The Problem of Inductive Inference*

BY JERZY NEYMAN
Statistical Laboratory, University of California, Berkeley

PART I. GENERAL IDEAS

1. Introduction

The words "inductive inference" in the title of the subject on which I was invited to speak, are used as a "prescientific concept" (Carnap) referring to all attempts at practical utilization of the results of observation. The nature of inductive inference, the possible ways of making the concept precise and of constructing its consistent theory have been the subject of a number of penetrating studies. The two latest are due to Rudolf Carnap [1] and R. B. Braithwaite [2]. Both authors devote a substantial amount of space to the ideas that prevail in most of the current statistical literature and, in particular, to the logical system which originated a quarter of a century ago out of my joint work with Egon S. Pearson. It is not unnatural that some of our ideas are presented by the two authors in a manner different from our own, at least different from the one that I would use now. In some cases, and this applies particularly to Professor Carnap's work, the differences are quite substantial and suggest that my previous writings must have been insufficiently clear. As a result, when Professor Carnap criticizes some attitudes which he represents as consistent with my point of view, I readily join him in his criticism without, however, accepting the responsibility for the criticized paragraphs.

The purpose of the present paper is, then, to make another attempt at a consistent presentation of the system of ideas that underly the sections of the theory of statistics for which I am at least partly responsible. In the second part of the paper these ideas are illustrated on the particular problem of homogeneity of neutral V-particles.

The point of view presented is substantially closer to that of Braithwaite than to that of Carnap. For this reason, and for the sake of brevity, most of the references given in the text will be addressed to Carnap's work, particularly to the points on which there is disagreement. An additional reason for this selection of material is that real progress in clarifying a situation is more frequently achieved by constructively discussing the existing differences of opinion rather than by emphasizing points on which there is a complete harmony.

*This paper was prepared with the support of the Office of Ordnance Research, U.S. Army under Contract DA-04-200-ORD-171.
2. Common Basis for Discussion

Non-dogmatic scientific discussion is possible only if the discussants agree on a common premise. In the present case, a suitable platform on which there appears to be complete agreement is described by Carnap both at the beginning and at the very end of his book. This is to the effect that a theory originates from "a body of generally accepted but more or less vague beliefs" expressed in terms of "inexact concepts." "We call the given concept (or the term used for it) the explicandum, and the exact concept proposed to take the place of the first (or the term proposed for it) the explicatum ... . The explicatum must be given by explicit rules for its use, for example, by a definition which incorporates it into a well constructed system of scientific either logicomathematical or empirical concepts." These general remarks are illustrated by examples of which we quote the following:

"For example, Euclid's axiom system of geometry was a rational reconstruction of the beliefs concerning spatial relations ... ."

To the present writer, the above passages illustrate the opinion that the general domain of human thought may be usefully divided into two broad spheres, that of prescientific inexact concepts and vague beliefs, the sphere of "explicanda" on the one hand, and that of "exact concepts" or "explicata" on the other hand. The second sphere, the contents of which constitute science, originates from the first but is essentially different. Clarity of discussion requires a sharp delineation between these two spheres of thought and continuous vigilance against confusion. For the sake of ease in reference the first sphere of thought will be called "phenomenal" and the second "conceptual."

3. Apparent Source of Disagreement with Rudolf Carnap

Such differences as exist between Professor Carnap's theory and the ideas of the present writer appear to stem out of an inequality in the emphasis we place on the requirement that the two spheres of thought, the phenomenal and the conceptual be carefully distinguished. This difference in emphasis may be illustrated by passages in Carnap's paragraph 42 in which he discusses the various possible ways of interpreting statements about probabilities.

"Suppose X asserts the following statement: (13) 'The probability of throwing an ace with this die is 0.15.'

"We want to determine in which sense this statement is meant by X. Here, as often, it is not advisable to ask direct questions like (i) 'What do you mean?' or (ii) 'What meaning does the word 'probability' have for you?' We ask instead: (iii) 'What is the basis for your assertion?' (iv) 'What observations led you to the value stated?' The frequentists emphasize the fact ... therefore, in order to fit our example to this conception, let us assume that X answers as follows: (14) 'I have made 1,000 throws with this die, of which 150 yielded an ace; no other results of throws with this die are known to me.'
"The frequentists will be inclined to take this answer as ... ."

Thereafter, a few more explanations, similar to (14), are suggested as possible answers of a "frequentist," eventually leading to

"'(19) 'There is a high probability, with respect to the evidence (14) for the prediction that the relative frequency of aces in a long series of future throws with this die will lie within an interval around 0.15 . . .'"

"Both (19) and (20) are statements of inductive logic."

If I am allowed to speak for the "frequentist," I would explain that every answer attributed by Carnap to the individual X is based on a confusion of the phenomenal and the conceptual spheres of thought. Without such confusion, and assuming that the word "probability" in question (13) represents the "explicatum" and not the vague "explicandum," the proper answers to question (iii) appear to be

A(iii) Either "because this value of the probability is a datum in a problem assigned to me" or "because the value 0.15 was obtained as a solution of a probabilistic problem assigned to me with numerical conditions."

To illustrate the point it may be useful to refer to the analogy with Euclidean geometry. Speaking in terms of this geometry (as distinguished from surveying) the only basis for the assertion that the distance between two points is equal to unity (or to any other specified number) is either that this number is included in data for a problem or that it is an outcome of calculations within the framework of Euclidean geometry, based on some other numerical data in a specified problem. For example, this problem may be:

Datum: (a) In a triangle ABC the sides AB and AC are equal; (b) the angle BAC is 90 degrees; (c) the length of the side BC is equal to the square root of 2.

Question: What is the common length of AB and AC?

Answer: Unity.

As to the question (iv) the proper answer appears to be:

A(iv). So long as we deal with the exact concept of probability, no observations are capable of producing the value of a given probability.

Further, always trying to speak for a hypothetical "frequentist," I would attempt to answer the questions (i) and (ii).

A(i). By the brief statement (13) relating to a real, phenomenal die, I mean that, when faced with a practical problem of a series of trials including one or more throws with this die (there may be other trials) I choose to treat this problem on the basis of the theory of probability (based on Kolmogoroff's system of axioms) with the value of the probability of throwing an ace equal to 0.15. In other words, I hope that the outcome of my computations so conducted
will be approximately comparable to the outcome of the real experiments. In still other words, I expect that in actual throws with the die concerned, the frequencies of various outcomes will behave, approximately, as implied by the theory.

It may be useful to compare this answer with the one that would be forthcoming to an analogous question relating to Euclidean geometry combined with Newtonian mechanics.

Suppose \( X \) is engaged in cosmological studies and asserts that the "Euclidean-Newtonian" distance \( D \) between the sun and the center of the cluster of galaxies in the constellation of Boötes is 450 million light years. The proper interpretation of this assertion is that, faced with some particular problem involving \( D \), the individual \( X \) intends to use Euclidean geometry and Newtonian mechanics and to base his calculations on \( D = 450 \times 10^6 \), and hopes that the outcomes of these calculations will not be too inconsistent with others based on similar premises.

A(ii). In answering question (ii), always assuming that the word "probability" is used as an exact concept, I would refer to the axioms of Kolmogoroff and/or to the relevant discussion of Braithwaite. In either case the exact concept of probability will appear inapplicable to any specified phenomenal die.

It will be noticed that, if instead of referring to probability of throwing an ace, assertion (13) referred to the estimate of this probability, then the answers to all the questions discussed would have been entirely different.

4. Theory of Inductive Inference Belongs to Conceptual Sphere of Thought

If one insists on consistency and on discriminating between the phenomenal and the conceptual spheres of thought, then every theory of "inductive inference" or, in order to avoid the use of a term that may lead to misunderstandings, every theory of utilization of observational data must be a system entirely included within the conceptual sphere of thought. Thus, prior to any discussion based on a theory of inductive inference and having some reference to phenomena, a theorizing step must be taken in order to substitute for the phenomena an appropriate conceptual counterpart. In other words, a strict application of any theory of inductive inference can be made only on the ground of a theoretical model of some phenomena, not on the ground of the phenomena themselves.

The phenomena selected for this purpose may be of different kinds. One possibility is, and this is used in my work [3, 4], to restrict the considerations to the frequently observed phenomena of stability of relative frequencies of results of repeated "trials." In this case the theory of probability based on the axioms of Kolmogoroff provides a convenient mathematical model. Another possibility, used by R. Carnap [1], B. O. Koopman [5], Bruno de Finetti [6], and L. J. Savage [7], is that of psychological phenomena connected with intensities of belief, or readiness to bet specified sums, etc. There may be others.
PROBLEM OF INDUCTIVE INFERENCE

However, whatever the choice of the phenomena, the conclusions of a consistent theory of inductive inference will always be applicable within the mathematical models of these phenomena and not within the domains of the phenomena themselves. Since, in many instances, the phenomena rather than their models are the subject of scientific interest, the transfer to the phenomena of an inductive inference reached within the model must be something like this: granting that the model M of phenomena P is adequate (or valid, or satisfactory, etc.) the conclusion reached within M applies to P.

5. The Theory of Statistics as the "Frequentist's" Theory of Inductive Inference

The last half century has been marked by a swing of all sciences towards indeterminism. Briefly, this tendency may be summarized as follows.

With many phenomena it appears difficult, if not impossible, to construct models in which all the particular quantities intervene as single valued functions of some specified arguments. That is, it is difficult to construct models of this kind that are anywhere near comparable to the phenomena in question. Furthermore, in many cases, instead of knowing the exact value of X characterizing a phenomenon in given circumstances, it appears sufficient to know how frequently a phenomenon of this kind is characterized by X = 1, how frequently it is characterized by X = 2, etc. For this reason, instead of continuing their efforts to construct models (deterministic models) in which all the quantities like X are treated as sure functions of certain arguments, scientists have embarked on a multitude of models (stochastic or indeterministic models) in which at least some of the intervening quantities are treated as random variables. Many of the models so constructed appear to be in excellent agreement with the actual phenomena and, in addition, offer a great deal of esthetic satisfaction.

The frequentist's theory of inductive inference, called either the theory of inductive behavior, or the theory of statistical decision functions, or, more simply, the modern theory of statistics, is meant for application in all those cases where a stochastic model has been adopted to represent a given class of phenomena. Problems of this theory are many and of varying difficulty and importance. In the present section a single example must suffice.

Use the letter P to denote phenomena of a certain category and the letter M to denote the stochastic model adopted to represent P. Assume that certain practical actions connected with P depend upon the value of an unknown number θ characterizing P. For example, P may denote the phenomenon of (i) the bacterial contamination of the domestic water supply in a city and (ii) of testing this water by bacteriologists. The unknown θ may stand for the average number of given bacteria per unit volume of water. If the value of θ is small, the water is fit for domestic use. Otherwise, the water must be subjected to decontamination (e.g. by adding chlorine) the indicated intensity of which must be adjusted to the value of θ.

The model M may postulate that at any given moment the bacteria are
Poisson-distributed in the water so that the number $X$ of bacteria in a sample to be taken from this water (sample of unit volume) is a random variable following the familiar distribution of Poisson.

Problems of modern statistical theory arise from the concurrence of the following conditions:

(a) A practical situation forces the selection of an action $a$ regarding the phenomena $P$.

(b) This action is to be selected out of a given set $A$ of possible actions.

(c) The relative desirability of any given action in the set $A$ depends on the value of a number $\theta$ characterizing $P$.

(d) This number $\theta$ is unknown.

(e) Within the stochastic model $M$ of the phenomena $P$ the value of $\theta$ characterizes the distribution of the observable random variables $X$.

In the present example, this quantity $X$ may be the number of bacteria in a sample taken from the total water supply. (In actual practice the observable variable is different, namely the number of fertile samples of a given size.) As postulated, $X$ is a Poisson variable with expectation equal to $\theta$.

The statistical problem consists in determining a function $f(X)$ that could serve as an appropriate estimate of $\theta$, that is to say, that could serve for the selection of an appropriate action with respect to the phenomena $P$.

In certain cases, the model $M$ implies that the parameter $\theta$ is itself a random variable with a specified distribution. In other cases, the stochastic element involved in the model $M$ is limited to the randomness of the variables $X$, and $\theta$ is treated as an unknown constant. The nature of the statistical problem of estimating $\theta$ depends very much on whether the model $M$ treats $\theta$ as a random variable or not.

In the particular model of contamination of water with bacteria, the only random element is $X =$ number of bacteria in a unit volume of water, and $\theta$ is an unknown constant. Taking into account the possible consequences of misestimation of $\theta$, one of the possible forms of the problem of estimation is as follows:

(i) The estimate $f(X)$ is required to have the property that, whatever be the value of $\theta > 0$, the probability that $f(X)$ will exceed $\theta$ has at least a pre-assigned value $\alpha$, close to unity,

$$P\{f(X) \geq \theta \mid \theta \} \geq \alpha.$$  

(1)

If this property is satisfied then, granting the validity of the model $M$, we may be sure that in the long run the contamination of water will be underestimated but very rarely.

(ii) Since overestimation of $\theta$ is unpleasant, at least because of the obnoxious taste of chlorinated water, we require from $f(X)$ that if $\theta'$ is greater than the actual value of $\theta$ then
be a minimum with respect to the class of estimates satisfying (i).

The solution of this problem is easy and the function \( f(X) \) provides the least upper bound of values of \( \theta \) assertable with preservation of condition (i).

If the factual situation justifies the substitution for the model \( M \) of another model \( M^* \), in which not only \( X \) but also \( \theta \) is a random variable with a specified distribution, then this additional datum is taken into account leading to another estimate \( f^*(X) \) of \( \theta \).

The point of this discussion and of this example is that the efforts of the representatives of modern statistical theory are directed towards solving problems that depend only on the stochastic model of the phenomena studied and on nothing else. Thus, the practical applicability of the solution reached within the model depends solely upon the adequacy of the model. Another essential point is that the solution reached is always unambiguously interpretable in terms of long range relative frequencies.

Returning to the comparison with the nonfrequentist theory of inductive inference it could be said that, if the solution of any problem in this theory is reached using certain elements that would not be usable in the frequentist theory, then the relation between the solution and the relevant phenomena is different than that just described. Specifically, if a problem of induction regarding phenomena \( P \), for which there is a stochastic model \( M \), is obtainable both within the frequentist and with a nonfrequentist theory of inductive inference, and if the latter is based on some elements \( E \) outside of the model \( M \) and, therefore, nonusable in the frequentist theory, then, the correspondence of the nonfrequentist solution and the phenomena \( P \) is bound to be looser than the correspondence between the frequentist solution and the same phenomena.

The above general ideas will now be illustrated on a particular problem which may have independent interest.

**PART II. PROBLEM OF HOMOGENEITY OF NEUTRAL V-PARTICLES**

**6. Phenomenon of Neutral V-Particles and the Stochastic Model of Their Decay**

Recent issues of physical journals contain a considerable number of papers [8, 9] given to the study of neutral V-particles. The description of the whole phenomenon would be too long for the present paper and, therefore, we shall limit ourselves to its particular aspect chosen to illustrate the general ideas presented in the first part of the paper.

For each particular neutral V-particle that decays in the cloud chamber the physicists are able to compute two numbers: \( t = \) the time interval between an
appropriately selected origin and the moment of decay, and \( T \), the time interval between the same origin and the moment when the undecayed particle would have escaped from observation (at least, reliable observation) by passing out of view. In other words, \( T \) is the potential maximum period of time during which the decay of the particle could have been observed and recorded. Thus, all the observations necessarily give \( 0 < t < T \).

The empirical values of \( t \) and \( T \) vary from one particle to the next. This variation is illustrated by the following data reproduced from a paper [8] by D. I. Page and J. A. Newth. They refer to 26 particles observed by the authors and classified as \( V \)-particles. The authors are very cautious in indicating a number of uncertainties relating to these data. However, for purposes of this article, these uncertainties will be ignored and the figures of Table I will be treated as reflecting accurately the 26 separate phenomena of disintegration of the neutral \( V \)-particles.

It is not impossible that some time in the future a theory will be constructed which, using some now unsuspected premises, would lead to the calculation of the time of decay of each individual particle. A theory of this kind would be analogous to Newtonian celestial mechanics that can be used, for example, to calculate the time of return of each individual comet to the vicinity of the sun and would represent a deterministic approach to the phenomenon of \( V \)-particles. To the present writer's knowledge no effort has been made, as yet, to treat the decay of particles deterministically, and it is doubtful whether this ever will be seriously attempted.

### Table I

<table>
<thead>
<tr>
<th>Particle number</th>
<th>( t = ) decay time ((10^{-16} \text{ sec.}))</th>
<th>( T = ) potential obs. time ((10^{-16} \text{ sec.}))</th>
<th>Particle number</th>
<th>( t = ) decay time ((10^{-16} \text{ sec.}))</th>
<th>( T = ) potential obs. time ((10^{-16} \text{ sec.}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>8.2</td>
<td>14</td>
<td>1.2</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>0.7</td>
<td>1.0</td>
<td>15</td>
<td>1.6</td>
<td>5.1</td>
</tr>
<tr>
<td>3</td>
<td>2.5</td>
<td>3.3</td>
<td>16</td>
<td>3.8</td>
<td>5.5</td>
</tr>
<tr>
<td>4</td>
<td>1.0</td>
<td>2.4</td>
<td>17</td>
<td>3.6</td>
<td>4.3</td>
</tr>
<tr>
<td>5</td>
<td>1.8</td>
<td>3.8</td>
<td>18</td>
<td>0.2</td>
<td>3.9</td>
</tr>
<tr>
<td>6</td>
<td>0.1</td>
<td>3.9</td>
<td>19</td>
<td>0.5</td>
<td>6.8</td>
</tr>
<tr>
<td>7</td>
<td>2.5</td>
<td>4.2</td>
<td>20</td>
<td>1.4</td>
<td>10.4</td>
</tr>
<tr>
<td>8</td>
<td>1.6</td>
<td>1.7</td>
<td>21</td>
<td>3.3</td>
<td>5.2</td>
</tr>
<tr>
<td>9</td>
<td>2.8</td>
<td>4.2</td>
<td>22</td>
<td>1.6</td>
<td>4.8</td>
</tr>
<tr>
<td>10</td>
<td>2.2</td>
<td>4.9</td>
<td>23</td>
<td>1.2</td>
<td>3.5</td>
</tr>
<tr>
<td>11</td>
<td>1.9</td>
<td>6.3</td>
<td>24</td>
<td>2.2</td>
<td>3.4</td>
</tr>
<tr>
<td>12</td>
<td>0.4</td>
<td>1.3</td>
<td>25</td>
<td>0.1</td>
<td>1.3</td>
</tr>
<tr>
<td>13</td>
<td>5.2</td>
<td>14.1</td>
<td>26</td>
<td>0.1</td>
<td>3.2</td>
</tr>
</tbody>
</table>
Instead of trying to study single particles with the intent of connecting deterministically with some other factors the individual times of their decay, we proceed indeterministically and try to determine how frequently particles of a given type decay within stated intervals of time since their birth. These frequencies of decay are, then, treated as characteristics of the particular types of particles.

As a result of many studies, a particular stochastic model of radioactive decay has become established, as described, for example, by Feller [10, 11]. According to this model, the time $X$ of decay of any particular particle is a random variable with exponential distribution

$$F_X(t) = P[X \leq t] = 1 - e^{-\lambda t} \quad \text{for} \quad 0 \leq t.$$

Formula (3) contains a parameter $\lambda$, which we shall label the decay parameter. This parameter, or its reciprocal $\theta = \lambda^{-1}$ termed the "average life" of the particle, characterizes the particular category to which the particle belongs. The operational meaning of formula (3) is that, if a considerable number of particles of the same category are observed, each for an indefinite time, then the relative frequencies of decays occurring between arbitrary moments $t_1 < t_2$ will be approximately equal to $F_X(t_2) - F_X(t_1)$.

In actual fact, the decay of a particle may be observed only if this decay occurs while the particle is passing through the cloud chamber. Furthermore, if a neutral $V$-particle fails to decay during its passage through the cloud chamber, then this particle escapes detection altogether. On the other hand, a decay occurring within the cloud chamber is observable and the observations yield not only the time $t$ of decay but also the maximum time $T$ within which the decay could have been observed.

In other words, the observations yield the particular values $t$ of the random variables $X$ corresponding not to all particles passing through the chamber, but only for those for which $X$ is less than the limit $T$ ascertainable for each particular particle. It follows that the random variables, for which the particular values are observable, do not follow the distribution of the form (3) but its "truncated" form

$$F_X(t) = \frac{1 - e^{-\lambda t}}{1 - e^{-\lambda T}} \quad \text{for} \quad 0 \leq t \leq T,$$

subject to the condition that $0 \leq X \leq T$.

In the following we will take formula (4) for granted and shall consider certain problems of inductive inference (really of inductive behavior) that face the physicists studying neutral $V$-particles. One such problem, that of estimating the value of the parameter $\lambda$, was recently studied by Bartlett [12]. The observable variables contemplated by Bartlett are exemplified by the figures in Table I. In addition to postulating (4) with respect to each particular particle, the main part of Bartlett's work postulates that the value of the decay
parameter $\lambda$ is the same for all the particles selected for the study. Our own problem will be, so to speak, of an earlier character and will concern the homogeneity of particles with respect to their values of the parameter $\lambda$.

7. Problem of Homogeneity of Decaying Neutral V-Particles

In this section we will visualize a physicist interested in the question whether a particular method of classifying neutral V-particles leads to classes homogeneous with respect to the decay parameter $\lambda$. His observational data are of the character exemplified in Table I. Upon analyzing these data he may adopt one of two possible actions: to consider his method of classification faulty and look for refinements which would permit the splitting of his original classes of particles into some subclasses. This may or may not lead to the discovery of new types of particles. Alternatively, the physicist may adopt the attitude that his classes of particles are homogeneous and, therefore, that his method of classification is satisfactory.

In statistical terminology the problem is as follows. A certain number of mutually independent random variables $X_1, X_2, \ldots, X_n$ ($= \mathbf{X}$) will be observed and it is taken for granted that the $i$-th of them follows a law characterized by the distribution function of the type (4) with a known $T = T_i$, and with an unknown value $\lambda_i$ of the decay parameter $\lambda$. The corresponding probability density evaluated at $X_i = x$ is, say

$$p_{X_i}(x | \lambda_i, T_i) = \frac{\lambda_i e^{-\lambda_i x}}{1 - e^{-\lambda_i T_i}}, \quad 0 \leq x \leq T_i.$$  

We denote by $\Omega_0$ the set of hypotheses specifying particular values of the $n$ unknown parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$. We are particularly interested in the subset of $\Omega_0$ characterized by the equalities

$$\lambda_1 = \lambda_2 = \cdots = \lambda_n.$$  

The logical sum of these hypotheses, to be denoted by $H_0$, is our hypothesis tested. Our problem is to deduce a test of $H_0$ which will have an easily interpretable optimum property. What is an optimum property is a subjective matter, lying outside of the theory of statistics. Our own approach is the following.

Denote by $\mathcal{X}$ the set of $n$-tuples $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, with $0 \leq x_i \leq T_i$, $i = 1, 2, \ldots, n$. This set will be called the sample space. We recognize that the definition of a method of testing the hypothesis $H_0$ amounts to the definition of a subset $w$ of $\mathcal{X}$, termed the critical region, and to the adoption of the rule to reject $H_0$ then and only when the sample point $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ determined by the observations falls within $w$. Thus, whatever the optimal property of the desired test may be, this property can be treated as that of the critical region $w$. In order to define this optimal property we shall consider the conditional probability $P(\mathbf{X} \notin w | H_0)$, given that the hypothesis $H_0$ is true, that the sample point $\mathbf{X}$ will fall within $w$ and thus that the test will reject $H_0$. 

Our first requirement will be that the critical region \( w \) be so chosen that this probability \( P\{X \in w | H_0\} \) have a small value \( \alpha \) chosen in advance. This number \( \alpha \) is termed the level of significance. Since the hypothesis \( H_0 \) does not specify the common value, say \( \lambda \), of the \( n \) decay parameters (6), this creates a certain difficulty in selecting \( w \). In fact, the critical region \( w \) must be what is termed "similar" to the sample space with respect to the parameter \( \lambda \). Using (5) and (6) the condition of similarity may be expressed by the identity

\[
P\{X \in w | H_0\} = \frac{\lambda^n}{\prod_{i=1}^{n} (1 - e^{-\lambda r_i})} \int_{w} \exp \left\{-\lambda \sum_{i=1}^{n} x_i \right\} dx_1 \cdots dx_n = \alpha.
\]

Assume for the moment that the problem of similar regions is solved and that we are able to construct effectively all the subsets \( w \) of \( \mathcal{X} \) that satisfy condition (7). Let \( \Phi(\alpha) \) denote the family of all such subsets. Our next step is to define the particular subset, say \( w_0 \), which will be selected for the actual test of the hypothesis \( H_0 \). For this purpose we exercise our imagination in order to define the particular way in which the hypothesis tested \( H_0 \) may be false.

In the present case two possibilities suggest themselves.

(i) A priori it may seem plausible that, if the decay parameter of a given category of \( V \)-particles is not constant for all these particles, then it may vary from one to another in a random manner, about some typical value. Thus we come to the necessity of considering a new random variable, say \( \Lambda \), of which the decay constants \( \lambda_1, \lambda_2, \ldots, \lambda_n \) of particles actually observed are particular values. Nothing whatever is known regarding the probability law of \( \Lambda \), but the intuition suggests the possibility that it would be one with a probability density and that the latter may be approximately represented by the Pearson Type III curve, say

\[
p_\Lambda(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \quad \text{for} \quad 0 < x
\]

where \( \alpha \) and \( \beta \) are unknown positive constants. The expectation, say \( \lambda \), of \( \Lambda \) and its coefficient of variation \( v \) are easily expressible in terms of \( \alpha \) and \( \beta \) and we have

\[
\alpha = \frac{1}{\lambda^v}, \quad \beta = \frac{1}{\lambda v^2}.
\]

The hypothesis tested \( H_0 \) is now expressible in terms of the parameter \( v \), namely, \( H_0 \) asserts that \( v = 0 \). As \( v \to 0 \), the probability density (8) degenerates and the random variable \( \Lambda \) tends to a constant \( \lambda \). Figure 1 gives graphs of the probability density (8) corresponding to a fixed value of \( \lambda = 0.271 \) and three values of the coefficient variation \( v = 0.1, v = 0.2 \) and \( v = 0.3 \). The heavily drawn vertical represents the common value \( \lambda \) of the expectation of \( \Lambda \).

The particular value \( \lambda = 0.271 \) was selected for illustration because it represents the estimate [11] of the decay parameter obtained using data in
Table I, on the assumption that the 26 particles referred to in this table are a sample of a homogeneous class.

If the intuition of the physicist suggests that the decay parameter of a \( V \)-particle may be a random variable with a distribution approximated by one of the curves in Figure 1, then he may choose to adjust his test of the hypothesis \( H_0 \) so as to be particularly sensitive to the deviations exemplified in Figure 1.

On the assumption that the decay parameter of a particle is a random variable following the density (8), the nontruncated probability density of the time \( X \) of decay of a particle is given by

\[
p_x(x \mid \lambda, \nu) = \int_0^\infty te^{-tx} \times \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} dt
\]

(10)

\[= \frac{\alpha \beta^\alpha}{(\beta + x)^{\alpha+1}} = \lambda(1 + \nu^2 x)^{-\frac{1}{\nu^2+1}}\]

where \( \alpha \) and \( \beta \) are connected with \( \lambda \) and \( \nu \) by means of formulae (9). The assumption that \( 0 \leq X \leq T \) leads to the truncated form of the distribution, namely

\[
p_x(x \mid \lambda, \nu, T) = \frac{\lambda(1 + \nu^2 x)^{-\frac{1}{\nu^2+1}}}{1 - (1 + \lambda^2 T)^{-\frac{1}{\nu^2}}} \quad \text{for} \quad 0 \leq x \leq T.
\]

Fig. 1. Graphs of probability density of \( \Lambda \) corresponding to four values of \( \nu \).
It is obvious that, as \( v \to 0 \), formula (11) converges to (5). Denote by \( \Omega_1 \) the set of hypotheses each assuming that the time of decay \( X \) is a random variable following the probability density (11) with some specified values of \( \lambda > 0 \) and \( v > 0 \). In the following we shall deduce a test of the hypothesis \( H_0 \) that will have an optimal property with respect to the set \( \Omega_1 \) of hypotheses alternative to \( H_0 \). The corresponding critical region, satisfying the condition of similarity, will be denoted by \( w_1 \).

(ii) Instead of visualizing the possibility that each neutral \( V \)-particle has a different value of the decay parameter varying continuously in a random fashion about a certain typical value \( \lambda \), the physicist may think of the possibility that the particles which appear to be all of the same type can actually be divided into two sharply different categories of which one is predominant. The particles of the predominant category are all characterized by a fixed value \( \lambda_1 \) of the decay parameter. The particles of the secondary category are all characterized by a fixed decay parameter \( \lambda_2 \) which, however, is a substantial multiple \( \rho \) of the first, \( \lambda_2 = \rho \lambda_1 \). The proportion of particles of the secondary category, say \( \gamma \), is a small number. Using the letter \( \lambda \) to represent the average decay parameter for all particles, we easily find

\[
\begin{align*}
\lambda_1 &= \frac{\lambda}{1 + (\rho - 1)\gamma}, \\
\lambda_2 &= \frac{\rho \lambda}{1 + (\rho - 1)\gamma}.
\end{align*}
\]

Remembering that the decay parameter is the reciprocal of the average time of the particle's life, it may be said that, in the case contemplated, all the particles are split into two categories, of which the predominant one is characterized by a reasonably long average life and the secondary by a short average life.

If the physicist expects that the particles with which he has to deal combine the predominant category with a small contamination by short-lived particles, he may choose to select a test of the hypothesis \( H_0 \) against the set \( \Omega_1 \) of alternatives defined by equations (12) with \( \lambda > 0, \gamma > 0 \) and with some specified value of \( \rho \), for example \( \rho = 5 \) or \( \rho = 10 \), etc. The hypothesis tested \( H_0 \) will be expressed by the condition \( \gamma = 0 \). On the assumption of two categories of particles, the truncated probability density of the time of decay \( X \) is given by, say,

\[
p^x_\gamma(x \mid \lambda, \rho, \gamma, T) = \frac{(1 - \gamma)\lambda_1 e^{-\lambda_1 x} + \gamma \lambda_2 e^{-\lambda_2 x}}{1 - (1 - \gamma)e^{-\lambda T} - \gamma e^{-\lambda_2 T}} \quad \text{for} \quad 0 \leq x \leq T
\]

where \( \lambda_1 \) and \( \lambda_2 \) are connected with \( \lambda \), \( \rho \) and \( \gamma \) by means of formulae (12).

8. Locally Best One-Sided Tests

In this section we define the concept of the locally best one-sided similar test. For this purpose we detach ourselves from the consideration of the hypothesis \( H_0 \) of homogeneity of neutral \( V \)-particles and consider a general case of an observable random variable \( Y \) known to possess a probability density
$p_r (y \mid \theta, \theta)$ depending upon $s + 1$ parameters with unspecified values. One parameter, $\theta$ will be assumed to have a value $\theta \geq \theta_0$. The totality of $s \geq 1$ other parameters is symbolized by a single letter $\theta$ and it is assumed that $\theta$ is known to belong to some specified set $\theta \in \Theta$. We shall consider the hypothesis that asserts $\theta = \theta_0$. Concerning the probability density of the variable $Y$ it will be assumed that, for all $\theta \in \Theta$, it is sufficiently regular to allow several differentiations with respect to $\theta$ under the sign of integral evaluated over an arbitrary set $w$ in the sample space.

In Section 7 we explained the concept of a region similar to the sample space with respect to a given parameter. In the present section we shall consider only regions $w$ that are similar with respect to $\theta$, so that

$$P \{ Y \in w \mid \theta = \theta_0 \} = \int_w p_r (y \mid \theta_0, \theta) \, dy \equiv \alpha$$

where $\alpha$ is a small preassigned number described as the level of significance. Now consider consecutive derivatives of $P \{ X \in w \mid \theta \}$ with respect to $\theta$ evaluated at $\theta = \theta_0$. It is possible that the first of them will have the same value for all the regions $w$ satisfying (14). Let $k$ be the order of the first derivative of $P \{ X \in w \mid \theta \}$ evaluated at $\theta = \theta_0$ the value of which depends on the region $w$.

**DEFINITION.** If the region $w_0$ is similar with respect to $\theta$ so that the identity (14) is satisfied and in addition if, whatever be $w$ satisfying (14),

$$\frac{\partial^k}{\partial \theta^k} P \{ X \in w_0 \mid \theta \} \bigg|_{\theta = \theta_0} \geq \frac{\partial^k}{\partial \theta^k} P \{ X \in w \mid \theta \} \bigg|_{\theta = \theta_0},$$

for every $\theta \in \Theta$, then we shall say that the region $w_0$ is the locally best one-sided similar region for testing the hypothesis $H$ against the set of alternative hypotheses $\theta > \theta_0$ and $\theta \in \Theta$.

In order to justify this terminology we consider the important concept of the power function of a test first introduced in 1933 [13]. This latter term refers to the probability $P \{ X \in w \mid \theta, \theta \}$ considered as a function, say $\beta (\theta)$, of $\theta$ while $\theta$ is kept fixed. If $w$ is adopted as the critical region for testing the hypothesis $H$, then $\beta (\theta)$ represents the probability that the test will reject $H$ in cases where $\theta$ is the true value of the parameter concerned. Obviously, when $\theta > \theta_0$ it is desirable to reject $H$ just as frequently as possible. Thus, it is desirable to use that particular region $w_0$ for which, as $\theta$ is increased from the initial value $\theta_0$, that is, as the hypothesis $H$ becomes more and more false, the power function increases faster than, or at least as fast as, that corresponding to any other similar region $w$.

If the probability density $p_r (y \mid \theta, \theta)$ is sufficiently regular to permit $k$ differentiations under the integral sign, then condition (15) can be rewritten in the form

$$\int_w \varphi (y) p_r (y \mid \theta_0, \theta) \, dy \geq \int_w \varphi (y) p_r (y \mid \theta_0, \theta) \, dy,$$
where \((y)\) is defined for all \(y\) at which \(p_r(y \mid \theta, \theta)\) does not vanish

\[
\phi(y) = \left. \frac{\partial^k p_r(y \mid \theta, \theta)}{\partial \theta^k} \right|_{\theta_0, \theta} / p_r(y \mid \theta_0, \theta).
\]

If \(k = 1\), then \(\phi(y)\) represents the logarithmic derivative of the probability density of \(Y\).

Returning to the particular problem of testing the hypothesis \(H_0\) that the \(V\)-particles of a given class are homogeneous with respect to their decay parameter \(\lambda\), we notice that the search for the best one-sided similar test must be split into two stages. First it is required to construct a class, if possible the complete class, say \(\Phi(\alpha)\) of regions \(w\) that are similar to the sample space with respect to the decay parameter \(\lambda\) assumed the same for all particles, but with an unspecified numerical value. The second step consists in determining within the class \(\Phi(\alpha)\) the particular element \(w_0\), if it exists, which satisfies condition (15).

9. Similar Regions for Testing Hypothesis \(H_0\)

The concept and the first systematic studies of similar regions were published in 1933 [13] and 1937 [14]. Of the further contributions those due to Lehmann and Scheffé deserve particular attention. The general results published in these papers [15, 16] imply that, if the observable random variables \((X_1, X_2, \ldots, X_n) = X\) have the joint probability density asserted by the hypothesis \(H_0\), namely

\[
p_x(x \mid \lambda) = \frac{\lambda \exp \left\{ -\lambda \sum_{i=1}^{n} x_i \right\}}{\prod_{i=1}^{n} (1 - e^{-\lambda T_i})}
\]

for \(0 \leq x_i \leq T_i\),

where \(T_1, T_2, \ldots, T_n\) are given numbers, then a region \(w\) in the sample space must possess a particular structure in order to be similar with respect to \(\lambda > 0\).

Let \(s\) be a positive number not exceeding the limit

\[
0 \leq s \leq \sum_{i=1}^{n} T_i = T^* \quad \text{(say)}.
\]

Let \(\mathcal{X}(s)\) stand for the set of points in the sample space \(\mathcal{X}\) satisfying the condition

\[
\sum_{i=1}^{n} x_i = s
\]

and \(w(s)\) for a measurable subset of \(\mathcal{X}(s)\). Since \(S_* = \sum_{i=1}^{n} X_i\) is a sufficient statistic for \(\lambda\), the conditional probability that \(X \in w(s)\) given that \(X \in \mathcal{X}(s)\), is independent of \(\lambda\). Denote by \(w(s, \alpha)\) a measurable subset of \(\mathcal{X}(s)\) so chosen that

\[
P\left\{ X \in w(s, \alpha) \mid \sum_{i=1}^{n} X_i = s \right\} = \alpha.
\]
Then, for the region \( w \) to be similar with respect to \( \lambda \), it is necessary and sufficient that it be a union of subsets \( w(s, \alpha) \)

\[ w = \bigcup_{0 < s < T} w(s, \alpha) \]

and that it be measurable.

It follows that the complete class \( \Phi(\alpha) \) of regions similar with respect to \( \lambda \) can be obtained by constructing unions (22) of subsets \( w(s, \alpha) \) subject to condition (21) and by eliminating those that are nonmeasurable.

In the following, it will be necessary to evaluate integrals taken over a similar region \( w \). Because of the particular structure that this region must possess, the handling of such integrals is exemplified by a transformation to a new system of variables, say \( s \) and \( y = (y_1, y_2, \ldots, y_{n-1}) \) such that \( s \) is given by (18) and \( y_1, y_2, \ldots, y_{n-1} \) are selected more or less arbitrarily with the familiar regularity restrictions. Let \( \mathbf{x}(s) \) and \( w^*(s) \) stand for the images of \( \mathbf{x}(s) \) and \( w(s, \alpha) \), respectively, in the space of the new variables. Then, for every integrable function \( f(x) \),

\[ \int \int f(x) \, dx = \int \int f(x(y)) \, J \, dy \]

where \( x(y) \) represents the vector function from \((s, y)\) to \( x \) and

\[ J = \frac{\partial(x_1, x_2, \ldots, x_n)}{\partial(s, y_1, y_2, \ldots, y_{n-1})} \]

Referring to condition (21) it can be said that \( w^*(s) \) is an arbitrary measurable subset of \( \mathbf{x}(s) \) subject to the condition that

\[ \int_{w^*(s)} p_x(x(y)) \, J \, dy \leq \alpha \int_{\mathbf{x}(s)} p_x(x(y)) \, J \, dy \]

and that the union

\[ w^* = \bigcup_{0 < s < T} w^*(s) \]

be measurable.

10. **Locally Best One-Sided Similar Tests of the Hypothesis \( H_0 \)**

In the following we shall deduce two similar regions \( w_1 \) and \( w_2 \) that are one-sided and locally best for testing the hypothesis \( H_0 \) against the two sets of alternatives \( \Omega_1 \) and \( \Omega_2 \), respectively. For this purpose it will be necessary for us to consider the probability densities of the observable times of decay \( X_1, X_2, \ldots, X_n \) as a family indexed by a single parameter \( \vartheta \) such that \( \vartheta = \vartheta_0 \) corresponds to the hypothesis tested \( H_0 \) and \( \vartheta > \vartheta_0 \) corresponds to an alternative hypothesis.

Reviewing the two preceding sections, it will be seen that the problem of determining the locally best similar region, say \( w_0 \), reduces to maximizing the
integral in the left hand side of (16) with respect to all regions that have structure (26) and satisfy (25). But (16) may now be rewritten in the form (23),

\[ \int_0^\infty ds \int_{u_1(s)} \varphi(x)p_x(x \mid \theta, \lambda) \mid J \mid dy, \]

where, for the sake of brevity, \( x \) stands for \( x(y) \). Now it is shown that in order to select a similar region maximizing the left hand side of (16), it is sufficient to select, for each \( s \) in the interval \( 0 < s \leq T^* \), an \( n - 1 \) dimensional region \( w_0(s) \) satisfying (25) and maximizing the integrand

\[ \int_{u_1(s)} \varphi(x)p_x(x \mid \theta, \lambda) \mid J \mid dy \]

in (27). This is easily accomplished by the application of the fundamental lemma of theory of testing hypotheses. In fact, the maximizing region \( w_0(s) \) is defined by the condition

\[ \varphi(x) \geq a(s) \]

where \( a(s) \) is determined for each \( s \) so that

\[ P\left\{ \varphi(X) \geq a(s) \mid \sum_{i=1}^n X_i = s \right\} = \alpha. \]

Here the probability on the left is computed on the assumption that the hypothesis tested is true. It follows from (26) that the critical region sought is defined by

\[ \varphi(x) \geq a\left( \sum_{i=1}^n x_i \right). \]

As will be seen below, in the problem considered, \( k = 1 \), so that \( \varphi(x) \) is simply the logarithmic derivative of the probability density \( p_x(x \mid \vartheta, \lambda) \) with respect to \( \vartheta \), evaluated at \( \vartheta = \theta_0 \). It follows that the procedure for determining the critical region sought reduces to the following steps:

1. Compute the logarithmic derivative \( \varphi(x) \), substitute the random variables \( X \) instead of \( x \) and deduce the conditional probability density of the random variable,

\[ Y_n = \varphi(X), \]

given that the random variable

\[ S_n = \sum_{i=1}^n X_i \]

has assumed a specified value \( s \).

2. For each \( s \) determine \( a(s) \), so as to satisfy (30). Once the function \( a(s) \) is determined, the application of the test consists in the rule of rejecting \( H_0 \) whenever the observed values of \( X \) satisfy (31).
After this general description, we proceed to develop the tests against
the sets of alternatives $\Omega_1$ and $\Omega_2$.

Considering first $\Omega_1$, we return to formula (11) and identify the square of
the coefficient of variation, $v^2$, with the parameter $\delta$ that appeared in the general
discussion above. Obviously $\delta_0 = 0$.

The joint probability density of the $n$ observable random variable
$(X_1, X_2, \ldots, X_n) = X$ all independent and each following the law (11), with an
appropriate value of $T$, is given by

$$p(x | \lambda, \nu) = \prod_{i=1}^n \frac{\lambda(1 + \nu^2 x_i)^{-(1/\nu + 1)}}{1 - (1 + \nu^2 x_i)^{-1/\nu}}$$

for $0 \leq x_i \leq T$, $i = 1, 2, \ldots, n$.

Differentiating the logarithm of (34) with respect to $\nu^2$, and passing to the limit
as $\nu \to 0$,

$$\varphi(x) = C + \lambda^2 \sum_{i=1}^n x_i^2,$$

where $C$ stands for a quantity independent of $x_1, x_2, \ldots, x_n$.

Referring to (29) and (30) it is easy to see that the locally best similar
region $w_1$ for testing the hypothesis $H_0$ against the alternatives $\Omega_1$ is defined
by, say

$$Y_* = \sum_{i=1}^n X_i^2 \geq b_1(\sum X_i)$$

where the function $b_1(s)$ is so selected that

$$P \{ Y_* \geq b_1(s) \mid S_* = s \} = \alpha.$$

In order to determine the critical region $w_2$ which is similar and locally
best for testing $H_0$ against the set of alternatives $\Omega_2$, we turn to (13) and write
down the joint probability density of the $n$ observable random variables
$X_1, X_2, \ldots, X_n$,

$$p^2(x | \lambda, \rho, \gamma, T) = \prod_{i=1}^n \frac{(1 - \gamma) \lambda e^{-\lambda_1 x_i} + \gamma \lambda e^{-\lambda_2 x_i}}{1 - (1 - \gamma) e^{-\lambda_1 x_i} - \gamma e^{-\lambda_2 x_i}},$$

with $\lambda_1$ and $\lambda_2$ given by (12), and identify the probability $\gamma$ with the parameter
$\delta$ that served in the general discussion of locally best similar tests. Here $\delta_0 = 0$.

The logarithmic derivative of (38) with respect to $\gamma$ evaluated at $\gamma = 0$ is,

$$\varphi^*(x) = C + (\rho - 1) \lambda \sum_{i=1}^n x_i + \rho \sum_{i=1}^n e^{-\lambda(s-1)x_i}$$

and it is seen that the critical region $w_2$ is defined by the inequality, analogous
to (36),

$$Z_* = \sum_{i=1}^n e^{-\lambda(s-1)x_i} \geq b_2(S_*)$$
where $b_2(s)$ must be so determined that, for $0 < s \leq T^*$, and on the hypothesis tested,

\[(41) \quad P\{Z_* \geq b_2(s) \mid S_* = s\} = \alpha.\]

It will be noticed that there is a very substantial difference between the definition of region $w_1$ and that of region $w_2$. Since on the hypothesis $H_0$ the conditional distribution, given $S_* = s$, of any function of the observable variable is independent of the unknown $\lambda$, the function $b_1(s)$ can be unambiguously determined and then the application of the test based on $w_1$ is straightforward: the hypothesis $H_0$ is rejected whenever the observed values of $X_1, X_2, \ldots, X_n$ satisfy the inequality (36). The situation is more complicated with the region $w_2$. The reason is that the criterion $Z_*$ defined by (40) depends on the value of $\lambda$ which is not specified by the hypothesis $H_0$. Thus, strictly speaking, the region $w_2$ cannot be used in practice. However, the difficulty may be approximately solved by substituting in (40) instead of $\lambda$ its maximum likelihood estimate $\hat{\lambda}$. This estimate may be obtained numerically by solving the equation

\[(42) \quad n + \sum_{i=1}^{\hat{\lambda}} (T_i - x_i) - \sum_{i=1}^{\hat{\lambda}} \frac{T_i}{1 - \exp[-\hat{\lambda}T_i]} = 0.\]

The estimate $\hat{\lambda}$ is not only consistent but also sufficient and, therefore, as will be shown in another publication, the criterion $Z_*$ of (40) may be replaced by the "working" criterion, say

\[(43) \quad Z_*^w = \sum_{i=1}^{\hat{\lambda}} \exp\{-\hat{\lambda}(\rho - 1)X_i\}\]

without any change in the asymptotic properties of the test.

The completion of the problem requires the evaluation of the two functions $b_1(s)$ and $b_2(s)$ so as to satisfy the conditions (37) and (41), respectively. Since the values of $T_1, T_2, \ldots, T_\ast$ are all different, this task is extremely laborious and, at least for purposes of the present paper, we shall adopt an approximate method. This is based on the following facts, the details of which are reserved for a separate publication. First we notice that, using the bivariate version of the central limit theorem, as $n \to \infty$, the two pairs of variables $(S_*, Y_*)$ and $(S_*, Z_*)$, properly normed, tend to be normally distributed with zero means, unit variances and with easily computable correlations. Second, and this is the more important circumstance, given the value of the normed variable $S_*$, the conditional density of the normed $Y_*$ and the conditional density of the normed $Z_*$ converge to the conditional density computed from the joint normal limiting distribution of the corresponding pair.

Denote by $\nu(\alpha)$ the normal deviate defined by the condition

\[(44) \quad \frac{1}{\sqrt{2\pi}} \int_{\nu(\alpha)}^{\infty} e^{-t^2/2} dt = \alpha.\]
Fix a number \( \tau \) and, using an obvious notation, denote by \( s_\tau \) the function of \( \tau \) defined by

\[
 s_\tau = E(S_\tau) + \tau \sigma_s.
\]

The results just quoted imply that, if the hypothesis tested is true, then, for every \( \tau \),

\[
 \lim_{n \to \infty} P \left( \frac{Y_\tau - E(Y_\tau) - \tau R_{Y,s} \sigma_Y}{\sigma_Y \sqrt{1 - R^2_{Y,s}}} \geq \nu(\alpha) \mid S_\tau = s_\tau(\tau) \right) = \alpha.
\]

It follows that, for large \( n \), the function \( b_\tau[s_\tau(\tau)] \) may be replaced by, say

\[
 C_\tau(\tau) = E(Y_\tau) + \tau R_{Y,s} \sigma_Y + \nu(\alpha) \sigma_Y \sqrt{1 - R^2_{Y,s}}.
\]

The region, say \( w_1(n) \), defined parametrically by

\[
 S_\tau = s_\tau(\tau),
\]

or explicitly by

\[
 Y_\tau \geq C_\tau(\tau),
\]

will then have, approximately, the same properties as the critical region \( w_1 \). A similarly defined region \( w_2(n) \), with \( Z_\tau \) replacing \( Y_\tau \) in formula (49), may be used instead of the region \( w_1 \). When \( n \) is large and the hypothesis \( H_0 \) of uniformity of particles happens to be true, both regions \( w_1(n) \) and \( w_2(n) \) will reject this hypothesis with a small probability close to \( \alpha \). Because of the fact that \( S_\tau \) is a sufficient statistic for \( \lambda \), the same conclusion will apply even though the true value of this parameter appearing in the left hand side of (49) is replaced by \( \hat{\lambda}_n \).

Now it is interesting to evaluate, approximately, the probabilities that these regions will reject \( H_0 \) when, in fact, this hypothesis is false and one of the alternatives, either from set \( \Omega_1 \) or from the set \( \Omega_2 \), happens to be true. This probability, a function of the alternative hypothesis on which it is computed, is called the power function of the test considered.

When one speaks of the approximate value of the probability, especially if this expression applies to a situation in which the number \( n \) of observations is described as "large," one has in mind a limit to which the probability in question converges as \( n \to \infty \). And here we come to a difficulty. If the tests contemplated are any good, then, as \( n \to \infty \) while the hypothesis tested is false and a fixed alternative is true, the probability of rejecting the hypothesis tested will tend to unity. This desirable property is described by the term "consistency."

While the consistency of a test is a very important property, one is interested in finding something more informative. A passage to the limit as \( n \to \infty \) was devised [17] in 1937 that permits the evaluation of the limiting form of the power
function of a consistent test which may be considered as an approximation to the actual power corresponding to large values of \( n \). Originally invented to apply to the smooth test for goodness of fit, (later studied by David [18] and extended by Scott [19] and by Barton [20]) the same method also served in determining the limiting power function of the \( \chi^2 \) test by Eisenhart [21] and, subsequently, in the study of more general problems by Wald [22]. Because of its wide range of applicability, the method is described in some detail in the following section.

11. Limiting Power Function of Tests Based on Sequences of Critical Regions \( \{w_1(n)\} \) and \( \{w_2(n)\} \)

Consider the general situation where the hypothesis tested \( H \) specifies a value \( \theta_0 \) of a parameter \( \theta \) involved in the probability law of the observable random variables \( X \), while the alternative hypotheses assert \( \theta > \theta_0 \). Suppose that for every number \( n \) of observations we have selected a critical region \( w_n \) corresponding, at least approximately, to a fixed level of significance \( \alpha \), and that the sequence \( \{w_n\} \) of these critical regions determines a consistent test. Let \( \beta(\theta \mid w_n) \) stand for the probability that the test based on \( w_n \) will reject \( H \), computed on the assumption that \( \theta \) is the true value of the parameter considered. This probability may well depend on the values of some other parameters unspecified by the hypothesis tested. However, we shall keep these values constant and consider the changes of \( \beta(\theta \mid w_n) \) corresponding to the variation of \( \theta \) alone.

Figure 2 represents plausible graphs of the power function of a consistent test.

Figure 2. Plausible graphs of the power function of a consistent test.
gation we have \( n_1 \) observations and that \( \theta' \) is the value of the parameter for which an approximate value of the power is sought. Upon inspecting Figure 2, it will be seen that there exists a sequence of values of \( \theta' \), say \( \{\theta'(n)\} \) such that

\[
\beta(\theta'(n) \mid w_n) = \beta(\theta' \mid w_n)
\]

and that this sequence converges to \( \theta_0 \). The exact nature of this sequence is, of course, unknown. However, in seeking an approximate value of \( \beta(\theta' \mid w_n) \) it is natural to try to approximate this sequence and, on a number of occasions, the sequence \( \{\theta_0 + (\theta' - \theta_0) \sqrt{n_1/n}\} \) appears to be useful. In fact, it often appears possible to prove that the limit

\[
\lim_{n \to \infty} \beta[\theta_0 + (\theta' - \theta_0) \sqrt{n_1/n} \mid w_n] = \beta_\infty(\theta' \mid w_n)
\]

exists and is easily computed. Once this limit is computed, it may be used to provide an approximation to the actual value of \( \beta(\theta' \mid w_n) \).

In relation to the two sequences \( \{w_1(n)\} \) and \( \{w_2(n)\} \) of critical regions defined in the preceding section, the method just described is applied as follows. In each of the two cases (i.e. for sequences \( \{w_1(n)\} \) and \( \{w_2(n)\} \)), the probability that the hypothesis \( H_0 \) is rejected coincides with the probability that an expression of the form (49) exceeds the fixed limit \( \nu(\alpha) \). In both cases the expression in question involves two sums of independent random variables, \( (Y_*, S_*) \) in one case and \( (Z_*, S_*) \) in the other. Generalizing, we will consider two sums of random variables

\[
U_* = \sum_{i=1}^s u_i \quad \text{to correspond to } S_*
\]

and

\[
V_* = \sum_{i=1}^s v_i \quad \text{to correspond to either } Y_* \text{ or } Z_*.\]

We shall assume that on every admissible hypothesis the pairs of variables \( \{u_i, v_i\} \) are independent, that the variables themselves have bounded variances which are bounded away from zero, and that otherwise they satisfy the conditions of the bivariate version of the central limit theorem. Denote by \( \xi_i \) and \( \eta_i \) the expectations of \( u_i \) and \( v_i \), respectively, as determined by the hypothesis tested \( H \). Consider a sequence \( \{h_n\} \) of alternative hypotheses such that, as \( n \to \infty \),

\[
E(u_i \mid h_n) = \xi_i + \frac{\delta_i}{\sqrt{n}} + O(1/\sqrt{n}) = \xi_i(n) \quad \text{(say)},
\]

\[
E(v_i \mid h_n) = \eta_i + \frac{\Delta_i}{\sqrt{n}} + O(1/\sqrt{n}) = \eta_i(n).
\]

Finally assume that, as \( n \to \infty \), the variances and the covariances of \( u_i \) and \( v_i \), computed on \( h_n \) converge to their values corresponding to \( H \). Let \( \sigma^2_{u}(h) \),
\( \sigma\nu(h) \) and \( \sigma u.v.(h) \) denote the variances and the covariances of the sums \( U_n \) and \( V_n \) computed under any hypothesis \( h \). Then, under the assumptions made, as \( n \to \infty \), the variances \( \sigma\nu(h) \) and \( \sigma u.v.(h) \) are of order of magnitude of \( n \) and the ratios

\[
\frac{\sigma\nu(h)}{\sigma\nu(H)}, \quad \frac{\sigma u.v.(h)}{\sigma u.v.(H)}, \quad \frac{\sigma u.v.(h)}{\sigma u.v.(H)}
\]

converge to unity. Now form the expression \( W_n \) analogous to that in the left hand side of (49),

\[
W_n = \frac{V_n - \sum \eta_i}{\sigma\nu(H) \sqrt{1 - R u.v.(H)}} - \frac{U_n - \sum \xi_i}{\sigma u.v.(H) \sqrt{1 - R u.v.(H)}},
\]

and consider the probabilities that \( W_n \geq \nu(\alpha) \). It will be seen that, on the assumption that \( H \) is true, \( W_n \) may be presented in the form of a normed sum of independent random variables satisfying the conditions of the central limit theorem. Thus, on the hypothesis \( H \), the probability that \( W_n \geq \nu(\alpha) \) converges to \( \alpha \).

In order to determine the limit of the probability \( P\{W_n \geq \nu(\alpha) | h_n\} \) we notice that the expression of \( W_n \) may be rewritten as

\[
W_n = W_n^* + G_n + \epsilon_n
\]

where \( \epsilon_n \) tends to zero as \( n \to \infty \) and

\[
W_n^* = \frac{V_n - \sum \eta_i(n)}{\sigma\nu(H) \sqrt{1 - R u.v.(H)}} - \frac{U_n - \sum \xi_i(n)}{\sigma u.v.(H) \sqrt{1 - R u.v.(H)}},
\]

\[
G_n = \left( \frac{\Delta}{\delta} - R u.v.(H) \frac{\delta}{\delta} \right) / \sqrt{1 - R u.v.(H)}.
\]

Here \( \Delta \) and \( \delta \) represent the arithmetic means of the \( \Delta_i \) and the \( \delta_i \), respectively. The other two new symbols \( \delta \) and \( \delta \) are defined so that their squares are equal to the arithmetic means of the variances of \( u_1, u_2, \ldots, u_n \) and of \( v_1, v_2, \ldots, v_n \), respectively.

It will be realized that in dealing with \( W_n^* \) under the corresponding hypothesis \( h_n \), we are faced with a sequence \( \{W_n^*\} \) of normed sums of independent random variables satisfying the conditions of the central limit theorem. It follows that

\[
\lim_{n \to \infty} (P\{W_n \geq \nu(\alpha) | h_n\} - P\{W_n^* \geq \nu(\alpha) - G_n\}) = 0
\]

and, finally,

\[
\lim_{n \to \infty} \left( P\{W_n \geq \nu(\alpha) | h_n\} - \frac{1}{\sqrt{2\pi}} \int_{\nu(\alpha)-G_n}^{\infty} e^{-t^2/2} dt \right) = 0.
\]
Consequently, an approximate value of the power function of an asymptotic test based on the criterion $W$, can be obtained from the tables of the normal integral as follows. We compute $G_*$, subtract it from $v(\alpha)$ and read from the table the value of the probability that the normed variable, with expectation zero and variance unity, will exceed the difference $v(\alpha) - G_*$. This probability is, then, the asymptotic value of the power function sought. Thus, the whole problem reduces to the calculation of $G_*$.

Suppose we contemplate a particular number $n_1$ of observations on $u$ and $v$, for which all the necessary computations have been performed, and also a particular value $\theta_1 > \theta_0$ of the parameter under test for which the approximate value of the power function is desired. Primarily, we shall be interested in this power corresponding to the number of observations available, that is, $n_1$. However, in many cases it is interesting to inquire what the power would have been of the test if, instead of $n_1$, we had some other number $n$ of observations, perhaps twice as many, and so forth. Since the answer depends not only on $n$ but also on the means, variances and covariances of the pairs $(u_i, v_i)$ to be observed, it seems natural to study the situation on the assumption that these means, variances and covariances will have the same average values as in the given set of $n_1$ observations.

We consider the hypothesis specifying $\theta = \theta_1$ as an element of an infinite sequence $\{h_n\}$, with $h_n$ asserting that $\theta = \theta_n$, where

$$\theta_n = \theta_0 + (\theta_1 - \theta_0) \sqrt{\frac{n_1}{n}}. \tag{62}$$

With reference to (54), this means setting

$$\delta_i = \frac{\partial E(u_i | \theta)}{\partial \theta} \bigg|_{\theta = \theta_0} (\theta_1 - \theta_0) \sqrt{\frac{n_1}{n}}, \tag{63}$$

$$\Delta_i = \frac{\partial E(v_i | \theta)}{\partial \theta} \bigg|_{\theta = \theta_0} (\theta_1 - \theta_0) \sqrt{\frac{n_1}{n}}.$$

As a result, the value $G_{n_1}$, corresponding to $n = n_1$ and to the specified value $\theta_1 > \theta_0$, is, say,

$$G_{n_1}(\theta_1) = \sqrt{n_1} \left(\theta_1 - \theta_0\right) \frac{\bar{\bar{w}}' - R_{u_v} \bar{\bar{x}}'}{\sqrt{1 - R_{u_v}^2}}. \tag{64}$$

where $\bar{x}'$ and $\bar{\bar{w}}'$ represent, respectively, the arithmetic means of the derivatives

$$\bar{x}' = \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{\partial E(u_i | \theta)}{\partial \theta} \bigg|_{\theta = \theta_0}, \tag{65}$$

$$\bar{\bar{w}}' = \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{\partial E(v_i | \theta)}{\partial \theta} \bigg|_{\theta = \theta_0}.$$

In order to obtain the value of $G_{n_1}(\theta_1)$ corresponding to the same $\theta_1$ and to
the same average properties of the \( u \) and \( v \), but to an arbitrary value of \( n \), it is sufficient to substitute this \( n \) into (64) for \( n_1 \).

Returning to the discussion of the sequences of critical regions \( \{ w_1(n) \} \) and \( \{ w_2(n) \} \) devised for testing the hypothesis of homogeneity of the neutral \( V \)-particles against the sets of alternatives \( \Omega_1 \) and \( \Omega_2 \), respectively, we see that the calculation of the asymptotic power functions involves the following.

(a) The evaluation, on the hypothesis tested, \( H_0 \) of the expectations of the variances and of the covariances of the pairs \( (X_i, X_i') \) and \( (X_i, \exp (-\lambda (\rho - 1) X_i)) \).

(b) The evaluation of the expectations of the three variables concerned, \( X_i, X_i' \) and \( \exp (-\lambda (\rho - 1) X_i) \) on the appropriate alternative hypotheses, and of the derivatives of these expectations with respect to the parameter concerned, either \( v^2 \) or \( \gamma \). The values of the derivatives needed are those at \( v = 0 \) and \( \gamma = 0 \).

The computations are tedious but routine and we shall limit ourselves to several formulae for the various expectations. We have

\[
E(X_i^n \mid H_0) = \frac{m!}{\lambda^n (1 - e^{-\lambda T_i})} \left(1 - e^{-\lambda T_i} \sum_{k=0}^{n} \frac{(\lambda T_i)^k}{k!}\right) \text{ for } m = 0, 1, 2, \ldots.
\]

Using this formula it is easy to compute the expectations of \( X_i \) and \( X_i' \), their variances and their covariance,

\[
E(e^{-\lambda(x-1)X_i} \mid H_0) = \frac{1 - e^{-\lambda(1 + m(\rho - 1)T_i)}}{[1 + m(\rho - 1)](1 - e^{-\lambda T_i})},
\]

\[
E(X_i e^{-\lambda(x-1)X_i} \mid H_0) = \frac{1 - e^{-\lambda\rho T_i}(1 + \lambda T_i)}{\lambda \rho (1 - e^{-\lambda T_i})}.
\]

Formulæ (67) and (68) combined with (66) provide the expectations, the variances and the covariance of the pair \( (X_i, \exp (-\lambda (\rho - 1) X_i)) \). On the alternative hypotheses, we need only the expectations of \( (X_i, X_i') \), based on (11) and of \( X_i \) and \( \exp (-\lambda (\rho - 1) X_i) \), based on (13). We have the recurrence formula,

\[
E(X^n \mid \lambda, v^2, T) = \frac{1}{1 - m v^2} \left(\frac{m \lambda E(X^{n-1} \mid \lambda, v^2, T) - \frac{T^n}{(1 + \lambda v^2 T)^{1/2}}}{1 - (1 - \gamma) e^{-\lambda T} - \gamma e^{-\lambda T}}\right),
\]

which solves the problem for the first pair of variables. For the second pair of variables, we have

\[
E(X \mid \lambda, \rho, \gamma, T)
\]

\[
= \frac{1 - \gamma}{\lambda_1} \left[1 - e^{-\lambda_1 T} (1 + \lambda_1 T)\right] - \frac{\gamma}{\lambda_2} \left[1 - e^{-\lambda_2 T} (1 + \lambda_2 T)\right]
\]

\[
= \frac{1 - (1 - \gamma) e^{-\lambda T} - \gamma e^{-\lambda T}}{1 - (1 - \gamma) e^{-\lambda T} - \gamma e^{-\lambda T}}.
\]
where \( \lambda_1 \) and \( \lambda_2 \) are connected with \( \lambda \) and \( \gamma \) by means of formulae (12). Also

\[
\begin{align*}
\langle 71 \rangle \quad E(e^{-(\lambda_1 + \lambda_2)x} | \lambda, \rho, \gamma, T) &= \frac{1 - \gamma}{\rho + (\rho - 1)^2 \gamma} \left[ 1 - \exp \left\{ -\lambda \frac{\rho + (\rho - 1)^2 \gamma}{1 + (\rho - 1)\gamma} T \right\} \right] \\
&+ \frac{2\rho - 1 + \gamma(\rho - 1)^2}{1 - (1 - \gamma)e^{-\lambda_1 T} - \gamma e^{-\lambda_2 T}} \left[ 1 - \exp \left\{ -\lambda \frac{2\rho - 1 + \gamma(\rho - 1)^2}{1 + (\rho - 1)\gamma} T \right\} \right].
\end{align*}
\]

12. Some Numerical Results Relating to the Data in Table I

In this section we give the numerical results of the application of the above theory to the data exhibited in Table I. I am indebted to Miss Mandakini J. Sane for deducing formulae leading to the final expressions of the test criterion \( W_0 \) of formula (56) and to the expression of \( G_0 \) of formula (64). Numerical com-

![Fig. 3. Asymptotic power function of the test of the hypothesis \( H_0 \) against the alternative \( \Omega_1 \). Quantity measured on the horizontal axis is \( \rho \), the coefficient of variation of \( A \).](image-url)
putations were performed mostly by Mrs. Jeanne Lovasich. I am very grateful to both Miss Sane and Mrs. Lovasich for their friendly help.

The values of the test criteria $W_*$, computed separately for the set of alternatives $\Omega_1$, as in (49), and separately for the set of alternatives $\Omega_2$, from an analogous formula, with $p = 10$, are, say,

$$W_{20}(\Omega_1) = -0.7605$$
$$W_{20}(\Omega_2) = 0.3754$$

respectively: Remembering that, if the hypothesis tested is true, then these numbers are particular values of random variables with zero means and having an approximately normal distribution with unit variance, and also remembering that $\nu(0.05) = 1.645$, it is seen that the data of Table I do not indicate the rejection of the hypothesis that the particles concerned are a sample from a homogeneous class.

Figures 3 and 4 give the graphs of the asymptotic power functions of the

![Diagram](image-url)

Fig. 4. Asymptotic power function of the test of the hypothesis $H_0$ against the alternative $\Omega_1$. Quantity measured on the horizontal axis is $\gamma$, the proportion of short-lived particles in the class.
two tests, corresponding to the level of significance \( \alpha = 0.05 \). In Figure 3 the power is plotted against \( v \), the coefficient of variation of the value of the decay constant \( \Lambda \), supposed to vary from one particle to the next in accordance with the density (8). In Figure 4, the independent variable is \( \gamma \), the proportion of short-lived particles supposed to be mixed with a predominant type of particles with longer life. In each case there are four curves. One of them, the lowest, corresponds to the amount of data actually available, with \( n_1 = 26 \) particles. The three others correspond to \( n = 100, 200 \) and \( n = 1000 \) and are meant to illustrate the situation which would prevail if the experimental results were substantially richer. It is seen that, in both cases, the values of the power of the tests, corresponding to moderate but significant deviations from the hypotheses of homogeneity, are very small. In fact, should one of the hypotheses of the set \( \Omega \), be true, with the coefficient of variation \( v \) equal to as much as fifty per cent, with the available amount of data, the chance of detecting that the hypothesis of homogeneity is at all false would be only 0.067!

Referring to Figure 1, which exhibits the probability densities of \( A \) corresponding to \( v = 0.1, v = 0.2 \) and \( v = 0.3 \), one is inclined to wonder how it could happen that a statistical test especially designed to be sensitive to detect this kind of variability of \( A \) could be as ineffective as shown in Figure 3. The answer to this question is that the test discussed does not deal directly with the variation of \( A \) (since this is unobservable) but with that of the observable random variables \( X_1, X_2, \cdots, X_n \). The probability density of any one such variable, computed on the assumption that the decay parameter of a particle is not constant but varies from one particle to the next, with densities exhibited in Figure 1, is given by formula (11). Figure 5 gives plots of this density corresponding to several values of the coefficient of variation \( v \) including \( v = 0 \). It is seen that, if the variability of the decay parameter is as much as 30 percent, the resulting change in the density of the observable \( X \) is of the order of magnitude of a “gnat’s eyelash.” This gives an intuitive explanation why any lack of homogeneity of \( V \)-particles is so difficult to detect.

13. Does the Failure of a Test to Reject the Hypothesis Tested Constitute a Confirmation of This Hypothesis?

In connection with the above numerical results it may be useful to return to the general discussion of principles outlined in the first part of the present paper. Specifically, I am concerned with the term “degree of confirmation” introduced by Carnap. More specifically, I am interested in the connotations that this term is likely to suggest, irrespective of Professor Carnap’s definition.

We have seen that the application of the locally best one-sided test to the data of Table I failed to reject the hypothesis that the 26 particles concerned come from a class that is homogeneous with respect to the decay parameter. The question is: does this result “confirm” the hypothesis that the particular class of particles is actually homogeneous? This question is pertinent because
in some sections of scientific literature the prevailing attitude is to consider that once a test, deemed to be reliable, fails to reject the hypothesis tested, then this means that the hypothesis is "confirmed."

The answer to the question depends very much on the exact meaning given to the words "confirmation," "confidence," etc. If one uses these words to describe one's intuitive feeling of confidence in the hypothesis tested $H_0$, then Figures 3 and 4 show that the attitude described is dangerous. As we emphasized a few lines above, the chance of detecting the presence of variability in the decay parameter, when only 26 observations are available, is extremely slim, even if the variation is considerable. Therefore, the failure of the test to reject $H_0$ cannot be reasonably considered as anything like a confirmation of $H_0$. The situation would have been radically different if the power function corresponding to $v = 0.1$ were, for example, greater than 0.95.

The general conclusion is that it is a little rash to base one's intuitive confidence in a given hypothesis on the fact that a test failed to reject this hypothesis. A more cautious attitude would be to form one's intuitive opinion only after studying the power function of the test applied.

However, the vague terms "confirmation" and "confidence" may be used to describe the choice of action of the physicist concerned with the classification of the $V$-particles. It will be remembered that Part II of this paper began with the contemplation of what course of action the physicist should choose. Clearly, however strong the temptation may be, he should refrain from any announcement of discovery that the decay parameter of the $V$-particles is not constant. Also, it is quite apparent that attempts at discovering variability in the decay constant, using data of the kind exhibited in Table I, should be postponed until such time as further observations will accumulate substantially more data. Finally, if it is necessary to use the probability density of the decay time $X$ in some further theoretical work, there is nothing against using the form asserted by the hypothesis tested $H_0$, which is much simpler than those asserted by the alternatives. This does not mean any sort of belief in the hypothesis of homogeneity, but merely reflects the conclusion drawn from the power functions in Figures 3 and 4 and of the comparison in Figure 5, to the general effect that, with moderate amounts of observations, it makes very little difference whether $H_0$ or one of the alternatives contemplated is true.


A symposium paper, particularly a paper on statistics presented at a symposium on applied mathematics cannot be considered complete without at least a cursory glance at the current developments in our discipline. There are many and they are important. However, because of the already considerable size of the paper, they must be treated in telegraphic style with unavoidable but regrettable omissions. Of the papers written with reference to a given direction or by a particular distinguished author, only a sample of one or two are mentioned.
After a start, some twenty years ago, from the simplest problems of testing statistical hypotheses and through the theory of interval estimation, the modern theory of statistics is concerned with substantially broader and more difficult multidecision problems. Also, along with fixed sample problems, we are now concerned with sequential problems, in which the size of the sample is a random variable. In all this, the influence of the talent of the late Professor Abraham Wald [23, 24] is likely to be felt for more than a generation. The most important representative of the Wald school appears to be Wolfowitz [25]. A remarkable exposition of further developments was recently given by Blackwell and Girshick [26]. Several new pathfinding papers are due to Stein [27]. A special aspect of the multidecision problem received a most satisfactory treatment by Scheffé [28] and Tukey [29]. A fundamental paper by Tukey is still in preparation. The ideas originated in these publications are likely to generate a substantial literature of which we mention the contribution of Roy and Bose [30].

A particular development in the theory of statistics is concerned with multivariate analysis of normally distributed correlated variables. Here the contributions of Hotelling [31], C. R. Rao [32] and T. W. Anderson [33], are of considerable importance.

A new direction of study, already counting many contributions by a number of authors, was started by the paper by Robbins and Monro [34]. The problem is a sequential experiment which will generate a stochastic process converging to an unknown number, such as a root of an equation.
Nonparametric tests form an important direction of modern statistical theory. In this domain, the contributions of Hoeffding, and Lehmann are very valuable. We will mention papers [35] and [36].

A separate and quite new direction of study is the problem of statistical inference in relation to stochastic processes. The pioneer work is due to Grenander [37] and Rosenblatt [38].

The comparison of the above literature with the general tone of twenty years ago suggests that the present stage of development of mathematical statistics may be compared with what happened in analysis after the appearance of Weierstrass. The prevailing tendency is directed towards much greater precision than was ever contemplated. Also, a number of contributions are given to the analysis of the basic concepts and to abolishing established routine of thought. In this connection we mention the papers by Barankin and Gurland [39], Hammersley [40], by J. L. Hodges, Jr. [41] by LeCam [42] and by Sverdrup [43].

Lastly, one should mention several series of papers on the fringes of theoretical statistics proper, directly concerned with stochastic models of various natural phenomena. Important contributions to the methodology of estimating biological populations are due to Chapman [44]. Two papers concerned with medical tests must be mentioned, one by Chiang [45] and an older one by Muench [46]. Important developments concerned specifically with bioassay, but having a much more general theoretical significance are due to Berkson [47]. Miss Fix has contributed to theory of follow-up studies of patients [48]. Miss Bates dealt similarly with the problem of contagion in accidents [49]. Interesting contributions to the statistical theory of learning are due to Bush and Mosteller [50]. Finally, there is a series of papers of Scott and Shane, with the participation of the present writer, published in nonstatistical journals and given to the problem of the spatial distribution of galaxies. The ultimate aim is to find evidence for deciding whether our universe is expanding or not. The two basic papers of this series are [51] and [52]. The most recent contribution [53] connected with this series, due to McVittie, represents an application to the same problem of kinematical and general relativity.

**Bibliography**


Received September 22, 1954.