CAUTION: THESE RECORDS WILL BE USED FOR OFFICIAL PURPOSES ONLY, DO NOT [redacted] NOR REVEAL CONTENTS [redacted] PERSON(S)

RECORDS CHARGE-OUT

10204

DATE OF REQUEST	SUSPENSE DATE
25 Jan 61	10 Feb 61

FILE OR SERIAL NUMBER AND SUBJECT

From File of Special Consultant (Friedman)
Statistical Methods in Cryptanalysis, Revised Edition.
Register No. 193
Serial No. 1008

	NAME AND EXTENSION OF PERSON REQUESTING FILE	ORGANIZATION, BUILDING, AND ROOM NUMBER	
TO	Mr. William Friedman LI 6-8520	310 2nd. Str., SE, Wash., D. C.	
		DATE RET'ND.	INITIAL HERE
RETURN TO	Mrs. Christian, AG-24, NSA, Ft. Geo. G. Meade, Md.		

INSTRUCTIONS WHEN TRANSFERRING FILE TO ANOTHER PERSON, COMPLETE SELF-ADDRESSED TRANSFER COUPON BELOW, DETACH, STITCH TO BLANK LETTER-SIZE PAPER AND PLACE IN OUT-GOING MAIL SERVICE.

2ND TRANSFER COUPON

10204

TO:

FILE (serial number and subject)

TRANSFERRED TO: (name and extension)

ORGANIZATION, BUILDING, AND ROOM NUMBER

DATE (sig) (ext.)

[redacted]

Register No. 193

WAR DEPARTMENT

OFFICE OF THE CHIEF SIGNAL OFFICER

WASHINGTON

STATISTICAL METHODS IN CRYPTANALYSIS

REVISED EDITION

[redacted]

30 April 1959

This document is declassified by authority
of the Director, National Security Agency.

Paul S. Willard
Colonel, AGC
Adjutant General

[redacted]

Statistical Methods in Cryptanalysis, - Revised Edition

[redacted]rity: IRS, dated 14 December 1948, signed Lt. Col. John A. Geddes,
[redacted]puty Chief, Army Security Agency, to Chief, Research
[redacted]

NO ACCOUNTING NECESSARY

REGISTRATION CANCELED
by
Authority Hqs, ASA ltr dated 27 Feb 46
2d Ind 11 Mar 46, signed:
HAROLD G. HAYES, Col., Signal Corps
Acting Chief, Army Security Agency

~~Confidential~~

Register Nº 193

WAR DEPARTMENT
OFFICE OF THE CHIEF SIGNAL OFFICER
WASHINGTON

STATISTICAL METHODS IN CRYPTANALYSIS

REVISED EDITION

TECHNICAL PAPER

By
SOLOMON KULLBACK, Ph. D.
Associate Cryptanalyst
SIGNAL INTELLIGENCE SERVICE

UNITED STATES
GOVERNMENT PRINTING OFFICE
WASHINGTON: 1938

PREFACE

It is here my pleasant task to acknowledge my indebtedness to Mr. William F. Friedman and to others of my associates in the OCSigO for their encouragement and assistance in the preparation of this book, and to the instructors and students of the Signal Intelligence School for their earnest efforts and cooperation.

In particular I must acknowledge the aid of Mr. Frank B. Rowlett, Dr. A. Sinkov, Lt. L. T. Jones, U. S. C. G., and Capt. H. G. Miller, Signal Corps, in carrying out observational tests of the theories and the numerical computation involved in the preparation of the charts and tables included herein.

S. K.

CONTENTS

STATISTICAL METHODS IN CRYPTANALYSIS, REVISED EDITION

SECTION I

INTRODUCTORY REMARKS

1. **Introduction.**—*a.* An examination of either plain-text or cryptographic text will convince the reader that the occurrences of the various textual elements do not follow a definite rigorous mathematical law.

b. In the solution of a cryptogram the cryptanalyst deals almost exclusively with *uncertainties* as regards the relationships of its textual elements. Accordingly he is concerned with the question: What is the probability of a certain event? Of course, there are certain causative or controlling factors which determine whether or not the event takes place and with sufficient information the answer to the question would be either: "It is certain to occur," or "It is certain not to occur."

c. The mathematical theory of probability and statistics is accordingly of importance to the cryptanalyst since it provides a means for the quantitative analysis of the uncertainties with which he deals. It also provides a means whereby he may study the behavior of groups of symbols and draw conclusions therefrom.

d. It is not very often that statistical analysis alone will enable the cryptanalyst to arrive at the solution of a cryptogram. *Statistical analysis will, however, enable the cryptanalyst to evaluate the desirability of pursuing certain procedures and will indicate the most likely order in which to try various possible steps in solution.*

e. Of fundamental importance in the application of statistical technique to cryptography are the various frequency tables relating to the characteristic frequencies of textual elements of different languages. A number of such tables will be found in section VIII.

f. It must be emphasized here that the methods and procedures to be discussed herein are a means to an end, and not an end in themselves.

2. **Purpose.**—This book has been prepared to provide cryptanalysts with an introduction to certain concepts and methods of the mathematical theory of statistics which are useful in cryptanalysis; and to provide the reader with certain formulas, charts, and tables which have been found to be of assistance in the solution of a variety of cryptanalytic problems.

3. **Arrangement of contents.**—*a.* The book is divided into two parts. In the first part, there are: (1) An exposition of the underlying theory; (2) A presentation of many useful formulas; (3) Procedures for the use of these formulas in the solution of problems; (4) Illustrations and examples.

b. In the second part are charts and tables which will assist in the application of the methods discussed in part 1, and a number of appendixes presenting the mathematical development of formulas presented in the first part. There is also a summary of all the formulas and definitions found throughout the book.

c. In keeping with the purpose as set forth in paragraph 2, no attempt has been made in the exposition of part 1 to present the mathematical analysis underlying the derivation of the formulas discussed.

PART 1

SECTION II

GENERAL CONSIDERATIONS OF PROBABILITY

4. **A priori probability.**—*a.* A complete discussion of the mathematical and philosophical implications involved in a logically rigorous approach to mathematical probability is beyond the purpose of this book. Herein it will suffice to use the following definition of *a priori* probability:

The probability that an event will occur is the ratio of the number of favorable cases to the number of total possible cases, all cases being equally likely to occur. By a favorable case, is meant one which will produce the event in question.

b. The probability for the occurrence of an event is always a positive fraction not exceeding 1. The numbers "1" and "0" are taken to represent certainty, since in those circumstances every case is either favorable or not favorable and will produce the event in question or will not produce the event in question. If the probability that an event will occur is p and the probability that it will not occur is q, then $p+q=1$. (It is certain that the event either will or will not occur.)

c. In cryptography the probability of occurrence of each of the letters of the alphabet in various languages is of interest. It is obviously impossible to apply the preceding definition of *a priori* probability, since that would involve a study of every conceivable message that might be sent. In this case, which illustrates the situation most frequently encountered in practical statistical work, there must be introduced the concept of *statistical* probability.

5. **Statistical probability.**—*a.* The fundamental basis in *statistical* probability is the fact that, for all practical purposes, the difference between the unknown *a priori* probability and the ratio of *observed* favorable cases to the *observed* total number of cases, can be made as small as we please by indefinitely increasing the total number of observed cases.[1] The limit of the ratio of the number of observed favorable cases to the total number of observed cases, as the latter number increases indefinitely, shall be called the probability that the event occurs.[1]

b. Thus, in order to find the probabilities of occurrence for each of the letters of the alphabet, it is necessary to examine a large amount of text. A study of 100,000 letters of English telegraphic text gave the result shown in figure 1. We thus find that the probability for the occurrence of A is 0.07189; for B it is 0.01146; for C it is 0.03345, etc.

c. It is usual to denote the numbers 7,189, 1,146, 3,345, etc. (i. e., the number of observed favorable cases) as the *absolute frequencies*, and the numbers 0.07189, 0.01146, 0.03345, etc. (the ratio of the number of observed favorable cases to the total number of observed cases) as the *relative frequencies*.

[1] See appendix A, p. 148.

Letter	Number of occurrences	Letter	Number of occurrences	Letter	Number of occurrences
A	7, 189	K	353	U	2, 993
B	1, 146	L	3, 549	V	1, 340
C	3, 345	M	2, 534	W	1, 401
D	4, 029	N	7, 558	X	469
E	12, 604	O	7, 408	Y	2, 099
F	2, 994	P	2, 661	Z	101
G	1, 795	Q	318		
H	3, 287	R	8, 256	Total	100, 000
I	7, 572	S	5, 759		
J	198	T	9, 042		

FIGURE 1.

6. **Combinations of probabilities.**—*a.* If an event under investigation is one of several mutually exclusive events, then the probability that it occurs is the sum of the probabilities of occurrence of each of the mutually exclusive events.

Example 1.—What is the probability that any one letter chosen at random from English telegraphic text is a vowel? Since the event in question is one of the mutually exclusive events "finding A, E, I, O, U, Y," the probability sought is $P_v=P_A+P_E+P_I+P_O+P_U+P_Y$ where P_v, P_A, P_E, P_I, P_O, P_U, P_Y, respectively, mean the probability for the occurrence of a vowel, the probability for the occurrence of A, etc. Adding the component probabilities, as found from figure 1, there results $P_v=0.39865$. It may be seen from this that approximately 40 percent of the letters of English telegraphic text are vowels.

b. If the event under study is the simultaneous occurrence of several events, or the successive occurrence of several events, then the probability that it will occur is the *product* of the probabilities of occurrence of the component events, provided the occurrence of one does not effect the occurrence of the others—or, as we shall say, provided the events are independent. Thus, the probability that two letters selected at random from English telegraphic text are vowels, is $0.4\times0.4=0.16$.

SECTION III

STATISTICS

7. **Definitions.**—*a.* By *statistical method* we mean the mathematical treatment of observational data in accordance with the fundamental laws of probability discussed in the preceding section.

b. By a *statistical variate* we mean a variable which may assume a finite or infinite number of different values in accordance with a certain law of probability. The sum of the probabilities corresponding to each of the different values must be one.

Example 2.—The variable θ, where θ is to represent any letter of the alphabet, is a statistical variate since θ will assume the values $A, B, C, \ldots, Z$ with probabilities corresponding to the values in figure 1.

c. In order to be able to study efficiently a mass of data, it is desirable that we be able to compute several numbers which will, to a certain extent, characterize the data and display its important properties.

d. By a *statistic* we mean any number computed from observed data in accordance with certain rules. The following are some of the more common statistics which are used to characterize a mass of data and which there will be occasion to use in the course of this work.

e. (1) The *arithmetic mean* or *average* of a sequence of numbers is the sum of the numbers divided by the number of items.

Example 3.—What is the average of 1, 2, 3, 4, 5? The average is $(1+2+3+4+5)/5=3$.

(2) The *weighted mean* or *average* of a series of numbers is the sum of the product of each number and its weight, divided by the sum of the weights. In general, in the study of observed data, the weight corresponds to the number of observed occurrences; in theoretical discussions, it corresponds to the probability of occurrence. It is usual to omit the adjective "weighted" since this definition reduces to the one first given.

(3) Symbolically we may express the foregoing as follows: If the numbers $x_1, x_2, \ldots, x_n$ have, respectively, the weights $w_1, w_2, \ldots, w_n$ (or occur respectively $w_1, w_2, \ldots, w_n$ times), then the average of $x_1, x_2, \ldots, x_n$ or symbolically $\bar{x}$ (read x bar) is given by

$$\bar{x}=\frac{w_1x_1+w_2x_2+\ldots+w_nx_n}{w_1+w_2+\ldots+w_n}$$

Example 4.—A study of 100 sets of English text, each of 50 letters, yielded the following as the number of occurrences of the letter A per set.

x_i	w_i
1	3
2	26
3	21
4	19
5	15
6	8
7	7
8	1
	100

(i. e., A occurred once in each of three sets; twice in each of 26 sets; three times in each of 21 sets, etc.). The average observed occurrence of A per set of 50 letters is therefore

$$\bar{x}=\frac{(3\times 1)+(26\times 2)+(21\times 3)+(19\times 4)+(15\times 5)+(8\times 6)+(7\times 7)+(1\times 8)}{3+26+21+19+15+8+7+1}$$

$$\bar{x}=374/100=3.74$$

(4) If x is a statistical variate, i. e., if x takes on the values $x_1, x_2, \ldots, x_n$ with the corresponding probabilities $p_1, p_2, \ldots, p_n$, respectively, then the average value of x is $\bar{x}=p_1x_1+p_2x_2+\ldots+p_nx_n$. (In this case the total weight $p_1+p_2+\ldots+p_n=1$).

f. The *mean square* of a series of numbers is the average of the squares of the numbers. Symbolically, if $x_1, x_2, \ldots, x_n$ is a sequence of numbers with corresponding weights $w_1, w_2, \ldots, w_n$, respectively, then

$$\text{mean square } x=\frac{w_1x_1^2+w_2x_2^2+\ldots+w_nx_n^2}{w_1+w_2+\ldots+w_n}$$
$$=f_1x_1^2+f_2x_2^2+\ldots+f_nx_n^2$$
$$\text{where } f_i=w_i/(w_1+w_2+\ldots+w_n) \qquad (i=1, 2, \ldots, n)$$ [2]

In the foregoing w_i $(i=1, 2, \ldots, n)$ is an absolute weight and f_i $(i=1, 2, \ldots, n)$ is a relative weight.

g. Let $x_1, x_2, \ldots, x_n$ be a sequence of numbers whose mean value is $\bar{x}$. The *deviation* of x_i from the mean is $x_i-\bar{x}$. The deviation will be negative, zero, or positive according as x_i is less than, equal to, or greater than $\bar{x}$.

h. The *variance* of a sequence of numbers is the mean square of the deviations from the mean, i. e.,

$$\text{variance}=v=\frac{w_1(x_1-\bar{x})^2+w_2(x_2-\bar{x})^2+\ldots+w_n(x_n-\bar{x})^2}{w_1+w_2+\ldots+w_n}$$
$$=f_1(x_1-\bar{x})^2+f_2(x_2-\bar{x})^2+\ldots+f_n(x_n-\bar{x})^2$$

where the x's, w's, and f's are defined as above.

The *positive square root* of the variance is called the *standard deviation.*

It may be shown that $v=f_1x_1^2+f_2x_2^2+\ldots+f_nx_n^2-(\bar{x})^2=$(Mean square of x) $-$ (square of the mean of x).

[2] The notation $(i=1, 2, \ldots, n)$ is a convenient way of indicating that i is to be replaced by all of the successive values $1, 2, 3, \ldots, n$, in turn.

i. In general, the average of a sequence of numbers is a *central value* about which the numbers tend to cluster; the variance is a measure of the *variation* about this central value.[3]

[3] The weighted sum of the deviations from the mean is not a suitable measure of the variation because it is in all cases equal to zero. The following simple algebra demonstrates this fact:

$$\frac{w_1(x_1-\bar{x})+w_2(x_2-\bar{x})+\ .\ .\ .\ +w_n(x_n-\bar{x})}{w_1+w_2+\ .\ .\ .\ +w_n}$$

$$=\frac{w_1x_1+w_2x_2+\ .\ .\ .\ +w_nx_n}{w_1+w_2+\ .\ .\ .\ +w_n}-\frac{\bar{x}(w_1+w_2+\ .\ .\ .\ +w_n)}{w_1+w_2+\ .\ .\ .\ +w_n}$$

$$=\bar{x}-\bar{x}=0$$

The next simple possible measure of the variation about the mean is the weighted sum of the absolute values of the deviations (the weighted sum of the arithmetical values of the deviations neglecting the sign). Symbolically this would be written as

$$\frac{w_1|x_1-\bar{x}|+w_2|x_2-\bar{x}|+\ .\ .\ .\ +w_n|x_n-\bar{x}|}{w_1+w_2+\ .\ .\ .\ +w_n}$$

However, because of the fact that the variance is more amenable to mathematical treatment and because of its relationship with the theory of least squares and the normal probability distribution the variance rather than the weighted sum of the absolute values of the deviations is the more commonly used measure of variation.

SECTION IV

FREQUENCY DISTRIBUTIONS

8. **Generalities.**—*a.* Some slight experience in cryptanalysis will soon convince one that an outstanding characteristic of the data studied is its variation. The data which are the object of statistical study always display variation in one or more respects.

b. The notion of a collection of data arranged in a *frequency distribution* with respect to one or more characteristics is fundamental in statistical work. If n observations originating from the same set of circumstances are made with respect to a statistical variate, and if the individual observations are arranged with respect to their magnitude, the result is said to form a frequency distribution; to each value of the variate, there corresponds an absolute frequency. In example 4 there is a frequency distribution of 100 observations of the number of occurrences of the letter A per set of 50 letters of English telegraphic text. Subsequent discussion in this section will introduce theoretical frequency distributions in which to each value of the variate will correspond a probability instead of a definite number of occurrences.

c. Frequency distributions may be discontinuous or continuous. In discontinuous distributions the statistical variate may assume a finite or infinite number of discontinuous values. (Values which are separated one from the other by finite quantities.) The distribution of the number of occurrences of the letter A per set of 50 letters given in example 4 page 5 is an illustration of a discontinuous distribution in which the statistical variate (the number of occurrences of the letter A per set) takes on a finite number of values. In continuous distributions the statistical variate may assume *all* possible values within its range of variation. In the latter case the frequency distribution may be expressed by stating the proportion of the data for which the variate is less than a given value or the proportion of the data for which the variate lies between given values.

d. It is presumed that the reader is already acquainted with instances of frequency distributions, e. g., the frequency distribution of single letters, digraphs, etc., of cryptograms.

e. The following is a frequency distribution of the lengths of words in a series of official telegrams; in all 10,000 words were studied.

Number of letters per word	Number of words	Number of letters	Number of letters per word	Number of words	Number of letters
X_i	F_i	X_iF_i	X_i	F_i	X_iF_i
1	390	390	10	288	2, 880
2	1, 028	2, 056	11	163	1, 793
3	1, 369	4, 107	12	86	1, 032
4	1, 745	6, 980	13	25	325
5	1, 457	7, 285	14	23	322
6	1, 169	7, 014	15	4	60
7	1, 039	7, 273			
8	735	5, 880		10, 000	51, 708
9	479	4, 311			

From this it is seen that the average number of letters per word of English telegraphic text is 5.17. For most purposes, assuming this value to be 5 will give a sufficiently accurate approximation. (This is one of the reasons why the arbitrary length of five characters per word has been adopted as standard for code or cipher text.)

f. It is very desirable to be able to characterize by means of a mathematical formula the relationship between the various values that a statistical variate may take, and the corresponding probabilities (or frequencies). Such a formulation simplifies the study of frequency distributions and enables valid judgments about sample distributions to be formed. The study of the possible formulas for frequency distributions has yielded a number of important results.

g. We shall here restrict ourselves to five types of frequency distributions which are of primary importance in cryptography, viz, the *binomial distribution*, the *normal distribution*, the *Poisson distribution*, the *modified Poisson distribution*, and the *multinomial distribution*.

9. **Binomial distribution.**[4]—*a.* The binominal distribution is the first example of a theoretical distribution to be established, and was discovered by Jacob Bernoulli about the end of the seventeenth century. It can be shown that if the probability that an event occurs is p, and the probability that it does not occur is q, $(q=1-p)$, then, if n independent observations are made, the probability that the event occurs exactly 0, 1, 2, . . . , n times is given by the respective term of the expansion of the binomial

$$(9.1)\qquad (q+p)^n=q^n+nq^{n-1}p+\frac{n(n-1)}{1\times 2}q^{n-2}p^2+\frac{n(n-1)(n-2)}{1\times 2\times 3}q^{n-3}p^3+\ .\ .\ .\ +p^n$$

Thus, the probability that the event occurs 0 times in n trials is $P_0=q^n$; the probability that the event occurs exactly one time in n trials is $P_1=nq^{n-1}p$; the probability that the event occurs exactly two times in n trials is $P_2=\frac{n(n-1)}{1\times 2}q^{n-2}p^2$; . . . ; the probability that the event occurs exactly x times (x an integer) in n trials ($x\leqq n$) is

$$P_x=\frac{n(n-1)(n-2)\ .\ .\ .\ (n-x+1)}{1\times 2\times 3\ .\ .\ .\ \times x}q^{n-x}p^x=\frac{n!}{x!(n-x)!}q^{n-x}p^x$$

where $x!$ (read x factorial) is equal to $x(x-1)(x-2)\ .\ .\ .\ 1$.

Example 5.—Using 0.1 as the probability for the occurrence of T in English text, what is the probability that T occurs zero times, exactly one time, exactly two times, . . . , exactly eight times in a set of 100 letters of English text? In this case $p=0.1$, $q=0.9$, $n=100$, so that the desired probabilities are:

that T occurs zero times $(0.9)^{100}=0.0000=P_0$

that T occurs exactly one time $100(0.9)^{99}(0.1)=0.0003=P_1$

that T occurs exactly two times $\frac{100\times 99}{1\times 2}(0.9)^{98}(0.1)^2=0.0016=P_2$

that T occurs exactly three times $\frac{100\times 99\times 98}{1\times 2\times 3}(0.9)^{97}(0.1)^3=0.0059=P_3$

that T occurs exactly four times $\frac{100\times 99\times 98\times 97}{1\times 2\times 3\times 4}(0.9)^{96}(0.1)^4=0.0159=P_4$

that T occurs exactly five times $\frac{100\times 99\times 98\times 97\times 96}{1\times 2\times 3\times 4\times 5}(0.9)^{95}(0.1)^5=0.0339=P_5$

[4] See appendix A, p. 148ff.

that T occurs exactly six times $\frac{100\times99\times98\times97\times96\times95}{1\times2\times3\times4\times5\times6}(0.9)^{94}(0.1)^{6}=0.0596=P_6$

that T occurs exactly seven times $\frac{100\times99\times98\times97\times96\times95\times94}{1\times2\times3\times4\times5\times6\times7}(0.9)^{93}(0.1)^{7}=0.0889=P_7$

that T occurs exactly eight times

$$\frac{100\times99\times98\times97\times96\times95\times94\times93}{1\times2\times3\times4\times5\times6\times7\times8}(0.9)^{92}(0.1)^{8}=0.1148=P_8$$

b. To find the probability that an event, whose possible occurrences are distributed in accordance with the foregoing distribution, occurs at least r times it is merely necessary to add the probabilities that the event occurs exactly $r, r+1, r+2, \ldots, n$ times. If then we use $P(r)$ to represent the probability for at least r occurrences we have

$$P(r)=\sum_{x=r}^{n}\frac{n!}{x!(n-x)!}q^{n-x}p^{x}=\sum_{x=r}^{n}P_x=1-\sum_{x=0}^{r-1}P_x$$ [5]

(The symbol $\sum_{x=r}^{n}$ means the sum of the terms for all integral values of x from r to n inclusive.)

Example 6.—Using 0.1 as the probability for the occurrence of T in English text, what is the probability that T occurs at least six times in a set of 100 letters? In order to find the desired probability it is necessary to subtract from 1 the sum of the probabilities that T occurs exactly 0, 1, . . . , 5 times. Using the values found in example 5, we have

$$\begin{aligned}P(6)&=1-(0.0000+0.0003+0.0016+0.0059+0.0159+0.0339)\\&=1-0.0576=0.9424\end{aligned}$$

c. For a statistical variate which takes on its possible values in accordance with the law of distribution given by the binomial distribution, it may be shown that the mean value$=\mu=np$, the mean square$=\mu_2=n^2p^2+npq$, and the variance $=\sigma^2=npq$. (See Appendix A, p. 148 ff).

Example 7.—Let us take as the probability for the occurrence of A in English text $p=0.072$. Then, the theoretical average value for the number of occurrences of A in a set of 50 letters of English text is $\mu=np=50(0.072)=3.6$; the theoretical value of the mean square of the number of occurrences (μ_2) is $\mu_2=n^2p^2+npq=(50)^2(0.072)^2+50(0.072)(0.928)=12.96+3.34=16.30$; the theoretical value of the variance (σ^2) is $\sigma^2=npq=50(0.072)(0.928)=3.34$. (In general, we shall use Greek letters for theoretical values and Roman letters for the corresponding observed values.)

Example 8.—It will be of interest to compare the theoretical values derived in example 7 with the observed values obtained from the observed occurrences of A in 100 sets of English text of 50 letters each, already considered in example 4. In example 4 it was found that $\bar{x}=3.74$. The mean square of the number of occurrences is given by (see p. 5).

$$m_2=\frac{3\times1^2+26\times2^2+21\times3^2+19\times4^2+15\times5^2+8\times6^2+7\times7^2+1\times8^2}{3+26+21+19+15+8+7+1}=\frac{1670}{100}=16.70$$

To find the variance we use the fact that variance$=$(mean square)$-$(square of mean), or $\sigma^2=\mu_2-\mu^2$. Thus $s^2=16.70-(3.74)^2=16.70-13.99=2.71$.

[5] Since $p+q=1$, (9.1) could be written as $\sum_{x=0}^{n}P_x=1$

A comparison of theoretical and observed values yields

	Theoretical	Observed
Mean (μ)	3.60	3.74
Mean square (μ_2)	16.30	16.70
Variance (σ^2)	3.34	2.71
Standard deviation (σ)	1.83	1.65

d. It should be clear that the values of the observed means of a sequence of samples will also be distributed in accordance with a certain law of distribution not necessarily the same as the law of distribution of the original observations. The distribution of means of samples of N from a population[6] distributed according to the terms of $(q+p)^n$ is given by the corresponding terms of $(q+p)^{nN}$ plotted to $1/N$ times the unit of the original binomial, i. e., the probability that the mean takes the value 0, $1/N$, $2/N$, $3/N$, . . . , nN/N is given by the corresponding term of the expansion of $(q+p)^{nN}$.

e. The mean of the distribution of means is given by np and the variance of the distribution of means is given by $\sigma_{\bar{x}}^2=\frac{npq}{N}$. The latter equation shows us then, that if σ^2 be the variance of a number of observations, the variance of the mean of N such observations is $\sigma_{\bar{x}}^2=\frac{\sigma^2}{N}$. This last result signifies that the sample means will show a smaller variation about the true (or population) mean than will the original observations. More exactly we may say that the mean of N observations is $\sqrt{N}$ times as reliable as any of the N original observations.

f. In order to apply the binomial distribution to numerical cases, it would be desirable that there be available tables giving the values of the several terms of the expansion of $(q+p)^n$ for various values of p and n. Unfortunately, such tables do not exist. However, since there are tables for other distributions, which will provide sufficiently close approximations to the binomial distribution for all our purposes, the lack of tables for the binomial distribution will not greatly inconvenience us.

10. **Normal distribution.**—*a.* In the case of the binomial distribution, we saw that the statistical variate took on only integral values. However, for the distribution now to be considered, such is not the case. A statistical variate is said to be normally distributed when it takes on all values between $-\infty$ (minus infinity) and $+\infty$ (plus infinity) with frequencies such that the logarithm of the frequency at any distance X from the mean of the distribution is less than the frequency at the mean of the distribution by a quantity proportional to X^2. A more precise expression of the foregoing is the following: The statistical variate normally distributed takes on all values between $-\infty$ (minus infinity) and $+\infty$ (plus infinity) in accordance with the following law of probability: The probability that the statistical variate lies between $X-\frac{\epsilon}{2}$ and $X+\frac{\epsilon}{2}$, where ϵ is a very small number is given by

$$p(X,\epsilon)=\frac{\epsilon}{\sigma\sqrt{2\pi}}e^{-(X-\mu)^2/2\sigma^2} \tag{10.1}$$

[6] By population we here mean the idealized aggregate of data from which the sample is supposed to have been drawn by chance.

In the preceding formula there are two parameters,[7] μ and σ. It may be shown that μ and σ^2 are the mean and variance respectively of a statistical variate with the normal law of distribution. (Hence the importance of the mean and variance or standard deviation, since a knowledge of them is all that is necessary *completely to determine the normal law of probability.*) In (10.1) $X-\mu$ is the distance of the observation X from the mean μ and σ measures in the same units the extent to which the individual observations are scattered.

b. For purposes of tabulation, it is usual to treat $((X-\mu)/\sigma)=x$ as the variate and to omit the factor ϵ/σ in (10.1); thus, in part 2 will be found tables giving the values of y for various values of x in accordance with the formula

(10.2) $$y=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

The curve corresponding to the formula (10.2) is the familiar normal probability curve, given in diagram 1 herewith. Geometrically, σ is the distance on either side of the mean (or center) of the steepest points, or points of inflection of the curve.

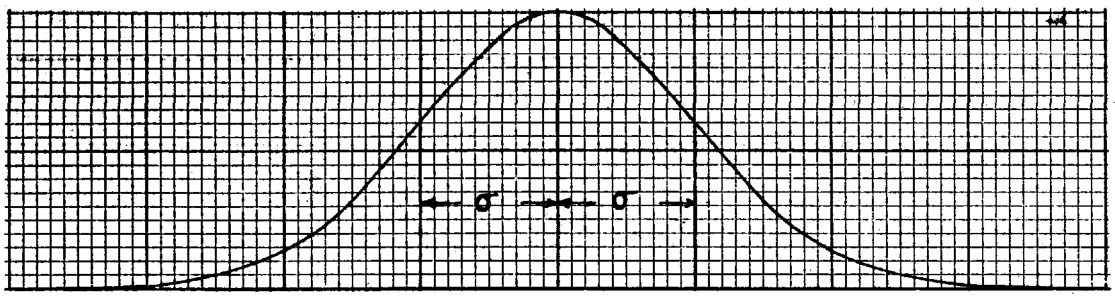


DIAGRAM 1.

Normal Probability Curve: $x=\frac{X-\mu}{\sigma}$

c. In practice it is more often necessary to know the probability, that a statistical variate satisfying the normal law, lies between two values say X_0 and X_1, where $X_1>X_0$. Tables have been calculated to enable this to be done readily. If we set $x_1=(X_1-\mu)/\sigma$ and $x_0=(X_0-\mu)/\sigma$, the desired result is given by[8]

(10.3) $$P(x_0, x_1)=\frac{1}{\sqrt{2\pi}}\int_{x_0}^{x_1} dx e^{-x^2/2}$$

The tables that have been calculated (shown in part 2) are for the value $x_0=-\infty$; that is, the tables give the probability that x is less than or equal to x_1. In order to obtain the result desired, use must then be made of the formula

(10.4) $$P(x_0, x_1)=P(-\infty, x_1)-P(-\infty, x_0)$$

[7] A parameter is a "variable constant" which enters into a mathematical formula. Thus in (10.1) μ and σ are constant for a given population but take on different values for different populations.

[8] The symbol $\int_{x_0}^{x_1}$ (read the integral from x_0 to x_1) may be traced back to the S of the word Sum. In essence the integral is the limit of the sum of the values of the integrand (the expression to be integrated) as x takes on values, between x_0 and x_1, which differ by smaller and smaller amounts. Thus the discussion in paragraph 10*c* is conceptually similar to the discussion in paragraph 9*c*.

A graphic description of the above will help clarify the matter. Assuming the total area under the curve to be unity, then the shaded area in diagram 2 is that which is desired in accordance with (10.3).

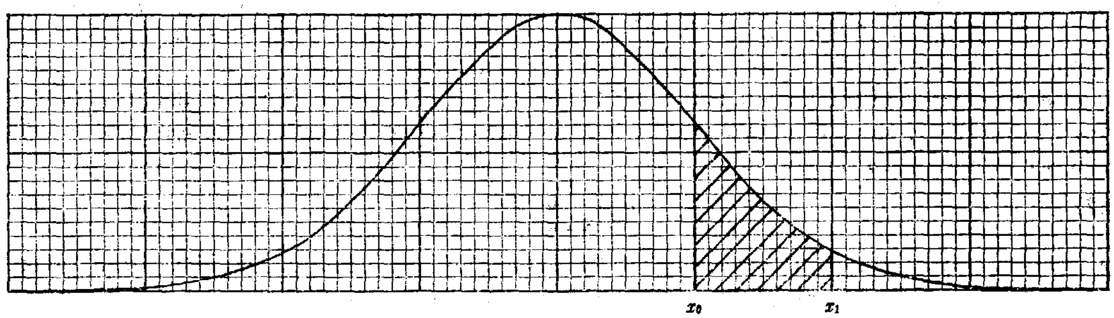


DIAGRAM 2.

The values that have been tabulated correspond to the shaded areas in diagrams 3*a* and 3*b*.

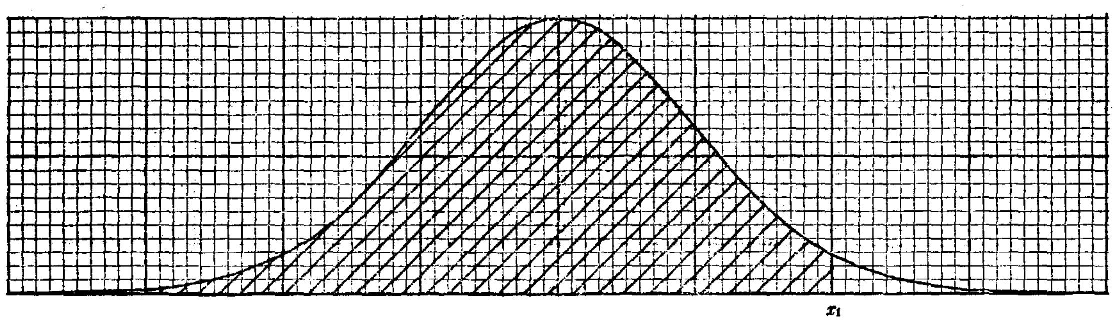


DIAGRAM 3*a*.

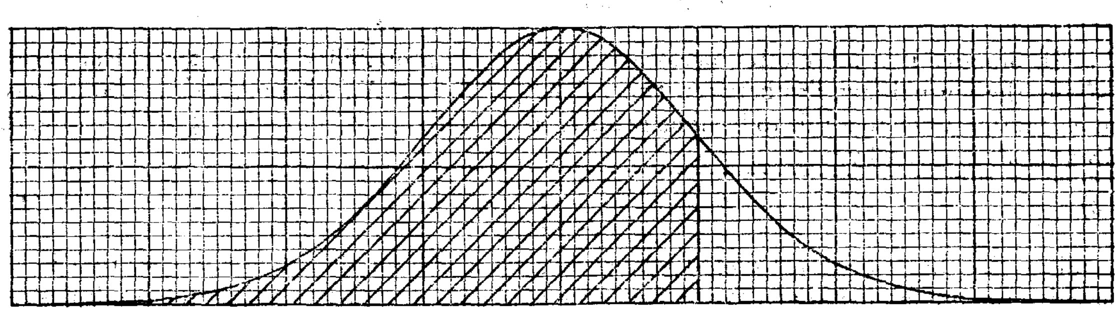


DIAGRAM 3*b*.

By subtracting the area shown in diagram 3*b* from that shown in diagram 3*a*, we get the desired area of diagram 2.

d. For the normal distribution 68 percent of the observations lie within a range of $\pm\sigma$ about the mean; 95 percent within a range of $\pm 2\sigma$ about the mean; 99.7 percent within a range of $\pm 3\sigma$ about the mean.

e. The means of sets of N observations distributed in accordance with the normal law of probability are also distributed normally; their mean is the same as that of the original observations, but with variance $1/N$ as large; i. e., if the mean and variance of the original distribution are μ and σ^2 respectively, then the mean and variance of the distribution of means are μ and σ^2/N respectively. The remarks made in paragraph 9*d* apply here too.

f. If in the binomial distribution p and q do not differ greatly and if n is large, then that distribution is given with a sufficient degree of approximation by a normal distribution with mean equal to np and variance equal to npq; i. e., under the conditions set forth above

$$\frac{n(n-1)(n-2)\ \cdot\ \cdot\ \cdot\ (n-x+1)}{1\times2\times3\times\ \cdot\ \cdot\ \cdot\ \times x}q^{n-x}p^{x}=\text{approx. }\frac{1}{\sqrt{2\pi npq}}e^{-(x-np)^2/2npq}$$

and

$$\sum_{x=0}^{r}\frac{n(n-1)\ \cdot\ \cdot\ \cdot\ (n-x+1)}{1\times2\times\ \cdot\ \cdot\ \cdot\ \times x}q^{n-x}p^{x}=\text{approx. }\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t}e^{-x^2/2}dx$$

where $t=(r-np)/npq$

g. To indicate the approximation of the binomial distribution by the normal distribution, there are listed on page 18 corresponding values as calculated from the binomial distribution, for $n=64$, $p=\frac{1}{2}$, and as given by the normal distribution.[9] (In the normal distribution we use $\mu=np=32$, and $\sigma^2=npq=16$).

Example 9.—What is the probability that in a set of 100 letters of English text, the number of vowels is between 35–45, inclusive? Taking as the probability for the occurrence of a vowel $p=0.40$, there is obtained from the binomial distribution, $\mu=np=40$ and $\sigma^2=npq=24$. $x_0=\frac{X_0-\mu}{\sigma}=\frac{35-40}{4.899}=-1.02$, $x_1=\frac{X_1-\mu}{\sigma}=\frac{45-40}{4.899}=1.02$. From the table of the normal distribution, it is found that $P(-\infty,\ 1.02)=0.8461$ and $P(-\infty,\ -1.02)=0.1539$ so that $P(-1.02,\ 1.02)=0.8461-0.1539=0.6922$. In other words, about 70 percent of sets of 100 letters each of English text will have between 35 and 45 vowels, inclusive.

h. Using the method employed in example 9, limits were calculated within which the number of vowels (A, E, I, O, U, Y), high-frequency consonants (D, N, R, S, T), medium-frequency consonants (B, C, F, G, H, L, M, P, V, W), and low-frequency consonants (J, K, Q, X, Z) would be expected to lie for messages up to 200 letters in length. The results have been graphed and may be found in charts 1, 2, 3, and 4. (See pp. 14, 15, 16, and 17.)

In chart 1, curve V_1 marks the lower limit of the number of vowels to be expected in a message of given length; curve V_2 marks the upper limit. Thus, for example, in a message of 100 letters in plain English there should be between 33 and 47 vowels.

In chart 2, curves H_1 and H_2 mark the lower and upper limits as regards the high-frequency consonants. In a message of 100 letters there should be between 28 and 42 high-frequency consonants.

In chart 3, curves M_1 and M_2 mark the lower and upper limits as regards the medium-frequency consonants. In a message of 100 letters there should be between 17 and 31 medium-frequency consonants.

In chart 4, curves L_1 and L_2 mark the lower and upper limits as regards the low-frequency consonants. In a message of 100 letters there should be between 0 and 3 low-frequency consonants.

---

[9] These values are taken from Yule, G. U., An Introduction to the Theory of Statistics, 9th Ed. Rev. London, 1929, ch. XV.

CHART No. 1

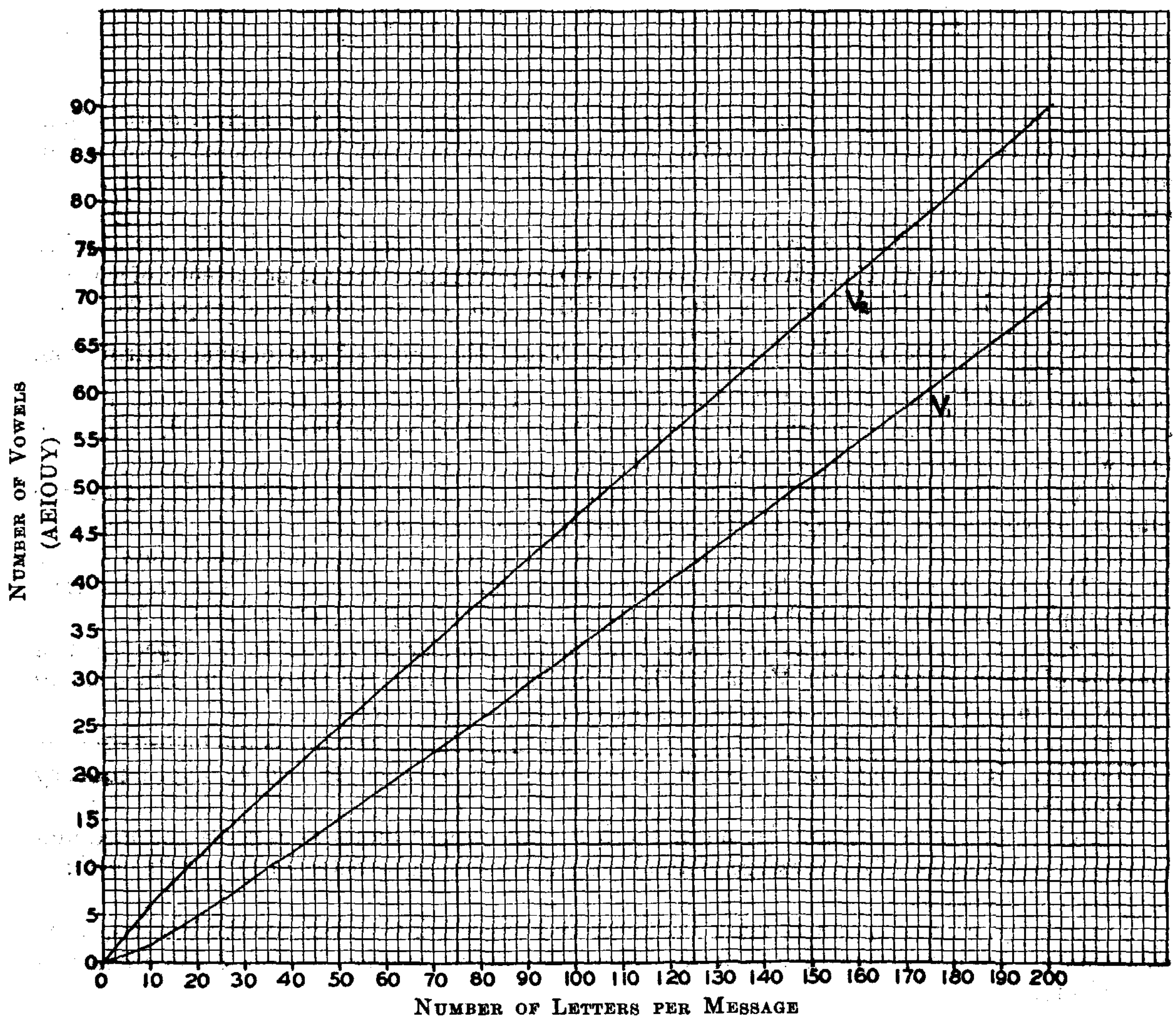

NUMBER OF VOWELS (AEIOUY)
NUMBER OF LETTERS PER MESSAGE
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200
V_2
V_1

CHART NO. 2

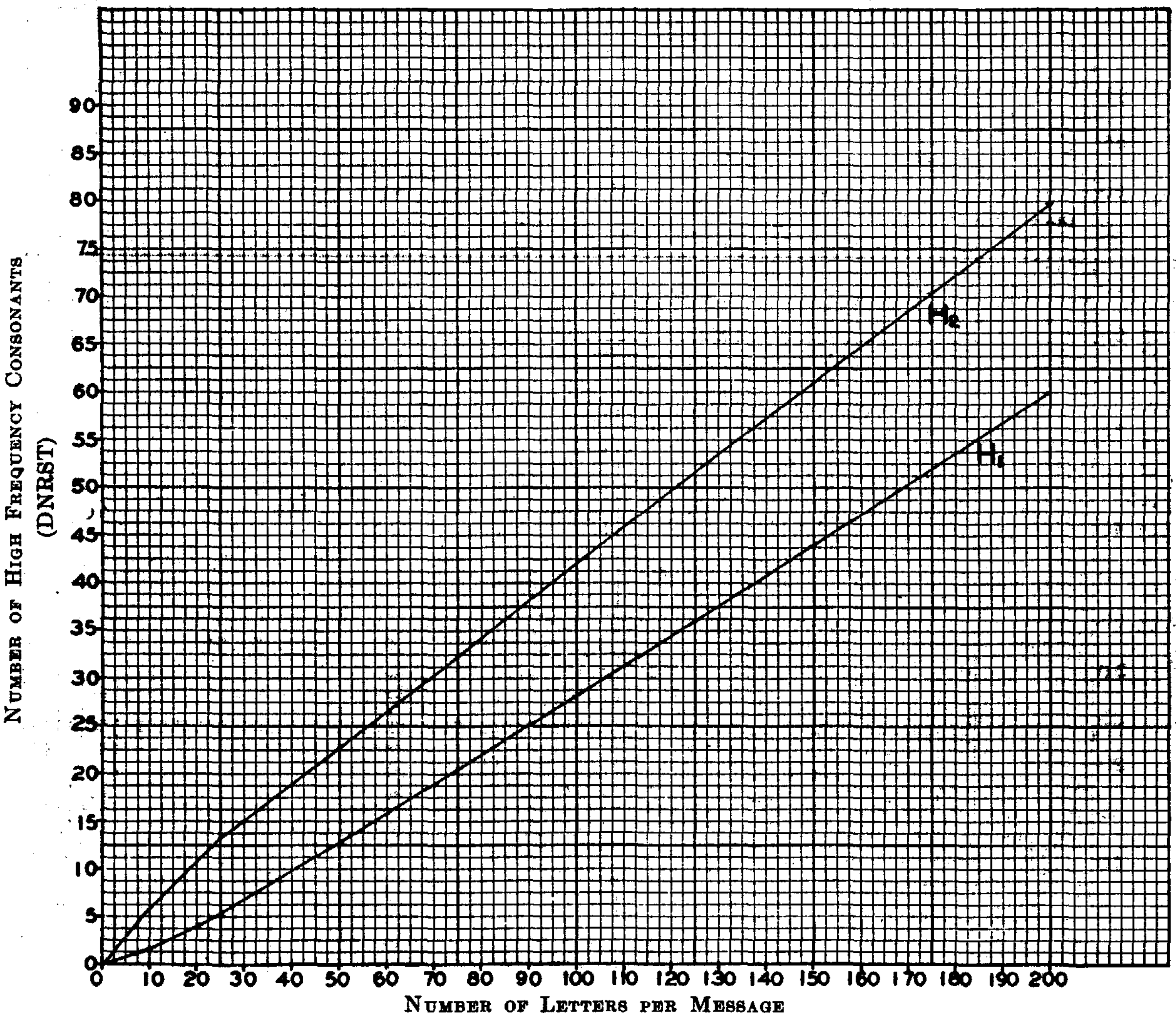

NUMBER OF HIGH FREQUENCY CONSONANTS (DNRST)
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
H_2
H_1
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200
NUMBER OF LETTERS PER MESSAGE

CHART No. 3

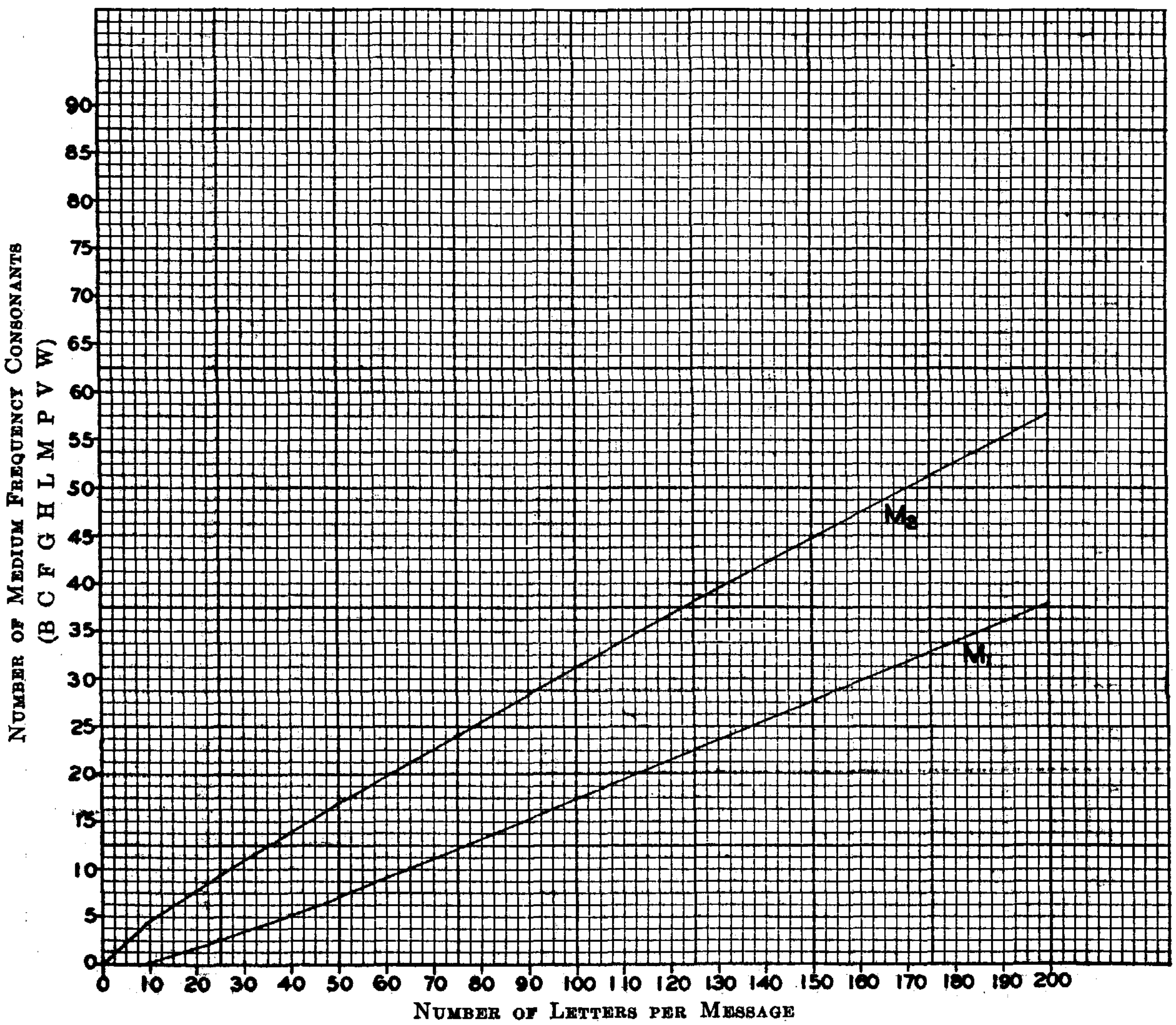
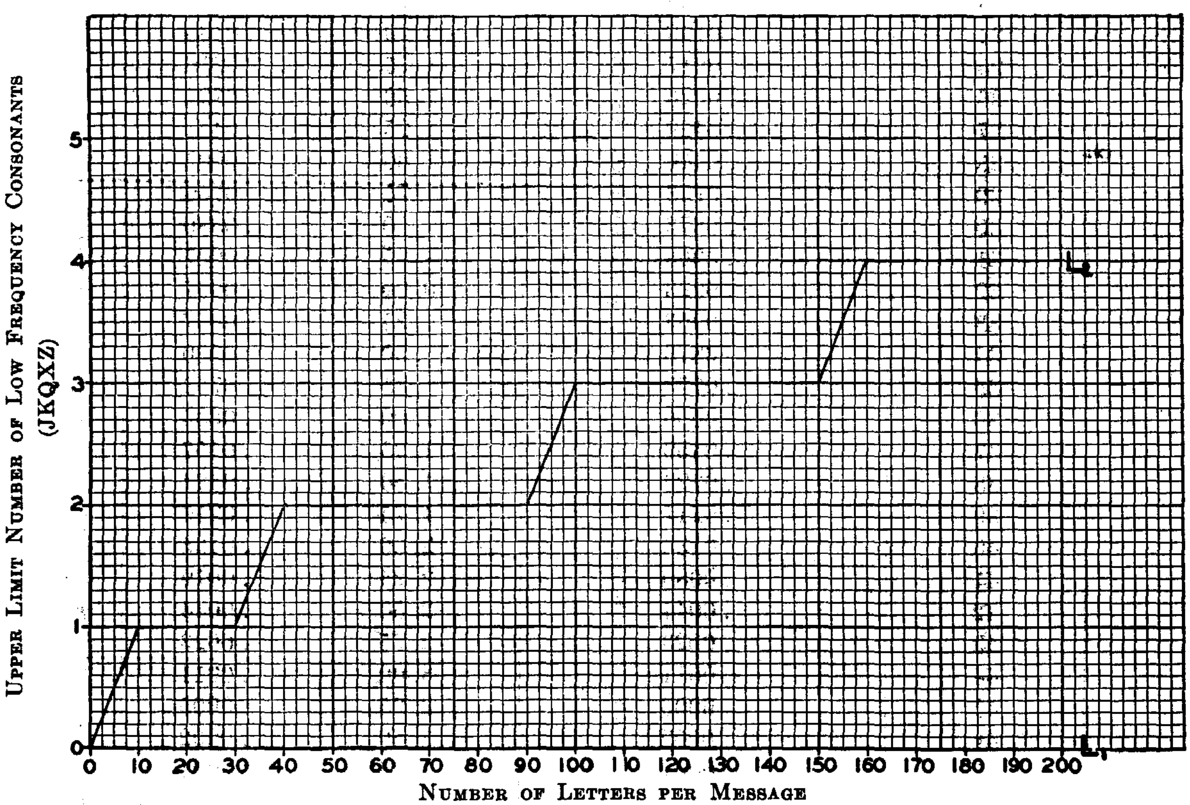

NUMBER OF MEDIUM FREQUENCY CONSONANTS
(B C F G H L M P V W)
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200
NUMBER OF LETTERS PER MESSAGE
M_2
M_1

CHART NO. 4


UPPER LIMIT NUMBER OF LOW FREQUENCY CONSONANTS (JKQXZ)
0
1
2
3
4
5
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200
NUMBER OF LETTERS PER MESSAGE

X	Binomial distribution		Normal distribution
	$\frac{64\times63\times\ldots\times(64-X+1)}{1\times2\times\ldots\times X}\left(\frac{1}{2}\right)^{n-X}\left(\frac{1}{2}\right)^{X}$	$x=\frac{X-32}{4}$	$\frac{1}{4\sqrt{2\pi}}e^{-x^2/32}$
17	0.0001	−3.75	0.0001
18	.0002	−3.50	.0002
19	.0005	−3.25	.0005
20	.0011	−3.00	.0011
21	.0023	−2.75	.0023
22	.0044	−2.50	.0044
23	.0080	−2.25	.0079
24	.0136	−2.00	.0135
25	.0217	−1.75	.0216
26	.0326	−1.50	.0324
27	.0459	−1.25	.0457
28	.0606	−1.00	.0605
29	.0753	−.75	.0753
30	.0873	−.50	.0880
31	.0963	−.25	.0967
32	.0993	0.00	.0997
33	.0963	.25	.0967
34	.0878	.50	.0880
35	.0753	.75	.0753
36	.0606	1.00	.0605
37	.0459	1.25	.0457
38	.0326	1.50	.0324
39	.0217	1.75	.0216
40	.0136	2.00	.0135
41	.0080	2.25	.0079
42	.0044	2.50	.0044
43	.0023	2.75	.0023
44	.0011	3.00	.0011
45	.0005	3.25	.0005
46	.0002	3.50	.0002
47	.0001	3.75	.0001

APPROXIMATION OF THE BINOMIAL DISTRIBUTION BY THE NORMAL DISTRIBUTION

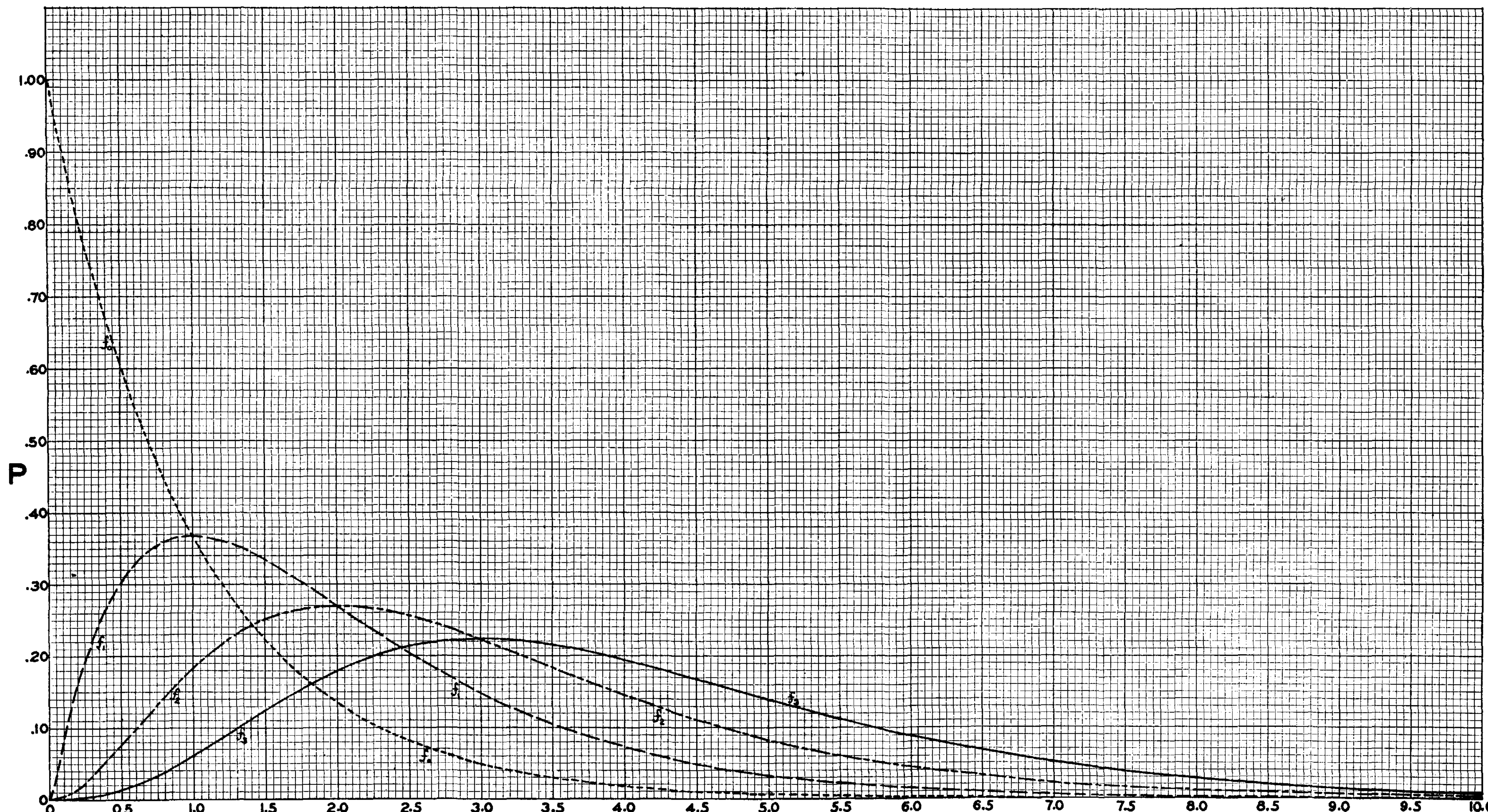
11. **Poisson distribution.**[10]—*a.* In both the binomial and normal distributions, it was seen that there are two parameters that play important roles; n and p in the binomial distribution, and μ and σ in the normal distribution. In the distribution now to be considered there enters but one parameter.

b. The Poisson distribution, known also as the Law of Small Numbers, the Law of Small Probabilities, and Poisson's Exponential Law, relates to a statistical variate which takes on positive integral values only, (0, 1, 2, . . .). According to this distribution, the probability that an event occurs zero, one, two, three, . . . x, . . . times is given by the corresponding term of the sequence

$$e^{-m},\ me^{-m},\ \frac{m^2e^{-m}}{2!},\ \frac{m^3e^{-m}}{3!},\ \ldots,\ \frac{m^xe^{-m}}{x!},\ \ldots$$

[10] See appendix B, p. 149 ff.

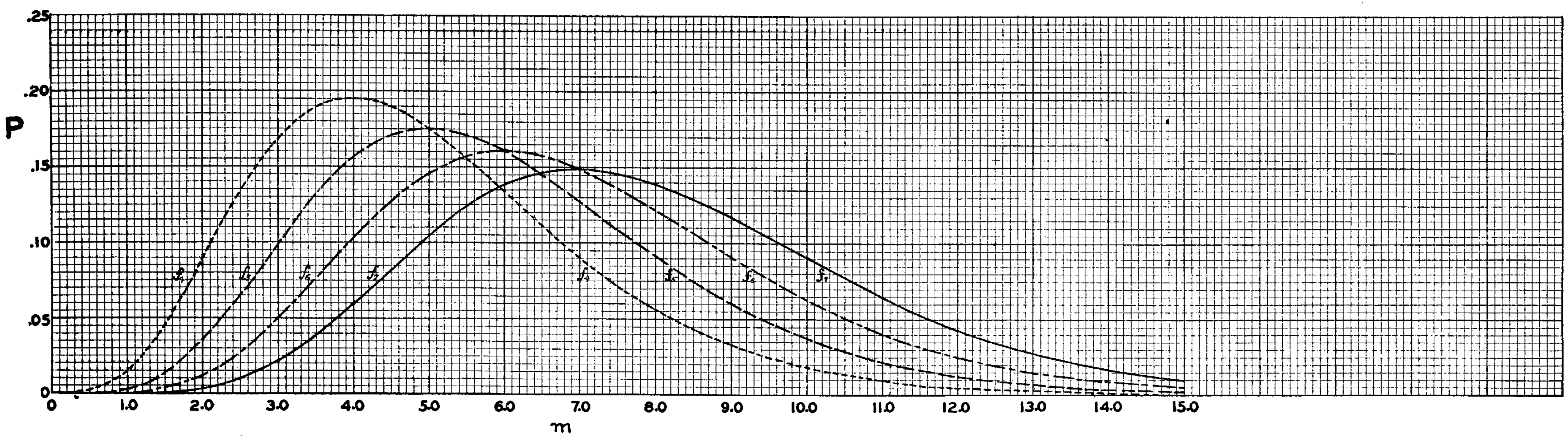
CHART No. 5.—POISSON EXPONENTIAL



was seen
ribution,
nters but

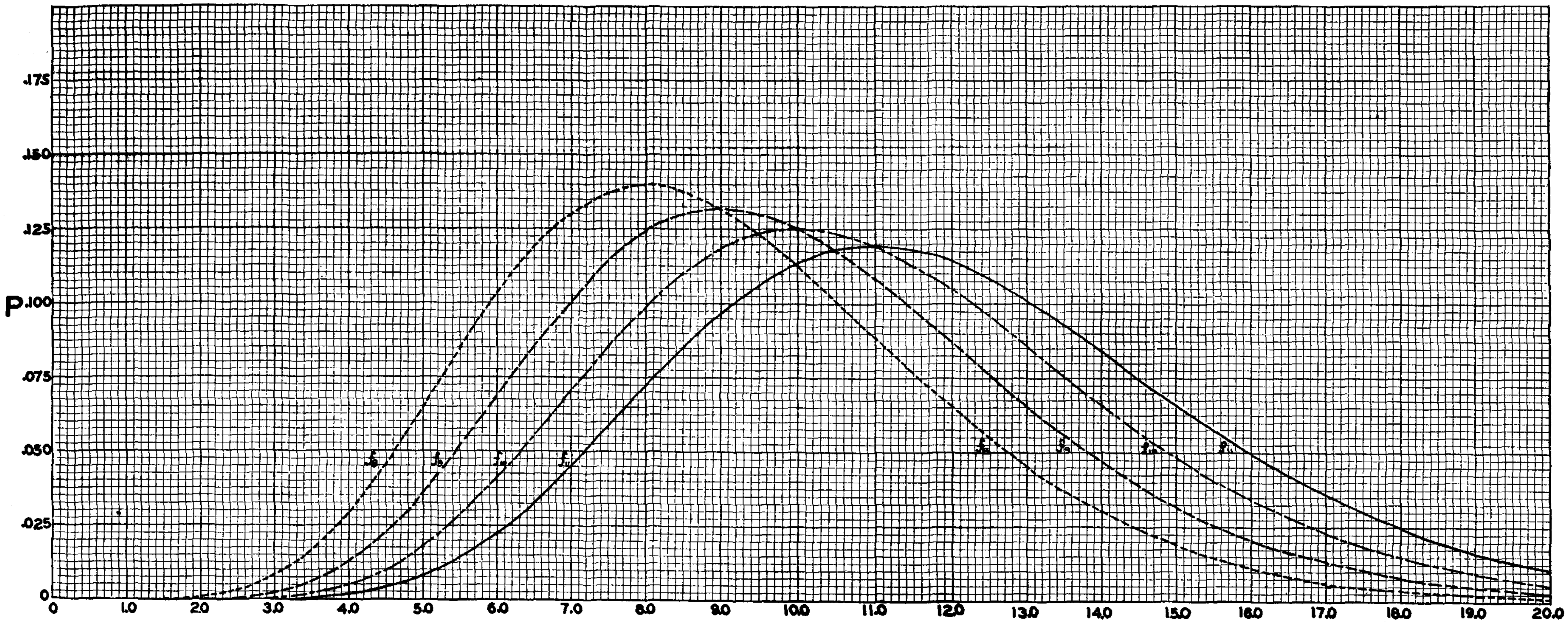
of Small
takes on
bability

CHART No. 6.—POISSON EXPONENTIAL



CURVES SHOWING PROBABILITY FOR 4, 5, 6, and 7 OCCURRENCES OF AN EVENT IN ACCORDANCE WITH THE POISSON EXPONENTIAL DISTRIBUTION

CHART No. 7.—POISSON EXPONENTIAL



CURVES SHOWING PROBABILITY FOR 8, 9, 10, AND 11 OCCURRENCES OF AN EVENT IN ACCORDANCE WITH THE POISSON EXPONENTIAL DISTRIBUTION

where $x!$ is factorial x, i. e., $x(x-1)(x-2)(x-3) \ldots 1$. The parameter m that enters into the distribution is the mean of the statistical variate.

c. The mean and variance of a statistical variate distributed in accordance with the Poisson distribution are equal i. e., $m=\sigma^2$. This may serve as an indication, but not a conclusive one, as to when this distribution may be used.

d. In paragraph 10*f* it was stated that the normal distribution will serve as an approximation to the binomial distribution if n is large and p and q nearly 0.5. If however, p (or q) is small, and n large, the Poisson distribution will provide a good approximation to the binomial distribution.

e. This distribution will be very useful in cryptanalysis since most of the probabilities that the cryptanalyst will consider are small. To facilitate the use of the Poisson distribution, tables have been prepared for this distribution for values of m from 0.1 to 15 by tenths and for the possible values of the statistical variate. These tables will be found in part 2. (See pp. 136–144).

For convenience in certain problems some of the tables have been prepared in graphic form and will be found in charts 5, 6, and 7. On the horizontal axis is plotted the value of the mean and on the vertical axis is plotted the value of the probability. The curves drawn are for 0, 1, 2, . . . , 11 occurrences. Thus in order to find the probability for three occurrences in a Poisson exponential with mean 6 one proceeds as follows: Find the value 6 on the horizontal or m axis; follow this value vertically until the curve f_3 is met; then proceed horizontally to the left where the value $P=0.09$ is found.

f. To indicate the approximation of the binomial distribution by the Poisson distribution, there are listed below values as calculated from the binomial distribution for $n=50$, $p=0.01$ and the corresponding values given by the Poisson distribution for $m=np=0.5$.

X	Binomial distribution $\frac{50\times49\times\ldots\times(50-X+1)}{1\times2\times\ldots\times X}(0.99)^{50-X}(0.01)^X$	X	Poisson distribution $e^{-0.5}(0.5)^X/X!$
0	0. 6050	0	0. 6065
1	. 3055	1	. 3033
2	. 0757	2	. 0758
3	. 0122	3	. 0126
4	. 0015	4	. 0016
5	. 0001	5	. 0002

Example 10.—A study of 100 sets of 50 letters each of English text yielded the following observed distribution for the number of B's per set of 50 letters:

X_i	F_i
0	66
1	29
2	5

(i. e., there were no B's in 66 of the sets, one B in each of 29 of the sets, and 2 B's in each of 5 of the sets). Compare this with the theoretical distribution to be expected according to the binomial distribution and the Poisson distribution, if $p=0.01$ is taken as the probability for the occurrence of B. Since 100 sets were observed, it is merely necessary to multiply the probabilities derived above for the binomial and the corresponding Poisson distribution by 100, in order to get the theoretical number of occurrences (or theoretical absolute frequencies). There is thus obtained:

X_i	Observed	Theoretical	
		Binomial	Poisson
0	66	60. 50	60. 65
1	29	30. 55	30. 33
2	5	7. 57	7. 58
3	0	1. 22	1. 26
4	0	. 15	. 16
5	0	. 01	. 02

12. **Modified Poisson distribution.**—*a.* It may be shown that under certain conditions any discontinuous frequency distribution, for which the variate takes on integral values, may be expressed as the sum of an infinite series of terms consisting of the Poisson exponential and its finite differences. That is to say if $F(x)$ $(x=0, 1, 2, \ldots)$ represents a discontinuous frequency distribution then

$$F(x)=P(x,m)+c_2\Delta^2P(x,m)+c_3\Delta^3P(x,m)+\ \ldots$$

where

$$P(x,m)=e^{-m}m^x/x! \qquad (x=0,\ 1,\ 2,\ \ldots)$$

$$\Delta P(x,m)=P(x,m)-P(x-1,m)$$

$$\Delta^2P(x,m)=\Delta P(x,m)-\Delta P(x-1,m)$$

etc.

and m and c_2, c_3, . . . are determined by $F(x)$. The foregoing series is known as the Poisson-Charlier frequency series or Charlier's type B frequency curves.

b. It has been seen thus far that the application of the binomial distribution is greatly aided by the fact that for values of p and q nearly 0.5 and n large, the normal distribution offers a suitable approximation, and that for p (or q) very small and n large, the Poisson distribution offers a good approximation. In order to find a suitable approximation to the binomial for intermediate values of p it is necessary to modify the Poisson distribution slightly. A satisfactory modification for this purpose is obtained by taking the first two terms of the series described in the preceding subparagraph.

c. According to this modified Poisson distribution, a good approximation for the probability that a statistical variate take the positive, integral value x under the conditions discussed in paragraph 12*b*, is given by

$$\text{(12.1)} \qquad \frac{n!}{x!\,(n-x)!}q^{n-x}p^x=\text{approx. } e^{-m}m^x/x!-\frac{np^2}{2}\Delta^2e^{-m}m^x/x!$$

where

$$\Delta e^{-m}m^x/x!=e^{-m}m^x/x!-e^{-m}m^{x-1}/(x-1)!$$

and $$\Delta^2 e^{-m}m^x/x! = \Delta e^{-m}m^x/x! - \Delta e^{-m}m^{x-1}/(x-1)!$$

The values of $\Delta e^{-m}m^x/x!$ and $\Delta^2 e^{-m}m^x/x!$ are easily obtained from the tables of the Poisson distribution by subtracting consecutive values.

d. To illustrate (12.1) consider the case for $n=100$ and $p=0.1$, so that $m=np=10$, and $np^2/2=0.5$.

In the following, the values in column 2 are taken directly from the tables of the Poisson distribution for $m=10$. The values in column 3 are obtained by subtracting from the corresponding value in column 2 the one just above it. The values in column 4 are obtained by subtracting from the corresponding value in column 3 the one just above it. The values in column 5 are obtained by multiplying the corresponding values of column 4 by $np^2/2=0.5$. Finally, column 6 gives the difference between the corresponding values of columns 2 and 5.

1	2	3	4	5	6
x	$e^{-10}(10)^x/x!$	$\Delta e^{-10}(10)^x/x!$	$\Delta^2 e^{-10}(10)^x/x!$	$0.5\Delta^2 e^{-10}(10)^x/x!$	$e^{-10}(10)^x/x! - 0.5\Delta^2 e^{-10}(10)^x/x!$
0	0.000045	[1] 0.000045	0.000045	0.000023	0.000022
1	.000454	.000409	.000364	.000182	.000272
2	.002270	.001816	.001407	.000704	.001566
3	.007567	.005297	.003481	.001741	.005826
4	.018917	.011350	.006053	.003027	.015890
5	.037833	.018916	.007566	.003783	.034050
6	.063055	.025222	.006306	.003153	.059902
7	.090079	.027024	.001802	.000901	.089178
8	.112599	.022520	−.004504	−.002252	.114851
9	.125110	.012511	−.010009	−.005005	.130115
10	.125110	.000000	−.012511	−.006256	.131366
11	.113736	−.011374	−.011374	−.005672	.119408
12	.094780	−.018956	−.007582	−.003791	.098571
13	.072908	−.021872	−.002916	−.001458	.074366
14	.052077	−.020831	.001041	.000521	.051556
15	.034718	−.017359	.003472	.001736	.032982
16	.021699	−.013019	.004340	.002170	.019529
17	.012764	−.008935	.004084	.002042	.010722
18	.007091	−.005673	.003262	.001631	.005160
19	.003732	−.003359	.002314	.001157	.002575
20	.001866	−.001866	.001493	.000747	.001119
21	.000889	−.000977	.000889	.000445	.000444
22	.000404	−.000485	.000492	.000246	.000158
23	.000176	−.000228	.000257	.000129	.000047
24	.000073	−.000103	.000125	.000063	.000010
25	.000029	−.000044	.000059	.000030	[2] .000000
26	.000011	−.000018	.000026	.000013	[2] .000000
27	.000004	−.000007	.000011	.000006	[2] .000000
28	.000001	−.000003	.000004	.000002	[2] .000000
29	.000001	.000000	.000003	.000002	[2] .000000

[1] The value of $e^{-m}m^x/x!$ for x a negative integer is zero.

[2] Even though these values come out negative they must be considered as 0.000000 since a negative probability has no meaning.

e. Let us now compare the corresponding values as given by the binomial distribution with $n=100$, $p=0.1$, the related Poisson distribution, and the modified Poisson distribution as just derived. The values for the binomial are taken from A. Fisher, Mathematical Theory of Probabilities, p. 268.

x	Binomial	Poisson	Modified Poisson	x	Binomial	Poisson	Modified Poisson
0	0.0001	0.0000	0.0000	12	0.0988	0.0948	0.0986
1	.0003	.0005	.0003	13	.0743	.0729	.0744
2	.0016	.0023	.0016	14	.0513	.0521	.0516
3	.0059	.0076	.0058	15	.0327	.0347	.0330
4	.0159	.0189	.0159	16	.0193	.0217	.0195
5	.0339	.0378	.0341	17	.0106	.0128	.0107
6	.0596	.0630	.0599	18	.0054	.0071	.0052
7	.0889	.0901	.0892	19	.0026	.0037	.0026
8	.1148	.1125	.1149	20	.0012	.0019	.0011
9	.1304	.1251	.1301	21	.0005	.0009	.0004
10	.1319	.1251	.1314	22	.0002	.0004	.0002
11	.1199	.1137	.1194	23	.0000	.0002	.0000

FIGURE 2.

Example 11.—A study of 100 sets of 100 letters each of English plain text yielded the following as the distribution of the occurrences of T.

x	F	x	F	x	F
2	1	7	10	12	10
3	2	8	12	13	7
4	2	9	14	14	3
5	4	10	13	15	2
6	8	11	10	16	2

(i. e., there were 2 T's in 1 set of 100 letters; 3 T's in each of 2 sets of 100 letters each; 4 T's in each of 2 sets of 100 letters each, etc.). Compare the the above distribution with the theoretical distribution to be expected according to the binomial, Poisson, and modified Poisson distributions, taking as the probability for the occurrence of T, $p=0.1$. Since 100 sets were observed it is necessary to multiply the probabilities derived in figure 2 by 100 to get the theoretical absolute frequencies. There thus results

x	Observed	Theoretical			x	Observed	Theoretical		
		Binomial	Poisson	Modified Poisson			Binomial	Poisson	Modified Poisson
0	0	0.01	0.00	0.00	12	10	9.88	9.48	9.86
1	0	.03	.05	.03	13	7	7.43	7.29	7.44
2	1	.16	.23	.16	14	3	5.13	5.21	5.16
3	2	.59	.76	.58	15	2	3.27	3.47	3.30
4	2	1.59	1.89	1.59	16	2	1.93	2.17	1.95
5	4	3.39	3.78	3.41	17	0	1.06	1.28	1.07
6	8	5.96	6.30	5.99	18	0	.54	.71	.52
7	10	8.89	9.01	8.92	19	0	.26	.37	.26
8	12	11.48	11.25	11.49	20	0	.12	.19	.11
9	14	13.04	12.51	13.01	21	0	.05	.09	.04
10	13	13.19	12.51	13.14	22	0	.02	.04	.02
11	10	11.99	11.37	11.94	23	0	0.00	.02	0.00

13. **Multinomial distribution.**[11]—*a.* The multinomial distribution is an extension of the binomial distribution. In the binomial distribution the possible event considered was one of two mutually exclusive categories: The event either did or did not occur. In the multinomial distribution the possible event may be one of r mutually exclusive categories with the respective probabilities of occurrence $p_1, p_2, \ldots, p_r$ where $p_1+p_2+\ldots+p_r=1$.

b. If an event may occur in one of r mutually exclusive ways with the corresponding probabilities $p_1, p_2 \ldots, p_r$ where $p_1+p_2+\ldots+p_r=1$, then in n observations the probability that the event has occurred exactly x_1 times the first way, exactly x_2 times the second way, $\ldots$, exactly x_r times the r th way where $x_1+x_2+\ldots+x_r=n$ is given by

$$P(x_1, x_2, \ldots, x_r)=\frac{n!}{x_1!x_2!\ldots x_r!}p_1^{x_1}p_2^{x_2}\ldots p_r^{x_r}$$

The sum of the foregoing expression for all positive integral values of $x_1, x_2, \ldots, x_r$ such that $x_1+x_2+\ldots+x_r=n$ is $(p_1+p_2+\ldots+p_r)^n$.

The binomial distribution is thus seen to be the preceding for $r=2$ with $p_1=p$ and $p_2=q$.

c. If the possible event be the selection of a letter from English telegraphic text then $r=26$ and the values of $p_1, p_2, \ldots, p_{26}$ are those listed in figure 1. The multinomial distribution will thus give the probability that in a selection of n letters of English telegraphic text there are exactly x_1 A's, x_2 B's, $\ldots$, x_{26} Z's where $x_1+x_2+\ldots+x_{26}=n$.

d. It may be shown that for the multinomial distribution $E(x_i)=np_i$.[12]

$$\begin{aligned}E(x_i^2)&=n^2p_i^2+np_i(1-p_i)\\ E(x_ix_j)&=n(n-1)p_ip_j=E(x_i)E(x_j)-np_ip_j \qquad (i\neq j;\, i, j=1, 2, \ldots, r)\\ &=E(x_i)E(x_j)-\frac{E(x_i)E(x_j)}{n}=\frac{n-1}{n}E(x_i)E(x_j)^{13}\end{aligned}$$

[11] See appendix C, p. 150.

[12] $E(\ )$ means the expected or average value of the expression in the parenthesis.

[13] For events which are independent in the sense of probability $E(x_ix_j)=E(x_i)E(x_j)$.

SECTION V

APPLICATIONS

14. **Repetitions.**—The importance of the role played by repetitions in the analysis of cryptograms is well understood, even by the amateur cryptanalyst. Repetitions in cryptographic text are basically of two sorts—causal and accidental. Causal repetitions are those which represent the encipherment of plain-text repetitions which have undergone the same cryptographic treatment. Accidental repetitions are those, which, through fortuitous circumstances, are the encipherments of different plain-text elements. In the case of most cryptograms of the substitution class, the finding of repetitions of sequences of fair length, say four, five, or more characters, usually leads to solution; because as the lengths of repetitions increase it becomes more certain that such repetitions are causal and not accidental in nature. However, it often happens in the case of the more complex types of cryptograms that repetitions are rather scarce and such as are found are short. In such cases it becomes very important to be able to judge whether the repetitions which are present are causal or are accidental. In the following we shall consider certain procedures and tests which will be of service in the evaluation of the cryptographic significance of repeated cipher elements.

15. **Expected number of blanks in random text.**—*a*. By random text is meant text in which the interplay of those factors which give rise to a particular cipher element is such that the cipher elements will occur with approximately the same probability, e. g., the cipher text produced by a polyalphabetic substitution of say 10 different alphabets would be random text insofar as the individual letters of the cryptogram were concerned. The uniliteral frequency distribution of such text would be "flat," i. e., there would be no pronounced crests and troughs.

b. Suppose there is at hand a selection of random text of N elements of a system in which there are n different elements possible, e. g., the text may consist of $N=50$ letters of an $n=26$ letter alphabet; or we may consider a text of $N=376$ digraphs where there are $n=676$ different possible digraphs, etc. Then the probability for the occurrence of a particular element is $1/n$. Not all of the n possible elements will necessarily occur in the text of N elements, and the number which does not appear is sometimes of significance. To take advantage of that number it would be necessary to know the theoretical distribution of the number of blanks, i. e., of the number of elements which do not appear. This distribution has been found to be

$$P_0(r)=\frac{n!}{r!}\sum_{x=0}^{n-r}(-1)^x\frac{1}{x!(n-r-x)!}\left(1-\frac{r+x}{n}\right)^N \tag{15.1}$$

where $P_0(r)$ represents the probability that there are exactly r blanks.

c. The values of (15.1) for $n=N=10$ are as follows:

r	$P_0(r)$	r	$P_0(r)$
0	0. 000362880	6	0. 017188920
1	. 016329600	7	. 000671760
2	. 136080000	8	. 000004599
3	. 355622400	9	. 000000001
4	. 345144240		
5	. 128595600		1. 000000000

A study of 200 sets of 10 random digits each, yielded the following as the distribution of the number of blanks per set of 10 digits.

Number of blanks	Theoretical frequency	Observed frequency	
r	$200P_0(r)$	f	rf
0	0. 08	0	0
1	3. 26	8	8
2	27. 22	22	44
3	71. 12	72	216
4	69. 02	72	288
5	25. 72	21	105
6	3. 44	4	24
7	. 14	1	7
8	0. 00	0	0
9	0. 00	0	0
	200. 00	200	692

From the foregoing it is seen that the observed average number of blanks per set of 10 digits is 692/200=3.46.

d. The average (or expected) number of blanks in a frequency distribution of random text of N elements of a system in which there are n different elements possible is given by [14]

(15.2) $$B_N=n(1-1/n)^N$$

For large values of n a good enough approximation is given by

(15.3) $$B_N=ne^{-N/n}$$

where $e=2.7183$ is the base of natural logarithms. For particular values of N and n, the value

[14] The value in (15.2) may be derived from the distribution given by (15.1) in accordance with the definition of the mean. However, the following simple considerations will lead to the same result. The probability that a particular element does not appear is $(1-1/n)$. In N observations, the probability that a particular element has not occurred is $(1-1/n)^N$. Since there are n different possible elements, the expected number of blanks is as in (15.2).

of B_N may be found from tables [15] of e^{-x}. For $n=26$, i. e., for monographic distributions a chart has been prepared whereby the value of B_N may be readily found for values of N from 0 to 200. This chart, No. 8, will be found on page 30.

Example 12.—How many blanks are to be expected in the digraphic distribution of a random text of 100 digraphs? In this case $N=100$ and $n=676$. Thus $B_{100}=676e^{-100/676}$; $100/676=0.148$; $e^{-0.148}=0.861$; $676\times 0.861=582$ or there are to be expected 582 blanks or $676-582=94$ different digraphs.

e. For large values of n (say $n\geqq 26$) it may be shown that the value in (15.1) is to a sufficient approximation given by

$$P_0(r)=\frac{n!}{r!(n-r)!}e^{-rN/n}(1-e^{-N/n})^{n-r} \tag{15.4}$$

In other words, the distribution of the number of blanks in random text of N elements of a system in which there are n elements possible is given by the binomial distribution with $p=e^{-N/n}$ and $n=n$, so that $\mu=ne^{-N/n}$ and $\sigma^2=ne^{-N/n}(1-e^{-N/n})$.

16. **Expected number of blanks in non-random text.**—*a.* By non-random text is meant text in which the elements have been properly allocated in accordance with their cryptographic treatment. Thus, the text of a cryptogram enciphered polyalphabetically with 10 alphabets, although random text in so far as the individual letters are concerned when considered as a whole, is non-random text when each letter is allocated to the proper alphabet. The text of a Playfair Cipher, for example, is non-random text when divided up into digraphs. Monoalphabetic text is an example of non-random text, closely akin to plain-text.

b. Suppose that the n possible elements of non-random text have different probabilities of occurrence, e. g., for monoalphabetic systems in English, the different probabilities of the various letters are those given in figure 1; for digraphic systems the different probabilities of the various digraphs are those given in section VIII. Let these n probabilities be $p_1, p_2, \ldots, p_n$. *In the following discussion the values of the probabilities only are of importance and not the correspondence between certain plain-text elements and certain probabilities.* In other words from a statistical viewpoint plain-text and non-random text are the same. If a text of N elements is considered, then all n possible elements will not necessarily appear. The theoretical distribution of the number of blanks is known for this case also.

c. If $P_0(r)$ represents the probability that there are exactly r blanks, then it may be shown that

$$\begin{aligned}
P_0(0)&=1-\sum_{i=1}^{n}(1-p_i)^N+\frac{1}{2!}\sum_{i,j=1}^{n}(1-p_i-p_j)^N-\frac{1}{3!}\sum_{i,j,k=1}^{n}(1-p_i-p_j-p_k)^N+\ldots\\
P_0(1)&=\sum_{i=1}^{n}(1-p_i)^N-\sum_{i,j=1}^{n}(1-p_i-p_j)^N+\frac{1}{2!}\sum_{i,j,k=1}^{n}(1-p_i-p_j-p_k)^N-\ldots\\
P_0(2)&=\frac{1}{2!}\left\{\sum_{i,j=1}^{n}(1-p_i-p_j)^N-\sum_{i,j,k=1}^{n}(1-p_i-p_j-p_k)^N-\ldots\right\}\\
P_0(3)&=\frac{1}{3!}\left\{\sum_{i,j,k=1}^{n}(1-p_i-p_j-p_k)^N-\ldots\right\}\\
&\text{etc.}
\end{aligned} \tag{16.1}$$

No special cases of (16.1) have been evaluated. If in (16.1) $p_1=p_2=\ldots=p_n=1/n$, then (16.1) reduces to (15.1) as is to be expected.

[15] Smithsonian Physical Tables. 7th Ed. Rev., pp. 48–53. The f_0 curve of Chart No. 5 may also be employed, since it is in reality the graph of e^{-m}.

d. If the n possible elements of a system have the probabilities of occurrence $p_1, p_2, \ldots, p_n$ respectively, then the average number of blanks in a text of N elements is given by [16]

$$B_N=(1-p_1)^N+(1-p_2)^N+\ldots+(1-p_n)^N \tag{16.2}$$

A good approximation to the formula in (16.2) is given by

$$B_N=e^{-Np_1}+e^{-Np_2}+\ldots+e^{-Np_n} \tag{16.3}$$

e. Using the values of $p_i(i=1, 2, \ldots, 26)$ for English text given in figure 3, (16.3) yields for the number of blanks in monoalphabetic (or plain) text, for values of N from 10 to 200, the results shown in figure 4.

$p_1=0.07189$	$p_{10}=0.00198$	$p_{19}=0.05754$
$p_2=.01146$	$p_{11}=.00353$	$p_{20}=.09042$
$p_3=.03345$	$p_{12}=.03549$	$p_{21}=.02993$
$p_4=.04290$	$p_{13}=.02534$	$p_{22}=.01340$
$p_5=.12604$	$p_{14}=.07558$	$p_{23}=.01401$
$p_6=.02994$	$p_{15}=.07408$	$p_{24}=.00469$
$p_7=.01795$	$p_{16}=.02661$	$p_{25}=.02099$
$p_8=.03287$	$p_{17}=.00318$	$p_{26}=.00101$
$p_9=.07592$	$p_{18}=.08256$	

FIGURE 3.

N	Average number of blanks		N	Average number of blanks
	Theoretical	Observed		Theoretical
10	18. 40	18. 50	110	5. 64
20	14. 27	14. 13	120	5. 46
30	11. 71	11. 55	130	5. 21
40	10. 06	10. 03	140	5. 04
50	8. 86	8. 84	150	4. 88
60	7. 95	7. 98	160	4. 78
70	7. 28	7. 33	170	4. 67
80	6. 75	6. 74	180	4. 56
90	6. 28	6. 29	190	4. 44
100	5. 98	5. 83	200	4. 40

FIGURE 4.

The observed values were obtained as the averages of 100 sets of text of 10, 20, . . . , 100 letters each. In view of the excellent correspondence between the observed and theoretical values, it was deemed unnecessary to continue this check for the cases $N=110$ to 200. The actual distributions of the observed number of blanks is given in figure 5.

---

[16] The value in (16.2) may be derived from the distribution given by (16.1) in accordance with the definition of the mean. However the following simple considerations will lead to the same result. The probability that the i th ($i=1, 2, \ldots, n$) element does not appear is $(1-p_i)$. The probability that the i th element does not occur in N observations is $(1-p_i)^N$. The expected number of blanks is thus as given in (16.2).

NUMBER OF LETTERS IN TEXT

Number of Blanks	10	20	30	40	50	60	70	80	90	100
26										
25										
24										
23										
22	1									
21	1									
20	16									
19	32									
18	33	1								
17	13	4								
16	4	12	1							
15		19	1							
14		34	5	1						
13		20	21	6	1					
12		4	25	11	5	2				
11		6	24	25	13	5	2	1		
10			15	17	17	13	9	4	1	1
9			5	22	18	15	13	12	7	4
8			1	12	24	22	19	14	13	7
7			2	4	13	24	25	21	24	20
6				2	8	15	20	26	24	31
5					1	4	9	15	22	17
4							3	5	5	15
3								1	2	2
2								1	2	1
1										2

FIGURE 5.

f. The graphs for the number of blanks given by (15.3) and (16.3) for monoalphabetic distributions in English have been plotted on one chart, chart number 8. Thus, given a text of N letters, one can estimate whether or not the text has been enciphered monoalphabetically, by comparing the observed number of blanks with the expected number of blanks in a text of N letters for both random and monoalphabetic text. The chart will be found on page 30 and also on page 163. A more accurate test as to whether or not the text were random would be to see whether the observed number of blanks could reasonably arise from the distribution given by (15.4).

g. The corresponding results for French, German, Italian, Portuguese, and Spanish are given below, in figure 6, the values of p_i used are given in Section VIII. Charts have been prepared so that the average number of blanks may be readily found for values of N from 10 to 200. These charts, charts Nos. 9, 10, 11, 12, and 13, will be found on pages 31–35 and also on pages 164–168.

N	Theoretical average number of blanks				
	French (25 letter alphabet)	German	Italian (21 letter alphabet)	Spanish	Portuguese (24 letter alphabet)
10	17. 87	18. 50	13. 62	16. 72	16. 58
20	13. 99	14. 37	9. 80	12. 75	12. 81
30	11. 59	11. 77	7. 53	10. 42	10. 39
40	9. 99	10. 01	6. 04	8. 78	8. 84
50	8. 85	8. 77	4. 98	7. 59	7. 73
60	7. 99	7. 74	4. 18	6. 69	6. 90
70	7. 31	7. 18	3. 57	5. 98	6. 24
80	7. 01	6. 63	3. 07	5. 38	5. 70
90	6. 25	6. 20	2. 66	4. 89	5. 26
100	5. 93	5. 80	2. 33	4. 41	4. 90
150	4. 65	4. 69		3. 02	3. 64
200	3. 97	4. 35		1. 22	2. 99

FIGURE 6.

17. **Expected number of elements occurring r times each.**—*a*. Results similar to those derived for the number of blanks are obtainable for the number of elements each of which occurs once, twice, three times, etc. Although the exact theoretical distributions have been found for each case, they will not be given here.

b. For random text of N elements, where there are n different possible elements, the average number of elements occurring once each is given by

(17.1) $$N(1-1/n)^{N-1}$$

the average number of elements occurring twice each is given by [17]

---

[17] *If $N>n$, it is certain that some elements will occur more than once.* If $N \leqq n$ it is possible that no element may occur more than once. Let us accordingly consider the problem, "In random text of N elements, where there are n elements possible and $N \leqq n$, what is the probability that at least one element occurs more than once?" The various possible forms that the distribution of the N elements may assume are given by the terms of the expansion of the multinomial $(p_1+p_2+ \ . \ . \ . \ +p_n)^N$ where $p_1=p_2= \ . \ . \ . \ =p_n=1/n$. The required probability is the sum of all those terms which contain at least one exponent greater than one (or the required probability is one minus the sum of all those terms having every exponent equal to one). Since $N \leqq n$ the number of terms in which every exponent is one is $n!/N!(n-N)!$ or the combination of n things taken N at a time. In accordance with the multinomial distribution, a sample of one of these terms is

$$\frac{N!}{1!1! \ . \ . \ . \ 1!}p_1p_2 \ . \ . \ . \ p_N.$$

Since $p_1=p_2= \ . \ . \ . \ =p_n=1/n$ we have that the sum of all those terms with each exponent equal to one is given by

$$\frac{n!}{N!(n-N)!}\cdot\frac{N!}{n^N}=\frac{n!}{(n-N)!n^N}.$$

Accordingly the probability that at least one element occurs more than once is given by $1-n!/(n-N)!n^N$. For large values of n a good approximation to $n!/(n-N)!n^N$ is given by $e^{-N(N-1)/2n}$, or the required probability is given by $1-e^{-N(N-1)/2n}$. As an example consider a random text of 100 letters. What is the probability that a digraphic distribution of the text will show at least one digraph occurring twice? Since there are 99 digraphs in the 100 letters, $N=99$, $n=676$. Thus, the required probability is $1-e^{99\times 98/2\times 676}$. $99\times 98=9702$; $2\times 676=1352$; $9702/1352=7.2$, $e^{-7.2}=0.0007$; $1-e^{-7.2}=0.9993$. It is practically certain that at least one digraph will occur more than once. For trigraphs the values are $N=98$, $n=17{,}576$. Thus, $98\times 97=9506$; $2\times 17{,}576=35{,}152$;

(17.2) $$N(N-1)n(1-1/n)^{N-2}/n^2 \cdot 2!$$

the average number of elements occurring r times each is given by

(17.3) $$N(N-1)\ \ldots\ (N-r+1)n(1-1/n)^{N-r}/n^r \cdot r!$$

For large values of n (17.1), (17.2), and (17.3) may respectively be approximated by

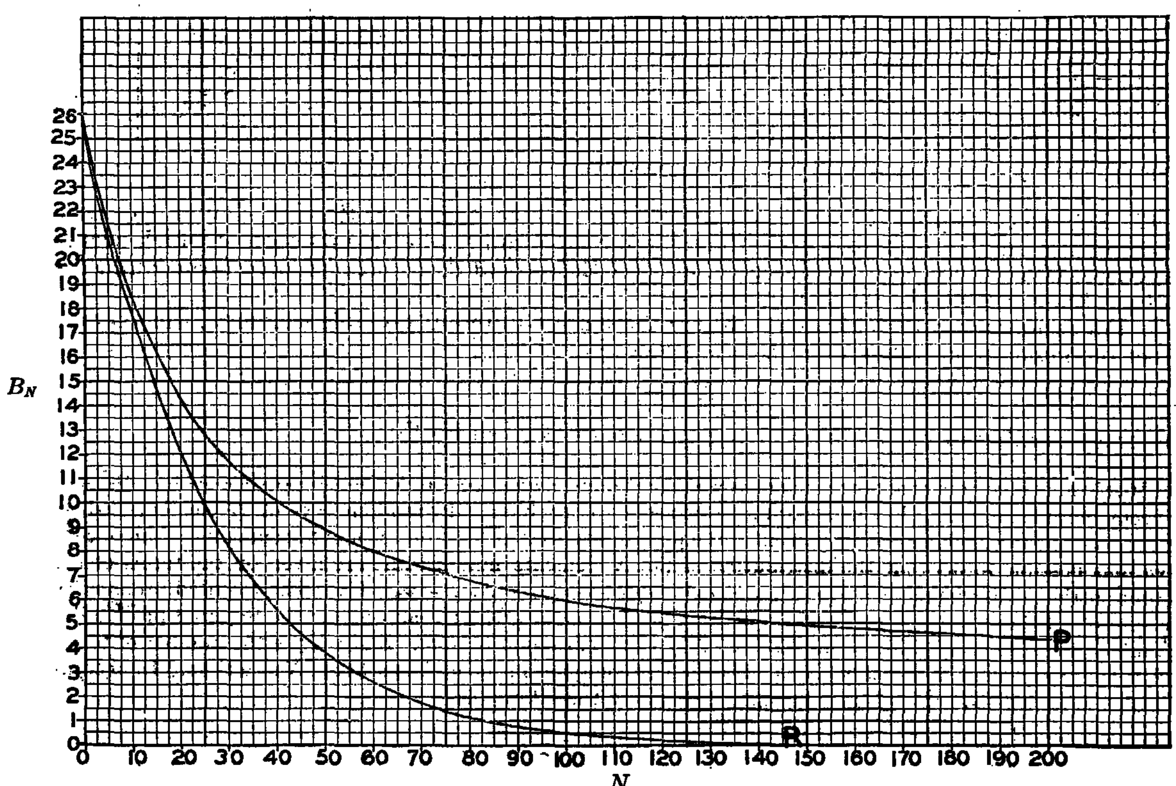
(17.4) $$n(N/n)e^{-N/n}$$

(17.5) $$n(N/n)^2(1/2!)e^{-N/n}$$

(17.6) $$n(N/n)^r(1/r!)e^{-N/n}$$

The numerical values of (17.4), (17.5), and (17.6) for special cases may be easily found by means of the tables for the Poisson Exponential distribution, wherein are given the values of $(1/r!)\ (N/n)^r e^{-N/n}$ for values of r from 0 to 37 and for values of $m=N/n$ by tenths from 0.1 to 15 or from Charts 5, 6, and 7.

CHART No. 8.—EXPECTED NUMBER OF BLANKS ENGLISH PLAIN TEXT (P) AND RANDOM TEXT (R)



NUMBER OF LETTERS PER MESSAGE

---

[17]—Continued.

9506/35,152=0.27; $e^{-0.27}=0.7642$; $1-e^{-0.27}=0.2358$. In other words, about 24 out of 100 such selections will show at least one trigraph occurring more than once. For tetragraphs, $N=97$, $n=456,976$ so that $97\times 96=9312$; $2\times 456,976=913,952$; $9312/913,952=0.01$; $e^{-0.01}=0.98$; $1-e^{-0.01}=0.02$. In other words about 2 out of 100 such selections would show at least one tetragraph occurring more than once. For pentagraphs $N=96$, $n=11,881,376$ so that $96\times 95=9120$; $2\times 11,881,376=23,762,752$; $9120/23,762,752=0.0004$; $e^{-0.0004}=0.9996$; $1-e^{-0.0004}=0.0004$. In other words it is almost certain that such a selection of text would not show a single pentagraph (or for that matter, a polygraph of more than five letters) occurring more than once.

CHART NO. 9.—EXPECTED NUMBER OF BLANKS FRENCH PLAIN TEXT

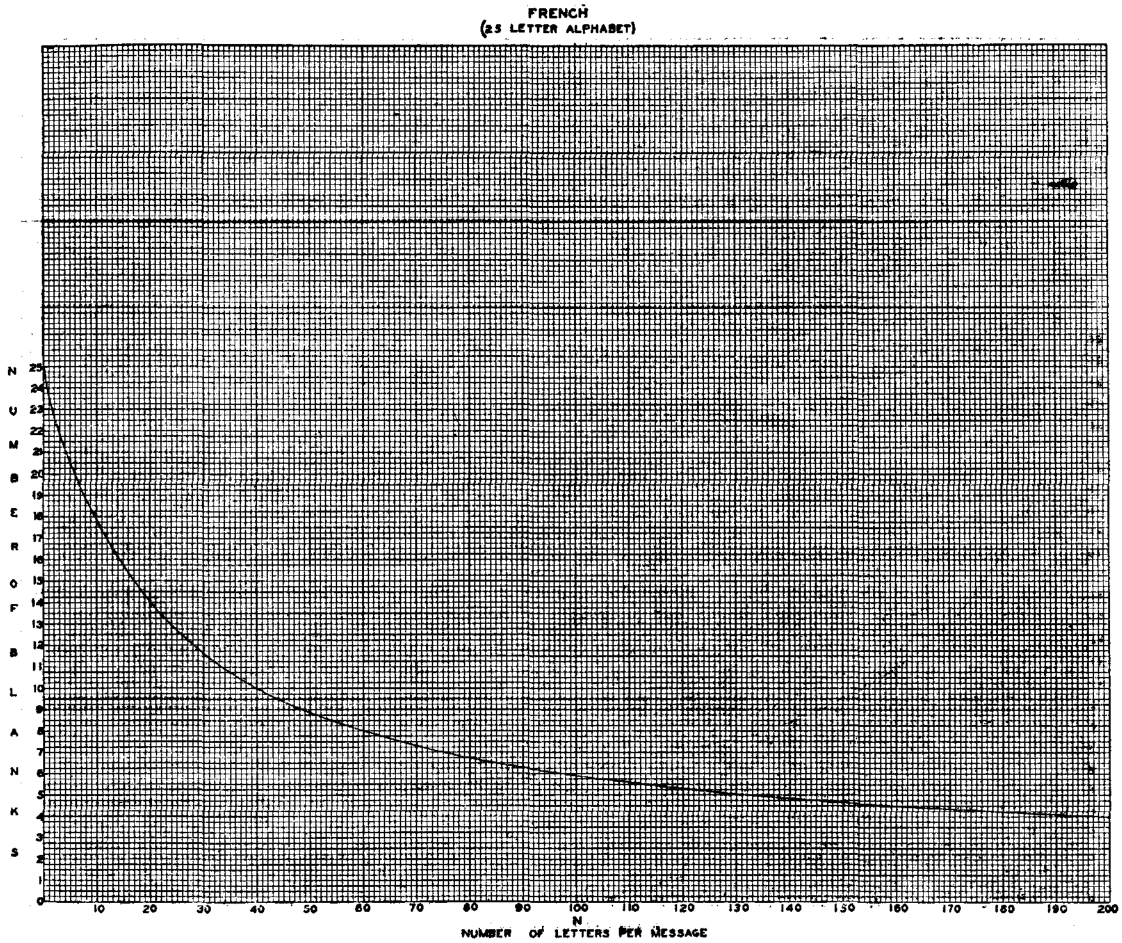


CHART No. 10.—EXPECTED NUMBER OF BLANKS GERMAN PLAIN TEXT

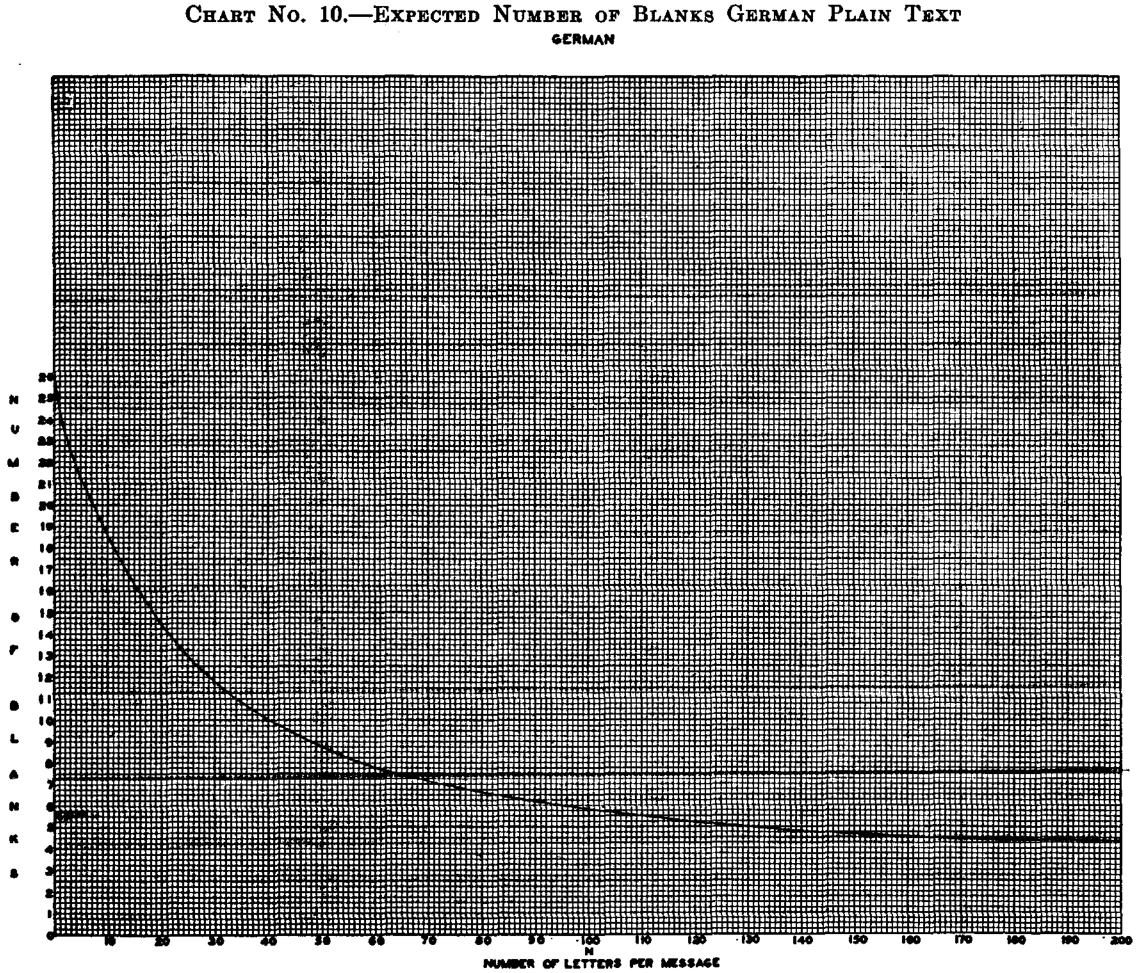


CHART No. 11.—EXPECTED NUMBER OF BLANKS ITALIAN PLAIN TEXT

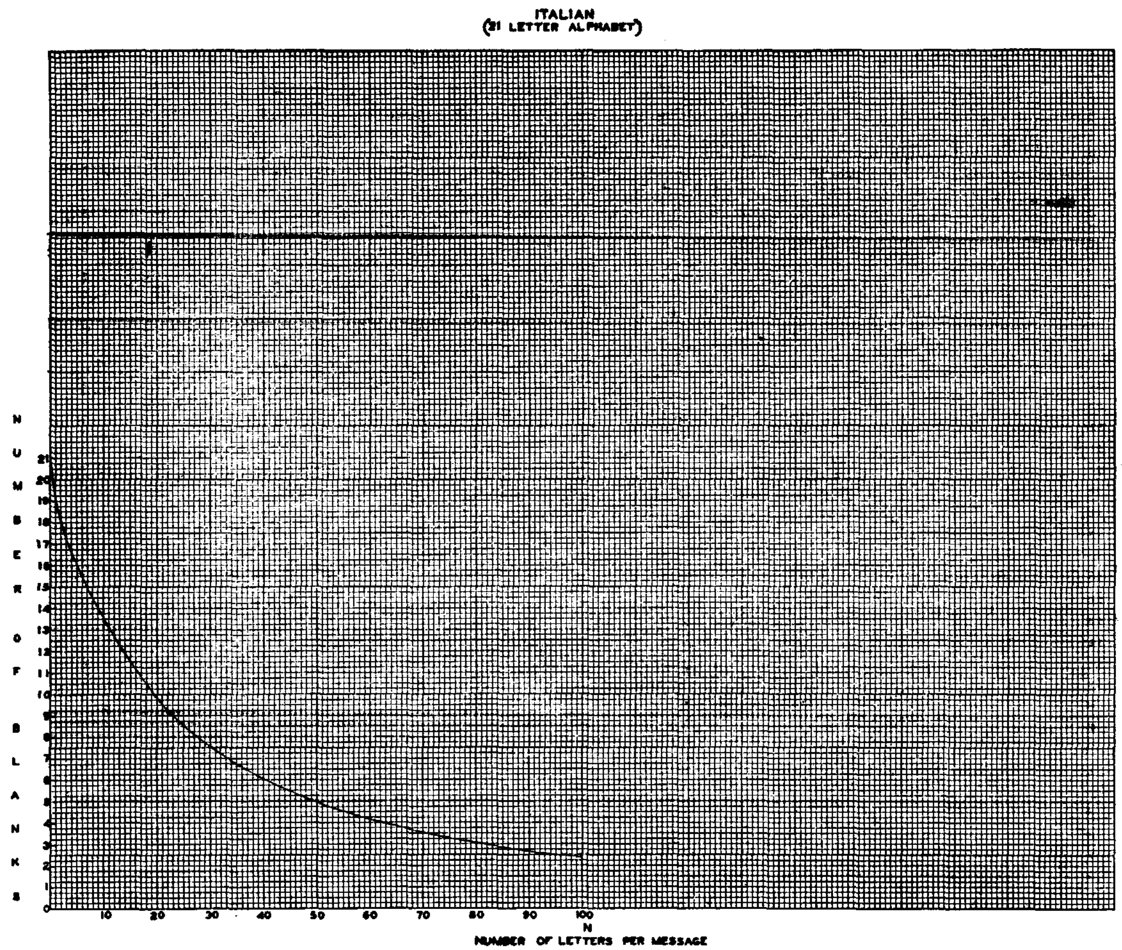


CHART No. 12.—EXPECTED NUMBER OF BLANKS PORTUGUESE PLAIN TEXT

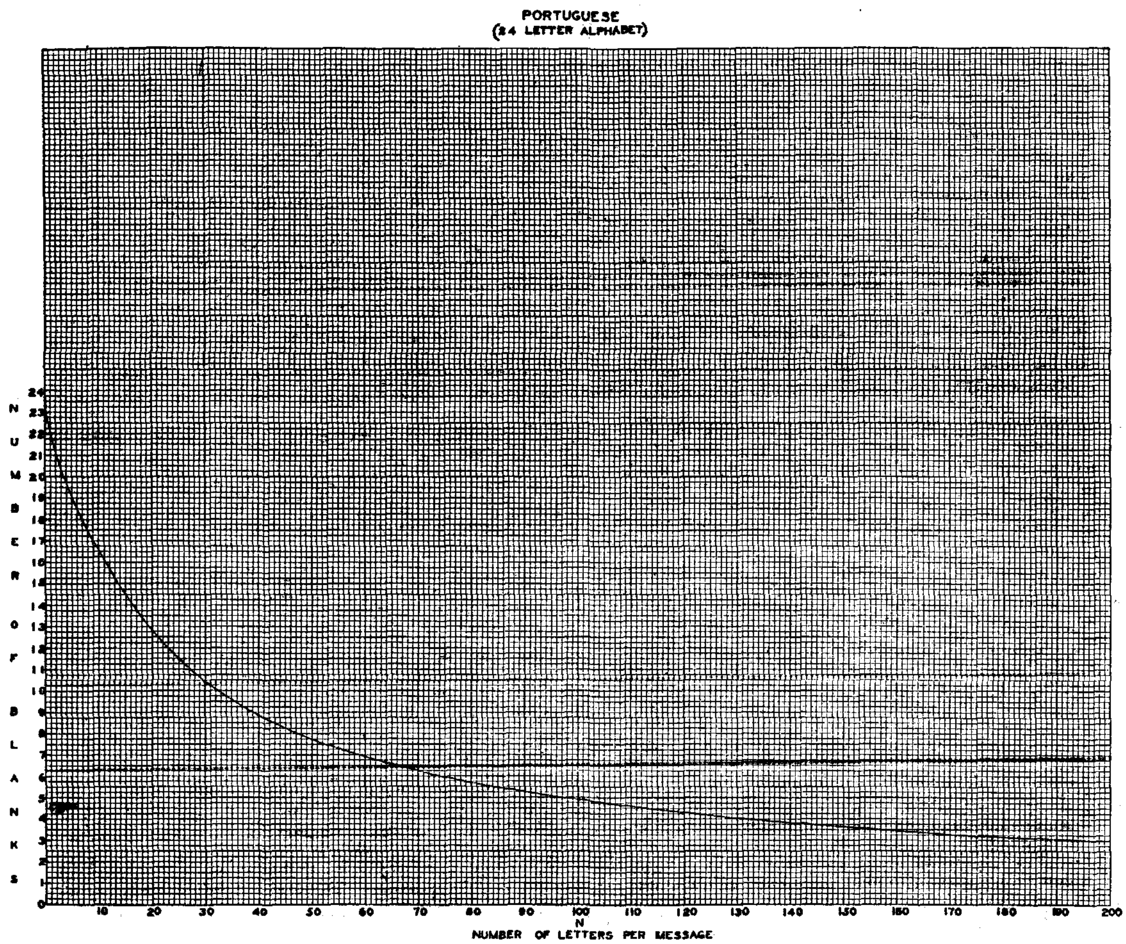
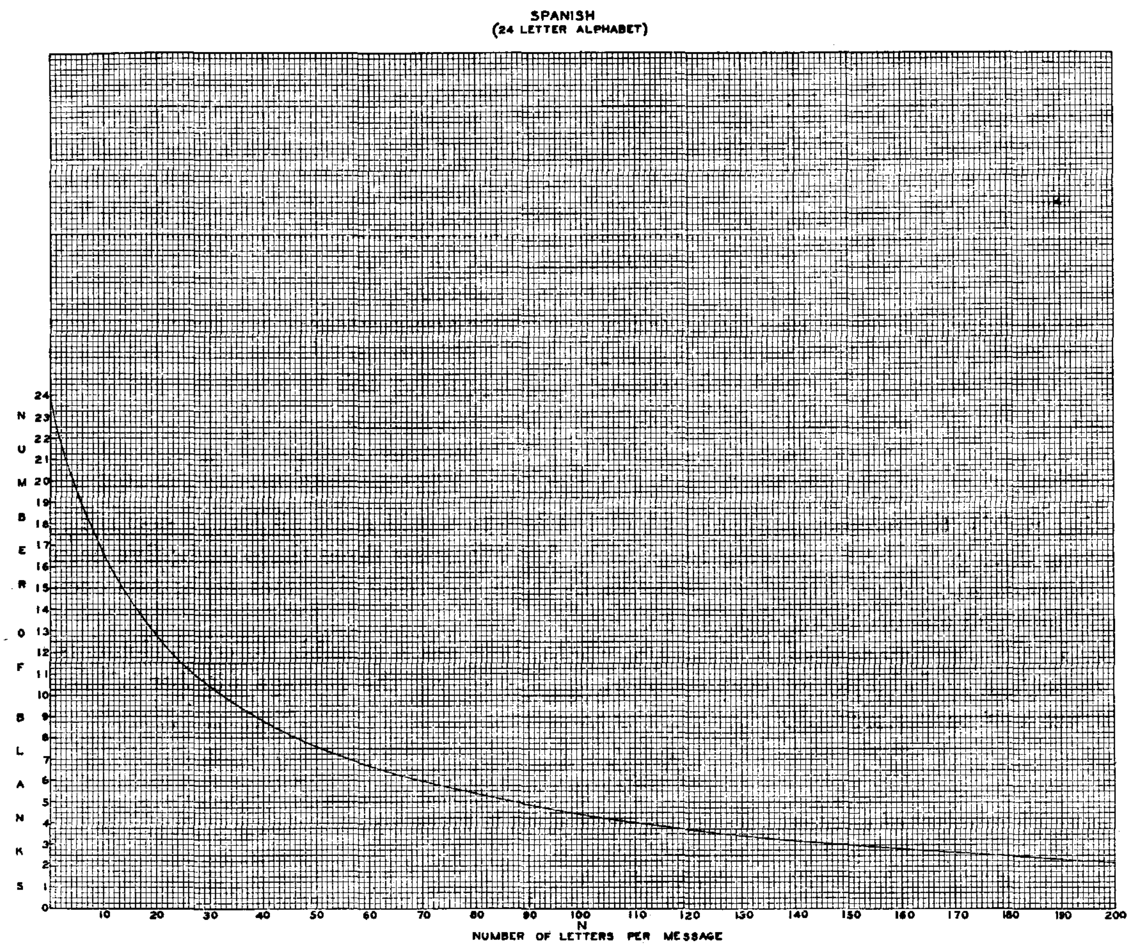


CHART No. 13.—EXPECTED NUMBER OF BLANKS SPANISH PLAIN TEXT



Example 13.—Given a random text of 104 letters, find the expected number of letters each of which occurs no times, once, twice, etc.

In this case $N=104$, $n=26$, so that $N/n=4$. The desired values are given below in the last column; the values in the middle column were obtained from the tables of the Poisson exponential distribution for $m=4$.

r	$(1/r!)(4)^re^{-4}$	$26(1/r!)(4)^re^{-4}$
0	0.018316	0.476216=0
1	.073263	1.904838=2
2	.146525	3.809650=4
3	.195367	5.079542=5
4	.195367	5.079542=5
5	.156293	4.063618=4
6	.104196	2.709096=3
7	.059540	1.548040=2
8	.029770	.774020=1
9	.013231	.344006=0
10	.005292	.137592=0
11	.001925	.050050=0
12	.000642	.016692=0
13	.000197	.005122=0
14	.000056	.001456=0

In other words, the average random text of 104 letters would show all letters occurring; two occurring once each; four occurring twice each; five occurring three times each; five occurring four times each; four occurring five times each; three occurring six times each; two occurring seven times and one occurring eight times.

c. In non-random text of N elements, where there are n possible different elements with the respective probabilities of occurrence $p_1, p_2, \cdots, p_n$, the average number of elements occurring once each is given by

(17.7) $$N\sum_{i=1}^{n}p_i(1-p_i)^{N-1};$$

the average number of elements occurring twice each is given by

(17.8) $$\frac{N(N-1)}{2!}\sum_{i=1}^{n}p_i^2(1-p_i)^{N-2};$$

the average number of elements occurring r times each is given by

(17.9) $$\frac{N(N-1)\cdots(N-r+1)}{r!}\sum_{i=1}^{n}p_i^r(1-p_i)^{N-r}.$$

The formulas (17.7), (17.8), and (17.9) may be respectively approximated by

(17.10) $$\sum_{i=1}^{n}(Np_i)e^{-Np_i}$$

(17.11) $$\sum_{i=1}^{n}(1/2!)(Np_i)^2e^{-Np_i}$$

(17.12) $$\sum_{i=1}^{n}(1/r!)(Np_i)^r e^{-Np_i}$$

The formulas in (17.10), (17.11), and (17.12) may also be evaluated by means of the tables for the Poisson exponential.

d. Charts giving the number of letters occurring r times each, for various values of N have *not* been prepared since these variations are to a large extent taken into account in formulas to be discussed now.

18. **The ϕ test for non-random character of text.**—*a.* It is to be expected, that the variation in the number of occurrences of the n possible elements of a text of N elements would be greater for non-random text than for random text. Some measure of this variation is desirable as a quantative test as to whether or not the text of a cryptogram has been properly arranged into its simplest component elements.

b. Consider a text of N elements in a system where there are n possible elements. Let us suppose that there are $f_1, f_2, \ldots, f_n$ respectively of each of the different possible elements in the text so that $f_1+f_2+\ldots+f_n=N$.

If we set $\phi=f_1(f_1-1)+f_2(f_2-1)+\ldots+f_n(f_n-1)$ then it is possible to show that

(18.1) $$E(\phi)=s_2N(N-1)$$

where $E(\ )$ means the average or expected value of the expression in the parenthesis, and s_2 is the sum of the squares of the probabilities of occurrence of each of the n possible elements in the system. (The definition of ϕ is not as arbitrary as may first appear, but is related to a most important concept, that of coincidences, which is discussed in Section VII. In paragraph 25*b* of that section is given a proof of (18.1)).

For monoalphabetic and digraphic distributions (18.1) yields the results shown below:

	E (ϕ)	
	Monoalphabetic text	Digraphic text
English	$0.0661N(N-1)$	$0.0069N(N-1)$
French	$.0778N(N-1)$	$.0093N(N-1)$
German	$.0762N(N-1)$	$.0112N(N-1)$
Italian	$.0738N(N-1)$	$.0081N(N-1)$
Japanese (Romaji)	$.0819N(N-1)$	$.0116N(N-1)$
Portuguese	$.0791N(N-1)$	
Russian	$.0529N(N-1)$	$.0058N(N-1)$
Spanish	$.0775N(N-1)$	$.0093N(N-1)$

For random text, $s_2=1/n$, so that (18.1) yields the results shown below:

$E(\phi)$

RANDOM TEXT

Monographic	Digraphic	Trigraphic
$0.038N(N-1)$	$0.0015N(N-1)$	$0.000057N(N-1)$

Example 14.—Does the following represent a selection of English text enciphered monoalphabetically?

IBMQO PBIUO MBBGA JCZOF MUUQB

A uniliteral distribution of the text yields the following:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

For this case the observed value of ϕ is $1\times0+5\times4+1\times0+1\times0+1\times0+2\times1+1\times0+3\times2+3\times2+1\times0+2\times1+3\times2+1\times0=42$. For monoalphabetic text in English the expected value is $0.066\times25\times24=39.6$; for random text the expected value is $0.038\times25\times24=22.8$. One must conclude that the cipher text is the result of a monoalphabetic substitution, since the observed value of ϕ (42) more closely approximates the expected value for English plain-text (39.6) than it does the expected value for random text (22.8).

Example 15.—Does the following represent a selection of English text enciphered monoalphabetically?

HKWZA RRPBQ BIVYS MPDMQ MVUDC

A uniliteral distribution of the text yields the following:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

For this case the observed value of ϕ is $1\times0+2\times1+1\times0+2\times1+1\times0+1\times0+1\times0+3\times2+2\times1+2\times1+2\times1+1\times0+1\times0+2\times1+1\times0+1\times0+1\times0=18$. As in example 14, the expected values for monoalphabetic and random text are 39.6 and 22.8 respectively. One must conclude that the text is *not* monoalphabetic.

For convenience we shall refer to the test described above as the ϕ (Phi) test.

c. From (18.1), there may be derived after some simple manipulation a formula for the expected value of the sum of the squares of the number of occurrences of each element. If we set $\psi=f_1^2+f_2^2+\ .\ .\ .\ +f_n^2$, then

(18.2) $$E(\psi)=s_2N^2+(1-s_2)N$$

The values of s_2 for monoalphabetic and digraphic text, for various languages are shown herewith:

	Monoalphabetic	Digraphic
English	0.0661	0.0069
French	.0778	.0093
German	.0762	.0112
Italian	.0738	.0081
Japanese (Romaji)	.0819	.0116
Portuguese	.0791	
Russian	.0529	.0058
Spanish	.0775	.0093

d. An idea of the variation of the observed values of $\phi=f_1(f_1-1)+f_2(f_2-1)+\ .\ .\ .\ +f_n(f_n-1)$ about its expected value is indicated by the variance which is

(18.3) $$\sigma_\phi^2=4N^3(s_3-s_2^2)+2N^2(5s_2^2+s_2-6s_3)+2N(4s_3-s_2-3s_2^2)$$

where s_2 and s_3 are respectively the sum of the squares and cubes of the probabilities of occurrence of each of the n possible elements.[18]

e. For English monoalphabetic text (18.3) becomes

(18.4) $$\sigma_\phi^2=0.004344N^3+0.110448N^2-0.114794N$$

For random monographic text (18.3) becomes

(18.5) $$\sigma_\phi^2=0.073964N(N-1)$$

f. The variance of the distribution of observed values of $\psi=f_1^2+f_2^2+\ .\ .\ .\ +f_n^2$ about the expected value is given by (18.3) also, so that the values in (18.4) and (18.5) for the special cases therein considered are also the same.[19]

g. We can approximate the distributions of ϕ and ψ by means of the normal distribution since we know both the mean and standard deviation (the positive square root of the variance).

h. The theoretical values obtained from (18.1), (18.2), and (18.3) for English monoalphabetic text, were compared with the corresponding values obtained from 100 sets of text for $N=10$, 20, . . . , 90, with the result shown below:

N	$E(\phi)$		$E(\psi)$		Standard deviation	
	Theoretical	Observed	Theoretical	Observed	Theoretical	Observed
10	5.9	6.5	15.9	16.5	3.8	4.0
20	25.1	25.9	45.1	45.9	8.8	10.5
30	57.4	57.6	87.4	87.6	14.6	17.1
40	103.0	103.6	143.0	143.6	22.2	22.7
50	161.7	161.5	211.7	211.5	28.5	29.2
60	233.6	236.6	293.6	296.6	36.5	34.8
70	318.8	323.5	388.8	393.5	44.9	43.2
80	417.1	423.5	497.1	503.5	54.1	50.3
90	528.7	534.0	618.7	624.0	63.7	58.5

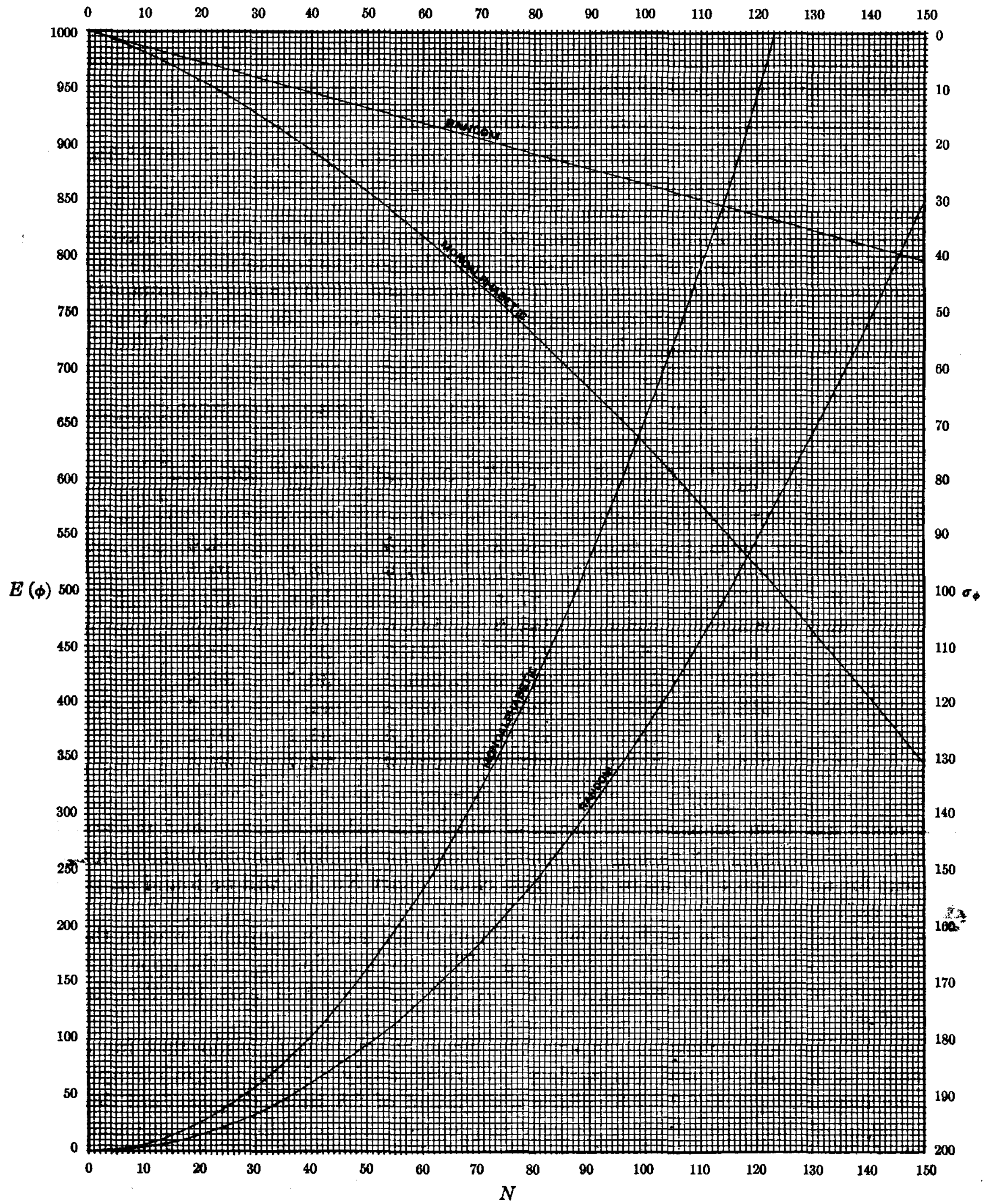
i. A chart has been prepared by means of which the values of (18.1) and the standard deviation as derived from (18.4) may be readily found for English monoalphabetic text and random text for all values of N up to 150. This chart, chart No. 14, will be found on page 40 and also on page 169.

The curves originating in the lower left-hand corner are used in conjunction with the scale on the left vertical axis for the expected value of ϕ. The curves originating in the upper left-hand corner are used in conjunction with the scale on the right vertical axis for the standard deviation of ϕ.

Let us consider again examples 14 and 15. From chart 14 it is found that for $N=25$, $\sigma_\phi=6.8$ and 11.5 for random and non-random text, respectively. If the text in example 14 is random, we have $(42-22.8)/6.8=2.8$. Since a deviation of 2.8 times the standard deviation from the mean of the normal curve is very improbable, our conclusion in example 14 is strengthened. If the text in example 15 is monoalphabetic, we have $(18-39.6)/11.5=-1.9$. If the text in example 15 is random, we have $(18-22.8)/6.8=-.7$. Thus our conclusion in example 15 is strengthened.

[18] See appendix D, p. 151–153.
[19] See appendix D, p. 151–153.

CHART NO. 14.—EXPECTED VALUE AND STANDARD DEVIATION OF ϕ



NUMBER OF LETTERS PER MESSAGE

j. Although a single ϕ test for small values of N would rarely give a reliable result, it is nevertheless possible to apply this test for small values of N, provided it is possible to obtain the average for a number of such tests.

Thus it is possible to determine the period of a polyalphabetic cipher, where the number of alphabets is large and the number of letters per distribution is small, even though there are no long repetitions.

k. Consider the following cryptogram which is known to be enciphered polyalphabetically with a number of alphabets between 40 and 50.

HSKUS	PMFHD	UJJIX	MSPTP	OIPCI	WKZVU
YPPNE	USAIG	BOOGA	OPGPR	HBOUC	SHPVG
HQXZS	ACKRK	VBGHM	VSFRY	YTKHK	VWZXV
LIJHW	ARLKF	IJSLT	MHKAH	QTUVT	XSMEC
FCSKT	GOOYB	XZVLI	JRYAC	DWEJM	SCAFP
IEAXO	KAQDW	EXPYP	QHDNO	JIXNZ	JGNUD
OARFU	ERJOY	BDOKE	IKDUV	TDVEV	LETDO
AFROU	NYNBD	VQOBE	GGSHQ	HXOPU	ZCOCU
KKZLT	PHKRT	CCOAS	BZUGB	UBBUN	OVTPO
VMIZD	EPQFV	KZ			

Assuming that 50 alphabets were used the message would be rewritten as in figure 7.

The value of ϕ for each alphabet is calculated and the result is as given in figure 7. The distribution of the observed values of ϕ is given below in figure 8.

For $N=6$ the expected value of ϕ for monoalphabetic text is $0.066\times6\times5=1.98$ and for random text is $0.038\times6\times5=1.14$. For $N=5$ the corresponding values are respectively $0.066\times5\times4=1.32$ and $0.038\times5\times4=0.76$. Since $\bar{\phi}$ is the mean of 32 and 18 observations, respectively, to find $\sigma_{\bar{\phi}}$ it is necessary to divide the standard deviations as obtained from (18.4) and (18.5) by $\sqrt{32}$ and $\sqrt{18}$ for $N=6$ and $N=5$, respectively. (See par. 10*e*.)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	H	S	K	U	S	P	M	F	H	D	U	J	J	I	X	M	S	P	T	P	O	I	P	C	I
2	H	B	O	U	C	S	H	P	V	G	H	Q	X	Z	S	A	C	K	R	K	V	B	G	H	M
3	I	J	S	L	T	M	H	K	A	H	Q	T	U	V	T	X	S	M	E	C	F	C	S	K	T
4	I	E	A	X	O	K	A	Q	D	W	E	X	P	Y	P	Q	H	D	N	O	J	I	X	N	Z
5	T	D	V	E	V	L	E	T	D	O	A	F	R	O	U	N	Y	N	B	D	V	Q	O	B	E
6	C	C	O	A	S	B	Z	U	G	B	U	B	B	U	N	O	V	T	P	O	V	M	I	Z	D
ϕ	4	0	2	2	2	0	2	0	2	0	2	0	0	0	0	0	2	0	0	2	6	2	0	0	0

	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	W	K	Z	V	U	Y	P	P	N	E	U	S	A	I	G	B	O	O	G	A	O	P	G	P	R
2	V	S	F	R	Y	Y	T	K	H	K	V	W	Z	X	V	L	I	J	H	W	A	R	L	K	F
3	G	O	O	Y	B	X	Z	V	L	I	J	R	Y	A	C	D	W	E	J	M	S	C	A	F	F
4	J	G	N	U	D	O	A	R	F	U	E	R	J	O	Y	B	D	O	K	E	I	K	D	U	V
5	G	G	S	H	Q	H	X	O	P	U	Z	C	O	C	U	K	K	Z	L	T	P	H	K	R	T
6	E	P	Q	F	V	K	Z																		
ϕ	2	2	0	0	0	2	2	0	0	2	0	2	0	0	0	2	0	2	0	0	0	0	0	0	2

FIGURE 7.

50 ALPHABETS

$N=6$			$N=5$		
ϕ_i	w_i	$\phi_i w_i$	ϕ_i	w_i	$\phi_i w_i$
	Number of occurrences			*Number of occurrences*	
0	17	0	0	13	0
2	13	26	2	5	10
4	1	4	4	0	0
6	1	6	6	0	0
	32	36		18	10

$\bar{\phi}=36/32=1.13.$ $\bar{\phi}=10/18=.56.$

FIGURE 8.

There thus results

		$N=6$	$N=5$
Monoalphabetic text	$\sigma_{\bar{\phi}}$	0. 36	0. 40
	$E(\phi)$	1. 98	1. 32
Observed $\bar{\phi}$		1. 13	. 56
Random text	$E(\phi)$	1. 14	. 76
	$\sigma_{\bar{\phi}}$	. 26	. 29

FIGURE 9.

From the values in figure 9 there is obtained the following:

N	Monoalphabetic text		Random text	
	$x=\frac{\bar{\phi}-E(\phi)}{\sigma_{\bar{\phi}}}$	$P(-\infty, x)$	$x=\frac{\bar{\phi}-E(\phi)}{\sigma_{\bar{\phi}}}$	$P(x, \infty)$
5	−1. 90	0. 0287	−0. 69	0. 7549
6	−2. 36	. 0092	−. 04	. 5160

The value $P(x, \infty)$ is obtained from the normal probability table page 135, by using the fact that $P(x, \infty)=P(-\infty, -x)$. The foregoing shows that for $N=5$ only 3 percent of monoalphabetic text would yield a value of $\bar{\phi}$ as small or smaller than that observed whereas 75 percent of random text would yield a value of $\bar{\phi}$ as big or bigger than that observed; for $N=6$ only 1 percent of monoalphabetic text would yield a value of $\bar{\phi}$ as small or smaller than that observed whereas 52 percent of random text would yield a value of $\bar{\phi}$ as big or bigger than that observed. We conclude that 50 alphabets were not used.

l. Assuming that 49 alphabets were used the message would be rewritten as in figure 10. The value of ϕ for each alphabet is again calculated and is given in figure 10. The distribution of the observed values of ϕ is given below in figure 11.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	H	S	K	U	S	P	M	F	H	D	U	J	J	I	X	M	S	P	T	P	O	I	P	C	I
2	R	H	B	O	U	C	S	H	P	V	G	H	Q	X	Z	S	A	C	K	R	K	V	B	G	H
3	K	F	I	J	S	L	T	M	H	K	A	H	Q	T	U	V	T	X	S	M	E	C	F	C	S
4	A	F	P	I	E	A	X	O	K	A	Q	D	W	E	X	P	Y	P	Q	H	D	N	O	J	I
5	K	D	U	V	T	D	V	E	V	L	E	T	D	O	A	F	R	O	U	N	Y	N	B	D	V
6	P	H	K	R	T	C	C	O	A	S	B	Z	U	G	B	U	B	B	U	N	O	V	T	P	O
ϕ	2	4	2	0	4	2	0	2	2	0	0	2	2	0	2	0	0	2	2	2	2	4	2	2	2

	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49
1	W	K	Z	V	U	Y	P	P	N	E	U	S	A	I	G	B	O	O	G	A	O	P	G	P
2	M	V	S	F	R	Y	Y	T	K	H	K	V	W	Z	X	V	L	I	J	H	W	A	R	L
3	K	T	G	O	O	Y	B	X	Z	V	L	I	J	R	Y	A	C	D	W	E	J	M	S	C
4	X	N	Z	J	G	N	U	D	O	A	R	F	U	E	R	J	O	Y	B	D	O	K	E	I
5	Q	O	B	E	G	G	S	H	Q	H	X	O	P	U	Z	C	O	C	U	K	K	Z	L	T
6	V	M	I	Z	D	E	P	Q	F	V	K	Z												
ϕ	0	0	2	0	2	6	0	0	0	4	2	0	0	0	0	0	6	0	0	0	2	0	0	0

FIGURE 10.

49 ALPHABETS

$N=6$			$N=5$		
ϕ_i	w_i	$\phi_i w_i$	ϕ_i	w_i	$\phi_i w_i$
	Number of occurrences			*Number of occurrences*	
0	14	0	0	10	0
2	18	36	2	1	2
4	4	16	4	0	0
6	1	6	6	1	6
	37	58		12	8

$\bar{\phi}=58/37=1.57$ $\bar{\phi}=8/12=0.67$

FIGURE 11.

	$N=6$	$N=5$
$E(\phi)$ Monoalphabetic text	1.98	1.32
Observed $\bar{\phi}$	1.57	.67
$E(\phi)$ Random text	1.14	.76

FIGURE 12.

We omit the detailed analysis used in the previous case as it seems quite obvious that 49 alphabets were not used.

m. A similar procedure yields the following results for 48 alphabets.

48 Alphabets

$N=6$			$N=5$		
ϕ_i	w_i	$\phi_i w_i$	ϕ_i	w_i	$\phi_i w_i$
	Number of occurrences			*Number of occurrences*	
0	19	0	0	5	0
2	19	38	2	1	2
4	3	12	4	0	0
6	1	6	6	0	0
	42	56		6	2

$\bar{\phi}=56/42=1.33$ $\qquad$ $\bar{\phi}=2/6=0.33$

Figure 13.

	$N=6$	$N=5$
$E(\phi)$ Monoalphabetic text	1.98	1.32
Observed $\bar{\phi}$	1.33	.33
$E(\phi)$ Random text	1.14	.76

Figure 14.

n. The results for 47 alphabets are given below.

47 Alphabets

$N=6$		
ϕ_i	w_i	$\phi_i w_i$
	Number of occurrences	
0	28	0
2	18	36
4	1	4
	47	40

$\bar{\phi}=40/47=0.85$

Figure 15.

	$N=6$
$E(\phi)$ Monoalphabetic text	1. 98
Observed $\bar{\phi}$	. 85
$E(\phi)$ Random text	1. 14

FIGURE 16.

o. Similar results are obtained by assuming 46, 45, . . . alphabets. Consider however, the results for an assumption of 43 alphabets. The message as written in 43 alphabets, and the values of ϕ for each alphabet are given in figure 17.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	H	S	K	U	S	P	M	F	H	D	U	J	J	I	X	M	S	P	T	P	O	I
2	G	A	O	P	G	P	R	H	B	O	U	C	S	H	P	V	G	H	Q	X	Z	S
3	W	Z	X	V	L	I	J	H	W	A	R	L	K	F	I	J	S	L	T	M	H	K
4	B	X	Z	V	L	I	J	R	Y	A	C	D	W	E	J	M	S	C	A	F	P	I
5	X	N	Z	J	G	N	U	D	O	A	R	F	U	E	R	J	O	Y	B	D	O	K
6	N	Y	N	B	D	V	Q	O	B	E	G	G	S	H	Q	H	X	O	P	U	Z	C
7	G	B	U	B	B	U	N	O	V	T	P	O	V	M	I	Z	D	E	P	Q	F	V
ϕ	2	0	2	4	4	4	2	4	2	6	4	0	2	4	2	4	6	0	4	0	4	4

	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
1	P	C	I	W	K	Z	V	U	Y	P	P	N	E	U	S	A	I	G	B	O	O
2	A	C	K	R	K	V	B	G	H	M	V	S	F	R	Y	Y	T	K	H	K	V
3	A	H	Q	T	U	V	T	X	S	M	E	C	F	C	S	K	T	G	O	O	Y
4	E	A	X	O	K	A	Q	D	W	E	X	P	Y	P	Q	H	D	N	O	J	I
5	E	I	K	D	U	V	T	D	V	E	V	L	E	T	D	O	A	F	R	O	U
6	O	C	U	K	K	Z	L	T	P	H	K	R	T	C	C	O	A	S	B	Z	U
7	K	Z																			
ϕ	4	6	2	0	14	8	2	2	0	4	2	0	4	2	2	2	4	2	4	6	2

FIGURE 17.

The distribution of the observed values of ϕ is given below in figure 18.

43 ALPHABETS

$N=7$			$N=6$		
ϕ_i	w_i	$\phi_i w_i$	ϕ_i	w_i	$\phi_i w_i$
	Number of occurrences			*Number of occurrences*	
0	4	0	0	3	0
2	6	12	2	9	18
4	11	44	4	4	16
6	3	18	6	1	6
	24	74	8	1	8
			14	1	14
				19	62

$\bar{\phi}=74/24=3.08$ $\bar{\phi}=62/19=3.26$

FIGURE 18.

	N=7	N=6
$E(\phi)$ Monoalphabetic text	2.77	1.98
Observed $\bar{\phi}$	3.08	3.26
$E(\phi)$ Random text	1.60	1.14

FIGURE 19.

We omit the detailed analysis used in subparagraph *k* above as it seems quite clear from figure 19 that the indications are that 43 alphabets were used in the encipherment of the message.

p. Writing out the generatrices for the first three columns on the assumption of normal alphabets,[20] there is obtained the following:

```
 H G W B X N G      S A Z X N Y B      K O X Z Z N U
 I H X C Y O H      T B A Y O Z C      L P Y A A O V
 J I Y D Z P I      U C B Z P A D      M Q Z B B P W
 K J Z E A Q J      V D C A Q B E      N R A C C Q X
 L K A F B R K      W E D B R C F      O S B D D R Y
 M L B G C S L      X F E C S D G      P T C E E S Z
 N M C H D T M      Y G F D T E H      Q U D F F T A
*O N D I E U N      Z H G E U F I      R V E G G U B
 P O E J F V O      A I H F V G J      S W F H H V C
 Q P F K G W P      B J I G W H K      T X G I I W D
 R Q G L H X Q      C K J H X I L      U Y H J J X E
 S R H M I Y R      D L K I Y J M      V Z I K K Y F
 T S I N J Z S      E M L J Z K N      W A J L L Z G
 U T J O K A T      F N M K A L O      X B K M M A H
 V U K P L B U      G O N L B M P      Y C L N N B I
 W V L Q M C V      H P O M C N Q      Z D M O O C J
 X W M R N D W      I Q P N D O R      A E N P P D K
 Y X N S O E X      J R Q O E P S      B F O Q Q E L
 Z Y O T P F Y      K S R P F Q T      C G P R R F M
 A Z P U Q G Z      L T S Q G R U      D H Q S S G N
 B A Q V R H A      M U T R H S V     *E I R T T H O
 C B R W S I B     *N V U S I T W      F J S U U I P
 D C S X T J C      O W V T J U X      G K T V V J Q
 E D T Y U K D      P X W U K V Y      H L U W W K R
 F E U Z V L E      Q Y X V L W Z      I M V X X L S
 G F V A W M F      R Z Y W M X A      J N W Y Y M T

                   *O N E . . .
                    N V I . . .
                    D U R . . .
                    I S T . . .
                    E I T . . .
                    U T H . . .
                    N W O . . .
```

FIGURE 20.

[20] See par. 45b—Elements of Cryptanalysis.

The complete message is as follows:

```
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43
ONEHUNDREDFIRSTFIELDARTILLERYFROMPOSITIONSI
NVICINITYOFBARLOWWILLBEINGENERALSUPPORTSTOP
DURINGATTACKSPECIALATTENTIONWILLBEPAIDTOASS
ISTINGADVANCEOFFIRSTBRIGADESTOPDURINGADVANC
EITWILLPLACECONCENTRATIONSONWOODSNORTHANDSO
UTHOFTHAYERFARMANDHILLSIXZEROEIGHTDASHAANDO
NWOODSEASTANDWESTTHEREOF
```

FIGURE 21.

SECTION VI

MATCHING ALPHABETS

19. **Nature of the problem.**—*a.* The analysis of the majority of cryptograms of the multi-alphabetic type reduces ultimately to a question of resolving the cryptographic text which is heterogeneous in composition (coming from several different cipher alphabets) into the homogeneous elements of monoalphabetic substitutions. If this can be accomplished the problem can practically always be solved, given sufficient time and patience.

b. Frequently, the reduction of the heterogeneous elements of the cryptogram to the simple terms of monoalphabetic substitutions involves the examination and detailed comparison (matching) of a multiplicity of frequency distributions to determine which of them present identical or nearly identical characteristics, (i. e., which match) and which can, therefore, be assumed to belong or apply to the same cipher alphabet. When the uniliteral frequency distributions are fairly large, say containing over 60 or 70 letters, this comparison is relatively easy for the experienced cryptanalyst and can be made by the eye; but when the distributions are small, each with a very limited number of letters, ocular examination and comparison is quite difficult and often inconclusive. In any event, the labor and time necessitated for the reduction of the text to its simplest terms, that is, the allocation of the letters to the respective cipher alphabets, is, in such cases, very considerable and makes the difference between a solution achieved in time to be of use and one that presents merely information of historical interest.

c. It will be shown that certain of the notions already discussed can be brought to bear upon this question, and thus by methods of mathematical comparison eliminate to a large degree the uncertainties of ocular examination and reduce the time required for cryptanalysis in many cases. It is advisable to emphasize at this point that there are limits to the size of alphabets to be matched below which the mathematical methods will not be effective. This is due to the fact that below a certain point the distribution of the values, calculated according to the tests to be described, for both properly matched and improperly matched alphabets overlap to such an extent that only very high or very low values are conclusive.

20. **Application of the ϕ test.**—*a.* If the uniliteral distributions of two monoalphabetic selections of text enciphered by means of the same substitution are aligned, they will show similar characteristics or match, i. e., frequent letters will correspond to frequent letters and infrequent letters or blanks will correspond to infrequent letters or blanks. In other words, the entire sequence of crests and troughs of the one distribution will correspond to the entire sequence of crests and troughs of the other distribution. If the two distributions are now combined into one by adding the frequencies of corresponding letters, the resulting distribution will be monoalphabetic in nature. Let us now extend this notion to the general case of non-random text. If there are aligned two non-random polygraphic frequency distributions, each of non-random polygraphic text, so that they match, then the resultant non-random polygraphic distribution obtained by combining the two distributions tested must also be non-random in nature. Accordingly the resultant non-random distribution should show the characteristic values discussed in paragraphs 16, 17, and 18. The value of N used is of course the sum of the number of elements in each of the two component non-random distributions.

b. Thus, if the two non-random distributions given by $f_1', f_2', \ldots, f_n'$, and $f_1'', f_2'', f_3'', \ldots, f_n''$, where $f_1'+f_2'+\ldots+f_n'=N_1$ and $f_1''+f_2''+\ldots+f_n''=N_2$ are combined into the one non-random distribution given by $f_1, f_2, \ldots, f_n$ where $f_1=f_1'+f_1''; f_2=f_2'+f_2''; \ldots; f_n=f_n'+f_n''$; and $f_1+f_2+\ldots+f_n=N_1+N_2=N$, then the expected value of $\phi=f_1(f_1-1)+f_2(f_2-1)+\ldots+f_n(f_n-1)$ is given by (18.1) and the variance of ϕ by (18.3).

For English monoalphabets these reduce to

(20.1) $$E(\phi)=0.066N(N-1)$$

(20.2) $$\sigma_\phi^2=0.004344N^3+0.110448N^2-0.114794N$$

(See chart 14, p. 169.)

c. If however the two distributions do not match, then

(20.3) $$E(\phi)=s_2N(N-1)-2N_1N_2(s_2-1/n)$$

(20.4) $$\sigma_\phi^2=(N_1^3+N_2^3)(4s_3-4s_2^2)+(N_1^2+N_2^2)(10s_2^2-12s_3+2s_2) + (N_1+N_2)(8s_3-6s_2^2-2s_2)+4N_1N_2\left[(N_1+N_2)\left(\frac{s_2}{n}-\frac{1}{n^2}\right) + \frac{1}{n}+\frac{1}{n^2}-\frac{2s_2}{n}\right]$$

where $N=N_1+N_2$, n is the number of different possible elements and s_2 and s_3 are the sum of the squares and cubes of the probabilities of occurrence of each of the possible elements.

For English monoalphabetic distributions which *do not match*, (20.3) and (20.4) become respectively,

(20.5) $$E(\phi)=0.066N(N-1)-0.056N_1N_2$$

(20.6) $$\sigma_\phi^2=0.004344(N_1^3+N_2^3)+0.110448(N_1^2+N_2^2)-0.114794(N_1+N_2) +4N_1N_2[(N_1+N_2)(0.001063)+0.034856]$$

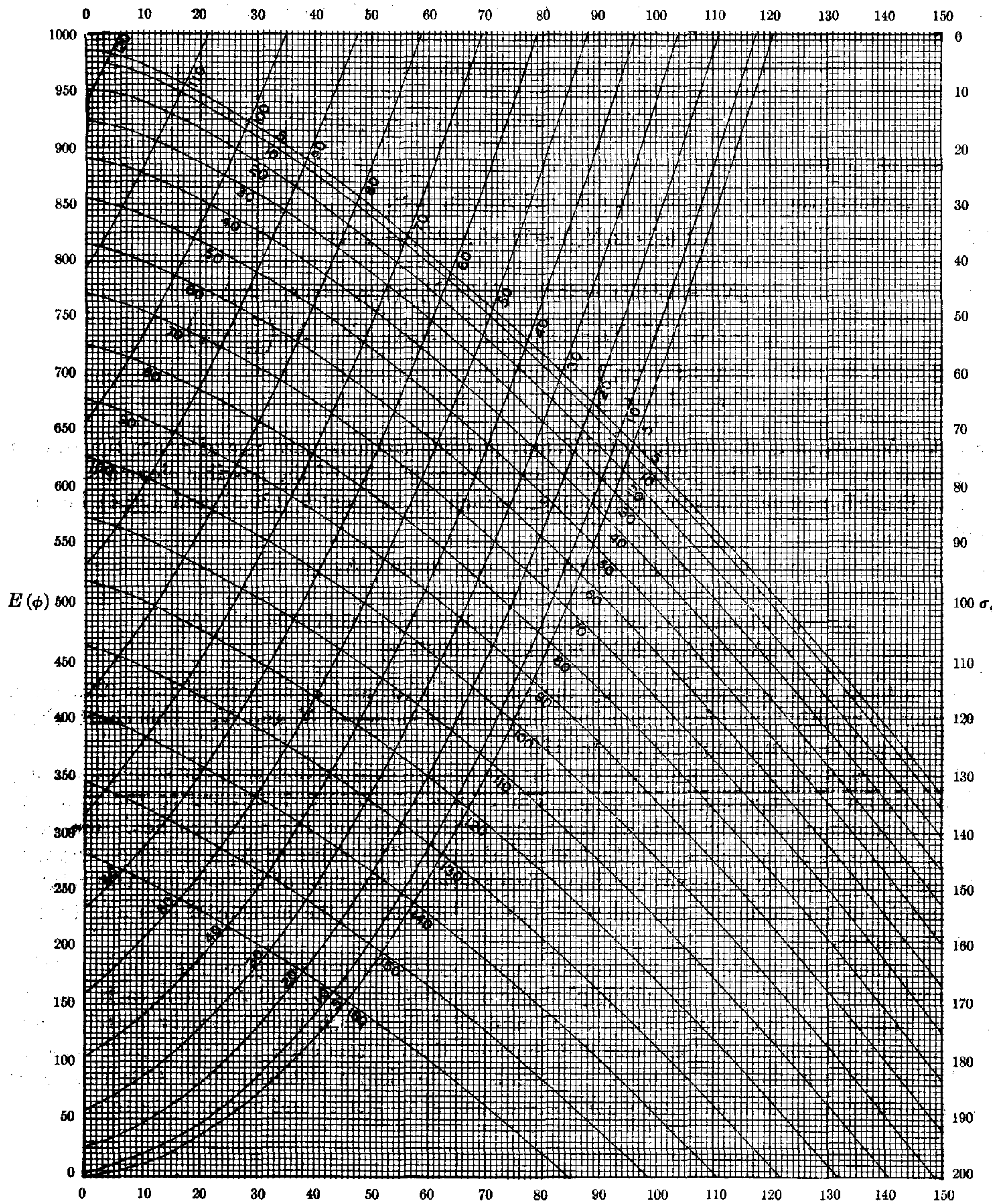
A chart has been prepared whereby the values of $E(\phi)$ and σ_ϕ as derived from (20.5) and (20.6) may be readily found for various combinations of values of N_1 and N_2. This chart, chart number 15, will be found on pages 50 and 170.

The curves proceeding upward to the right are used in conjunction with the scale on the left vertical axis for the expected value of ϕ. The curves proceeding downward to the right are used in conjunction with the scale on the right vertical axis for the standard deviation of ϕ.

The values of N_1 are given on the horizontal axis. The value of N_2 is given on the particular one of the family of curves corresponding thereto. Because of the symmetrical relation of N_1 and N_2 in the formulas, the value of N_2 may be read on the horizontal axis and that of N_1 on the curves.

d. In order to illustrate and, to a certain extent, check the preceding results experimentally, a study was made of 100 pairs of monoalphabetic distributions of 15 and 20 letters each. In one case the monoalphabetic distributions of 15 and 20 letters each were properly matched and combined to yield 100 monoalphabetic distributions of 35 letters each. In the second case the distributions were improperly matched and combined to yield a distribution of 35 letters made up of two different monoalphabetic distributions, one of 20 letters and one of 15 letters.

CHART NO. 15.—EXPECTED VALUE AND STANDARD DEVIATION OF ϕ
NON-MATCHING PAIRS OF MONOALPHABETS



(The value of N_2 is given on the curve corresponding thereto)

e. When the monoalphabetic distributions were properly matched and the value of $\phi=\sum_{i=1}^{n} f_i(f_i-1)$ calculated, the following were the observed values.

ϕ	Number of occurences	ϕ	Number of occurrences	ϕ	Number of occurrences
38	1	74	9	102	1
50	1	76	4	104	1
52	1	78	5	106	2
54	2	80	2	110	2
58	1	82	3	112	2
60	2	84	4	114	1
62	3	86	4	116	1
64	2	88	3	118	1
66	5	92	1	128	1
68	10	96	2		
70	7	98	5		100
72	7	100	4		

FIGURE 22.

From the above distribution it is calculated that the observed average value of ϕ is 79.7 and the observed standard deviation is 16.8. Using the value $N=35$, (20.1) and (20.2) yield as the expected mean and the expected standard deviation 78.5 and 18.0 respectively.

f. When the monoalphabetic distributions were improperly matched and the value of $\phi=\sum_{i=1}^{n} f_i(f_i-1)$ calculated, the following were the observed values.

ϕ	Number of occurrences	ϕ	Number of occurrences	ϕ	Number of occurrences
34	1	58	5	78	1
38	2	60	7	80	1
40	1	62	4	82	1
42	3	64	5	84	2
46	3	66	11	90	3
48	5	68	8	104	1
50	8	70	5	110	1
52	5	72	1		
54	2	74	2		100
56	8	76	4		

FIGURE 23.

From the foregoing distribution it is calculated that the observed average value of ϕ is 61.8 and the observed standard deviation is 13.4. From (20.5) and (20.6), for $N_1=20$ and $N_2=15$, it is found that the theoretical mean and standard deviation are 61.7 and 14.2 respectively.

g. The following table, figure 24, shows the overlapping of the distribution of observed values of ϕ as calculated from the correctly and incorrectly matched distributions. (The number of cases are progressively summed or given cumulatively.) (As the size of the distributions matched increases, the overlapping becomes smaller.) In other words, from figure 24 it is seen for example that 10 incorrectly matched distributions gave a value of $\phi=46$ or less whereas only 1 correctly matched pair gave a value of $\phi=46$ or less; 50 incorrectly matched distributions gave a value of $\phi=60$ or less whereas only 8 incorrectly matched distributions gave a value of $\phi=60$ or less.

ϕ	Correctly matched	Incorrectly matched	ϕ	Correctly matched	Incorrectly matched
34	0	1	78	60	91
38	1	3	80	62	92
40	1	4	82	65	93
42	1	7	84	69	95
46	1	10	86	73	95
48	1	15	88	76	95
50	2	23	90	76	98
52	3	28	92	77	98
54	5	30	96	79	98
56	5	38	98	84	98
58	6	43	100	88	98
60	8	50	102	89	98
62	11	54	104	90	99
64	13	59	106	92	99
66	18	70	110	94	100
68	28	78	112	96	
70	35	83	114	97	
72	42	84	116	98	
74	51	86	118	99	
76	55	90	128	100	

FIGURE 24.

21. **Cross-product sum or χ test.**—*a.* Suppose that the two distributions which it is desired to test for matching are given respectively by $f_1, f_2, \ldots, f_n$ and $f_1', f_2', \ldots, f_n'$ where $f_1+f_2+\ldots+f_n=N_1$ and $f_1'+f_2'+\ldots+f_n'=N_2$. Consider then the statistic χ defined by

(21.1) $$\chi=f_1f_1'+f_2f_2'+\ldots+f_nf_n'$$

(The definition of χ is not as arbitrary as may first appear, but is also related to the concept of coincidences which is discussed in Section VII. In paragraph 25*c* of that section the cross-product sum is again considered.)

It may be shown that if the two distributions are properly aligned and match, then [21]

(21.2) $$E(\chi)=s_2N_1N_2$$

and

(21.3) $$\sigma_\chi^2=N_1N_2[(N_1+N_2)(s_3-s_2^2)+s_2^2+s_2-2s_3]$$

[21] See appendix E, p. 154 ff.

where s_2 and s_3 are defined as in paragraph 20c. For English monoalphabetic text (21.2) and (21.3) become

(21.4) $$E(\chi)=0.066N_1N_2$$

(21.5) $$\sigma_\chi^2=N_1N_2[0.001086(N_1+N_2)+0.059569]$$

b. If the two distributions are not properly aligned and do not match then [22]

(21.6) $$E(\chi)=N_1N_2/n$$

(21.7) $$\sigma_\chi^2=N_1N_2\left[(N_1+N_2)\left(\frac{s_2}{n}-\frac{1}{n^2}\right)+\frac{1}{n}+\frac{1}{n^2}-\frac{2s_2}{n}\right]$$

For non-matching English monoalphabetic distributions (21.6) and (21.7) become

(21.8) $$E(\chi)=0.038N_1N_2$$

(21.9) $$\sigma_\chi^2=N_1N_2[0.001063(N_1+N_2)+0.034856]$$

Charts have been prepared to enable the values of $E(\chi)$ and σ_χ as derived from (21.4), (21.5), (21.8), and (21.9) for various combinations of N_1 and N_2 to be found readily. These charts, charts numbers 16 and 17, will be found on pages 54, 55 and 171, 172.

The curves originating in the lower left hand corner are used in conjunction with the scale on the left vertical axis for the expected value of χ. The curves originating in the upper left hand corner are used in conjunction with the scale on the right vertical axis for the standard deviation of χ.

The values of N_1 are given on the horizontal axis. The value of N_2 is given on the particular one of the family of curves corresponding thereto. Because of the symmetrical relation of N_1 and N_2 in the formulas, the value of N_2 may be read on the horizontal axis and that of N_1 on the curves.

c. If the test is applied to two random distributions, then

(21.10) $$E(\chi)=N_1N_2/n$$

(21.11) $$\sigma_\chi^2=N_1N_2[1/n-1/n^2]$$

For $n=26$, (21.10) and (21.11) become

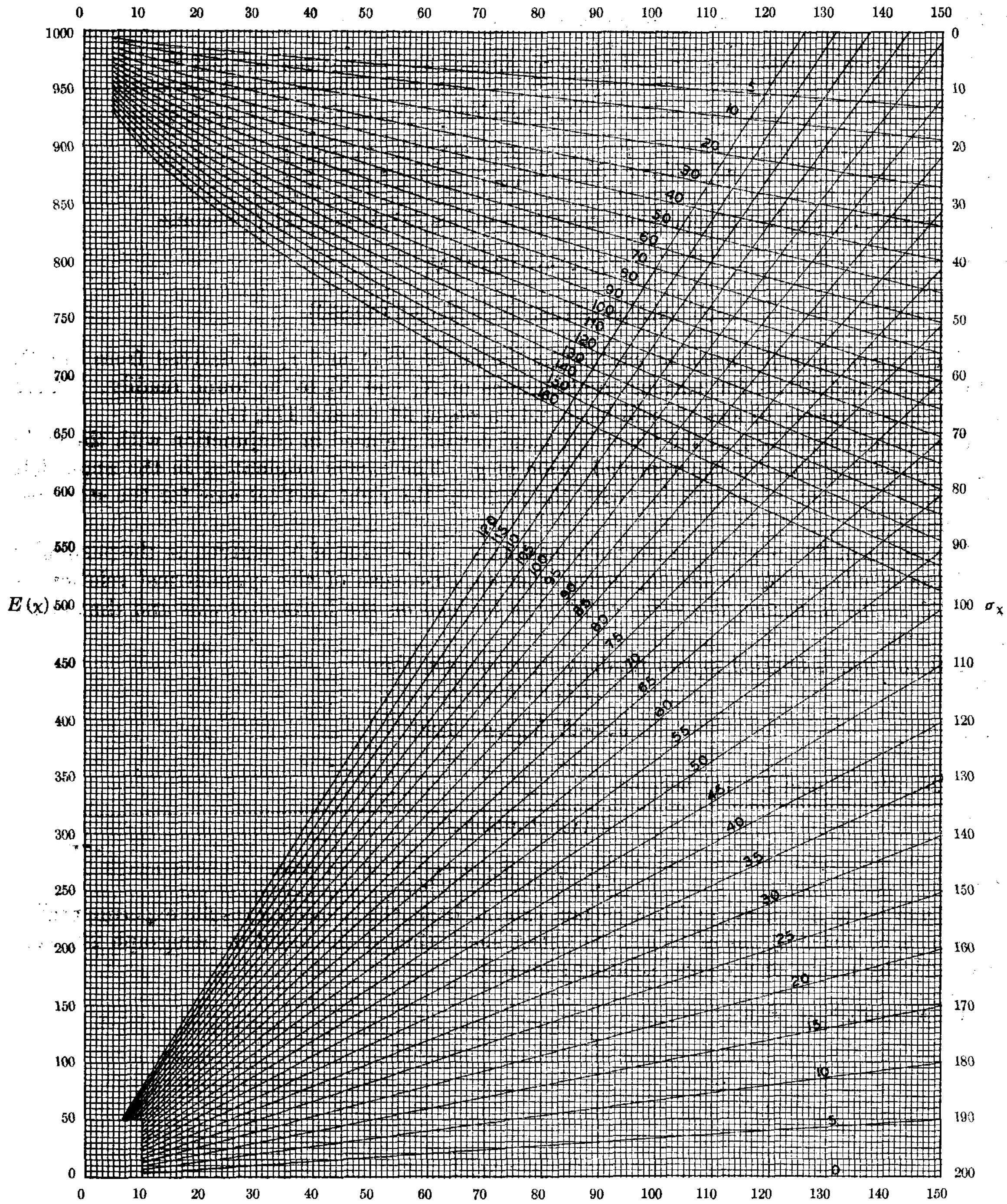
(21.12) $$E(\chi)=0.038N_1N_2$$

(21.13) $$\sigma_\chi^2=0.036982N_1N_2$$

d. In order to illustrate, and to a certain extent check, the preceding results experimentally, the 100 sets of distributions of 15 and 20 letters each, already discussed in paragraph 20, were also studied by means of the cross product sum test.

[22] See appendix F, p. 155 ff.

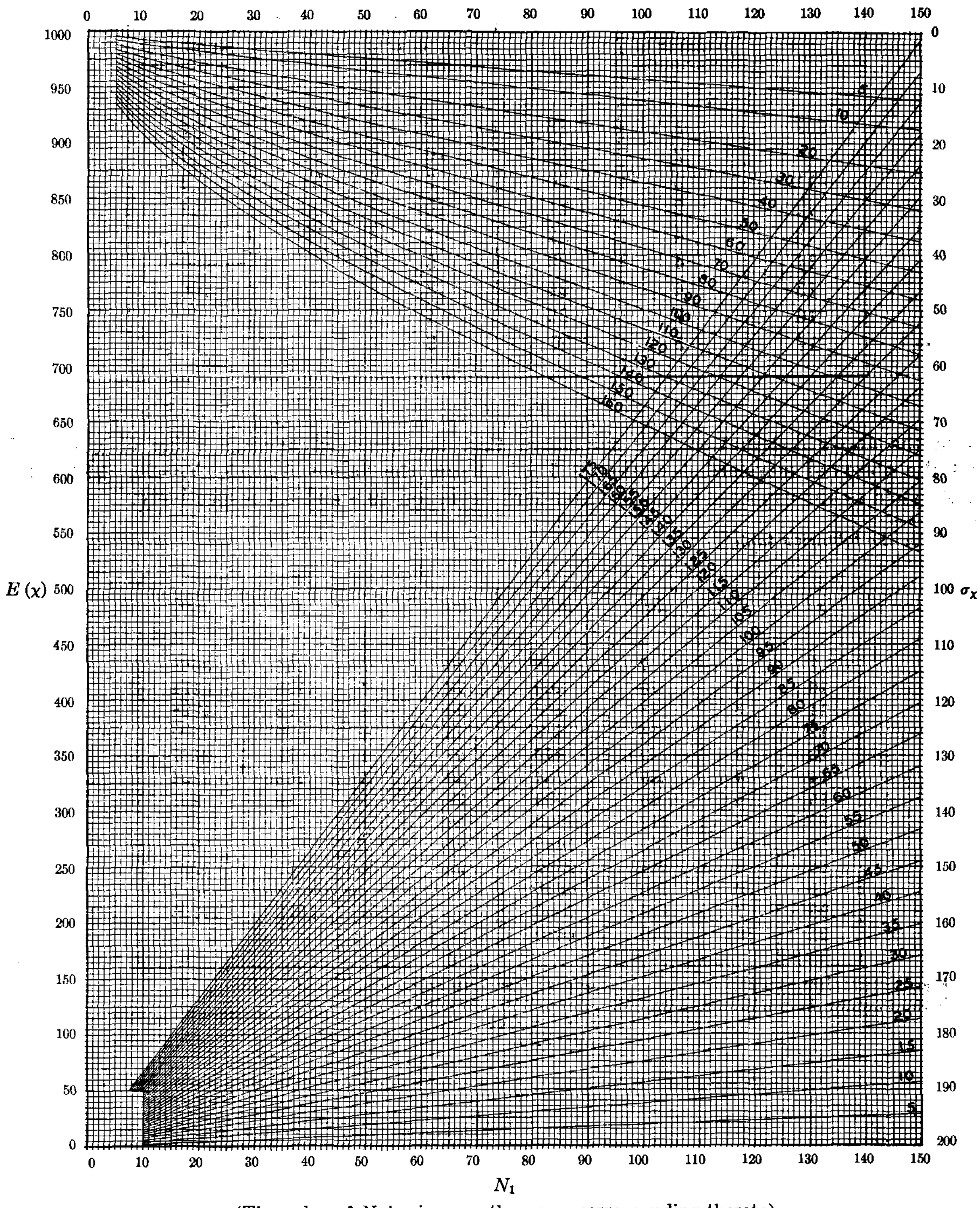
Chart No. 16.—Expected Value and Standard Deviation of χ, Matching Pairs of Monoalphabets



N_1

(The value of N_2 is given on the curve corresponding thereto)

CHART NO. 17.—EXPECTED VALUE AND STANDARD DEVIATION OF χ, NON-MATCHING PAIRS OF MONOALPHABETS



(The value of N_2 is given on the curve corresponding thereto)

e. When the alphabets were properly matched, and the value of $\chi=\sum_{i=1}^{n} f_i f_i'$ calculated, the following results were obtained:

χ	Number of occurrences	χ	Number of occurrences	χ	Number of occurrences
7	1	18	8	28	6
10	2	19	10	29	1
11	3	20	7	31	1
12	2	21	10	33	1
13	4	22	7	35	1
14	4	23	6		
15	4	24	3		100
16	7	25	4		
17	5	27	3		

FIGURE 25.

From the above distribution it is calculated that the observed average value of χ is 19.7 and the observed standard deviation is 5.3.[23] Using the values $N_1=15$, $N_2=20$, (21.3) and (21.4) yield as the expected mean and standard deviation 19.8 and 5.4 respectively.

[23] See the following table:

χ_i	f_i	$\chi_i f_i$	$\chi_i^2 f_i$	χ_i	f_i	$\chi_i f_i$	$\chi_i^2 f_i$
7	1	7	49	22	7	154	3388
10	2	20	200	23	6	138	3174
11	3	33	363	24	3	72	1728
12	2	24	288	25	4	100	2500
13	4	52	676	27	3	81	2187
14	4	56	784	28	6	168	4704
15	4	60	900	29	1	29	841
16	7	112	1792	31	1	31	961
17	5	85	1445	33	1	33	1089
18	8	144	2592	35	1	35	1225
19	10	190	3610				
20	7	140	2800		100	1971	41706
21	10	210	4410				

Mean = 1971/100 = 19.71.
Mean square = 41706/100 = 417.06.

$\sigma^2 = 417.06 - (19.71)^2$
$\sigma^2 = 417.06 - 388.4841$
$\sigma^2 = 28.5759$
$\sigma = 5.34$

f. When the alphabets were improperly matched, and the value of $\chi=\sum_{i=1}^{n} f_i f_i'$ calculated, the following were the observed values.

χ	Number of occurrences	χ	Number of occurrences	χ	Number of occurrences
2	1	10	11	17	3
4	2	11	10	18	3
5	5	12	5	20	1
6	7	13	10	23	2
7	8	14	5		
8	6	15	5		100
9	10	16	6		

FIGURE 26.

From the above distribution it is calculated that the observed average value of χ is 10.9 and the observed standard deviation is 4.1.[24] Using the values of $N_1=15$ and $N_2=20$, (21.8) and (21.9) yield as the expected mean and standard deviation 11.4 and 4.7 respectively.

g. The following table (fig. 27) shows the overlapping of the distributions of observed values of χ as calculated from the correctly and incorrectly matched distributions. The number of cases is given cumulatively.

In other words, from figure 27 it is seen, for example, that 23 incorrectly matched pairs gave a value of $\chi=7$ or less, whereas only 1 correctly matched pair gave a value of $\chi=7$ or less;

[24] See the following table:

χ_i	f_i	$\chi_i f_i$	$\chi_i^2 f_i$	χ_i	f_i	$\chi_i f_i$	$\chi_i^2 f_i$
2	1	2	4	13	10	130	1690
4	2	8	32	14	5	70	980
5	5	25	125	15	5	75	1125
6	7	42	252	16	6	96	1536
7	8	56	392	17	3	51	867
8	6	48	384	18	3	54	972
9	10	90	810	20	1	20	400
10	11	110	1100	23	2	46	1058
11	10	110	1210				
12	5	60	720		100	1093	13657

Mean $=1093/100=10.93$.
Mean square $=13657/100=136.57$

$\sigma^2=136.57-(10.93)^2=136.57-119.4649$
$\sigma^2=17.1051$
$\sigma=4.13$

100 incorrectly matched pairs gave a value of $\chi=23$ or less, whereas only 80 correctly matched pairs gave a value of $\chi=23$ or less.

χ	Correctly matched	Incorrectly matched	χ	Correctly matched	Incorrectly matched
2	0	1	19	50	97
3	0	1	20	57	98
4	0	3	21	67	98
5	0	8	22	74	98
6	0	15	23	80	100
7	1	23	24	83	
8	1	29	25	87	
9	1	39	26	87	
10	3	50	27	90	
11	6	60	28	96	
12	8	65	29	97	
13	12	75	30	97	
14	16	80	31	98	
15	20	85	32	98	
16	27	91	33	99	
17	32	94	34	99	
18	40	97	35	100	

FIGURE 27.

22. **Comparison of the two tests.**—*a.* It is desirable to compare the two tests just described to decide whether one of them is, in general, a better one to use for the particular purpose of matching alphabets. To do so we shall compare the results obtained from the same pairs of alphabets by both tests; the overlapping of the distributions for correctly and incorrectly matched pairs; and also the closeness with which the observed distributions are approximated by the theoretical normal distribution.

b. In figures 28 and 29 are given the values of χ and ϕ for the same pairs of alphabets for both correct and incorrect matching. Thus, for correctly matched pairs, one pair gave a value for χ of 7 and a value for ϕ of 68; two pairs gave a value for χ of 16 and a value for ϕ of 68; etc. Qualitatively, it may be seen from both tables that small values of ϕ correspond to small values of χ and that large values of ϕ correspond to large values of χ.

c. The lines drawn between 110 and 112 of the ϕ coordinates and between 23 and 24 of the χ coordinates represent the limits beyond which the observed values of ϕ and χ for the *improperly* matched pairs did not occur. It is most interesting that all the values of ϕ above 110 correspond to values of χ above 23. Furthermore, only 6 of the observed values of ϕ for correctly matched pairs lie beyond the upper limit of observed values of ϕ for incorrectly matched pairs, whereas 20 of the observed values of χ lie beyond the upper limit of observed values of χ for incorrectly matched pairs. In other words, if we used 112 as a lower limit for values of ϕ indicating a correct match or an upper limit of ϕ indicating an incorrect match, and 24 as a lower limit for values of χ indicating a correct match or an upper limit of χ indicating an incorrect match, the χ test would have yielded 14 more pairs than the ϕ test. The evidence here is in favor of the χ test.

Correct Match

ϕ

x	38	50	52	54	58	60	62	64	66	68	70	72	74	76	78	80	82	84	86	88	92	96	98	100	102	104	106	110	112	114	116	118	128	
7										1																								1
10	1										1																							2
11					1			1		1																								3
12									1			1																						2
13			1	1					1						1																			4
14						1	1		1				1																					4
15				1		1				1			1																					4
16		1					1		1	2								1		1														7
17											1	1	1			1			1															5
18							1			2	1	1	2		1																			8
19									1	2	2		1	1	1		1	1																10
20								1		1		2	1				1		1															7
21											1	1	1	2	2				1				1	1										10
22											1	1	1							1			2					1						7
23																	1	2	1			1	1											6
24														1		1					1													3
25																				1		1				1						1		4
27																							1	1			1							3
28																								2	1		1	1			1			6
29																													1					1
31																														1				1
33																													1					1
35																																	1	1
	1	1	1	2	1	2	3	2	5	10	7	7	9	4	5	2	3	4	4	3	1	2	5	4	1	1	2	2	2	1	1	1	1	100

Figure 28.

Incorrect Match

ϕ

χ	34	38	40	42	46	48	50	52	54	56	58	60	62	64	66	68	70	72	74	76	78	80	82	84	90	104	110	
2								1																				1
4								1				1																2
5	1				1				1	1					1													5
6		1			1		1	1		1	1				1													7
7				2			1	1			1	1		1		1												8
8			1	1		1									1		1			1								6
9						2	2	1		1		2		1		1												10
10		1			1		1			2				2	2				1	1								11
11						1	3		1	1		1				1	1								1			10
12										1	1		1			1	1											5
13										1	1		3		1	1		1		1	1							10
14											1			1	1	1	1											5
15						1										2						1	1					5
16												2			2									1	1			6
17															2		1											3
18																			1	1					1			3
20																											1	1
23																								1		1		2
	1	2	1	3	3	5	8	5	2	8	5	7	4	5	11	8	5	1	2	4	1	1	1	2	3	1	1	100

Figure 29.

d. Using the theoretical values for the mean and standard deviation the corresponding normal distributions were calculated and compared with the observed distributions. Here again, the observed χ distributions were much better approximated by the theoretical distributions than were the ϕ distributions. The result for the distribution of χ is given in figure 30.

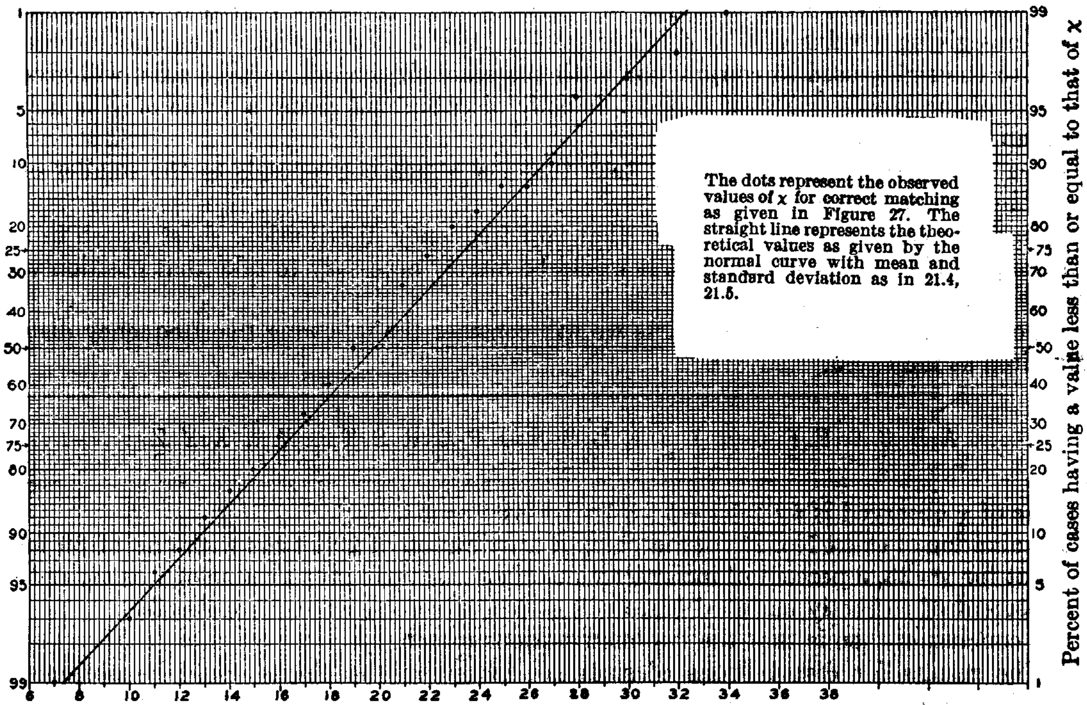


FIGURE 30.

e. It is quite clear from figures 24 and 27 that the distributions of ϕ for properly and improperly matched alphabets overlap to a greater extent than do the distributions of χ for properly and improperly matched alphabets. This is to be expected, since the theoretical mean values of ϕ for properly and improperly matched alphabets do not differ relatively as much as do the theoretical mean values of χ for properly and improperly matched alphabets. It may also be added that the ϕ test when used for matching alphabets involves the determination as to whether a given distribution is made up of one or two monoalphabets.

f. We must therefore conclude that the χ test is to be preferred to the ϕ test insofar as matching alphabets is concerned.

23. **Application of the cross-product sum test.**—*a.* In order to facilitate the use of the χ test certain charts have been prepared. These charts were prepared on specially ruled paper so designed that the graph of the normal probability curve is a straight line. The values of the means and standard deviations used were obtained from (21.4), (21.8), and (21.5), (21.9) respectively. These charts tell for certain sizes of paired alphabets what percentage of incorrectly matched alphabets would yield a value of χ as large or larger than that observed; and what percentage of correctly matched alphabets would yield a value of χ as small or smaller than that observed. In other words, given an observed value of χ, the charts will enable the cryptanalyst to judge at a glance the relative position of the matched alphabets as regards the distributions of correct and incorrectly matched alphabets and thus enable him to estimate the validity of his matching. These charts, charts Nos. 18–35 inclusive, will be found on pages 63–80.

Each chart is for a particular size of one of the matched distributions and the values for the size of the other of the matched distributions are indicated on the particular curve of the family corresponding thereto. The observed value of the cross product sum is to be found on the horizontal axis. The lines proceeding downwards to the right, used in conjunction with the scale on the left vertical axis, will give the percentage of correctly matched monoalphabetic distributions giving a value of χ as small or smaller than that observed. The lines proceeding upwards to the right, used in conjunction with the scale on the left vertical axis, will give the percentage of incorrectly matched monoalphabetic distributions giving a value of χ as large or larger than that observed.

In using the charts, it is necessary to take that one which corresponds to the smaller of the two distributions matched.

b. In those cases where it is known that two frequency distributions *must* match in some one of the possible relative alignments, it would be merely necessary to take that position which yields the greatest value for χ.

c. To illustrate the use of the charts we will consider the following two frequency distributions, paired as indicated.

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

The value of χ observed is $4+2+1+1+8+4+1+4+8=33$.

Examination of chart No. 21 for matching a distribution of 20 letters with one of 30 letters shows that for $\chi=33$, 8 percent of incorrectly matched cases will give a value of χ as *big or bigger* and 21 percent of correctly matched cases will give a value of χ as *small or smaller.*

The conclusion then is that the two distributions match.

d. The results for distributions of sizes not given by the charts could be obtained by interpolation from the charts.

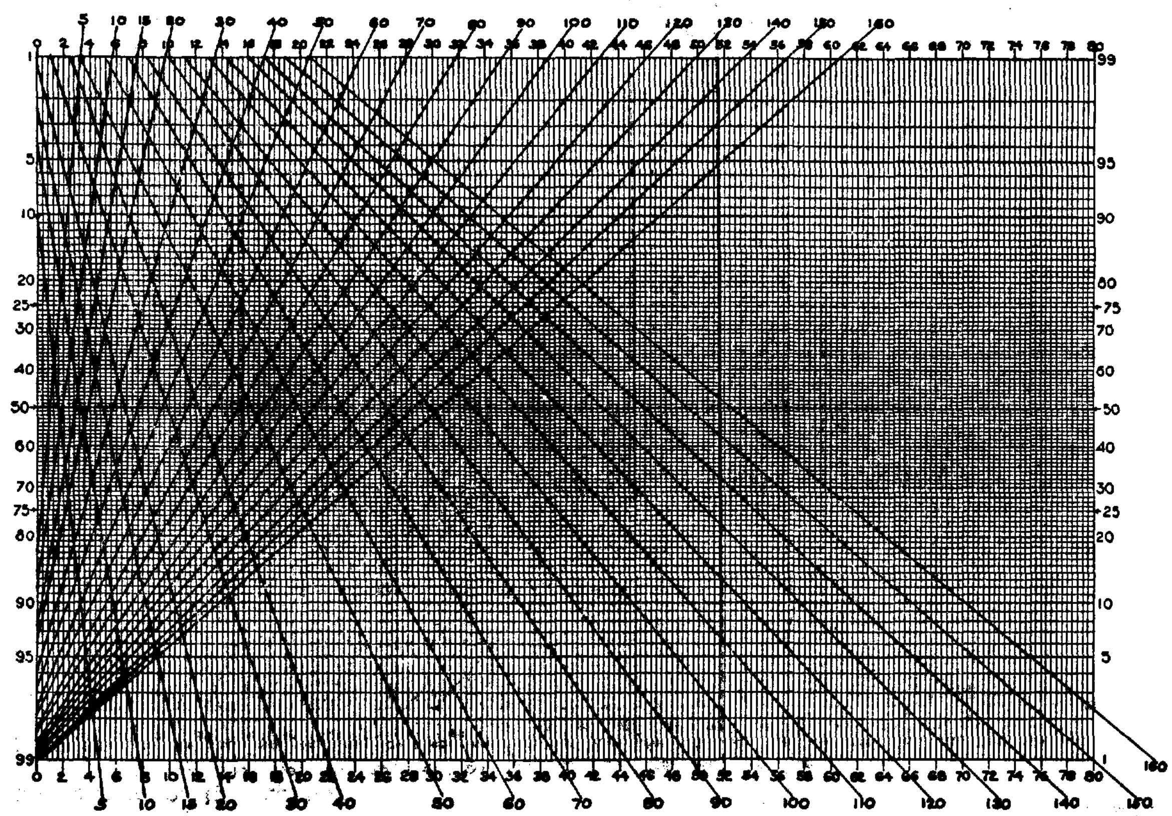
CHART No. 18

NUMBER OF LETTERS SECOND DISTRIBUTION
INCORRECT MATCH.

DISTRIBUTION OF 5 LETTERS.

OBSERVED VALUE OF
CROSS PRODUCT SUM.

PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.



DISTRIBUTION OF 5 LETTERS.

NUMBER OF LETTERS SECOND DISTRIBUTION
CORRECT MATCH.

PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.

PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.

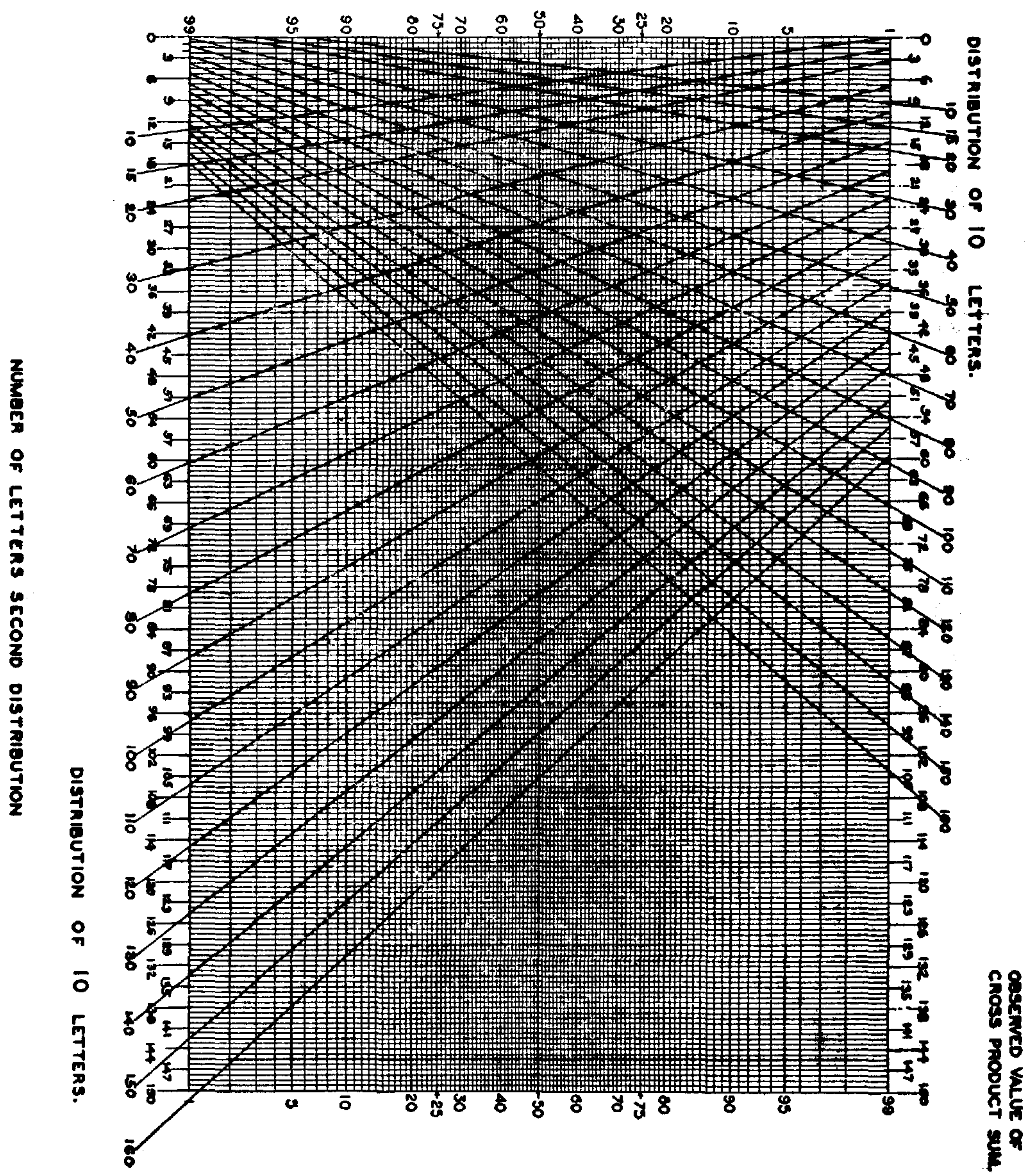


CHART No. 19

CHART No. 20

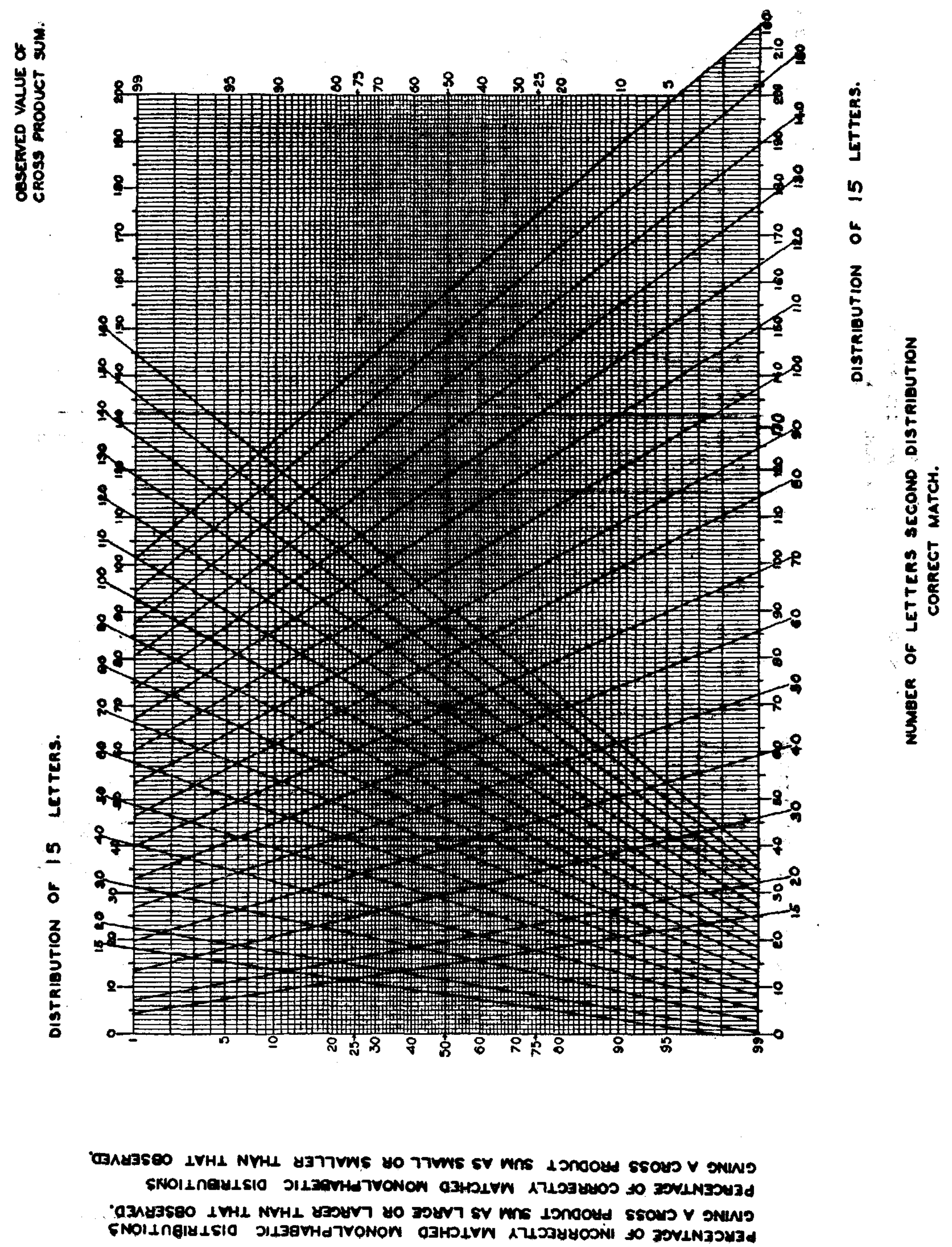

NUMBER OF LETTERS SECOND DISTRIBUTION
INCORRECT MATCH.
DISTRIBUTION OF 15 LETTERS.
OBSERVED VALUE OF
CROSS PRODUCT SUM.
DISTRIBUTION OF 15 LETTERS.
NUMBER OF LETTERS SECOND DISTRIBUTION
CORRECT MATCH.
PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.

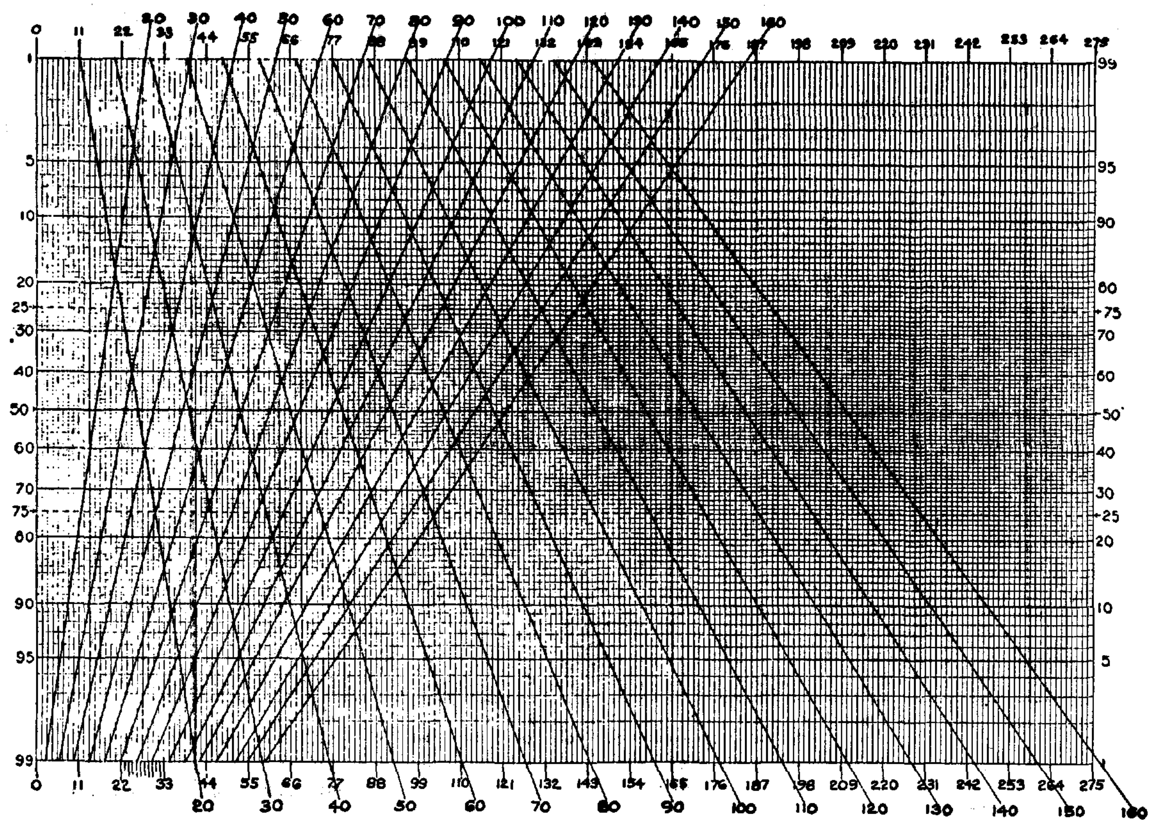
CHART No. 21

NUMBER OF LETTERS SECOND DISTRIBUTION
INCORRECT MATCH.

DISTRIBUTION OF 20 LETTERS.

OBSERVED VALUE OF
CROSS PRODUCT SUM.

PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.

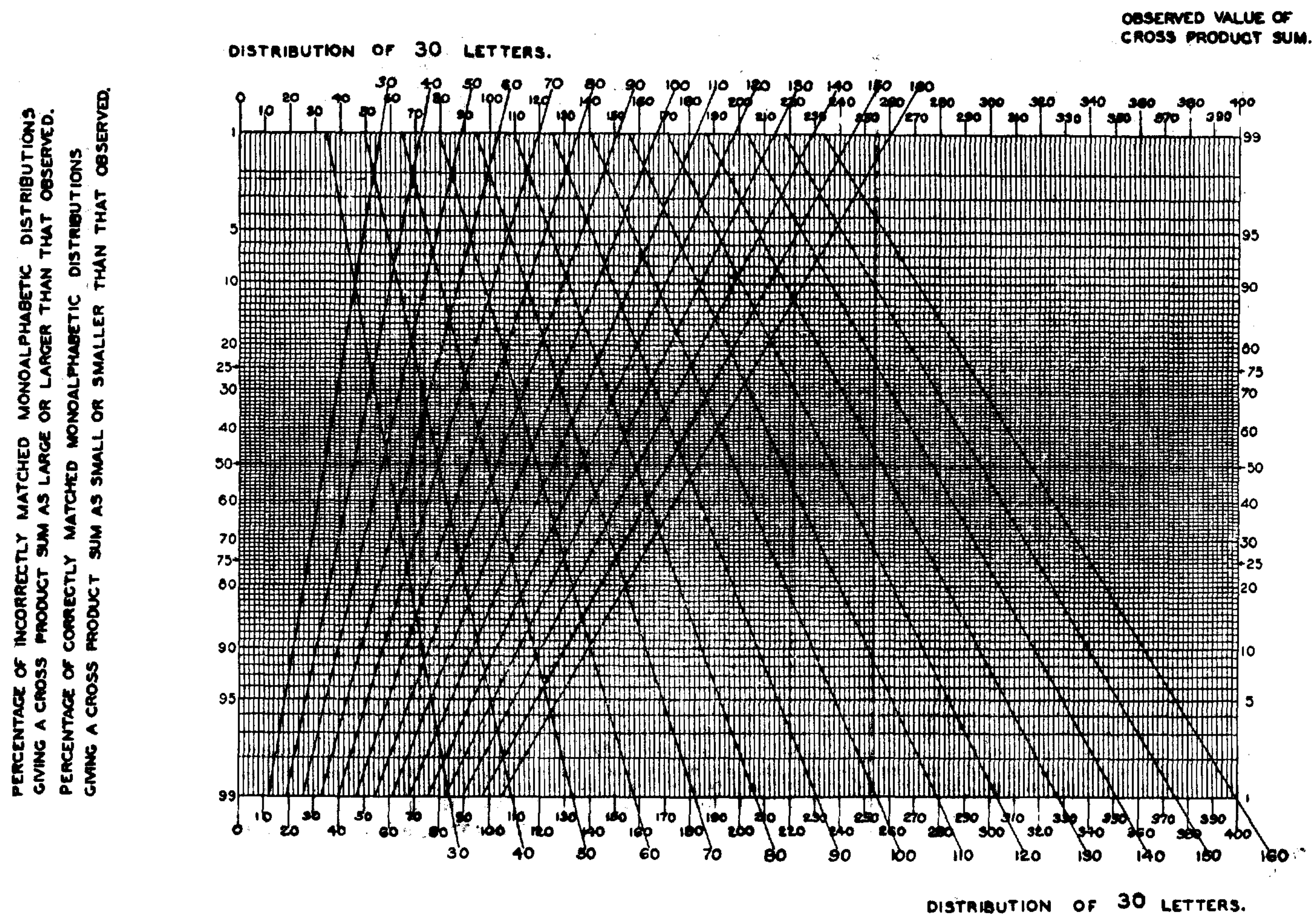


DISTRIBUTION OF 20 LETTERS.

NUMBER OF LETTERS SECOND DISTRIBUTION
CORRECT MATCH.

CHART No. 22

NUMBER OF LETTERS SECOND DISTRIBUTION
INCORRECT MATCH.



NUMBER OF LETTERS SECOND DISTRIBUTION
CORRECT MATCH.

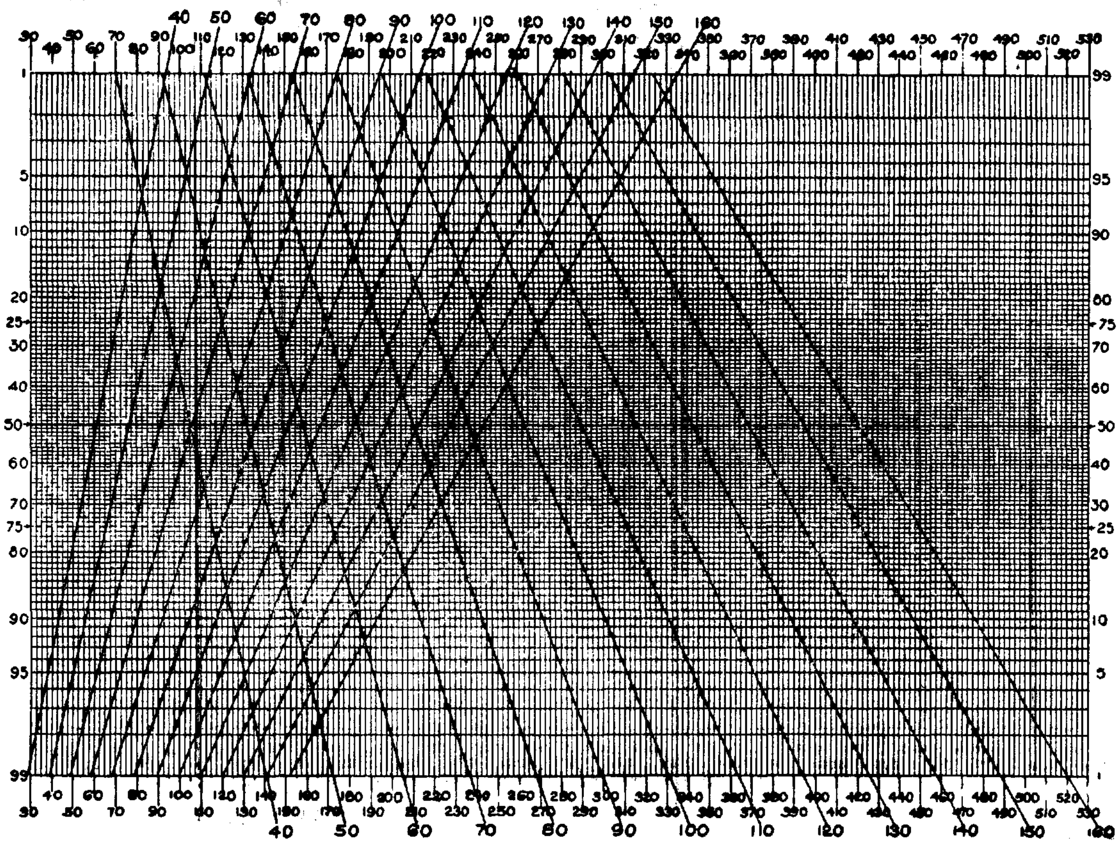
CHART No. 23

NUMBER OF LETTERS SECOND DISTRIBUTION
INCORRECT MATCH.

OBSERVED VALUE OF
CROSS PRODUCT SUM.

DISTRIBUTION OF 40 LETTERS.

PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.



DISTRIBUTION OF 40 LETTERS.

NUMBER OF LETTERS SECOND DISTRIBUTION
CORRECT MATCH.

CHART No. 24

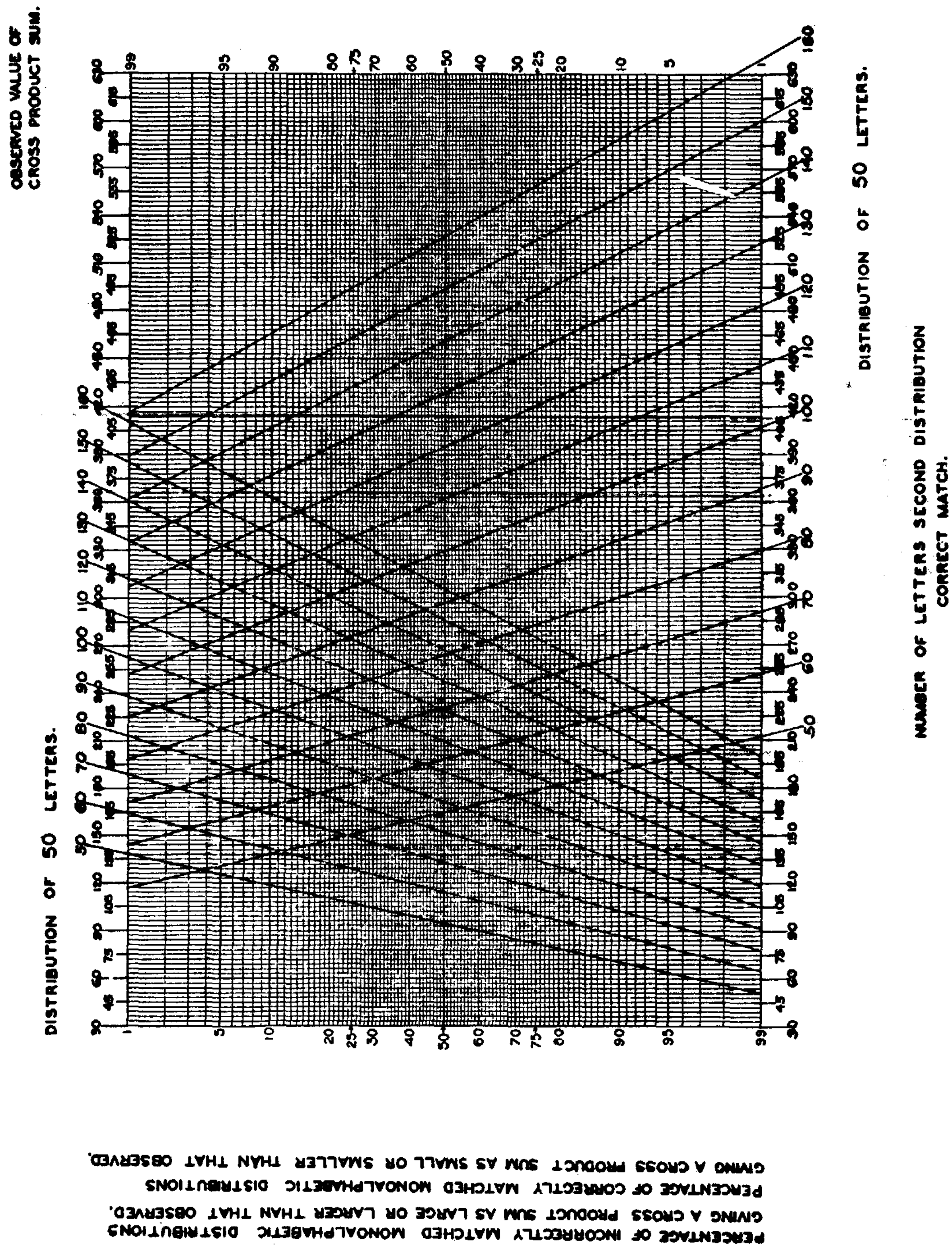
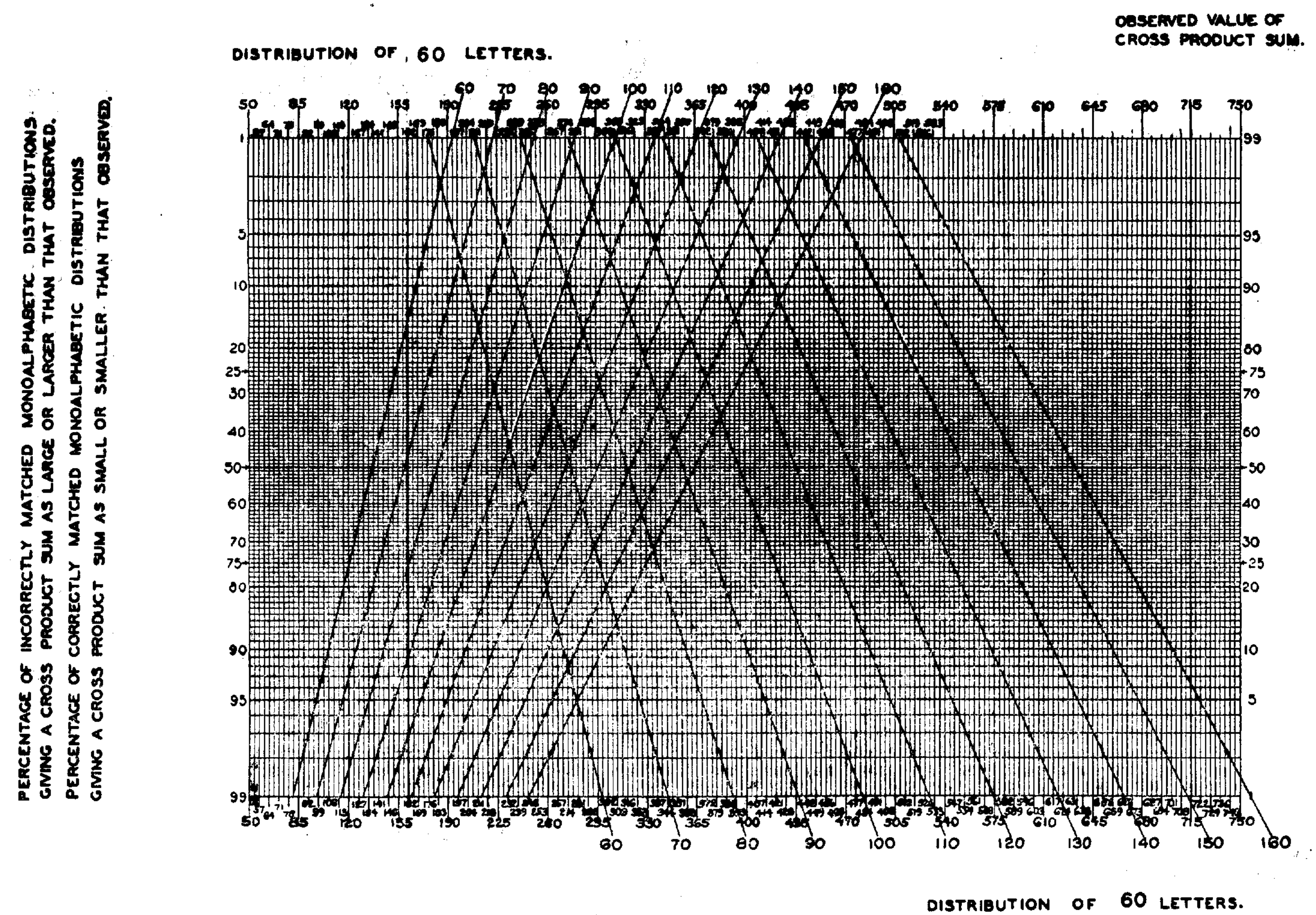

NUMBER OF LETTERS SECOND DISTRIBUTION INCORRECT MATCH.
DISTRIBUTION OF 50 LETTERS.
OBSERVED VALUE OF CROSS PRODUCT SUM.
DISTRIBUTION OF 50 LETTERS.
NUMBER OF LETTERS SECOND DISTRIBUTION CORRECT MATCH.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.
PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.

CHART No. 25

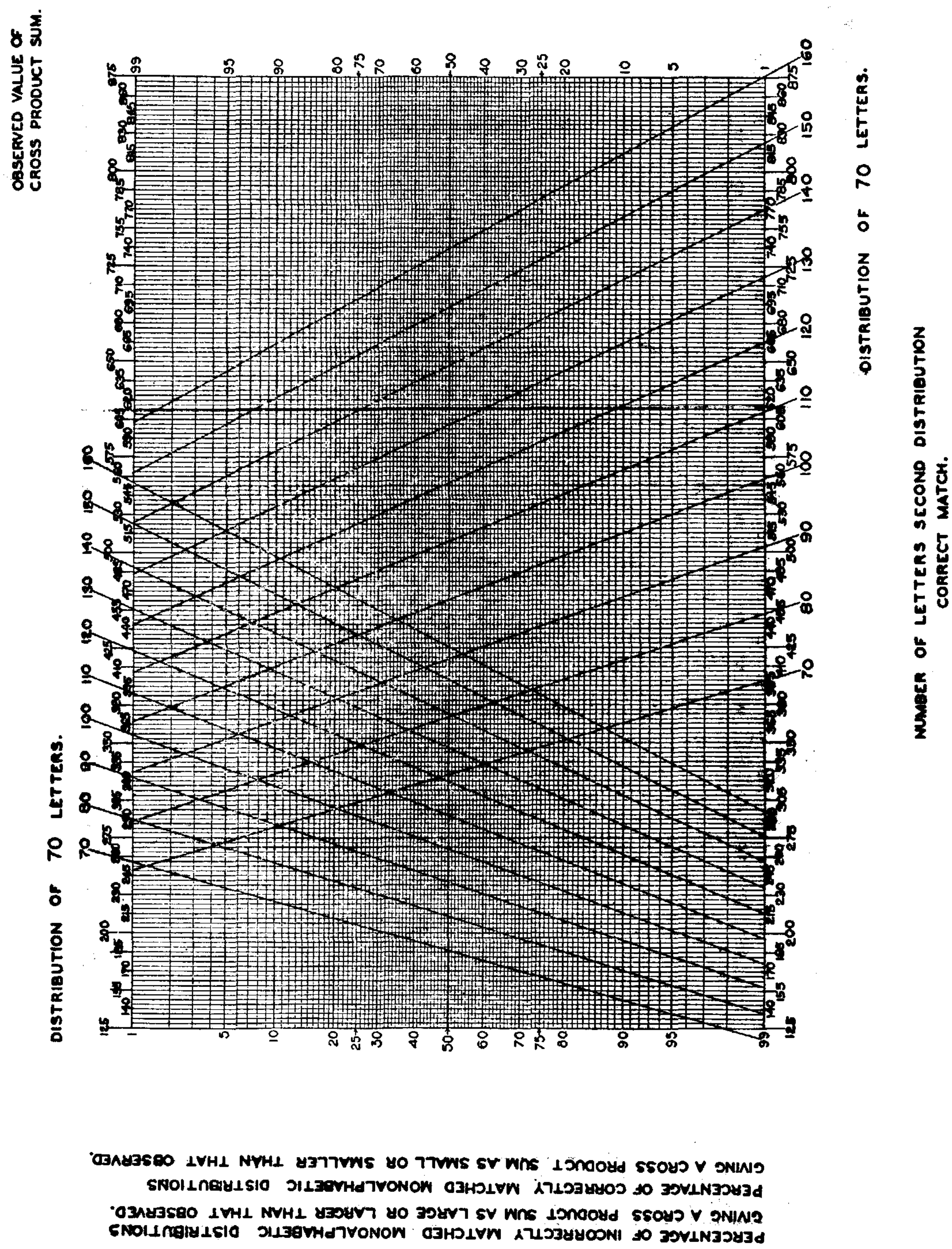
NUMBER OF LETTERS SECOND DISTRIBUTION
INCORRECT MATCH.



NUMBER OF LETTERS SECOND DISTRIBUTION
CORRECT MATCH.

CHART NO. 26

NUMBER OF LETTERS SECOND DISTRIBUTION INCORRECT MATCH.



PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.

PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.

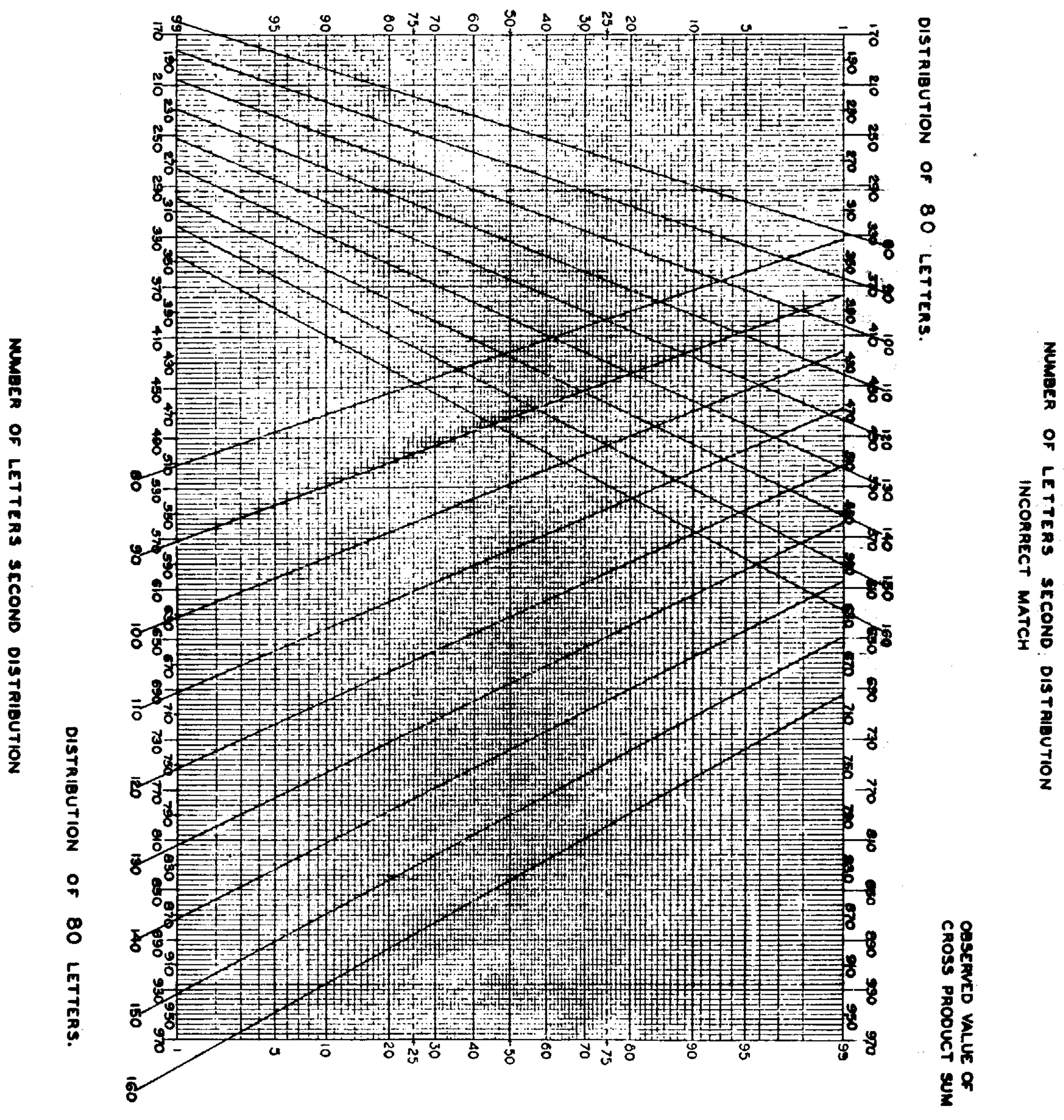
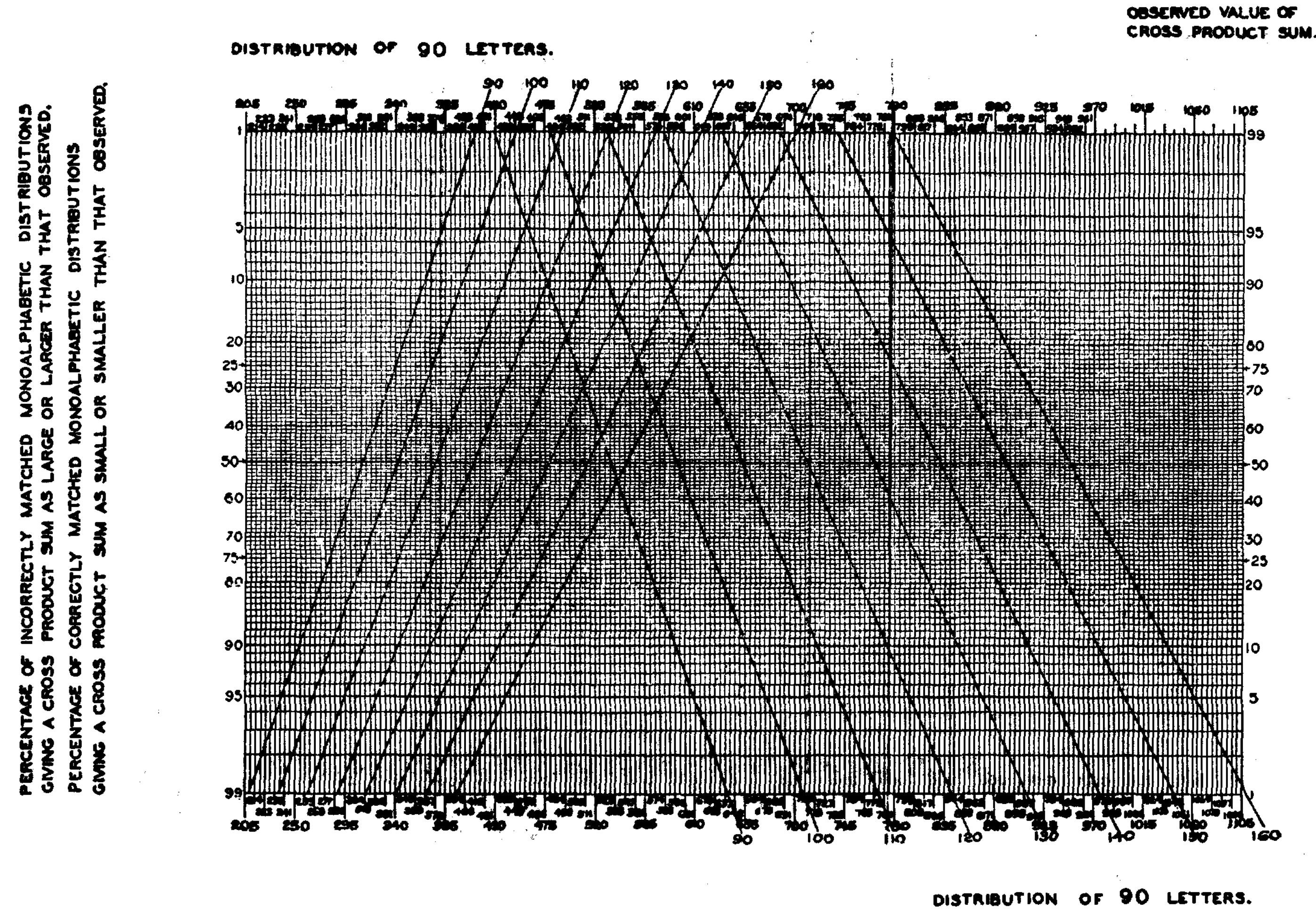


CHART No. 28

NUMBER OF LETTERS SECOND DISTRIBUTION
INCORRECT MATCH.



NUMBER OF LETTERS SECOND DISTRIBUTION
CORRECT MATCH.

CHART No. 29

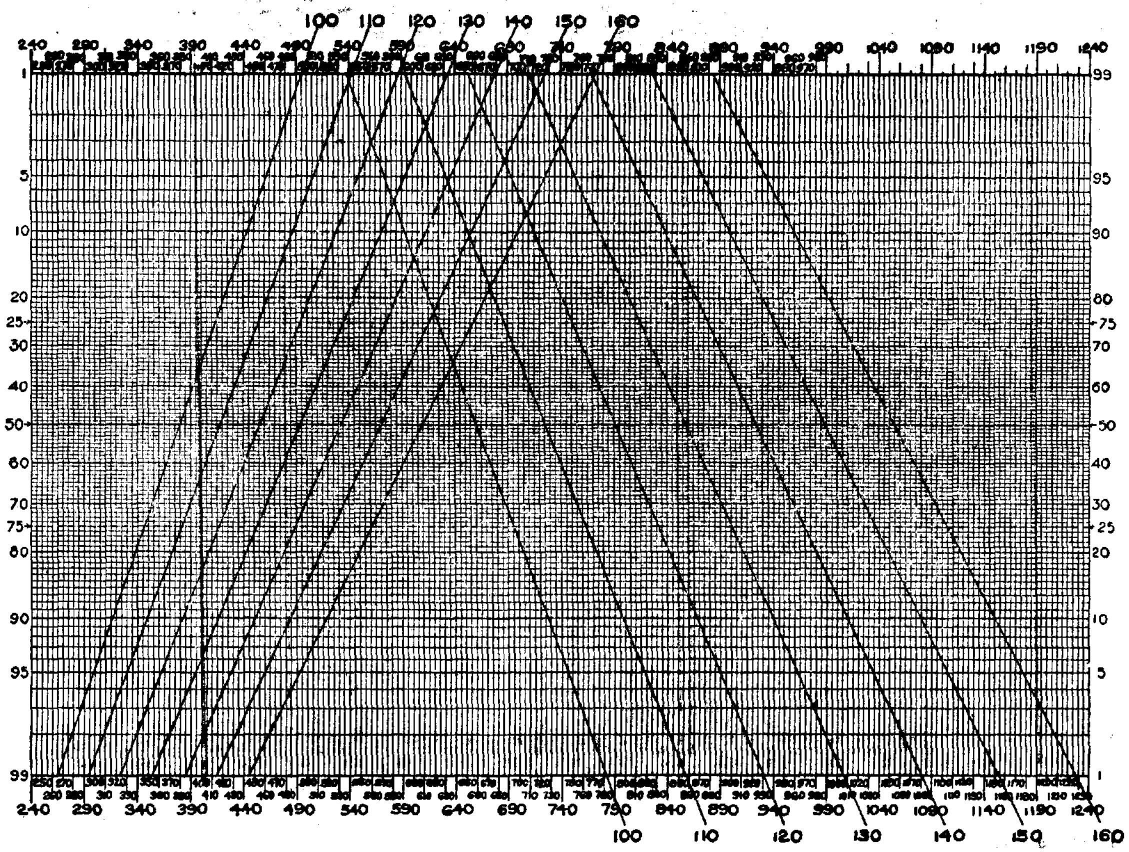
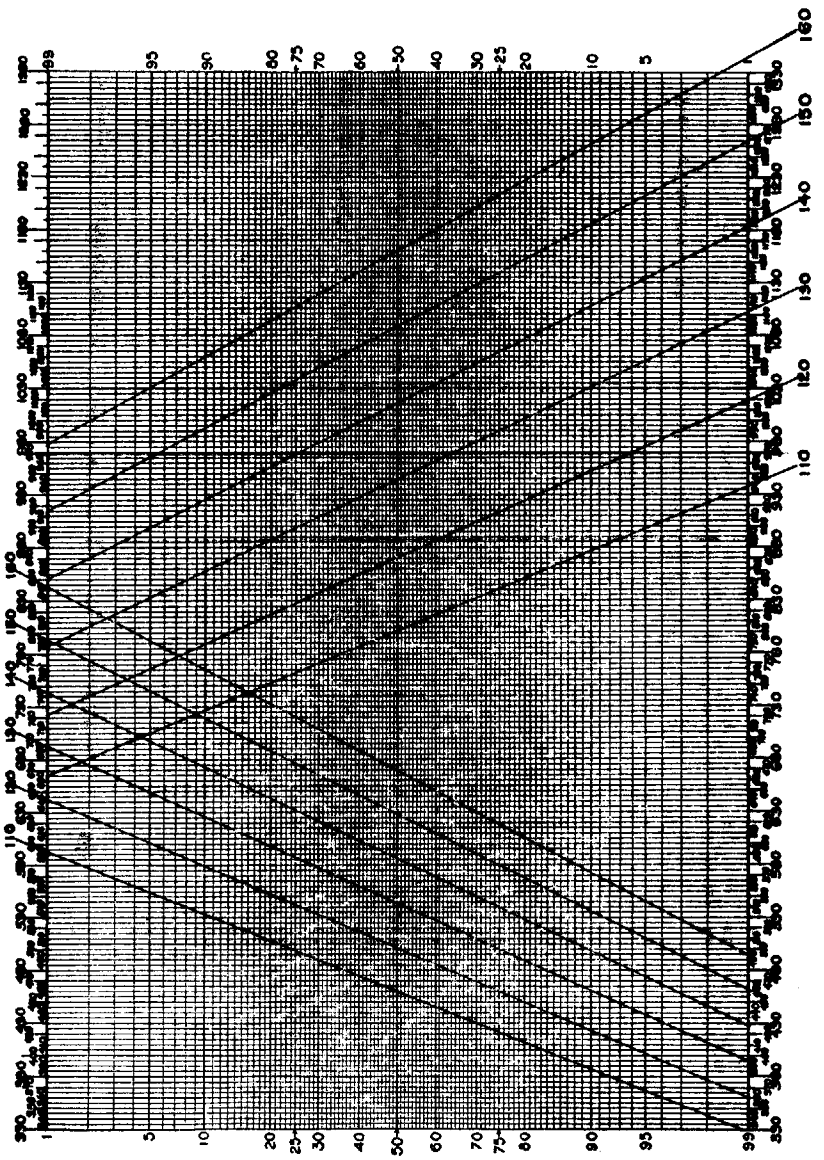

NUMBER OF LETTERS SECOND DISTRIBUTION
INCORRECT MATCH.
OBSERVED VALUE OF
CROSS PRODUCT SUM.
DISTRIBUTION OF 100 LETTERS.
100 110 120 130 140 150 160
240 290 340 390 440 490 540 590 640 690 740 790 840 890 940 990 1040 1090 1140 1190 1240
1 5 10 20 25 30 40 50 60 70 75 80 90 95 99
99 95 90 80 75 70 60 50 40 30 25 20 10 5 1
100 110 120 130 140 150 160
DISTRIBUTION OF 100 LETTERS.
NUMBER OF LETTERS SECOND DISTRIBUTION
CORRECT MATCH.
PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.

CHART No. 30

NUMBER OF LETTERS SECOND DISTRIBUTION
INCORRECT MATCH.

DISTRIBUTION OF 110 LETTERS.

OBSERVED VALUE OF
CROSS PRODUCT SUM.



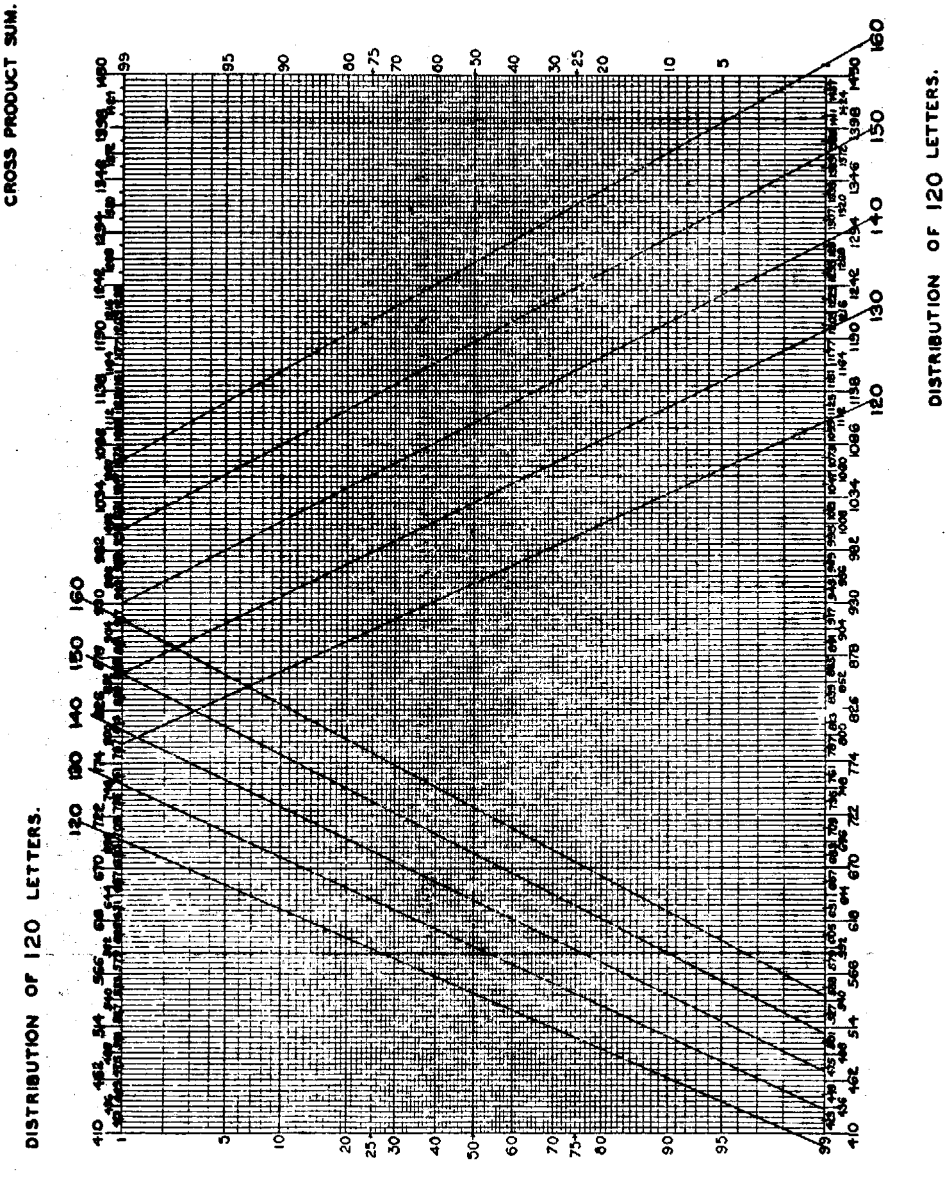
DISTRIBUTION OF 110 LETTERS.

NUMBER OF LETTERS SECOND DISTRIBUTION
CORRECT MATCH.

PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.

PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.

CHART No. 31



PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.

PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.

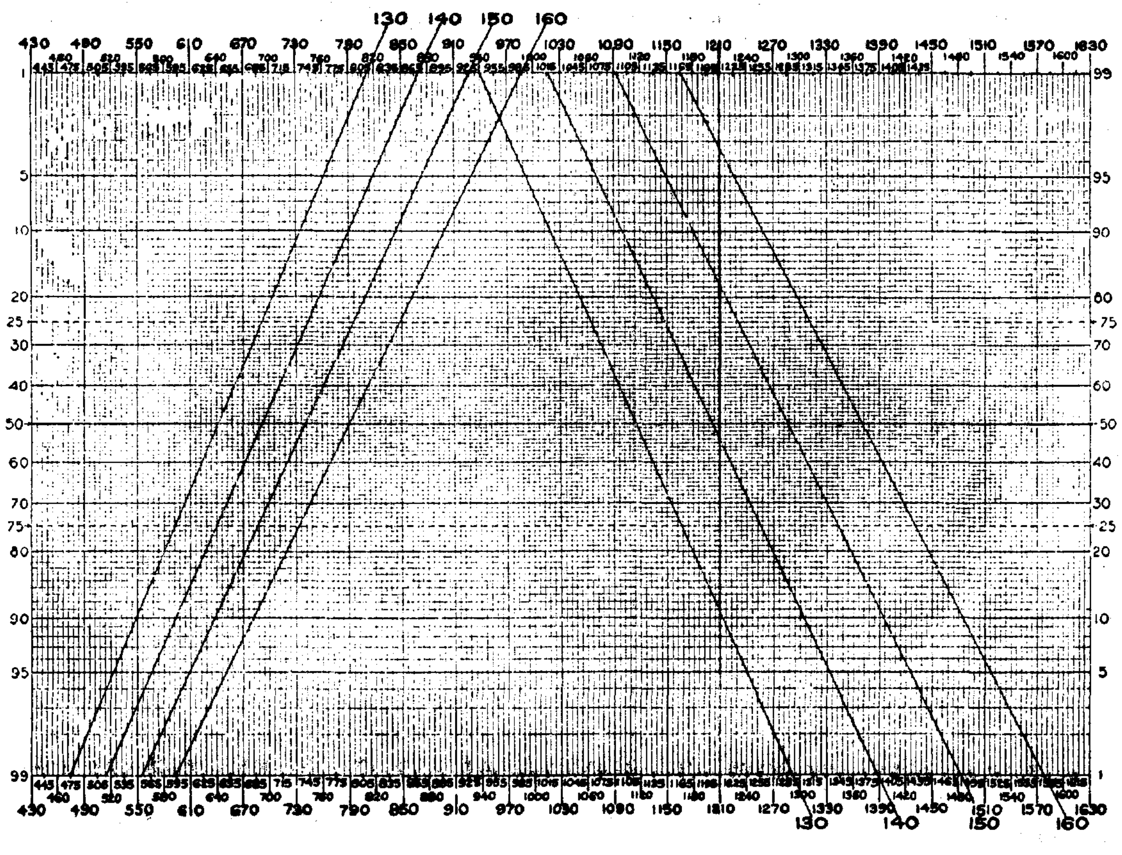
CHART No. 32

NUMBER OF LETTERS SECOND DISTRIBUTION
INCORRECT MATCH.

OBSERVED VALUE OF
CROSS PRODUCT SUM.

DISTRIBUTION OF 130 LETTERS.

PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.



DISTRIBUTION OF 130 LETTERS.

NUMBER OF LETTERS SECOND DISTRIBUTION
CORRECT MATCH.

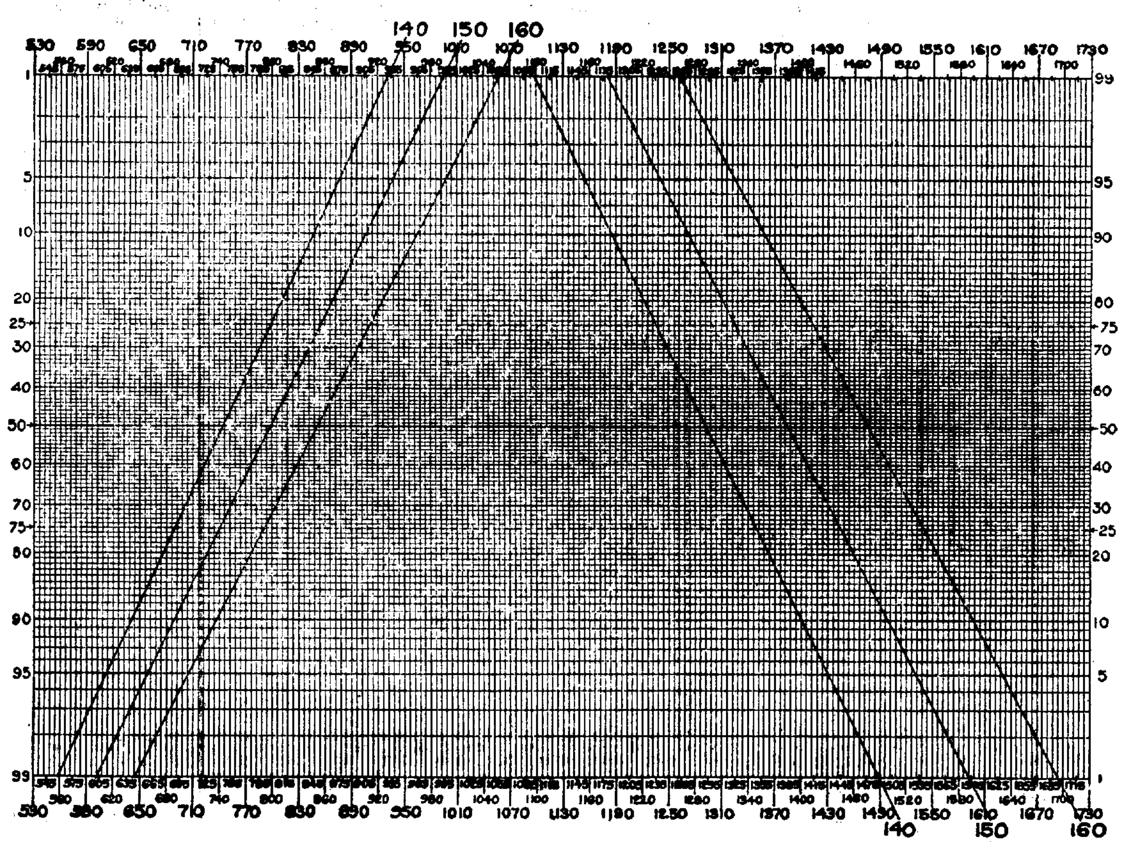
CHART No. 33

NUMBER OF LETTERS SECOND DISTRIBUTION
INCORRECT MATCH.

OBSERVED VALUE OF
CROSS PRODUCT SUM.

DISTRIBUTION OF 140 LETTERS.

PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.



DISTRIBUTION OF 140 LETTERS.

NUMBER OF LETTERS SECOND DISTRIBUTION
CORRECT MATCH.

CHART No. 34

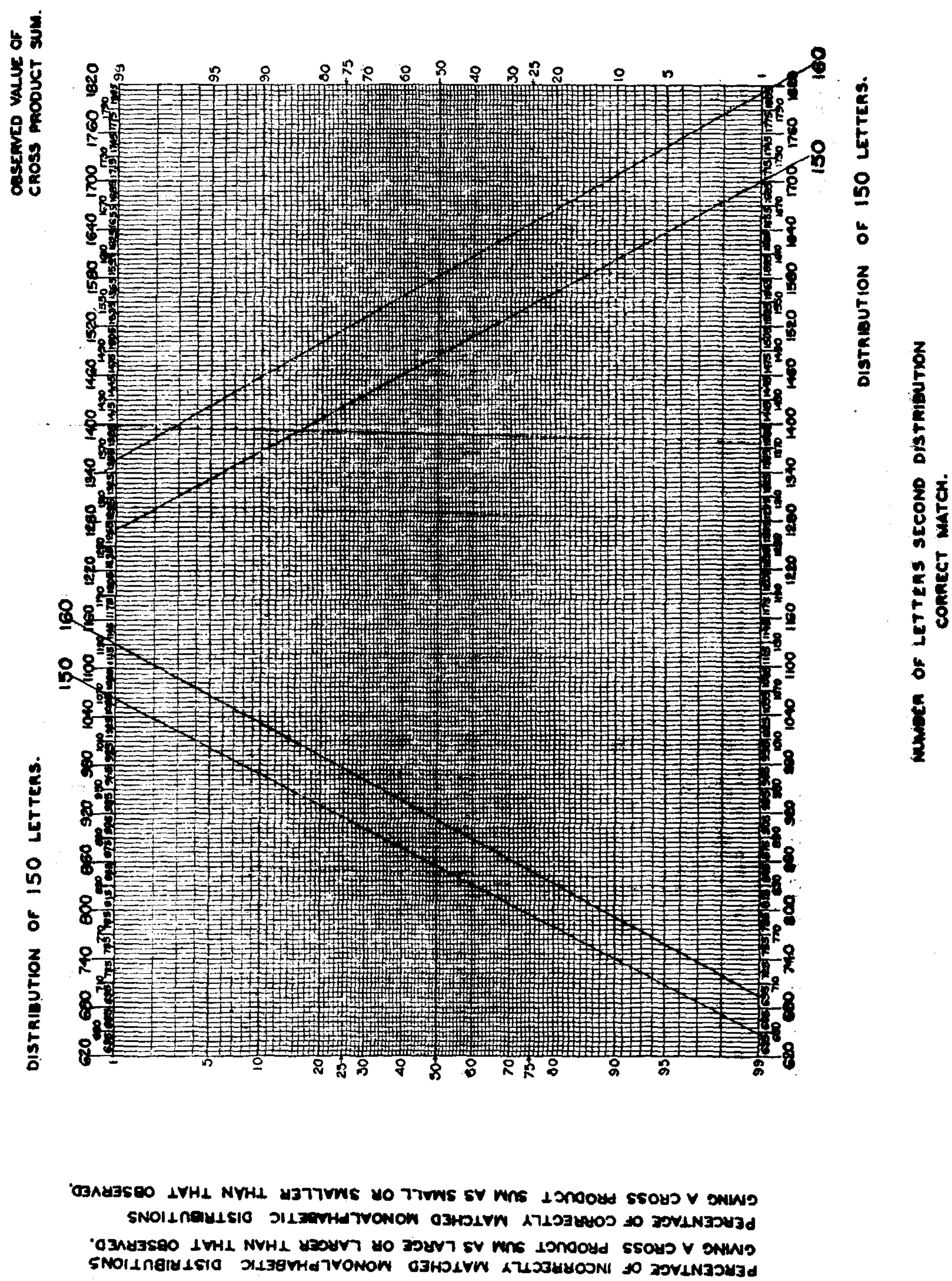

NUMBER OF LETTERS SECOND DISTRIBUTION INCORRECT MATCH.
DISTRIBUTION OF 150 LETTERS.
OBSERVED VALUE OF CROSS PRODUCT SUM.
DISTRIBUTION OF 150 LETTERS.
NUMBER OF LETTERS SECOND DISTRIBUTION CORRECT MATCH.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.
PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.

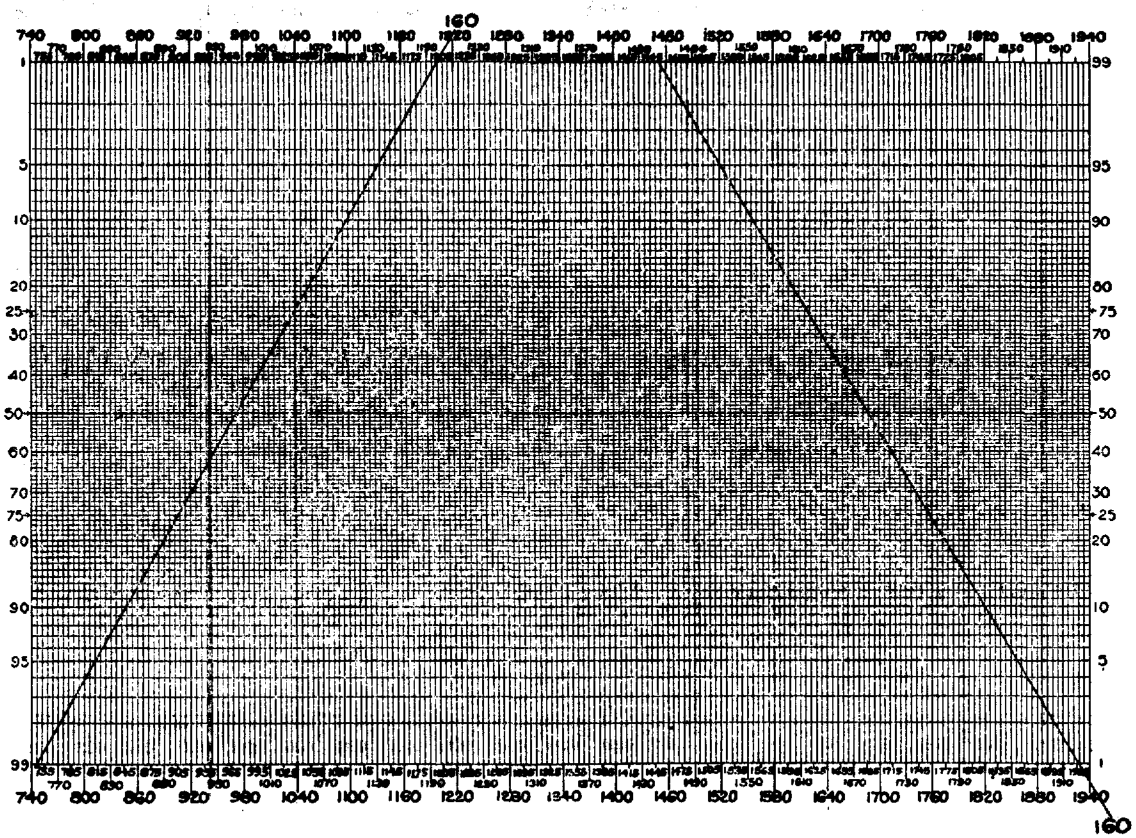
CHART No. 35

NUMBER OF LETTERS SECOND DISTRIBUTION
INCORRECT MATCH.

OBSERVED VALUE OF
CROSS PRODUCT SUM

DISTRIBUTION OF 160 LETTERS.

PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.



DISTRIBUTION OF 160 LETTERS.

NUMBER OF LETTERS SECOND DISTRIBUTION
CORRECT MATCH.

SECTION VII

COINCIDENCES

24. **General considerations.**—*a.* The concept of coincidences discussed in this section is a fundamental one in cryptanalysis and the application of statistical technique thereto. If any two selections of plain-text are superimposed, it will be found that a certain number of the letters in corresponding positions of the two messages are identical. If the text is written out in digraphs, trigraphs, etc., before superimposition, it will be found that a certain number of the digraphs, trigraphs, etc., in corresponding positions of the two messages are identical. Suppose now, that the selections of text are enciphered by a substitution system in such a manner that textual elements the same distance from the beginnings of the messages undergo the same enciphering process. If the resulting cryptograms are now superimposed the superimposed cipher texts will show identical elements in corresponding positions just as did the original text.[25]

b. The following considerations lead to the determination of the expected value of the ratio of coincidences (i. e. identical pairs) to the total possible number of pairs.

The probability of occurrence of a specified single letter in random text employing a 26-letter alphabet is $p=1/26=0.0385$. If a considerable volume of such text is written on a large sheet of paper and a pencil is directed at random toward this text, the probability that the pencil point will hit the letter A, *or any other letter which may be specified in advance,* is 0.0385. Now suppose two pencils are directed simultaneously toward the sheet of paper. The probability that both pencil points will hit two A's is $1/26 \times 1/26 = 1/26^2 = 0.00148$, since in this case one is dealing with the probability of the simultaneous occurrence of two events which are independent. The probability of hitting two B's, two C's, . . . , two Z's is likewise $1/26^2$. Hence, if no particular letter is specified, and merely this question is asked: "What is the probability that both pencil points will hit the same letter?" the answer must be the sum of the separate probabilities for simultaneously hitting 2 A's, 2 B's, and so on, for the whole alphabet, which is $26 \times 1/26^2 = 1/26 = 0.0385$. This, then, is the probability that *any* two letters selected *at random* in random text of a 26-letter alphabet will be identical or will *coincide.* Since this value remains the same so long as the number of alphabetic elements remains fixed, it may be said that *the probability of monographic coincidence in random text of a 26-element* alphabet is 0.0385. The foregoing italicized expression is important enough to warrant assigning a special symbol to it, *viz,* κ_r (read "kappa sub-*r*"). For a 26-element alphabet, then, $\kappa_r=0.0385$.

For random text employing *n* possible elements the probability of getting a particular element is $1/n$. The probability for the simultaneous occurrence of two of the same particular element is $1/n^2$. Accordingly the sum of the probabilities for the simultaneous occurrence of two of the same element is $n \times 1/n^2$ and $\kappa_r = 1/n$.

[25] It is interesting to note that a similar concept is the basis for the solution of transposition messages of identical length by anagramming. In transposition messages however it is not the property of "correspondence in value" which is invariant, but the property of "correspondence in position" which is invariant. Indeed, we might venture to define cryptanalysis as the solution of cryptograms by an analysis and application of the "invariant" characteristics of the cryptographic system employed. A cryptographic system which has no invariant characteristics would be secure against unauthorized decipherment.

Now consider the matter of monographic coincidence in English plain text. Following the same reasoning outlined above, the probability of coincidence of two A's in plain text is the square of the probability of occurrence of the single letter A in such text. The probability of coincidence of two B's is the square of the probability of occurrence of the single letter B, and so on. The sum of these squares for all the letters of the alphabet, as shown in the following table, is found to be 0.066. This then is the probability that any two letters selected at random in a large volume of normal English telegraphic plain text will coincide. Since this value remains the same so long as the character of the language does not change radically, it may be said that *the probability of monographic coincidence in English telegraphic plain text is* 0.066, or $\kappa_p = 0.066$.

p_i	p_i^2	p_i^3	p_i^4
0. 072	5184	373248	26832400
. 011	121	1331	14641
. 033	1089	35937	1188100
. 043	1849	79507	3422500
. 126	15876	2000376	252810000
. 030	900	27000	810000
. 018	324	5832	104976
. 033	1089	35937	1188100
. 076	5776	438976	33408400
. 002	4	8	16
. 004	16	64	256
. 035	1225	42875	1512900
. 025	625	15625	390625
. 076	5776	438976	33408400
. 074	5476	405224	30030400
. 027	729	19683	531441
. 003	9	27	81
. 083	6889	571787	47472100
. 058	3364	195112	11289600
. 090	8100	729000	65610000
. 030	900	27000	810000
. 013	169	2197	28561
. 014	196	2744	38416
. 005	25	125	625
. 020	400	8000	160000
. 001	1	1	1
1. 001	. 066112	. 005457	. 000511

c. The sum of the squares of the probabilities of occurrence of the various single letters, digraphs, etc., of a particular language is thus an important cryptographic property, and yields the probability for monographic coincidence, digraphic coincidence, etc.

d. In figure 31 are listed the probabilities for monographic and digraphic coincidence for plain text in several languages.

	κ_p Monographic	κ_p^2 Digraphic
English	0.0661	0.0069
French	.0778	.0093
German	.0762	.0112
Italian	.0738	.0081
Japanese (Romaji)	.0819	.0116
Portuguese	.0791	
Russian	.0529	.0058
Spanish	.0775	.0093

FIGURE 31.

For convenience the following values of the reciprocals of various numbers from 20 to 36, and of the reciprocals of the squares, cubes, and 4th powers of these numbers are listed:

x	$1/x$	$1/x^2$	$1/x^3$	$1/x^4$
20	0.0500	0.002500	0.000125	0.00000625
21	.0476	.002266	.000108	.00000514
22	.0455	.002070	.000094	.00000429
23	.0435	.001892	.000082	.00000358
24	.0417	.001739	.000073	.00000302
25	.0400	.001600	.000064	.00000256
26	.0385	.001482	.000057	.00000220
27	.0370	.001369	.000051	.00000187
28	.0357	.001274	.000046	.00000162
29	.0345	.001190	.000041	.00000142
30	.0333	.001109	.000037	.00000123
31	.0323	.001043	.000034	.00000109
32	.0313	.000980	.000031	.00000096
33	.0303	.000918	.000028	.00000084
34	.0294	.000864	.000025	.00000075
35	.0286	.000818	.000023	.00000067
36	.0278	.000773	.000021	.00000060

e. The distribution of the number of coincidences, for text properly superimposed, is in accordance with the binomial distribution $(p+q)^N$ where N is the total possible number of pairs and the values of p are as given in figure 31.

f. As we have already seen, the Poisson distribution or modified Poisson distribution offers a good approximation to the binomial distribution $(p+q)^N$ for values of p ranging as in the table above and N not very large, so that $m=Np \leqq 15$. For large values of N, the normal distribution with $m=Np$ and $\sigma^2=Npq$ will give a good enough approximation.

g. If the superimposed texts bear no relationship to one another, then the number of coincidences will be distributed in accordance with the binomial $(p+q)^N$ where N is the total possible number of pairs and $p=1/n$, with n the number of possible elements. For a 26-letter alphabet $p=1/26=0.038$ for single letters, $p=1/676=0.0015$ for digraphs, and $p=1/26^3=0.000057$ for trigraphs. The Poisson distribution offers a good approximation to the binomial $(p+q)^N$ for values of p corresponding to those just indicated.

h. The considerations outlined above thus enable the cryptanalyst to avail himself of repetitions of single letters and to evaluate the significance of such repetitions.

25. **Related tests.**—*a.* The tests already given for studying the random or non-random character of text and for matching alphabets are related to the concept of coincidences.

b. For, consider a monographic distribution. If a letter occurs f_i times, it is equivalent to $_i(f_i-1)/2$ coincidences. (The combinations of f_i things taken two at a time.) If there is a total of N letters in the distribution then there is possible a total of $N(N-1)/2$ pairs. Accordingly the expected value of

$$\frac{\frac{f_1(f_1-1)}{2}+\frac{f_2(f_2-1)}{2}+\cdots+\frac{f_n(f_n-1)}{2}}{\frac{N(N-1)}{2}}=s_2=\kappa_p \tag{25.1}$$

or as in (18.1)

$$E(\phi)=s_2N(N-1) \tag{25.2}$$

c. Consider now the problem of matching alphabets. If a letter occurs f_1 times in one distribution and f_1' times in the other, then the number of coincidences of that particular letter between the two distributions is f_1f_1'. Using the same notation as in paragraph 21, it is seen that

$$\chi=f_1f_1'+f_2f_2'+\cdots+f_nf_n' \tag{25.3}$$

gives the number of coincidences between the two distributions and that the total possible number of pairs is N_1N_2. Thus the expected value of

$$\frac{f_1f_1'+f_2f_2'+\cdots+f_nf_n'}{N_1N_2}=s_2=\kappa_p \tag{25.4}$$

or as in (21.2)

$$E(\chi)=s_2N_1N_2. \tag{25.5}$$

d. In the two cases discussed above, the distribution of the "number of coincidences" is not the binomial because the various coincidences are interrelated and are not independent as is required by the assumptions giving rise to the binomial distribution. Thus, in order to find the standard deviation of χ and ϕ it is necessary to apply a procedure which involves the fact that the simultaneous distribution of $f_1, f_2, \cdots, f_n$ is given by the multinomial distribution.

26. **Applications.**—*a.* In paragraph 18*j* it was indicated how the average of a number of ϕ tests could be employed to determine the number of alphabets used in a polyalphabetic message in which the number of alphabets is large and the number of letters per alphabet is small. We shall now show that the discussion in paragraphs 24*e*, 24*f*, and 24*g* is also directly applicable to the above mentioned problem.

b. Consider a rectangular array of letters of N columns and r rows. If the array of letters represents the polyalphabetic encipherment of English plain text with N alphabets then the expected number of coincidences between a pair of rows is Np where $p=0.066$. If the columns are random text the expected number of coincidences between a pair of rows is $N/26$. The r rows yield $R=r(r-1)/2$ pairs of rows so that we are enabled to find the average number of coincidences of R sets of N pairs of letters each. In accordance with the discussion in paragraph 9e we then have that the distribution of the average number of coincidences thus found is given by $(p+q)^{NR}$ with unit $1/R$, and that the mean and standard deviation are respectively given by $\mu=Np$ and $\sigma^2=Npq/R$.

c. Let us apply the foregoing to the message already considered in paragraph 18k.

In figure 7 there are found an array of 32 columns by 6 rows and an array of 18 columns by 5 rows. The observed average number of coincidences between the $R=6\times5/2=15$ sets of 32 pairs each is $\bar{c}=18/15=1.2$ and between the $R=5\times4/2=10$ sets of 18 pairs each is $\bar{c}=5/10=0.5$.[26]

Accordingly we have

N	R	Observed $\bar{c}$	50 alphabets		$n(n\neq50)$ alphabets	
			$E(\bar{c})$	$\sigma_{\bar{c}}$	$E(\bar{c})$	$\sigma_{\bar{c}}$
32	15	1.2	2.112	0.36	1.231	0.28
18	10	.5	1.188	.33	.692	.26

N	R	50 alphabets		$n(n\neq50)$ alphabets	
		$x=\dfrac{\bar{c}-E(\bar{c})}{\sigma_{\bar{c}}}$	$P(-\infty, x)$	$x=\dfrac{\bar{c}-E(\bar{c})}{\sigma_{\bar{c}}}$	$P(x, \infty)$
32	15	−2.53	0.0063	−0.11	0.5438
18	10	−2.08	.0179	−.74	.7704

The results obtained above are thus comparable to the results obtained in paragraph 18k and our conclusion is the same viz 50 alphabets were not used.

We will not continue the application of this procedure to the remaining cases, but suggest that the reader carry out the procedure for several possibilities.

[26] The value 18 is easily obtained as one-half the sum of the ϕ values of the first 32 columns and the value 5 as one-half the sum of the ϕ values of the last 18 columns.

d. Consider the 50 lines of text in figure 32.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	K	F	G	B	R	P	S	Y	K	C	N	F	R	V	H	T	X	C	E	Y	W	J	U	B	B	V
2	W	H	V	M	B	N	H	O	S	U	R	J	R	Q	S	Z	F	D	I	J	U	D	U	K	Y	H
3	P	V	B	W	X	P	I	Y	O	X	N	Y	B	A	S	O	Z	I	P	W	B	Y	C	Z	I	H
4	W	F	B	I	K	L	C	Z	Q	R	R	F	O	K	A	M	M	E	S	T	J	D	C	J	B	G
5	V	C	E	M	R	N	J	P	O	O	R	F	Q	C	K	S	D	E	M	M	V	L	B	Q	Y	R
6	V	G	U	M	L	B	M	X	A	A	N	N	Y	C	V	T	N	A	F	N	B	L	K	M	M	R
7	P	F	R	M	R	T	L	C	W	K	O	B	C	E	U	S	H	P	E	I	B	X	S	M	G	S
8	D	S	F	W	B	P	S	U	N	P	K	H	M	N	W	P	K	L	N	W	E	N	A	L	I	Q
9	K	G	E	E	A	N	J	Y	K	J	K	A	R	D	S	G	N	S	R	U	W	E	Q	U	H	R
10	P	C	H	E	A	X	L	Z	L	W	V	B	O	C	U	I	S	D	N	Y	C	Z	T	X	G	I
11	J	C	P	H	A	S	P	O	O	A	U	Q	L	G	V	C	I	Q	U	I	C	G	G	J	Y	R
12	Q	F	I	M	L	B	M	B	X	W	E	F	F	G	A	Y	Y	D	K	T	B	Y	T	X	W	B
13	I	G	U	M	Z	N	J	K	E	W	P	H	F	E	V	C	S	Q	G	T	B	X	U	X	M	Q
14	P	G	U	D	I	X	Q	Y	Z	V	L	L	B	C	B	J	X	R	U	U	O	Q	B	Z	F	S
15	D	B	Q	M	I	K	Z	E	Z	O	O	B	B	X	A	A	F	W	N	B	C	Q	I	X	B	V
16	G	F	E	M	A	X	L	U	N	J	P	K	U	U	F	G	D	B	K	U	S	G	Q	L	S	V
17	I	M	T	U	F	N	X	V	L	A	O	R	L	X	C	I	X	O	W	T	B	E	B	X	O	D
18	G	G	F	E	R	P	R	P	L	J	U	A	R	W	K	I	J	A	E	I	B	Y	S	K	J	J
19	P	Y	F	H	F	D	S	O	N	M	T	B	O	G	H	S	V	T	M	E	C	Y	W	W	H	R
20	C	C	U	B	I	F	D	R	P	W	A	G	H	J	S	N	W	W	N	T	C	R	I	M	H	U
21	L	H	T	N	B	B	W	E	O	L	K	L	I	W	A	C	I	Q	J	I	W	J	L	Y	A	D
22	G	C	E	N	I	M	V	V	A	I	U	A	H	T	A	J	V	J	H	J	P	C	Q	L	H	R
23	Q	B	V	R	A	S	P	M	S	C	K	F	M	O	A	Z	D	G	V	U	I	R	T	X	M	I
24	S	H	T	S	M	L	Z	F	Q	W	N	W	Q	W	C	E	V	G	E	C	B	Y	T	O	B	Q
25	L	D	F	F	R	L	G	O	D	U	Q	N	L	J	S	F	K	W	R	Y	L	D	Z	K	J	S
26	Q	V	L	L	A	P	A	Z	L	K	Q	G	A	I	O	P	U	V	F	W	G	U	G	V	C	K
27	J	O	E	R	U	S	F	Z	G	I	J	L	N	V	M	S	Q	G	P	J	L	O	M	W	O	X
28	N	P	E	H	X	X	B	I	L	C	Q	Q	X	I	S	X	M	V	O	M	Z	H	X	A	C	K
29	J	Y	P	C	D	S	D	K	Q	R	H	A	K	I	K	U	I	H	I	A	F	R	W	T	K	I
30	P	G	G	Y	O	Z	Q	B	O	K	S	W	B	I	C	L	U	J	K	S	O	J	G	O	P	I
31	I	T	Y	D	L	H	S	Q	E	N	C	Z	P	W	S	Q	J	H	X	Y	F	F	W	Y	S	I
32	Q	J	T	V	U	F	W	S	T	P	Y	G	B	Z	M	F	Z	K	X	Y	P	B	X	A	I	P
33	U	D	P	Z	O	F	E	V	N	I	Z	R	M	F	R	F	B	K	K	X	H	K	P	N	U	U
34	T	X	C	R	E	Z	G	I	U	N	Y	C	Q	W	S	C	F	E	Z	S	O	Q	R	U	V	A
35	T	V	U	V	E	N	D	H	T	U	O	V	R	T	S	Q	M	J	K	Z	R	Q	W	L	H	I
36	T	X	C	R	U	B	P	Q	A	U	N	Y	X	Y	I	C	K	O	D	S	O	W	G	A	W	I
37	B	W	R	X	J	G	S	M	L	G	C	L	A	Z	S	U	I	M	P	J	L	U	M	E	D	Y
38	A	G	K	D	A	X	W	C	J	H	H	D	D	V	M	S	U	E	F	M	H	Z	D	Y	V	P
39	L	W	E	U	O	B	Q	M	L	G	C	L	A	Z	S	U	U	G	P	M	S	U	Y	C	O	G
40	E	N	P	C	I	B	G	I	X	B	Q	Q	X	M	K	F	I	M	P	H	P	M	M	D	D	G
41	O	Q	E	K	N	N	G	I	U	E	R	J	B	Q	C	I	S	J	I	X	Z	N	D	L	Y	I
42	W	D	G	X	Z	N	S	B	O	M	Q	U	P	O	R	F	U	V	O	Z	L	U	C	D	O	L
43	W	D	P	C	S	C	S	O	L	G	U	X	A	Z	N	A	J	E	F	S	O	F	W	L	I	K
44	A	T	F	D	L	N	S	J	L	N	Y	G	L	E	C	I	I	P	B	Z	C	Q	P	V	B	X
45	W	G	L	E	W	G	E	K	X	Z	N	Q	R	E	Y	M	Y	J	W	T	D	K	C	A	V	I
46	Q	V	H	U	U	R	H	H	O	M	N	E	H	I	T	C	F	E	B	L	E	R	N	D	Q	X
47	X	J	H	C	R	P	F	K	V	G	N	A	B	A	S	X	P	O	A	Q	M	U	A	F	G	E
48	T	J	T	X	X	K	Z	R	G	U	B	H	B	G	K	Y	B	K	B	M	O	M	W	D	I	I
49	D	T	P	C	P	C	Y	S	T	I	O	Z	P	J	T	Q	X	E	F	U	D	U	Z	W	F	K
50	X	J	H	C	L	X	D	Q	O	J	Q	D	U	W	S	I	S	K	P	S	O	R	R	W	V	I

FIGURE 32.

e. It has been determined by a study of the underlined repetitions that lines 1 to 24 are in one polyalphabetic substitution and lines 26 to 50 are in another polyalphabetic substitution. Moreover, it is known that line 25 belongs in one of these two polyalphabets; it remains only to determine to which polyalphabet line 25 belongs. To do so we observe the number of coincidences between line 25 and each of the lines 1 to 24 and 26 to 50. There thus results:

Number of coincidences between line 25 and

Line No.	Number of coincidences	Line No.	Number of coincidences	Line No.	Number of coincidences
1	2	18	4	35	2
2	4	19	2	36	1
3	1	20	3	37	2
4	2	21	1	38	0
5	0	22	0	39	2
6	1	23	0	40	3
7	2	24	1	41	1
8	2	26	1	42	3
9	2	27	1	43	2
10	1	28	2	44	2
11	2	29	0	45	0
12	0	30	0	46	0
13	0	31	2	47	2
14	0	32	2	48	1
15	1	33	2	49	2
16	0	34	2	50	2
17	1				

FIGURE 33.

Rearranging the data in figure 33 there is obtained—

Line 25 and lines 1–24			Line 25 and lines 26–50		
c_i	w_i	c_iw_i	c_i	w_i	c_iw_i
0	7	0	0	5	0
1	7	7	1	5	5
2	7	14	2	13	26
3	1	3	3	2	6
4	2	8			
				25	37
	24	32			

$\bar{c}=32/24=1.33$ $\bar{c}=37/25=1.48$

FIGURE 34.

The expected number of coincidences for 26 pairs of letters "monoalphabetically" related is $0.066\times26=1.72$ and the expected number of coincidences for 26 pairs of letters "randomly" related is $0.038\times26=0.99$. The evidence here points to the conclusion that line 25 must go with lines 26–50.

f. A problem somewhat similar is involved in the solution of an M–94 type cipher with unknown alphabets.

A possible method of procedure for the solution of such a problem is the following. The unknown text is first arranged into lines of 25 letters each. Then all the lines are studied for repetitions in corresponding positions in order to get together a set of lines all enciphered on the same generatrix. Having this set of lines, additional lines may be added to it by testing each line of text against the set for coincidences.

The following considerations must however be kept in mind in order to avoid any difficulty. Suppose that there are 800 lines of text to be studied and that we have been fortunate enough to get together, on the basis of repetitions, 7 lines of a generatrix that has been used 50 times.[27] For a line from the correct generatrix, the expected number of coincidences with the set is $0.066\times25\times7=11.6$. For a line from some other generatrix the expected number of coincidences with the set is $0.038\times25\times7=6.6$.

The distribution of the number of coincidences of every remaining correct generatrix with the set of seven lines, and every incorrect generatrix with the set of seven lines, is in accordance with the Poisson distribution with means 11.6 and 6.6, respectively. Since, however, there are 43 additional lines from the correct generatrix and 750 lines from the other generatrices the probabilities given in the tables must be multiplied by 43 and 750, respectively, to get the absolute frequencies of the distributions. The results are given in figure 35.

[27] These values correspond with the observations made during the solution of such a problem. Theoretically, each generatrix should occur $800/25=32$ times.

Mean 11.6			Mean 6.6		
x_i	f_i	$43f_i$	x_i	f_i'	$750f_i'$
0	0.000009	0.000387	0	0.001360	1.02000
1	.000106	.004558	1	.008978	6.73350
2	.000617	.026531	2	.029629	22.22175
3	.002385	.102555	3	.065183	48.88725
4	.006915	.297345	4	.107553	80.66475
5	.016043	.689849	5	.141969	106.47675
6	.031017	1.333731	6	.156166	117.12450
7	.051400	2.210200	7	.147243	110.43225
8	.074529	3.204747	8	.121475	91.10625
9	.096060	4.130500	9	.089082	66.81150
10	.111430	4.791490	10	.058794	44.09550
11	.117508	5.052844	11	.035276	26.45700
12	.113591	4.884413	12	.019402	14.55150
13	.101358	4.358394	13	.009850	7.37850
14	.083982	3.611226	14	.004644	3.48300
15	.064946	2.792678	15	.002043	1.53225
16	.047086	2.024698	16	.000843	.63225
17	.032129	1.381547	17	.000327	.24525
18	.020706	.890358	18	.000120	.09000
19	.012641	.543563	19	.000042	.03150
20	.007332	.315276	20	.000014	.01050
21	.004050	.174150	21	.000004	.00300
22	.002136	.091848	22	.000001	.00075
23	.001077	.046311			
24	.000521	.022403			
25	.000242	.010406			
26	.000108	.004644			
27	.000046	.001978			
28	.000019	.000817			
29	.000008	.000344			
30	.000003	.000129			
31	.000001	.000043			

FIGURE 35.

In other words we might expect the following:

Number of coincidences	Number of occurrences		Number of coincidences	Number of occurrences	
	Correct generatrix	Other generatrices		Correct generatrix	Other generatrices
0	0	1	10	5	44
1	0	7	11	5	26
2	0	22	12	5	15
3	0	49	13	4	7
4	0	81	14	4	3
5	1	106	15	3	2
6	1	117	16	2	1
7	2	110	17	1	0
8	3	91	18	1	0
9	4	67	19	1	0

FIGURE 36.

It is thus seen that in order to avoid including an incorrect line it is necessary to take only those lines which yield 17 or more coincidences.

The values in figure 35 also enable us to answer the question, "What are the probabilities that a line which gives x coincidences with the set of 7 is a correct generatrix? an incorrect generatrix?" The probability that a line is from the correct generatrix is $43f_i/(43f_i+750f_i')$; the probability that a line is from an incorrect generatrix is $750f_i'/(43f_i+750f_i')$ where $43f_i$ and $750f_i'$ are taken from the row corresponding to the observed number of coincidences.

Thus, the probability that a line which has 14 coincidences with the set is from the correct generatrix is $3.611/(3.611+3.483)=0.51$ and the probability that the line is from an incorrect generatrix is $3.483/(3.611+3.483)=0.49$.

The ratio of the probability that a given line is from the correct generatrix to the probability that the given line is from an incorrect generatrix is $43f_i/750f_i'$.

g. In the preceding subparagraph, when considering the number of coincidences between pairs of lines no distinction was made as to whether the coincident letters were consecutive or separated. It may be of some interest to break down the number of coincidences into the various possible cases.

If lines of 25 letters each are considered, it may be shown that for all pairs of lines having two coincidences 8 percent will be consecutive or digraphs and 92 percent will be separated letters; for three coincidences 1 percent will be consecutive or trigraphs, 22 percent will consist of a digraph plus a single letter and 77 percent will be separated letters; for four coincidences, 0.18 percent will be consecutive or tetragraphs, 3.7 percent will consist of a trigraph plus a single letter, 3.7 percent will consist of two digraphs, 36.5 percent will consist of a digraph plus two separate letters, and 55.92 percent will consist of separated letters.

Now using the Poisson distribution with means $m=0.066\times25=1.65$ and $m=0.038\times25=0.95$ respectively, we have as the probability for 0, 1, 2, 3, and 4 coincidences between a pair of lines each of 25 letters for correct and incorrect matching respectively, the following:

Number of coincidences	Correct ($m=1.65$)	Incorrect ($m=0.95$)
0	0. 190290	0. 387224
1	. 316798	. 366896
2	. 261203	. 174300
3	. 143707	. 055355
4	. 059353	. 013221

If now, the number of coincidences is broken down into its various possibilities and the proper percentage of the probability taken, there results the following:

Coincidences	Correct ($m=1.65$)	Incorrect ($m=0.95$)
None	0. 190290	0. 387224
One	. 316798	. 366896
Digraph	. 020896	. 013944
2 separated	. 240307	. 160356
Trigraph	. 014371	. 005536
Digraph and single letter	. 031616	. 012178
3 separated	. 110654	. 042623
Tetragraph	. 000107	. 000024
Trigraph and single letter	. 002196	. 000489
Two digraphs	. 002196	. 000489
Digraph and 2 separated	. 021664	. 004826
4 separated	. 033190	. 007393

If the values above are rearranged in order with respect to the magnitude of the corresponding probability there results the following:

Coincidences	Correct ($m=1.65$)	Coincidences	Incorrect ($m=0.95$)
One	0. 316798	None	0. 387224
2 separated	. 240307	One	. 366896
None	. 190290	2 separated	. 160356
3 separated	. 110654	3 separated	. 042623
4 separated	. 033190	Digraph	. 013944
Digraph and single letter	. 031616	Digraph and single letter	. 012178
Digraph and 2 separated	. 021664	4 separated	. 007393
Digraph	. 020896	Trigraph	. 005536
Trigraph	. 014371	Digraph and 2 separated	. 004826
Trigraph and single letter	. 002196	Trigraph and single letter	. 000489
Two digraphs	. 002196	Two digraphs	. 000489
Tetragraph	. 000107	Tetragraph	. 000024

63301—38——7

In general, if lines of n letters each are matched the number of coincidences is distributed as follows:

Coincidences		Fraction
2	Digraph	$2/n$
	2 separated	$(n-2)/n$
3	Trigraph	$6/n(n-1)$
	Digraph and single letter	$6(n-3)/n(n-1)$
	3 separated	$(n-3)(n-4)/n(n-1)$
4	Tetragraph	$24/n(n-1)(n-2)$
	Trigraph and single letter	$24(n-4)/n(n-1)(n-2)$
	Digraph and digraph	$24(n-4)/n(n-1)(n-2)$
	Digraph and 2 separated	$12(n-4)(n-5)/n(n-1)(n-2)$
	4 separated	$(n-4)(n-9)/n(n-1)$

27. **Summary.**—It is thus seen that the concept of "coincidences" is of far reaching importance in cryptanalysis. We will not however include further illustrations of its use here, as it is felt that such further illustrations are more suitable for cryptanalytic discussions. We trust that the reader will be able to avail himself of the theories and procedures herein discussed in any further applications of the concept he may encounter.

PART 2

SECTION VIII

FREQUENCY DATA

TABLE 1-A.—*Absolute frequencies of letters appearing in five sets of Governmental plain-text telegrams, each set containing 10,000 letters, arranged alphabetically*

Set No. 1		Set No. 2		Set No. 3		Set No. 4		Set No. 5	
Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency
A	738	A	783	A	681	A	740	A	741
B	104	B	103	B	98	B	83	B	99
C	319	C	300	C	288	C	326	C	301
D	387	D	413	D	423	D	451	D	448
E	1, 367	E	1, 294	E	1, 292	E	1, 270	E	1, 275
F	253	F	287	F	308	F	287	F	281
G	166	G	175	G	161	G	167	G	150
H	310	H	351	H	335	H	349	H	349
I	742	I	750	I	787	I	700	I	697
J	18	J	17	J	10	J	21	J	16
K	36	K	38	K	22	K	21	K	31
L	365	L	393	L	333	L	386	L	344
M	242	M	240	M	238	M	249	M	268
N	786	N	794	N	815	N	800	N	780
O	685	O	770	O	791	O	756	O	762
P	241	P	272	P	317	P	245	P	260
Q	40	Q	22	Q	45	Q	38	Q	30
R	760	R	745	R	762	R	735	R	786
S	658	S	583	S	585	S	628	S	604
T	936	T	879	T	894	T	958	T	928
U	270	U	233	U	312	U	247	U	238
V	163	V	173	V	142	V	133	V	155
W	166	W	163	W	136	W	133	W	182
X	43	X	50	X	44	X	53	X	41
Y	191	Y	155	Y	179	Y	213	Y	229
Z	14	Z	17	Z	2	Z	11	Z	5
Total	10, 000		10, 000		10, 000		10, 000		10, 000

TABLE 2-A.—*Absolute frequencies of letters appearing in the combined five sets of messages totalling 50,000 letters, arranged alphabetically*

A	3, 683	G	819	L	1, 821	Q	175	V	766
B	487	H	1, 694	M	1, 237	R	3, 788	W	780
C	1, 534	I	3, 676	N	3, 975	S	3, 058	X	231
D	2, 122	J	82	O	3, 764	T	4, 595	Y	967
E	6, 498	K	148	P	1, 335	U	1, 300	Z	49
F	1, 416								

TABLE 1–B.—*Absolute frequencies of letters appearing in five sets of Government plain-text telegrams, each set containing 10,000 letters, arranged according to frequency*

Set No. 1		Set No. 2		Set No. 3		Set No. 4		Set No. 5	
Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency
E	1, 367	E	1, 294	E	1, 292	E	1, 270	E	1, 275
T	936	T	879	T	894	T	958	T	928
N	786	N	794	N	815	N	800	R	786
R	760	A	783	O	791	O	756	N	780
I	742	O	770	I	787	A	740	O	762
A	738	I	750	R	762	R	735	A	741
O	685	R	745	A	681	I	700	I	697
S	658	S	583	S	585	S	628	S	604
D	387	D	413	D	423	D	451	D	448
L	365	L	393	H	335	L	386	H	349
C	319	H	351	L	333	H	349	L	344
H	310	C	300	P	317	C	326	C	301
U	270	F	287	U	312	F	287	F	281
F	253	P	272	F	308	M	249	M	268
M	242	M	240	C	288	U	247	P	260
P	241	U	233	M	238	P	245	U	238
Y	191	G	175	Y	179	Y	213	Y	229
G	166	V	173	G	161	G	167	W	182
W	166	W	163	V	142	V	133	V	155
V	163	Y	155	W	136	W	133	G	150
B	104	B	103	B	98	B	83	B	99
X	43	X	50	Q	45	X	53	X	41
Q	40	K	38	X	44	Q	38	K	31
K	36	Q	22	K	22	K	21	Q	30
J	18	J	17	J	10	J	21	J	16
Z	14	Z	17	Z	2	Z	11	Z	5
Total	10, 000		10, 000		10, 000		10, 000		10, 000

TABLE 1–C.—*Absolute frequencies of vowels, high frequency consonants, medium frequency consonants, and low frequency consonants appearing in five sets of Government plain-text telegrams, each set containing 10,000 letters*

Set No.	Vowels	High Frequency Consonants	Medium Frequency Consonants	Low Frequency Consonants
1	3, 993	3, 527	2, 329	151
2	3, 985	3, 414	2, 457	144
3	4, 042	3, 479	2, 356	123
4	3, 926	3, 572	2, 358	144
5	3, 942	3, 546	2, 389	123
Total [1]	19, 888	17, 538	11, 889	685

[1] Grand total, 50,000.

TABLE 2-B.—*Absolute frequencies of letters appearing in the combined five sets of messages totalling 50,000 letters arranged according to frequencies*

E	6,498	I	3,676	C	1,534	Y	967	X	231
T	4,595	S	3,058	F	1,416	G	819	Q	175
N	3,975	D	2,122	P	1,335	W	780	K	148
R	3,788	L	1,821	U	1,300	V	766	J	82
O	3,764	H	1,694	M	1,237	B	487	Z	49
A	3,683								

TABLE 2-C.—*Absolute frequencies of vowels, high frequency consonants, medium frequency consonants, and low frequency consonants appearing in the combined five sets of messages totalling 50,000 letters*

Vowels	19,888
High Frequency Consonants (D, N, R, S, and T)	17,538
Medium Frequency Consonants (B, C, F, G, H, L, M, P, V, and W)	11,889
Low Frequency Consonants (J, K, Q, X, and Z)	685
Total	50,000

TABLE 2-D.—*Absolute frequencies of letters as initial letters of 10,000 words found in Government plain-text telegrams*

(1) ARRANGED ALPHABETICALLY

A	905	G	109	L	196	Q	30	V	77
B	287	H	272	M	384	R	611	W	320
C	664	I	344	N	441	S	965	X	4
D	525	J	44	O	646	T	1,253	Y	88
E	390	K	23	P	433	U	122	Z	12
F	855								
								Total	10,000

(2) ARRANGED ACCORDING TO ABSOLUTE FREQUENCIES

T	1,253	R	611	M	384	L	196	J	44
S	965	D	525	I	344	U	122	Q	30
A	905	N	441	W	320	G	109	K	23
F	855	P	433	B	287	Y	88	Z	12
C	664	E	390	H	272	V	77	X	4
O	646								
								Total	10,000

TABLE 2-E.—*Absolute frequencies of letters as final letters of 10,000 words found in Government plain-text telegrams*

(1) ARRANGED ALPHABETICALLY

A	269	G	225	L	354	Q	8	V	4
B	22	H	450	M	154	R	769	W	45
C	86	I	22	N	872	S	962	X	116
D	1,002	J	6	O	575	T	1,007	Y	866
E	1,628	K	53	P	213	U	31	Z	9
F	252							Total	10,000

(2) ARRANGED ACCORDING TO ABSOLUTE FREQUENCIES

E	1,628	R	769	F	252	C	86	I	22
T	1,007	O	575	G	225	K	53	Z	9
D	1,002	H	450	P	213	W	45	Q	8
S	962	L	354	M	154	U	31	J	6
N	872	A	269	X	116	B	22	V	4
Y	866								
								Total	10,000

TABLE 3.—*Relative frequencies of letters appearing in 1,000 letters based upon Table 2-B*

(1) ARRANGED ALPHABETICALLY

A	73.66	G	16.38	L	36.42	Q	3.50	V	15.32
B	9.74	H	33.88	M	24.74	R	75.76	W	15.60
C	30.68	I	73.52	N	79.50	S	61.16	X	4.62
D	42.44	J	1.64	O	75.28	T	91.90	Y	19.34
E	129.96	K	2.96	P	26.70	U	26.00	Z	.98
F	28.32								
								Total	1,000.00

(2) ARRANGED ACCORDING TO FREQUENCY

E	129.96	I	73.52	C	30.68	Y	19.34	X	4.62
T	91.90	S	61.16	F	28.32	G	16.38	Q	3.50
N	79.50	D	42.44	P	26.70	W	15.60	K	2.96
R	75.76	L	36.42	U	26.00	V	15.32	J	1.64
O	75.28	H	33.88	M	24.74	B	9.74	Z	.98
A	73.66								
								Total	1,000.00

(3) VOWELS

A	73.66
E	129.96
I	73.52
O	75.28
U	26.00
Y	19.34
Total	397.76

(4) HIGH-FREQUENCY CONSONANTS

D	42.44
N	79.50
R	75.76
S	61.16
T	91.90
Total	350.76

(5) MEDIUM-FREQUENCY CONSONANTS

B	9.74
C	30.68
F	28.32
G	16.38
H	33.88
L	36.42
M	24.74
P	26.70
V	15.32
W	15.60
Total	237.78

(6) LOW-FREQUENCY CONSONANTS

X	4.62
Q	3.50
K	2.96
J	1.64
Z	.98
Total	13.70

Total (3), (4), (5), (6) 1,000.00

TABLE 4.—*Frequency distribution for 10,000 letters of literary English, as compiled by Hitt*[1]

(1) ALPHABETICALLY ARRANGED

A	778	G	174	L	372	Q	8	V	112
B	141	H	595	M	288	R	651	W	176
C	296	I	667	N	686	S	622	X	27
D	402	J	51	O	807	T	855	Y	196
E	1,277	K	74	P	223	U	308	Z	17
F	197								

(2) ARRANGED ACCORDING TO FREQUENCY

E	1,277	R	651	U	308	Y	196	K	74
T	855	S	622	C	296	W	176	J	51
O	807	H	595	M	288	G	174	X	27
A	778	D	402	P	223	B	141	Z	17
N	686	L	372	F	197	V	112	Q	8
I	667								

TABLE 5.—*Frequency distribution for 10,000 letters of telegraphic English as compiled by Hitt*

(1) ALPHABETICALLY ARRANGED

A	813	G	201	L	392	Q	38	V	136
B	149	H	386	M	273	R	677	W	166
C	306	I	711	N	718	S	656	X	51
D	417	J	42	O	844	T	634	Y	208
E	1,319	K	88	P	243	U	321	Z	6
F	205								

(2) ARRANGED ACCORDING TO FREQUENCY

E	1,319	S	656	U	321	F	205	K	88
O	844	T	634	C	306	G	201	X	51
A	813	D	417	M	273	W	166	J	42
N	718	L	392	P	243	B	149	Q	38
I	711	H	386	Y	208	V	136	Z	6
R	677								

[1] Hitt, Capt. Parker. *Manual for the Solution of Military Ciphers.* Army Service Schools Press, Fort Leavenworth, Kansas, 1916.

Table 6.—*Frequency distribution of digraphs—Based on 50,000 letters of Government plain-text telegrams; reduced to 5,000 digraphs*

Second Letter

First Letter	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	3	6	14	27	1	4	6	2	17	1	2	32	14	64	2	12		44	41	47	13	7	3		12	
B	4				18				2	1		6	1		4			2	1	1	2				7	
C	20		3	1	32	1		14	7		4	5	1	1	41			4	1	14	4		1		1	
D	32	4	4	8	33	8	2	2	27	1		3	5	4	16	5	2	12	13	15	5	3	4		1	
E	35	4	32	60	42	18	4	7	27	1		29	14	111	12	20	12	87	54	37	3	20	7	7	4	1
F	5		2	1	10	11	1		39			2	1		40	1		9	3	11	3		1		1	
G	7		2	1	14	2	1	20	5	1		2	1	3	6	2		5	3	4	2		1			
H	20	1	3	2	20	5			33			1	2	3	20	1	1	17	4	28	8		1		1	
I	8	2	22	6	13	10	19				2	23	9	75	41	7		27	35	27		25		15		2
J	1				2										2						2					
K	1		1		6				2			1		1					1							
L	28	3	3	9	37	3	1	1	20			27	2	1	13	3		2	6	8	2	2	2		10	
M	36	6	3	1	26	1		1	9				13		10	8		2	4	2	2				2	
N	26	2	19	52	57	9	27	4	30	1	2	5	5	8	18	3	1	4	24	82	7	3	3		5	
O	7	4	8	12	3	25	2	3	5	1	2	19	25	77	6	25		64	14	19	37	7	8	1	2	
P	14	1	1	1	23	2		3	6			13	4	1	17	11		18	6	8	3	1	1		1	
Q													1					1			15					
R	39	2	9	17	98	6	7	3	30	1	1	5	9	7	28	13		11	31	42	5	5	4		9	
S	24	3	13	5	49	12	2	26	34		1	2	3	4	15	10		5	19	63	11	1	4		1	
T	28	3	6	6	71	7	1	78	45			5	6	7	50	2	1	17	19	19	5		36		41	1
U	5	3	3	3	11	1	8		5			6	5	21	1	2		31	12	12		1				
V	6				57				12						1					1						
W	12				22			4	13			1		2	19			1	1						1	
X	2		2	1	1	1		1	2					1	1	2		1	1	7						
Y	6	2	4	4	9	11	1	1	3			2	2	6	10	3		4	11	15	1		1			
Z	1				2				1																	

TABLE 7-A.—*The 438 different digraphs of table 6 arranged according to their absolute frequencies*

EN	111
RE	98
ER	87
NT	82
TH	78
ON	77
IN	75
TE	71
AN	64
OR	64
ST	63
ED	60
NE	57
VE	57
ES	54
ND	52
TO	50
SE	49
	[1] 1,249
AT	47
TI	45
AR	44
EE	42
RT	42
AS	41
CO	41
IO	41
TY	41
FO	40
FI	39
RA	39
ET	37
OU	37
LE	37
MA	36
TW	36
EA	35
IS	35
SI	34
DE	33
HI	33
AL	32
CE	32
DA	32
EC	32
RS	31
UR	31
NI	30
RI	30
EL	29
HT	28
LA	28
RO	28
TA	28
	[2] 2,495
LL	27
AD	27
DI	27
EI	27
IR	27
IT	27
NG	27
ME	26
NA	26
SH	26
IV	25
OF	25
OM	25
OP	25
NS	24
SA	24
IL	23
PE	23
IC	22
WE	22
UN	21
CA	20
EP	20
EV	20
GH	20
HA	20
HE	20
HO	20
LI	20
SS	19
TT	19
IG	19
NC	19
OL	19
OT	19
TS	19
WO	19
BE	18
EF	18
NO	18
PR	18
AI	17
HR	17
PO	17
RD	17
TR	17
DO	16
DT	15
IX	15
QU	15
SO	15
YT	15
AC	14
AM	14
CH	14
CT	14
EM	14
GE	14
OS	14
PA	14
PL	13
RP	13
SC	13
WI	13
MM	13
DS	13
AU	13
IE	13
LO	13
	[3] 3,745
AP	12
DR	12
EQ	12
AY	12
EO	12
OD	12
SF	12
US	12
UT	12
VI	12
WA	12
FF	11
PP	11
RR	11
UE	11
FT	11
SU	11
YF	11
YS	11
YO	10
FE	10
IF	10
LY	10
MO	10
SP	10
YE	9
FR	9
IM	9
LD	9
MI	9
NF	9
RC	9
RM	9
RY	9
DD	8
NN	8
DF	8
IA	8
HU	8
LT	8
MP	8
OC	8
OW	8
PT	8
UG	8
AV	7
BY	7
CI	7
EH	7
OA	7
EW	7
EX	7

[1] The 18 digraphs above this line compose 25% of the total.
[2] The 53 digraphs above this line compose 50% of the total.
[3] The 117 digraphs above this line compose 75% of the total.

TABLE 7–A.—*The 438 different digraphs of table 6 arranged according to their absolute frequencies*—Continued

GA	7	SD	5	DV	3	KI	2
IP	7	SR	5	AA	3	LM	2
NU	7	TL	5	EU	3	LR	2
OV	7	TU	5	OE	3	LU	2
RG	7	UM	5	YI	3	LV	2
RN	7	AF	4	FS	3	LW	2
TE	7	BA	4	FU	3	MR	2
TN	7	BO	4	GN	3	MT	2
XT	7	CK	4	GS	3	MU	2
AB	6	CR	4	HC	3	MY	2
AG	6	CU	4	HN	3	NB	2
BL	6	DB	4	LB	3	NK	2
OO	6	DC	4	LC	3	OG	2
YA	6	DN	4	LF	3	OK	2
GO	6	DW	4	LP	3	PF	2
ID	6	EB	4	MC	3	RB	2
KE	6	EG	4	NP	3	SG	2
LS	6	EY	4	NV	3	SL	2
MB	6	GT	4	NW	3	TP	2
PI	6	HS	4	OH	3	UP	2
PS	6	MS	4	AH	2	WN	2
RF	6	NH	4	AK	2	XA	2
TC	6	NR	4	BI	2	XC	2
TD	6	OB	4	BR	2	XI	2
TM	6	PM	4	BU	2	XP	2
UL	6	RW	4	DG	2	YB	2
VA	6	SN	4	DH	2	YL	2
YN	6	SW	4	DO	2	YM	2
CL	5	WH	4	AO	2	ZE	2
DM	5	YC	4	OY	2	GG	1
DP	5	YD	4	FC	2	AJ	1
DU	5	YR	4	FL	2	BJ	1
OI	5	PH	3	GC	2	BM	1
UA	5	PU	3	GF	2	BS	1
UI	5	RH	3	GL	2	BT	1
FA	5	SB	3	GP	2	CD	1
GI	5	SM	3	GU	2	CF	1
GR	5	TB	3	HD	2	CM	1
HF	5	UB	3	HM	2	CN	1
NL	5	UC	3	IB	2	CS	1
NM	5	UD	3	IK	2	CW	1
NY	5	YP	3	IZ	2	CY	1
RL	5	CC	3	JE	2	DJ	1
RU	5	AW	3	JO	2	DY	1
RV	5	DL	3	JU	2	EJ	1

TABLE 7-A.—*The 138 different digraphs of table 6 arranged according to their absolute frequencies*—Continued

AE	1	HY	1	PD	1	WL	1
UO	1	JA	1	PN	1	WR	1
YU	1	KA	1	PV	1	WS	1
EZ	1	KC	1	PW	1	WY	1
FD	1	KL	1	PY	1	XD	1
FG	1	KN	1	QM	1	XE	1
FM	1	KS	1	QR	1	XF	1
FP	1	LG	1	RJ	1	XH	1
FW	1	LH	1	RK	1	XN	1
FY	1	LN	1	SK	1	XO	1
GD	1	MD	1	SV	1	XR	1
GJ	1	MF	1	SY	1	XS	1
GM	1	MH	1	TG	1	YG	1
GW	1	NJ	1	TQ	1	YH	1
HB	1	NQ	1	TZ	1	YW	1
HL	1	OJ	1	UF	1	ZA	1
HP	1	OX	1	UV	1	ZI	1
HQ	1	PB	1	VO	1		
HW	1	PC	1	VT	1	Total	5,000

TABLE 7-B.—*The 18 digraphs composing 25% of the digraphs in Table 6 arranged alphabetically according to their initial letters*

(1) AND ACCORDING TO THEIR FINAL LETTERS

AN	64	ON	77
		OR	64
ED	60	RE	98
EN	111		
ER	87	SE	49
ES	54	ST	63
		TE	71
IN	75	TH	78
		TO	50
ND	52	VE	57
NE	57		
NT	82	Total	1,249

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

AN	64	ON	77
		OR	64
EN	111	RE	98
ER	87		
ED	60	SE	49
ES	54	ST	63
		TH	78
IN	75	TE	71
		TO	50
NT	82	VE	57
NE	57		
ND	52	Total	1,249

TABLE 7–C.—*The 53 digraphs composing 50% of the 5,000 digraphs of Table 6, arranged alphabetically according to their initial letters*

(1) AND ACCORDING TO THEIR FINAL LETTERS

AL	32	MA	36
AN	64		
AR	44	ND	52
AS	41	NE	57
AT	47	NI	30
		NT	82
CE	32		
CO	41	ON	77
		OR	64
DA	32	OU	37
DE	33		
		RA	39
EA	35	RE	98
EC	32	RI	30
ED	60	RO	28
EE	42	RS	31
EL	29	RT	42
EN	111		
ER	87	SE	49
ES	54	SI	34
ET	37	ST	63
		TA	28
FI	39	TE	71
FO	40	TH	78
		TI	45
HI	33	TO	50
HT	28	TW	36
		TY	41
IN	75		
IO	41	UR	31
IS	35		
		VE	57
LA	28		
LE	37	Total	2, 495

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

AN	64	MA	36
AT	47		
AR	44	NT	82
AS	41	NE	57
AL	32	ND	52
		NI	30
CO	41		
CE	32	ON	77
		OR	64
DE	33	OU	37
DA	32		
		RE	98
EN	111	RT	42
ER	87	RA	39
ED	60	RS	31
ES	54	RI	30
EE	42	RO	28
ET	37		
EA	35	ST	63
EC	32	SE	49
EL	29	SI	34
		TH	78
FO	40	TE	71
FI	39	TO	50
		TI	45
HI	33	TY	41
HT	28	TW	36
		TA	28
IN	75		
IO	41	UR	31
IS	35		
		VE	57
LE	37		
LA	28	Total	2, 495

TABLE 7-D.—*The 117 digraphs composing 75% of the 5,000 digraphs of Table 6, arranged alphabetically according to their initial letters—*

(1) AND ACCORDING TO THEIR FINAL LETTERS

AC	14	EP	20	LO	13	RI	30
AD	27	ER	87			RO	28
AI	17	ES	54	MA	36	RS	31
AL	32	ET	37	ME	26	RT	42
AM	14	EV	20				
AN	64			NA	26	SA	24
AR	44	FI	39	NC	19	SE	49
AS	41	FO	40	ND	52	SH	26
AT	47			NE	57	SI	34
AU	13	GE	14	NG	27	SO	15
		GH	20	NI	30	SS	19
BE	18			NO	18	ST	63
		HA	20	NS	24		
CA	20	HE	20	NT	82	TA	28
CE	32	HI	33			TE	71
CH	14	HO	20	OF	25	TH	78
CO	41	HR	17	OL	19	TI	45
CT	14	HT	28	OM	25	TO	50
				ON	77	TR	17
DA	32	IC	22	OP	25	TS	19
DE	33	IE	13	OR	64	TT	19
DI	27	IG	19	OS	14	TW	36
DO	16	IL	23	OT	19	TY	41
DS	13	IN	75	OU	37		
DT	15	IO	41			UN	21
		IR	27	PA	14	UR	31
EA	35	IS	35	PE	23		
EC	32	IT	27	PO	17	VE	57
ED	60	IV	25	PR	18		
EE	42	IX	15			WE	22
EF	18			QU	15	WO	19
EI	27	LA	28				
EL	29	LE	37	RA	39	YT	15
EM	14	LI	20	RD	17		
EN	111	LL	27	RE	98	Total	3,745

Table 7–D, Concluded.—*The 117 digraphs comprising 75% of the 5,000 digraphs of Table 6, arranged alphabetically according to their initial letters—*

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

AN	64	EI	27	MA	36	RI	30
AT	47	EP	20	ME	26	RO	28
AR	44	EV	20			RD	17
AS	41	EF	18	NT	82		
AL	32	EM	14	NE	57	ST	63
AD	27			ND	52	SE	49
AI	17	FO	40	NI	30	SI	34
AC	14	FI	39	NG	27	SH	26
AM	14			NA	26	SA	24
AU	13	GH	20	NS	24	SS	19
		GE	14	NC	19	SO	15
BE	18			NO	18		
		HI	33			TH	78
CO	41	HT	28	ON	77	TE	71
CE	32	HA	20	OR	64	TO	50
CA	20	HE	20	OU	37	TI	45
CH	14	HO	20	OF	25	TY	41
CT	14	HR	17	OM	25	TW	36
				OP	25	TA	28
DE	33	IN	75	OL	19	TS	19
DA	32	IO	41	OT	19	TT	19
DI	27	IS	35	OS	14	TR	17
DO	16	IR	27				
DT	15	IT	27	PE	23	UR	31
DS	13	IV	25	PR	18	UN	21
		IL	23	PO	17		
EN	111	IC	22	PA	14	VE	57
ER	87	IG	19				
ED	60	IX	15	QU	15	WE	22
ES	54	IE	13			WO	19
EE	42			RE	98		
ET	37	LE	37	RT	42	YT	15
EA	35	LA	28	RA	39		
EC	32	LL	27	RS	31	Total	3,745
EL	29	LI	20				
		LO	13				

Table 7–E.—*All the 438 digraphs of Table 6, arranged first alphabetically according to their initial letters and then alphabetically according to their final letters.*

(SEE TABLE 6.—READ ACROSS THE ROWS)

TABLE 8.—*The 438 different digraphs of Table 6, arranged first alphabetically according to their initial letters, and then according to their absolute frequencies under each initial letter* [1]

AN	64	CT	14	ED	60	GH	20
AT	47	CI	7	ES	54	GE	14
AR	44	CL	5	EE	42	GA	7
AS	41	CK	4	ET	37	GO	6
AL	32	CR	4	EA	35	GI	5
AD	27	CU	4	EC	32	GR	5
AI	17	CC	3	EL	29	GT	4
AC	14	CD	1	EI	27	GN	3
AM	14	CF	1	EP	20	GS	3
AU	13	CM	1	EV	20	GC	2
AP	12	CN	1	EF	18	GF	2
AY	12	CS	1	EM	14	GL	2
AV	7	CW	1	EO	12	GP	2
AB	6	CY	1	EQ	12	GU	2
AG	6			EH	7	GD	1
AF	4	DE	33	EW	7	GG	1
AA	3	DA	32	EX	7	GJ	1
AW	3	DI	27	EB	4	GM	1
AH	2	DO	16	EG	4	GW	1
AK	2	DT	15	EY	4		
AO	2	DS	13	EU	3		
AE	1	DR	12	EJ	1		
AJ	1	DD	8	EZ	1	HI	33
		DF	8			HT	28
BE	18	DM	5	FO	40	HA	20
BY	7	DP	5	FI	39	HE	20
BL	6	DU	5	FF	11	HO	20
BA	4	DB	4	FT	11	HR	17
BO	4	DC	4	FE	10	HU	8
BI	2	DN	4	FR	9	HF	5
BR	2	DW	4	FA	5	HS	4
BU	2	DL	3	FS	3	HC	3
BJ	1	DV	3	FU	3	HN	3
BM	1	DG	2	FC	2	HD	2
BS	1	DH	2	FL	2	HM	2
BT	1	DQ	2	FD	1	HB	1
		DJ	1	FG	1	HL	1
CO	41	DY	1	FM	1	HP	1
CE	32			FP	1	HQ	1
CA	20	EN	111	FW	1	HW	1
CH	14	ER	87	FY	1	HY	1

[1] For arrangement alphabetically first under intial letters and then under final letters, see Table 6.

TABLE 8, Contd.—*The 438 different digraphs of Table 6, arranged first alphabetically according to their initial letters, and then according to their absolute frequencies under each initial letter* [1]

IN	75	LI	20	NE	57	OA	7
IO	41	LO	13	ND	52	OV	7
IS	35	LY	10	NI	30	OO	6
IR	27	LD	9	NG	27	OI	5
IT	27	LT	8	NA	26	OB	4
IV	25	LS	6	NS	24	OE	3
IL	23	LB	3	NC	19	OH	3
IC	22	LC	3	NO	18	OG	2
IG	19	LF	3	NF	9	OK	2
IX	15	LP	3	NN	8	OY	2
IE	13	LM	2	NU	7	OJ	1
IF	10	LR	2	NL	5	OX	1
IM	9	LU	2	NM	5		
IA	8	LV	2	NY	5	PE	23
IP	7	LW	2	NH	4	PR	18
ID	6	LG	1	NR	4	PO	17
		LH	1	NP	3	PA	14
IB	2	LN	1	NV	3	PL	13
IK	2			NW	3	PP	11
IZ	2	MA	36	NB	2	PT	8
		ME	26	NK	2	PI	6
JE	2	MM	13	NJ	1	PS	6
JO	2	MO	10	NQ	1	PM	4
JU	2	MI	9			PH	3
JA	1	MP	8	ON	77	PU	3
		MB	6	OR	64	PF	2
KE	6	MS	4	OU	37	PB	1
KI	2	MC	3	OF	25	PC	1
KA	1	MR	2	OM	25	PD	1
KC	1	MT	2	OP	25	PN	1
KL	1	MU	2	OL	19	PV	1
KN	1	MY	2	OT	19	PW	1
KS	1	MD	1	OS	14	PY	1
		MF	1				
LE	37	MH	1	OD	12	QU	15
LA	28			OC	8	QM	1
LL	27	NT	82	OW	8	QR	1

[1] For arrangement alphabetically first under initial letters and then under final letters, see Table 6.

TABLE 8, Concluded.—*The 438 different digraphs of Table 6, arranged first alphabetically according to their initial letters, and then according to their absolute frequencies under each initial letter* [1]

RE	98	SR	5	US	12	XI	2
RT	42	SN	4	UT	12	XP	2
RA	39	SW	4	UE	11	XD	1
RS	31	SB	3	UG	8	XE	1
RI	30	SM	3	UL	6	XF	1
RO	28	SG	2	UA	5	XH	1
RD	17	SL	2	UI	5	XN	1
RP	13	SK	1	UM	5	XO	1
RR	11	SV	1	UB	3	XR	1
RC	9	SY	1	UC	3	XS	1
RM	9			UD	3		
RY	9	TH	78	UP	2	YT	15
RG	7	TE	71	UF	1	YF	11
RN	7	TO	50	UO	1	YS	11
RF	6	TI	45	UV	1	YO	10
RL	5	TY	41			YE	9
RU	5	TW	36	VE	57	YA	6
RV	5	TA	28	VI	12	YN	6
RW	4	TS	19	VA	6	YC	4
RH	3	TT	19	VO	1	YD	4
RB	2	TR	17	VT	1	YR	4
RJ	1	TF	7			YI	3
RK	1	TN	7	WE	22	YP	3
		TC	6	WO	19	YB	2
ST	63	TD	6	WI	13	YL	2
SE	49	TM	6	WA	12	YM	2
SI	34	TL	5	WH	4	YG	1
SH	26	TU	5	WN	2	YH	1
SA	24	TB	3	WL	1	YU	1
SS	19	TP	2	WR	1	YW	1
SO	15	TG	1	WS	1		
SC	13	TQ	1	WY	1	ZE	2
SF	12	TZ	1			ZA	1
SU	11			XT	7	ZI	1
SP	10	UR	31	XA	2		
SD	5	UN	21	XC	2	Total	5,000

[1] For arrangement alphabetically first under initial letters and then under final letters, see Table 6.

TABLE 9–A.—*The 438 different digraphs of Table 6, arranged first alphabetically according to their final letters, and then according to their absolute frequencies*

RA	39	EC	32	RE	98	GF	2
MA	36	IC	22	TE	71	PF	1
EA	35	NC	19	NE	57	CF	2
DA	32	AC	14	VE	57	MF	1
LA	28	SC	13	SE	49	UF	1
TA	28	RC	9	EE	42	XF	1
NA	26	OC	8	LE	37		
SA	24	TC	6	DE	33		
CA	20	DC	4	CE	32	NG	27
HA	20	YC	4	ME	26	IG	19
PA	14	CC	3	PE	23	UG	8
WA	12	HC	3	WE	22	RG	7
IA	8	LC	3	HE	20	AG	6
GA	7	MC	3	BE	18	EG	4
OA	7	UC	3	GE	14	DG	2
VA	6	FC	2	IE	13	OG	2
YA	6	GC	2	UE	11	SG	2
FA	5	XC	2	FE	10	FG	1
UA	5	KC	1	YE	9	GG	1
BA	4	PC	1	KE	6	LG	1
AA	3			OE	3	TG	1
XA	2			JE	2	YG	1
JA	1	ED	60	ZE	2		
KA	1	ND	52	AE	1		
ZA	1	AD	27	XE	1		
		RD	17			TH	78
AB	6	OD	12			SH	26
MB	6	LD	9			GH	20
DB	4	DD	8	OF	25	CH	14
EB	4	ID	6	EF	18	EH	7
OB	4	TD	6	SF	12	NH	4
LB	3	SD	5	FF	11	WH	4
SB	3	YD	4	YF	11	OH	3
TB	3	UD	3	IF	10	PH	3
UB	3	HD	2	NF	9	RH	3
IB	2	CD	1	DF	8	AH	2
NB	2	FD	1	TF	7	DH	2
RB	2	GD	1	RF	6	LH	1
YB	2	MD	1	HF	5	MH	1
HB	1	PD	1	AF	4	XH	1
PB	1	XD	1	LF	3	YH	1

TABLE 9–A, Contd.—*The 438 different digraphs of Table 6, arranged first alphabetically according to their final letters, and then according to their absolute frequencies*

TI	45
FI	39
SI	34
HI	33
NI	30
RI	30
DI	27
EI	27
LI	20
AI	17
WI	13
VI	12
MI	9
CI	7
PI	6
GI	5
OI	5
UI	5
YI	3
BI	2
KI	2
XI	2
ZI	1
AJ	1
BJ	1
DJ	1
EJ	1
GJ	1
NJ	1
OJ	1
RJ	1
CK	4
AK	2
IK	2
NK	2
OK	2
RK	1
SK	1
AL	32
EL	29
LL	27
IL	23
OL	19
PL	13
BL	6
UL	6
CL	5
NL	5
RL	5
TL	5
DL	3
FL	2
GL	2
SL	2
YL	2
HL	1
KL	1
WL	1
OM	25
AM	14
EM	14
MM	13
IM	9
RM	9
TM	6
DM	5
NM	5
UM	5
PM	4
SM	3
HM	2
LM	2
YM	2
BM	1
CM	1
FM	1
GM	1
QM	1
EN	111
ON	77
IN	75
AN	64
UN	21
NN	8
RN	7
TN	7
YN	6
DN	4
SN	4
GN	3
HN	3
WN	2
CN	1
KN	1
LN	1
PN	1
XN	1
TO	50
CO	41
IO	41
FO	40
RO	28
HO	20
WO	19
NO	18
PO	17
DO	16
SO	15
LO	13
EO	12
MO	10
YO	10
GO	6
OO	6
BO	4
AO	2
JO	2
UO	1
VO	1
XO	1
OP	25
EP	20
RP	13
AP	12
PP	11
SP	10
MP	8
IP	7
DP	5
LP	3
NP	3
YP	3
GP	2
TP	2
UP	2
XP	2
FP	1
HP	1
EQ	12
DQ	2
HQ	1
NQ	1
TQ	1
ER	87
OR	64
AR	44
UR	31
IR	27
PR	18
HR	17
TR	17
DR	12
RR	11
FR	9
GR	5
SR	5
CR	4
NR	4
YR	4
BR	2
LR	2
MR	2
QR	1
WR	1
XR	1

TABLE 9-A, Concluded.—*The 438 different digraphs of Table 6, arranged first alphabetically according to their final letters, and then according to their absolute frequencies*

ES	54	OT	19	JU	2	PW	1
AS	41	TT	19	LU	2	YW	1
IS	35	DT	15	MU	2		
RS	31	YT	15	YU	1	IX	15
NS	24	CT	14			EX	7
SS	19	UT	12	IV	25	OX	1
TS	19	FT	11	EV	20		
OS	14	LT	8	AV	7	TY	41
DS	13	PT	8	OV	7	AY	12
US	12	XT	7	RV	5	LY	10
YS	11	GT	4	DV	3	RY	9
LS	6	MT	2	NV	3	BY	7
PS	6	BT	1	LV	2	NY	5
HS	4	VT	1	PV	1	EY	4
MS	4			SV	1	MY	2
FS	3	OU	37	UV	1	OY	2
GS	3	QU	15			CY	1
BS	1	AU	13	TW	36	DY	1
CS	1	SU	11	OW	8	FY	1
KS	1	HU	8	EW	7	HY	1
WS	1	NU	7	DW	4	PY	1
XS	1	DU	5	RW	4	SY	1
		RU	5	SW	4	WY	1
NT	82	TU	5	AW	3		
ST	63	CU	4	NW	3	IZ	2
AT	47	EU	3	LW	2	EZ	1
RT	42	FU	3	CW	1	TZ	1
ET	37	PU	3	FW	1		
HT	28	BU	2	GW	1	Total	5,000
IT	27	GU	2	HW	1		

TABLE 9-B.—*The 18 digraphs composing 25% of the 5,000 digraphs of Table 6, arranged alphabetically according to their final letters—*

(1) AND ACCORDING TO THEIR INITIAL LETTERS

ED	60	IN	75
ND	52	ON	77
NE	57	TO	50
RE	98		
SE	49	ER	87
TE	71	OR	64
VE	57		
		ES	54
TH	78	NT	82
		ST	63
AN	64		
EN	111	Total	1, 249

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

ED	60	IN	75
ND	52	AN	64
RE	98	TO	50
TE	71		
NE	57	ER	87
VE	57	OR	64
SE	49	ES	54
TH	78	NT	82
		ST	63
EN	111		
ON	77	Total	1, 249

TABLE 9-C.—*The 53 digraphs composing 50% of the 5,000 digraphs of Table 6, arranged alphabetically according to their final letters—*

(1) AND ACCORDING TO THEIR INITIAL LETTERS

DA	32	RE	98	EN	111	IS	35
EA	35	SE	49	IN	75	RS	31
LA	28	TE	71	ON	77		
MA	36	VE	57			AT	47
RA	39			CO	41	ET	37
TA	28	TH	78	FO	40	HT	28
				IO	41	NT	82
EC	32	FI	39	RO	28	RT	42
		HI	33	TO	50	ST	63
		NI	30				
ED	60	RI	30	AR	44	OU	37
ND	52	SI	34	ER	87		
		TI	45	OR	64	TW	36
CE	32			UR	31		
DE	33	AL	32			TY	41
EE	42	EL	29				
LE	37			AS	41		
NE	57	AN	64	ES	54	Total	2, 495

TABLE 9–C, Concluded.—*The 53 digraphs composing 50% of the 5,000 digraphs of Table 6, arranged alphabetically according to their final letters—*

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

RA	39	LE	37	ON	77	IS	35
MA	36	DE	33	IN	75	RS	31
EA	35	CE	32	AN	64		
DA	32					NT	82
LA	28	TH	78	TO	50	ST	63
TA	28			CO	41	AT	47
		TI	45	IO	41	RT	42
EC	32	FI	39	FO	40	ET	37
ED	60	SI	34	RO	28	HT	28
ND	52	HI	33				
		NI	30	ER	87	OU	37
RE	98	RI	30	OR	64		
TE	71			AR	44	TW	36
NE	57	AL	32	UR	31		
VE	57	EL	29			TY	41
SE	49			ES	54		
EE	42	EN	111	AS	41	Total	2, 495

TABLE 9–D.—*The 117 digraphs composing 75% of the 5,000 digraphs of Table 6, arranged alphabetically according to their final letters—*

(1) AND ACCORDING TO THEIR INITIAL LETTERS

CA	20	ND	52	EF	18	SI	34
DA	32	RD	17	OF	25	TI	45
EA	35						
HA	20	BE	18	IG	19	AL	32
LA	28	CE	32	NG	27	EL	29
MA	36	DE	33			IL	23
NA	26	EE	42	CH	14	LL	27
PA	14	GE	14	GH	20	OL	19
RA	39	HE	20	SH	26		
SA	24	IE	13	TH	78	AM	14
TA	28	LE	37			EM	14
		ME	26	AI	17	OM	25
AC	14	NE	57	DI	27		
EC	32	PE	23	EI	27		
IC	22	RE	98	FI	39	AN	64
NC	19	SE	49	HI	33	EN	111
		TE	71	LI	20	IN	75
AD	27	VE	57	NI	30	ON	77
ED	60	WE	22	RI	30	UN	21

TABLE 9–D, Contd.—*The 117 digraphs composing 75% of the 5,000 digraphs of Table 6, arranged alphabetically according to their final letters—*

(1) AND ACCORDING TO THEIR INITIAL LETTERS—Continued

CO	41	AR	44	OS	14	YT	15
DO	16	TR	17	IS	35		
FO	40	UR	31	RS	31	AU	13
HO	20	ER	87			OU	37
IO	41	OR	64	AT	47	QU	15
LO	13	PR	18	CT	14	EV	20
NO	18	HR	17	DT	15	IV	25
PO	17	IR	27	ET	37		
RO	28			HT	28	TW	36
SO	15	AS	41	IT	27		
TO	50	SS	19	NT	82	IX	15
WO	19	TS	19	OT	19		
		DS	13	RT	42	TY	41
EP	20	ES	54	ST	63		
OP	25	NS	24	TT	19	Total	3,745

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

RA	39	TE	71	TH	78	AM	14
MA	36	NE	57	SH	26	EM	14
EA	35	VE	57	GH	20		
DA	32	SE	49	CH	14	EN	111
LA	28	EE	42			ON	77
TA	28	LE	37	TI	45	IN	75
NA	26	DE	33	FI	39	AN	64
SA	24	CE	32	SI	34	UN	21
CA	20	ME	26	HI	33		
HA	20	PE	23	NI	30		
PA	14	WE	22	RI	30		
		HE	20	DI	27	TO	50
EC	32			EI	27	CO	41
IC	22	BE	18	LI	20	IO	41
NC	19	GE	14	AI	17	FO	40
AC	14	IE	13			RO	28
				AL	32	HO	20
ED	60	OF	25	EL	29	WO	19
ND	52	EF	18	LL	27	NO	18
AD	27			IL	23	PO	17
RD	17			OL	19	DO	16
		NG	27			SO	15
RE	98	IG	19	OM	25	LO	13

TABLE 9-D, Concluded.—*The 117 digraphs composing 75% of the 5,000 digraphs of Table 6, arranged alphabetically according to their final letters*

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES—Continued

OP	25	ES	54	AT	47	QU	15
EP	20	AS	41	RT	42	AU	13
		IS	35	ET	37		
		RS	31	HT	28	IV	25
ER	87	NS	24	IT	27	EV	20
OR	64	SS	19	OT	19		
AR	44	TS	19	TT	19	TW	36
UR	31	OS	14	DT	15	IX	15
IR	27	DS	13	YT	15		
PR	18			CT	14	TY	41
HR	17	NT	82				
TR	17	ST	63	OU	37	Total	3,745

TABLE 9-E.—*All the 438 different digraphs of Table 6 arranged alphabetically first according to their final letters and then according to their initial letters*

(SEE TABLE 6.—READ DOWN THE COLUMNS)

TABLE 10-A.—*The 56 trigraphs appearing 100 or more times in the 50,000 letters of Government plain-text telegrams arranged according to their absolute frequencies*

ENT	569	TOP	174	EIG	135
ION	260	NTH	171	FIV	135
AND	228	TWE	170	MEN	131
ING	226	TWO	163	SEV	131
IVE	225	ATI	160	ERS	126
TIO	221	THR	158	UND	125
FOR	218	NTY	157	NET	118
OUR	211	HRE	153	PER	115
THI	211	WEN	153	STA	115
ONE	210	FOU	152	TER	115
NIN	207	ORT	146	EQU	114
STO	202	REE	146	RED	113
EEN	196	SIX	146	TED	112
GHT	196	ASH	143	ERI	109
INE	192	DAS	140	HIR	106
VEN	190	IGH	140	IRT	105
EVE	177	ERE	138	DER	101
EST	176	COM	136	DRE	100
TEE	174	ATE	135		

TABLE 10–B. *The 56 trigraphs appearing 100 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their initial letters and then according to their absolute frequencies*

AND	228
ATI	160
ASH	143
ATE	135
COM	136
DAS	140
DER	101
DRE	100
ENT	569
EEN	196
EVE	177
EST	176
ERE	138
EIG	135
ERS	126
EQU	114
ERI	109
FOR	218
FOU	152
FIV	135
GHT	196
HRE	153
HIR	106
ION	260
ING	226
IVE	225
INE	192
IGH	140
IRT	105
MEN	131
NIN	207
NTH	171
NTY	157
NET	118
OUR	211
ONE	210
ORT	146
PER	115
REE	146
RED	113
STO	202
SIX	146
SEV	131
STA	115
TIO	221
THI	211
TEE	174
TOP	174
TWE	170
TWO	163
THR	158
TER	115
TED	112
UND	125
VEN	190
WEN	153

TABLE 10–C.—*The 56 trigraphs appearing 100 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their central letters and then according to their absolute frequencies*

DAS	140
EEN	196
VEN	190
TEE	174
WEN	153
REE	146
MEN	131
SEV	131
NET	118
PER	115
TER	115
RED	113
TED	112
DER	101
IGH	140
THI	211
GHT	196
THR	158
TIO	221
NIN	207
SIX	146
EIG	135
FIV	135
HIR	106
ENT	569
AND	228
ING	226
ONE	210
INE	192
UND	125
ION	260
FOR	218
TOP	174
FOU	152
COM	136

TABLE 10–C, Concluded.—*The 56 trigraphs appearing 100 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their central letters and then according to their absolute frequencies*

EQU	114
HRE	153
ORT	146
ERE	138
ERS	126
ERI	109
IRT	105

DRE	100
EST	176
ASH	143
STO	202
NTH	171
ATI	160
NTY	157
ATE	135

STA	115
OUR	211
IVE	225
EVE	177
TWE	170
TWO	163

TABLE 10–D.—*The 56 trigraphs appearing 100 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their final letters and then according to their absolute frequencies*

STA	115
AND	228
UND	125
RED	113
TED	112
IVE	225
ONE	210
INE	192
EVE	177
TEE	174
TWE	170
HRE	153
REE	146
ERE	138
ATE	135
DRE	100
ING	226
EIG	135
NTH	171
ASH	143

IGH	140
THI	211
ATI	160
ERI	109
COM	136
ION	260
NIN	207
EEN	196
VEN	190
WEN	153
MEN	131
TIO	221
STO	202
TWO	163
TOP	174
FOR	218
OUR	211
THR	158
PER	115

TER	115
HIR	106
DER	101
DAS	140
ERS	126
ENT	569
GHT	196
EST	176
ORT	146
NET	118
IRT	105
FOU	152
EQU	114
FIV	135
SEV	131
SIX	146
NTY	157

TABLE 11-A.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Government plain-text telegrams arranged according to their absolute frequencies*

TION	218	THIR	104	ASHT	64
EVEN	168	EENT	102	HUND	64
TEEN	163	REQU	98	DRED	63
ENTY	161	HIRT	97	RIOD	63
STOP	154	COMM	93	IVED	62
WENT	153	QUES	87	ENTS	62
NINE	153	UEST	87	FFIC	62
TWEN	152	EQUE	86	FROM	59
THRE	149	NDRE	77	IRTY	59
FOUR	144	OMMA	71	RTEE	59
IGHT	140	LLAR	71	UNDR	59
FIVE	135	OLLA	70	NAUG	56
HREE	134	VENT	70	OURT	56
EIGH	132	DOLL	68	UGHT	56
DASH	132	LARS	68	STAT	54
SEVE	121	THIS	68	AUGH	52
ENTH	114	PERI	67	CENT	52
MENT	111	ERIO	66	FICE	50

TABLE 11-B.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Government plain-text telegrams, arranged first alphabetically according to their initial letters, and then according to their absolute frequencies*

ASHT	64	HREE	134	REQU	98
AUGH	52	HIRT	97	RIOD	63
		HUND	64	RTEE	59
COMM	93				
CENT	52	IGHT	140	STOP	154
		IVED	62	SEVE	121
DASH	132	IRTY	59	STAT	54
DOLL	68				
DRED	63	LLAR	71		
		LARS	68	TION	218
EVEN	168			TEEN	163
ENTY	161	MENT	111	TWEN	152
EIGH	132			THRE	149
ENTH	114	NINE	153	THIR	104
EENT	102	NDRE	77	THIS	68
EQUE	86	NAUG	56		
ERIO	66			UEST	87
ENTS	62	OMMA	71	UNDR	59
		OLLA	70	UGHT	56
FOUR	144	OURT	56		
FIVE	135				
FFIC	62	PERI	67	VENT	70
FROM	59				
FICE	50	QUES	87	WENT	153

TABLE 11–C.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their second letters and then according to their absolute frequencies*

DASH	132
LARS	68
NAUG	56
NDRE	77
TEEN	163
WENT	153
SEVE	121
MENT	111
EENT	102
REQU	98
UEST	87
VENT	70
PERI	67
CENT	52
FFIC	62
IGHT	140
UGHT	56
THRE	149
THIR	104

THIS	68
TION	218
NINE	153
FIVE	135
EIGH	132
HIRT	97
RIOD	63
FICE	50
LLAR	71
OLLA	70
OMMA	71
ENTY	161
ENTH	114
ENTS	62
UNDR	59
FOUR	144
COMM	93
DOLL	68

EQUE	86
HREE	134
ERIO	66
DRED	63
FROM	59
IRTY	59
ASHT	64
STOP	154
RTEE	59
STAT	54
QUES	87
HUND	64
OURT	56
AUGH	52
EVEN	168
IVED	62
TWEN	152

TABLE 11–D.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their third letters and then according to their absolute frequencies*

LLAR	71
STAT	54
FICE	50
UNDR	59
EVEN	168
TEEN	163
TWEN	152
HREE	134
QUES	87
DRED	63
IVED	62
RTEE	59

EIGH	132
AUGH	52
IGHT	140
ASHT	64
UGHT	56
THIR	104
THIS	68
ERIO	66
FFIC	62
OLLA	70
DOLL	68

COMM	93
OMMA	71
WENT	153
NINE	153
MENT	111
EENT	102
VENT	70
HUND	64
CENT	52
TION	218
STOP	154
RIOD	63
FROM	59

Table 11–D, Concluded.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their third letters and then according to their absolute frequencies*

REQU	98	OURT	56	IRTY	59
		DASH	132	FOUR	144
THRE	149	UEST	87	EQUE	86
HIRT	97			NAUG	56
NDRE	77	ENTY	161		
LARS	68	ENTH	114	FIVE	135
PERI	67	ENTS	62	SEVE	121

Table 11–E.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their final letters and then according to their absolute frequencies*

OMMA	71	DASH	132	QUES	87
OLLA	70	EIGH	132	THIS	68
		ENTH	114	LARS	68
		AUGH	52	ENTS	62
FFIC	62				
		PERI	67		
				WENT	153
HUND	64	DOLL	68	IGHT	140
DRED	63			MENT	111
RIOD	63	COMM	93	EENT	102
IVED	62	FROM	59	HIRT	97
				UEST	87
		TION	218	VENT	70
NINE	153	EVEN	168	ASHT	64
THRE	149	TEEN	163	UGHT	56
FIVE	135	TWEN	152	OURT	56
HREE	134			STAT	54
SEVE	121	ERIO	66	CENT	52
EQUE	86				
NDRE	77	STOP	154		
RTEE	59			REQU	98
FICE	50	FOUR	144		
		THIR	104		
		LLAR	71	ENTY	161
NAUG	56	UNDR	59	IRTY	59

TABLE 12.—*Average length of words and messages*

Number of letters in word x	Number of times x-letter word appears	Number of letters
1	378	378
2	973	1, 946
3	1, 307	3, 921
4	1, 635	6, 540
5	1, 410	7, 050
6	1, 143	6, 858
7	1, 009	7, 063
8	717	5, 736
9	476	4, 284
10	274	2, 740
11	161	1, 771
12	86	1, 032
13	23	299
14	23	322
15	4	60
	9, 619	50, 000

(1) Average length of words .... 5.2 Letters.
(2) Average length of messages .... 217 Letters.
(3) Modal (most frequent) length .... 105–114 Letters.
(4) It is extremely unusual to find 5 consecutive letters without at least one vowel.
(5) The average number of letters between vowels is 2.

TABLE 13.—*Frequency of letters of*

	French		German		Italian		Spanish		Portuguese		Japanese (Romaji)	
	f	f^2	f	f^2	f	f^2	f	f^2	f	f^2	f	f^2
A	73. 5	5, 402	46	2, 116	102	10, 400	130	16, 900	27	729	17	289
B	9. 0	81	19	361	9	81	10	100	1	1	3	9
C	35. 2	1, 239	31	961	42	1, 764	42	1, 764	8	64	1	1
D	46. 2	2, 134	55	3, 025	37	1, 369	46	2, 116	11	121	3	9
E	171. 0	29, 240	180	32, 400	125	15, 625	144	20, 736	25	625	11	121
F	13. 1	172	15	225	8	64	7	49	2	4	1	1
G	7. 0	49	30	900	20	400	10	100	2	4	3	9
H	5. 0	25	44	1, 936	22	484	9	81	2	4	10	100
I	69. 3	4, 802	72	5, 184	115	13, 225	71	5, 041	12	144	25	625
J	3. 0	9	6	36	0	0	3	9	0	0	2	4
K	0	0	13	169	0	0	0	0	0	0	16	256
L	49. 2	2, 421	37	1, 369	65	4, 225	55	3, 025	5	25	0	0
M	31. 2	973	20	400	26	676	25	625	9	81	4	16
N	83. 5	6, 972	94	8, 836	65	4, 225	64	4, 096	11	121	14	196
O	66. 3	4, 422	25	625	86	7, 396	84	7, 056	23	529	30	900
P	28. 1	790	7	49	32	1, 024	33	1, 089	6	36	1	1
Q	7. 0	49	0	0	6	36	15	225	2	4	0	0
R	69. 4	4, 816	76	5, 776	66	4, 356	70	4, 900	16	256	9	81
S	69. 3	4, 802	63	3, 969	60	3, 600	77	5, 929	18	324	15	225
T	67. 3	4, 529	66	4, 356	60	3, 600	44	1, 936	9	81	11	121
U	67. 3	4, 529	51	2, 601	30	900	40	1, 600	8	64	17	289
V	18. 1	328	10	100	15	225	7	49	3	9	0	0
W	0	0	16	256	0	0	0	0	0	0	2	4
X	5. 0	25	0	0	0	0	1	1	1	1	0	0
Y	3. 0	9	0	0	0	0	10	100	0	0	4	16
Z	3. 0	9	24	576	9	81	3	9	1	1	1	1
	1, 000. 0	77, 827	1, 000	76, 226	1, 000	73, 756	1, 000	77, 536	202	3, 228	200	3, 274
	$\kappa_p=0.0778$		$\kappa_p=0.0762$		$\kappa_p=0.0738$		$\kappa_p=0.0775$		$\kappa_p=0.0791$		$\kappa_p=0.0819$	

Russian		
	f	f^2
А	15	225
Б	4	16
В	10	100
Г	4	16
Д	6	36
ЕЭ	16	256
Ж	2	4
З	3	9
ІИ	14	196
Й	2	4
К	6	36
Л	8	64
М	6	36
Н	15	225
О	22	484
П	6	36
Р	10	100
С	11	121
Т	12	144
У	5	25
ФѲ	1	1
Х	2	4
Ц	1	1
Ч	3	9
Ш	2	4
Щ	1	1
ЪЬ	4	16
Ы	4	16
Ѣ	4	16
Ю	2	4
Я	4	16
	205	2, 221
	$\kappa_p=0.0529$	

63301—38——9

TABLE 14.—*Czech digraphic table*

[Based on 10,000 digraphs]

First Letters	Second Letters																									
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	10	26	46	70	1	5		16	4	33	50	83	22	120	7	32		33	74	65	8	61			1	46
B	25	1		4	27				14	1		9		4	22	3		37	2	6	31				43	
C	21		2		48		5	87	69	1	37	3	1	22	6	1				20	5	5			3	1
D	23	2	4	7	67			1	32	4	7	14	5	53	45	5		14	4	2	21	5			18	1
E	19	49	92	72		5	8	32	6	39	34	106	76	142	32	46		70	86	86	13	37			3	52
F					9				3						4			3							1	
G	11	1			2				3			1		2	2			4			1					2
H	15	2		3	9					2	4	30	3	13	57	8		6	5	6	11	7			5	1
I	18	17	64	29	8	2	6	9	6	22	30	44	48	62	14	23		17	79	80	6	52			6	29
J	16	1		5	104			1	42	1		1	4	6	15	3			26	4	4				2	
K	47		4	4	42		2		5	4	2	20	3	4	65	4		11	2	28	43	4			55	2
L	54	10	2	2	139	1	2	2	55	2	9	1	2	25	55	2			9	4	27	3			22	6
M	41	5	1	2	42			1	51	2	3	3	8	14	43	10		4	6	6	22	6				11
N	96		9	1	153	2	3		150		23	4	1	10	66	4		3	12	11	35	5			68	2
O	10	63	37	41	4	3	1	15	3	55	25	31	33	35	12	32		46	89	76	77	102			1	49
P	21				18				14			16		1	105	1		127		1	8				2	
Q																										
R	109	1	4	7	97		6	1	72		3	3	12	15	99	1		1	7	3	25	5			19	4
S	18		1	2	74	2		1	58	2	40	14	4	16	34	37		1	2	200	20	20			1	
T	79	4	2	3	166	1			54	8	10	14	7	26	94	13		94	19	5	23	28			24	8
U	23	11	19	32	2		1	7	2	27	17	12	19	27	7	37		12	49	40	4	26				38
V	94	5	4	1	106			3	29		1	9	1	24	42	2		7	16	6	7				51	6
W																										
X																										
Y	13	15	40	7				7	1	12	25	20	25	14	12	26		7	32	20	7	25				20
Z	49	18	1	25	58			5	19	2	14	2	9	27	7	9		3	6	7	9	14			1	

TABLE 15.—*French digraphic table*

SECOND LETTERS

FIRST LETTERS	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	5	21	37	20	3	11	25		103	4		52	24	117		45	11	102	29	69	50	43		1	1	
B	17				8				7			28			14			8	4		2					
C	40	3	13	3	73			27	17			10	3		77			9	3	19	23					
D	62			4	236			2	50				3		19			17	4		44					
E	55	10	84	119	70	34	17	13	16	7		129	117	255	8	68	21	156	294	104	98	34		6	1	6
F	30		1	1	16	13		2	7			8			24			15		2	5					
G	9				37			1	6			1	1	18	1			17		2	5					
H	8				36			1	1						12						6					
I	12	6	10	18	86	7	23		2			69	19	74	45	4	13	45	73	120		24		19		
J	2				8										12						1					
K																										
L	107	1	4	8	338	1		1	32	2		73	4	9	27		4		10	2	24	1			6	
M	34	20		1	119				42				17		23	24					8					
N	44	3	54	58	183	14	7		26	1		4	2	57	47	10	12	9	95	183	17	14			8	1
O		8	8	4			2		48			6	38	167		6		61	34	13	134	3			7	
P	48				46	2		4	6			12			54	10		41	12	7	13					
Q			2		1														1		94					
R	85		23	38	169	4	9	2	64			44	20	10	57	13	9	14	30	46	14	19				
S	58	4	40	84	145	12	1	1	70	2		51	12	5	43	27	16	10	62	64	39	5			1	
T	82		7	71	140		2	5	99	3		31	10	15	30	15	13	69	43	37	26	2				2
U	29	4	24	12	86	13	3		66	7		25	19	55	5	30		93	53	41	1	26		25		1
V	34				56				32						32			9			4					
W																										
X		1	3	5	8			5	4			8	3	1		2		1	5	4						
Y	4		1		8	3							1		1		2		1		1					
Z	5				3	1						2			2			1								

Table 16.—*German digraphic table*

Second Letters

First Letters	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	2	36	33	2		7	22	30	6	1	1	51	21	111		6		47	44	51	94	1				1
B	37	4	1	4	131		1	1	14			11			3			17	13	3	8	2				2
C								248			20															
D	60	5		24	241	4	5	2	115		3	3	2	4	7	4		20	12	2	24	3	5			3
E	19	47	21	51	35	35	41	40	225	5	11	91	42	441	5	10	1	380	159	65	43	11	24			11
F	27	2		8	52	11	5	2	7		2	3		2	6			7	1	14	26					4
G	22	2		13	181	4	4	3	8	2	6	16	1	1	5			11	11	5	10	4	4			8
H	45	6		5	64	2	5	23	15	1	3	30	9	16	14			58	10	54	11	7	8			2
I	5	8	71	16	186	3	41	10			2	27	21	145	10			16	54	79	3	6	1			6
J	4				14										5						2					
K	11				26		2		2			7			15			13		3	11	2	2			
L	45	7	1	20	75	2	8		48		4	48		6	11			2	17	26	24	2	2			2
M	42	6		12	37	7	3	4	35		2	3	22	2	17	4		1	2	11	13	2	1			5
N	68	34	3	237	123	19	102	12	51	5	15	10	18	37	26	8		8	74	68	41	16	25			27
O	2	19	6	8		14	12	14	1		1	16	22	60		4		34	28	10	4	2	3			
P	15				8	8		3	10			3			7	9		20	4	4	1					
Q																					3					
R	57	24	15	66	129	23	14	11	54	3	17	18	22	39	40	8	1	11	64	44	33	16	14			9
S	36	13	68	37	107	4	13	1	46	5	9	7	6	9	41	20		5	72	111	29	9	9			5
T	63	11		28	224	4	21	13	34	1	2	8	5	2	9			35	40	31	27	18	11			43
U		17	21	5	22	29	18	13	3	1	1	5	19	152		8		51	64	20	2	3				2
V	3				59				11						33				2							1
W	33			1	37		1		38						9			2			10					
X																										
Y																			1							
Z	5	1		2	39		1		15			4			3				4	20	50		7			

TABLE 17.—*Italian digraphic table (military text)*

[Based on 10,000 letters]

SECOND LETTERS

FIRST LETTERS	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	25	20	55	80	40	15	30		25			124	65	124	15	40	15	109	85	95	2	40				30
B	15	20			20				20			10									10					
C	60		20		40			100	60						119		2	15			20					
D	30			2	138				138						40						25					
E	40	15	80	85	35	10	55		50			148	35	144	10	114	20	192	172	60	10	25				5
F	10				10				20						10			10			10					
G	15				20		20	5	65			40		5	20			10			10					
H	5				70				30						2											
I	104	15	100	60	70	25	40	5	20			95	60	114	109	50	10	50	100	70	35	20				5
J																										
K																										
L	133	2	20	10	114		10		100			124	15		40	20		15	10	25	15	2				
M	70	2			80				25				10		30	35					5					
N	80	2	40	30	109	5	15		55			2	5	25	85	2			15	124	10	5				40
O	?0	10	40	70	15	15	20	2	35			80	30	172	10	50	10	104	90	25	10	40				
P	60				65				25						70	25		45			25					
Q																					55					
R	124	2	25	20	152		5		90			15	15	15	95	5		20	25	35	15	5				5
S	30		35	5	90		5		139				10		60	15			90	95	30	2				
T	119				100				139						114			60		55	15					
U	35	10	10	10	40	2	5		20			20	5	45	25	10		25	15	20	5					
V	25				45				40						20			10			2	10				
W																										
X																										
Y																										
Z	30				5				55						2											2

Table 18.—*Japanese digraphic table*

[Based on 10,000 letters]

Note.—Long vowel sound indicated by double letter

Second Letter

First Letter	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	6	7	2	20	7	6	17	13	201	8	138		22	167	10	4		124	61	59	4		16		7	16
B	16			1	24				32					1	16					1	21				2	
C	1							24							2											
D	43				16		1		2						45						2					
E	2	3	1	5	1	2	6	3	129	3	47		5	134	12	2		46	19	47	1		14		8	9
F	1			1	1				2						1			1			58					
G	58				23				38						32						20				6	
H	51				9				50						123						42				10	
I	8	14	6	21	12	30	54	50	36	40	149		30	212	74	7		27	165	221	4		28		31	20
J	2				1				46						27						36		1		1	
K	200				60				89		43				160						202				52	
L																										
M	34				37				34		1			2	60				1	1	17				1	
N	94	15	4	20	20	5	23	14	191	17	47		17	64	183	9		15	85	29	4		29		22	17
O	15	38	7	24	10	10	37	43	42	23	237		53	187	356	2		105	158	111	8		45		46	9
P	7				2				2						13	9		1		1	1				3	
Q																										
R	47				42				103	1	1		1		31			1	2		163				18	
S	37		1		108			87	194				1		57			1	8	2	67				1	
T	122				102				52				1		155				3	17	106		1			
U	8	24	5	12	4	10	33	25	25	12	130		47	139	29	1		85	106	67	98		26		22	10
V																										
W	102				1				1						55						1					
X																										
Y	20				4		1		2		1		1	1	147			1			58					
Z	22				29				1						8				1		24				1	

TABLE 19.—*Polish digraphic table*

[Based on 10,000 digraphs]

SECOND LETTERS

FIRST LETTERS	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	5	12	113	77	7	3	25		11	50	65	110	46	107	17	34		69	65	51	12		82		1	52
B	16		2		8				19	2		2		2	17	2		15	1		10				25	
C	22	1	2		31			138	59	26	16	1	2	12	6	4			8	6	6		8		31	118
D	37		2	12	26				4	1	3	26	7	28	65	5		14	11	2	8		10		25	89
E	18	13	15	30	2	2	32	4	9	33	17	15	30	36	16	9		30	28	7	3		16		1	23
F	17				2				8						10			2								
G	11			7	8				13	3		10	3	8	65			9			5		1		1	1
H	5	1	6	3	5	1	1	1	15	2	3	8	5	13	17	14		5	8	7	3		16		1	3
I	76	5	48	15	300	1	3		11	6	16	32	12	28	31	34		7	40	30	6		23			23
J	81	3	2	8	70				26	5	1		5	12	18	4		6	28	1	8		10			8
K	37		5		14	1			83	4	3	5		6	104	6		27	10	29	26		12			2
L	87	4	3	9	69		2		49	1	22			26	38	6			26	9	30		10		14	4
M	33	4	11	4	15	1			67	5	7	3	6	3	44	9		2	10	1	17		7		15	8
N	137	2	21	3	53		1		239		20		1	15	46	4			25	18	4		2		62	3
O	4	35	28	73		8	28		6	47	39	94	29	68	13	38		72	128	31	3		151		1	51
P	32		4	1	9				20			3		2	142			85		1	8		1			
Q																										
R	86		6	3	53		1		2	11	4	5	14	5	103	7			10	20	23		7		21	96
S	19		38		7	2	1		77	13	80	12	5	9	23	30		1	2	117	8		11		10	70
T	84		2	3	62	1	1	3	3	2	12			10	62	4		36	3	2	30		24		37	5
U	7	7	13	13		2	10	4	10	7	13	8	36	4	16			18	36	12	4		22			16
V																										
W	114	1	14	13	40		2		102	3	12	10	3	34	73	11		11	40	8			10		61	10
X																										
Y	2	8	91	4			4	1	1	20	9	38	41	28	4	22		8	28	26	2		29			14
Z	119	12	20	14	117	2	5		58	16	19	13	10	52	46	18		9	7	12	8		16		81	6

TABLE 20.—*Spanish diagraphic table*

SECOND LETTERS

FIRST LETTERS	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	30	47	95	124	42	11	25	15	17	6		108	56	130	10	52	30	157	151	36	32	12			16	9
B	28				16				27	1		14		2	6			20	2	2	6					
C	68		13		49			28	100			7		2	107			5		26	43					
D	74			9	139				58			3	3	2	83	2		3	1		10	2			1	
E	34	8	84	67	48	10	34	19	4	15		155	54	233	12	29	9	196	236	36	17	16		18	6	11
F	13				19	1			22			1		2	10			9		1	15					
G	33				19				8					2	19			11			33					
H	41				14				11						21											
I	92	6	81	53	77	12	21		4	11		36	36	74	113	15	2	42	69	55	5	11			3	12
J	8				17				3						4						4					
K																										
L	184		14	12	72	2	14		31			31	10	3	75	27	5	3	8	13	13	5			1	
M	57	11			50				50					2	53	39		1			12					
N	60	3	52	56	62	15	9	4	60	4		34	7	3	67	16	7	1	34	105	19	8			2	7
O	24	19	27	53	47	10	5	7	6	2		41	49	159	6	40	10	102	202	32	5	21		1	18	3
P	52		1		63				28			17			68			60		2	28					
Q																					111					
R	156	2	25	13	148	4	5	1	77			29	14	16	72	12	12	29	24	37	17	7			5	7
S	84	4	33	46	106	15	3	7	64	2		12	24		47	60	24	4	22	123	38	6		1	22	2
T	106		1		94				84						83	2	1	71	2	1	17					
U	25	6	13	16	153	3	9		16			20	13	39	5	10	3	15	24	23					1	
V	16				30				29						19											
W																										
X									6				3			12				2						
Y	19	3	7	8	20	2						3	6		8	3	3	2	3			1			3	
Z	20		4												3		1									

TABLE 21.—*Swedish diagraphic table*

SECOND LETTERS

FIRST LETTERS	A	B	C	D	E	F	G	H	I	J	K	L	N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	34	15	4	103	17	37	57	30	11	6	48	84	98	166	15	24		238	73	140	14	81			1	
B	23	4			50			1	32			15			5			12	1						15	
C					4			60			39				2											
D	83	5		25	281	14	1	3	30		3	9	5	15	18	2		12	29	11	4	8			1	
E	31	7	6	81	18	51	26	12	11		36	127	18	186	20	9		190	57	162	9	18		12		1
F	35	1		3	15	9			11			13			125			27	1	5	13				1	
G	85	5	1	10	69	4	20	8	10	3	8	2	5	24	19	3		25	21	31	1	4			2	
H	74			3	33	2		1	1	4	4	1	1		26	4		6	6	3	13	2			1	
I	4	4	29	45	19	5	107	7			15	72	2	112	27	3		8	39	23	1	11				
J	20			1	2							4		1	7				1	3	3					
K	78	1			33		1	1	8		1	11	5	6	22			14	2	22	13	29			2	
L	82	10	1	15	82	11	3	2	62	7	10	108	10	17	14	5		3	49	25	15	14			13	
M	78	1	1	4	102	12	3	4	6		4	10	22	3	35	3		5	10	9	4				3	
N	159	16	2	137	44	16	93	19	60	3	15	3	25	55	42	11		4	71	49	15	18			9	2
O	4	10	46	22	3	7	21		5	5	7	11	77	44	2	9		167	9	22	2	29				
P	35			1	23		3	1	1			10	7	1	7	44		11	3	6						
Q																										
R	161	44		36	115	30	9	11	74	3	9	22	55	45	38	1		8	35	66	39	14			8	
S	75	5	8	6	63	10	2	4	49	2	46	18	7	6	29	9		2	22	127	9	14			2	
T	135	26	1	10	136	28	6	11	99	10	4	13	18	14	53	11		50	54	142	27	20			24	
U	2			3	1	1	2	1	1		1	16	5	36		24		12	9	60		5				
V	105			19	38	3		2	56		3	1	2	1	12			13	12	7	3					
W																										
X																										
Y	2		5	10	1	1	16	1			3	1	3	6		1		1	13	13						
Z					1				1										1							

TABLE 22.—*Checkerboard individual frequencies* [1]

[Based on a count of 5,000 digraphs]

		P_1					C_1		
A	B	C	D	E	244	225	375	394	197
F	G	H	I J	K	125	98	193	271	95
L	M	N	O	P	229	199	188	350	251
Q	R	S	T	U	148	162	258	427	295
V	W	X	Y	Z	42	12	34	91	97
212	317	358	308	249	A	B	C	D	E
120	108	216	256	85	F	G	H	I J	K
216	140	152	435	269	L	M	N	O	P
206	121	306	364	284	Q	R	S	T	U
38	29	21	147	43	V	W	X	Y	Z
		C_2					P_2		

[1] The numbers in the C_1 C_2 squares represent the frequency of the individual components of the cipher digraph used to replace a given P_1 P_2 digraph in accordance with a digraphic checkerboard system where P_1 and P_2 are the plain-text squares.

Section IX

STATISTICAL TABLES

Table I [1]

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

x	0	1	2	3	4	5	6	7	8	9
0.0	0.3989	0.3989	0.3989	0.3988	0.3986	0.3984	0.3982	0.3980	0.3977	0.3973
0.1	0.3970	0.3965	0.3961	0.3956	0.3951	0.3945	0.3939	0.3932	0.3925	0.3918
0.2	0.3910	0.3902	0.3894	0.3885	0.3876	0.3867	0.3857	0.3847	0.3836	0.3825
0.3	0.3814	0.3802	0.3790	0.3778	0.3765	0.3752	0.3739	0.3725	0.3712	0.3697
0.4	0.3683	0.3668	0.3653	0.3637	0.3621	0.3605	0.3589	0.3572	0.3555	0.3538
0.5	0.3521	0.3503	0.3485	0.3467	0.3448	0.3429	0.3410	0.3391	0.3372	0.3352
0.6	0.3332	0.3312	0.3292	0.3271	0.3251	0.3230	0.3209	0.3187	0.3166	0.3144
0.7	0.3123	0.3101	0.3079	0.3056	0.3034	0.3011	0.2989	0.2966	0.2943	0.2920
0.8	0.2897	0.2874	0.2850	0.2827	0.2803	0.2780	0.2756	0.2732	0.2709	0.2685
0.9	0.2661	0.2637	0.2613	0.2589	0.2565	0.2541	0.2516	0.2492	0.2468	0.2444
1.0	0.2420	0.2396	0.2371	0.2347	0.2323	0.2299	0.2275	0.2251	0.2227	0.2203
1.1	0.2179	0.2155	0.2131	0.2107	0.2083	0.2059	0.2036	0.2012	0.1989	0.1965
1.2	0.1942	0.1919	0.1895	0.1872	0.1849	0.1826	0.1804	0.1781	0.1758	0.1736
1.3	0.1714	0.1691	0.1669	0.1647	0.1626	0.1604	0.1582	0.1561	0.1539	0.1518
1.4	0.1497	0.1476	0.1456	0.1435	0.1415	0.1394	0.1374	0.1354	0.1334	0.1315
1.5	0.1295	0.1276	0.1257	0.1238	0.1219	0.1200	0.1182	0.1163	0.1145	0.1127
1.6	0.1109	0.1092	0.1074	0.1057	0.1040	0.1023	0.1006	0.0989	0.0973	0.0957
1.7	0.0940	0.0925	0.0909	0.0893	0.0878	0.0863	0.0848	0.0833	0.0818	0.0804
1.8	0.0790	0.0775	0.0761	0.0748	0.0734	0.0721	0.0707	0.0694	0.0681	0.0669
1.9	0.0656	0.0644	0.0632	0.0620	0.0608	0.0596	0.0584	0.0573	0.0562	0.0551
2.0	0.0540	0.0529	0.0519	0.0508	0.0498	0.0488	0.0478	0.0468	0.0459	0.0449
2.1	0.0440	0.0431	0.0422	0.0413	0.0404	0.0396	0.0387	0.0379	0.0371	0.0363
2.2	0.0355	0.0347	0.0339	0.0332	0.0325	0.0317	0.0310	0.0303	0.0297	0.0290
2.3	0.0283	0.0277	0.0270	0.0264	0.0258	0.0252	0.0246	0.0241	0.0235	0.0229
2.4	0.0224	0.0219	0.0213	0.0208	0.0203	0.0198	0.0194	0.0189	0.0184	0.0180
2.5	0.0175	0.0171	0.0167	0.0163	0.0158	0.0154	0.0151	0.0147	0.0143	0.0139
2.6	0.0136	0.0132	0.0129	0.0126	0.0122	0.0119	0.0116	0.0113	0.0110	0.0107
2.7	0.0104	0.0101	0.0099	0.0096	0.0093	0.0091	0.0088	0.0086	0.0084	0.0081
2.8	0.0079	0.0077	0.0075	0.0073	0.0071	0.0069	0.0067	0.0065	0.0063	0.0061
2.9	0.0060	0.0058	0.0056	0.0055	0.0053	0.0051	0.0050	0.0048	0.0047	0.0046
3	0.0044	0.0033	0.0024	0.0017	0.0012	0.0009	0.0006	0.0004	0.0003	0.0002
4	0.0001	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
x	0	1	2	3	4	5	6	7	8	9

[1] Copied from, Vorlesungen über Die Grundzüge der Mathematischen Statistik, C. V. L. Charlier, Lund 1920.

Table II[1]

$$P(-\infty, x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} dx e^{-x^2/2}$$

x	0	1	2	3	4	5	6	7	8	9
−2	0.0228	0.0179	0.0139	0.0107	0.0082	0.0063	0.0047	0.0035	0.0026	0.0019
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0321	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
+0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
+0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
+0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
+0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
+0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
+0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
+0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
+0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
+0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
+0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
+1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
+1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
+1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
+1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
+1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
+1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
+1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
+1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
+1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
+1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
+2	0.9772	0.9821	0.9861	0.9893	0.9918	0.9937	0.9953	0.9965	0.9974	0.9981
x	0	1	2	3	4	5	6	7	8	9

Example: For $x = -1.53$, $P(-\infty, -1.53) = 0.0630$

[1] Copied from, Vorlesungen über Die Grundzüge der Mathematischen Statistik, C. V. L. Charlier, Lund 1920.

TABLE III.[1] *Tables of $e^{-m}m^x/x!$: General Term of Poisson's Exponential Expansion ("Law of Small Numbers").*

x	m = 0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0	x
0	·904837	·818731	·740818	·670320	·606531	·548812	·496585	·449329	·406570	·367879	0
1	·090484	·163746	·222245	·268128	·303265	·329287	·347610	·359463	·365913	·367879	1
2	·004524	·016375	·033337	·053626	·075816	·098786	·121663	·143785	·164661	·183940	2
3	·000151	·001092	·003334	·007150	·012636	·019757	·028388	·038343	·049398	·061313	3
4	·000004	·000055	·000250	·000715	·001580	·002964	·004968	·007669	·011115	·015328	4
5	—	·000002	·000015	·000057	·000158	·000356	·000696	·001227	·002001	·003066	5
6	—	—	·000001	·000004	·000013	·000036	·000081	·000164	·000300	·000511	6
7	—	—	—	—	·000001	·000003	·000008	·000019	·000039	·000073	7
8	—	—	—	—	—	—	·000001	·000002	·000004	·000009	8
9	—	—	—	—	—	—	—	—	—	·000001	9

x	1·1	1·2	1·3	1·4	1·5	1·6	1·7	1·8	1·9	2·0	x
0	·332871	·301194	·272532	·246597	·223130	·201897	·182684	·165299	·149569	·135335	0
1	·366158	·361433	·354291	·345236	·334695	·323034	·310562	·297538	·284180	·270671	1
2	·201387	·216860	·230289	·241665	·251021	·258428	·263978	·267784	·269971	·270671	2
3	·073842	·086744	·099792	·112777	·125510	·137828	·149587	·160671	·170982	·180447	3
4	·020307	·026023	·032432	·039472	·047067	·055131	·063575	·072302	·081216	·090224	4
5	·004467	·006246	·008432	·011052	·014120	·017642	·021615	·026029	·030862	·036089	5
6	·000819	·001249	·001827	·002579	·003530	·004705	·006124	·007809	·009773	·012030	6
7	·000129	·000214	·000339	·000516	·000756	·001075	·001487	·002008	·002653	·003437	7
8	·000018	·000032	·000055	·000090	·000142	·000215	·000316	·000452	·000630	·000859	8
9	·000002	·000004	·000008	·000014	·000024	·000038	·000060	·000090	·000133	·000191	9
10	—	·000001	·000001	·000002	·000004	·000006	·000010	·000016	·000025	·000038	10
11	—	—	—	—	—	·000001	·000002	·000003	·000004	·000007	11
12	—	—	—	—	—	—	—	—	·000001	·000001	12

x	2·1	2·2	2·3	2·4	2·5	2·6	2·7	2·8	2·9	3·0	x
0	·122456	·110803	·100259	·090718	·082085	·074274	·067206	·060810	·055023	·049787	0
1	·257159	·243767	·230595	·217723	·205212	·193111	·181455	·170268	·159567	·149361	1
2	·270016	·268144	·265185	·261268	·256516	·251045	·244964	·238375	·231373	·224042	2
3	·189012	·196639	·203308	·209014	·213763	·217572	·220468	·222484	·223660	·224042	3
4	·099231	·108151	·116902	·125409	·133602	·141422	·148816	·155739	·162154	·168031	4
5	·041677	·047587	·053775	·060196	·066801	·073539	·080360	·087214	·094049	·100819	5
6	·014587	·017448	·020614	·024078	·027834	·031867	·036162	·040700	·045457	·050409	6
7	·004376	·005484	·006773	·008255	·009941	·011836	·013948	·016280	·018832	·021604	7
8	·001149	·001508	·001947	·002477	·003106	·003847	·004708	·005698	·006827	·008102	8
9	·000268	·000369	·000498	·000660	·000863	·001111	·001412	·001773	·002200	·002701	9
10	·000056	·000081	·000114	·000158	·000216	·000289	·000381	·000496	·000638	·000810	10
11	·000011	·000016	·000024	·000035	·000049	·000068	·000094	·000126	·000168	·000221	11
12	·000002	·000003	·000005	·000007	·000010	·000015	·000021	·000029	·000041	·000055	12
13	—	·000001	·000001	·000001	·000002	·000003	·000004	·000006	·000009	·000013	13
14	—	—	—	—	—	·000001	·000001	·000001	·000002	·000003	14
15	—	—	—	—	—	—	—	—	—	·000001	15

[1] Copied from, Tables for Statisticians and Biometricians, Edited by Karl Pearson, Part I, 2nd Ed., Cambridge University.

TABLE III—(*continued*).

x	*m* 3·1	3·2	3·3	3·4	3·5	3·6	3·7	3·8	3·9	4·0	*x*
0	·045049	·040762	·036883	·033373	·030197	·027324	·024724	·022371	·020242	·018316	0
1	·139653	·130439	·121714	·113469	·105691	·098365	·091477	·085009	·078943	·073263	1
2	·216461	·208702	·200829	·192898	·184959	·177058	·169233	·161517	·153940	·146525	2
3	·223677	·222616	·220912	·218617	·215785	·212469	·208720	·204588	·200122	·195367	3
4	·173350	·178093	·182252	·185825	·188812	·191222	·193066	·194359	·195119	·195367	4
5	·107477	·113979	·120286	·126361	·132169	·137680	·142869	·147713	·152193	·156293	5
6	·055530	·060789	·066158	·071604	·077098	·082608	·088102	·093551	·098925	·104196	6
7	·024592	·027789	·031189	·034779	·038549	·042484	·046568	·050785	·055115	·059540	7
8	·009529	·011116	·012865	·014781	·016865	·019118	·021538	·024123	·026869	·029770	8
9	·003282	·003952	·004717	·005584	·006559	·007647	·008854	·010185	·011643	·013231	9
10	·001018	·001265	·001557	·001899	·002296	·002753	·003276	·003870	·004541	·005292	10
11	·000287	·000368	·000467	·000587	·000730	·000901	·001102	·001337	·001610	·001925	11
12	·000074	·000098	·000128	·000166	·000213	·000270	·000340	·000423	·000523	·000642	12
13	·000018	·000024	·000033	·000043	·000057	·000075	·000097	·000124	·000157	·000197	13
14	·000004	·000006	·000008	·000011	·000014	·000019	·000026	·000034	·000044	·000056	14
15	·000001	·000001	·000002	·000002	·000003	·000005	·000006	·000009	·000011	·000015	15
16	—	—	—	·000001	·000001	·000001	·000001	·000002	·000003	·000004	16
17	—	—	—	—	—	—	—	—	·000001	·000001	17
x	4·1	4·2	4·3	4·4	4·5	4·6	4·7	4·8	4·9	5·0	*x*
0	·016573	·014996	·013569	·012277	·011109	·010052	·009095	·008230	·007447	·006738	0
1	·067948	·062981	·058345	·054020	·049990	·046238	·042748	·039503	·036488	·033690	1
2	·139293	·132261	·125441	·118845	·112479	·106348	·100457	·094807	·089396	·084224	2
3	·190368	·185165	·179799	·174305	·168718	·163068	·157383	·151691	·146014	·140374	3
4	·195127	·194424	·193284	·191736	·189808	·187528	·184925	·182029	·178867	·175467	4
5	·160004	·163316	·166224	·168728	·170827	·172525	·173830	·174748	·175290	·175467	5
6	·109336	·114321	·119127	·123734	·128120	·132270	·136167	·139798	·143153	·146223	6
7	·064040	·068593	·073178	·077775	·082363	·086920	·091426	·095862	·100207	·104445	7
8	·032820	·036011	·039333	·042776	·046329	·049979	·053713	·057517	·061377	·065278	8
9	·014951	·016805	·018793	·020913	·023165	·025545	·028050	·030676	·033416	·036266	9
10	·006130	·007058	·008081	·009202	·010424	·011751	·013184	·014724	·016374	·018133	10
11	·002285	·002695	·003159	·003681	·004264	·004914	·005633	·006425	·007294	·008242	11
12	·000781	·000943	·001132	·001350	·001599	·001884	·002206	·002570	·002978	·003434	12
13	·000246	·000305	·000374	·000457	·000554	·000667	·000798	·000949	·001123	·001321	13
14	·000072	·000091	·000115	·000144	·000178	·000219	·000268	·000325	·000393	·000472	14
15	·000020	·000026	·000033	·000042	·000053	·000067	·000084	·000104	·000128	·000157	15
16	·000005	·000007	·000009	·000012	·000015	·000019	·000025	·000031	·000039	·000049	16
17	·000001	·000902	·000002	·000003	·000004	·000005	·000007	·000009	·000011	·000014	17
18	—	—	·000001	·000001	·000001	·000001	·000002	·000002	·000003	·000004	18
19	—	—	—	—	—	—	—	·000001	·000001	·000001	19
x	5·1	5·2	5·3	5·4	5·5	5·6	5·7	5·8	5·9	6·0	*x*
0	·006097	·005517	·004992	·004517	·004087	·003698	·003346	·003028	·002739	·002479	0
1	·031093	·028686	·026455	·024390	·022477	·020708	·019072	·017560	·016163	·014873	1
2	·079288	·074584	·070107	·065852	·061812	·057982	·054355	·050923	·047680	·044618	2
3	·134790	·129279	·123856	·118533	·113323	·108234	·103275	·098452	·093771	·089235	3

TABLE III—(*continued*).

x	*m* 5·1	5·2	5·3	5·4	5·5	5·6	5·7	5·8	5·9	6·0	*x*
4	·171857	·168063	·164109	·160020	·155819	·151528	·147167	·142755	·138312	·133853	4
5	·175294	·174785	·173955	·172821	·171401	·169711	·167770	·165596	·163208	·160623	5
6	·149000	·151480	·153660	·155539	·157117	·158397	·159382	·160076	·160488	·160623	6
7	·108557	·112528	·116343	·119987	·123449	·126717	·129782	·132635	·135268	·137677	7
8	·069205	·073143	·077077	·080991	·084871	·088702	·092470	·096160	·099760	·103258	8
9	·039216	·042261	·045390	·048595	·051866	·055192	·058564	·061970	·065398	·068838	9
10	·020000	·021976	·024057	·026241	·028526	·030908	·033382	·035943	·038585	·041303	10
11	·009273	·010388	·011591	·012882	·014263	·015735	·017298	·018952	·020696	·022529	11
12	·003941	·004502	·005119	·005797	·006537	·007343	·008216	·009160	·010175	·011264	12
13	·001546	·001801	·002087	·002408	·002766	·003163	·003603	·004087	·004618	·005199	13
14	·000563	·000669	·000790	·000929	·001087	·001265	·001467	·001693	·001946	·002228	14
15	·000191	·000232	·000279	·000334	·000398	·000472	·000557	·000655	·000766	·000891	15
16	·000061	·000075	·000092	·000113	·000137	·000165	·000199	·000237	·000282	·000334	16
17	·000018	·000023	·000029	·000036	·000044	·000054	·000067	·000081	·000098	·000118	17
18	·000005	·000007	·000008	·000011	·000014	·000017	·000021	·000026	·000032	·000039	18
19	·000001	·000002	·000002	·000003	·000004	·000005	·000006	·000008	·000010	·000012	19
20	—	—	·000001	·000001	·000001	·000001	·000002	·000002	·000003	·000004	20
21	—	—	—	—	—	—	—	·000001	·000001	·000001	21

x	6·1	6·2	6·3	6·4	6·5	6·6	6·7	6·8	6·9	7·0	*x*
0	·002243	·002029	·001836	·001662	·001503	·001360	·001231	·001114	·001008	·000912	0
1	·013682	·012582	·011569	·010634	·009772	·008978	·008247	·007574	·006954	·006383	1
2	·041729	·039006	·036441	·034029	·031760	·029629	·027628	·025751	·023990	·022341	2
3	·084848	·080612	·076527	·072595	·068814	·065183	·061702	·058368	·055178	·052129	3
4	·129393	·124948	·120530	·116151	·111822	·107553	·103351	·099225	·095182	·091226	4
5	·157860	·154936	·151868	·148674	·145369	·141969	·138490	·134946	·131351	·127717	5
6	·160491	·160100	·159461	·158585	·157483	·156166	·154648	·152939	·151053	·149003	6
7	·139856	·141803	·143515	·144992	·146234	·147243	·148020	·148569	·148895	·149003	7
8	·106640	·109897	·113018	·115994	·118815	·121475	·123967	·126284	·128422	·130377	8
9	·072278	·075707	·079113	·082484	·085811	·089082	·092286	·095415	·098457	·101405	9
10	·044090	·046938	·049841	·052790	·055777	·058794	·061832	·064882	·067935	·070983	10
11	·024450	·026456	·028545	·030714	·032959	·035276	·037661	·040109	·042614	·045171	11
12	·012429	·013669	·014986	·016381	·017853	·019402	·021028	·022728	·024503	·026350	12
13	·005832	·006519	·007263	·008064	·008926	·009850	·010837	·011889	·013005	·014188	13
14	·002541	·002887	·003268	·003687	·004144	·004644	·005186	·005774	·006410	·007094	14
15	·001033	·001193	·001373	·001573	·001796	·002043	·002317	·002618	·002949	·003311	15
16	·000394	·000462	·000540	·000629	·000730	·000843	·000970	·001113	·001272	·001448	16
17	·000141	·000169	·000200	·000237	·000279	·000327	·000382	·000445	·000516	·000596	17
18	·000048	·000058	·000070	·000084	·000101	·000120	·000142	·000168	·000198	·000232	18
19	·000015	·000019	·000023	·000028	·000034	·000042	·000050	·000060	·000072	·000085	19
20	·000005	·000006	·000007	·000009	·000011	·000014	·000017	·000020	·000025	·000030	20
21	·000001	·000002	·000002	·000003	·000003	·000004	·000005	·000007	·000008	·000010	21
22	—	—	·000001	·000001	·000001	·000001	·000002	·000002	·000003	·000003	22
23	—	—	—	—	—	—	—	·000001	·000001	·000001	23

TABLE III—(*continued*).

x	*m* 7·1	7·2	7·3	7·4	7·5	7·6	7·7	7·8	7·9	8·0	*x*
0	·000825	·000747	·000676	·000611	·000553	·000500	·000453	·000410	·000371	·000335	0
1	·005858	·005375	·004931	·004523	·004148	·003803	·003487	·003196	·002929	·002684	1
2	·020797	·019352	·018000	·016736	·015555	·014453	·013424	·012464	·011569	·010735	2
3	·049219	·046444	·043799	·041282	·038889	·036614	·034455	·032407	·030465	·028626	3
4	·087364	·083598	·079934	·076372	·072916	·069567	·066326	·063193	·060169	·057252	4
5	·124057	·120382	·116703	·113031	·109375	·105742	·102142	·098581	·095067	·091604	5
6	·146800	·144458	·141989	·139405	·136718	·133940	·131082	·128156	·125171	·122138	6
7	·148897	·148586	·148074	·147371	·146484	·145421	·144191	·142802	·141264	·139587	7
8	·132146	·133727	·135118	·136318	·137329	·138150	·138783	·139232	·139499	·139587	8
9	·104249	·106982	·109596	·112084	·114440	·116660	·118737	·120668	·122449	·124077	9
10	·074017	·077027	·080005	·082942	·085830	·088661	·091427	·094121	·096735	·099262	10
11	·047774	·050418	·053094	·055797	·058521	·061257	·063999	·066740	·069473	·072190	11
12	·028267	·030251	·032299	·034408	·036575	·038796	·041066	·043381	·045736	·048127	12
13	·015438	·016754	·018137	·019586	·021101	·022681	·024324	·026029	·027794	·029616	13
14	·007829	·008616	·009457	·010353	·011304	·012312	·013378	·014502	·015684	·016924	14
15	·003706	·004136	·004603	·005107	·005652	·006238	·006867	·007541	·008260	·009026	15
16	·001644	·001861	·002100	·002362	·002649	·002963	·003305	·003676	·004078	·004513	16
17	·000687	·000788	·000902	·001028	·001169	·001325	·001497	·001687	·001895	·002124	17
18	·000271	·000315	·000366	·000423	·000487	·000559	·000640	·000731	·000832	·000944	18
19	·000101	·000119	·000141	·000165	·000192	·000224	·000259	·000300	·000346	·000397	19
20	·000036	·000043	·000051	·000061	·000072	·000085	·000100	·000117	·000137	·000159	20
21	·000012	·000015	·000018	·000021	·000026	·000031	·000037	·000043	·000051	·000061	21
22	·000004	·000005	·000006	·000007	·000009	·000011	·000013	·000015	·000018	·000022	22
23	·000001	·000002	·000002	·000002	·000003	·000004	·000004	·000005	·000006	·000008	23
24	—	—	·000001	·000001	·000001	·000001	·000001	·000002	·000002	·000003	24
25	—	—	—	—	—	—	—	·000001	·000001	·000001	25
x	8·1	8·2	8·3	8·4	8·5	8·6	8·7	8·8	8·9	9·0	*x*
0	·000304	·000275	·000249	·000225	·000203	·000184	·000167	·000151	·000136	·000123	0
1	·002459	·002252	·002063	·001889	·001729	·001583	·001449	·001326	·001214	·001111	1
2	·009958	·009234	·008560	·007933	·007350	·006808	·006304	·005836	·005402	·004998	2
3	·026885	·025239	·023683	·022213	·020826	·019517	·018283	·017120	·016025	·014994	3
4	·054443	·051740	·049142	·046648	·044255	·041961	·039765	·037664	·035656	·033737	4
5	·088198	·084854	·081576	·078368	·075233	·072174	·069192	·066289	·063467	·060727	5
6	·119067	·115967	·112847	·109716	·106581	·103449	·100328	·097224	·094143	·091090	6
7	·137778	·135848	·133805	·131659	·129419	·127094	·124693	·122224	·119696	·117116	7
8	·139500	·139244	·138823	·138242	·137508	·136626	·135604	·134446	·133161	·131756	8
9	·125550	·126866	·128025	·129026	·129869	·130554	·131084	·131459	·131682	·131756	9
10	·101696	·104031	·106261	·108382	·110388	·112277	·114043	·115684	·117197	·118580	10
11	·074885	·077550	·080179	·082764	·085300	·087780	·090197	·092547	·094823	·097020	11
12	·050547	·052993	·055457	·057935	·060421	·062909	·065393	·067868	·070327	·072765	12
13	·031495	·033426	·035407	·037435	·039506	·041617	·043763	·045941	·048147	·050376	13
14	·018222	·019578	·020991	·022461	·023986	·025565	·027196	·028877	·030608	·032384	14
15	·009840	·010703	·011615	·012578	·013592	·014657	·015773	·016941	·018161	·019431	15
16	·004981	·005485	·006025	·006604	·007221	·007878	·008577	·009318	·010102	·010930	16
17	·002373	·002646	·002942	·003263	·003610	·003985	·004389	·004823	·005289	·005786	17
18	·001068	·001205	·001356	·001523	·001705	·001904	·002121	·002358	·002615	·002893	18
19	·000455	·000520	·000593	·000673	·000763	·000862	·000971	·001092	·001225	·001370	19
20	·000184	·000213	·000246	·000283	·000324	·000371	·000423	·000481	·000545	·000617	20

TABLE III—(*continued*).

x	*m*										*x*
	8·1	*8·2*	*8·3*	*8·4*	*8·5*	*8·6*	*8·7*	*8·8*	*8·9*	*9·0*	
21	·000071	·000083	·000097	·000113	·000131	·000152	·000175	·000201	·000231	·000264	*21*
22	·000026	·000031	·000037	·000043	·000051	·000059	·000069	·000081	·000093	·000108	*22*
23	·000009	·000011	·000013	·000016	·000019	·000022	·000026	·000031	·000036	·000042	*23*
24	·000003	·000004	·000005	·000006	·000007	·000008	·000009	·000011	·000013	·000016	*24*
25	·000001	·000001	·000002	·000002	·000002	·000003	·000003	·000004	·000005	·000006	*25*
26	—	—	—	·000001	·000001	·000001	·000001	·000001	·000002	·000002	*26*
27	—	—	—	—	—	—	—	—	·000001	·000001	*27*
x	*9·1*	*9·2*	*9·3*	*9·4*	*9·5*	*9·6*	*9·7*	*9·8*	*9·9*	*10·0*	*x*
0	·000112	·000101	·000091	·000083	·000075	·000068	·000061	·000055	·000050	·000045	*0*
1	·001016	·000930	·000850	·000778	·000711	·000650	·000594	·000543	·000497	·000454	*1*
2	·004624	·004276	·003954	·003655	·003378	·003121	·002883	·002663	·002459	·002270	*2*
3	·014025	·013113	·012256	·011452	·010696	·009987	·009322	·008698	·008114	·007567	*3*
4	·031906	·030160	·028496	·026911	·025403	·023969	·022606	·021311	·020082	·018917	*4*
5	·058069	·055494	·053002	·050593	·048266	·046020	·043855	·041770	·039763	·037833	*5*
6	·088072	·085091	·082154	·079262	·076421	·073632	·070899	·068224	·065609	·063055	*6*
7	·114493	·111834	·109147	·106438	·103714	·100981	·098246	·095514	·092790	·090079	*7*
8	·130236	·128609	·126883	·125065	·123160	·121178	·119123	·117004	·114827	·112599	*8*
9	·131683	·131467	·131113	·130623	·130003	·129256	·128388	·127405	·126310	·125110	*9*
10	·119832	·120950	·121935	·122786	·123502	·124086	·124537	·124857	·125047	·125110	*10*
11	·099133	·101158	·103090	·104926	·106661	·108293	·109819	·111236	·112542	·113736	*11*
12	·075176	·077555	·079895	·082192	·084440	·086634	·088770	·090843	·092847	·094780	*12*
13	·052623	·054885	·057156	·059431	·061706	·063976	·066236	·068481	·070707	·072908	*13*
14	·034205	·036067	·037968	·039904	·041872	·043869	·045892	·047937	·050000	·052077	*14*
15	·020751	·022121	·023540	·025006	·026519	·028076	·029677	·031319	·033000	·034718	*15*
16	·011802	·012720	·013683	·014691	·015746	·016846	·017992	·019183	·020419	·021699	*16*
17	·006318	·006884	·007485	·008123	·008799	·009513	·010266	·011058	·011891	·012764	*17*
18	·003194	·003518	·003867	·004242	·004644	·005074	·005532	·006021	·006540	·007091	*18*
19	·001530	·001704	·001893	·002099	·002322	·002563	·002824	·003105	·003408	·003732	*19*
20	·000696	·000784	·000880	·000986	·001103	·001230	·001370	·001522	·001687	·001866	*20*
21	·000302	·000343	·000390	·000442	·000499	·000563	·000633	·000710	·000795	·000889	*21*
22	·000125	·000144	·000165	·000189	·000215	·000245	·000279	·000316	·000358	·000404	*22*
23	·000049	·000057	·000067	·000077	·000089	·000102	·000118	·000135	·000154	·000176	*23*
24	·000019	·000022	·000026	·000030	·000035	·000041	·000048	·000055	·000064	·000073	*24*
25	·000007	·000008	·000010	·000011	·000013	·000016	·000018	·000022	·000025	·000029	*25*
26	·000002	·000003	·000003	·000004	·000005	·000006	·000007	·000008	·000010	·000011	*26*
27	·000001	·000001	·000001	·000001	·000002	·000002	·000002	·000003	·000004	·000004	*27*
28	—	—	—	—	·000001	·000001	·000001	·000001	·000001	·000001	*28*
29	—	—	—	—	—	—	—	—	—	·000001	*29*
x	*10·1*	*10·2*	*10·3*	*10·4*	*10·5*	*10·6*	*10·7*	*10·8*	*10·9*	*11·0*	*x*
0	·000041	·000037	·000034	·000030	·000028	·000025	·000023	·000020	·000018	·000017	*0*
1	·000415	·000379	·000346	·000317	·000289	·000264	·000241	·000220	·000201	·000184	*1*
2	·002095	·001934	·001784	·001646	·001518	·001400	·001291	·001190	·001097	·001010	*2*
3	·007054	·006574	·006125	·005705	·005313	·004946	·004603	·004283	·003984	·003705	*3*

TABLE III—(*continued*).

x	*m*										*x*
	10·1	*10·2*	*10·3*	*10·4*	*10·5*	*10·6*	*10·7*	*10·8*	*10·9*	*11·0*	
4	·017811	·016764	·015773	·014834	·013946	·013107	·012313	·011564	·010856	·010189	*4*
5	·035979	·034199	·032492	·030855	·029287	·027786	·026350	·024978	·023667	·022415	*5*
6	·060565	·058139	·055777	·053482	·051252	·049089	·046991	·044960	·042995	·041095	*6*
7	·087387	·084716	·082072	·079458	·076878	·074334	·071830	·069367	·066949	·064577	*7*
8	·110326	·108013	·105668	·103296	·100902	·098493	·096072	·093646	·091218	·088794	*8*
9	·123810	·122415	·120931	·119364	·117720	·116003	·114219	·112375	·110475	·108526	*9*
10	·125048	·124863	·124559	·124139	·123606	·122963	·122215	·121365	·120418	·119378	*10*
11	·114817	·115782	·116633	·117368	·117987	·118492	·118882	·119159	·119323	·119378	*11*
12	·096637	·098415	·100110	·101719	·103239	·104667	·106003	·107243	·108386	·109430	*12*
13	·075080	·077218	·079318	·081375	·083385	·085344	·087248	·089094	·090877	·092595	*13*
14	·054165	·056259	·058355	·060450	·062539	·064618	·066683	·068730	·070754	·072753	*14*
15	·036471	·038256	·040071	·041912	·043777	·045663	·047567	·049485	·051415	·053352	*15*
16	·023022	·024388	·025795	·027243	·028729	·030252	·031810	·033403	·035026	·036680	*16*
17	·013678	·014633	·015629	·016666	·017744	·018863	·020022	·021220	·022458	·023734	*17*
18	·007675	·008292	·008943	·009629	·010351	·011108	·011902	·012732	·013600	·014504	*18*
19	·004080	·004451	·004848	·005271	·005720	·006197	·006703	·007237	·007802	·008397	*19*
20	·002060	·002270	·002497	·002741	·003003	·003285	·003586	·003908	·004252	·004618	*20*
21	·000991	·001103	·001225	·001357	·001502	·001658	·001827	·002010	·002207	·002419	*21*
22	·000455	·000511	·000573	·000642	·000717	·000799	·000889	·000987	·001093	·001210	*22*
23	·000200	·000227	·000257	·000290	·000327	·000368	·000413	·000463	·000518	·000578	*23*
24	·000084	·000096	·000110	·000126	·000143	·000163	·000184	·000208	·000235	·000265	*24*
25	·000034	·000039	·000045	·000052	·000060	·000069	·000079	·000090	·000103	·000117	*25*
26	·000013	·000015	·000018	·000021	·000024	·000028	·000032	·000037	·000043	·000049	*26*
27	·000005	·000006	·000007	·000008	·000009	·000011	·000013	·000015	·000017	·000020	*27*
28	·000002	·000002	·000003	·000003	·000004	·000004	·000005	·000006	·000007	·000008	*28*
29	·000001	·000001	·000001	·000001	·000001	·000002	·000002	·000002	·000003	·000003	*29*
30	—	—	—	—	—	·000001	·000001	·000001	·000001	·000001	*30*

x	*11·1*	*11·2*	*11·3*	*11·4*	*11·5*	*11·6*	*11·7*	*11·8*	*11·9*	*12·0*	*x*
0	·000015	·000014	·000012	·000011	·000010	·000009	·000008	·000008	·000007	·000006	*0*
1	·000168	·000153	·000140	·000128	·000116	·000106	·000097	·000089	·000081	·000074	*1*
2	·000931	·000858	·000790	·000727	·000670	·000617	·000568	·000522	·000481	·000442	*2*
3	·003445	·003202	·002976	·002764	·002568	·002385	·002214	·002055	·001907	·001770	*3*
4	·009559	·008965	·008406	·007879	·007382	·006915	·006476	·006062	·005674	·005309	*4*
5	·021221	·020082	·018997	·017963	·016979	·016043	·015153	·014307	·013504	·012741	*5*
6	·039259	·037487	·035778	·034130	·032544	·031017	·029549	·028137	·026782	·025481	*6*
7	·062253	·059979	·057755	·055584	·053465	·051400	·049388	·047432	·045530	·043682	*7*
8	·086376	·083970	·081579	·079206	·076856	·074529	·072231	·069962	·067725	·065523	*8*
9	·106531	·104496	·102427	·100328	·098204	·096060	·093900	·091728	·089548	·087364	*9*
10	·118249	·117036	·115743	·114374	·112935	·111430	·109863	·108239	·106562	·104837	*10*
11	·119324	·119164	·118899	·118533	·118068	·117508	·116854	·116110	·115281	·114368	*11*
12	·110375	·111220	·111964	·112607	·113149	·113591	·113933	·114175	·114320	·114363	*12*
13	·094243	·095820	·097322	·098747	·100093	·101358	·102539	·103636	·104647	·105570	*13*
14	·074721	·076656	·078553	·080409	·082219	·083982	·085694	·087350	·088950	·090489	*14*
15	·055294	·057236	·059177	·061110	·063035	·064946	·066841	·068716	·070567	·072391	*15*
16	·038360	·040065	·041793	·043541	·045306	·047086	·048877	·050678	·052484	·054293	*16*
17	·025047	·026396	·027780	·029198	·030648	·032129	·033639	·035176	·036739	·038325	*17*
18	·015446	·016424	·017440	·018492	·019581	·020706	·021865	·023060	·024288	·025550	*18*
19	·009023	·009682	·010372	·011095	·011852	·012641	·013465	·014322	·015212	·016137	*19*

TABLE III—(*continued*).

x	*m*										*x*
	11·1	*11·2*	*11·3*	*11·4*	*11·5*	*11·6*	*11·7*	*11·8*	*11·9*	*12·0*	
20	·005008	·005422	·005860	·006324	·006815	·007332	·007877	·008450	·009051	·009682	*20*
21	·002647	·002892	·003153	·003433	·003732	·004050	·004388	·004748	·005129	·005533	*21*
22	·001336	·001472	·001620	·001779	·001951	·002136	·002334	·002547	·002774	·003018	*22*
23	·000645	·000717	·000796	·000882	·000975	·001077	·001187	·001307	·001435	·001575	*23*
24	·000298	·000335	·000375	·000419	·000467	·000521	·000579	·000642	·000712	·000787	*24*
25	·000132	·000150	·000169	·000191	·000215	·000242	·000271	·000303	·000339	·000378	*25*
26	·000057	·000065	·000074	·000084	·000095	·000108	·000122	·000138	·000155	·000174	*26*
27	·000023	·000027	·000031	·000035	·000041	·000046	·000053	·000060	·000068	·000078	*27*
28	·000009	·000011	·000012	·000014	·000017	·000019	·000022	·000025	·000029	·000033	*28*
29	·000004	·000004	·000005	·000006	·000007	·000008	·000009	·000010	·000012	·000014	*29*
30	·000001	·000002	·000002	·000002	·000003	·000003	·000003	·000004	·000005	·000005	*30*
31	—	·000001	·000001	·000001	·000001	·000001	·000001	·000002	·000002	·000002	*31*
32	—	—	—	—	—	—	—	·000001	·000001	·000001	*32*
x	*12·1*	*12·2*	*12·3*	*12·4*	*12·5*	*12·6*	*12·7*	*12·8*	*12·9*	*13·0*	*x*
0	·000006	·000005	·000005	·000004	·000004	·000003	·000003	·000003	·000002	·000002	*0*
1	·000067	·000061	·000056	·000051	·000047	·000042	·000039	·000035	·000032	·000029	*1*
2	·000407	·000374	·000344	·000317	·000291	·000268	·000246	·000226	·000208	·000191	*2*
3	·001641	·001522	·001412	·001309	·001213	·001124	·001042	·000965	·000894	·000828	*3*
4	·004966	·004643	·004341	·004057	·003791	·003541	·003307	·003088	·002882	·002690	*4*
5	·012017	·011330	·010679	·010062	·009477	·008924	·008400	·007905	·007436	·006994	*5*
6	·024233	·023037	·021892	·020794	·019744	·018740	·017781	·016864	·015988	·015153	*6*
7	·041889	·040151	·038467	·036836	·035258	·033733	·032259	·030837	·029464	·028141	*7*
8	·063358	·061230	·059142	·057095	·055091	·053129	·051212	·049339	·047511	·045730	*8*
9	·085181	·083000	·080828	·078665	·076515	·074381	·072266	·070171	·068100	·066054	*9*
10	103069	·101261	·099418	·097544	·095644	·093720	·091777	·089819	·087849	·085870	*10*
11	·113376	·112308	·111168	·109959	·108686	·107352	·105961	·104516	·103023	·101483	*11*
12	·114321	·114180	·113947	·113624	·113215	·112720	·112142	·111484	·110749	·109940	*12*
13	·106406	·107153	·107811	·108380	·108860	·109251	·109554	·109769	·109897	·109940	*13*
14	·091965	·093376	·094720	·095994	·097197	·098326	·099381	·100360	·101263	·102087	*14*
15	·074185	·075946	·077670	·079355	·080997	·082594	·084143	·085641	·087086	·088475	*15*
16	·056103	·057909	·059709	·061500	·063279	·065043	·066788	·068513	·070213	·071886	*16*
17	·039932	·041558	·043201	·044859	·046529	·048208	·049895	·051586	·053279	·054972	*17*
18	·026843	·028167	029521	·030903	·032312	·033746	·035204	·036683	·038183	·039702	*18*
19	·017095	·018086	·019111	·020168	·021258	·022379	·023531	·024713	·025925	·027164	*19*
20	·010342	·011033	·011753	·012504	·013286	·014099	·014942	·015816	·016721	·017657	*20*
21	·005959	·006409	·006884	·007383	·007908	·008459	·009036	·009640	·010272	·010930	*21*
22	·003278	·003554	·003849	·004162	·004493	·004845	·005216	·005609	·006023	·006459	*22*
23	·001724	·001885	·002058	·002244	·002442	·002654	·002880	·003122	·003378	·003651	*23*
24	·000869	·000958	·001055	·001159	·001272	·001393	·001524	·001665	·001816	·001977	*24*
25	·000421	·000468	·000519	·000575	·000636	·000702	·000774	·000852	·000937	·001028	*25*
26	·000196	·000219	·000246	·000274	·000306	·000340	·000378	·000420	·000465	·000514	*26*
27	·000088	·000099	·000112	·000126	·000142	·000159	·000178	·000199	·000222	·000248	*27*
28	·000038	·000043	·000049	·000056	·000063	·000071	·000081	·000091	·000102	·000115	*28*
29	·000016	·000018	·000021	·000024	·000027	·000031	·000035	·000040	·000046	·000052	*29*
30	·000006	·000007	·000009	·000010	·000011	·000013	·000015	·000017	·000020	·000022	*30*
31	·000002	·000003	·000003	·000004	·000005	·000005	·000006	·000007	·000008	·000009	*31*
32	·000001	·000001	·000001	·000002	·000002	·000002	·000002	·000003	·000003	·000004	*32*
33	—	—	—	·000001	·000001	·000001	·000001	·000001	·000001	·000002	*33*
34	—	—	—	—	—	—	—	—	—	·000001	*34*

TABLE III—(*continued*)

x	*m*										*x*
	13·1	*13·2*	*13·3*	*13·4*	*13·5*	*13·6*	*13·7*	*13·8*	*13·9*	*14·0*	
0	·000002	·000002	·000002	·000002	·000001	·000001	·000001	·000001	·000001	·000001	*0*
1	·000027	·000024	·000022	·000020	·000019	·000017	·000015	·000014	·000013	·000012	*1*
2	·000175	·000161	·000148	·000136	·000125	·000115	·000105	·000097	·000089	·000081	*2*
3	·000766	·000709	·000657	·000608	·000562	·000520	·000481	·000445	·000411	·000380	*3*
4	·002510	·002341	·002183	·002035	·001897	·001768	·001648	·001535	·001429	·001331	*4*
5	·006575	·006180	·005807	·005455	·005123	·004810	·004514	·004236	·003974	·003727	*5*
6	·014356	·013596	·012872	·012183	·011526	·010902	·010308	·009743	·009206	·008696	*6*
7	·026867	·025639	·024458	·023322	·022230	·021181	·020173	·019207	·018280	·017392	*7*
8	·043994	·042304	·040661	·039064	·037512	·036007	·034547	·033132	·031762	·030435	*8*
9	·064036	·062046	·060088	·058161	·056269	·054410	·052588	·050802	·049054	·047344	*9*
10	·083887	·081901	·079916	·077936	·075963	·073998	·072046	·070107	·068185	·066282	*10*
11	·099901	·098281	·096626	·094940	·093227	·091489	·089730	·087953	·086162	·084359	*11*
12	·109059	·108109	·107094	·106017	·104880	·103687	·102441	·101146	·099804	·098418	*12*
13	·109898	·109773	·109566	·109279	·108914	·108473	·107957	·107370	·106713	·105989	*13*
14	·102833	·103500	·104087	·104595	·105024	·105373	·105644	·105836	·105951	·105989	*14*
15	·089807	·091080	·092291	·093439	·094522	·095539	·096488	·097369	·098181	·098923	*15*
16	·073530	·075141	·076717	·078255	·079753	·081208	·082618	·083981	·085295	·086558	*16*
17	·056661	·058345	·060019	·061683	·063333	·064966	·066580	·068173	·069741	·071283	*17*
18	·041237	·042786	·044348	·045920	·047500	·049086	·050675	·052266	·053856	·055442	*18*
19	·028432	·029725	·031043	·032385	·033750	·035135	·036539	·037962	·039400	·040852	*19*
20	·018623	·019619	·020644	·021698	·022781	·023892	·025030	·026193	·027383	·028597	*20*
21	·011617	·012332	·013074	·013846	·014645	·015473	·016329	·017213	·018125	·019064	*21*
22	·006917	·007399	·007904	·008433	·008987	·009565	·010168	·010797	·011452	·012132	*22*
23	·003940	·004246	·004571	·004913	·005275	·005656	·006057	·006478	·006921	·007385	*23*
24	·002151	·002336	·002533	·002743	·002967	·003205	·003457	·003725	·004008	·004308	*24*
25	·001127	·001233	·001348	·001470	·001602	·001744	·001895	·002056	·002229	·002412	*25*
26	·000568	·000626	·000689	·000758	·000832	·000912	·000998	·001091	·001191	·001299	*26*
27	·000275	·000306	·000340	·000376	·000416	·000459	·000507	·000558	·000613	·000674	*27*
28	·000129	·000144	·000161	·000180	·000201	·000223	·000248	·000275	·000305	·000337	*28*
29	·000058	·000066	·000074	·000083	·000093	·000105	·000117	·000131	·000146	·000163	*29*
30	·000025	·000029	·000033	·000037	·000042	·000047	·000053	·000060	·000068	·000076	*30*
31	·000011	·000012	·000014	·000016	·000018	·000021	·000024	·000027	·000030	·000034	*31*
32	·000004	·000005	·000006	·000007	·000008	·000009	·000010	·000012	·000013	·000015	*32*
33	·000002	·000002	·000002	·000003	·000003	·000004	·000004	·000005	·000006	·000006	*33*
34	·000001	·000001	·000001	·000001	·000001	·000001	·000002	·000002	·000002	·000003	*34*
35	—	—	—	—	—	·000001	·000001	·000001	·000001	·000001	*35*

x	*14·1*	*14·2*	*14·3*	*14·4*	*14·5*	*14·6*	*14·7*	*14·8*	*14·9*	*15·0*	*x*
0	·000001	·000001	·000001	·000001	·000001	—	—	—	—	—	*0*
1	·000011	·000010	·000009	·000008	·000007	·000007	·000006	·000006	·000005	·000005	*1*
2	·000075	·000069	·000063	·000058	·000053	·000049	·000045	·000041	·000038	·000034	*2*
3	·000352	·000325	·000300	·000277	·000256	·000237	·000219	·000202	·000186	·000172	*3*
4	·001239	·001153	·001073	·000999	·000929	·000864	·000803	·000747	·000694	·000645	*4*
5	·003494	·003275	·003070	·002876	·002694	·002523	·002362	·002211	·002069	·001936	*5*
6	·008212	·007752	·007316	·006902	·006510	·006139	·005787	·005454	·005138	·004839	*6*
7	·016541	·015726	·014946	·014199	·013486	·012804	·012152	·011530	·010937	·010370	*7*
8	·029153	·027913	·026715	·025559	·024443	·023367	·022330	·021331	·020370	·019444	*8*
9	·045673	·044040	·042447	·040894	·039380	·037907	·036472	·035078	·033723	·032407	*9*
10	·064399	·062537	·060700	·058887	·057101	·055343	·053614	·051915	·050247	·048611	*10*
11	·082547	·080730	·078910	·077089	·075270	·073456	·071648	·069850	·068062	·066287	*11*

TABLE III—*(continued)*

x	*m*										*x*
	14·1	*14·2*	*14·3*	*14·4*	*14·5*	*14·6*	*14·7*	*14·8*	*14·9*	*15·0*	
12	·096993	·095530	·094034	·092507	·090951	·089371	·087769	·086148	·084510	·082859	*12*
13	·105200	·104349	·103437	·102469	·101446	·100371	·099247	·098076	·096862	·095607	*13*
14	·105951	·105839	·105654	·105396	·105069	·104672	·104209	·103681	·103089	·102436	*14*
15	·099594	·100195	·100723	·101181	·101567	·101881	·102125	·102298	·102402	·102430	*15*
16	·087768	·088923	·090021	·091063	·092045	·092967	·093827	·094626	·095361	·096034	*16*
17	·072795	·074277	·075724	·077135	·078509	·079842	·081133	·082380	·083581	·084736	*17*
18	·057023	·058596	·060158	·061708	·063243	·064761	·066259	·067735	·069187	·070613	*18*
19	·042317	·043793	·045277	·046768	·048264	·049763	·051263	·052762	·054257	·055747	*19*
20	·029834	·031093	·032373	·033673	·034992	·036327	·037678	·039044	·040422	·041810	*20*
21	·020031	·021025	·022045	·023090	·024161	·025256	·026375	·027517	·028680	·029865	*21*
22	·012838	·013570	·014329	·015114	·015924	·016761	·017623	·018511	·019424	·020362	*22*
23	·007870	·008378	·008909	·009462	·010039	·010640	·011264	·011911	·012584	·013280	*23*
24	·004624	·004957	·005308	·005677	·006065	·006472	·006899	·007345	·007812	·008300	*24*
25	·002608	·002816	·003036	·003270	·003518	·003780	·004057	·004348	·004656	·004980	*25*
26	·001414	·001538	·001670	·001811	·001962	·002123	·002294	·002475	·002668	·002873	*26*
27	·000739	·000809	·000884	·000966	·001054	·001148	·001249	·001357	·001473	·001596	*27*
28	·000372	·000410	·000452	·000497	·000546	·000598	·000656	·000717	·000784	·000855	*28*
29	·000181	·000201	·000223	·000247	·000273	·000301	·000332	·000366	·000403	·000442	*29*
30	·000085	·000095	·000106	·000118	·000132	·000147	·000163	·000181	·000200	·000221	*30*
31	·000039	·000044	·000049	·000055	·000062	·000069	·000077	·000086	·000096	·000107	*31*
32	·000017	·000019	·000022	·000025	·000028	·000032	·000035	·000040	·000045	·000050	*32*
33	·000007	·000008	·000009	·000011	·000012	·000014	·000016	·000018	·000020	·000023	*33*
34	·000003	·000003	·000004	·000005	·000005	·000006	·000007	·000008	·000009	·000010	*34*
35	·000001	·000001	·000002	·000002	·000002	·000002	·000003	·000003	·000004	·000004	*35*
36	—	·000001	·000001	·000001	·000001	·000001	·000001	·000001	·000002	·000002	*36*
37	—	—	—	—	—	—	—	·000001	·000001	·000001	*37*

SECTION X

SUMMARY OF FORMULAS AND DEFINITIONS

The *a priori probability* that an event will occur is the ratio of the number of favorable cases to the number of total possible cases, all cases being equally likely to occur. (See par. 4.)

The *statistical probability* that an event occur is the limit of the ratio of the number of observed favorable cases to the total number of observed cases as the latter number increases indefinitely. (See par. 5.)

Statistical method is the mathematical treatment of observational data in accordance with the fundamental laws of probability. (See par. 7.)

A *statistical variate* is a variable which may assume a finite or infinite number of different values in accordance with a certain law of probability. (See par. 7.)

A *statistic* is any number computed from observed data in accordance with certain rules. (See par. 7.)

A *frequency distribution* is a collection of data arranged with respect to one or more characteristics. (See par. 8.)

The symbol $\sum_{x=r}^{n}$ means the sum for all integral values of x from r to n inclusive.

A (statistical) population is an idealized aggregate of data from which a sample is supposed to have been drawn by chance.

Random text is text in which the interplay of those factors which give rise to a particular cipher element is such that the cipher elements will occur with approximately the same frequency. (See par. 15.)

Non-random text is text in which the elements have been properly allocated in accordance with their cryptographic treatment. (See par. 16).

$$\bar{x}=\frac{w_1x_1+w_2x_2+\cdot\cdot\cdot+w_nx_n}{w_1+w_2+\cdot\cdot\cdot+w_n} \qquad \text{(See par. 7.)}$$

$$\text{mean square } x=\frac{w_1x_1^2+w_2x_2^2+\cdot\cdot\cdot+w_nx_n^2}{w_1+w_2+\cdot\cdot\cdot+w_n} \qquad \text{(See par. 7.)}$$

$$\text{variance}=v=\frac{w_1(x_1-\bar{x})^2+w_2(x_2-\bar{x})^2+\cdot\cdot\cdot+w_n(x_n-\bar{x})^2}{w_1+w_2+\cdot\cdot\cdot+w_n} \qquad \text{(See par. 7.)}$$

$$\text{Standard deviation}=\sigma=\sqrt{\text{variance}} \qquad \text{(See par. 7.)}$$

$$n!=n(n-1)(n-2)\cdot\cdot\cdot 2\times 1$$

BINOMIAL DISTRIBUTION (See par. 9.)

$$(q+p)^n=q^n+nq^{n-1}p+\frac{n(n-1)}{1\times 2}q^{n-2}p^2+\cdot\cdot\cdot+\frac{n!}{x!(n-x)!}q^{n-x}p^x+\cdot\cdot\cdot+p^n$$

$$\mu=np,\ \sigma^2=npq,\ \mu_2=n^2p^2+npq$$

$$\mu_{\bar{x}}=np,\ \sigma_{\bar{x}}^2=npq/N=\sigma^2/N$$

NORMAL DISTRIBUTION (See par. 10.)

$$p(X,\epsilon)=(\epsilon/\sigma\sqrt{2\pi})e^{-(X-\mu)^2/2\sigma^2}$$

$$\mu_{\bar{x}}=\mu \qquad \sigma_{\bar{x}}^2=\sigma^2/N$$

$$P(x_0,x_1)=\frac{1}{\sqrt{2\pi}}\int_{x_0}^{x_1}e^{-x^2/2}dx$$

POISSON DISTRIBUTION (See par. 11.)

$$e^{-m},\ me^{-m},\ m^2e^{-m}/2!,\ \cdots,\ m^xe^{-m}/x!,\ \cdots$$

$$m=\sigma^2$$

Expected number of blanks, random text

$$B_N=n(1-1/n)^N$$

$$B_N=ne^{-N/n} \qquad \text{(See par. 15.)}$$

Expected number of blanks, non-random text

$$B_N=(1-p_1)^N+(1-p_2)^N+\cdots+(1-p_n)^N$$

$$B_N=e^{-Np_1}+e^{-Np_2}+\cdots+e^{-Np_n} \qquad \text{(See par. 16.)}$$

Expected number of elements occurring r times each, random text

$$N(N-1)\cdots(N-r+1)n(1-1/n)^{N-r}/n^r r!$$

or

$$n(N/n)^r(1/r!)e^{-N/n} \qquad \text{(See par. 17.)}$$

Expected number of elements occurring r times each, non-random text.

$$\frac{N(N-1)\ \ldots\ (N-r+1)}{r!}\sum_{i=1}^{n}p_i^r(1-p_i)^{N-r}$$

or

$$\sum_{i=1}^{n}(1/r!)(Np_i)^re^{-Np_i} \qquad \text{(See par. 17.)}$$

$$\phi=f_1(f_1-1)+f_2(f_2-1)+\cdots+f_n(f_n-1) \qquad \text{(See par. 18.)}$$

$$E(\phi)=s_2N(N-1) \qquad \text{(See par. 18.)}$$

$$\psi=f_1^2+f_2^2+\cdots+f_n^2 \qquad \text{(See par. 18.)}$$

$$\psi=\phi+N$$

$$E(\psi)=s_2N^2+(1-s_2)N \qquad \text{(See par. 18.)}$$

$$\sigma_\phi^2=\sigma_\psi^2=4N^3(s_3-s_2^2)+2N^2(5s_2^2+s_2-6s_3)+2N(4s_3-s_2-3s_2^2) \qquad \text{(See par. 18.)}$$

Non-matching distributions

$$E(\phi)=s_2N(N-1)-2N_1N_2(s_2-1/n) \qquad \text{(See par. 20.)}$$

$$\sigma^2_\phi=(N_1^3+N_2^3)(4s_3-4s_2^2)+(N_1^2+N_2^2)(10s_2^2-12s_3+2s_2)$$
$$+(N_1+N_2)(8s_3-6s_2^2-2s_2)+4N_1N_2[(N_1+N_2)(s_2/n-1/n^2)$$
$$+1/n+1/n^2-2s_2/n] \qquad \text{(See par. 20.)}$$

$$\chi=f_1f_1'+f_2f_2'+\cdot\cdot\cdot+f_nf_n' \qquad \text{(See par. 21.)}$$

Properly matched distributions

$$E(\chi)=s_2N_1N_2 \qquad \text{(See par. 21.)}$$

$$\sigma_\chi^2=N_1N_2[(N_1+N_2)(s_3-s_2^2)+s_2^2+s_2-2s_3] \qquad \text{(See par. 21.)}$$

Non-matching distributions

$$E(\chi)=N_1N_2/n \qquad \text{(See par. 21.)}$$

$$\sigma_\chi^2=N_1N_2[(N_1+N_2)(s_2/n-1/n^2)+1/n+1/n^2-2s_2/n] \qquad \text{(See par. 21.)}$$

Random distributions

$$E(\chi)=N_1N_2/n \qquad \text{(See par. 21.)}$$

$$\sigma_\chi^2=N_1N_2(1/n-1/n^2) \qquad \text{(See par. 21.)}$$

Probability for monographic and digraphic coincidence, plain text

	κ_p	κ_p^2
English	0.0661	0.0069
French	.0778	.0093
German	.0762	.0112
Italian	.0738	.0081
Japanese (Romaji)	.0819	.0116
Portuguese	.0791	
Russian	.0529	.0058
Spanish	.0775	.0093

SECTION XI

APPENDIXES

In these appendixes we shall include a more detailed mathematical discussion of some of the theories and procedures given in part 1.

APPENDIX A

BINOMIAL DISTRIBUTION

Empirical assumption.—If an event which can happen in two different ways be repeated a great number of times under the same essential conditions, the ratio of the number of times that it happens in one way, to the total number of trials, will approach a definite limit, as the latter number increases indefinitely.[1]

Definition.—The limit described in the empirical assumption shall be called the probability that the event shall happen in the first way under those conditions.[1]

Theorem of compound probability.—If a compound event consists in the conjunction of any number of independent events, the probability of the compound event is the product of the probabilities for the individual events.[2]

Thus suppose that in N independent sets of n independent observations each an event occurs $x_1, x_2, \cdots, x_N$ times respectively. Then if p is the probability that the event occur, in accordance with the definition

$$\lim_{N\to\infty} \frac{x_1+x_2+\cdots+x_N}{n\cdot N}=p \tag{1}$$

We shall use the notation $E(x/n)$ to represent the left member of equation (1).

If an event occurs x_1 times in n observations, then there are $x_1(x_1-1)/2$ pairs of occurrences in $n(n-1)/2$ pairs of observations; $x_1(x_1-1)(x_1-2)/3!$ triplets of occurrences in $n(n-1)(n-2)/3!$ triplets of observations, etc. Using the theorem of compound probability, we write this as

$$\begin{aligned} &E(x/n)=p \\ &E(x(x-1)/n(n-1))=p^2 \\ &E(x(x-1)(x-2)/n(n-1)(n-2))=p^3 \\ &\quad \text{etc.} \end{aligned} \tag{2}$$

or since n is a constant

$$\begin{aligned} &E(x)=np \\ &E(x(x-1))=n(n-1)p^2 \\ &E(x(x-1)(x-2))=n(n-1)(n-2)p^3 \\ &\quad \text{etc.} \end{aligned} \tag{3}$$

[1] J. L. Coolidge, An Introduction to Mathematical Probability, 1925, p. 4.

[2] J. L. Coolidge, Op. cit., p. 18.

Since $E(x(x-1))=n(n-1)p^2$ we have

(4) $$E(x^2-x)=E(x^2)-E(x)=n(n-1)p^2$$

(5) $$E(x^2)=n(n-1)p^2+np=n^2p^2+npq \text{ where } q=1-p$$

Since $\sigma^2=E(x^2)-[E(x)]^2$ we have

(6) $$\sigma^2=n^2p^2+npq-n^2p^2=npq$$

If $\bar{x}=(x_1+x_2+\cdots+x_N)/N$, where $x_1, x_2, \cdots, x_N$ are the number of occurrences of an event in each of N independent sets of n independent observations each, then

(7) $$\begin{aligned} E(\bar{x})&=E(x_1/N)+E(x_2/N)+\cdots+E(x_N/N)\\ &=np/N+np/N+\cdots+np/N=Nnp/N=np \end{aligned}$$

Since $(\bar{x})^2=\frac{1}{N^2}\sum_{i=1}^{N}x_i^2+\frac{2}{N^2}\sum_{i,j=1}^{N}x_ix_j$ $(i\neq j)$

(8) $$E((\bar{x})^2)=\frac{1}{N^2}\sum_{i=1}^{N}E(x_i^2)+\frac{2}{N^2}\sum_{i,j=1}^{N}E(x_ix_j) \qquad (i\neq j)$$

Since the observations are independent $E(x_ix_j)=E(x_i)E(x_j)$. Using (3) and (5) there is obtained

(9) $$\begin{aligned} E((\bar{x})^2)&=\frac{N}{N^2}(n^2p^2+npq)+\frac{N(N-1)}{2}\cdot\frac{2}{N^2}n^2p^2\\ &=n^2p^2/N+npq/N+n^2p^2-n^2p^2/N=npq/N+n^2p^2 \end{aligned}$$

(10) $$\sigma_{\bar{x}}^2=E((\bar{x})^2)-[E(\bar{x})]^2=npq/N+n^2p^2-n^2p^2=npq/N, \text{ or}$$

(11) $$\sigma_{\bar{x}}^2=\sigma^2/N$$

If we set $M_r=E[x(x-1)(x-2)\ldots(x-r+1)]$, then it may be shown that for discontinuous distributions, the probability that there are exactly r occurrences is given by

(12) $$P(r)=\frac{1}{r!}\left[M_r-M_{r+1}+\frac{M_{r+2}}{2!}-\frac{M_{r+3}}{3!}+\cdots\right]$$

From (3) it is seen that $M_k=n(n-1)\ldots(n-k+1)p^k=(n!/(n-k)!)p^k$

(13) $$\begin{aligned} P(r)&=\frac{n!}{r!}\left[\frac{p^r}{(n-r)!}-\frac{p^{r+1}}{(n-r-1)!}+\frac{p^{r+2}}{2!(n-r-2)!}+\cdots\right]\\ &=\frac{n!}{r!}\frac{p^r}{(n-r)!}\left[1-(n-r)p+\frac{(n-r)(n-r-1)}{2!}p^2-\cdots\right]\\ &=\frac{n!}{r!(n-r)!}p^r(1-p)^{n-r}=\frac{n!}{r!(n-r)!}p^rq^{n-r}, \text{ where } q=1-p. \end{aligned}$$

Appendix B

POISSON EXPONENTIAL DISTRIBUTION

We shall here derive the Poisson exponential distribution by treating the binomial distribution as $n\to\infty$ with $\lim_{n\to\infty} np=m$ where m is finite.

From (3) we thus obtain

$$
\begin{aligned}
&E(x)=m\\
(14)\qquad &E(x(x-1))=n^2p^2(1-1/n)=m^2\\
&E(x(x-1)(x-2))=n^3p^3(1-1/n)(1-2/n)=m^3\\
&\text{etc.}
\end{aligned}
$$

Thus

$$(15)\qquad E(x^2-x)=E(x^2)-E(x)=m^2$$

$$(16)\qquad E(x^2)=m^2+m$$

$$(17)\qquad \sigma^2=E(x^2)-[E(x)]^2=m^2+m-m^2=m$$

From (14) we have that $M_k=m^k$, thus

$$
\begin{aligned}
(18)\qquad P(r)&=\frac{1}{r!}\left(m^r-m^{r+1}+\frac{m^{r+2}}{2!}-\frac{m^{r+3}}{3!}+\cdots\right)\\
&=\frac{m^r}{r!}\left(1-m+\frac{m^2}{2!}-\frac{m^3}{3!}+\cdots\right)\\
&=\frac{m^r}{r!}e^{-m}
\end{aligned}
$$

Appendix C

MULTINOMIAL DISTRIBUTION

If a possible event is one of n mutually exclusive events, then a simple extension of the treatment in appendix A will apply to this case.

If in N observations the event has occurred x_1 times the first way, x_2 times the second way, $\cdots$, x_n times the nth way such that $x_1+x_2+\cdots+x_n=N$

$$
\begin{aligned}
&E(x_1/N)=p_1,\ E(x_2/N)=p_2,\ \cdots,\ E(x_n/N)=p_n\\
(19)\qquad &E(x_ix_j/N(N-1))=p_ip_j && (i\neq j,\ i,j=1,2,\cdots,n)\\
&E(x_ix_jx_k/N(N-1)(N-2))=p_ip_jp_k && (i\neq j\neq k,\ i,j,k=1,2,\cdots,n)\\
&E(x_i(x_i-1)x_j/N(N-1)(N-2))=p_i^2p_j && (i\neq j,\ i,j=1,2,\cdots,n)\\
&\text{etc.}
\end{aligned}
$$

The values in (19) follow from the following considerations: If the event has occurred x_i times the ith way and x_j times the jth way then the number of pairs of occurrences of both ith and jth ways is x_ix_j. However the total number of possible pairs is $N(N-1)$.

Since the occurrences of $x_1, x_2, \ldots, x_n$ are not mutually independent

$$(20)\qquad E(x_ix_j)\neq E(x_i)E(x_j)$$

Indeed from (19) we find

$$
\begin{aligned}
(21)\qquad E(x_ix_j)&=N(N-1)p_ip_j=Np_iNp_j-Np_ip_j\\
&=E(x_i)E(x_j)-Np_ip_j
\end{aligned}
$$

APPENDIX D

THE DERIVATION OF THE STANDARD DEVIATION OF ψ AND ϕ

The standard deviation of a statistical variate Y is defined by

(1) $$\sigma^2 = E(Y^2) - [E(Y)]^2$$

Thus, the standard deviation of

(2) $$\psi = f_1^2 + f_2^2 + \ldots + f_n^2$$

is given by

(3) $$\sigma_\psi^2 = E(\psi^2) - [E(\psi)]^2$$

In (2) ψ is the sum of the squares of the occurrences of the n possible elements of a cryptogram of N elements; in other words,

(4) $$f_1 + f_2 + \ldots + f_n = N$$

From (2), we have that

(5) $$E(\psi) = E(f_1^2) + \ldots + E(f_n^2)$$

Furthermore, also from (2)

(6) $$\psi^2 = f_1^4 + f_2^4 + \ldots + f_n^4 + 2f_1^2f_2^2 + 2f_1^2f_3^2 + \ldots + 2f_{n-1}^2 f_n^2$$

So that

(7) $$E(\psi^2) = E(f_1^4) + E(f_2^4) + \ldots + E(f_n^4) + 2E(f_1^2f_2^2) + \ldots + 2E(f_{n-1}^2 f_n^2)$$

From (5) and (7), it is clear that we must find $E(f_i^2)$, $E(f_i^4)$, $E(f_i^2f_j^2)$, which we now proceed to do.

If the probabilities of occurrence of the n possible elements are respectively $p_1, p_2, \ldots, p_n$, then

(8) $$\begin{aligned} &E(f_i) = Np_i \\ &E(f_i(f_i-1)) = N(N-1)p_i^2 \\ &E(f_i(f_i-1)(f_i-2)) = N(N-1)(N-2)p_i^3 \\ &E(f_i(f_i-1)(f_i-2)(f_i-3)) = N(N-1)(N-2)(N-3)p_i^4 \end{aligned}$$

(See appendixes A and C.)

Since $f_i^2 = f_i(f_i-1) + f_i$ we have

(9) $$E(f_i^2) = E(f_i(f_i-1)) + E(f_i) = N(N-1)p_i^2 + Np_i$$

Since $$f_i^4 = f_i(f_i-1)(f_i-2)(f_i-3) + 6f_i(f_i-1)(f_i-2) + 7f_i(f_i-1) + f_i$$

we have

(10) $$E(f_i^4) = N(N-1)(N-2)(N-3)p_i^4 + 6N(N-1)(N-2)p_i^3 + 7N(N-1)p_i^2 + Np_i$$

Since $$\begin{aligned} f_i^2f_j^2 &= [f_i(f_i-1)+f_i][f_j(f_j-1)+f_j] \\ &= f_i(f_i-1)f_j(f_j-1) + f_i(f_i-1)f_j + f_if_j(f_j-1) + f_if_j \end{aligned}$$

we have that

(11) $$E(f_i^2f_j^2) = N(N-1)(N-2)(N-3)p_i^2p_j^2 + N(N-1)(N-2)p_i^2p_j + N(N-1)(N-2)p_ip_j^2 + N(N-1)p_ip_j$$

From (5) and (9) we have

(12) $$E(\psi)=N(N-1)p_1^2+Np_1+\ .\ .\ .\ +N(N-1)p_n^2+Np_n$$

But $$p_1^2+p_2^2+\ .\ .\ .\ +p_n^2=s_2 \text{ and } p_1+p_2+\ .\ .\ .\ +p_n=1$$
so that

(13) $$E(\psi)=N(N-1)s_2+N=N^2s_2+(1-s_2)N$$

From (7), (10), and (11), we have

(14) $$\begin{aligned}E(\psi^2)=&N(N-1)(N-2)(N-3)\Sigma p_i^4+6N(N-1)(N-2)\Sigma p_i^3\\&+7N(N-1)\Sigma p_i^2+N\Sigma p_i+2N(N-1)(N-2)(N-3)\Sigma p_i^2p_j^2\\&+2N(N-1)\Sigma p_ip_j+2N(N-1)(N-2)\Sigma p_i^2p_j\end{aligned}$$

For convenience, let us write

(15) $$\begin{aligned}s_2&=p_1^2+p_2^2+\ .\ .\ .\ +p_n^2\\s_3&=p_1^3+p_2^3+\ .\ .\ .\ +p_n^3\\s_4&=p_1^4+p_2^4+\ .\ .\ .\ +p_n^4\end{aligned}$$

Now $$(p_1^2+p_2^2+\ .\ .\ .\ +p_n^2)(p_1^2+p_2^2+\ .\ .\ .\ +p_n^2)=\Sigma p_i^4+2\Sigma p_i^2p_j^2$$ so that

(16) $$2\Sigma p_i^2p_j^2=s_2^2-s_4;$$

also $$(p_1+p_2+\ .\ .\ .\ +p_n)(p_1+p_2+\ .\ .\ .\ +p_n)=\Sigma p_i^2+2\Sigma p_ip_j$$ so that

(17) $$2\Sigma p_ip_j=1-s_2; \text{ also}$$

$$(p_1+p_2+\ .\ .\ .\ +p_n)(p_1^2+p_2^2+\ .\ .\ .\ +p_n^2)=\Sigma p_i^3+\Sigma p_i^2p_j$$ so that

(18) $$\Sigma p_i^2p_j=s_2-s_3$$

In accordance with the above, we can therefore write

(19) $$E(\psi)=N(N-1)s_2+N$$

(20) $$\begin{aligned}E(\psi^2)=&N(N-1)(N-2)(N-3)s_4+6N(N-1)(N-2)s_3+7N(N-1)s_2+N\\&+N(N-1)(N-2)(N-3)(s_2^2-s_4)+2N(N-1)(N-2)(s_2-s_3)+N(N-1)(1-s_2)\end{aligned}$$

Therefore

(21) $$\begin{aligned}E(\psi^2)-[E(\psi)]^2=&N(N-1)(N-2)(N-3)s_4+6N(N-1)(N-2)s_3+7N(N-1)s_2\\&+N+N(N-1)(N-2)(N-3)s_2^2-N(N-1)(N-2)(N-3)s_4\\&+2N(N-1)(N-2)s_2-2N(N-1)(N-2)s_3+N(N-1)\\&-N(N-1)s_2-N(N-1)N(N-1)s_2^2-2N^2(N-1)s_2-N^2\end{aligned}$$

(22) $$\begin{aligned}\sigma_\psi^2=&s_3(6N^3-18N^2+12N-2N^3+6N^2-4N)\\&+s_2(7N^2-7N+2N^3-6N^2+4N-N^2+N-2N^3+2N^2)\\&+s_2^2(N^4-6N^3+11N^2-6N-N^4+2N^3-N^2)\\&+N+N(N-1)-N^2\end{aligned}$$

(23) $$\begin{aligned}\sigma_\psi^2=&s_3(4N^3-12N^2+8N)+s_2(2N^2-2N)\\&+s_2^2(-4N^3+10N^2-6N)\end{aligned}$$

(24) $$\begin{aligned}\sigma_\psi^2=&N^3(4s_3-4s_2^2)+N^2(-12s_3+2s_2+10s_2^2)\\&+N(8s_3-2s_2-6s_2^2)\end{aligned}$$

As may be easily computed from (15), the values of s_2, s_3, and s_4 for English monoalphabetic text are

(25) $$s_2=0.066112,\ s_3=0.005457,\ s_4=0.000511$$

So that, finally

(26) $$\sigma_\psi^2=N^3(0.004344)+N^2(0.110448)-N(0.114794)$$

From the result above, one can readily derive the standard deviation of ϕ

(27) $$\phi=f_1(f_1-1)+f_2(f_2-1)+\ .\ .\ .\ +f_n(f_n-1)$$

(28) $$\phi=f_1^2-f_1+f_2^2-f_2+\ .\ .\ .\ +f_n^2-f_n$$

(29) $$\phi=\psi-N$$

(30) $$E(\phi)=E(\psi)-E(N)=N(N-1)s_2+N-N=N(N-1)s_2$$

From (29)

(31) $$\phi^2=\psi^2-2N\psi+N^2$$

Therefore

(32) $$E(\phi^2)=E(\psi^2)-2NE(\psi)+N^2$$

so that

(33) $$\sigma_\phi^2=E(\phi^2)-[E(\phi)]^2=E(\psi^2)-2NE(\psi)+N^2-[E(\psi)]^2+2NE(\psi)-N^2$$

(34) $$\sigma_\phi^2=E(\psi^2)-[E(\psi)]^2=\sigma_\psi^2$$

Thus (26) will also give the standard deviation for ϕ.

The corresponding results for random text may be easily derived from the preceding results. For random text $p_i=1/n$, so that

(35) $$\begin{aligned} s_2&=\Sigma p_i^2=n\ (1/n^2)=1/n \\ s_3&=\Sigma p_i^3=n\ (1/n^3)=1/n^2 \\ s_4&=\Sigma p_i^4=n\ (1/n^4)=1/n^3 \end{aligned}$$

Substituting these values in (24), there is obtained

(36) $$\sigma_\psi^2=N^3(4/n^2-4/n^2)+N^2(-12/n^2+2/n+10/n^2)+N(8/n^2-2/n-6/n^2)$$

(37) $$\sigma_\psi^2=2N^2(1/n-1/n^2)-2N(1/n-1/n^2)$$

(38) $$\sigma_\psi^2=2N(N-1)(n-1)/n^2$$

For $n=26$ (38) becomes

(39) $$\sigma_\psi^2=0.073964N(N-1)$$

From (13) there is obtained for $n=26$

(40) $$E(\psi)=\frac{N(N-1)}{26}+N=0.038N^2+0.962N$$

and from (30),

(41) $$E(\phi)=\frac{N(N-1)}{26}=0.038N(N-1)$$

APPENDIX E

THE STANDARD DEVIATION FOR THE PRODUCT-SUM MATCHING TEST

Consider two non-random distributions of N_1 and N_2 elements respectively, where the occurrences of corresponding elements are given by

$$f_1, f_2, \cdots, f_n \text{ and } f_1', f_2', \cdots, f_n'$$

so that

$$f_1+f_2+\cdots+f_n=N_1; f_1'+f_2'+\cdots+f_n'=N_2$$

The frequencies of the two distributions are of course independent of one another.

Let us now consider the statistic defined by

$$\chi=f_1f_1'+f_2f_2'+\cdots+f_nf_n' \tag{1}$$

From (1) there is obtained

$$E(\chi)=E(f_1f_1')+E(f_2f_2')+\cdots+E(f_nf_n') \tag{2}$$

Since f_i and f_i' are independent, $(i=1, 2, \cdots, n)$

$$E(\chi)=E(f_1)E(f_1')+E(f_2)E(f_2')+\cdots+E(f_n)E(f_n') \tag{3}$$

$$E(\chi)=N_1p_1N_2p_1+N_1p_2N_2p_2+\cdots+N_1p_nN_2p_n=s_2N_1N_2 \tag{4}$$

In (4) we have assumed, of course, that the two distributions represent encipherments by means of the *same* substitution.

Since

$$\sigma_\chi^2=E(\chi^2)-[E(\chi)]^2 \tag{5}$$

we proceed to obtain χ^2 from (1).

Thus

$$\chi^2=f_1^2f_1'^2+\cdots+f_n^2f_n'^2+2\Sigma f_if_i'f_jf_j' \tag{6}$$

Since the f's and f''s are independent, we have

$$E(\chi^2)=E(f_1^2)E(f_1'^2)+\cdots+E(f_n^2)E(f_n'^2)+2\Sigma E(f_if_j)E(f_i'f_j') \tag{7}$$

But, as in the previous appendices

$$E(f_i^2)=N_1(N_1-1)p_i^2+N_1p_i \tag{8}$$

$$E(f_if_j)=N_1(N_1-1)p_ip_j, \text{ so that} \tag{9}$$

$$E(\chi^2)=\Sigma(N_1(N_1-1)p_i^2+N_1p_i)(N_2(N_2-1)p_i^2+N_2p_i)+2\Sigma N_1(N_1-1)N_2(N_2-1)p_i^2p_j^2 \tag{10}$$

$$\begin{aligned}E(\chi^2)=&N_1(N_1-1)N_2(N_2-1)\Sigma p_i^4+N_1(N_1-1)N_2\Sigma p_i^3+N_1N_2(N_2-1)\Sigma p_i^3\\&+N_1N_2\Sigma p_i^2+2N_1(N_1-1)N_2(N_2-1)\Sigma p_i^2p_j^2\end{aligned} \tag{11}$$

If we again write $s_2=\Sigma p_i^2$, $s_3=\Sigma p_i^3$, $s_4=\Sigma p_i^4$ then since $(p_1^2+p_2^2+\cdots+p_n^2)(p_1^2+p_2^2+\cdots+p_n^2)=\Sigma p_i^4+2\Sigma p_i^2p_j^2$, $2\Sigma p_i^2p_j^2=s_2^2-s_4$. Thus we have from (11):

(12) $$E(\chi^2)=N_1(N_1-1)N_2(N_2-1)s_4+N_1(N_1-1)N_2s_3+N_1N_2(N_2-1)s_3 +N_1N_2s_2+N_1(N_1-1)N_2(N_2-1)s_2^2-N_1(N_1-1)N_2(N_2-1)s_4$$

(13) $$E(\chi^2)=N_1(N_1-1)N_2s_3+N_1N_2(N_2-1)s_3+N_1N_2s_2+N_1(N_1-1)N_2(N_2-1)s_2^2$$

Therefore

(14) $$\sigma_\chi^2=E(\chi^2)-[E(\chi)]^2=N_1(N_1-1)N_2s_3+N_1N_2(N_2-1)s_3+N_1N_2s_2 +N_1(N_1-1)N_2(N_2-1)s_2^2-N_1^2N_2^2s_2^2$$

(15) $$\sigma_\chi^2=N_1N_2\{s_3(N_1+N_2-2)+s_2+s_2^2(1-N_1-N_2)\}$$

For English monoalphabets we get

(16) $$\sigma_\chi^2=N_1N_2\{(N_1+N_2-2)(0.005457)+0.066112-(N_1+N_2-1)(0.004371)\}$$

(17) $$\sigma_\chi^2=N_1N_2\{(N_1+N_2)(0.001086)+0.059569\}$$

For random text, $p_i=1/n$, so that $s_2=1/n$, $s_3=1/n^2$, $s_4=1/n^3$, $s_2^2=1/n^2$, and (4) becomes

(18) $$E(\chi)=N_1N_2/n$$

From (15) we get

(19) $$\sigma_\chi^2=N_1N_2\{N_1/n^2+N_2/n^2-2/n^2+1/n+1/n^2-N_1/n^2-N_2/n^2\}$$

(20) $$\sigma_\chi^2=N_1N_2(1/n-1/n^2)$$

For $n=26$, (18) and (20) become $E(\chi)=0.038N_1N_2$, $\sigma_\chi^2=0.036982N_1N_2$

Appendix F

STANDARD DEVIATION FOR PRODUCT-SUM MATCHING TEST—NON-MATCHING DISTRIBUTIONS

We proceed exactly as in the case for correct matching, appendix E, except that the corresponding probabilities will, now, not be the same.

(1) $$\chi=f_1f_1'+f_2f_2'+\ .\ .\ .\ +f_nf_n'$$

(2) $$E(\chi)=E(f_1)E(f_1')+E(f_2)E(f_2')+\ .\ .\ .\ +E(f_n)E(f_n')$$

(3) $$E(\chi)=N_1p_1N_2\pi_1+N_1p_2N_2\pi_2+\ .\ .\ .\ +N_1p_nN_2\pi_n$$

Where $p_1, p_2, \ .\ .\ .\ p_n$ and $\pi_1, \pi_2, \ .\ .\ .\ , \pi_n$ are two different arrangements of the probabilities of occurrence for the n possible elements.

Now

(4) $$\begin{aligned}(p_1+p_2+\ .\ .\ .\ p_n)(\pi_1+\pi_2+\ .\ .\ .\ +\pi_n)=&p_1\pi_1+p_2\pi_1+\ .\ .\ .\ +p_n\pi_1\\ &+p_1\pi_2+p_2\pi_2+\ .\ .\ .\ +p_n\pi_2\\ &+\ .\ .\ .\ .\ .\ .\ .\ .\ .\\ &+p_1\pi_n+p_2\pi_n+\ .\ .\ .\ +p_n\pi_n\end{aligned}$$

so that in general

(5) $$p_1\pi_1+p_2\pi_2+\ .\ .\ .\ +p_n\pi_n=(1/n)(p_1+p_2+\ .\ .\ .\ +p_n)(\pi_1+\pi_2+\ .\ .\ .\ +\pi_n)=1/n$$

Therefore

(6) $$E(\chi)=N_1N_2/n$$

From (1), we have

(7) $$\chi^2=f_1^2f_1'^2+\ .\ .\ .\ +f_n^2f_n'^2+2\Sigma f_if_i'f_jf_j'$$

(8) $$E(\chi^2)=\Sigma E(f_i^2)E(f_i'^2)+2\Sigma E(f_if_j)E(f_i'f_j')$$

As in the former case

$$E(\chi^2)=\Sigma(N_1(N_1-1)p_i^2+N_1p_i)(N_2(N_2-1)\pi_i^2+N_2\pi_i) \\ +2\Sigma N_1(N_1-1)p_ip_jN_2(N_2-1)\pi_i\pi_j \tag{9}$$

$$E(\chi^2)=N_1N_2(N_1-1)(N_2-1)\Sigma p_i^2\pi_i^2+N_1N_2(N_1-1)\Sigma p_i^2\pi_i \\ +N_1N_2(N_2-1)\Sigma p_i\pi_i^2+N_1N_2\Sigma p_i\pi_i \\ +2N_1N_2(N_1-1)(N_2-1)\Sigma p_ip_j\pi_i\pi_j \tag{10}$$

If we again write $s_2=p_1^2+p_2^2+\cdots+p_n^2=\pi_1^2+\pi_2^2+\cdots+\pi_n^2$

then $(p_1^2+p_2^2+\cdots+p_n^2)(\pi_1^2+\pi_2^2+\cdots+\pi_n^2)=p_1^2\pi_1^2+p_2^2\pi_1^2+\cdots+p_n^2\pi_1^2$

$$+\ \cdots\cdots\cdots\cdots$$

$$+p_1^2\pi_n^2+p_2^2\pi_n^2+\cdots+p_n^2\pi_n^2$$

so that in general $\Sigma p_i^2\pi_i^2=s_2^2/n$ and

$$\Sigma p_i^2\pi_i=s_2/n=\Sigma\pi_i^2p_i\text{, also}$$

$$(p_1\pi_1+p_2\pi_2+\cdots+p_n\pi_n)^2=\Sigma p_i^2\pi_i^2+2\Sigma p_ip_j\pi_i\pi_j \text{ so that}$$

$$2\Sigma p_ip_j\pi_i\pi_j=1/n^2-s_2^2/n$$

Substituting these values in (10)

$$E(\chi^2)=N_1N_2(N_1-1)(N_2-1)s_2^2/n+N_1N_2(N_1-1)s_2/n \\ +N_1N_2(N_2-1)s_2/n+N_1N_2/n \\ +N_1N_2(N_1-1)(N_2-1)(1/n^2-s_2^2/n) \tag{11}$$

$$\sigma_\chi^2=E(\chi^2)-[E(\chi)]^2 \tag{12}$$

$$\sigma_\chi^2=N_1N_2(N_1-1)s_2/n+N_1N_2(N_2-1)s_2/n+N_1N_2/n \\ +N_1N_2(N_1-1)(N_2-1)/n^2-N_1^2N_2^2/n^2 \tag{13}$$

$$\sigma_\chi^2=N_1N_2[(N_1+N_2-2)s_2/n+1/n-(N_1+N_2-1)/n^2] \tag{14}$$

$$\sigma_\chi^2=N_1N_2[(N_1+N_2)(s_2/n-1/n^2)+1/n-2s_2/n+1/n^2] \tag{15}$$

For $n=26$ (15) becomes for English text

$$\sigma_\chi^2=N_1N_2[(N_1+N_2)(0.001063)+0.034856] \tag{16}$$

Appendix G

STANDARD DEVIATION OF ϕ AND ψ. NON-MATCHING DISTRIBUTIONS

Consider two distributions of N_1 and N_2 elements, respectively, where the occurrences of corresponding elements are given by $f_1', f_2', \cdots, f_n'$ and $f_1'', f_2'', \cdots, f_n''$ so that $f_1'+f_2'+\cdots+f_n'=N_1$ and $f_1''+f_2''+\cdots+f_n''=N_2$. Suppose that these two distributions are combined by adding the frequencies of corresponding elements and let the frequencies of the resultant distribution be given by $f_1, f_2, \cdots, f_n$ so that $f_1=f_1'+f_1''$; $f_2=f_2'+f_2''$; $\cdots$; $f_n=f_n'+f_n''$ and $f_1+f_2+\cdots+f_n=N_1+N_2=N$.

If the two distributions match, then the discussion regarding

(1) $$\psi = f_1^2 + f_2^2 + \cdots + f_n^2 \text{ and}$$

(2) $$\phi = f_1(f_1-1) + f_2(f_2-1) + \cdots + f_n(f_n-1)$$

is identical with that already given in appendix D.

If the two distributions do not match, then a modification is necessary and we proceed as follows:

(3) $$\begin{aligned}\phi &= \Sigma f_i(f_i-1) = \Sigma(f_i'+f_i'')(f_i'+f_i''-1)\\ &= \Sigma f_i'(f_i'-1) + \Sigma f_i''(f_i''-1) + 2\Sigma f_i'f_i''\end{aligned}$$

From the discussion in Appendix D we have that

(4) $$E(\Sigma f_i'(f_i'-1)) = s_2N_1(N_1-1)\ ;\ E(\Sigma f_i''(f_i''-1)) = s_2N_2(N_2-1)$$

and from appendix F

(5) $$E(\Sigma f_i'f_i'') = N_1N_2/n.$$

There thus results

(6) $$E(\phi) = s_2[N_1(N_1-1) + N_2(N_2-1)] + 2N_1N_2/n$$

Since $N = N_1 + N_2$

(7) $$N(N-1) = (N_1+N_2)(N_1+N_2-1) = N_1(N_1-1) + N_2(N_2-1) + 2N_1N_2$$

and we may also write (6) as

(8) $$E(\phi) = s_2N(N-1) - 2N_1N_2(s_2 - 1/n)$$

If we let $\phi_1 = \Sigma f_i'(f_i'-1)$ and $\phi_2 = \Sigma f_i''(f_i''-1)$ then (3) may also be written as

(9) $$\phi = \phi_1 + \phi_2 + 2\chi$$

The discussion in paragraph 25 of the text has pointed out the relation of ϕ and χ to the concept of coincidences. In (9) ϕ_1 and ϕ_2 are related to the number of coincidences within each respective message and χ is related to the number of coincidences between the two messages. Since the messages are independent and the number of coincidences within one of the messages is independent of the number of coincidences between the messages ϕ in (9) has been expressed as the sum of three independent variables. Accordingly

(10) $$\sigma_\phi^2 = \sigma_{\phi_1}^2 + \sigma_{\phi_2}^2 + 4\sigma_\chi^2$$

From appendix D we find that

(11) $$\sigma_{\phi_1}^2 = N_1^3(4s_3 - 4s_2^2) + N_1^2(-12s_3 + 2s_2 + 10s_2^2) + N_1(8s_3 - 2s_2 - 6s_2^2)$$

(12) $$\sigma_{\phi_2}^2 = N_2^3(4s_3 - 4s_2^2) + N_2^2(-12s_3 + 2s_2 + 10s_2^2) + N_2(8s_3 - 2s_2 - 6s_2^2)$$

and from appendix F we have that

(13) $$\sigma_\chi^2 = N_1N_2[(N_1+N_2)(s_2/n - 1/n^2) + 1/n - 2s_2/n + 1/n^2]$$

There thus results

(14) $$\begin{aligned}\sigma_\phi^2 = {} & (N_1^3+N_2^3)(4s_3 - 4s_2^2) + (N_1^2+N_2^2)(10s_2^2 - 12s_3 + 2s_2)\\ & + (N_1+N_2)(8s_3 - 6s_2^2 - 2s_2)\\ & + 4N_1N_2[(N_1+N_2)(s_2/n - 1/n^2) + 1/n + 1/n^2 - 2s_2/n]\end{aligned}$$

Section XII

CHARTS

CHART No. 1

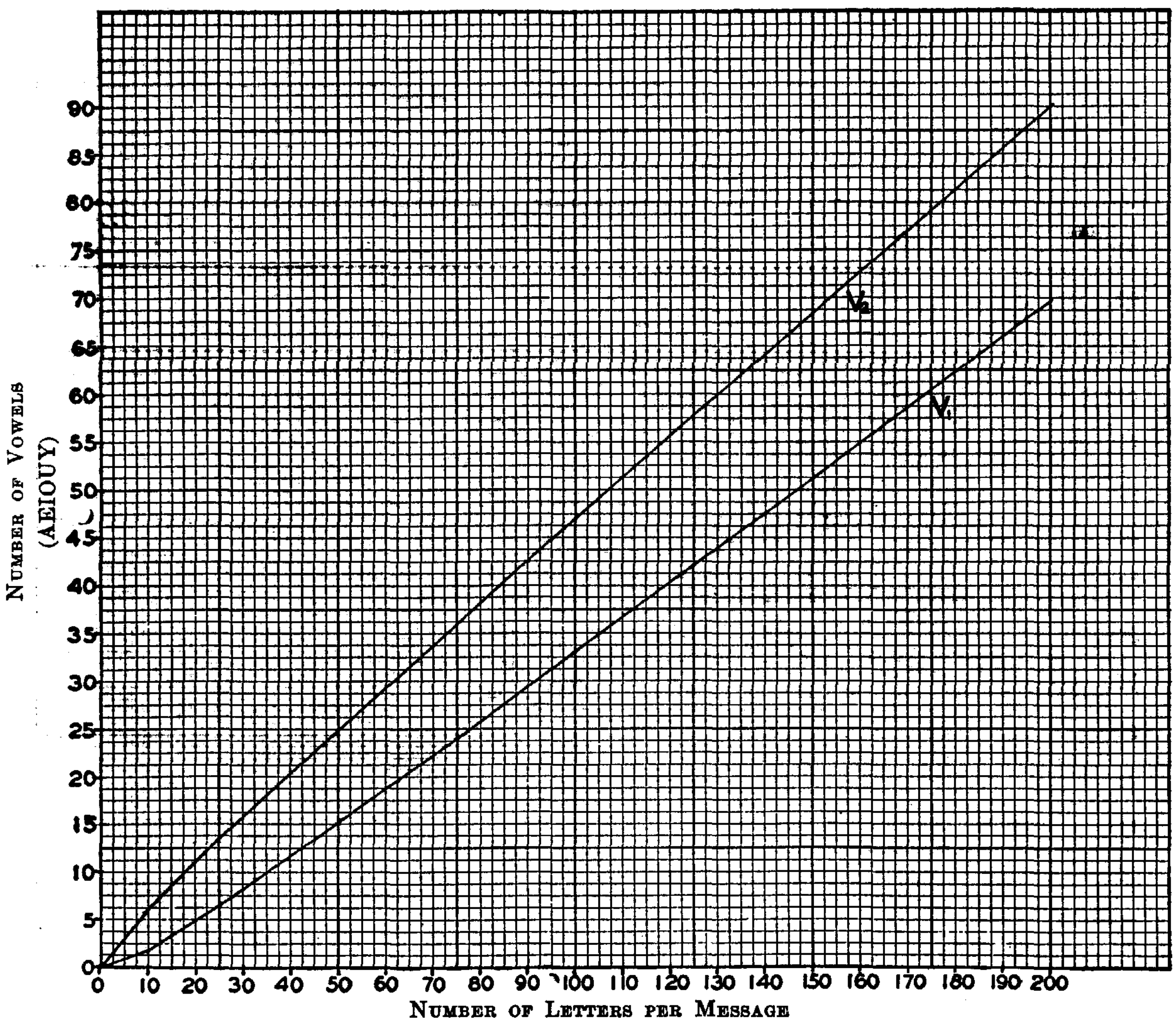

NUMBER OF VOWELS
(AEIOUY)
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
V2
V1
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200
NUMBER OF LETTERS PER MESSAGE

CHART NO. 2

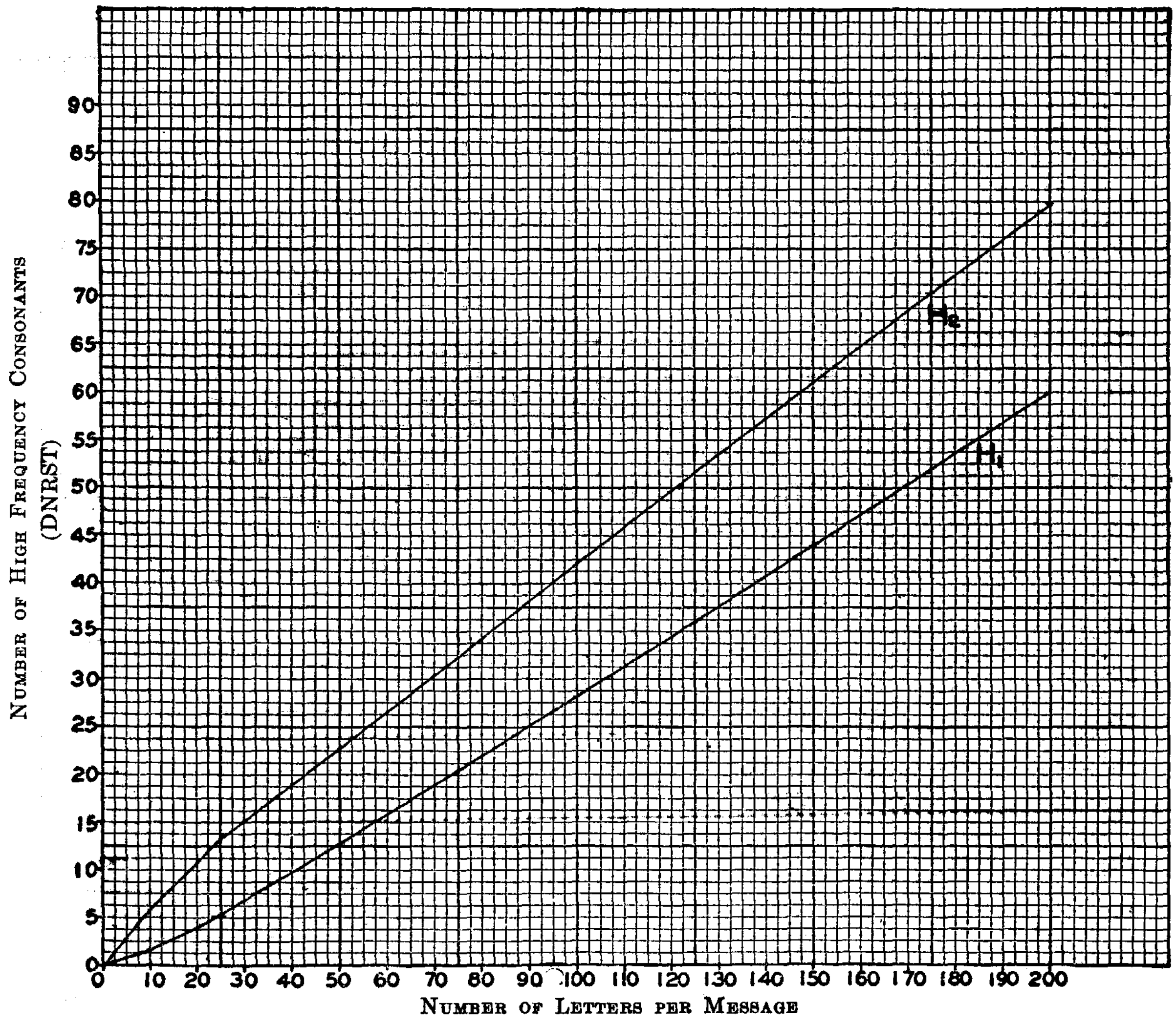

NUMBER OF HIGH FREQUENCY CONSONANTS (DNRST)
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
H_2
H_1
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200
NUMBER OF LETTERS PER MESSAGE

CHART No. 3

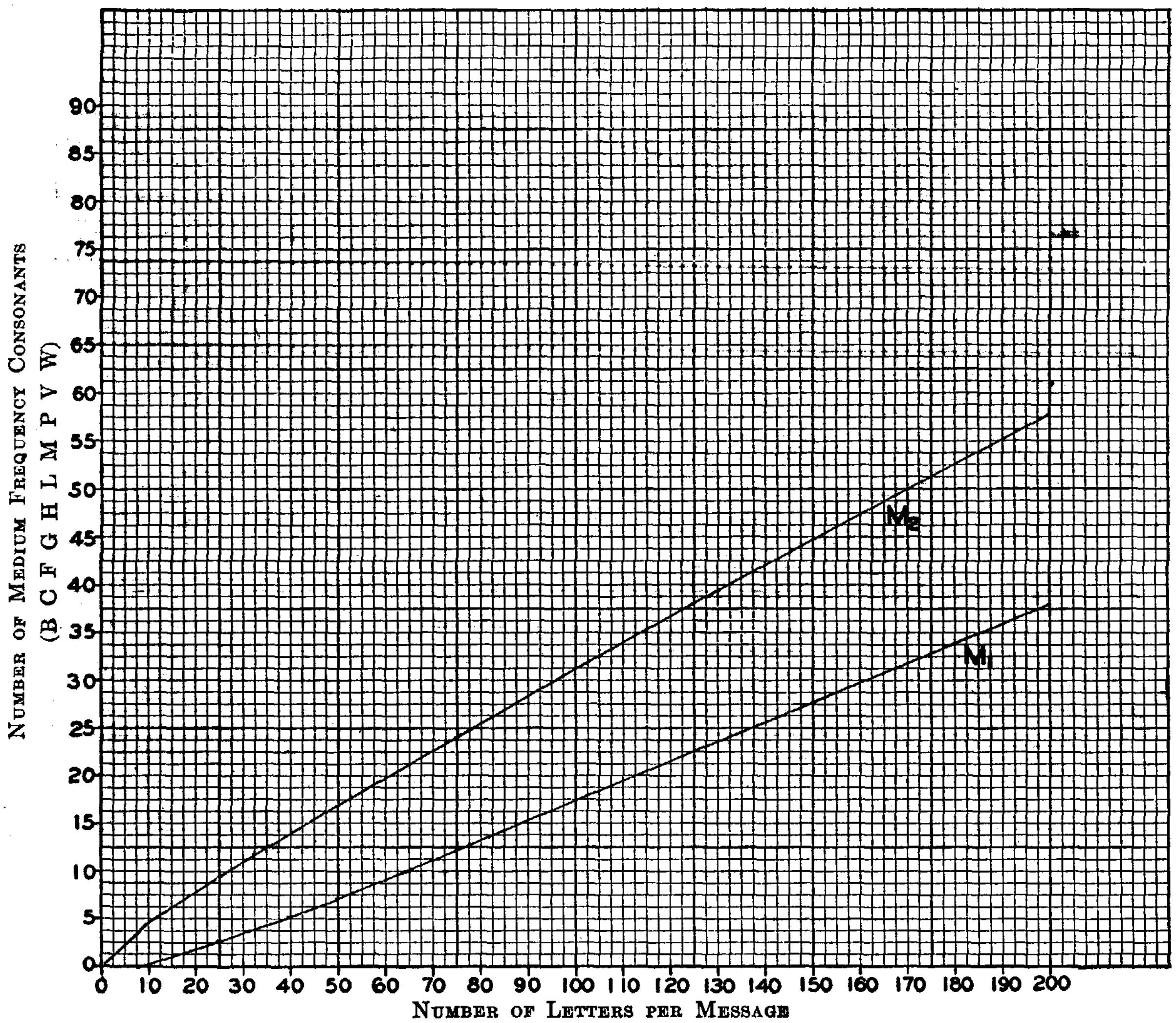

NUMBER OF MEDIUM FREQUENCY CONSONANTS
(B C F G H L M P V W)
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200
NUMBER OF LETTERS PER MESSAGE
M_2
M_1

CHART No. 4

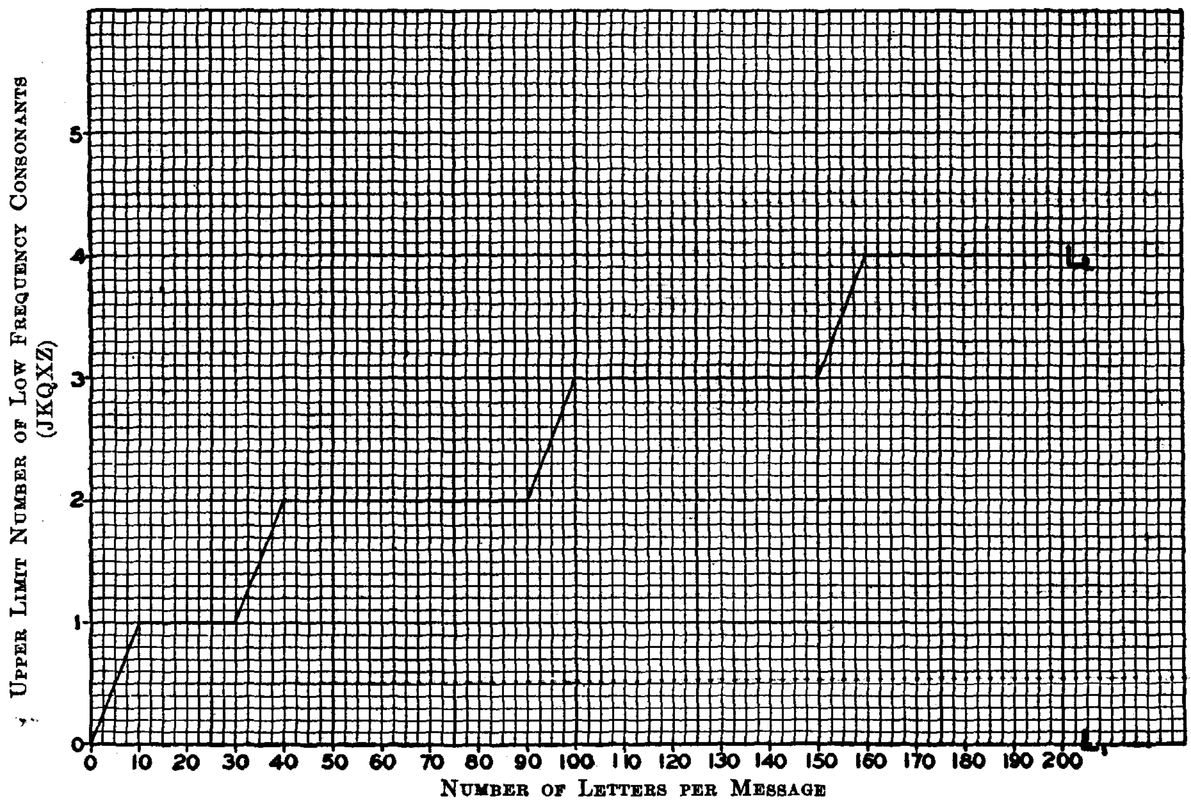
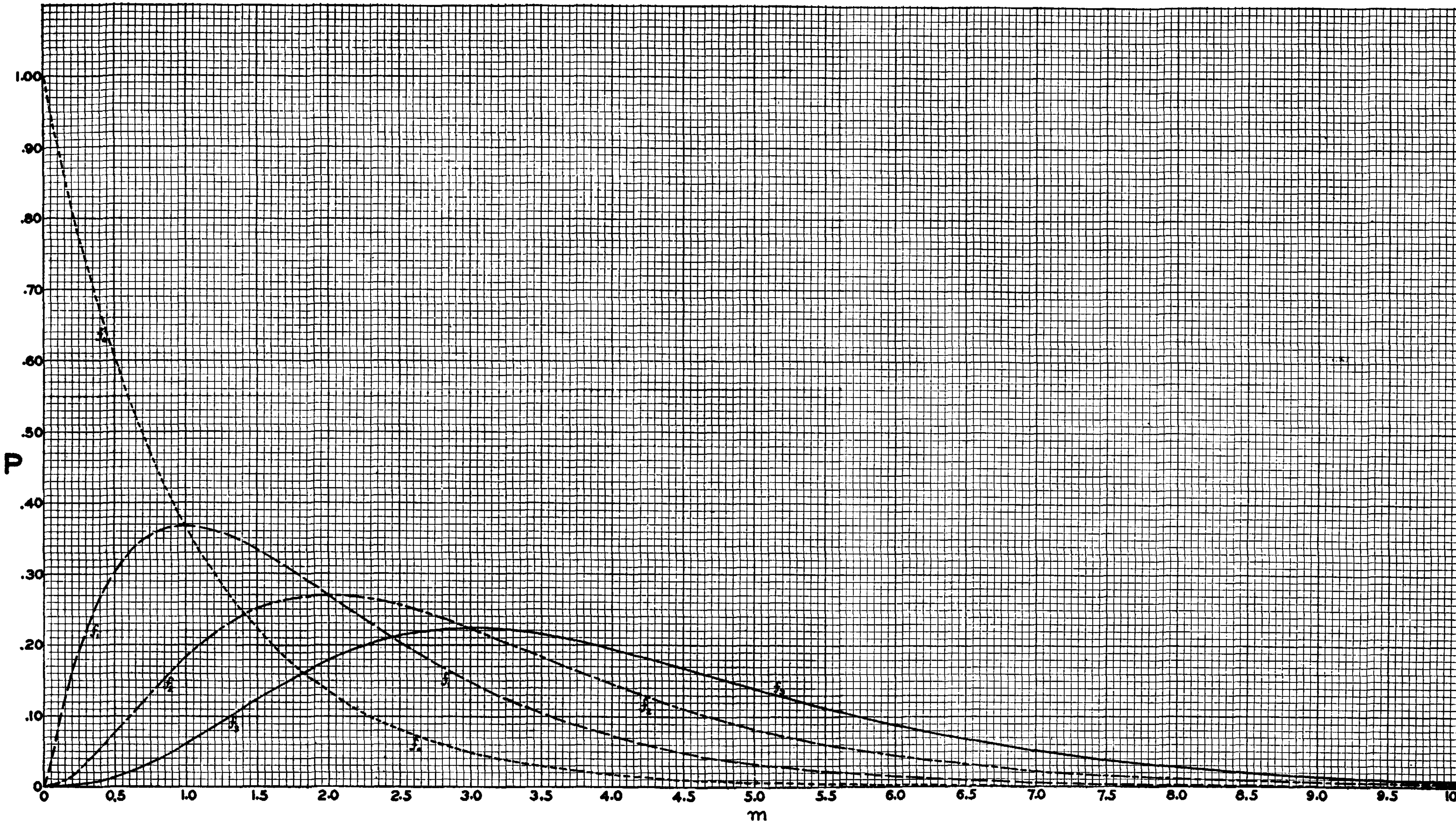

UPPER LIMIT NUMBER OF LOW FREQUENCY CONSONANTS (JKQXZ)
0
1
2
3
4
5
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200
NUMBER OF LETTERS PER MESSAGE

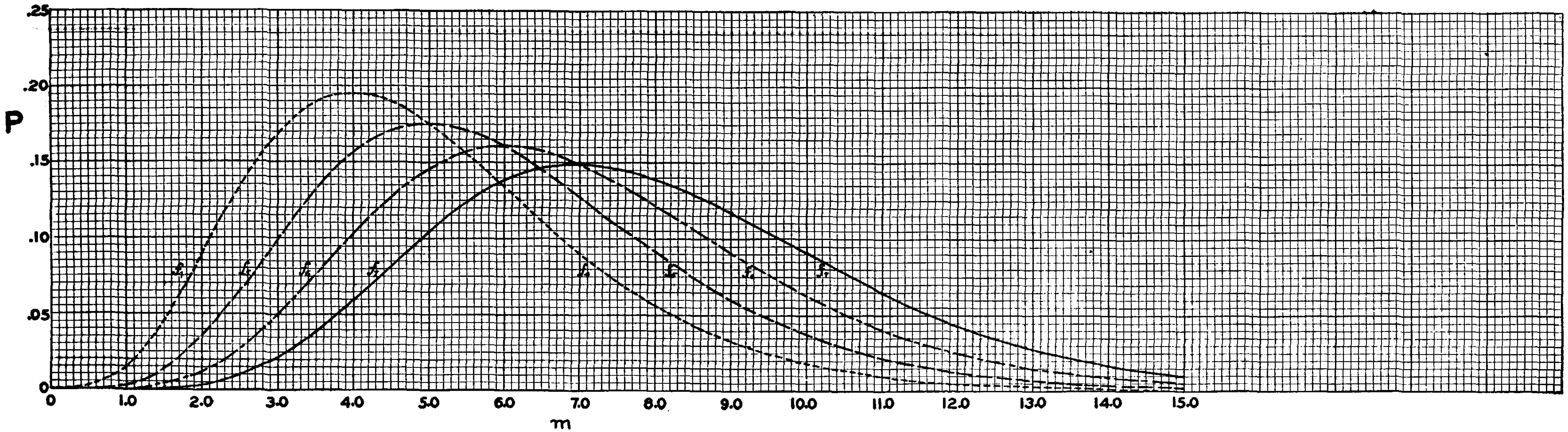
CHART No. 5.—POISSON EXPONENTIAL



CURVES SHOWING PROBABILITY FOR 0, 1, 2, AND 3 OCCURRENCES OF AN EVENT IN ACCORDANCE WITH THE POISSON EXPONENTIAL DISTRIBUTION

REF ID:A58459

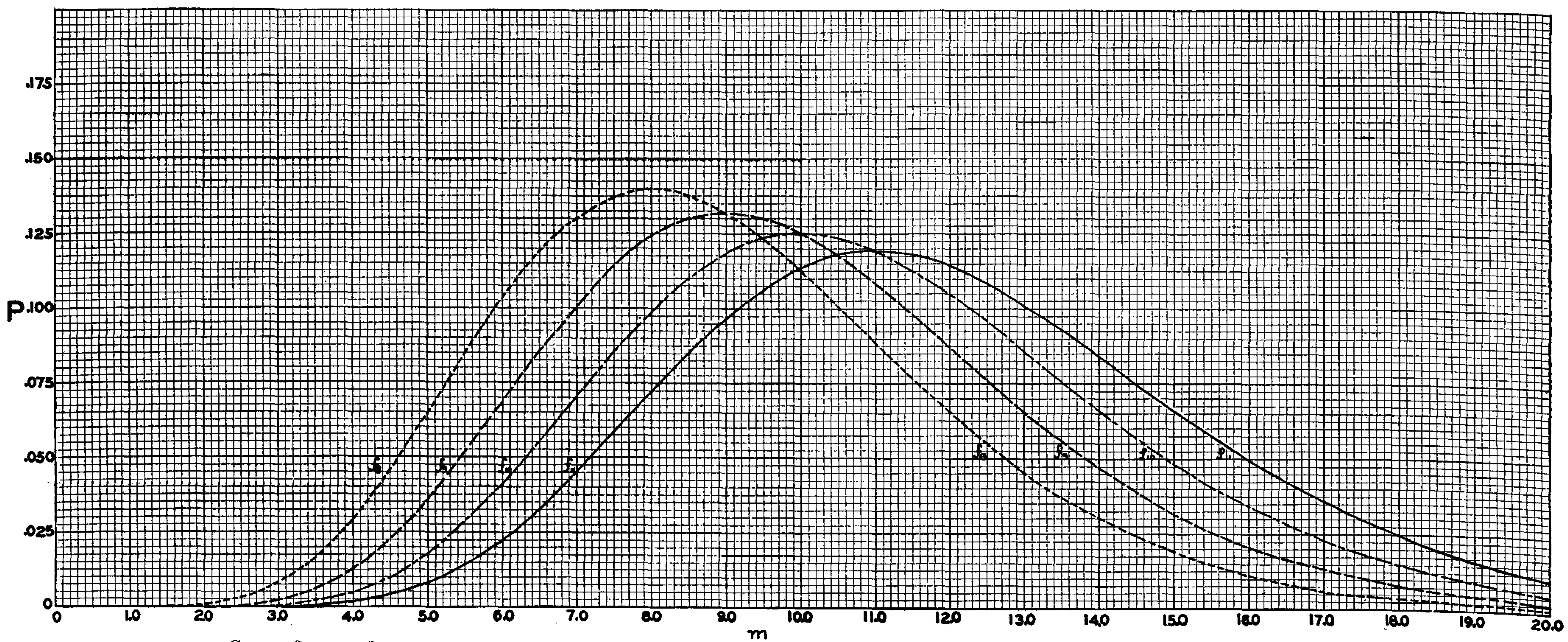
CHART NO. 6.—POISSON EXPONENTIAL



CURVES SHOWING PROBABILITY FOR 4, 5, 6, AND 7 OCCURRENCES OF AN EVENT IN ACCORDANCE WITH THE POISSON EXPONENTIAL DISTRIBUTION

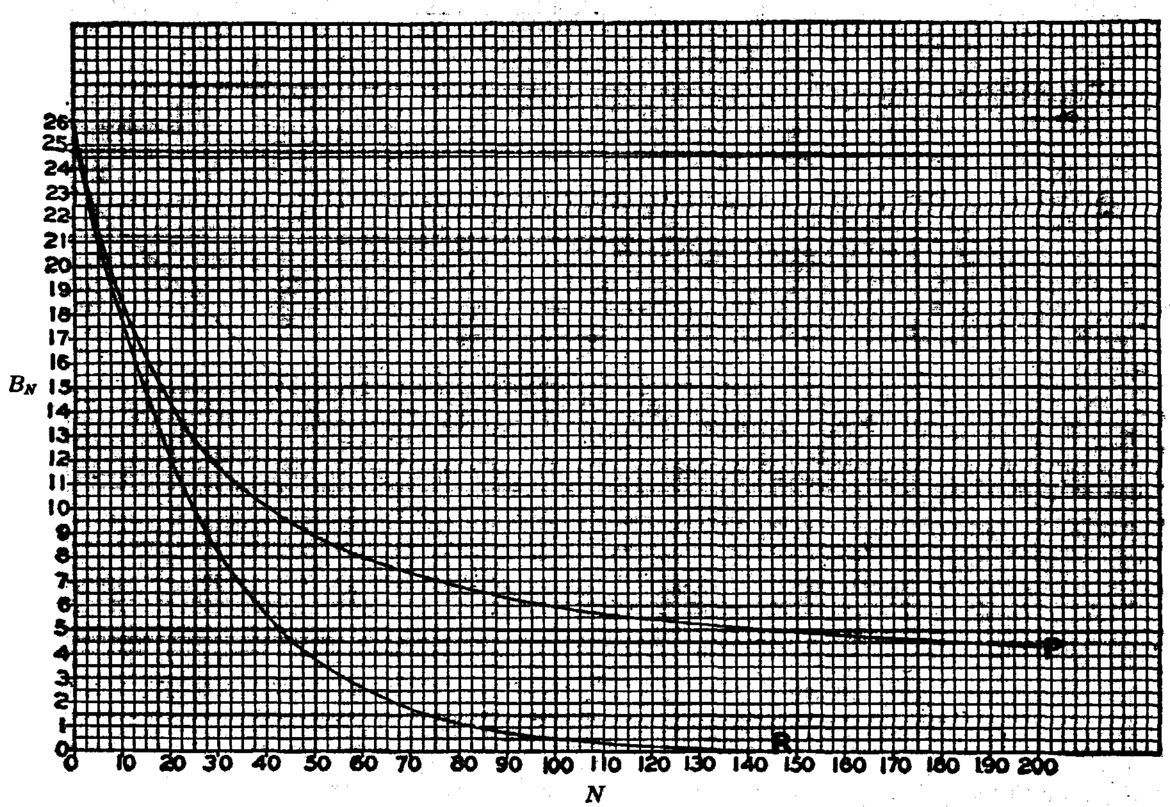
REF ID:A58459

CHART No. 7.—POISSON EXPONENTIAL



CURVES SHOWING PROBABILITY FOR 8, 9, 10, AND 11 OCCURRENCES OF AN EVENT IN ACCORDANCE WITH THE POISSON EXPONENTIAL DISTRIBUTION

CHART No. 8.—EXPECTED NUMBER OF BLANKS ENGLISH PLAIN TEXT (P) AND RANDOM TEXT (R)



NUMBER OF LETTERS PER MESSAGE

CHART NO. 9.—EXPECTED NUMBER OF BLANKS FRENCH PLAIN TEXT

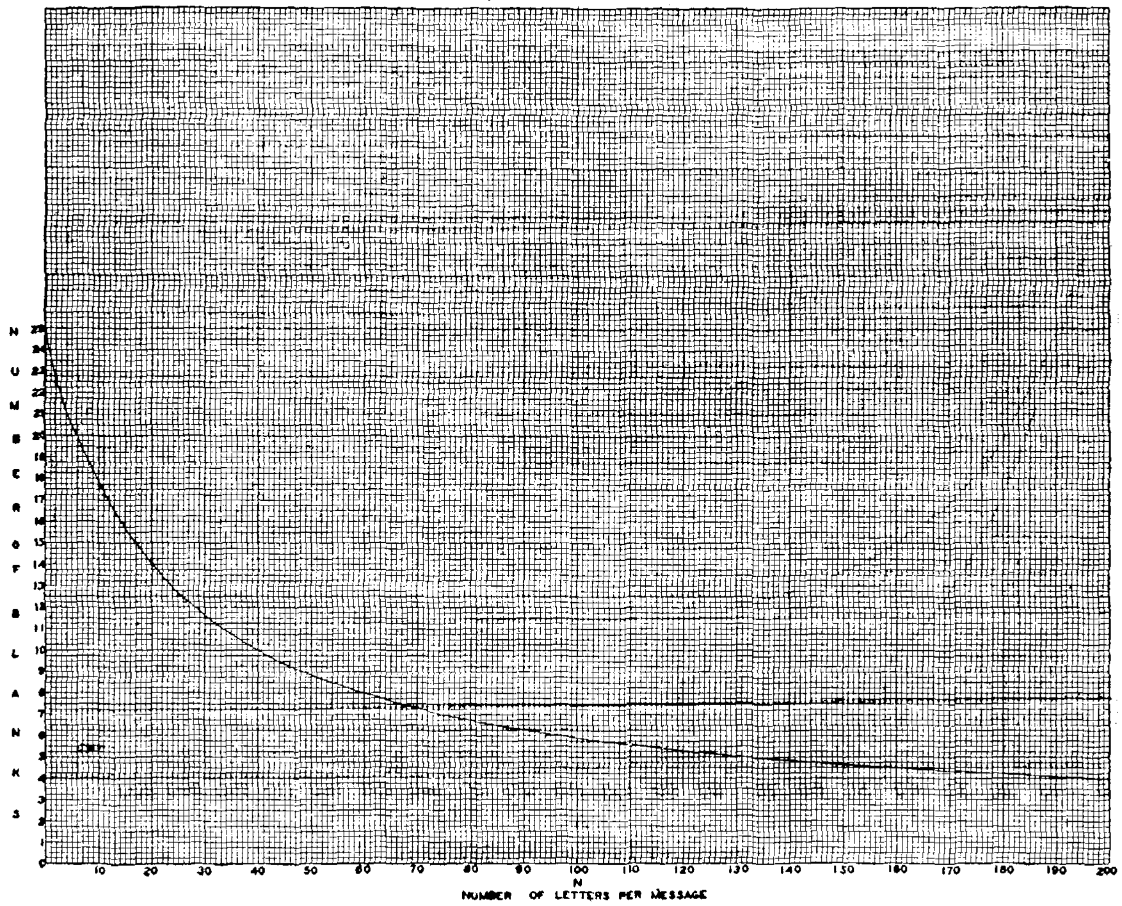


CHART No. 10.—EXPECTED NUMBER OF BLANKS GERMAN PLAIN TEXT

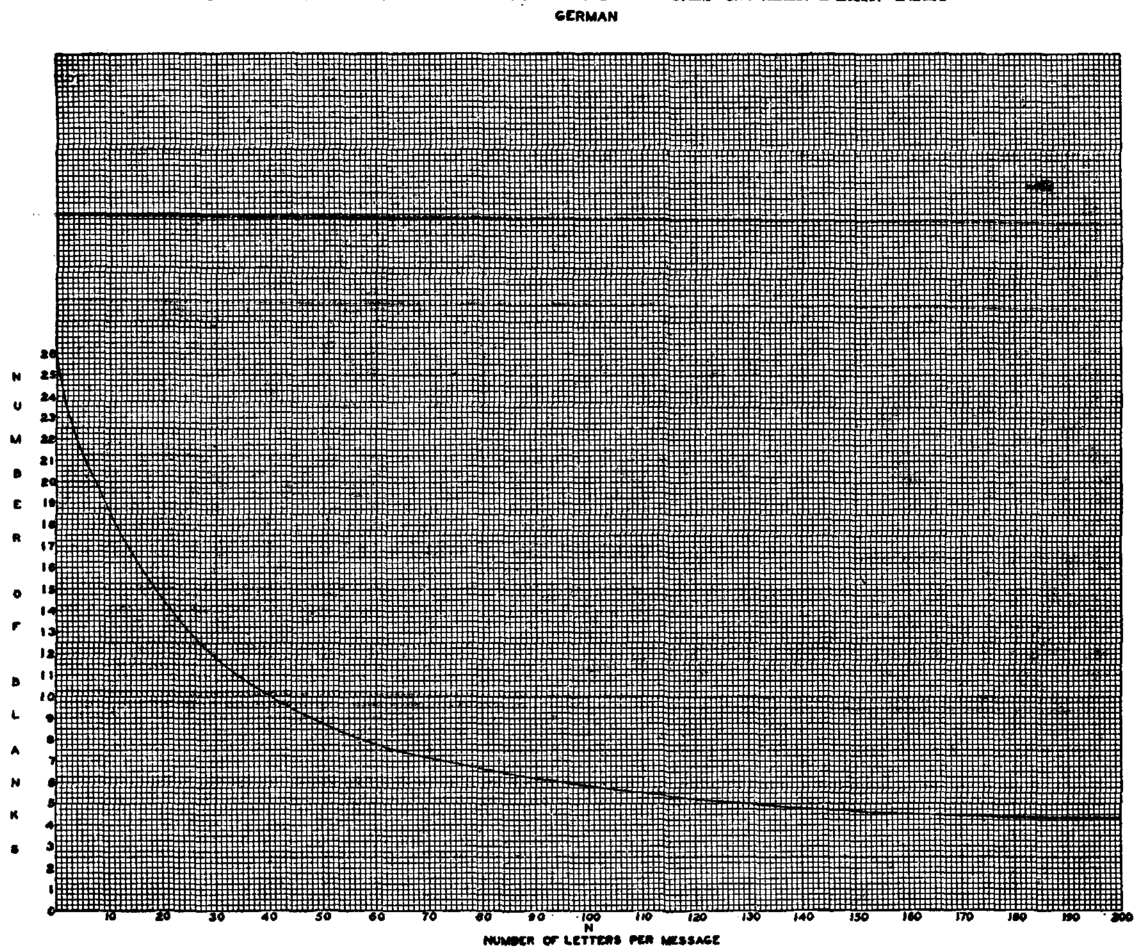


CHART No. 11.—EXPECTED NUMBER OF BLANKS ITALIAN PLAIN TEXT

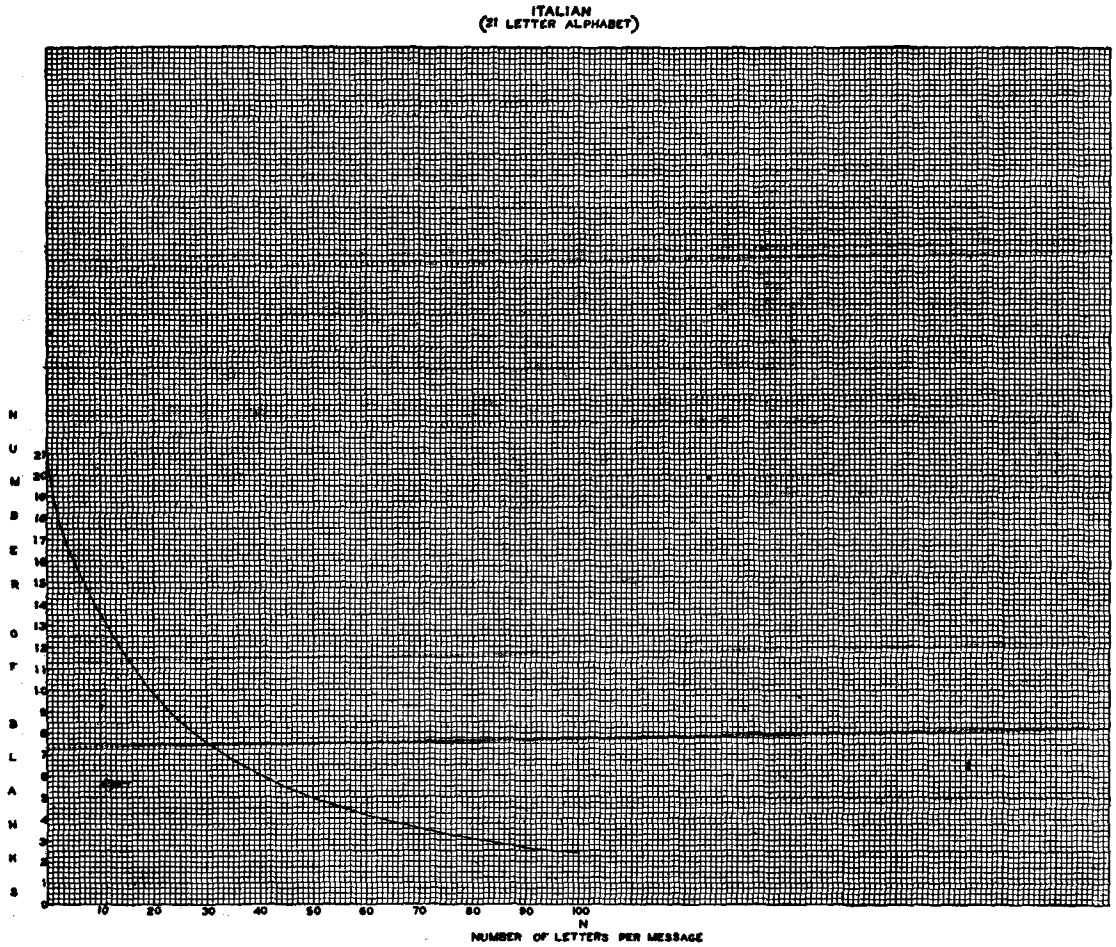


CHART No. 12.—EXPECTED NUMBER OF BLANKS PORTUGUESE PLAIN TEXT

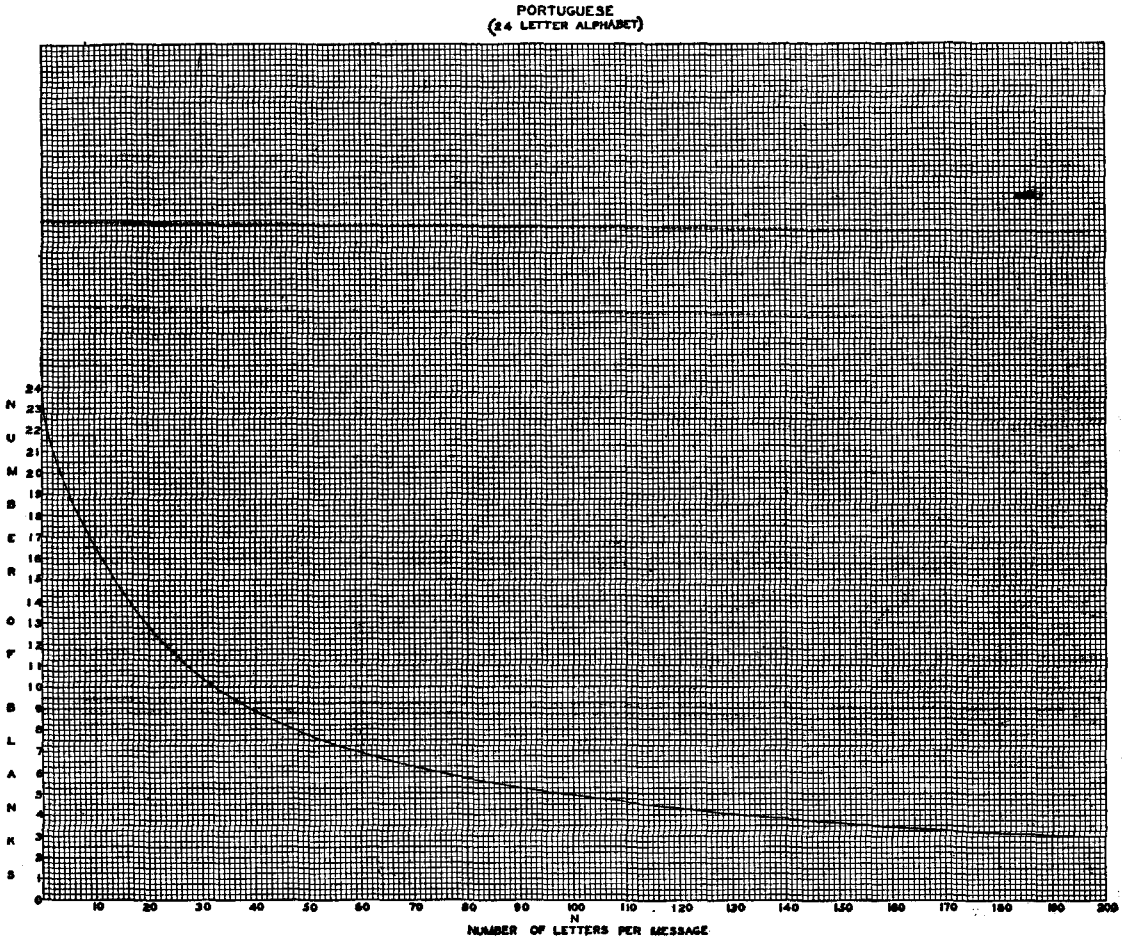


CHART No. 13.—EXPECTED NUMBER OF BLANKS SPANISH PLAIN TEXT

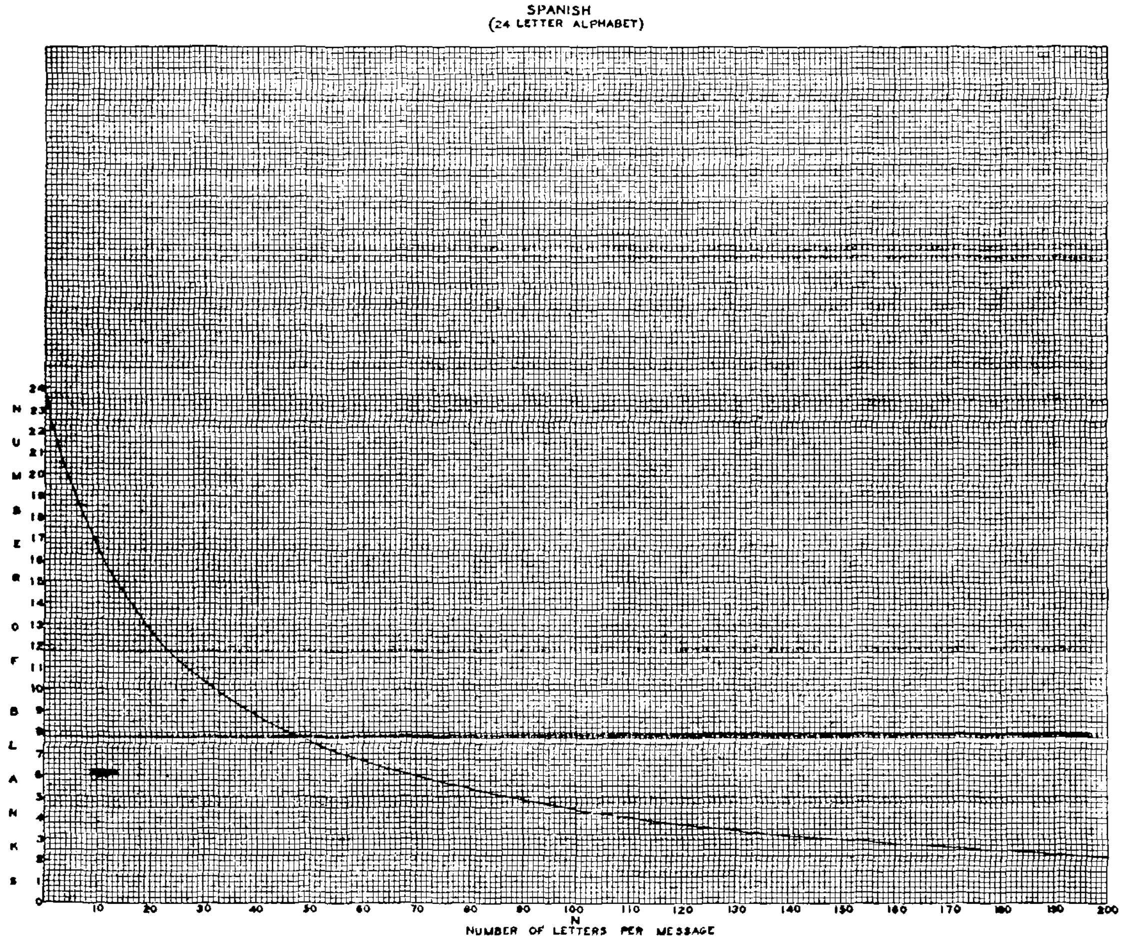


CHART NO. 14.—EXPECTED VALUE AND STANDARD DEVIATION OF ϕ

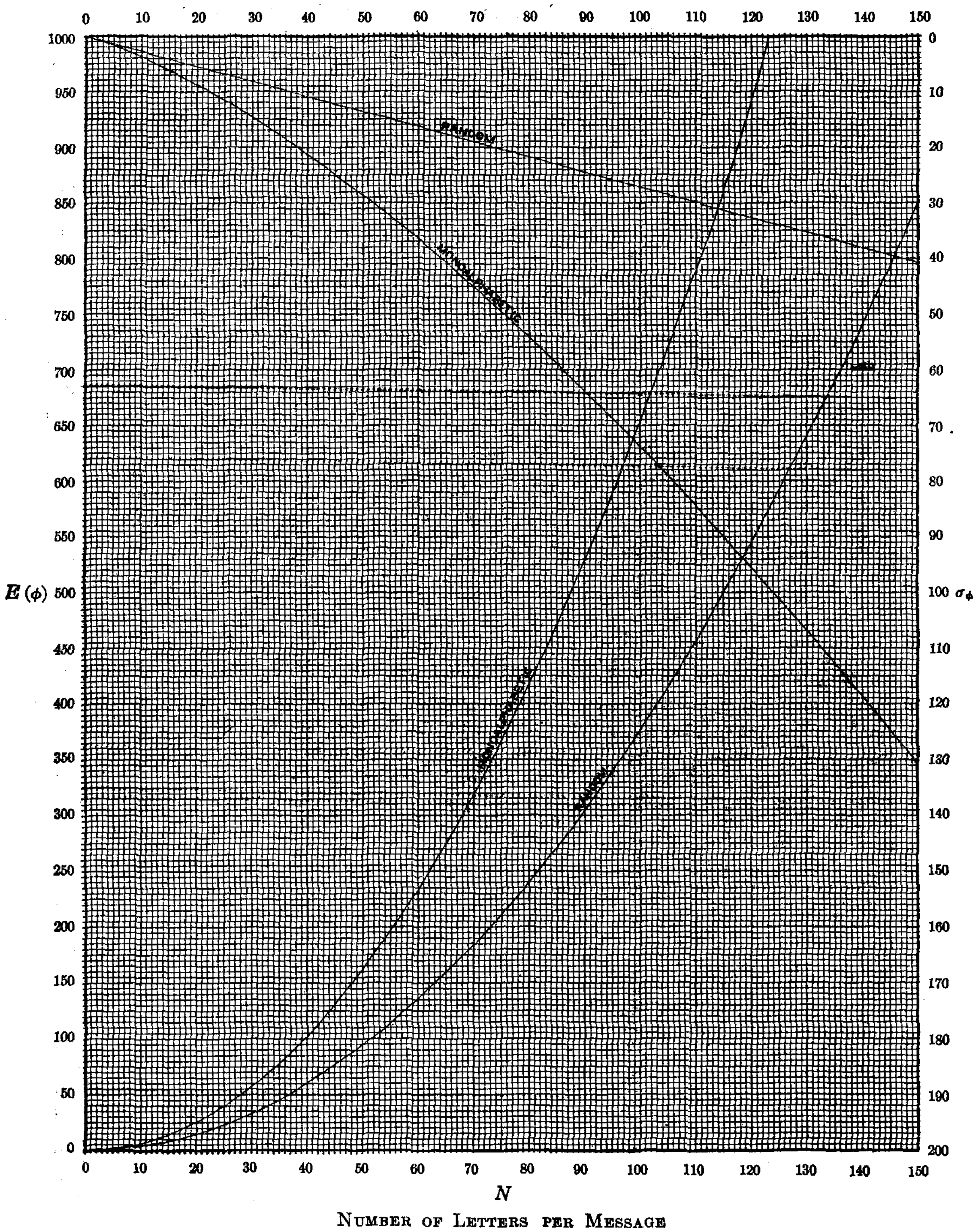
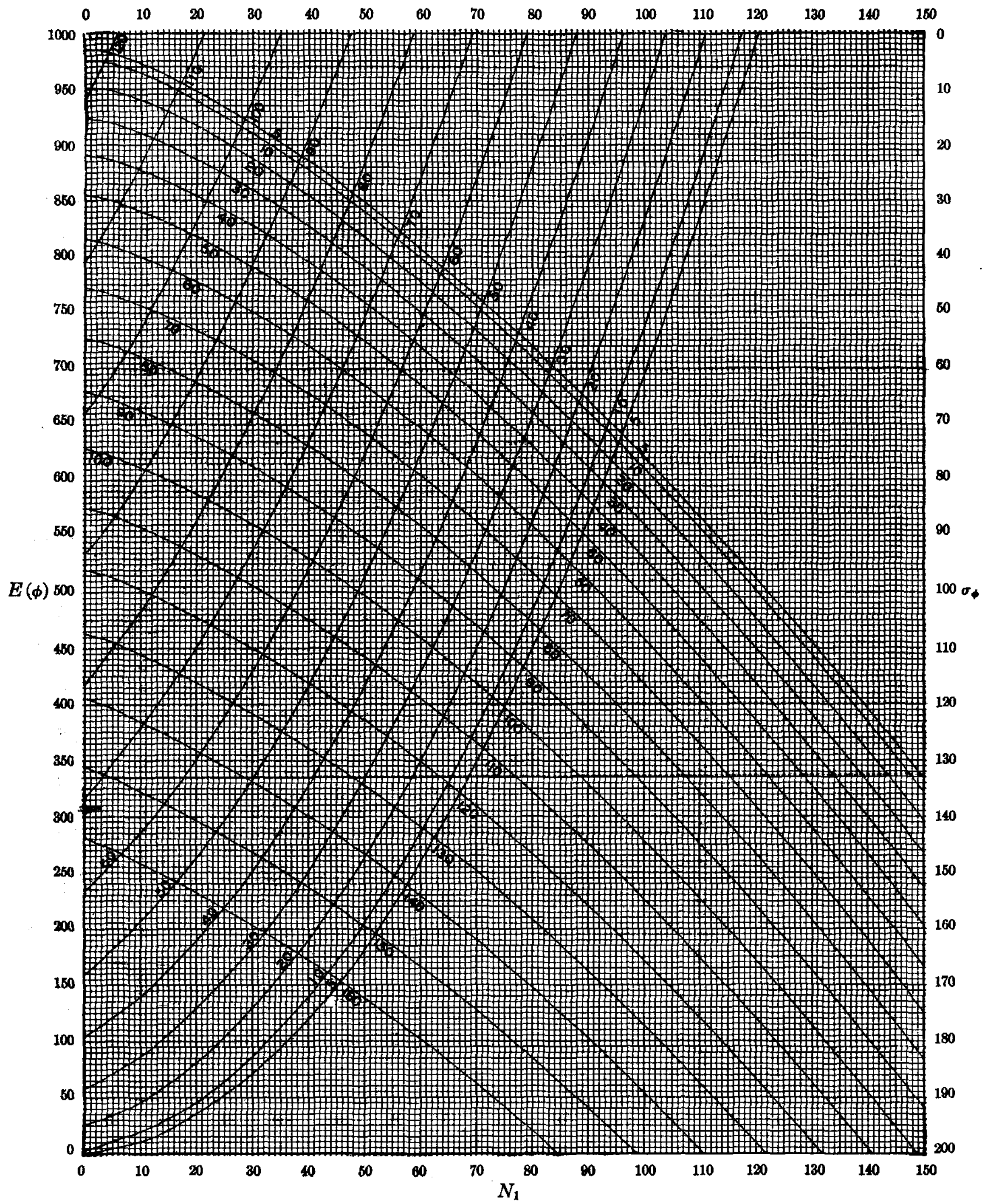
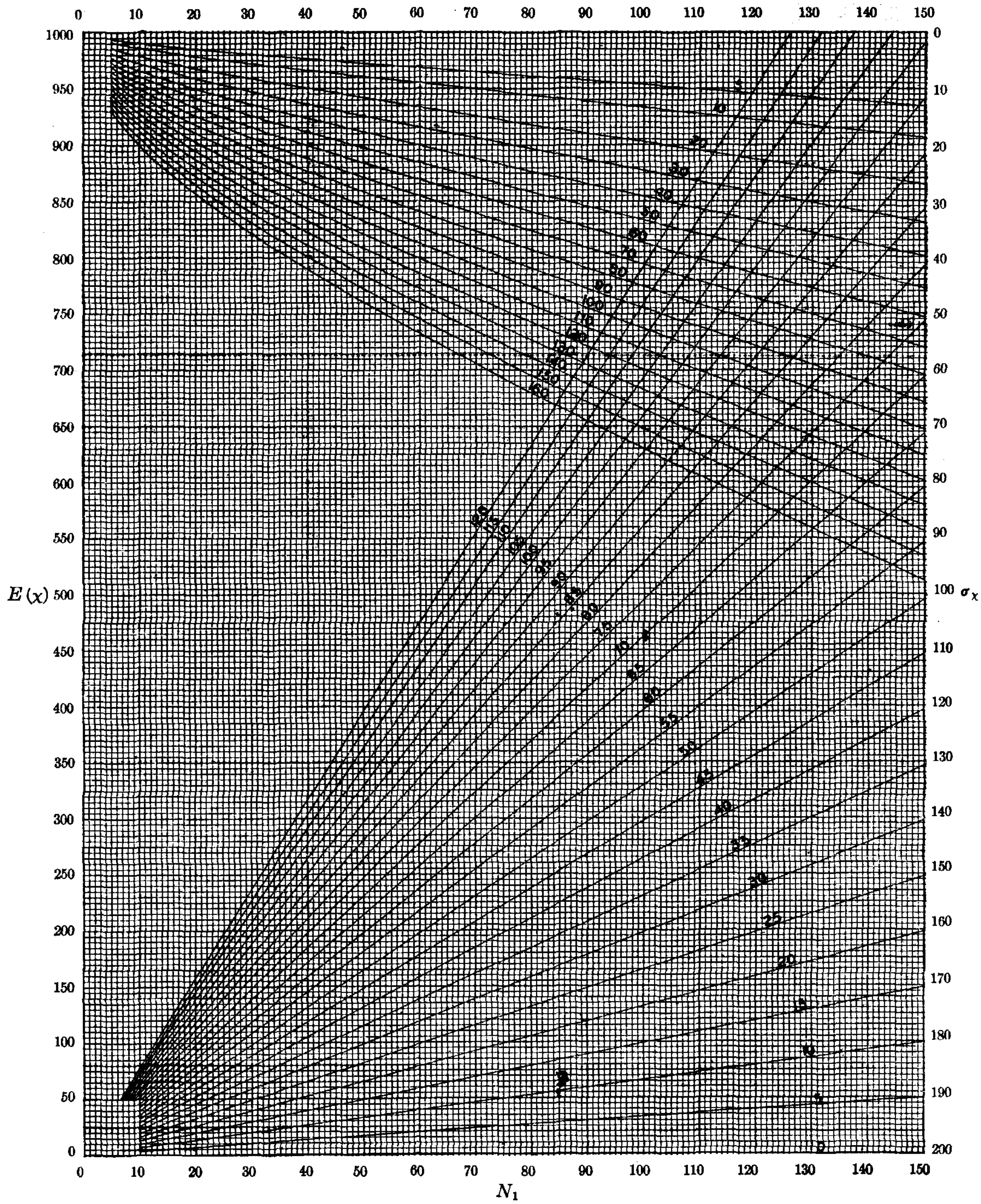


CHART No. 15.—EXPECTED VALUE AND STANDARD DEVIATION OF ϕ
NON-MATCHING PAIRS OF MONOALPHABETS



(The value of N_2 is given on the curve corresponding thereto)

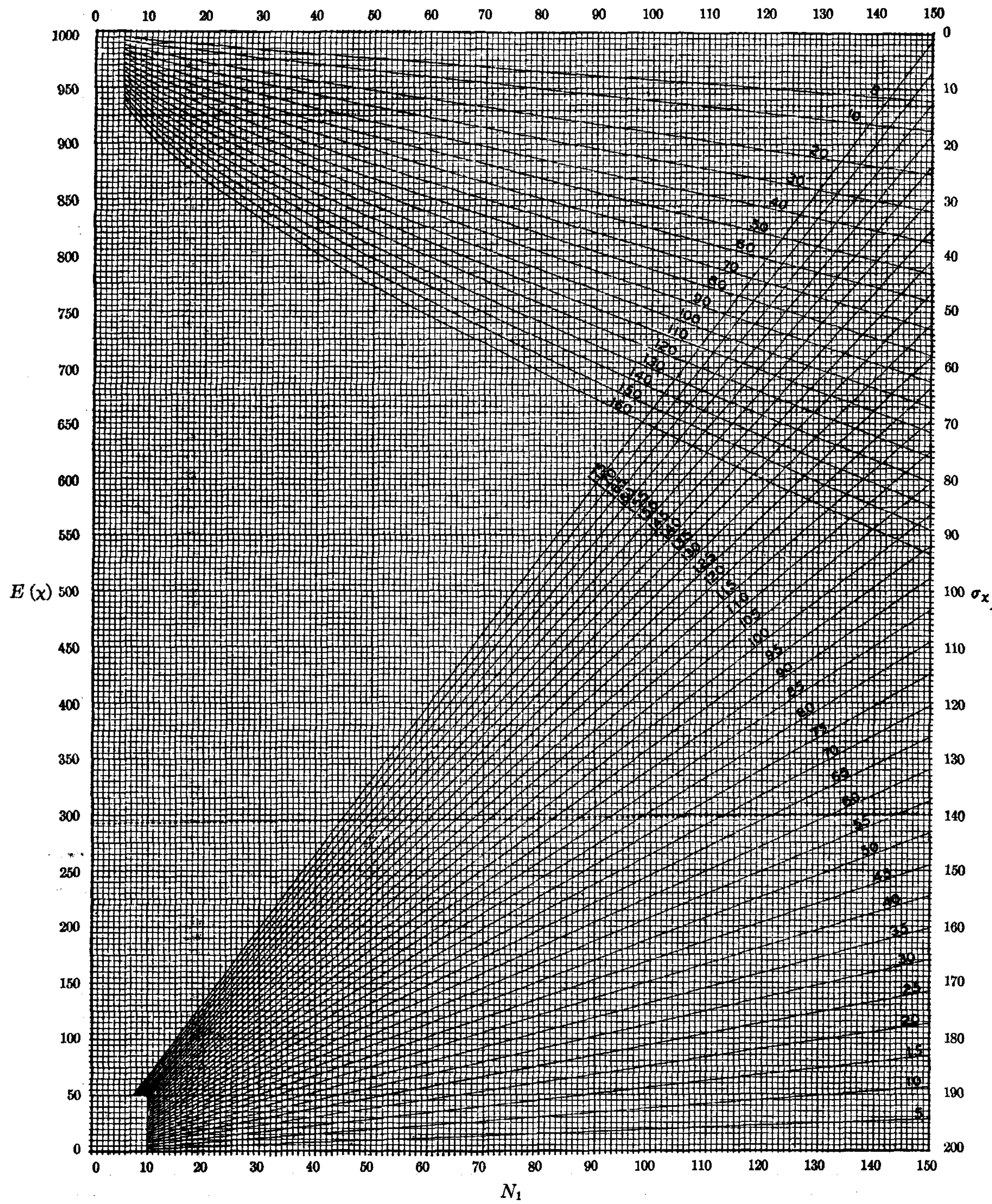
CHART No. 16.—EXPECTED VALUE AND STANDARD DEVIATION OF χ, MATCHING PAIRS OF MONOALPHABETS



(The value of N_2 is given on the curve corresponding thereto)

63301—38——12

CHART NO. 17.—EXPECTED VALUE AND STANDARD DEVIATION OF χ, NON-MATCHING PAIRS OF MONOALPHABETS



(The value of N_2 is given on the curve corresponding thereto)

CHART No. 18

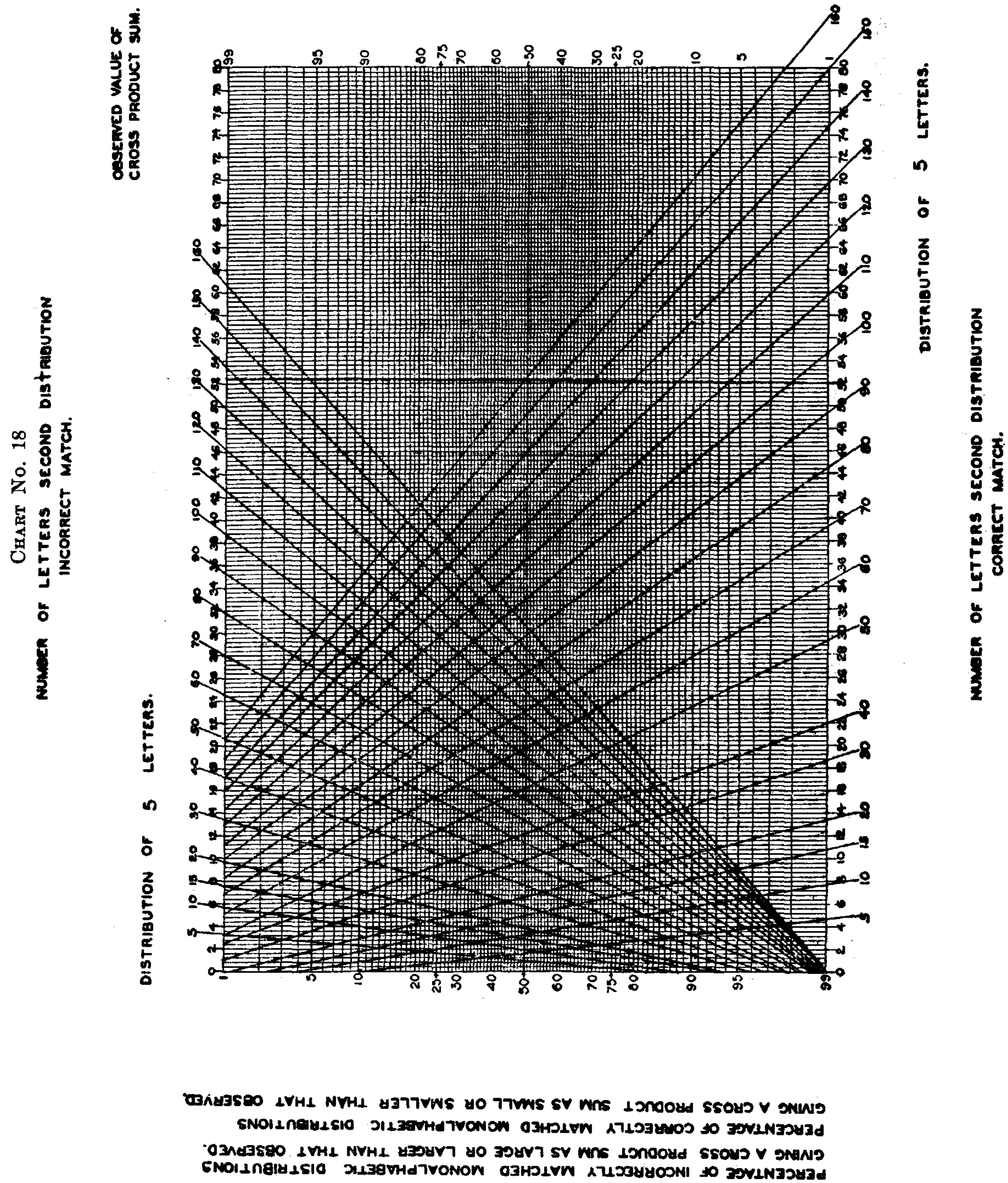
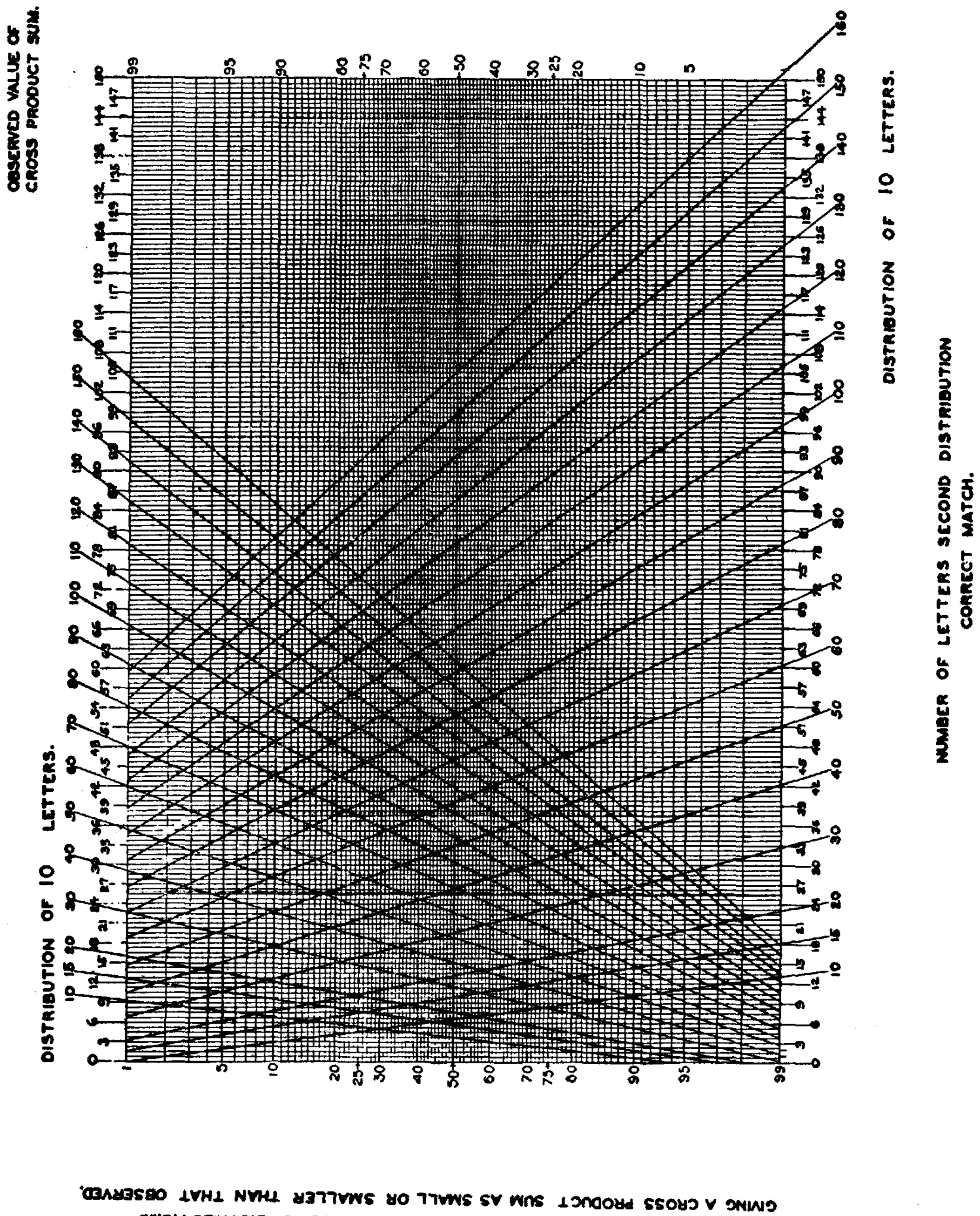

NUMBER OF LETTERS SECOND DISTRIBUTION INCORRECT MATCH.
DISTRIBUTION OF 5 LETTERS.
OBSERVED VALUE OF CROSS PRODUCT SUM.
DISTRIBUTION OF 5 LETTERS.
NUMBER OF LETTERS SECOND DISTRIBUTION CORRECT MATCH.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.
PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.

CHART No. 19


NUMBER OF LETTERS SECOND DISTRIBUTION INCORRECT MATCH.
DISTRIBUTION OF 10 LETTERS.
OBSERVED VALUE OF CROSS PRODUCT SUM.
DISTRIBUTION OF 10 LETTERS.
NUMBER OF LETTERS SECOND DISTRIBUTION CORRECT MATCH.
PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.

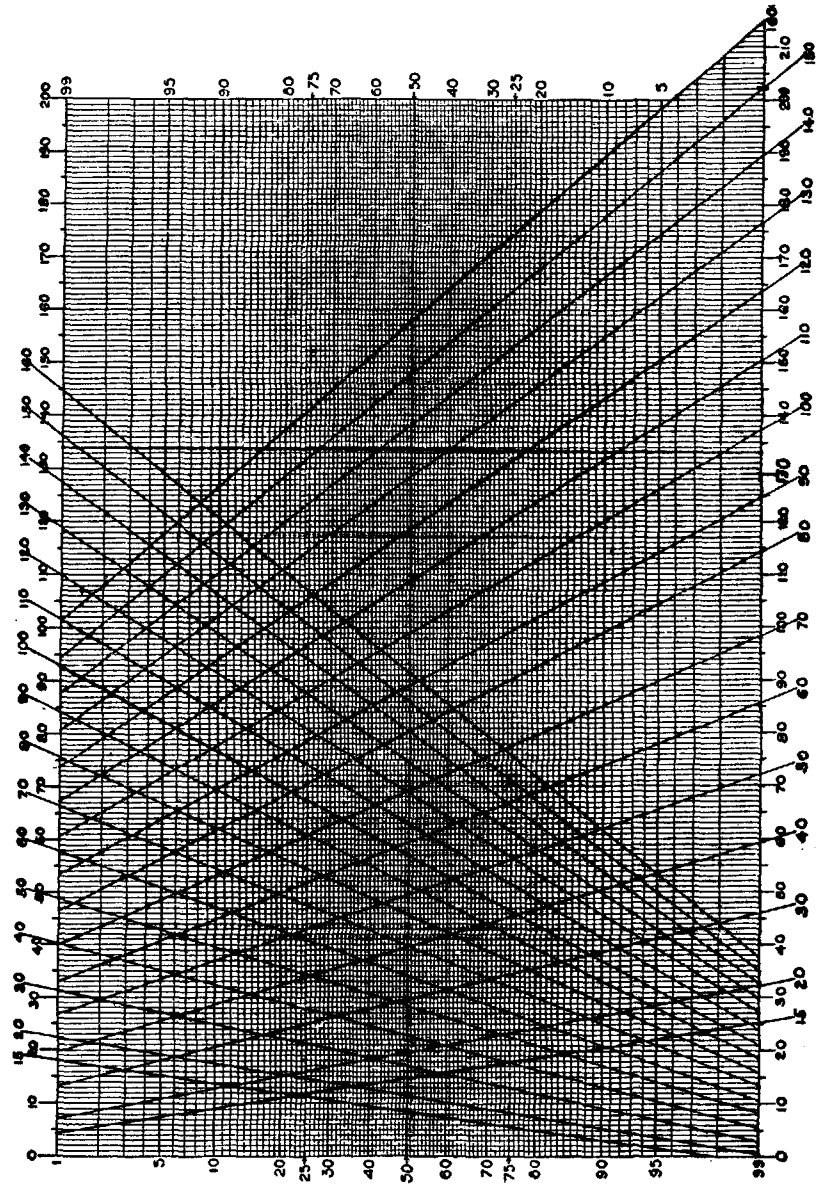


CHART No. 20

NUMBER OF LETTERS SECOND DISTRIBUTION INCORRECT MATCH.

OBSERVED VALUE OF CROSS PRODUCT SUM.

DISTRIBUTION OF 15 LETTERS.

DISTRIBUTION OF 15 LETTERS.

NUMBER OF LETTERS SECOND DISTRIBUTION CORRECT MATCH.

PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.

PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.

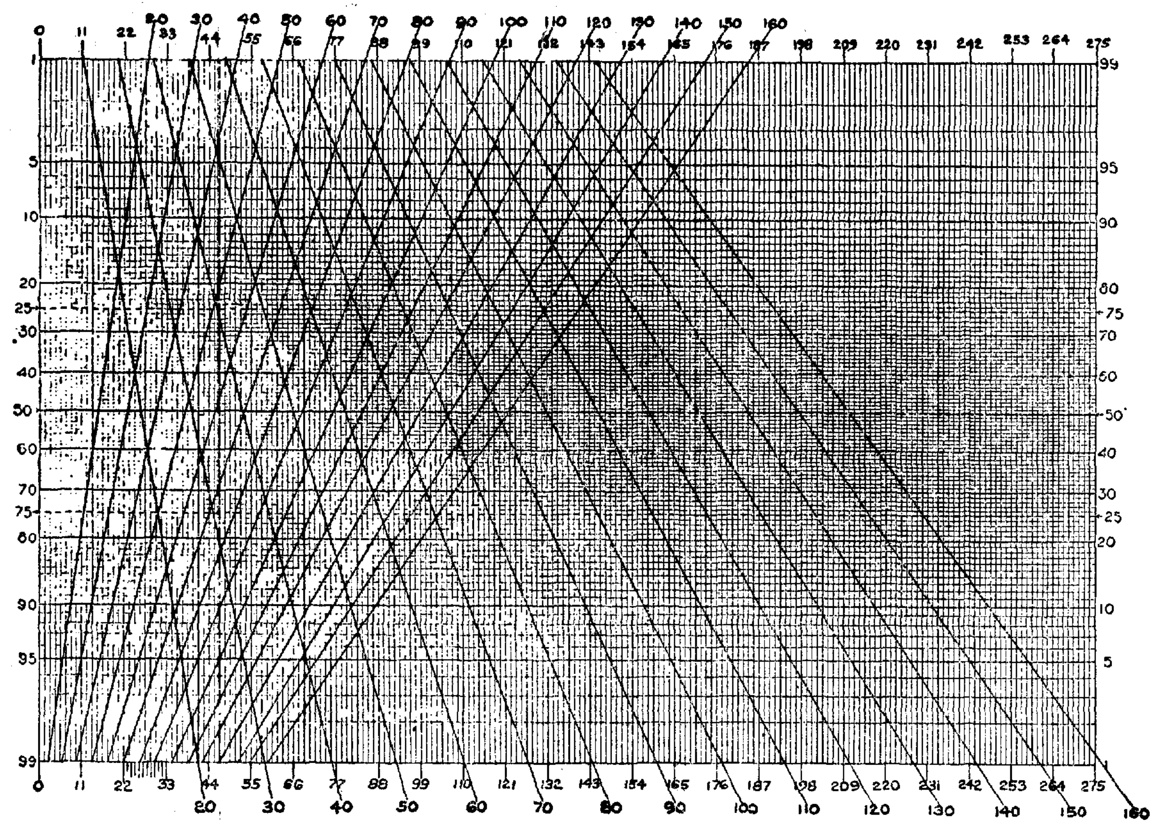
CHART No. 21

NUMBER OF LETTERS SECOND DISTRIBUTION
INCORRECT MATCH.

OBSERVED VALUE OF
CROSS PRODUCT SUM.

DISTRIBUTION OF 20 LETTERS.

PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.



DISTRIBUTION OF 20 LETTERS.

NUMBER OF LETTERS SECOND DISTRIBUTION
CORRECT MATCH.

CHART No. 22

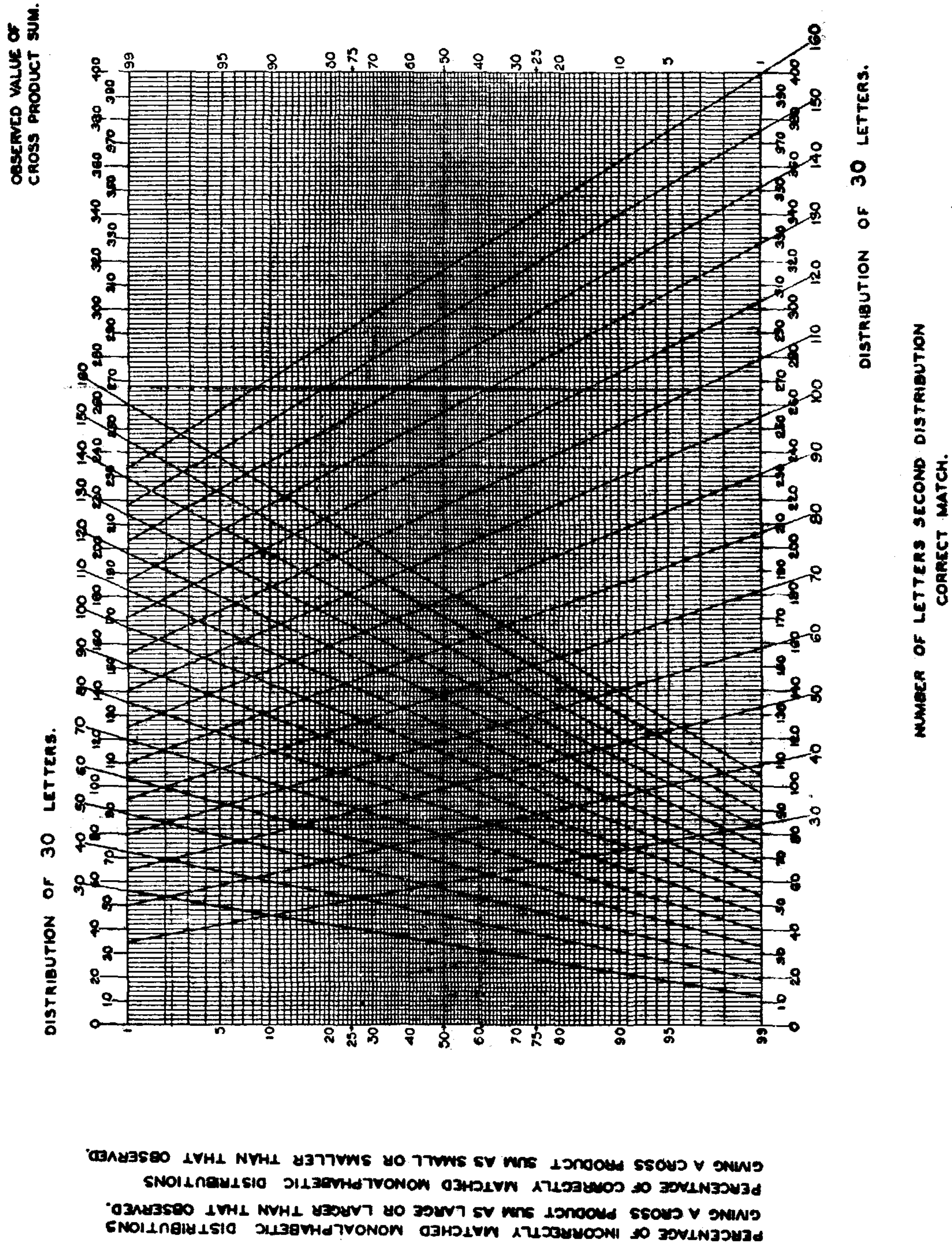

NUMBER OF LETTERS SECOND DISTRIBUTION INCORRECT MATCH.
DISTRIBUTION OF 30 LETTERS.
OBSERVED VALUE OF CROSS PRODUCT SUM.
DISTRIBUTION OF 30 LETTERS.
NUMBER OF LETTERS SECOND DISTRIBUTION CORRECT MATCH.
PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.

Chart No. 23

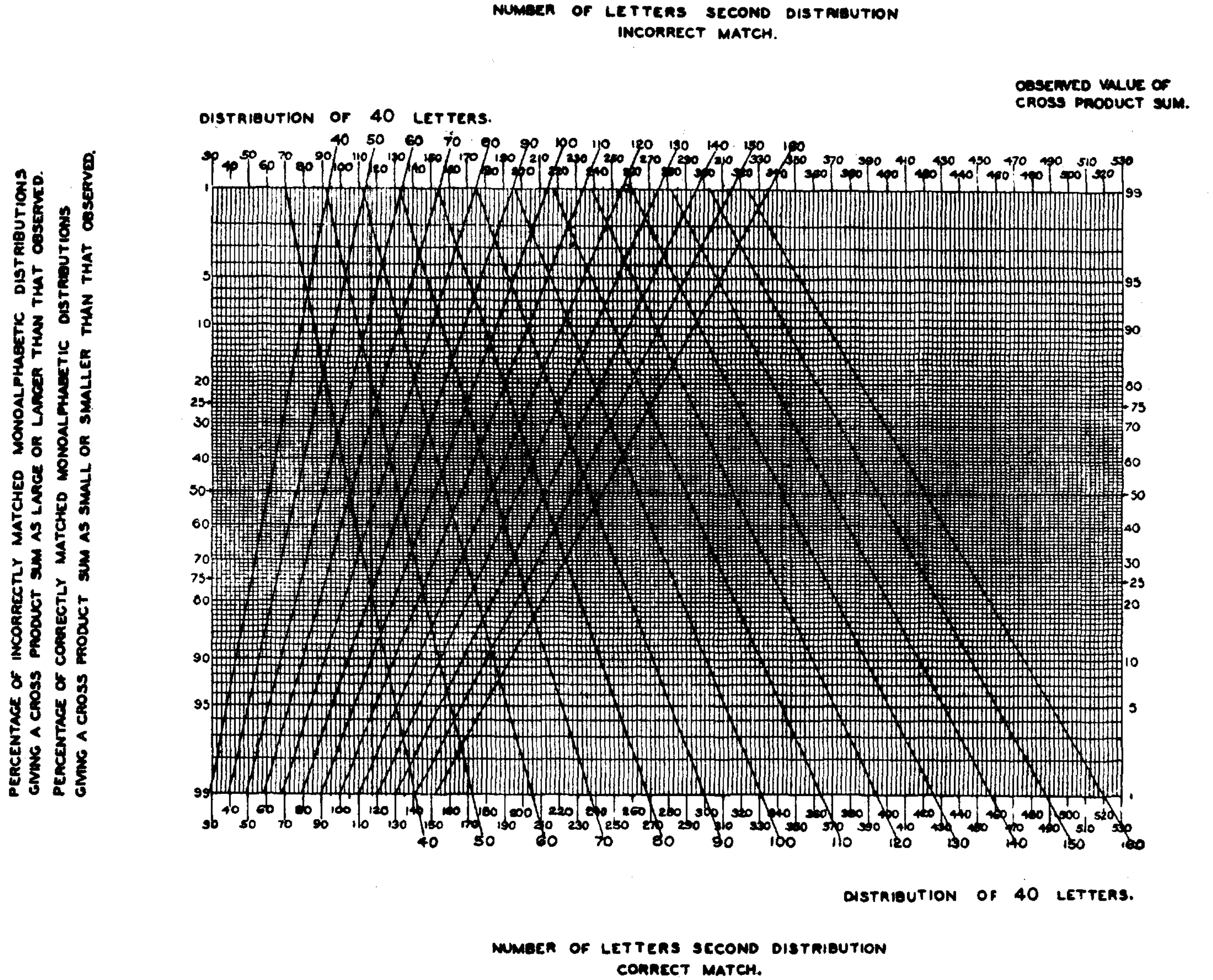
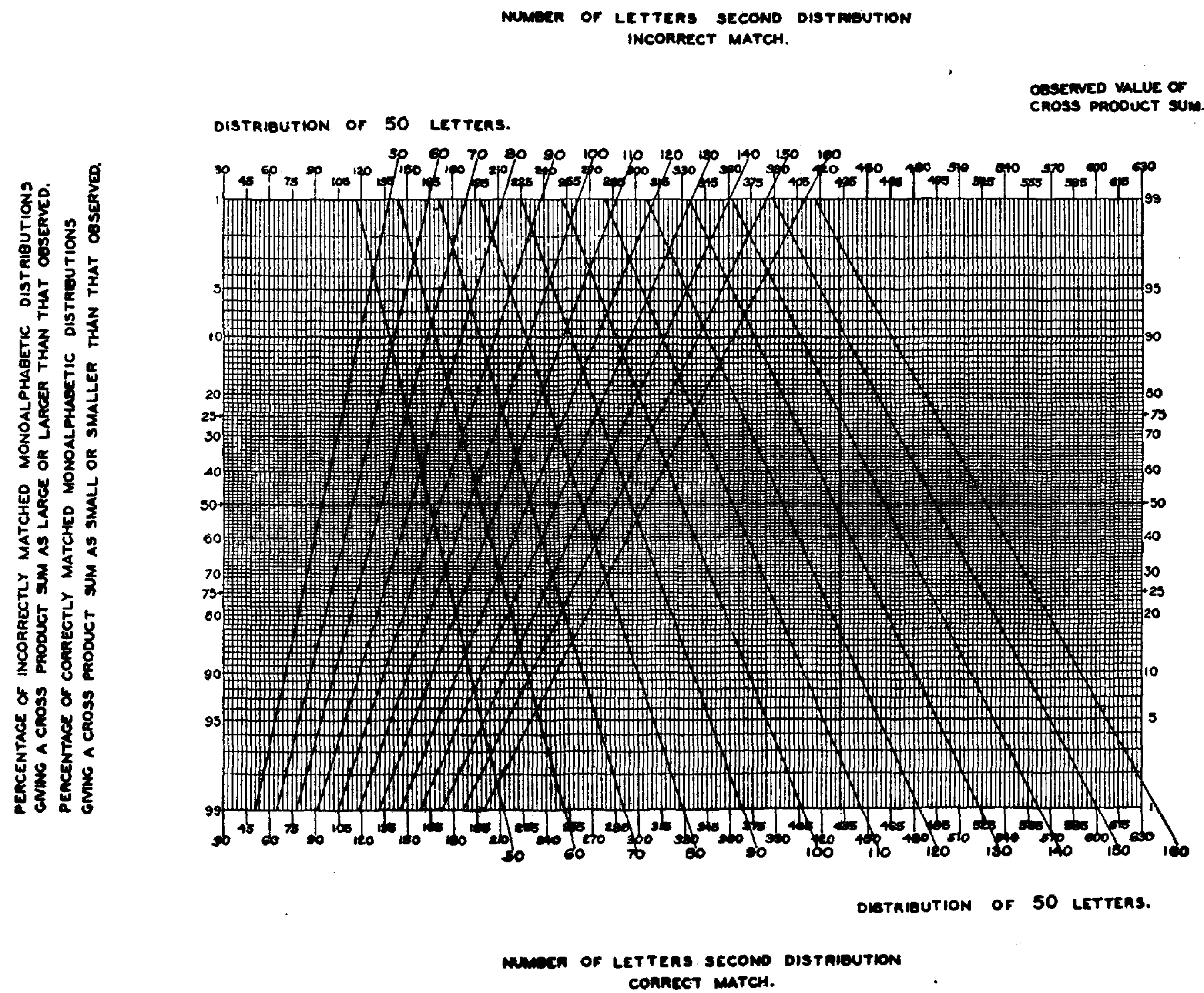

NUMBER OF LETTERS SECOND DISTRIBUTION
INCORRECT MATCH.
OBSERVED VALUE OF
CROSS PRODUCT SUM.
DISTRIBUTION OF 40 LETTERS.
PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.
DISTRIBUTION OF 40 LETTERS.
NUMBER OF LETTERS SECOND DISTRIBUTION
CORRECT MATCH.

CHART No. 24



NUMBER OF LETTERS SECOND DISTRIBUTION
INCORRECT MATCH.
OBSERVED VALUE OF
CROSS PRODUCT SUM.
DISTRIBUTION OF 50 LETTERS.
PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.
DISTRIBUTION OF 50 LETTERS.
NUMBER OF LETTERS SECOND DISTRIBUTION
CORRECT MATCH.

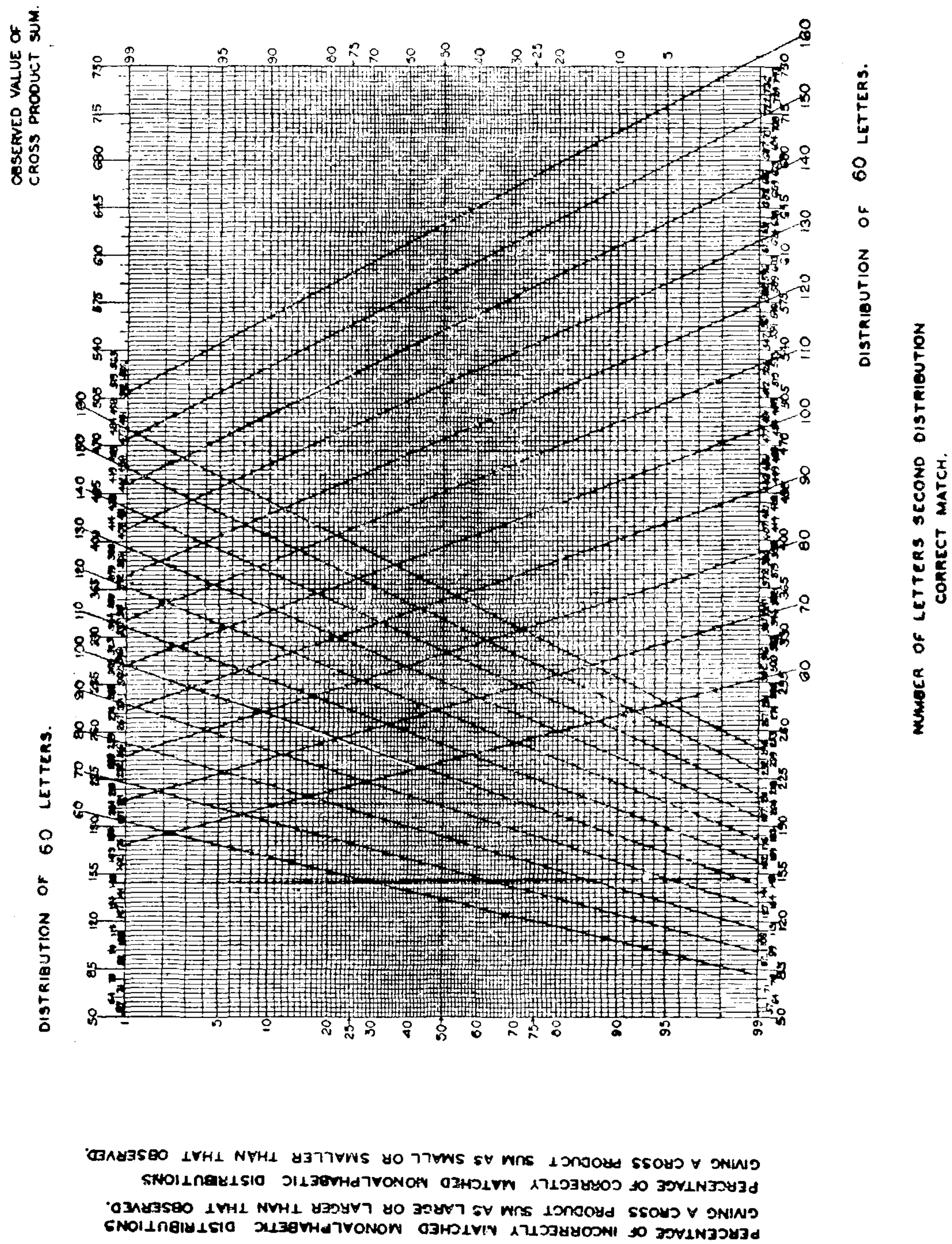


CHART No. 25

NUMBER OF LETTERS SECOND DISTRIBUTION INCORRECT MATCH.

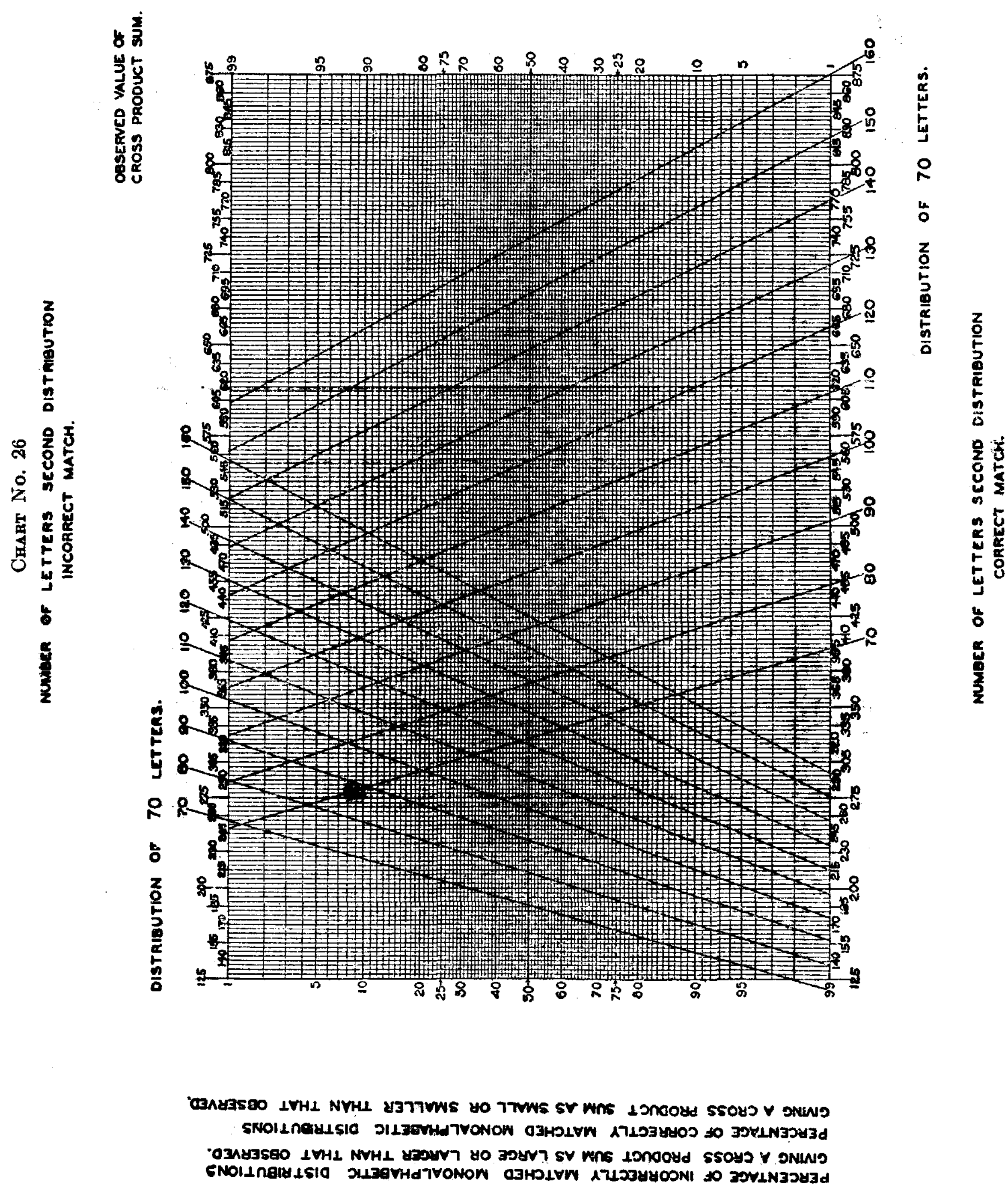


CHART No. 26
NUMBER OF LETTERS SECOND DISTRIBUTION INCORRECT MATCH.
DISTRIBUTION OF 70 LETTERS.
OBSERVED VALUE OF CROSS PRODUCT SUM.
DISTRIBUTION OF 70 LETTERS.
NUMBER OF LETTERS SECOND DISTRIBUTION CORRECT MATCH.
PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.

CHART No. 27

NUMBER OF LETTERS SECOND DISTRIBUTION INCORRECT MATCH.

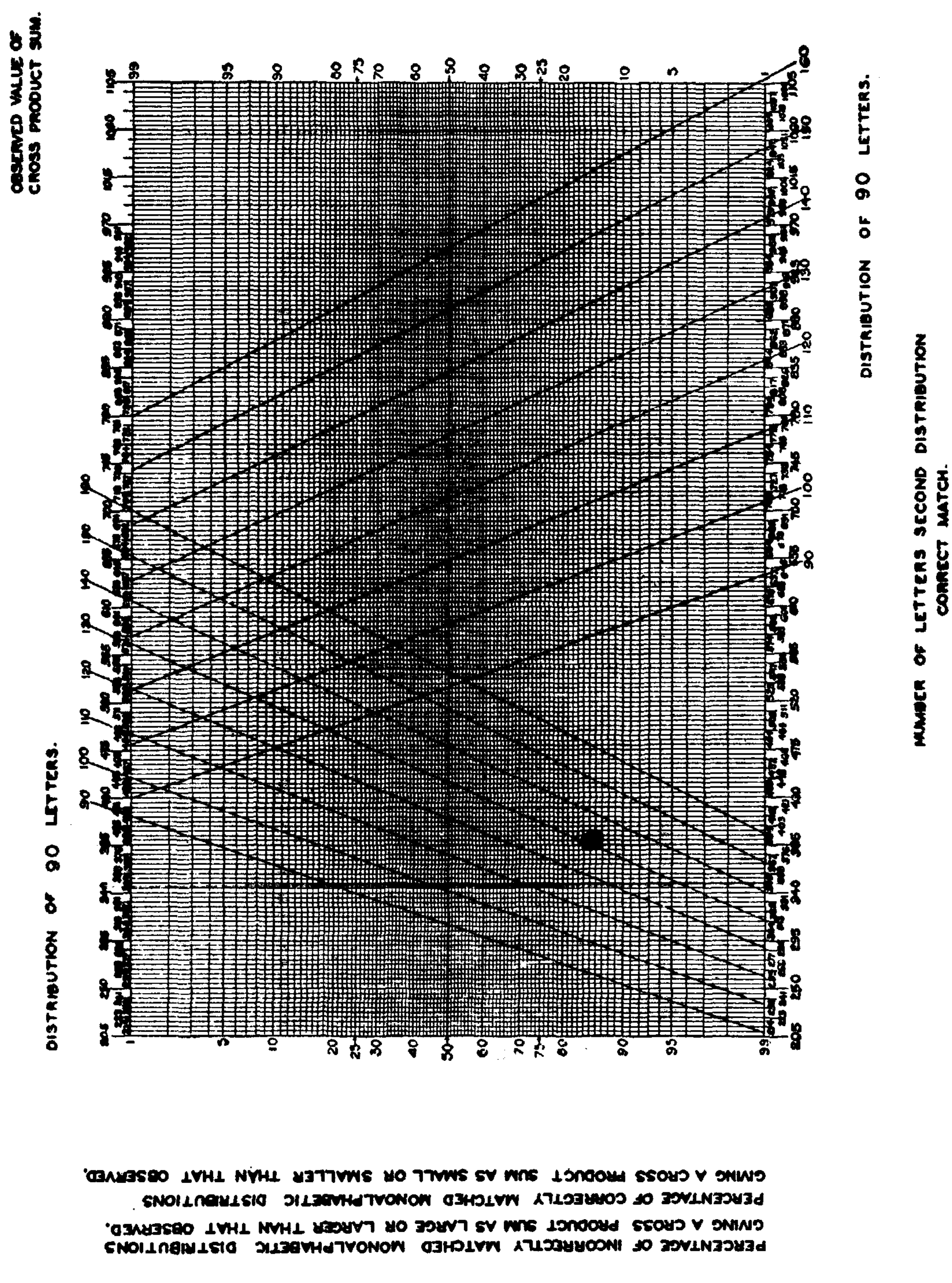


CHART No. 28

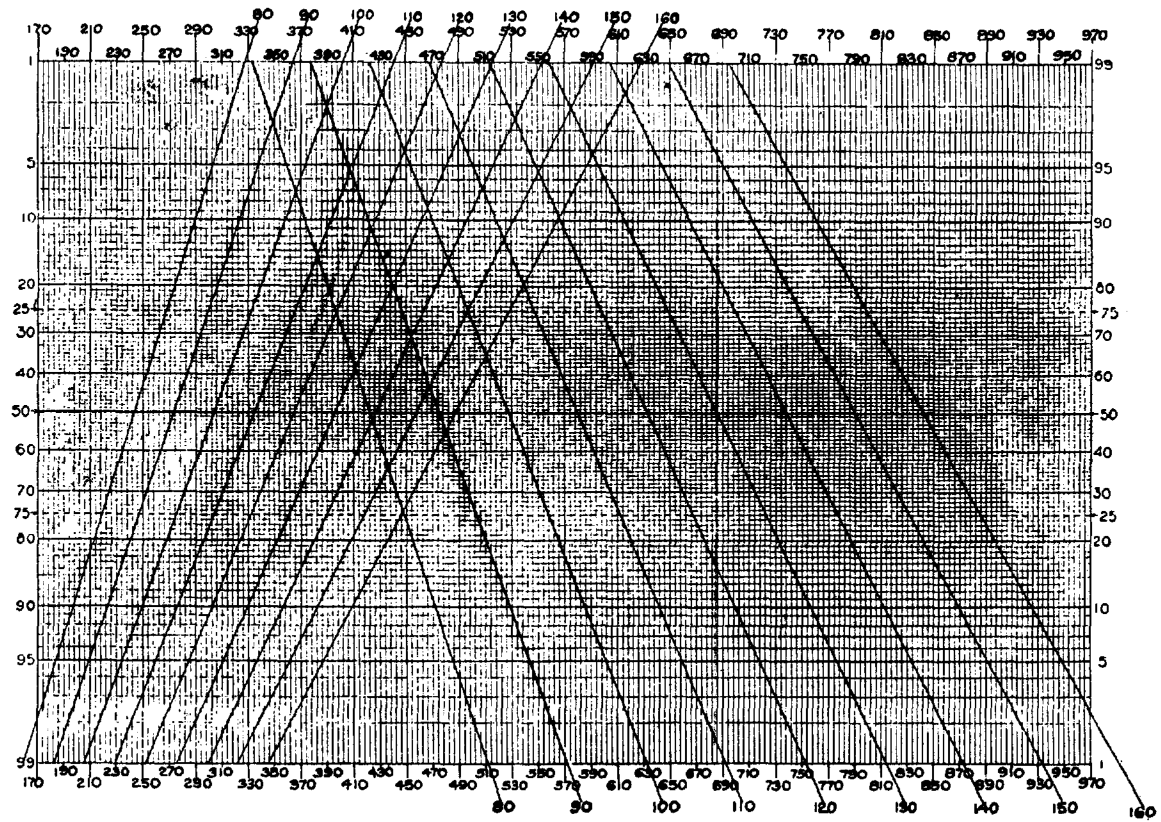
NUMBER OF LETTERS SECOND DISTRIBUTION
INCORRECT MATCH

OBSERVED VALUE OF
CROSS PRODUCT SUM

DISTRIBUTION OF 80 LETTERS.

PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.

PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.



DISTRIBUTION OF 80 LETTERS.

NUMBER OF LETTERS SECOND DISTRIBUTION
CORRECT MATCH.

CHART No. 29

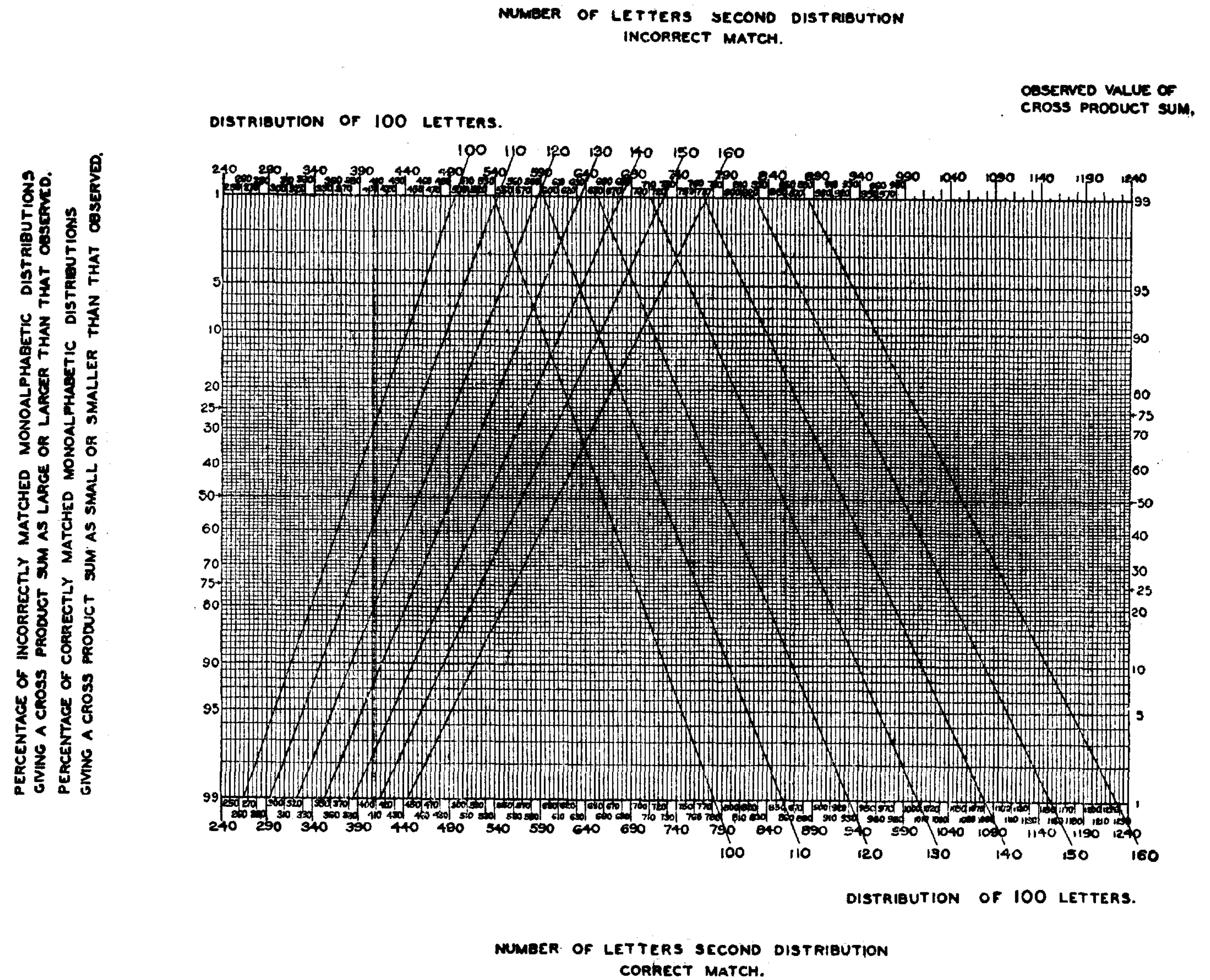
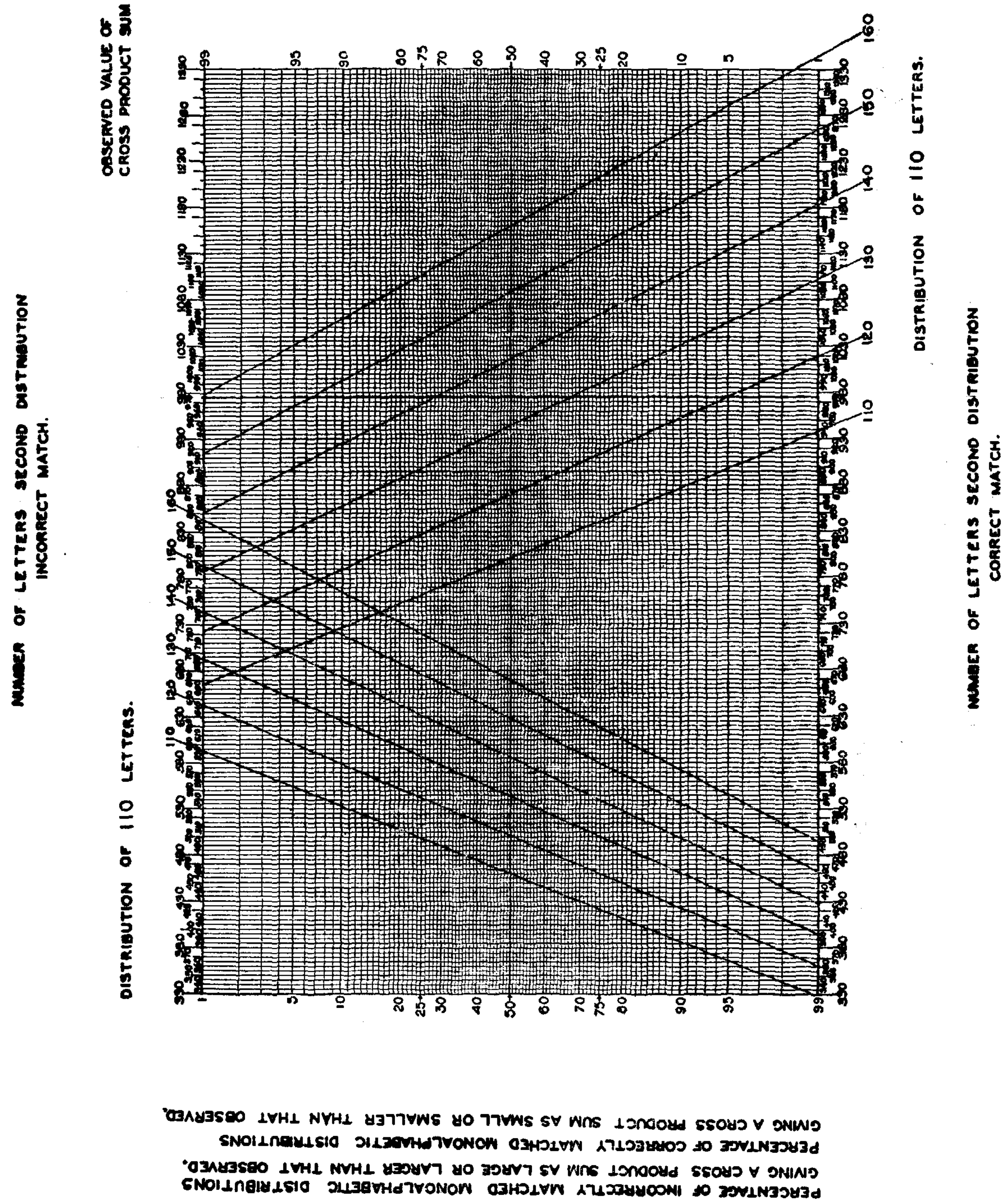

NUMBER OF LETTERS SECOND DISTRIBUTION
INCORRECT MATCH.
OBSERVED VALUE OF
CROSS PRODUCT SUM.
DISTRIBUTION OF 100 LETTERS.
100 110 120 130 140 150 160
240 290 340 390 440 490 540 590 640 690 740 790 840 890 940 990 1040 1090 1140 1190 1240
1 5 10 20 25 30 40 50 60 70 75 80 90 95 99
99 95 90 80 75 70 60 50 40 30 25 20 10 5 1
240 290 340 390 440 490 540 590 640 690 740 790 840 890 940 990 1040 1090 1140 1190 1240
100 110 120 130 140 150 160
DISTRIBUTION OF 100 LETTERS.
NUMBER OF LETTERS SECOND DISTRIBUTION
CORRECT MATCH.
PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS
GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.

CHART No. 30


NUMBER OF LETTERS SECOND DISTRIBUTION INCORRECT MATCH.
DISTRIBUTION OF 110 LETTERS.
OBSERVED VALUE OF CROSS PRODUCT SUM
DISTRIBUTION OF 110 LETTERS.
NUMBER OF LETTERS SECOND DISTRIBUTION CORRECT MATCH.
PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.

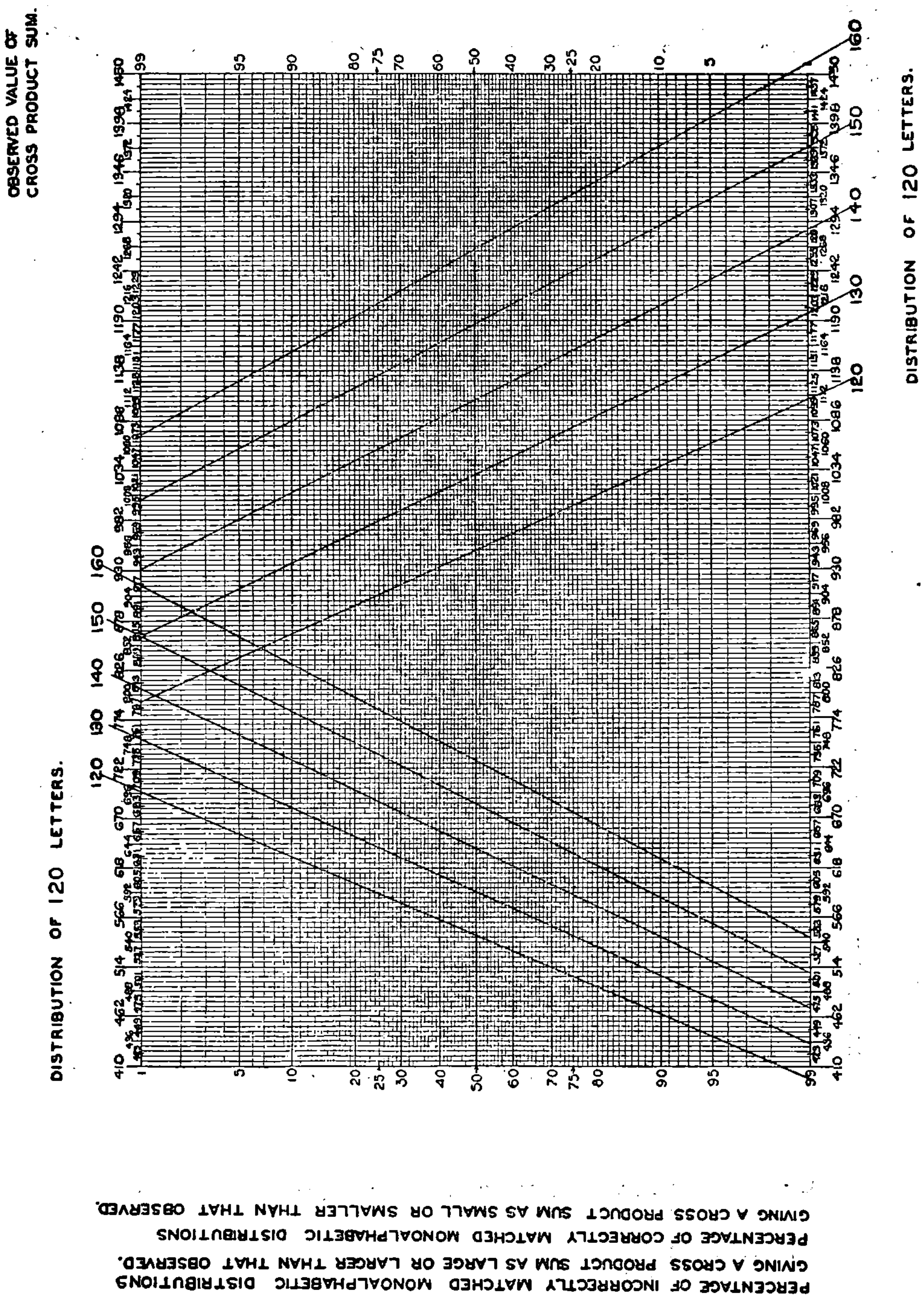


CHART No. 31

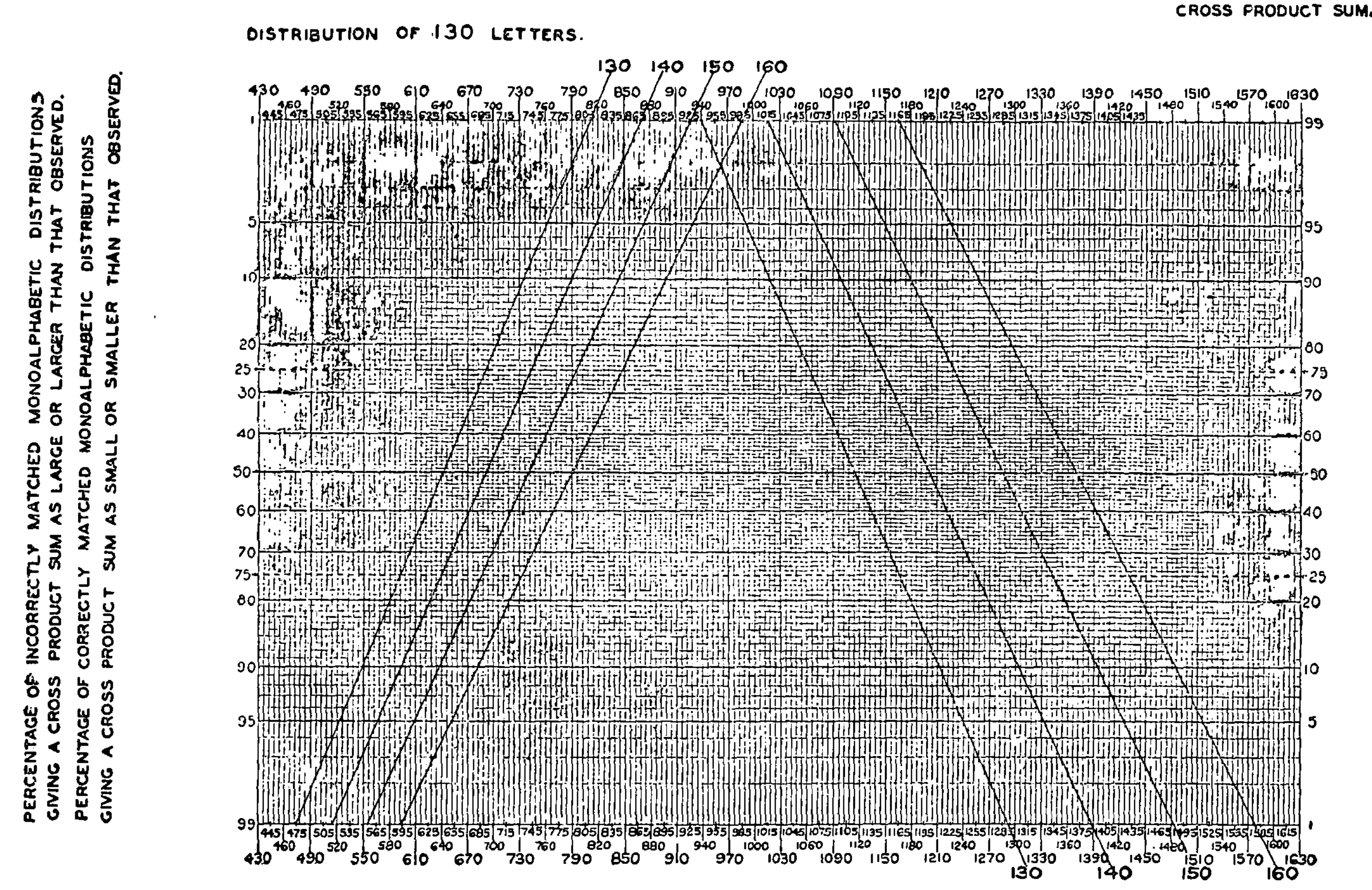
NUMBER OF LETTERS SECOND DISTRIBUTION INCORRECT MATCH.

NUMBER OF LETTERS SECOND DISTRIBUTION CORRECT MATCH.

63301—38——13

Chart No. 32

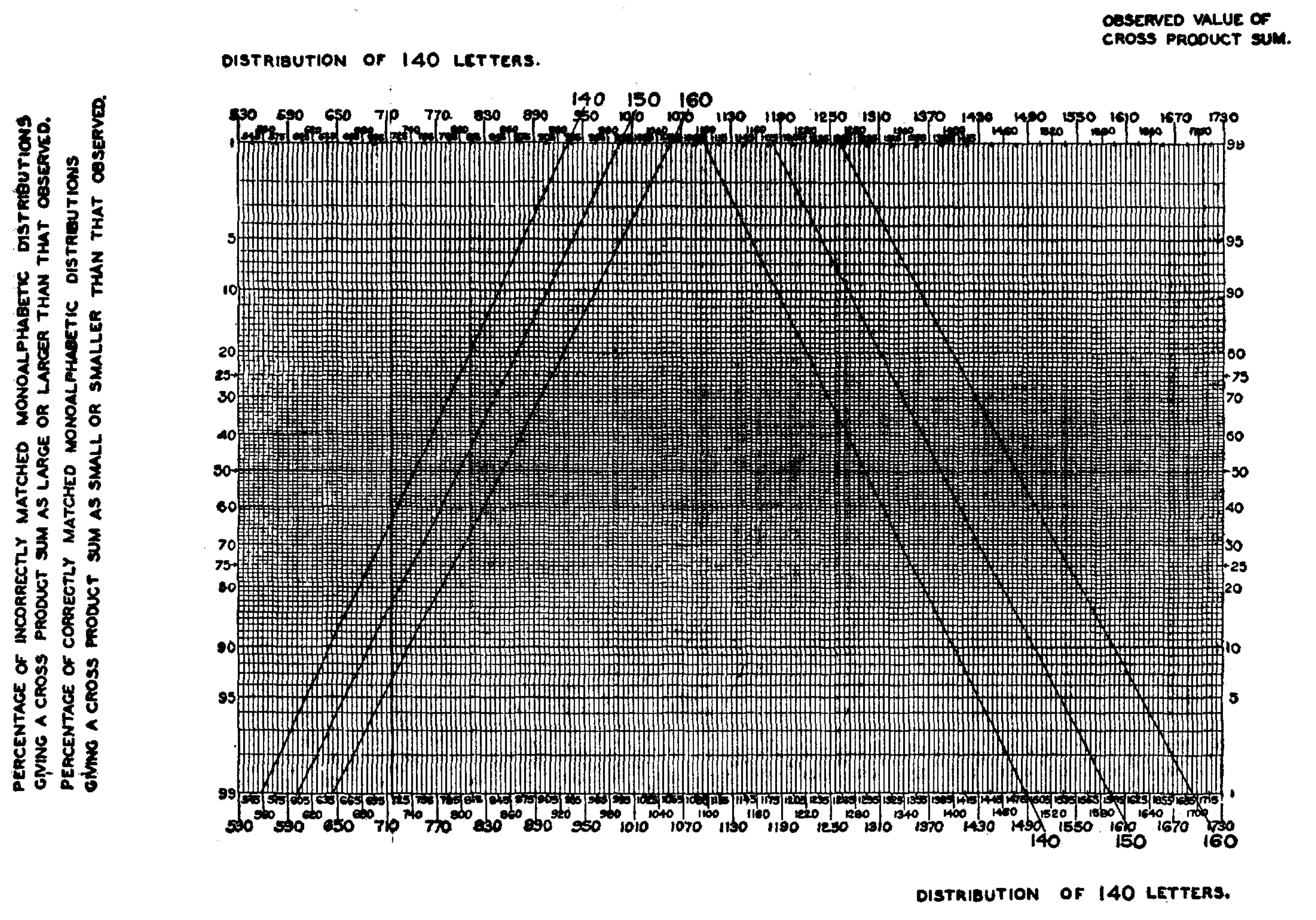
NUMBER OF LETTERS SECOND DISTRIBUTION
INCORRECT MATCH.



NUMBER OF LETTERS SECOND DISTRIBUTION
CORRECT MATCH.

CHART No. 33

NUMBER OF LETTERS SECOND DISTRIBUTION
INCORRECT MATCH.



NUMBER OF LETTERS SECOND DISTRIBUTION
CORRECT MATCH.

CHART No. 34

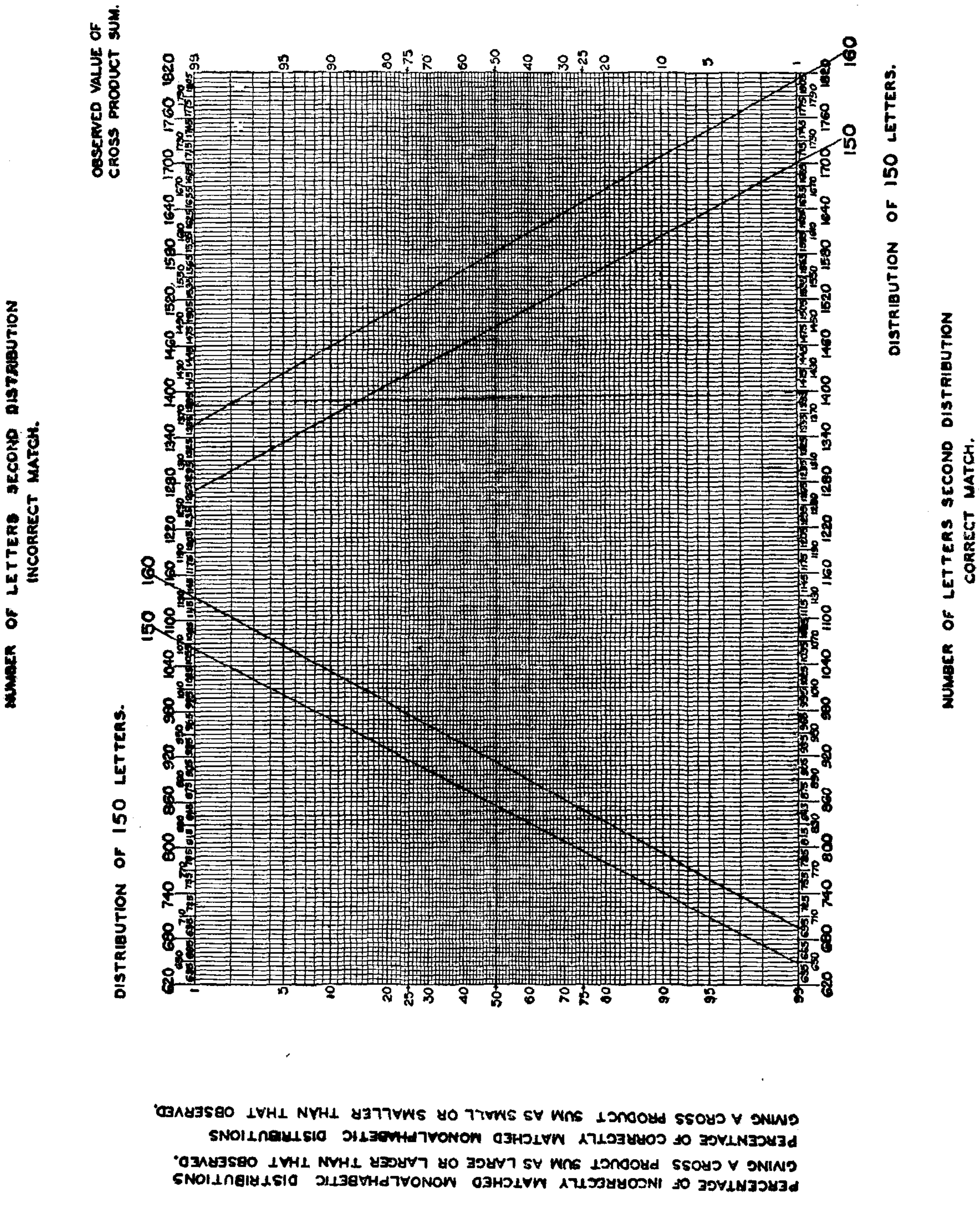
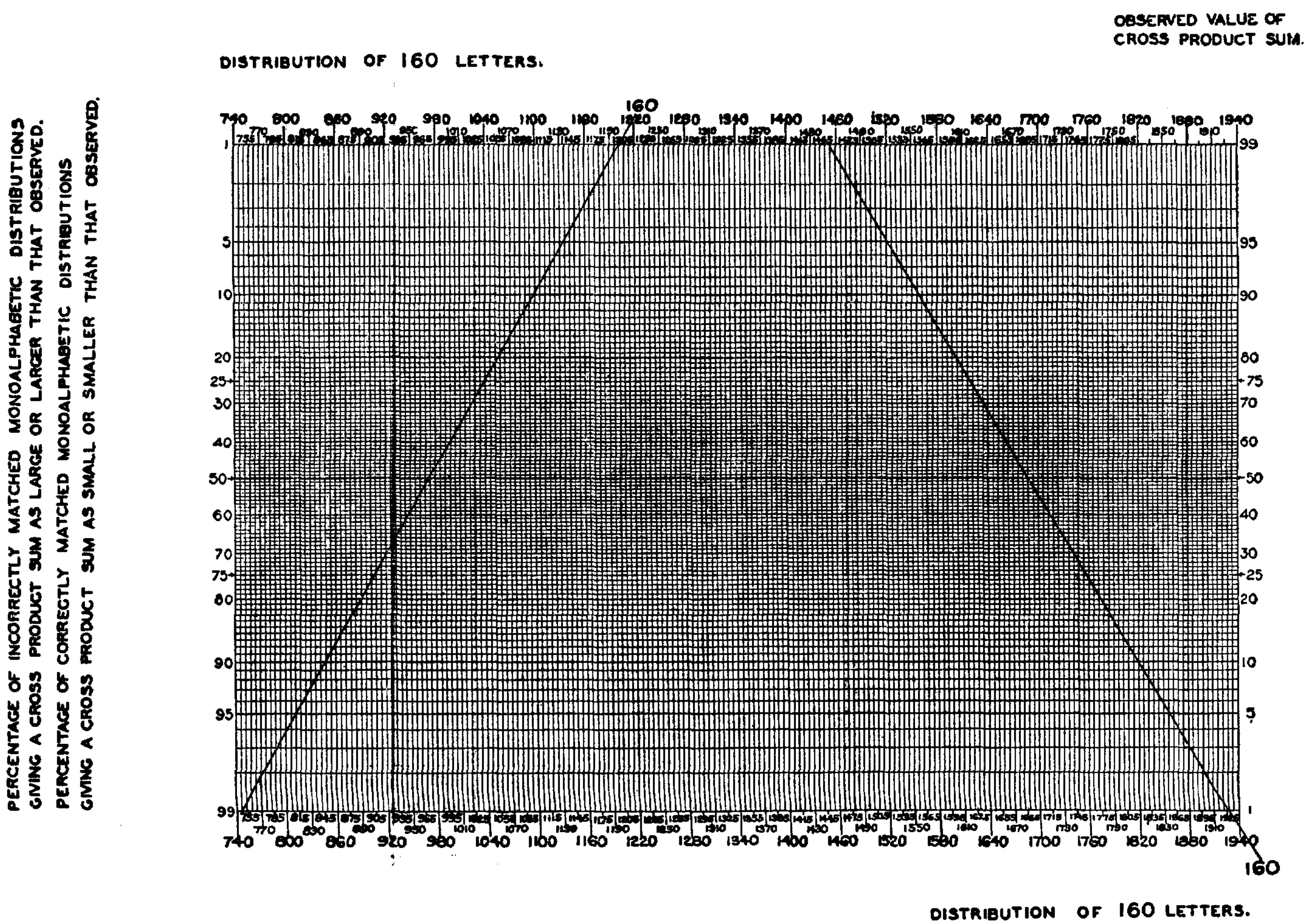

NUMBER OF LETTERS SECOND DISTRIBUTION INCORRECT MATCH.
OBSERVED VALUE OF CROSS PRODUCT SUM.
DISTRIBUTION OF 150 LETTERS.
DISTRIBUTION OF 150 LETTERS.
NUMBER OF LETTERS SECOND DISTRIBUTION CORRECT MATCH.
150
160
150
160
PERCENTAGE OF INCORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS LARGE OR LARGER THAN THAT OBSERVED.
PERCENTAGE OF CORRECTLY MATCHED MONOALPHABETIC DISTRIBUTIONS GIVING A CROSS PRODUCT SUM AS SMALL OR SMALLER THAN THAT OBSERVED.

CHART No. 35

NUMBER OF LETTERS SECOND DISTRIBUTION INCORRECT MATCH.



NUMBER OF LETTERS SECOND DISTRIBUTION CORRECT MATCH.

INDEX