18. One point should be made clear here. If the numbering of the universe is carried out in such a way as to be independent of the other characteristics of the universe, any set of numbers will serve to draw a random sample. The randomness in such a case lies, so to speak, in the allocation of ordinals to the universe, not in deciding which ordinals to select for the sample. But in practice a procedure of this kind is of no value, since it only throws back to the difficulty of numbering the universe "at random." The usual course is to number the universe in any convenient way, related to the characteristics or not, and then seek for a set of numbers which are a random set from the possible ordinals of the universe.

19. This procedure has the additional advantage that it saves a great deal of labour and narrows the field of inquiry enormously.

Instead of having to examine on every occasion whether our sampling method is really independent of the properties of the universe, we can confine ourselves to the single question whether we can construct a set of digits which will provide a random sample of any desired size from any finite set of integers, it being almost certain that the arrangement of digits in the sampling numbers will be independent of the characteristics of the universe. This is the fundamental problem in the construction of sets of " random sampling numbers." As is shown below, it does not admit of a complete solution.

Random Sampling Numbers

20. Any set of digits whatsoever is random in the sense that it might arise in random sampling from a infinite universe of digits; but such a set is not necessarily one which can be used for random sampling, as, for instance, if it consists of a block of zeros. A set of Random Sampling Numbers (by which is meant a set of numbers which can be used for random sampling, not necessarily a set obtained by random methods) must therefore conform to certain requirements other than that of having been chosen at random.

21. If a random method is repeatedly applied to a universe consisting of a finite number of distinct species of object, each species being present in equal and infinite amount, we expect that in a large number of trials each type of object will appear an approximately equal number of times. It follows that in a set of N Random Sampling Numbers we should expect each digit to occur in approximately $N/10$ cases, for the set may be used to draw the ten digits, 0 to 9, of which it is composed. By an extension of this result we should also expect each *pair* of digits to occur an approximately equal number of times, or, put another way, we should expect that no digit would tend to be followed by any other; and so on.

22. As has been stated, any given set of N need not conform to such expectations; but a Random Sampling set should do so. We accordingly call such a set *locally random*, and define it as a set which conforms approximately to expectation among all possible sets of N.

23. To complete the definition it is necessary to make precise the meaning of " approximately " in the statement that certain frequencies are approximately 10 per cent. of the total, etc., in a locally random set. To decide this point we have to consider a given set of N as a member of a universe of sets of N. The proportions in such a universe will give a frequency distribution, and in particular the divergences of expectation from actuality can be combined into a χ^2 distribution. On purely arbitrary grounds we take the χ^2 test to give a measure of permissible divergence from

expectation. That is to say, if we have expectations $\tilde{m}$ and actual frequencies m

$$\chi^2 = S\left\{\frac{(\tilde{m} - m)^2}{m}\right\}$$

$$P = \frac{\int_{\chi}^{\infty} e^{-\frac{\chi^2}{2}} \chi^{\nu - 1} d\chi}{\int_{0}^{\infty} e^{-\frac{\chi^2}{2}} \chi^{\nu - 1} d\chi}$$

where ν is the number of degrees of freedom and the m's will be permissible if, say, $0{\cdot}01 < P < 0{\cdot}99$. In such circumstances the frequencies are " approximately " in accordance with expectation.

Four Tests for Local Randomness.

24. For practical purposes in deciding whether a given set is locally random, we have found that the following four tests are useful and searching. They are, however, not sufficient to establish the existence of local randomness, although they are necessary.

(*a*) The first and most obvious test to be applied is that all the digits shall occur an approximately equal number of times. This test we call the *frequency test*;

(*b*) Secondly, if the series is locally random, no digit shall tend to be followed by any other digit. If therefore we form a bivariate table showing the distribution of pairs of digits in the series, arranged in the rows according to the first digit, and in the columns according to the second digit, we should get frequencies which are approximately equal in all the cells. This test we refer to as the *serial test*;

(*c*) Thirdly, if the digits are arranged in blocks of, say, five, there will be certain expectation of the numbers in which the five digits are all the same, the numbers in which there are four of one kind, and so on. This test we refer to as the *poker test*, from an analogy with the card game;

(*d*) Finally, there are certain expectations in regard to the gaps occuring between the same digits in the series. For instance, if we take one digit, say, zero, in about one-tenth of the cases the first zero will be followed immediately by a second zero, and there will be no gap. In about nine-hundredths of the cases there will be one digit between two zeros. In about eighty-one-thousandths of the cases there will be a gap of two digits between successive zeros, and so on. This we call the *gap test*.

25. These four tests taken together are very powerful. It is comparatively easy to form series that evade the first three. For example, the recurring series, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, evades the frequency test, the series 1–2–3–4–, etc., evades the frequency test, and the serial test if the dashes are filled with random digits. We have, however, not succeeded in constructing a series which would certainly evade the gap test. Such a series would, it appears, have to have a very peculiar bias indeed, such as would hardly ever rise in practice.

26. The gap test may be extended. Not only will there be an expectation of the frequency of the gaps, but there will also be an expectation of the gaps between gaps of the same kind; these in turn will have expected gaps between them, and so on. There is thus an infinite sequence of the gap tests no one of which includes the others. All these tests are necessary for local randomness, though we have not established their sufficiency. It appears, therefore, that there is no finite set of tests of this character which is sufficient to demonstrate the local randomness of all finite sets of numbers.

27. There is clearly something conventional about the permissible limits of local randomness. To make the concept more precise, we should perhaps speak of a series as being locally random in a certain test Domain. In this connection it should be noticed that if two series S and T are each locally random in a test Domain, it does not follow that the sum of the series $S + T$ is also locally random in that Domain. Further, if a series S is locally random in a Domain, it does not follow that any part of S is locally random in that Domain.

28. This last result is very important. A Random Sampling set may be locally random, and therefore satisfactory, when taken as a whole; but a section of it may not. The familiar process of " taking a page " from a book of random numbers therefore requires some further examination.

Consider an extreme case. Suppose we construct a Random Sampling set consisting of $10^{10^{10}}$ digits. The chance that any given block of a million consists of zeros is 10^{-10^6}. Such a set should therefore occur at least $10^{(10^{10}-10^6-6)}$ times, *i.e.*, is practically certain. If it did not occur, we should probably reject the whole set as not being locally random. Clearly, however, this block of a million zeros is not locally random, and if we turned to the pages in the table at which it occurred, we should not be able to use it to draw random samples which only required that block.

29. It follows that a Random Sampling set cannot be used indiscriminately to draw any given number of samples of any given size

from any given universe until it has been tested for local randomness not only as a whole, but in the parts which it is proposed to use separately. It follows further from the previous paragraph that a perfect Random Sampling set is impossible—there can never be compiled a set of Random Sampling Numbers which are adequate for all requirements. A set may be useful over a certain range—for instance, a set of a million might be locally random for sampling which required a batch of ten digits or more; but as the size of the set increases, there are bound to appear bad patches which are, in themselves, not locally random.

Production of Locally Random Series.

30. It has long been held that mechanical methods of producing random series of integers do not give satisfactory results. Dice-throwing, for example, to give a random series of the integers 1 to 6, notoriously results in bias. Nor are roulette tables much better. Karl Pearson has shown * by analysis of the gaming results at Monte Carlo that the odds against the absence of bias are exceedingly large. The source of this bias is not altogether clear, but if we exclude the possibilities of deliberate falsification, it would appear to arise from small imperfections in the roulette wheel which direct the ball into some compartments in preference to others.

31. So far as we are aware, the only tables of random numbers at present in existence are those due to L. H. C. Tippett, who compiled them by taking digits "at random" from census reports. This amounts to an abandonment of the mechanical method in favour of one which may reasonably be supposed to be free from bias. The reliability of such numbers must, however, depend on the results which they give, and it is stated (in the introduction to these tables) that Tippett's numbers have been found by experiment to give results in accordance with expectation. In view of the position in regard to local randomness revealed by the foregoing analysis, it would appear desirable to examine these tables again to determine the ranges within which they are locally random. We refer below to some tests we have ourselves carried out on those numbers.†

32. The selection of digits from apparently random distributions must be made with some caution. We have attempted to construct a random series by selecting digits from the London Telephone Directory. In order to exclude bias as far as possible, pages were taken by opening the book haphazardly; numbers of less than four digits were ignored; numbers associated with names printed in heavy type were also ignored; and only the two right-hand digits were taken.

* *Chances of Death*, Vol. I. See also *Phil. Mag.*, 5th Series, Vol. I, 1900, p. 157.

† See also the note by Mr. Udny Yule following this paper.

It was found that a series of this kind was significantly biased. There appeared a deficiency of fives and nines, as the following table shows :

TABLE I

Distribution of 10,000 *Digits taken Haphazard from the London Telephone Directory*

Digit	0	1	2	3	4	5	6	7	8	9	Total
Frequency	1026	1107	997	966	1075	933	1107	972	964	853	10,000

$\chi^2 = 58{\cdot}582.$ $P =$ less than 0·000,000,1.

The reasons for this effect are complicated, and are not confined to the obvious one that Telephone Engineers would avoid fives and nines because of their assonance. It thus appears that the London Directory is useless as a source of random digits.

The Randomising Machine.

33. We have designed an apparatus which, so far as we have been able to test, appears to give locally random sequences of random digits. Essentially the machine consists of a disc divided into ten equal sections, on which the digits 0 to 9 are inscribed. The disc rotates rapidly at a speed which can, if necessary, be made constant to a high degree of approximation by means of a tuning-fork. The experiment is conducted in a dark room, and the disc is illuminated from time to time by an electric spark or by a flash of a neon lamp, which is of such short duration that the disc appears to be at rest. At each flash a number is chosen from the apparently stationary disc by means of a pointer fixed in space.

In the actual experiment, the disc was rotated by an electric motor at about 250 revolutions per minute. It was illuminated by a neon lamp in parallel with a condenser in an independent electric circuit which was broken by means of a key. Owing to experimental conditions, the time between the making of the circuit and the passing of the flash varied, but to add an extra element of randomness the key was tapped irregularly by the experimenter. Flashes occurred, on the average, about once in three or four seconds.

Test of Results.

34. The first five thousand digits obtained in this way are given in the Appendix.* We have applied to them the four tests previously mentioned with the following results :—

* It should perhaps be pointed out that all these numbers were taken from the machine by the same experimenter. As opportunity permits, we propose to examine the unlikely possibility that other observers would produce non-locally random sets.

TABLE II

Serial Test of the 5,000 Digits of the Appendix

First Digit

Second Digit	0	1	2	3	4	5	6	7	8	9	Total
0	47	61	44	52	44	56	39	39	50	52	484
1	51	34	33	62	45	44	49	49	46	53	466
2	43	56	64	50	45	43	54	42	51	55	503
3	47	46	63	50	68	41	55	50	47	47	514
4	45	49	54	40	46	59	56	51	51	48	499
5	47	41	59	44	44	54	67	58	42	55	511
6	58	43	56	58	58	52	57	44	51	45	522
7	50	48	35	56	41	48	46	42	65	56	487
8	52	49	45	53	58	44	43	49	48	52	493
9	44	39	50	49	50	70	56	63	42	58	521
Total	484	466	503	514	499	511	522	487	493	521	5000

The above table was formed by entering a pair XY in the Xth column and the Yth row; *e.g.*, the first three numbers 231 give rise to a unit in the second column and the third row, and one in the third column and first row. To make the 5,000 digits yield exactly 5,000 pairs, the last digit was taken with the first to form the pair 02.

It will be seen that the total of the Xth row is the same as the total for the Xth column.

Frequency Test.

The frequencies of digits are those given in the marginal row or column. The expectation is 500 in each cell, and the number of degrees of freedom is 9.

We find $\chi^2 = 5{\cdot}76$. $P = 0{\cdot}76$ approx.
and the test is favourable.

Serial Test.

For this test we compare the frequency in each cell of the body of the table with the theoretical frequency 50. It should be noted that the table is different from an ordinary contingency table, although resembling it in form, for the theoretical frequencies are not calculated from the marginal frequencies. The number of degrees of freedom, ν, is 90. There are 10 linear constraints imposed by the condition that totals of corresponding rows and columns are the same, but only 9 of these are independent; and there is one constraint imposed by the total frequency.

We find

$$\chi^2 = 110{\cdot}44,\ \sqrt{2\chi^2} = 14{\cdot}86$$
$$\sqrt{2\nu - 1} = \sqrt{179} = 13{\cdot}38.$$

The deviation of $\sqrt{2\chi^2}$ is 1·48, giving a probability P of 0·07—within permissible limits.

Poker Test.

The expected and observed frequencies in the 1,000 groups of fives are as follows :

Type	Observed Frequency	Expected Frequency
aaaaa	1	0·1
aaaab	6	4·5
aaabb	9	9
aaabc	75	72
aabbc	101	108
aabcd	501	504
abcde	307	302·4
	1,000	1,000·0

Running together the first three frequencies, we find $\chi^2 = 1{\cdot}09$, $\nu = 4$, $P = 0{\cdot}9$ approx., a favourable result.

Gap Test.

The gap test carried out on the zeros of the table gives the following :—

Size of Gap	Observed Frequency	Expected Frequency
0	48	48·3
1	43	43·5
2	38	39·1
3	39	35·2
4	25	31·7
5	28	28·5
6	23	25·7
7	31	23·1
8	22	20·8
9	17	18·7
10	15	16·8
11	17	15·2
12	12	13·6
13	10	12·3
14	10	11·0
15	13	9·9
16–20 inclusive	33	36·7
21–25 „	18	21·6
26 and over	41	31·2
	483	482·9

$\chi^2 = 11{\cdot}22$, $\nu = 18$, $P = 0{\cdot}88$, again favourable.

The set of 5,000 is thus locally random in the fourfold test domain. We may add that the individual sets of 1,000 are also locally random.

35. It might be thought that the pursuit of a mechanical method for giving random series is akin to that of a machine for achieving perpetual motion. No matter how good the machine is and how efficient the technique under which the experiment is conducted, there are always conceivable sources of bias which may deprive the results of randomness. This objection, however, could be lodged against any method of constructing Random Sampling Numbers, and does not in our opinion affect the value of a series obtained mechanically, since we are not looking for a machine to construct a perfectly random series, whatever that may mean, but only one in which the bias, if any, should fall below the threshold of permissible limits imposed by the nature of the tests which we apply—in short, that it should be *locally* random. The series we have obtained appears to conform to these requirements at least for sets of 5,000. Further experiments would be desirable before it can be regarded as established that the machine will give locally random sets of greater number; but the above results are encouraging and provide at least a *prima facie* case.

Test of Tippett's Numbers.

36. We may compare the analysis of the five thousand digits of Appendix I with that of the first thousand and the last thousand of Tippett's numbers, which are as follows :—

TABLE III

Serial Test on Tippett's First Thousand

First Digit

Second Digit	0	1	2	3	4	5	6	7	8	9	Total
0	11	9	8	14	12	13	10	13	2	4	96
1	14	17	9	10	13	8	8	9	10	12	110
2	11	10	5	8	15	10	9	7	10	11	96
3	4	13	8	7	12	4	16	13	10	12	99
4	4	11	12	16	12	10	9	8	8	12	102
5	13	14	6	10	4	8	13	8	10	16	102
6	7	10	14	8	11	15	14	15	7	10	111
7	10	12	13	9	6	13	10	7	9	10	99
8	15	9	8	6	8	8	11	5	8	5	83
9	7	5	13	11	9	13	11	14	9	10	102
Total	**96**	**110**	**96**	**99**	**102**	**102**	**111**	**99**	**83**	**102**	**1000**

For the *frequency test*, $\chi^2 = 5{\cdot}56$, $P =$ about 0·78.
For the *serial test*, $\chi^2 = 100{\cdot}4$, $P =$ about 0·21.
The *poker test* for groups of four gives—

Type	Observed Frequency	Theoretical Frequency
aaaa	0	0·25
aaab	10	9
aabb	9	6·75
aabc	103	108
abcd	128	126
	250	250·00

Running together the first two frequencies, $\chi^2 = 1{\cdot}07$, $\nu = 3$, $P =$ about 0·79.

The *gap test* for the 96 zeros gives—

Size of Gap	Observed Frequency	Theoretical Frequency
0 to 2 (inclusive)	25	26·0
3 „ 5 „	19	19·0
6 „ 8 „	13	13·8
9 „ 11 „	10	10·1
18 and over	28	26·1
	95	95·0

TABLE IV

Serial Test for the Last 1,000 *of Tippett's Numbers*

First Digit

Second Digit	0	1	2	3	4	5	6	7	8	9	Total
0	10	9	12	17	9	9	10	8	10	13	107
1	13	7	11	6	10	5	7	5	12	13	89
2	6	10	16	9	8	15	14	9	8	7	102
3	15	10	10	9	8	15	8	13	9	10	107
4	12	8	10	8	7	8	14	5	14	8	94
5	10	8	9	17	14	9	5	12	10	11	105
6	7	9	8	14	10	10	9	10	16	10	103
7	11	8	7	4	11	8	12	6	9	10	86
8	12	9	11	15	7	15	8	7	7	11	102
9	11	11	8	8	10	11	16	11	7	12	105
Total	107	89	102	107	94	105	103	86	102	105	1000

$\chi^2 = 0{\cdot}2242$, $P = 0{\cdot}994$. This is probably admissible, in view of the small frequency and the approximate nature of the theoretical frequencies.

We conclude that the set of 1,000 is locally random so far as these tests are concerned.

For the frequency test $\chi^2 = 5{\cdot}18$, $P =$ about $0{\cdot}82$.

For the serial text $\chi^2 = 83{\cdot}6$, $P =$ about $0{\cdot}67$.

For the poker test we find, with a similar grouping to that employed for the first thousand, $\chi^2 = 0{\cdot}634$, $P =$ about $0{\cdot}89$.

For the gap test applied to zeros, we find $\chi^2 = 1{\cdot}16$, $P =$ about $0{\cdot}88$.

37. These results indicate that the first and last thousand of Tippett's tables are locally random sets. But the fact that two chosen at random have been found to be locally random affords little evidence that each of the other 39 thousands is locally random; nor does it follow that the whole 41,600 form a locally random set. To apply the serial test to the whole table would require an amount of labour for which we have not the time. In connection with some other experiments we have, however, obtained certain results which have a bearing on the question.

38. In these experiments 2,000 random permutations of the digits 0 to 9 were obtained from Tippett's tables in the following way: the observer went through the tables from the beginning, writing down the digits as they occurred, but omitting those which had occurred already. When nine out of the ten possible digits had been obtained, the last was filled in without reference to the tables, and the observer began again on a new permutation. The whole 2,000 permutations used about 39,000 Tippett digits, *i.e.*, practically the whole table.

In the permutations so obtained the occurrences of the six possible arrangements of certain triads were counted. The results for the first 1,920 permutations with the digits 1, 2, 3 are as follows :—

Permutation	Observed Frequency
1 2 3	302
1 3 2	296
2 1 3	339
2 3 1	327
3 1 2	323
3 2 1	333

The expected frequency in each class is 320, and $\chi^2 = 4{\cdot}65$, $P =$ about $0{\cdot}46$.

So far as this limited test is concerned, the table appears to be locally random.

Necessity for More Extended Tables of Random Numbers.

39. Whether Tippett's tables are locally random or not, there is some need for a more extended locally random set. Tippett's numbers are very limited in size, there being only about forty thousand of them. For many sampling purposes this is sufficient, but if we wish to draw a large number of samples from a large universe Tippett's numbers would, we think, be unsatisfactory.

Suppose, for example, we wished to apply a random series of numbers to draw a thousand random samples of 100 from a universe of one million individuals.* In order to specify the individuals, it is necessary to use six digits. The thousand samples of 100, therefore, require six hundred thousand digits, or a table about fifteen times as large as Tippett's.

40. In practice, a situation of this kind has sometimes been met by using Tippett's tables over and over again with variations, *e.g.*, by going through the tables once taking digits horizontally, then taking digits vertically, or taking every other digit, and so on. Although this procedure might give satisfactory results in some cases, the total set of digits so obtained may very well not be locally random. In fact, any test which uses the same set of digits over and over again must be invalid if the set is used sufficient times.

To prove this result, consider a set N used m times, and consider the gap test applied to the zeros of the table. A gap greater than $2N$ is impossible, and hence the observed frequencies of gaps of $2N$ or over must be zero. But as m increases, the theoretical frequency of these gaps increases, and hence the contribution to χ^2 increases without limit; and as the number of degrees of freedom remains constant, there must come a time when the set mN falls outside the test-domain, however generously the latter is determined.

41. It appears, therefore, that the practice of using a Random Sampling set more than once must be regarded with distrust until it has been shown that the multiple of the set is locally random. This is true whether the set is taken in a different order or not. The labour of demonstration in such a case would be enormous; but until it is carried out the practice will be surrounded by doubt.

Note.—Since the above paper was written we have seen a paper by W. O. Kermack and A. G. McKendrick on " Tests for Randomness in a Series of Numerical Observations " (*Proc. Roy. Soc., Ed.*, vol. 57, 1936–7, Part 3, pp. 228–40). Kermack and McKendrick give a test for periodicities in series of numbers which is distinct in type from those proposed in the foregoing. They apply this test to

* This is not in the least a fantastic case. A problem of the same magnitude would arise if one wished to take a sample of 100,000 from the agricultural holdings over one acre in England and Wales, which number about 400,000.

Tippett's numbers (though apparently only to about 1,500 of them) and to certain telephone numbers, finding it favourable to randomness in both cases. The latter result stands in contrast to our results with London telephone numbers, but Kermack and McKendrick are apparently dealing with a five-figure Scottish exchange.

Appendix

Random Sampling Numbers Produced by the Machine

1st Thousand

23157	54859	01837	25993	76249	70886	95230	36744
05545	55043	10537	43508	90611	83744	10962	21343
14871	60350	32404	36223	50051	00322	11543	80834
38976	74951	94051	75853	78805	90194	32428	71695
97312	61718	99755	30870	94251	25841	54882	10513
11742	69381	44339	30872	32797	33118	22647	06850
43361	28859	11016	45623	93009	00499	43640	74036
93806	20478	38268	04491	55751	18932	58475	52571
49540	13181	08429	84187	69538	29661	77738	09527
36768	72633	37948	21569	41959	68670	45274	83880
07092	52392	24627	12067	06558	45344	67338	45320
43310	01081	44863	80307	52555	16148	89742	94647
61570	06360	06173	63775	63148	95123	35017	46993
31352	83799	10779	18941	31579	76448	62584	86919
57048	86526	27795	93692	90529	56546	35065	32254
09243	44200	68721	07137	30729	75756	09298	27650
97957	35018	40894	88329	52230	82521	22532	61587
93732	59570	43781	98885	56671	66826	95996	44569
72621	11225	00922	68264	35666	59434	71687	58167
61020	74418	45371	20794	95917	37866	99536	19378
97839	85474	33055	91718	45473	54144	22034	23000
89160	97192	22232	90637	35055	45489	88438	16361
25966	88220	62871	79265	02823	52862	84919	54883
81443	31719	05049	54806	74690	07567	65017	16543
11322	54931	42362	34386	08624	97687	46245	23245

2nd Thousand

64755	83885	84122	25920	17696	15655	95045	95947
10302	52289	77436	34430	38112	49067	07348	23328
71017	98495	51308	50374	66591	02887	53765	69149
60012	55605	88410	34879	79655	90169	78800	03666
37330	94656	49161	42802	48274	54755	44553	65090
47869	87001	31591	12273	60626	12822	34691	61212
38040	42737	64167	89578	39323	49324	88434	38706
73508	30908	83054	80078	86669	30295	56460	45336
32623	46474	84061	04324	20628	37319	32356	43969
97591	99549	36630	35106	62069	92975	95320	57734
74012	31955	59790	96982	66224	24015	96749	07589
56754	26457	13351	05014	90966	33674	69096	33488
49800	49908	54831	21998	08528	26372	92923	65026
43584	89647	24878	56670	00221	50193	99591	62377
16653	79664	60325	71301	35742	83636	73058	87229
48502	69055	65322	58748	31446	80237	31252	96367
96765	54692	36316	86230	48296	38352	23816	64094
38923	61550	80357	81784	23444	12463	33992	28128
77958	81694	25225	05587	51073	01070	60218	61961
17928	28065	25586	08771	02641	85064	65796	48170
94036	85978	02318	04499	41054	10531	87431	21596
47460	60479	56230	48417	14372	85167	27558	00368
47856	56088	51992	82439	40644	17170	13463	18288
57616	34653	92298	62018	10375	76515	62986	90756
08300	92704	66752	66610	57188	79107	54222	22013

3rd Thousand

89221	02362	65787	74733	51272	30213	92441	39651
04005	99818	63918	29032	94012	42363	01261	10650
98546	38066	50856	75045	40645	22841	53254	44125
41719	84401	59226	01314	54581	40398	49988	65579
28733	72489	00785	25843	24613	49797	85567	84471
65213	83927	77762	03086	80742	24395	68476	83792
65553	12678	90906	90466	43670	26217	69900	31205
05668	69080	73029	85746	58332	78231	45986	92998
39302	99718	49757	79519	27387	76373	47262	91612
64592	32254	45879	29431	38320	05981	18067	87137
07513	48792	47314	83660	68907	05336	82579	91582
86593	68501	56638	99800	82839	35148	56541	07232
83735	22599	97977	81248	36838	99560	32410	67614
08595	21826	54655	08204	87990	17033	56258	05384
41273	27149	44293	69458	16828	63962	15864	35431
00473	75908	56238	12242	72631	76314	47252	06347
86131	53789	81383	07868	89132	96182	07009	86432
33849	78359	08402	03586	03176	88663	08018	22546
61870	41657	07468	08612	98083	97349	20775	45091
43898	65923	25078	86129	78491	97653	91500	80786
29939	39123	04548	45985	60952	06641	28726	46473
38505	85555	14388	55077	18657	94887	67831	70819
31824	38431	67125	25511	72044	11562	53279	82268
91430	03767	13561	15597	06750	92552	02391	38753
38635	68976	25498	97526	96458	03805	04116	63514

4th Thousand

02490	54122	27944	39364	94239	72074	11679	54082
11967	36469	60627	83701	09253	30208	01385	37482
48256	83465	49699	24079	05403	35154	39613	03136
27246	73080	21481	23536	04881	89977	49484	93071
32532	77265	72430	70722	86529	18457	92657	10011
66757	98955	92375	93431	43204	55825	45443	69265
11266	34545	76505	97746	34668	26999	26742	97516
17872	39142	45561	80146	93137	48924	64257	59284
62561	30365	03408	14754	51798	08133	61010	97730
62796	30779	35497	70501	30105	08133	00997	91970
75510	21771	04339	33660	42757	62223	87565	48468
87439	01691	63517	26590	44437	07217	98706	39032
97742	02621	10748	78803	38337	65226	92149	59051
98811	06001	21571	02875	21828	83912	85188	61624
51264	01852	64607	92553	29004	26695	78583	62998
40239	93376	10419	68610	49120	02941	80035	99317
26936	59186	51667	27645	46329	44681	94190	66647
88502	11716	98299	40974	42394	62200	69094	81646
63499	38093	25593	61995	79867	80569	01023	38374
36379	81206	03317	78710	73828	31083	60509	44091
93801	22322	47479	57017	59334	30647	43061	26660
29856	87120	56311	50053	25365	81265	22414	02431
97720	87931	88265	13050	71017	15177	06957	92919
85237	09105	74601	46377	59938	15647	34177	92753
75746	75268	31727	95773	72364	87324	36879	06802

5th Thousand

29935	06971	63175	52579	10478	89379	61428	21363
15114	07126	51890	77787	75510	13103	42942	48111
03870	43225	10589	87629	22039	94124	38127	65022
79390	39188	40756	45269	65959	20640	14284	22960
30035	06915	79196	54428	64819	52314	48721	81594
29039	99861	28759	79802	68531	39198	38137	24373
78196	08108	24107	49777	09599	43569	84820	94956
15847	85493	91442	91351	80130	73752	21539	10986
36614	62248	49194	97209	92587	92053	41021	80064
40549	54884	91465	43862	35541	44466	88894	74180
40878	08997	14286	09982	90308	78007	51587	16658
10229	49282	41173	31468	59455	18756	08908	06660
15918	76787	30624	25928	44124	25088	31137	71614
13403	18796	49909	94404	64979	41462	18155	98335
66523	94596	74908	90271	10009	98648	17640	68909
91665	36469	68343	17870	25975	04662	21272	50620
67415	87515	08207	73729	73201	57593	96917	69699
76527	96996	23724	33448	63392	32394	60887	90617
19815	47789	74348	17147	10954	34355	81194	54407
25592	53587	76384	72575	84347	68918	05739	57222
55902	45539	63646	31609	95999	82887	40666	66692
02470	58376	79794	22482	42423	96162	47491	17264
18630	53263	13319	97619	35859	12350	14632	87659
89673	38230	16063	92007	59503	38402	76450	33333
62986	67364	06595	17427	84623	14565	82860	57300