Deviation of Random Samples from Average Conditions and Significance to Traffic Men

By E. C. MOLINA and R. P. CROWELL

THE traffic executive deals with questions which lead him into the consideration of problems of widely differing natures. At almost every turn he is confronted by the fact that his decisions and programs in relation to these different phases of the work must be based on records which are seldom continuous and in most cases are merely "samples." These sample records are assumed to measure the characteristics of the entire volume of facts or data of which they are taken to be representative. In the use and analysis of these records there are a number of perplexing questions which come to his mind if he allows himself the luxury of a little theoretical speculation.

Practically all of his information regarding the efficiency with which his office is run and on which he must base his plans for continued efficiency is obtained from the peg counts. These peg counts are records of the number of calls handled and are taken on two or three days out of each month. At the same time that the calls are counted, the number of employee hours used in the handling of the traffic is counted. The results of these peg counts are used to represent the performance of that office for the month. When the inquiring traffic man meditates a little on the subject of these peg counts he soon begins to wonder how nearly representative they are of his every day performance. He can—and sometimes does—think up a number of things which will explain any poor results which show up.

One of the means taken to insure the accuracy of the peg count is to observe the counting of 25 to 50 calls each by as many of the operators as possible, with the idea of determining how accurately the operators count. In this way from 1,500 to 3,000 observations are made on the accuracy of the operators' counting, in a period of two or three days. The traffic man occasionally questions whether he can rely on the results of this comparatively small number of checking observations to give him an indication of the accuracy of the count as a whole.

In order that comparisons may be made of the performance of different offices and the cost of handling different kinds of calls, it is the practice to translate all the work done into terms of traffic units (representing the relation of the labor value of the different operations to a fixed value arbitrarily selected). In order to do this, at longer intervals than the regular peg counts, the traffic is counted in more detail. From certain classifications and subdivisions of these supplementary counts, coefficients or equating factors are developed which are applied to the regular counts to develop units. The speculative traffic man ponders over these and wonders how representative the supplementary counts are of the every day distribution of traffic.

This speculation leads him also to question the labor values which have been assigned to the different operations and which have been furnished him for the purpose of equating his traffic. He knows that because of the impossibility of making continuous stop watch observations on his operators, he has to accept the results of such observations made on a considerable number of calls handled in a similar manner at some time in the past and probably in some other place, as being representative of the work involved in handling those types of calls at the present time in his office.

After thus puzzling himself over peg counts and similar records, the traffic man may turn his attention to some of the service problems and begins to scrutinize with considerable skepticism the records which are maintained of this feature of his work. Among the most valuable records of the way in which the service at his office is being handled, are the records developed as a part of the central office instruction routines. These are observations taken on ten calls handled by each of the operators on the force, periodically. He looks over the latest detail sheets and observes that the results of these tests on two particular operators show that the one he considered a very careful and methodical girl has made a high proportion of mistakes while the operator whom he thinks is the more careless shows an absolutely perfect test. Because of his other knowledge he suspects these records and decides to check them up by examining the summaries of similar tests taken for some months past. These summaries show figures which bear out his original estimate of the ability of the two operators, which relieves his mind but leaves him still puzzled as to why the averaging of a series of figures which are not representative, makes the summary more nearly representative.

There is another set of figures which the traffic man consults in connection with the quality of the service and which causes him a good deal of worry. These are the figures obtained from central office speed of answer tests, tests of the speed of answer to recall signals, etc. The speed of answer tests, for example, are made by an employee in the central office who causes signals to appear and with a stop watch determines how long it takes the operators to answer each signal. The signals used in making these tests are distributed in all parts of the switchboard and the number of tests made in each hour is roughly proportional to the amount of traffic handled. The results of these tests are summarized in such a manner as to show the percentage of tests which are not answered within 5, 10 and 20 seconds. The traffic man who gives this matter thought, is concerned to know how much reliance he can place on the results of these tests as being representative of the percentage of slow answers applying to all the calls handled in the office.

The speculative traffic man by this time is in a frame of mind which either leads him to doubt all figures or to feel that there must be something in the figures which he cannot explain but which makes certain of them quite representative, although there are certain others about which he does not feel the same way. He is sure that some of them are representative because decisions and programs based on them produce the results desired. He is also sure that some of them are not representative because they imply things which he knows are not so, as a result of observation. Just how far he can rely upon the figures which he is using, and where to draw the line is a question which only long experience or an understanding of the reasons which lie behind the taking of these records can solve. It will probably be of interest to discuss, from the purely theoretical angle, certain simple traffic data with the idea of noticing how the application of a certain mathematical procedure can aid in drawing accurate conclusions from them.

The type of traffic problem which will be considered may be stated as follows:

A group of 50,000 calls originated in an exchange area. An unknown number of them were delayed more than 10 seconds. Observations were made on 300 of the calls and of these 9, or 3 per cent., were delayed more than 10 seconds. With this information is it a safe bet that the unknown percentage for the entire 50,000 calls is below 5? Or better yet, are we justified in betting 99 in 100 that the unknown percentage for the 50,000 calls is below 5? Or again, may we bet 8 in 10 that the unknown percentage is between 0.5 and 5? It is taken for granted that the observer is justified in believing that the calls under consideration fulfill the conditions of random sampling such as that each call is independent of every other call, or that an appreciable number of the calls is not due to the occurrence of some unusual event,—the opening of the first game of the world series, for example.

Assuming that the reader is unfamiliar with the theory of probability, a digression becomes necessary and in order that he may enter into the spirit of the theory the reader is requested to forget for the present the telephone problem. Of course, only a bird's-eye view of the theory will be given here. Several lacunæ will be encountered; the filling in of any one of them would call for a volume of not very small dimensions.

Introduction to the Theory of "A Posteriori" Probability

The problem to be dealt with belongs to the class of problems which gave rise to that branch of the Theory of Probability which is known as "A Posteriori Probability" or "Probability of Causes." It is frequently referred to as the Theory of Sampling.

To bring out certain of the ideas involved it will be helpful to consider what may appear as a very extreme example from the traffic man's point of view, but which is nevertheless typical of the type of problem in which a consideration of a posteriori probability enters. We are told that at a student gathering a particular young man won 7 out of 15 times. Our informant refuses to divulge what is going on at the gathering. What probabilities should we assign to the following hypotheses?

1. He threw heads 7 times out of 15 throws with a coin.
2. He threw 7 aces out of 15 throws with a 6 face die.
3. He won on points 7 rounds in a fifteen round bout.
4. The aggregate of all other hypotheses.

A little careful consideration will make it clear that with reference to each hypothesis (or aggregate of hypotheses) two essential questions must be answered before we can determine the a posteriori probability. Consider the six face die hypothesis; we must know:

1st —What is the relative frequency or probability with which gambling with a 6 face die is indulged in at student gatherings?

2nd—Given a six face die, what is the probability of throwing an ace 7 times in 15 throws?

Quoting Mr. Arne Fisher[1] we may restate these two questions as follows:

1st —What is the a priori *existence* probability in favor of the 6 face die hypothesis?

2nd—What is the *productive* probability for the observed event given by the hypothesis of a 6 face die?

[1] Arne Fisher—The Mathematical Theory of Probabilities—2nd Edition—Art. 41.

In most problems of this type the determination of the *productive* probability for each hypothesis is a question of pure mathematics. But when we proceed to evaluate the a priori *existence* probability for each hypothesis or cause, common sense and guessing must frequently be resorted to. The history of the applications of a posteriori probability is so full of paradoxes resulting from appeals to common sense that to some high authorities the whole theory is a fallacy. Prof. George Chrystal [2] closes a severe attack on Laplace's Theorie Analytique with the statement—"The indiscretions of great men should be quietly allowed to be forgotten." Nevertheless, the writers will assume the Laplacian view of the subject, especially as it has been defended by such authorities as Karl Pearson and E. T. Whittaker.

The above typical problem has been introduced because its mere statement leads us immediately to the conceptions of existence and productive probabilities with reference to different possible hypotheses. But, it is not our intention to bring any notoriety on the young man by answering the questions raised. Moreover, the hypotheses made, differ qualitatively, whereas, our telephone problem involves various hypotheses which differ only quantitatively. We, therefore, proceed to another typical problem, a solution of which will give us at once the solution of the telephone problem.

A bag contains 1,000 balls; an unknown number of these are white and the rest not white. Of 100 balls drawn 7 are found to be white. What light does this information throw on the value of the unknown number of white balls? What is the probability that there are 70 white? Is it a safe bet that the number of white balls lies between 60 and 80?

Two cases of this problem may be considered:

Case 1. After a ball is drawn it is replaced and the bag is shaken thoroughly before the next drawing is made.

Case 2. A drawn ball is not replaced before another ball is drawn.

These two cases become essentially identical if the total number of balls in the bag is very large compared with the number drawn.[3] In the following discussion Case 1 is assumed.

The information at hand is that 100 drawings resulted in 7 whites. Obviously the bag contains at least one white, but we are free to choose between 999 possible hypotheses.

[2] Transactions of the Acturial Society of Edinburgh—Vol. II, No. 13—On Some Fundamental Principles in the Theory of Probabilities.

[3] For the application to practice herein contemplated it is thought that the number of balls in the bag should be at least ten times the number drawn.

1—The bag contains 1 white and 999 not white.
2—The bag contains 2 white and 998 not white.
3—The bag contains 3 white and 997 not white.
....
K—The bag contains K white and (1,000-K) not white.
....
997—The bag contains 997 white and 3 not white.
998—The bag contains 998 white and 2 not white.
999—The bag contains 999 white and 1 not white.

Let $W(K)$ be the existence probability for the K'th hypothesis. By "existence probability" is meant the likelihood that the bag contains exactly K white balls when the circumstances of the drawing, but not the actual results of the drawing, are fully taken into account. Its exact value may often be in doubt either because we do not have complete knowledge of the circumstances preceding the drawing or because we are not able to deduce its exact value from this knowledge. It is obvious, however, that there must be some such value and we must, therefore, introduce a symbol to represent it.

Let $B(7,100,K)$ = productive probability for the K'th hypothesis; by this is meant the probability of obtaining the observed event (7 white in 100 drawings) if the bag contains K white balls and 1,000-K that are not white.

Then the a posteriori probability in favor of the K'th hypothesis (meaning thereby the probability in favor of the K'th hypothesis *after* the 7 white balls were drawn) is [4]

$$P_k = \frac{W(K)B(7,100,\,K)}{\sum_{s=1}^{s=999} W(S)B(7,100,\,S)}. \tag{1}$$

Now to say that the bag with a total of 1,000 balls contains K white balls is equivalent to saying that the *ratio* of white to total balls is

$$p_k = K/1000$$

and that the *ratio* of not white to total balls is

$$q_k = 1 - p_k = (1000 - K)/1000.$$

[4] This is the celebrated Laplacian generalization of Bayes' formula. No attempt to demonstrate it will be made here. The subject is dealt with at length by Laplace in the Théorie Analytique des Probabilités and by Poisson in the Recherches Sur La Probabilité des Jugements. A beautiful and relatively short demonstration is given by Poincaré in his Calcul des Probabilités.

We may, therefore, rewrite (1) as follows:

$$P_k = \frac{W'(p_k)B'(7,100,\ p_k)}{\sum_{s=1}^{s=999} W'(p_s)B'(7,100,\ p_s)}, \tag{2}$$

where W', B' are the forms assumed by the functions W, B, respectively, when the ratio p_k is used instead of the number K.

The interpretation of the terms of the expansion of the binomial $(p+q)^{100}$ tells us that

$$B'(7,100,\ p) = \binom{100}{7} p^7(1-\mathrm{p})^{93} = \binom{100}{7} p^7 q^{93}$$

where $\binom{100}{7}$ is a symbol for the number of combinations of 100 things 7 at a time.

Substituting in (2) and canceling from numerator and denominator the common factor $\binom{100}{7}$ gives

$$P_k = \frac{W'(p_k)p_k{}^7(1-p_k)^{93}}{\sum_1^{999} W'(p_s)p_s{}^7(1-p_s)^{93}}. \tag{3}$$

From (3) we obtain for the a posteriori probability that the ratio of white balls does not exceed $K_2/1{,}000$,

$$P(K \not> K_2) = \sum_1^{K_2} P_k.$$

Likewise, the a posteriori probability that the ratio is not less than $K_1/1{,}000$ is

$$P(K \not< K_1) = \sum_{K_1}^{999} P_k.$$

Finally, the a posteriori probability that the ratio is not less than $K_1/1{,}000$ or greater than $K_2/1{,}000$ is

$$P(K_1 \leq K \leq K_2) = \sum_{K_1}^{K_2} P_k = \frac{\sum_{K_1}^{K_2} W'(p_s)p_s{}^7(1-p_s)^{93}}{\sum_1^{999} W'(p_s)p_s{}^7(1-p_s)^{93}}. \tag{4}$$

Solution of the Telephone Problem

Obviously the telephone problem is analogous to the problem of the bag containing an unknown ratio of white balls. The corresponding elements in the two problems may be tabulated as follows:

1st—1,000 balls in bag versus 50,000 calls originated.

2nd—100 balls drawn versus 300 calls observed.

3rd—7 white balls drawn versus 9 calls delayed more than 10 seconds (*i.e.*, defective with reference to a particular characteristic).

4th—To the 999 possible hypotheses with reference to the unknown per cent. of white balls correspond 49,999 possible hypotheses with reference to the unknown per cent. of calls delayed more than 10 seconds.

The problems differ in that a ball drawn from the bag is returned before another drawing is made, whereas an observed call is comparable to a ball being drawn and not returned. With the numbers involved, however, the discrepancy may be ignored.

A formula of the same form as (4) will, therefore, give the answer to our question. We may, however, substitute definite integrals in place of the finite summations since the difference between any two consecutive possible values for the unknown ratio is very small. The integrals together with some desirable transformations of them will be found in the appendix to this article. We will mention here, however, that the transformations made involve an arbitrary assumption as to how the a priori *existence* probability for the different hypotheses varies. As stated above in connection with Prof. Chrystal's views, this is the phase of the subject which lends itself to considerable difference of opinion. The reader who contemplates using the curves embodied in this article should read the appendix with special reference to the assumptions made.

The attached curves Fig. 1 show graphically the conclusions to be drawn from the mathematical analysis. A glance at the right hand end of the curves will show that they are associated in pairs. The upper curve of a pair slopes downward from left to right while its mate slopes upward.

Consider the pair of curves marked .03. For the abscissa 300 they give as ordinates the values .0625 and .014. The interpretation of these figures is as follows: if 300 observations gave 3 per cent. of calls delayed then we may bet

1st—99 in 100 that the unknown percentage of calls delayed is *not greater* than 6.25.

2nd—99 in 100 that it is *not less* than 1.4 per cent.
3rd—98 in 100 that it lies between 1.4 per cent. and 6.25 per cent.

Likewise, considering the curves marked .06 if 1,000 observations gave 6 per cent. of calls delayed, then we may bet

1st—99 in 100 that the unknown percentage of calls delayed is not greater than 8.05.
2nd—99 in 100 that it is not less than 4.4 per cent.
3rd—98 in 100 that it lies between 4.4 per cent. and 8.05 per cent.

It is obvious from the shape of the curves that a few hundred observations do not give more than a vague idea as to the unknown per cent. of calls delayed. On the other hand, the gain in accuracy obtained by making more than 10,000 observations would hardly justify the expense involved. The number of observations which safety requires in any particular problem must be determined by the conditions of the problem itself. If we are willing to take a chance of 9 in 10 or 8 in 10 instead of 99 in 100 or 98 in 100, respectively, the curves of Fig. 2 will give us an idea of the range within which the unknown percentage of defectives lies.

APPENDIX

Case No. 1—Infinite Source of Samples

An inspection of n samples has given c defectives. The observed frequency is then c/n. Let p be the unknown true frequency and p_1 the frequency of delayed calls which has been arbitrarily chosen as being the maximum permissible.

The a posteriori probability that $p \ngtr p_1$ is

$$P=\frac{\int_o^{p_1} W(x)x^c(1-x)^{n-c}\,dx}{\int_o^1 W(x)x^c(1-x)^{n-c}\,dx}, \tag{1}$$

where $W(x)$ is the a priori existence probability that $p=x$. This formula is unmanageable if the form of $W(x)$ is unknown.

Assume first that $W(x)$ is a constant b for $o<x<g$, where $g>p_1$. Then

$$P=\frac{\int_o^{p_1} x^c\,(1-x)^{n-c}dx}{\int_o^g x^c\,(1-x)^{n-c}\,dx+\int_g \frac{W(x)}{b}\,x^c(1-x)^{n-c}dx}. \tag{2}$$

Now assume that

$$\int_g^1 \frac{W(x)}{b} x^c(1-x)^{n-c}dx,$$

is negligible compared with

$$\int_o^g x^c(1-x)^{n-c}dx,$$

and also assume that g, c and $(n-c)$ are such that approximately

$$\int_o^g x^c(1-x)^{n-c}dx = \int_o^1 x^c(1-x)^{n-c}dx.$$

Then, finally,

$$P = \frac{\int_o^{p_1} x^c(1-x)^{n-c}dx}{\int_o^1 x^c(1-x)^{n-c}dx} = \frac{(n+1)!}{c!(n-c)!}\int_o^{p_1} x^c(1-x)^{n-c}dx, \tag{3}$$

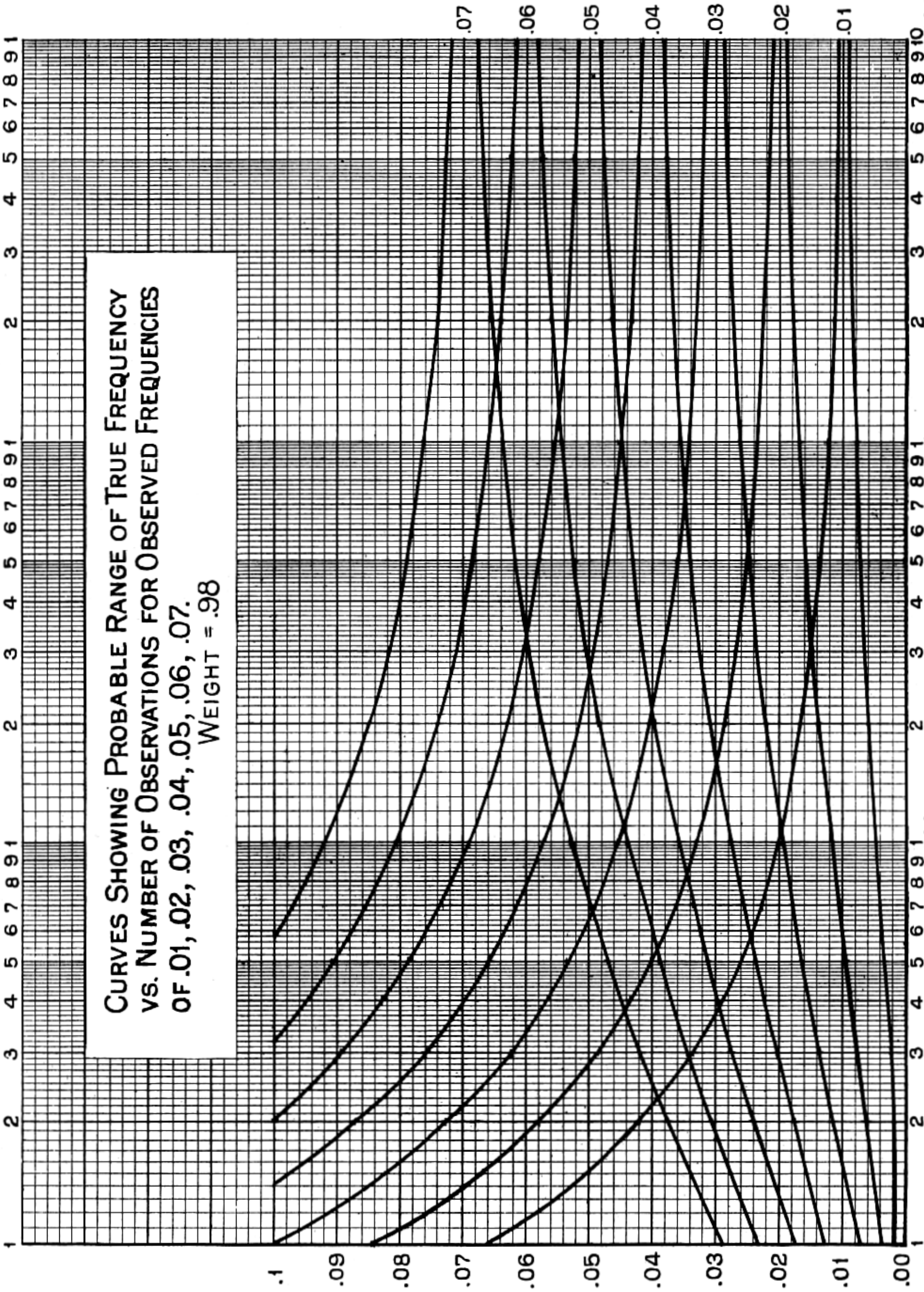
This well known formula might have been obtained by assuming *ab initio* that $W(x)$ is independent of x. It should be particularly noted that this independence is not identical with the assumptions made above. In the applications which are here contemplated the values of p_1, c and n are such that g need be but a small fraction of the range o to 1.

In the "Théorie Analytique" Laplace transforms (3) so that it can be evaluated in terms of the Laplace-Bernoulli integral

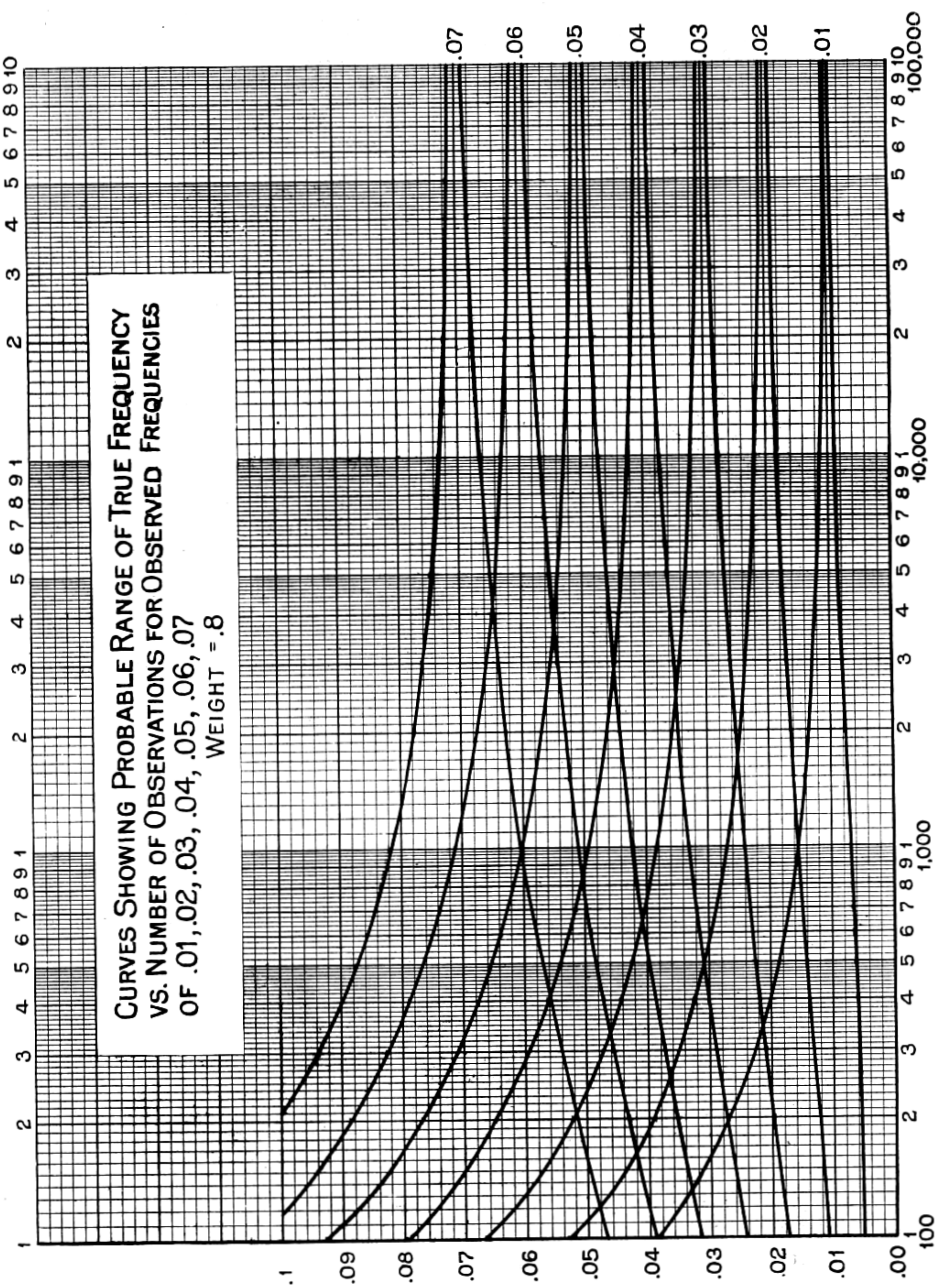
$$\frac{2}{\sqrt{\pi}}\int_o^k e^{-t^2}dt,$$

where k is a function of p_1, c and n. This transformation is most valuable when p_1 is in the neighborhood of 1/2. For small values of p_1 the transformation which converts the binomial expansion to Poisson's exponential binomial limit is more appropriate and gives, writing $(n\,p_1) = a_1$,

$$P = \frac{1}{c!}\int^{a_1} y^c e^{-y}dy = P(c+1, a_1). \tag{4}$$



Curves Showing Probable Range of True Frequency
vs. Number of Observations for Observed Frequencies
of .01, .02, .03, .04, .05, .06, .07.
Weight = .98
100
1,000
10,000
100,000
.1
.09
.08
.07
.06
.05
.04
.03
.02
.01
.00



CURVES SHOWING PROBABLE RANGE OF TRUE FREQUENCY
VS. NUMBER OF OBSERVATIONS FOR OBSERVED FREQUENCIES
OF .01, .02, .03, .04, .05, .06, .07
WEIGHT = .8
.07
.06
.05
.04
.03
.02
.01
.1
.09
.08
.07
.06
.05
.04
.03
.02
.01
.00
100
1,000
10,000
100,000