The Continuum of Inductive Methods

By

RUDOLF CARNAP

THE UNIVERSITY OF CHICAGO PRESS

PREFACE

The purpose of this monograph is the development of a new approach to the study of inductive methods; they are methods for determining the degree of confirmation (or probability in the inductive sense) of a given hypothesis on the basis of a given body of evidence. It is shown that from any method of this kind we can derive a method for estimating unknown magnitudes (here restricted to frequencies). In distinction to *deductive* logic, where there is practically general agreement on method, in *inductive* logic we find a great variety of incompatible methods. They are here incorporated into an infinite system of possible methods, called the continuum of inductive methods. Within this system, the methods are characterized by parameters. Essentially, one parameter is sufficient for a complete characterization of each method. A procedure is developed for comparing various methods, not on the basis of philosophical arguments offered for them or of intuitive judgments on their immediate plausibility, but rather from the objective point of view of their success in various possible universes (here restricted to those describable in a certain simple language). This leads to a procedure for determining the optimum inductive method in any given possible universe. The comparative analysis comes to the result that some widely used methods for estimating frequencies (e.g., Reichenbach's principle of induction, R. A. Fisher's maximum likelihood principle, the principle of unbiased estimates, and Wald's minimax principle) have certain disadvantages, which are avoided by other inductive methods in the continuum.

The present investigations are based on a conception of inductive logic which was systematically developed in *Logical Foundations of Probability* (Vol. I of *Probability and Induction* (1950)); it is explained in nontechnical terms in *The Nature and Application of Inductive Logic* (1951), which is a reprint of six sections from the former book. Knowledge of either of these books is, however, not presupposed; the concepts used in this monograph will be explained so that it can be read independently.

The substance of this monograph is to become part of Volume II of *Probability and Induction*. I shall welcome critical reactions.

I want to thank Professors Gerhard Tintner and John Myhill, who read an earlier version of the manuscript and made many helpful, critical comments.

RUDOLF CARNAP

UNIVERSITY OF CHICAGO
November 1951

CONTENTS

SUMMARY

Part I. The λ-system. Inductive methods are here understood as methods of confirmation or of estimation. A method of confirmation determines the degree of confirmation of any hypothesis h (on the basis of anybody of evidence e (c(h,e)). A method of estimation determines an estimate for a magnitude x concerning the universe (or population) on the basis of evidence e concerning a sample from the universe (e(x,e)). The discussion in this monograph is restricted to some simple language-systems (in technical terms, the functional calculus of first order with identity). Measurable quantities are not expressible in these systems, but only properties of individuals; the only quantities occurring are cardinal numbers and relative frequencies of properties. Therefore, the discussion of estimation is here restricted to the estimation of the relative frequency of a property M in a finite universe (e(rf,M,e)). It is shown that there is a one-one correspondence between the confirmation-functions c and the estimate functions for relative frequency e. A given function c determines completely the corresponding function e and vice versa. A complete inductive method consists of a function c and the corresponding function e. The purpose of this monograph is to construct an ordered system of the inductive methods, a system which contains not only those methods which have so far been proposed or discussed by authors or probability or statistics but also an infinity of other methods. This system is a continuum; the transition from one method to another consists in the continuous variation of certain features. A system of this kind is practically useful only if each method can be characterized by a small number of parameters. An analysis of the c- and e-functions leads to the surprising result that one parameter λ is sufficient; in other words, the continuum of inductive methods is one-dimensional. The possible values of λ are the nonnegative real numbers and ∞. Any value of λ determines completely a confirmation function c_λ and the corresponding estimate-function e_λ in the following sense. If a language-system L and a number λ are given, we can calculate the value of $c_\lambda(h,e)$ for any pair of sentences h,e (e must not be self-contradictory) in L and the value of $e_\lambda(\text{rf},\text{M},e)$ for any M and e in L. The meaning of λ may be indicated in nontechnical terms as follows. The values of c and e are influenced by two particular factors, one of which is of an empirical nature, the other of a logical nature. λ is the relative weight given to the logical factor, in comparison to the empirical factor, in the determination of the values of c_λ and e_λ. Let e_M describe a sample of s individuals of which s_M have the property M; let h_M be the hypothesis that an individual outside the sample has the property M. It is shown that, for any λ, $e_\lambda(\text{rf},M, e_M) = c_\lambda(h_M, e_M)$. For λ= 0, it is found that (1) $c_0(h_M, e_M) = s_M/s$, and (2) $e_0(\text{rf},M, e_M) = s_M/s$. We call (1) the straight rule of confirmation and (2) the straight rule of estimation. The latter says that the estimate of the relative frequency of M in the universe is equal to the relative frequency of M in the sample. The great majority of contemporary statisticians accept this rule; it is implied, e.g., by R. A. Fisher's principle of maximum likelihood, by the principle of the preferability of unbiased estimate-functions, and by Reichenbach's rule of induction. Laplace's rule of succession, if we modify it in a suitable way so as to eliminate its inconsistency, leads to λ = 2. Some theoreticians (e.g., C. S. Peirce, Keynes, and Wittgenstein) have declared that all individual distributions (called 'constitutions') should be given equal probability a priori. This principle leads to λ = ∞. It is shown that the values of $c_\infty(h_M, e_M)$ and $e_\infty(\text{rf},M, e_M)$ are independent of e_M. To let the esti-

mate of the relative frequency of *M* in the universe not be influenced by past observations concerning *M* obviously cannot be regarded as sound inductive reasoning. Therefore, the method characterized by $\lambda = \infty$ is unacceptable. It is found that the method of the straight rule ($\lambda = 0$) also has certain inadequacies, though not such serious ones as the method of $\lambda = \infty$.

Part II. Comparison of the success of inductive methods. Let a language-system L for a finite universe and in it a state-description *k* be given, i.e., a complete description of a possible state (not necessarily the actual state) of the universe. We regard an inductive method characterized by λas the more successful in *k*, the smaller the mean square error of the estimates supplied by the function e_λ, for the relative frequency of the strongest properties expressible in L (the *Q*-properties) on the basis of all possible samples of the fixed size *s*. For any state-description *k*, we can easily determine that value of λ for which the mean square error has its minimum; we denote this value by 'λ^Δ'. The method characterized by λ^Δ is called the optimum method for *k*. It is found that λ^Δ is independent of *s*; thus, for any given *k*, there is one optimum method of estimation, irrespective of the size of the samples to which it is applied. And, conversely, for any method characterized by any λ, there are state-descriptions for which this method is the optimum one. If a state-description is homogeneous, i.e., all individuals are completely alike so that there is maximum uniformity or degree of order, then $\lambda^\Delta = 0$; in other words, the straight rule is the optimum method. If a state-description is such that the *Q*-properties have equal frequencies (minimum uniformity), then $\lambda^\Delta = \infty$. In any other case, with intermediate uniformity, λ^Δ has an intermediate value, i.e., it is positive and finite. Since an observer knows at any time only a sample, not the whole universe, he cannot determine the optimum method for the actual universe. However, as soon as he observes any nonhomogeneous sample, he knows that the universe is nonhomogeneous. It is shown that he can then determine a positive λ' such that, in the universe as a whole, the mean square error for λ' is smaller than that for $\lambda = 0$; in other words, the method of λ' is, with deductive certainty, more successful in the actual universe than the method of the straight rule. This result seems to be a serious argument against the straight rule, and thereby against the customary preference for unbiased estimate-functions and against the principle of maximum likelihood.

In the Appendix, the estimate-function for relative frequency to which *Wald's minimax principle* leads is examined. It is shown that it has some serious disadvantages. A related function is defined, which is free of the disadvantages and which belongs to the λ-system.

PART I

THE λ-SYSTEM

§ 1. The Situation in the Theory of Inductive Methods

Any reasoning or inference in science belongs to one of two kinds: either it yields certainty in the sense that the conclusion is necessarily true, provided that the premises are true, or it does not. The first kind is that of deductive inference including all transformations or calculations in pure mathematics (arithmetic, algebra, analysis, etc.). The second kind will here be called '*inductive inference*'. Thus this term is used here in a much wider sense than in traditional terminology; it covers all nondeductive inference.

Inductive inference is clearly of fundamental importance in any field of science from physics to history. Any specific investigation in any field is focused in a *hypothesis* or a set of competitive hypotheses. A hypothesis may be a prediction of a single event (the weather for tomorrow, the result of an experiment, the outcome of the next presidential election) or of a general trend (decrease in the rate of death by cancer, increase in unemployment) ; it may be a law, e.g., in physics or, physiology or economics; its form may be deterministic or statistical (in terms of frequencies, averages, or the like). The scientist takes as a basis the observational evidence at hand or makes efforts to gather new evidence relevant for his hypothesis. Then he tries to arrive at a judgment concerning the hypothesis based on the evidence and to decide whether to accept or reject the hypothesis. The judgment is obviously of an inductive nature; the scientist is aware that at any future time when new evidence is found he may have to revise his judgment concerning the status of the hypothesis in question. The new evidence may be favorable or unfavorable or irrelevant to the hypothesis. That it is favorable means that the hypothesis is now confirmed more strongly than before; if the increase in confirmation is large enough, a previously rejected hypothesis may now be accepted. Instead of examining a given hypothesis, the scientist may try to arrive at a quantitative *estimate* for the unknown value of some magnitude. For example, lie may wish to estimate, on the basis of his evidence, the *amount* of rain for tomorrow, the *number* of votes for a certain candidate, the rate of increase in unemployment. Estimation is likewise an inductive procedure, because there is no guaranty that the actual value of the magnitude will turn out to be close, let alone exactly equal, to the estimated value.

This monograph will be concerned with the two kinds of inductive methods just indicated, methods of confirmation and methods of estimation, both taken in a form that leads to numerical values. We say that a person *X* possesses a *method of confirmation* if he is able to determine in some way (not necessarily by explicitly formulated rules), at least in certain cases of a hypothesis *h* and a body of evidence *e*, a numerical value which he regards as the degree of confirmation of *h* with respect to *e*; in other words, if there is a numerical function c(*h*,*e*,) which has values for at least some pairs of sentences *h*,*e* and which represents to *X* the degree of confirmation of *h* on the basis of *e*. (I use the term 'sentence' as synonymous with 'declarative sentence' or 'statement'.) For reasons to be explained soon, our present discussions of *methods of estimation* will be restricted to the estimation of one kind of magnitude, viz., the relative frequency (henceforth abbreviated by 'rf') of a property *M* in a class *K* with respect to a given body of evidence *e*; the elements of *K* are "unobserved", i.e., not described in *e*. We say that *X* possesses a method of estimation of this kind if he knows a way of determining, at least for certain cases of *e*, *M*, and *K*, a numerical value which he accepts as an estimate of rf of *M* in *K* with respect to *e*; in other words, if there is an estimate-function e(rf,*M*,*K*,*e*) which has values for at least some triples *M*,*K*,*e* of the kind specified, and which represents to *X* the estimate of the rf of *M* in *K* on the basis of *e*.

Instead of 'degree of confirmation', the term '*probability*' is often used.[1] Here I shall avoid. the latter term bee: use of its ambiguity. It seems to me that, in scientific contexts, the word 'probability' is chiefly used in the following two senses.[2] In its first sense it denotes a logical relation between two propositions or sentences, expressing the degree of confirmation or strength of support that is given to a hypothesis by a body of evidence. The term is used in this sense by John Maynard Keynes,[3] Harold Jeffreys,[4] C. I. Lewis,[5] Donald Williams,[6] and others; the same

1. R. Carnap, *Logical foundations of probability* (*Probability and induction*, Vol. 1) (Chicago: University of Chicago Press, 1950), henceforth referred to by '[I]'. A second volume, referred to by '[II]', is in preparation.

2. See [I], § 9, or the earlier paper "The two concepts of probability," *Philos. and Phen. Research*, 5 (1915), 513-32, reprinted in: H. Feigl and W. Sellars (eds.), *Readings in philosophical analysis* (New York, 1949).

3. J. M. Keynes, *A treatise on probability* (London and New York, 1921; 2d ed., 1929).

4. H. Jeffreys, *Theory of probability* (Oxford, 1939).

5. C. I. Lewis, *An analysis of knowledge and valuation* (LaSalle, Ill., 1946).

6. D. Williams, *The ground of induction* (Cambridge, Mass., 1947).

holds in my opinion (in agreement with the authors just mentioned) also for the classical authors,[7] especially Bernoulli, Bayes, and Laplace. The word 'probability' in its second sense means an empirical relation between a property *M* and a reference-class *K*, representing the relative frequency in the long run (sometimes defined as the limit in an infinite sequence) of *M* in *K*. It is used in this sense by Richard von Mises,[8] Hans Reichenbach,[9] and in contemporary mathematical statistics (e.g., by R. A. Fisher,[10] Harald Cramér,[11] S. S. Wilks,[12] Jerzy Neyman,[13,14] E. S. Pearson,[14] A. Wald,[15] DI. G. Kendall,[16] and others). It seems to me that there is no point in debating the question as to which of the two is "the right sense" of the word. Both concepts, though fundamentally different, have an important function in science under whatever name; the first has its place in inductive logic and hence in the methodology of science, the second in mathematical statistics and its applications.

Suppose that X_1 uses a method of confirmation represented by the function $c_1(h,e)$, and X_2 uses one represented by $c_2(h,e)$. We say that these two methods are *incompatible* if they lead sometimes to different numerical results; i.e., if there is at least one pair of sentences *h*,*e* such that $c_1(h,e) \neq c_2(h,e)$. Analogously, let X_1 use a method of estimation for rf, represented by $e_1(\mathrm{rf},M,K,e)$, and X_2 one represented by $e_2(\mathrm{rf},M,K,e)$. Here again we regard the two given methods as *incompatible* if they sometimes yield different values, i.e., if there is at least one triple *M*,*K*,*e*, such that $e_1(\mathrm{rf},M,K,e) \neq e_2(\mathrm{rf},M,K,e)$.

If we look at the contemporary situation in the field of inductive logic, the theory of inductive inferences, we notice the remarkable fact that a variety of mutually incompatible inductive methods are proposed and

7. Compare [I], § 12B.

8. R. von Mises, *Wahrscheinlichkeitsecchnung* (Wien, 1931), *Probability, statistics, and truth* (New York, 1939).

9. H. Reichenbach, *The theory of probability* (Berkeley, 1949).

10. R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Philos. Transactions of the Royal Society*, Ser. A, 222 (1922), 309-68; *The design of experiments* (Edinburgh and London, 1935, and later editions).

11. H. Cramér, *Mathematical methods of statistics* (Princeton, 1946).

12. S. S. Wilks, *Mathematical statistics* (Princeton, 1943).

13. J. Neyman, *Lectures and conferences on mathematical statistics* (Washington, D.C., 1938).

14. J. Neyman and E. S. Pearson, "Contributions to the theory of testing statistical hypotheses," *Statistical Research Memoirs*, 1 (1936), 1-37; 2 (1938), 25-57.

15. A. Wald, *On the principles of statistical inference* (Notre Dame, 1942).

16. M. G. Kendall, *Advanced theory of statistics*, Vol. I (London, 1943; 4th ed., 1948), Vol. II (1946; 2d ed., 1948).

discussed by authors of theoretical treatises and applied in practical work by scientists and statisticians. None of the authors is able to convince the others that their methods are invalid. I shall not try to decide the difficult question whether the situation in inductive logic is in this respect fundamentally different from that in deductive logic, including mathematics. One might point out the fact that in the latter field, likewise, we find differences of opinion concerning the validity of certain inferences. (For example, most mathematicians would regard as valid the indirect proof of an existential sentence of the form 'There is a number with the property *P*' based on the derivation of a contradiction from the sentence 'All numbers have the property non-*P*; but intuitionists like L. E. J. Brouwer and H. Weyl would reject it.) Furthermore, there are alternative systems of multivalued logic. On the other hand, one might perhaps regard differences of these kinds as based on different interpretations of the basic logical terms rather than as genuine differences in opinion, i.e., incompatible answers to the same question. Whatever the solution of this philosophical problem may be, it seems to me that there can hardly be any doubt about the historical fact that, as matters stand today, the differences of opinion concerning the validity of inductive methods go much deeper and are much more extensive in their scope than the differences in deductive logic. Most of the latter differences seem to reduce either to the case that the same sentence is interpreted in two different ways or to the case that only one of the two parties assigns any interpretation to a given formula, while the other declares it meaningless, i.e., not a genuine sentence. (This is, for example, done by intuitionists with respect to the existential sentence in the earlier example as long as no instance for it is constructed.) An analogous situation occurs frequently in inductive logic: one method assigns a value of degree of confirmation to a given pair of sentences *h*,*e*, while another method assigns no value and thus declares 'c(*h*,*e*)' as meaningless. A difference of this kind is rather harmless. But here, in the field of inductive logic, we frequently find differences which are far more critical, namely, cases in which two methods assign different values. It even occurs sometimes that one method assigns to a given pair *h*,*e* the maximum value of c (1, according to a customary convention), while another method assigns to the same pair the minimum value (0). [For example, let λ be a universal factual sentence in an infinite universe (e.g., 'All ravens are black'), and *e* describe a finite sample of positive instances (e.g., a sample of ten ravens, all of them black). Then one of the methods to be discussed later (the straight rule) takes c(*h*,*e*) = 1, and another one (c*) assigns the value 0,] In view of the situation described, another deep philosophical problem arises. The fact that there are at present time irreconcilable differences may be merely a matter of historical contingency due to the present lack of knowledge in the field of inductive logic. If so, it would be conceivable that at some future time, on the basis of deeper insight, all will agree that a certain inductive method is the only valid one. On the other hand, it may be that the multiplicity of mutually incompatible methods is an essential characteristic of inductive logic, so that it would be meaningless to talk of "the one valid method." This does not necessarily imply that the choice of an inductive method is merely a matter of whim. It may still be possible to judge inductive methods as being more or less adequate. However, questions o this kind would then not be purely theoretical but rather of a pragmatic nature. A method would here be judged as a more or less suitable instrument for a certain purpose. I shall not try to discuss this problem here still less to solve it; but I may indicate that at the present time I am more inclined to think in the direction of the second answer. In the later part of this monograph some of the points which are relevant for judgment concerning the adequacy of inductive methods will be discussed.

2. Our Task: The Construction of a Continuum of Inductive Methods

In view of the multiplicity of inductive methods, the scientist who wants to apply a method is compelled to make a choice. Now the main point is that the scientist should be given, as a basis for his choice, not simply the list of those inductive methods which have been proposed or considered so far but *a systematic survey of all possible inductive methods*. It would not do merely to enlarge the list, which now comprises perhaps a dozen or so methods, by adding some dozens or hundreds of new methods. *The system of possible inductive methods is a continuum*. The few methods now known represent just a few points picked out from the continuum by the hazards of historical and psychological circumstances.

It will be our task to order the possible inductive methods into a system in such a manner that the nature of any particular method determines its place in the system. If possible, it would be especially useful to find a system for which the inverse also holds: the position of a method in the system, described by the values of the co-ordinates or parameters, uniquely determines the method. An n-dimensional system of this kind would fulfil the following two conditions: (1) if any method is given, we can determine for it a particular set of n numerical values $p_1, p_2, \ldots, p_n$, called the parameters of the method; and (2) any such set of n parameter values characterizes uniquely and completely an inductive method with respect to the values of c and e. The meaning of this complete characterization will later be made more exact (§ 10).

It is clear that it would have great advantages if a parameter system of this kind could be found with a number *n* which was not only finite but sufficiently small for easy practical manipulation. The investigation of entities of some kind in any branch of science applying the quantitative method has often been greatly helped by the introduction of parameters characteristic of the entities. A procedure of this kind not only facilitates the comparison of entities already known but supplies a convenient survey of the totality of possible entities of the kind in question and thus leads to the study of new entities by varying the parameter values of known ones. Special advantages result if the system is continuous, that is, if all real numbers or those of given intervals serve as parameter values. In this case, it is possible to study the change in the nature of the entities due to a continuous variation of the parameters; the convenient and fruitful methods of the differential and integral calculus are applicable; it is easily possible to find those entities which possess a certain characteristic in a minimum or maximum degree; and so on. In the case of inductive methods a parameter system would enable us not only to compare any two of the historically given methods in a more exact way than was possible so far but also to study new methods quantitatively. It would be easy to discover one or several new methods which fulfil any given condition or which are most useful for a specified purpose.

It is the *purpose of this monograph* to discover a way which leads to the construction of a parameter system of inductive methods. The resulting system will be called the λ-system. No attempt is made to include in the system all conceivable inductive methods. This would be useless, because most of the methods which one could arbitrarily construct in the form of a c- or e-function would be entirely inadequate for the purpose of inductive application. No one would even consider them as possible methods of confirmation or of estimation because their use would be in conflict with the implicitly accepted basic principles of inductive reasoning (e.g., any method of estimation represented by a function e(rf,*M*,*K*,*e*) of such a kind that its value would be lower, the higher the rf of *M* in the sample described in *e*). However, the methods contained in the λ-system, called the λ -methods, do not constitute a narrow selection but an infinite class and, moreover, a continuum which contains, among others, the historically known methods and those others which are related to them by sharing with them their fundamental features but differ from them through a continuous variation of certain of their characteristics.

3. Preliminary Explanations

Before we begin the analysis of inductive methods which is to lead to the λ-system, a few concepts have to be explained. These explanations are necessarily brief and not quite exact but, I hope, practically sufficient for an understanding of this monograph. For exact definitions and detailed explanations the reader is referred to [I]. A more detailed treatment of the λ-system will be given in [II].

The present monograph, like my previous publications, studies inductive logic in application not to the whole language of science but to simple language systems L.[17] A system L_N^π of this kind contains a finite number N of *individual constants* ('a_1', 'a_2', . . . , a_N'), which stand for *individuals* (things, events, or positions), and a finite number π of *primitive predicates* ('P_1', 'P_2', . . . , 'P_π') which designate *primitive properties* of the individuals. For the sake of simplicity, we shall disregard in the present context the infinite system L_∞. This does not involve an essential restriction in generality, because we assume that in any inductive method dealing with an infinite universe of discourse the value of c is the limit of the values in finite systems L_N for $N \rightarrow \infty$,[18] and analogously for the values of e. It seems that this assumption is in accord with the historically given methods.

An *atomic sentence* consists of a primitive predicate and an individual constant (e.g., 'P_2a_5', 'the individual a_5 has the property P_2'). Other *molecular sentences* are formed out of atomic ones with the help of the customary *connectives* of negation ('~', 'not'), disjunction ('$\vee$', 'or'), and conjunction ('.', 'and'). Analogously, other molecular predicate expressions are formed out of primitive predicates with the help of connectives (e.g., '$P_1 . \sim P_2$', 'P_1 and not P_2'). A sentence consisting of a predicate or a compound predicate expression and an individual constant is called a *full sentence* of the predicate or predicate expression or of the property designated by it. The systems contain *individual variables* with *quantifiers* for the formulation of universal and existential sentences ('for every individual, . . .', 'then is an individual such that …'). '=' is a sign of identity between individuals. With its help, numerical statements can be formulated (e.g., 'there are at least two individuals with the property M' is rendered in the form 'there are individuals x, y, such that x is M and y is M and x is not identical with y'). No other primitive symbols occur in the systems L. Thus measurable quantities (like length, mass, etc.) cannot be expressed but absolute frequencies (cardinal numbers of classes or properties) and hence relative frequencies can be expressed. For this reason we restrict

17. [I], §§ 15, 16. 18. See [I], §§ 54A, 56.

the discussion of estimation to the estimation of relative frequency (rf). This is, of course, a very narrow restriction. It is, however, to be noted that rf belongs to the most basic and most important magnitudes for statistical inference and estimation. Certain coefficients measuring the dependence of properties, e.g., the coefficient of association of two properties and, for a contingency table, Pearson's coefficient of mean square contingency or other related coefficients, are, of course, determined by the relative frequencies of the properties involved. Whether and how the basic ideas of this monograph can be transferred to more comprehensive language-systems involving measurable quantities remains a question for the future.

Any sentence is either *L-true* (logically true, analytic, e.g., '$Pa \vee \sim Pa$') , or *L-false* (logically false, self -contradictory, e.g., '$Pa . \sim Pa$'} or *factual* (synthetic, e.g., '$P_1a_3 \vee P_2a_6$').[19] Correspondingly, the molecular predicate expressions and the properties designated by them are divided into three kinds: (1) *L-universal* (e.g., '$P \vee \sim P$'; every full sentence is L-true), (2) *L-empty* (e.g., '$P . \sim P$'; every full sentence is L-false), (3) *factual* (e.g., '$P_1 . \sim P_2$'; every full sentence is factual).[20] Furthermore, logical relations among sentences can be defined, e.g., *L-implication* (logical implication, deducibility), *L-equivalence* (mutual logical implication), and *L-exclusion* (mutual logical incompatibility). We shall use 't' as an abbreviation for a particular *tautology* (i.e., an L-true molecular sentence, e.g., '$Pa \vee \sim$Pa').

Of especial importance are those molecular properties which are defined by a conjunction in which every primitive predicate occurs either unnegated or negated. They are called *Q*-properties and are designated by 'Q_1', 'Q_2', … , 'Q_K'.[21] Obviously, their number $\kappa = 2^\pi$. For example, for $\pi = 3$ there are 8 *Q*'s:

$P_1 . P_2 . P_3$	Q_1	$\sim P_1 . P_2 . P_3$	Q_5
$P_1 . P_2 . \sim P_3$	Q_2	$\sim P_1 . P_2 . \sim P_3$	Q_6
$P_1 . \sim P_2 . P_3$	Q_3	$\sim P_1 . \sim P_2 . P_3$	Q_7
$P_1 . \sim P_2 . \sim P_3$	Q_4	$\sim P_1 . \sim P_2 . \sim P_3$	Q_8

The *Q*-properties of a system L are the strongest non-L-empty properties expressible in L. If *M* is any non-L-empty molecular property in L, then *M* is uniquely analyzable into a disjunction of *Q*'s or one *Q*. Let *w* be the number of these *Q*'s. *w* is called the (logical) *width* of *M*; if *M* is L-empty,

19. [1], § 20.
20. [1], § 25.
21. [1], § 31.

we assign to it the width $w = 0$. If w is the width of M, w/κ is its *relative width*.[22] The following table gives examples of predicate-expressions in the system L^3, i.e., for $\pi = 3$, $\kappa = 8$.

Predicate Expression	Logical Nature	Transformed in Terms of Q's	Width w	Relative Width w/K
$P_2 . \sim P_2$	L-empty		0	0
$P_1 . P_2 . P_3$		Q_1	1	⅛
$P_1 . \sim P_2$		$Q_1 \vee Q_2$	2	¼
$P_1 . (P_2 \vee P_3)$		$Q_1 \vee Q_2 \vee Q_3$	3	⅜
P_1	Factual	$Q_1 \vee Q_2 \vee Q_3 \vee Q_4$	4	½
$P_1 \vee (P_2 . P_3)$		$Q_1 \vee Q_2 \vee Q_3 \vee Q_4 \vee Q_5$	5	⅝
$P_1 \vee P_2$		$Q_1 \vee Q_2 \vee Q_3 \vee Q_4 \vee Q_5 \vee Q_6$	6	¾
$P_1 \vee P_2 \vee P_3$		$Q_1 \vee Q_2 \vee Q_3 \vee Q_4 \vee Q_5 \vee Q_6 \vee Q_7$	7	⅞
$P_1 \vee \sim P_1$	L-universal	$Q_1 \vee Q_2 \vee Q_3 \vee Q_4 \vee Q_5 \vee Q_6 \vee Q_7 \vee Q_8$	8	1

A *state-description*[23] in a system L_N is a conjunction containing as components for every atomic sentence either it, or its negation, but not both, and no other sentences. Thus a state-description describes completely a possible state of the universe of discourse consisting of the N individuals of the system in question. A state-description can be transformed into another, L-equivalent form, called its Q-form; this is a conjunction of N Q-sentences, one for each of the N individual constants of the system. For any sentence j in the system, the class of those state-descriptions in which j holds, i.e., each of which L-implies j, is called the *range* of j.[24] The range of j is null if and only if j is L-false; in any other case j is L-equivalent to the disjunction of the state-descriptions in its range.

§ 4. The Characteristic Function of an Inductive Method

In preparation for the parameter system to be constructed later, we shall now make a study of possible inductive methods. This will lead to a class of mathematical functions such that every inductive method can be characterized by exactly one function of this class.

We begin with *methods of confirmation*; methods of estimation will analyzed later (§ 6). A method of confirmation is represented by a function c(h,e). This is not a mathematical function; although its values are numbers, the two arguments are not numbers but sentences. Our first aim is to find another way of characterizing any given method of conformation a way that uses a mathematical function instead of a c-function.

22. [1], § 32.

23,24. [1], § 18.

A c-function assigns a real number to every pair of sentences *h*,*e* of its domain. We presuppose, as is customary, that *e* is not L-false.[25] [Some customary methods require that *e* be factual; thus they exclude not only L-false but also L-true sentences as evidence. We shall, however, not demand this condition.]

We assume that any method of confirmation which is to be included in our λ-system can be represented by a c-function which fulfils ten conditions Cl-C10. C1-C9 will be stated in this section, C10 in § 8.

C1. If *h* and h′ are L-equivalent, c(*h*,*e*) = c(*h′*,*e*).

C2. If e and e′ are L-equivalent, c(*h*,*e*) = c(*h*,*e′*).

C3. General multiplication principle: c(*h* . *h′*,*e*) = c(*h*,*e*) × c(*h′*,*e* . *h*).

C4. Special addition principle: If *h* and *h″* are L-exclusive on the basis of *e* (i.e., *e* . *h* . *h′* is L-false), then c(*h* ∨ *h′*,*e*)= c(*h*,*e*) + c(*h′*,*e*).

C5. 0 ≤ c(*h*,*e*) ≤ l.

These principles have generally been accepted since the times of the classical calculus of probability.[26] The first two are obvious, since L-equivalent sentences have the same factual content. C3 and C4 are essential. C5 is an inessential stipulation concerning the range of values of c.

We shall now introduce some notations for sentences of special forms, which will be used throughout the subsequent discussions. Let e_Q be a conjunction of *s* *Q*-sentences with *s* distinct individual constants, among them s_1 with 'Q_1', s_2 with 'Q_2', ... s_K with 'Q_K'; hence $s_1 + s_2 + \ldots + s_K = s$. Let h_1 be a full sentence of 'Q_1' with an individual constant not occurring in e_Q; and $h_2, \ldots, h_K$ full sentences of '$Q_2, \ldots, Q_K$', respectively, with the same individual constant. e_Q may be regarded as a description of an observed sample of size *s*, and any hypothesis of the form h_i ($i = 1, 2, \ldots, K$) as a prediction concerning an individual not yet observed. Let e_1 be formed from e_Q by replacing every *Q*-predicate different from 'Q_1' by '$\sim Q_1$' (e.g., '$Q_3 a_5$' is replaced by '$\sim Q_1 a_5$'). Let e_2, ..., e_K be formed analogously. Thus any sentence of the form e_i describes the same sample as e_Q but in a less specific way; it ascribes the property Q_i to the same s_i individuals as e_Q does; but of the other $s - s_i$ individuals it says merely that they are not Q_i, without specifying which of the other *Q*'s each of them has. Let *M* be a factual molecular property with the width *w* ($0 < w < K$); hence M is a disjunction of *w* *Q*'s, of which we say that they are "in *M*". Let e_M be formed from e_Q by replacing every *Q*-predicate which is in *M* by '*M*' and every other one by '~*M*'.

25. [I], §§ 52, 55A.

26. [I], § 53.

Then e_M is a conjunction of s_M full sentences of '*M*' and $s - s_M$ negations of full sentences of '*M*', and

(4-1) $s_M = \sum_{i \text{ in } M} s_i$, where the sum runs over the subscripts *i* of those *Q*'s which are in *M*.

Thus e_M describes the same sample as e_Q, but it says of each individual only whether it is *M* or not. (Any e_i is a special case of e_M with Q_i as *M*.) Let h_M be the full sentence of '*M*' with the same individual constant as in h_1. Hence h_M is L-equivalent to a disjunction of *w* hypotheses h_i with the *Q*'s in *M*. Since these hypotheses are L-exclusive of one another, we have, according to the special addition principle (C4)

(4-2) For any *e*, $c(h_M,e) = \sum_{i \text{ in } M} c(h_i,e)$.

For the sake of greater generality, we admit also the case of an empty sample, that is, $s = 0$; in this case e_Q (and likewise any e_i and e_M) gives no factual information; it is an L-true sentence, e.g., the tautology '*t*'.

The inductive inference from an evidence describing a sample to a hypothesis concerning an individual not belonging to the sample, in other words, the determination of the value of $c(h_1,e_1)$, $c(h_M,e_M)$, and the like, is known as the *singular predictive inference*.[27] It is of basic importance for inductive logic. As we shall see later (§ 5), all other cases are reducible to this one in the sense that, if the values of a c-function for cases of the kind $c(h_i,e_i)$ are given, all other values can be derived.

We shall now lay down some further assumptions concerning the c-functions of the λ-system. In accordance with all historically given methods, we may assume that it has no influence on the value of $c(h,e)$, where *h* and *e* contain no variables, whether in addition to the individuals mentioned in *h* and *e* there are still other individuals in the universe of discourse or not. We shall state in C6 this condition for the special case of $c(h_i,e_i)$; the general condition just given follows from C6.

C6. For given sentences h_i and e_i (and given π primitive predicates), the value of $c(h_i,e_i)$ is the same in all systems L_N^π in which the two sentences occur, independently of *N*.[28]

In agreement with all known methods, we may assume that c treats all individuals on a par so that only their numbers influence the values of c. This assumption is stated in C7 in technical terms, as a condition for the admission of c to our λ-system.

27. See [I], § 44B,

28. [I], § 57B.

C7. c is *symmetrical with respect to the individual constants*, in the following sense.[29] If h' and e' are formed from h and e, respectively, by exchanging two individual constants (i.e., by replacing each occurrence of the one constant in h and in e by the other constant and vice versa), then $c(h,e) = c(h',e')$.

In an analogous way, we may assume that c treats all Q's on a par. This condition is formulated in technical terms in C8, which is analogous to C7:

C8. c is symmetrical with respect to the Q-predicates, in the following sense. If h' and e' are formed from h and e, respectively, by exchanging two Q-predicates, then $c(h,e) = c(h',e')$.

C9. For any M, $c(h_M,e_M) = c(h_M,e_Q)$.

This condition seems to be accepted generally, in most cases tacitly. For example, the probability (in the logical sense) that the next throw of a given die will yield an odd number is generally regarded as depending merely on the number of odd results and the number of even results obtained so far with this die; it is assumed that it makes no difference whether a particular previous result is recorded merely as 'odd' or more specifically as '1' or '3' or '5', whether merely as 'even' or rather as '2' or '4' or '6'. (Note that the additional information in e_Q is regarded as irrelevant only for the particularly hypothesis h_M; it is, of course, quite relevant for certain other hypotheses.)

From C9 with Q_i for M:

(4-4) For any i, $c(h_i,e_i) = c(h_i,e_Q)$.

From C9 and (4-2)

(4-5) For any M, $c(h_M,e_M) = \sum_{i \text{in} M} c(h_i,e_Q) \sum_{i \text{in} M} c(h_i,e_i)$.

Suppose that a method of confirmation is given, applicable to sentences of the forms e_1 and h_1. It cannot be given by an enumeration of the values of c for a list of cases because the number of cases is infinite. Therefore, it must be given by general rules which constitute a definition of the function c. The value of c for any two sentences of the forms h_1 and e_1 in a system L^{π}_{N} must thus depend upon—in other words, be a function of—certain magnitudes determined by those two sentences and the system. Let us examine the magnitudes which may possibly be arguments of the function in question. First, there are the two numbers N and π which are

29. [I], §§ 90, 91.

characteristic of the language system and hence are given with it. However, N is irrelevant for $c(h_1,e_1)$ according to C6. The number κ of Q's in the system is uniquely determined by π ($\kappa = 2^\pi$) and vice versa. Hence we may take κ instead of π as an argument. κ may be regarded as a logical, i.e., nonempirical, magnitude. There are, furthermore, empirical magnitudes involved, viz., the numbers s and s_1, by which the observed sample is described in e_1. Since e_1 contains individual constants, it tells us not only that s individuals belong to the sample and s_1 of them are Q_1, but also *which* particular individuals belong to the sample and *which* of them are Q_1. But this information is irrelevant in virtue of C7. The one individual constant occurring in h_1 is distinct from those in e_1. However, according to C7, it does not matter otherwise *which* constant this is. Therefore, the consideration of h_1 does not introduce any additional magnitude. Thus we arrive at the result that the value of $c(h_1,e_1)$ is uniquely determined by κ, s, and s_1. An analogous result holds for any other Q_i instead of Q_1. Hence:

(**4-6**) For any function c in the λ-system there is a mathematical function G such that for any Q_i and any pair of sentences h_i,e_i in a system L^π_N with any values of N, κ, s, and s_i, $c(h_i,e_i) = G(\kappa,s,s_i)$.

We obtain from (4-6) with (4-5):

(**4-7**) For any M, $c(h_M,e_M) = \sum_{i \text{ in } M} G(\kappa,s,s_i)$,

and with (4-4):

(**4-8**) For any i, $c(h_i,e_Q) = G(\kappa,s,s_i)$.

We call *G the characteristic function* of the given method of confirmation or of its c-function. If various methods of confirmation are given, represented by several functions c, c′, etc., then we can determine their several characteristic functions G, G', etc., with the help of (4-6). We shall soon find that these G-functions deserve the attribute 'characteristic' inasmuch as each of them characterizes completely and uniquely: the corresponding c-function (§ 5). A G-function is a mathematical function from triples of integers to real numbers of the closed interval (0,1).

Let κ sentences h_i (i = 1 to κ) be given, with any individual constant. Let h be their disjunction. Since h is L-true, its c on any evidence is 1. The sentences h_i are mutually L-exclusive. Therefore, according to the special addition principle (C4):

$$\sum_{i=1}^{\kappa} c(h_i,e_Q) = c(h,e_Q) = 1.$$

Hence with 't' for e_Q ($s = s_i = 0$):

(4-9) $\sum c(h_i,t) = 1$

Now, according to C8:

(4-10) The values $c(h_i,t)$ ($i = 1$ to K) are equal.

Hence with (4-9)

(4-11) For any c in the λ-system, any Q_i and any h_i, $c(h_i,t) = 1/K$.

Hence with (4-5)

(4-12). For any c and any M with width w, $c(h_M,t) = w/K$.

From (4-11) and (4-6), for $s = s_i = 0$:

(4-13) For any characteristic function G in the λ -system, $G(K,0,0) = 1/K$.

If in our subsequent discussions a G-function is defined in such a way that the definition does not supply a value for $G(K,0,0)$, then it will be understood that this value is to be determined according to the following convention:

(4-14) *Limit convention for G-functions*. We take as value of $G(K,0,0)$ $\lim_{\delta \to 0} (K,\delta,\delta)$.

The value determined in this way must then, according to (4-13), be equal to $1/K$ in order to qualify the given G-function for the λ -system.

§ 5. A Characteristic Function Gives a Complete Characterization

We shall now show that a characteristic function $G(K,s,s_i)$ characterizes completely its method of confirmation. This means that, if a function G is given, we can determine the value of its c not only for the special sentences h_i and e_i (according to (4-6)) but for any pair of sentences h,e (e non-L-false) in any system L_n^π. The proof of this result is somewhat more technical than oar former discussions; a reader who is satisfied with the result itself may skip over this section.

Let $G(K,s,s_i)$ be a given characteristic function. We shall show how the values of a confirmation function c can be determined with the help of G.

The c-values for any sentences are reducible, as we shall see, to c-values with respect to the tautological evidence 't' ("probability a priori" in the classical terminology). For the latter values we introduce the notation 'm'

(5-1) $m(h) = c(h,t)$.

Thus, for every c-function, there is a corresponding m-function defined by (5-1). The m-value of any sentence is reducible to the m-values of state-descriptions, as we shall see. Therefore, we begin with a consideration of state-descriptions.

We consider a state-description in *Q*-form, i.e:, a conjunction *k* of *N* *Q*-sentences with the *N* individual constants of the system in question. We arrange the components of *k* in the following order: first, the full sentences of 'Q_1', if any, in an arbitrary order; let their number be N_1 ($0 \leq N_1 \leq N$); then the N_2 full sentences of 'Q_2', etc.; finally, the N_K full sentences of 'Q_K'. The numbers N_i (i = 1 to K) are called the *Q-numbers* in *k*; their sum is *N*. Let the *Q*-sentences in this order be $j_1, j_2, \ldots, j_N$; hence *k* is $j_1 \cdot j_2 \cdot \ldots \cdot j_N$. Let k_n (n = 1 to N) be the conjunction of the *n* first *j*-sentences: $j_1 \cdot \ldots \cdot j_N$. Then, for $n \geq 2$, k_n is $k_{n-1} \cdot j_n$. Therefore, according to the general multiplication principle (C3, § 4):

(5-2) $$c(k_n,t) = c(k_{n-1},t) \times c(j_n,k_{n-1}),$$

since $k_{n-1} \cdot t$ is L-equivalent to k_{n-1}. Applying this result successively to $n = N, N-1, \ldots, 2$, and noting that k_1 is j_1 and k_N is k, we obtain;

(5-3) $$m(k) = c(k,t) = c(j_1,t) \times c(j_2,k_1) \times c(j_3,k_2) \ldots \times c(j_N, k_{N-1}).$$

Now we consider a particular predicate 'Q_i' occurring in the conjunction *k*. *k* contains N_i sentences with this predicate. Let m_i be the number of those components in *k* which precede the first *Q*-sentence (hence, if $i = 1$, $m_i = 0$, and, for $i > 1$, $m_i = \sum_{p=1}^{i-1} N_p$). The N_i Q_i-sentences are $j_{m_i+1}, j_{m_i+2}, \ldots, j_{m_i+N_i}$. Consider the terms for these *j*-sentences in the product on the right-hand side of (5-3). Let us first assume that 'Q_i' is not the first *Q*-predicate: occurring in *k*; hence $m_i > 0$. Then the first of the terms mentioned is c(j_{m_i+1}, k_{m_i}). k_{m_i} is a conjunction of m_i *Q*-sentences, none of them with 'Q_i'. j_{m_i+1} is a full sentence of 'Q_i'with an individual constant not occurring in k_{m_i}. Thus the two arguments of c here have the forms of h_i and e_Q, with $s = m_i$ and $s_i = 0$. Hence the first term is, according to (4-8), $G(K, m_i,0)$. If, on the other hand, 'Q_i' is the first *Q*-predicate in *k*, and hence $m_i = 0$, then the, first term is c(j_1,t); this, by (4-11), is always $1/K$. Here, $s = 0$ and $s_i = 0$; thus the first term can be stated as $G(K,0.0)$, which agrees with the former result $G(K,m_i,0)$ for $m_i = 0$. The second term, no matter whether $m_i = 0$ or $m_i > 0$, is c(j_{m_i+2}, k_{m_i+1}). Here the situation is analogous but with $s = m_i + 1$ and $s_i = 1$. Hence the second term is $G(K,m_i + 1,1)$. In this way it goes on; at each step, both, s and s_i are increased by 1. For the last term with 'Q_i', $s = m_i + N_i - 1$, and $s_i = N_i - 1$; hence the term is $G(K,m_i + N_i - 1, N_i - 1)$. Thus the partial product in (5-3) containing the terms with 'Q_i' is

$\prod_{p=1}^{N_i} G(\kappa, m_i + p - 1, p - 1)$. The results for the partial products with other Q's are analogous. Hence from (5-3):

(5-4) For any state-description k with the Q-numbers N_i, $\mathrm{m}(k) = \prod_i \prod_{p=1}^{N_i} G(\kappa, m_i + p - 1, p - 1)$, where the first product runs through those values of i (from 1 to κ) for which $N_i > 0$. The first term in the total product is always $1/\kappa$.

Thus G determines the m-values for all state-descriptions. With their help the m-values of all other sentences can be determined, because the following can be proved with the special addition principle (C4, § 4):[30]

(5-5) Let j be any sentence in the system L_n^π. If j is L-false, $\mathrm{m}(j) = 0$. Otherwise, $\mathrm{m}(j)$ is the sum of the m-values for the state-descriptions in the range of j.

Further, all c-values are reducible to m-values as follows:[31]

(5-6) For any sentences h,e ($\mathrm{m}(e) \neq 0$), $\mathrm{c}(h,e) = \dfrac{\mathrm{m}(e \cdot h)}{\mathrm{m}(e)}$.

Thus a characteristic function G determines the value of its c for any pair of sentences h,e in any system L_n^π, provided that e is not L-false.

It can, moreover, be shown that a characteristic function G determines *uniquely* its function c in the following sense. Let c be any c-function which fulfils the conditions C1-C9 (§ 4) and which has values for all sentence pairs h,e in a system L_n^π (e not L-false). Let the corresponding function $G(\kappa,s,s_i)$ be determined according to (4-6). If we now apply the procedure described above in (5-4), (5-5), and (5-6) for any sentence pair h,e, then (5-6) yields just that value of $\mathrm{c}(h,e)$ which the given function c has for h and e. In particular, if we take a pair of sentences of the forms h_i and e_i, then (5-6) supplies just that value $\mathrm{c}(h_i,e_i)$ which was used, according to (4-6), for determining $G(\kappa,s,s_i)$. This can easily be shown with the help of the theory of c-functions developed in [I]. The result means, in effect, that any G-function determines at most one c-function fulfilling the conditions. (Obviously, not every arbitrarily chosen function $G(\kappa,s,s_i)$ determines a c-function fulfilling the conditions.)

§6. Methods of Estimation

We have assigned to each method of confirmation a characteristic function G. The analogous problem for the methods of estimation can now be solved easily because we can simply transfer the G-functions from the former to the latter methods. The possibility of this procedure is due to a

30. [I], § 54C. 31. [I], § 54B.

very close relationship between the two kinds of methods. A one-one correspondence can be established which correlates with each method of confirmation a general method of estimation. This fact is of fundamental im. portance for the general theory of estimation but has so far not been sufficiently utilized.

Let us briefly indicate the correspondence in its general form before we consider the special form which it obtains in application to the systems L. Suppose that a scientist *X* has chosen a method of confirmation applicable to the language of science as a whole or in his field of work, a language also containing quantitative magnitudes of various kinds. Let this method be represented by a particular function c applicable to the sentences of the language. Suppose further that *X* is now in search of a general method of estimation which could be applied to all quantitative magnitudes expressible in his language. We saw earlier that *X* has to choose from a great number, theoretically speaking, an infinite number, of mutually incompatible possible methods of estimation, many of which have actually been applied or discussed. There is, however, for any method of confirmation represented by a function c, one method of estimation which is based on c in the following way.[32] We wish to define a function e such that e(*f,u,e*) is the estimate of the unknown value of a function *f* for an argument *u* (e.g., the length of the rod *u*, or the number of persons who will use a certain bus line at the date *u*) with respect to given evidence *e*. Let $r_1, r_2, \dots, r_n$ be *n* numbers which include all values of *f*(*u*) which are possible on the basis of *e*. Let h'_l ($l = 1, 2, \dots, n$) be the hypothesis that $f(u) = r_1$. Then we take as the estimate of *f*(*u*) the weighted mean of the possible values r_1 with the c for these values as weights:

(6-1)
$$e(f,u,e) = \sum_{i=1}^{n} [r_1 \times c(h'_l,e)].$$

The function e defined in this way on c is called the c-*mean estimate-function*. It is a general estimate-function, i.e., applicable to any kind of magnitude *f*. [We have assumed here for the sake of simplicity, that the set of possible values is finite. If it is infinite, e.g., an interval of the continuous scale of a measurable magnitude, then a modified definition is to be taken involving an integral instead of the sum. The basic idea, however, remains essentially the same.[33]] Since *X* has chosen the function c as his method of confirmation, it seems natural for him to choose the function e based on c as his method of estimation. He is, however, in no way compelled to make this particular choice. There is no incompatibility between any c-function and any e-function. Therefore, *X* may choose an estimate

32. [I], §§ 41D, 98-100. 33. [I], § 100A.

function e which is not defined on the basis of his function c or which is defined on c in a manner different from (6-1).

The general procedure just outlined will now be applied to our language systems L. As mentioned earlier, we shall here apply estimation only to the relative frequency (rf) of a property *M* in a class *K*. *K* is here supposed to be defined by the enumeration of its elements;[34] hence the cardinal number of *K*, say *m*, is given by the definition of *K*. There are $m + 1$ possible values for the absolute frequency of *M* in *K* (viz., 0, 1, 2, . . . , *m*) and likewise $m + 1$ possible values for the rf of *M* in *K* (viz., 0, $1/m$, $2/m$, … , 1). Therefore, we can lay down here a definition of the form (6-1):[35]

(6-2) $$\mathrm{e}(\mathrm{rf},M,K,e) = \sum_{l=1}^{m} [(l/m) \times \mathrm{c}(h'_l,e)].$$

Here, h'_l is the hypothesis that exactly *l* elements of *K* have the property *M*. Numerical hypotheses of this form can be formulated in the systems L. If two functions e and c fulfil (6-2), we say that e is *based upon* c.

We discussed earlier (§ 4) cases of the singular predictive inference, i.e., the application of any function c to sentences h_M,e_M, of a specified form concerning a property *M*. Now for cases of this kind there is an interesting connection between confirmation and (predictive) estimation of rf: the estimate of the rf of *M*, expressed by the function e based upon c, is equal to the value of c for h_M,e_M. The following theorem states this connection in a more general way.[36]

(6-3) Let c be a function fulfilling the conditions C1-C9 (§ 4). Let e be the estimate-function for rf based on c. Let *e* be any non-L-false sentence. Let *M* be a molecular property, and *h* a sentence ascribing *M* to an individual not mentioned in *e*. Let *K* be any finite non-null class of individuals not mentioned in *e*. Then

$$\mathrm{e}(\mathrm{rf},M,K,e) = \mathrm{c}(h,e)\ .$$

This general theorem leads to the following results, if e is hased on c:

(6-4) $\mathrm{e}(\mathrm{rf},M,K,e_M) = \mathrm{c}(h_M,e_M)$,

(6-5) $= \mathrm{c}(h_M,e_Q)$ (C9),

(6-6) $= \sum_{i \text{ in} M} G(K,s,s_i)$ (4-7),

$= \sum_{i \text{ in} M} \mathrm{c}(h_i,e_Q) = \sum_{i \text{ in} M} \mathrm{c}(h_i,e_i)$ (4-5),

(6-7) $\mathrm{e}(\mathrm{rf},M,K,e_Q) = \mathrm{c}(h_M,e_Q)$ (6-3),

(6-8) $= \mathrm{e}(\mathrm{rf},M,K,e_M)$ (6-5).

34. [I], § 15A. 35. [I] § 104. 36. [I], § 106, T106-1c.

From (6-4) with Q_i as M:

(6-9)	e(rf,Q_i,K,e_i)	= c(h_i,e_i),	
(6˜10)		= c(h_i,e_Q)	(4-4),
(6-11)	e(rf,Q_i,K,e_Q)	= c(h_i,e_Q)	(6-3),
(6-12)		= e(rf,Q_i,K,e_i)	(6-10),

From (6-8), (6-6), and (6-11)

(6-13) $$\mathrm{e}(\mathrm{rf},M,K,e_Q) = \sum_{i \,\mathrm{in}\, M} \mathrm{e}(\mathrm{rf},Q_i,K,e_Q).$$

The result (6-3) suggests a simple procedure for transferring our characteristic functions from methods of confirmation to methods of estimation for rf. Suppose that G is the characteristic function for a given c. This means, according to (4-6), that:

(6-15) $$\mathrm{c}(h_i,e_i)= G(K,s,s_i).$$

Hence we obtain with the help of (6-9):

(6-16) If G is the characteristic function of c, and e is based upon c, then

$$\mathrm{e}(\mathrm{rf},Q_i,K,e_i) = G(K,s,s_i).$$

Therefore, we may regard the same function G also as the characteristic function for that estimate-function e which is based upon c. It determines the value of e for the predictive estimation of the rf of a Q.

The result (6-16) holds only for a function e based on a given function c. It is not applicable to an estimate-function e for rf which is defined without reference to any c-function. But this result makes it appear natural to define in any case the characteristic function G of e by the same equation as in (6-16):

(6-17) For any estimate-function e for rf, its characteristic function G is determined as follows:

$$G(K,s,s_i) = \mathrm{e}(\mathrm{rf},Q_i,K,e_i)$$

That each of the magnitudes K, s, s_i might conceivably influence the value of e(rf,Q_i,K,e_i)is easy to see in analogy to the previous consideration concerning c (§ 4). The irrelevance of N is likewise easily seen. One might perhaps think that here the cardinal number m of K could be of influence. However, it seems that all historically know, estimate-functions for rf are independent of m. And certainly any e-function based upon any c-function which fulfils the conditions C1-C9 is independent of m.[37]

37. [I], § 106, T106-1d.

Therefore, we require the same of all e-functions to be admitted to the λ-system.

If any function e is given, no matter whether its definition does or does not involve a c-function, we can determine its characteristic function *G* by (6-17). Now this function *G* determines uniquely and completely a function c, as we have seen (§ 5). Hence for any function e, there is a unique function c such that both have the same characteristic function and therefore e is based upon c.

Thus the c-functions and the e-functions within our system are matched in pairs. A combined inductive method is represented by a function c and a function e based upon c and is characterized by a function *G* common to c and e.

It can be shown that, for any estimate-function based upon a c-function, the estimate for the sum of two magnitudes is equal to the sum of the estimates of the two magnitudes ([I], T100-2). Since the rf of the disjunction of two L-exclusive properties is the sum of their rf's, the following holds:

(**6-18**) Let e be any estimate-function of our system. Let M_1 and M_2 be L-exclusive molecular properties and $M_{1,2}$ their disjunction. Then $\text{e}(\text{rf},M_{1,2},K,e) = \text{e}(\text{rf},M_1,K,e) + \text{e}(\text{rf},M_2,K,e)$.

This follows also from (6-3) and the special addition principle for c (C4). (6-13) is a special case of (6-18).

Since we apply the c-functions of the λ-system also in cases where the evidence is tautological, in other words, the sample is empty, we shall do the same with our estimate-functions e, although this is not customary in present-day mathematical statistics. In particular, we obtain from (4-11) and (6-9):

(**6-19**) $$\text{e}(\text{rf},Q_i,K,t) = 1/\kappa,$$

and from (4-12) and (6-7):

(**6-20**) For any *M* with width *w*, $\text{e}(\text{rf},M,K,t) = w/\kappa$.

The result (6-19) seems quite plausible in view of the fact that the sum of the rf's of the κ *Q*'s in *K* is obviously 1, and hence the mean rf is $1/\kappa$ (compare [I], § 107).

§7. The Empirical Factor and the Logical Factor

We have solved the first part of our problem: for each method of confirmation or of estimation of rf, there is a mathematical function *G* which characterizes that method completely. Now we approach the second, more difficult part of the problem: we wish to find a set of parameters for the complete characterization of any function G and thereby of the corresponding inductive method. The number of parameters involved must be finite and even small, because otherwise the procedure would not serve the intended purpose: the possibility of investigating old and new inductive methods by studying their characters determined completely by their parameters.

Let us briefly look again at the ranges for the arguments and values of the functions G. π is the number of primitive predicates in a given system; it is a positive integer. Hence $\kappa = 2^\pi$, the number of Q's, is a positive power of 2. s is the number of individuals in the sample described by e_i, hence a nonnegative integer $< N$ (usually positive, but we admit the case $s = 0$). s_i is a nonnegative integer $\leq s$. The values of G are values of c or of an estimate of rf and hence are real numbers of the closed interval (0,1). (The range of values of a particular function G need not, of course, cover the whole interval; it may, e.g., be restricted to the rational numbers of the interval or to a part of them; but every real number of the interval may occur as value of some function G.)

Now, if all possible mathematical functions with the ranges of arguments and values just specified could serve as characteristic functions of inductive methods, then our problem could not be solved. It is not even possible to characterize by a finite set of parameters all rational integral functions of the specified kind, since, e.g., the number of coefficients in a polynomial of one argument ($a_0 + a_{1u} + a_2u^2 + a_3u^3 + \ldots$) may go beyond any bound. Thus, although the functions G have as arguments only three integers, it is not immediately apparent how it should be possible to find a finite and even small number of parameters for their characterization.

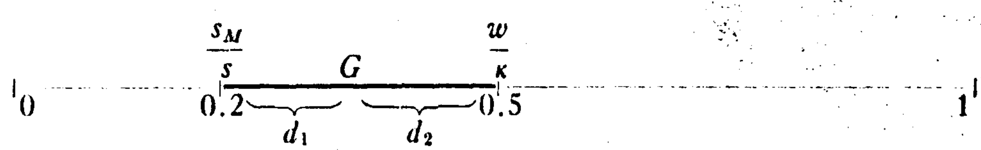
A close examination of the situation shows, however, that the task is not so hopeless as it might appear at first glance. The decisive point is that the functions G are not any arbitrary mathematical functions (for the specified ranges of arguments and values) but characteristic functions for functions c and e. $G(K,s,s_i)$ is intended to give the values of c(h_i,e_i) and e(rf,Q_i, κ,e_i). Now the functions c and e characterized by G must somehow be in accord with sound ways of inductive reasoning. We have seen earlier that there is a great variety of inductive methods acknowledged and used by careful thinkers. Nevertheless, these methods have certain general characteristics in common. And there are, on the other hand, traits of such a kind that any function c or e showing one of them would be regarded as unacceptable by practically everybody. (For example, consider a function G such that, for fixed K and s, $G(K,s,s_i)$ decreases with increasing s_i. Then the functions c and e characterized by this function G would be such that (1) the degree of confirmation for the prediction that the next thing is Q_1 and (2) the estimate of the rf of Q_1 among future observations would be lower, the more frequently Q_1 occurs in the observed sample. Practically everybody would regard these results as absurd and hence reject these functions c and e.) Now we intend to be rather liberal in the admission of inductive methods to the projected λ-system. We shall not restrict it to those methods which appear plausible to a majority of thinkers but shall also admit methods which I and presumably many others would regard as unsatisfactory. On the other hand, we intend to exclude those methods which practically everybody would reject. Our task is to see whether we can find a convenient parameter system for a class of G-functions restricted in a plausible way in view of their meaning as characterizations of inductive methods.

Suppose that a scientist X has made certain observations concerning the individuals of a given sample whirls he describes in e_M and that he wants to find the value of c for a singular prediction h_M concerning the property M and an estimate for the rf of M in an unobserved class K. Let us look at two particular factors in the situation which may possibly (not necessarily, and certainly not exclusively) influence the values of c and e. We take as the first the rf of M in the observed sample described in e_M, that is, s_M /s. This is an *empirical factor*. X cannot know this magnitude a priori, he learns about it from the factual circumstances concerning the sample, either by his own observations or by a report about the observations of others. It is clear in which direction this factor influences the values of c and e. Other things (including s) being equal, those values are higher, the greater s_M /s. This has often been stated explicitly and may be regarded as one of the fundamental and generally accepted characteristics of inductive reasoning. Moreover, all known methods of confirmation or estimation for rf agree that in any case of a sufficiently large sample the value of c or e, respectively, is either equal to or close to s_M /s.

The second factor to be considered is the relative width of M, that is, w/K. This is a logical factor. X knows it, independently of any observation of facts, from a logical analysis of the given language system L_n^π, which determines $K = 2^\pi$, and the given definition of 'M', which enables him to transform 'M' into a disjunction of certain Q's, whose number is w. As to the direction of the influence of this factor, a result analogous to the former one holds here. Other things (including K) being equal, if w/K is greater in one case than in another, the value of c is higher or equal, and likewise that of e. I believe that the stronger statement with "higher" instead of "higher or equal" also holds. [For example, let M be Q_1, hence $w = 1$, let h_M be 'Mc', i.e., 'Q_1c'. Let M' be $Q_1 \vee Q_2$, hence its width is $w' = 2$; let h'_M be '$M'c$', i.e., '$Q_1c \vee Q_2c$'. Let $s_2 = 0$, i.e., 'Q_2' does not occur in e_Q. Then $s_{M'} = s_1 + s_2 = s_1 = s_M$. Thus in the two cases with M and M' the values of K, s, and s_M are the same, but w' (= 2) > w (= 1): According to the special addition principle (C4, § 4), $c(h'_M,e_Q) = c(Q_1c,e_Q) + c(Q_2c,e_Q) = c(h_M,e_Q) + c(Q_2c,e_Q) \geq c(h_M,e_Q)$. I think, moreover, that, for any adequate c, $c(Q_2c,e_Q) > 0$, because the evidence e_Q does not exclude the possibility that c is Q_2; therefore, $c(h'_M,e_Q) > c(h_M,e_Q)$.]

§ 8. An Interval for Values of *G*

The two factors s_M/s and w/K are of particular importance for our problem because the value of $c(h_M,e_M)$, and hence likewise that of $e(rf,M,K,e_M)$, must always lie between those two values or be equal to one of them. It seems that this condition is fulfilled by practically all historically given methods (for an exception see the Appendix). The condition also seems plausible as a necessary condition of adequacy.[38] For example, let M be a primitive predicate. Then, no matter what is the value of K, $w = K/2$ and hence $w/K = ½$.[39] Let e_M say that, in a sample of 10 individuals, 2 have been found to be M; hence $s_M/s = 0.2$. The above condition says that in this case the value of c and e lies in the closed interval (0.2, 0.5), which is marked in the accompanying diagram by the heavy segment of the whole line representing the interval (0,1). Many people (namely, those



who accept the straight rule to be discussed later) would take in this case 0.2 as the value for c or e; in particular, the majority of contemporary statisticians would take 0.2 as the estimate of rf. Others would take other values. The decisive point is that, if someone takes in this case a value different from 0.2, he will choose a higher, not a lower, value, and, further, he will not go above 0.5. To choose any value below 0.2 would appear as

38. The kinds of c-functions studied in [I] (regular c-functions and symmetrical c-functions) are very comprehensive and also contain functions which do not fulfil the condition mentioned above. Therefore, this condition, in contradistinction to C1-C7, is not stated as a theorem in [I]; the same holds for C8 and C9 (14).

39. [I], § 32, T32-4c.

entirely arbitrary; if anybody deviates from 0.2, he does so because he thinks (or feels instinctively) that the value of c or e should somehow be influenced by the value ½, that it should somehow tend in the direction toward ½. On the other hand, to take a value > ½ when the observed rf is < ½ would appear as entirely unacceptable. For those who wish to apply their function c (or e) also to the tautology 't' as evidence, a plausible value for c(h_M,t) (or for e(rf,M,K,t))—and, indeed, the only acceptable value—would appear to be $1/K$ In the case of a Q (see (4-11) and (6-19)) and hence w/K in the case of any M with width w (see (4-12) and (6-20)); thus, in particular, in the case of a primitive predicate as in the above example. For a factual evidence e_M, the value of c or e would then move with increasing sample size s (if s_M/s remains the same) more and more away from w/K in the direction toward s_M/s, asymptotically approaching the latter value.

We shall now state as C10 a special case of the general condition just discussed, with Q_i as M and hence $w = 1$. We shall see that the general condition can then be derived. C10 is the last of our conditions of admission to the λ-system. Thus this system includes all those c-functions which fulfil the conditions C1-C10.

C10. c is such that for any sentences h_i and e_i with any values of K, s, and s_i, either $s_i/s \leq$ c(h_i,e_i) $\leq 1/K$ or $1/K \leq$ c(h_i,e_i) $\leq s_i/s$.

The general condition for M discussed above follows from C10 with (4-1) and (4-5):

(**8-1**) For any M, any sentences h_M and e_M with any values of K, w, s, and s_M, either $s_M/s \leq$ c(h_M,e_M) $\leq w/K$ or $w/K \leq$ c(h_M,e_M) $\leq s_M/s$.

In the special case in which s_M/s happens to be $= w/K$ (e.g., if in the earlier example $s_M = 5$), there would presumably be general agreement that this common value is the only acceptable value for c or e.

The preceding analysis was made for the purpose of finding suitable restricting conditions for our characteristic functions G. What effect in this respect does the acceptance of C10 have? Since c(h_i,e_i) gives the value of $G(K,s,s_i)$, C10 leads to the following result:

(**8-2**) For any characteristic function G of the λ-system and any values of K, s, and s_i, either $s_i/s \leq G(K,s,s_i) \leq 1/K$ or $1/K \leq G(K,s,s_i) \leq s_i/s$.

This result in itself may appear as rather insignificant, since it does no more than restrict to a still narrower interval the possible values of G, which were confined anyway to the interval (0,1). It does not say anything about the *form* of the functions G, which, of course, involves the more important problem. We shall, however, see later that the step here made opens the way for other, more important steps toward our goal, the parameter system.

§ 9. The λ-Functions

We found that the values of any G-function for cases of a certain kind must lie in a certain interval. In situations of this kind, where a value u must lie in an interval determined by the boundaries u_1 and u_2 (both included), we can characterize the location of u in relation to u_1 and u_2 in various ways. We might do it, e.g., by specifying the distances $d_1 = |u - u_1|$ and $d_2 = |u - u_2|$. Another way, often used and in the present case more convenient, consists in representing u as weighted mean of u_1 and u_2. That u is the *weighted mean* of u_1 and u_2 with the weights W_1 and W_2, respectively, means that

(9-1)
$$u = \frac{W_1 u_1 + W_2 u_2}{W_1 + W_2}.$$

If $W_1 = W_2$, u lies in the middle between u_1 and u_2. If $W_1 > W_2$, u lies nearer to u_1 than to u_2, i.e., $d_1 < d_2$. More specifically, the ratio of the distances is always equal to the inverse ratios of the weights:

(9-2)
$$\frac{d_1}{d_2} = \frac{W_2}{W_1}.$$

We admit as possible weights not only finite positive values but also 0 and ∞, to be used in the cases that u coincides with one. of the boundary values. If $W_1 = \infty$ and W_2 is finite (or W_1 is finite and $W_2 = 0$), then $d_1 = 0$, and hence u coincides with u_1. If $W_1 = 0$ and W_2 is finite (or W_1 is finite and $W_2 = \infty$), then $d_2 = 0$, and hence u coincides with u_2.

If the two boundaries u_1 and u_2 are given, then it is sufficient for the location of u to specify the ratio of the weights; their actual values are irrelevant. Therefore, we may, if we wish, standardize the value of one of the two weights, choosing, say, 1 or any other finite positive value that might seem convenient; then the value of the other weight (which runs now from 0 to ∞) determines the value of u.

We shall now apply this procedure to our situation. We are going to represent the value of $G(K,s,s_i)$ as a weighted mean of the values s_i/s and $1/K$. We decide to standardize by convention the weight of the empirical factor s_i/s; then the weight of the logical factor $1/K$ determines the value of $G(K,s,s_i)$. We choose as standardized weight of s_i/s the number s of individuals in the observed sample. This choice is not in need of a *theo-*

retical justification, since it does not involve any assertion. It is merely a convention, which in no way restricts the possible values of G. The choice is suggested (merely suggested, not imposed upon us) by the reflection that it might seem natural to give to the empirical factor determined by e_i more influence, the larger the sample, in other words, the more factual information is conveyed by e_i. The *practical* justification of the choice lies in the fact that it leads to an especially simple form of the parameter system, as we shall see.

After the weight of the empirical factor has been standardized, the weight to be attributed to the logical factor $1/\kappa$ is characteristic of the function G. For a given function G, this weight need not necessarily be always the same but may change from case to case, depending possibly upon the values κ, s, and s_i. Therefore, we represent it as a function of these three arguments: $\lambda(\kappa,s,s_i)$. The letter 'λ' is meant to suggest 'logical', since the value of the λ-function indicates the relative weight given to the logical factor. For example, if we find that for a certain inductive method in a given case $\lambda = 3$, this shows that the weight given to the logical factor in this case is equivalent to the weight given to the observation of three individuals. Thus the value of G is a weighted mean of s_i/s and $1/\kappa$, the weight of the first being s, and that of the second being $\lambda(\kappa,s,s_i)$. Hence we have, according to (9-1):

(9-3) $$G(\kappa,s,s_i) = \frac{s_i + \lambda(\kappa,s,s_i)/\kappa}{s + \lambda(\kappa,s,s_i)}.$$

This yields:

(9-4) $$\lambda(\kappa,s,s_i) = \frac{s \times G(\kappa,s,s_i) - s_i}{1/\kappa - G(\kappa,s,s_i)}.$$

Thus, if any inductive method is given in the form of a function c or e, then we can first determine its characteristic function G by the procedure explained earlier and then its function λ with the help of (9-4). On the other hand, if a function λ is given, then we can determine the corresponding function G by (9-3) and then the functions c and e by the earlier procedure. Thus a G-function and the corresponding λ-function determine each other, and hence the latter may be regarded as a characteristic function of the corresponding inductive method just as well as the former. What is gained by the construction of this second characteristic function? Both are mathematical functions of the same three arguments. But there is a decisive difference. It is not possible to simplify the general form of the functions G by dropping any one of the three arguments, because the value of $c(h_i,e_i)$, which determines the value of $G(\kappa,s,s_i)$, in general depends upon all three arguments. This holds, indeed, for almost all inductive methods, with the exception of only two which. occupy, so to speak, two opposite extreme positions (the straight rule and the function ct to be explained later). On the other hand, the use of the functions λ leads to a great simplification, because here we may drop several of the three arguments.

If, for any one of the customary methods, both of confirmation and of estimation of rf, we construct its function G according to the earlier procedure and then its function λ by (9-4), we find that the latter never contains the arguments s and s_i at all. This will soon be shown for several examples. The same holds likewise for those methods, infinite in number, which have the same fundamental character as the customary methods and can be constructed from them by a variation of some of their features. The result may at first appear surprising and perhaps even strange and implausible; but this impression will change after a little reflection. The fact that a given function λ is not dependent upon s and s_i must not be interpreted as meaning that the inductive method characterized by this λ-function fails to take into consideration the values s and s_i for the determination of $c(h_i,e_i)$. As (9-3) shows, s and s_i appear anyway in the function G, which determines $c(h_i,e_i)$, even if they do not appear in the λ-function. In particular, the nonappearance of s in the λ-function is due to our choice of s as the standardized weight of the empirical factor. (This is easily shown by the reappearance of s in the λ-function if any choice of a weight not involving s is made.) Thus the nonappearance of s shows simply that our choice of the weight fits well a common feature of all customary methods which give any influence at all to both the empirical and the logical factors.

While we admit to the λ-system all those inductive methods which fulfil the conditions C1-C10 (§§ 4 and 8), we shall restrict our further investigations for the sake of simplicity to those methods whose λ-functions are independent of s and s_i. (We shall, however, mention later a λ-function dependent upon s; see (20-10).) If λ is independent of s and s_i, then the same holds for the right-hand side of (9-4). Thus we obtain the following formulation concerning c, if we note that $G(\kappa,s,s_i)$ gives the value of $c(h_i,e_i)$.

C11. In the subsequent discussions it is presupposed that c is such that, if a system L_n^π is given and hence $\kappa = 2^\pi$ is fixed,

$$\frac{s \times c(h_i,e_i) - s_i}{1/\kappa - c(h_i,e_i)}$$

has always the same value, for any values of s and s_i and any sentences h_i and e_i.

That the conditions C1-C11 form a consistent system is easily seen from the fact that there are c-functions fulfilling all of them. This holds, e.g., for those examples of c-functions of the λ-system which we shall explain later (§§ 12-15).

Owing to C11, of the three arguments of the function λ, only κ remains. Therefore, we may now write simply '$\lambda(\kappa)$' instead of '$\lambda(\kappa,s,s_i)$'. Then (9-3), (6-15), and (6-16) yield:

(9-5) $G(\kappa,s,s_i)$

(9-6) $c(h_i,e_i)$

(9-7) $e(\text{rf},Q_i,K,e_i)$

$$\left.\begin{matrix} G(\kappa,s,s_i) \\ c(h_i,e_i) \\ e(\text{rf},Q_i,K,e_i) \end{matrix}\right\} = \frac{s_i+\lambda(\kappa)/\kappa}{s+\lambda(\kappa)}$$

Hence with (4-5), (4-1), and (6-4):

(9-8) $c(h_M,e_M)$

(9-9) $e(\text{rf},M,K,e_M)$

$$\left.\begin{matrix} c(h_M,e_M) \\ e(\text{rf},M,K,e_M) \end{matrix}\right\} = \frac{s_M+(w/\kappa)\lambda(\kappa)}{s+\lambda(\kappa)}$$

These two results show that likewise in this case, for a property M with the width w, the value of c for h_M and the estimate of the rf of M may be represented as a weighted mean of the empirical factor, which here is s_M/s, and the, logical factor, which here is w/κ, with the same weights s and $\lambda(\kappa)$ as before.

§10. A λ-Function Gives a Complete Characterization

The usefulness of the functions λ depends on whether such a function characterizes its inductive method completely. We shall now show that this is the case. This can easily be shown because we found earlier the analogous result with respect to the characteristic functions G. Suppose that a function $\lambda(\kappa)$ is given. Then we can express the G-function occurring in (5-4) in terms of $\lambda(\kappa)$ by (9-5). Thus we obtain the following (writing for simplicity 'λ' instead of '$\lambda(\kappa)$'):

(10-2) $$m(k) = \prod_i \prod_{p=1}^{N_i} \frac{p-1+\lambda/\kappa}{m_i+p-1+\lambda},$$

(where the quotient for $m_i = 0$ and $p = 1$ is always to be taken as $1/\kappa$).

Hence:

(10-3) $$m(k) = \frac{\prod_i A_i}{\prod_i B_i},$$

where A_i and B_i are as follows:

(10-4) $$A_i = \prod_{p=1}^{N_i} (\lambda/K + p - 1),$$

(10-5) $$B_i = \prod_{p=1}^{N_i} (\lambda + m_i + p - 1)$$

From (10-4):

(10-6) $$A_i = \lambda/K(\lambda/K + 1)(\lambda/K + 2) \ldots (\lambda/K + N_i - 1).$$

Let i_1 and i_2 be two values of i such that $i_1 < i_2$, $N_{i_1}, > 0$, $N_{i_2} > 0$, while for the intermediate values of i, if any, $N_i = 0$. For $i > 0$, $m_i = \prod_{p=1}^{i-1} N_p$. Therefore, $m_{i_2} = m_{i_1} + N_{i_1}$. According to (10-5), for any i, B_i is a product of N_i ascending factors, each greater by 1 than the preceding one. The greatest factor in B_{i_1} is λ + na$_i$, + $m_{i_1} - 1$. The smallest factor in B_{i_2} is $\lambda + m_{i_2} = \lambda + m_{i_1} + N_{i_1}$, hence greater by 1 than the greatest factor in B_{i_1}. Therefore, $\prod_i B_i$ is a product of $\sum N_i = N$ ascending factors from λ to $\lambda + N - 1$. Thus from (10-3) and (10-6):

(10-7) $$\mathrm{m}(k) = \frac{\prod_i [\lambda/K(\lambda/K + 1)(\lambda/K + 2) \ldots (\lambda/K + N_\iota - 1)]}{\lambda(\lambda + 1)(\lambda + 2) \ldots (\lambda + N - 1)}.$$

Here it is convenient to use the function $\left[{r \atop n}\right]$, where r is a real number and n a nonnegative integer, defined by the following recursive definition:[40]

(10-8) (*a*) $\left[{r \atop 0}\right] = 1.$

(*b*) $\left[{r \atop n+1}\right] = \left[{r \atop n}\right](r - n).$

This definition immediately yields the following theorems:

(10-9) (*a*) $\left[{r \atop 1}\right] = 1$

(*b*) For n > 0, $\left[{r \atop n}\right] = r(r - 1)(r - 2) \ldots (r - n + 1);$

hence $\left[{r \atop n}\right]$ is a product of n descending factors beginning with r.

40. In [I], D40-3, a function $\left[{m \atop n}\right]$ was defined, which is a specialization of the present one, with an integer m. For the present function, the notation '$r^{(n)}$' is sometimes used.

(In the cases of our applications, we have always $r \geq n$; hence every factor and the whole product is positive.)

(*c*) $\left[{n \atop n}\right] = n!$.

(*d*) $\left[{n+1 \atop n}\right] = (n+1)!$.

Using this notation, (10-7) becomes:

(10-10)
$$\mathrm{m}(k) = \frac{\prod_i \left[{\lambda/\kappa + N_i - 1 \atop N_i}\right]}{\left[{\lambda + N - 1 \atop N}\right]}.$$

The formula (10-7) (or (10-10)) determines the values of m for any state-descriptions in any given system L_n^π. According to the procedure outlined in §5, this determines the values of m for all sentences and the values of c for all pairs of sentences *h*,*e* (*e* not L-false). The function c, in turn, determines all values of the estimate-function e based upon it. Thus any given λ-function characterizes completely a combined inductive method consisting of a function c and a function e.

§ 11. Inductive Methods of the First Kind: λ Is Independent of *κ*

It seems that almost all historically given inductive methods are such that their λ-functions are not even dependent upon *κ* but have a constant numerical value. As far as I am aware at present, the only exception is the function c* which I proposed in earlier publications. We now divide the inductive methods here investigated into two kinds; the first kind comprises those whose λ does not depend upon *κ*, the second those whose λ is a function of *κ*. We shall discuss each of these two kinds in turn and examine examples of known methods for them.

Suppose that an inductive method is given either by a c-function or by an e-function. We determine first its *G*-function (§§ 4, 6) and then its λ-function by (9-4). If we find that the latter is independent, not only of *s* and s_i, but also of *κ*, the method belongs to the first kind. Thus here the λ-function degenerates into a constant, which we simply denote by λ. We regard λ as the *parameter* characterizing the method in question.

The relation between the parameter λ and the corresponding function *G* is expressed by the following formulas, which are merely simplifications of (9-3) and (9-4):

(11-1)
$$G(\kappa,s,s_i) = \frac{s_i + \lambda/\kappa}{s + \kappa},$$

(11-2)
$$\lambda = \frac{s \times G(\kappa,s,s_i) - s_i}{1/\kappa - G(\kappa,s,s_i)}.$$

We consider as possible values of λ not only all finite positive real numbers (corresponding to the interior points of the interval (w/K, s_M/s) discuss!d in § 8) but also 0 and ∞, corresponding to the end points of the interval. Negative values are excluded, because they would correspond to points outside the interval and hence violate the condition C10 (§ 8). In our further discussions, various functions of λ will occur. The following conventions will be applied to them.

(11-3) *Limit conventions* for any function $f(\lambda)$.

(a) If the definition of a function $f(\lambda)$ does not assign to it a value for $\lambda = 0$, we take as this value $\lim_{\lambda \to 0} f(\lambda)$.

(b) We take as value of $f(\lambda)$ for $\lambda = \infty$ $\lim_{\lambda \to \infty} f(\lambda)$

Let λ be a number of the specified range. We shall now introduce notations for the functions characterized by the given λ. G_λ is the G-function determined by λ according to (11-1); c_λ is the c-function characterized by G_λ; m_λ is the m-function corresponding to c_λ; e_λ is the e-function based upon c_λ.

The formulas (9-6) to (9-9) here take the following simpler forms:

(11-4) $c_\lambda(h_i,e_i)$
(11-5) $e_\lambda(\mathrm{rf},Q_i,K,e_i)$ $\Big\} = \dfrac{s_i + \lambda/K}{s+\lambda}$

(11-6) $c_\lambda(h_M,e_M)$
(11-7) $e_\lambda(\mathrm{rf},M,K,e_M)$ $\Big\} = \dfrac{s_M + (w/K)\lambda}{s+\lambda}$

In the case that the evidence is tautological and hence the sample is empty, we found previously the following results ((4-11), (6-19), (4-12), (6-20)):

(11-8) $m_\lambda(h_i) = c_\lambda(h_i,t)$
(11-9) $e_\lambda(\mathrm{rf},Q_i,K,t)$ $\Big\} = 1/K$
(11-10) $m_\lambda(h_M) = c_\lambda(h_M,t)$
(11-11) $e_\lambda(\mathrm{rf},M,K,t)$ $\Big\} = w/K$

We see now that these values are in accord with the above results (11-4) to (11-7). For every positive, finite λ they follow immediately from those results for $s = s_i = s_M = 0$. For $\lambda = 0$ and $\lambda = \infty$, our limit conventions must be applied. Obviously, both limits for 0 and for ∞ are in the case of (11-4) and (11-5) $1/K$ and in the case of (11-6) and (11-7) w/K.

§ 12. The Nonextreme Methods

The two inductive methods of the first kind characterized by the extreme values 0 and ∞ of λ may be called *the extreme methods*. We shall consider them later. Here we shall establish some common characters of the others, *the nonextreme methods*. Thus we suppose now that λ is a given positive and finite number. Our previous discussion (§ 10) shows that the number λ determines completely the functions m_λ, c_λ, and e_λ. According to (10-7) we have, for a state-description k with the Q-numbers N_i:

(12-1) $$m_\lambda(K) = \frac{\prod_i [\lambda/\kappa(\lambda/\kappa+1)...(\lambda/\kappa+N_i-1)]}{\lambda(\lambda+1)(\lambda+2)...(\lambda+N-1)},$$

(here 'λ' is not short for 'λ(K)' but denotes a number).

(The alternative form (10-10) could likewise be used here.) This determines, as explained previously (§ 5) the values of m_λ for all sentences, and the values of c_λ for all pairs of sentences h,e (e not L-false). Since λ is here positive and finite, the same holds for every factor and every product on the right-hand side of (12-1), hence for $m_\lambda(K)$ itself and therefore also for the value of m_λ for every non-L-false sentence. Let e be a non-L-false sentence and h a factual sentence logically independent of e, i.e., such that neither h nor its negation is L-implied by e; hence both e . ~h and e . h are non-L-false and have positive m_λ-values. Now $c_\lambda(h,e) = m_\lambda(e\ .\ h)/\ m_\lambda(e)$; and $m_\lambda(e) = m_\lambda(e\ .\ h) + m_\lambda(e\ .\ {\sim}h) > m_\lambda(e\ .\ h)$. Therefore:

(12-2) $$0 < c_\lambda(h,e) < 1\ .$$

Sentences of the forms h_i,e_i and h_M,e_M fulfil the conditions here stated for h,e. Therefore:

(12-3) $$0 < c_\lambda(h_i,e_i) < 1\ ,$$
(12-4) $$0 < c_\lambda(h_M,e_M) < 1\ .$$

Hence with (6-9) and (6-4):

(12-5) $$0 < e_\lambda(\mathrm{rf},Q_i,K,e_i) < 1$$
(12-6) $$0 < e_\lambda(\mathrm{rf},M,K,e_M) < 1\ .$$

On the basis of these results, the following can be shown (with the help of the theory of c-functions developed in [I] and the results mentioned at the end of § 5). If any positive, finite real number is chosen as λ, then this determines one and only one c-function c_λ, and this function c_λ fulfils the conditions C1-C11. On the basis of our later explanations in §§ 13 and 14, the same holds also for λ = 0 and λ = ∞. Hence we arrive easily at the following results. Let $\lambda(K)$ be an arbitrarily chosen function of K whose values for possible values of K (which are positive finite numbers) are nonnegative numbers (or possibly ∞). Then, for any system L_n^π, $\lambda(K)$ has a fixed nonnegative value (possibly ∞) and therefore determines for this system one and only one c-function, and this c-function fulfils the conditions C1-C11.

We shall now consider two examples of nonextreme methods.

1. *The modified Laplace method*. Laplace's famous *rule of succession* concerns the case of the singular predictive inference. Therefore, it applies to the sentences h_M, e_M. It says, expressed in our notation, that

(12-7) $$c(h_M, e_M) = \frac{s_M + 1}{s + 2}.$$

If this is applied to all factual properties M, it leads to contradictions. For example, in a system with $K = 4$, let e_Q be such that $s = 4$, $s_1 = s_2 = s_3 = s_4 = 1$ (i.e., the sample contains one individual for each of the four Q's). Then, according to Laplace's rule, for every i, $c(h_i, e_Q) = (1 + 1)/(4 + 2) = ⅓$. Hence we obtain, on the one hand, by using the special addition principle (C4, § 4), which is, of course, accepted by Laplace: $c(h_1 \vee h_2, e_Q) = ⅔$. On the other hand, applying Laplace's rule directly to $Q_1 \vee Q_2$ (which is P_1), we obtain $c(h_1 \vee h_2, e_Q) = (2 + 1)/(4 + 2) = ½$, which is incompatible with the first result. If the first procedure (with C4) is applied to the disjunction of the four hypotheses h_i, we should even obtain $c = 4/3 > 1$, which is, of course, impossible. However, Laplace's method becomes consistent if we restrict its application to a suitable kind of properties. We shall do this now in a special way. We apply Laplace's rule directly only to the primitive properties (of a given system); then we choose the values of $c(h_M, e_M)$ for other properties M in a certain way such that they fit well to the values supplied by the rule. We call the method constructed in this way the *modified Laplace method*. A primitive property has always the relative width $w/K = ½$ and the width $w = K/2$. We obtain from (11-6) for this value:

(12-8) $$c(h_M, e_M) = \frac{s_M + \lambda/2}{s + \lambda}$$

Comparing this with (12-7), we see that Laplace's rule, if applied to primitive properties or, more generally, to any M with $w = K/2$, is characterized by $\lambda = 2$. We decide now to apply in the case of all other factual properties the same value $\lambda = 2$. This assures that the values of $c(h_M, e_M)$ for any properties are in accord with the values determined by the original rule for the primitive properties. We obtain from (11-4) to (11-7) for λ = 2:

(12-9) $c_2(h_i,e_i)$
(12-10) $e_2(\text{rf},Q_i,K,e_i)$ $\Bigg\} = \dfrac{s_i + 2/\kappa}{s+2}$,

(12-11) $c_2(h_M,e_M)$
(12-12) $e_2(\text{rf},M,K,e_M)$ $\Bigg\} = \dfrac{s_M + 2w/\kappa}{s+2}$.

(12-11) is our modification of Laplace's rule (12-7); it coincides with the latter for $w = \kappa/2$. The subsequent table (12-19) gives, in the column λ = 2, numerical examples for (12-11) and (12-12). According to (12-1), the in for a state-description k is here:

(12-13) $$m_2(k) = \frac{\prod_i [2/\kappa(2/\kappa+1)...(2/\kappa+N_i-1)]}{(N+1)!},$$

2. *The method with* λ = 1. If λ is a positive integer, the form of (12-1) becomes simple, since the denominator is then (N + λ − 1)!/(λ − 1) !, The modified Laplace method is an example, with λ = 2. Another method of this kind, which deserves serious consideration, is that characterized by λ = 1. This method gives to the logical factor as much weight as to the observation of one individual. Here (12-1) becomes:

(12-14) $$m_1(k) = \frac{\prod_i [1/\kappa(1/\kappa+1)(1/\kappa+2)(1/\kappa+N_i-1)]}{N!}.$$

(11-4) to (11-7) we obtain:

(12-15) $c_1(h_i,e_i)$
(12-16) $e_1(\text{rf},Q_i,K,e_i)$ $\Bigg\} = \dfrac{s_i + 1/\kappa}{s+1}$,

(12-17) $c_1(h_M,e_M)$
(12-18) $e_1(\text{rf},M,K,e_M)$ $\Bigg\} = \dfrac{s_M + w/\kappa}{s+1}$.

For numerical examples see table (12-19), column λ = 1.

In order to illustrate some selected methods by numerical examples, the subsequent table gives the values of $c_\lambda(h_M,e_M)$, which are also the values of $e_\lambda(\text{rf},M,K,e_M)$ for a primitive property M (hence w/κ = ½) and a sample of size s = 10, according to (11-6) and (11-7). The values are given only for s_M = 0 to 5; note that the value for s_M is here, with w/κ = ½, the same as that for $10 - s_M$. The values are stated for those methods of the first kind which are characterized by λ = 0 (the straight rule, see, below, § 14), λ = 1 (just mentioned), λ = 2 (the modified Laplace method), λ = 4, 8, 16, 32, and λ = ∞ (the first extreme method, see the next section). [As we shall see later, § 15, the same values hold also for methods of the second kind in systems with a suitable K. In particular, the values stated here for λ = 2, 4, 8, 16, 32 hold for the method of c* and e*, characterized by $\lambda = K$, in a system with $K = 2, 4, 8, 16, 32$, respectively, hence with 1, 2, 3, 4, or 5 primitive predicates, respectively.]

(12-19) Numerical values of $c_\lambda(h_M,e_M)$and $e_\lambda(\text{rf},M,K,e_M)$ for $w/K = ½$ and $s = 10$, as functions of λ and s_M.

s_M	λ =							
	0	1	2	4	8	16	32	∞
0	0	0.0455	0.0833	0.1429	0.2222	0.3077	0.3810	0.5
1	0.1	0.1364	0.1667	0.2143	0.2778	0.3462	0.4047	0.5
2	0.2	0.2273	0.2500	0.2857	0.3333	0.3846	0.4286	0.5
3	0.3	0.3182	0.3333	0.3571	0.3889	0.4231	0.4524	0.5
4	0.4	0.4091	0.4167	0.4286	0.4444	0.4615	0.4762	0.5
5	0.5	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5

§ 13. The First Extreme Method: λ = ∞

We shall now investigate the two extreme methods among those of the first kind, i.e., those characterized by λ = 0 and λ = ∞. We begin with . the latter.

The functions characterized by λ = ∞ are G_∞, c_∞, m_∞, and e_∞. According to the limit convention (11-3b), we obtain from (11-1):

(13-1) $$(G_\infty\ K,s,s_i) = \lim_{\lambda\to\infty} \frac{s_i + \lambda / K}{s + \lambda} = 1/K.$$

Thus $G_\infty$is equal to the logical factor, independently of s and s_i. The empirical factor s_i/s is entirely disregarded. From (13-1) or directly from (11-4) to (11-7):

(13-2) $c_\infty(h_i,e_i)$
(13-3) $e_\infty(\text{rf},Q_i,K,e_i)$ } $= 1/K$,
(13-4) $c_\infty(h_M,e_M)$
(13-5) $e_\infty(\text{rf},M,K,e_M)$ } $= w/K$.

Thus $c_\infty$is here equal to the relative width of the property in question, irrespective of any information concerning an observed sample. The values stated in (13-2) to (13-5) are the same as those given by this and any other method on the. basis of the tautological evidence (see (11-8) to (11-11)). Thus this method clings always to that value of c or e which is determined before any factual information is available. This value is never changed, no matter what factual information is obtained (as long as it does not concern the individual mentioned in the hypothesis). Thus accepting this method means refusing to give any regard to experience, to the results of observations, in making expectations or estimations. This is in gross contrast to what is generally regarded as sound inductive reasoning. Therefore, the method characterized by **λ** = ∞ seems unacceptable as an inductive method in science.

From (13-2) with (4-4):

(13-6) $$c_\infty(h_i, e_Q) = 1/K.$$

Let k be any state-description with N individuals. Each of the N factors on the right-hand side of (5-3) has for c. the form as in (13-6) and hence is equal to 1/rc. Therefore:

(13-7) $$m_\infty(k) = 1/K^N.$$

Let m† be defined as that m-function which, in application to a given system L^π_n, has the same value for all state-descriptions of that system; let c† be based upon m†.[41] The number of state-descriptions is K^N.[42] The sum of their m-values must. be 1.[43] Therefore:

(13-8) $$m\dagger(k) = 1/K^N.$$

Hence the function m† is the same as m_∞; therefore, c† is the same as c_∞, a function which we found to be objectionable. The definition of c† does not immediately reveal the unacceptable consequences to which it leads. On the contrary, it may appear at first glance as quite plausible. This is presumably the reason why this function ct has actually been proposed by several authors, as we shall see.

By an *individual distribution*[44] (sometimes called a "constitution") with respect to given individuals and given properties, which form an exhaustive and nonoverlapping division, we mean a specific description assigning to each of the individuals one of the properties (e.g., e_Q and e_M). By a *statistical distribution*[44] we mean a weaker description which states for each of the properties merely how many individuals have it, but not which individuals. The question of equal a priori probabilities of distributions

41. [I], §110A.
42. [I], § 35, T35-1b.
43. [I], § 55, D55-1b.
44. [I], § 26B.

has frequently been discussed in connection with the classical or related conceptions of probability and the principle of indifference. Probabilities a priori are represented in our theory by the values of c with respect to the tautological evidence, hence by the values of m (§ 5). The controversy concerned the question as to which of the two following, rules should be accepted:

(A) individual distributions have equal m-values.
(B) Statistical distributions have equal m-values.

Now it can easily be shown that either rule leads to contradictions if taken in the given unrestricted form and hence applied to *all* divisions. [For example, in a system with five individuals and two primitive predicates, hence four *Q*'s, consider two divisions D_1 and D_2: D_1 consists of the four *Q*'s, and D_2 of Q_1-or-Q_2 (which is the first primitive property), Q_3, and Q_4. The number of individual distributions for D_1 is $I_1 = 1024$, for D_2, $I_2 = 243$ (see [I], T40-31a); for our discussion we need only the obvious result that $I_1 \neq I_2$. The number of statistical distributions for D_1 is $S_1 = 56$, for D_2, $S_2 = 21$ (see [I]; T40-33a); again, we need only the obvious result that $S_1 \neq S_2$. Let m_A satisfy rule (A) and m_B rule (B). Let *h* be the conjunction of the five full sentences of 'Q_4' with the five individual constants. Then *h* is an individual distribution for both D_1 and D_2. Therefore, rule (A) leads, if applied to D_1, to the result $m_A(h) = 1/I_1$ and, if applied to D_2, to $m_A(h) = 1/I_2$. These two results are incompatible. Let h′ be the statistical distribution corresponding to *h*, saying that the cardinal number of Q_4 is five. (h′ is L-equivalent to *h*.) Then h′ is a statistical distribution for both D_1 and D_2. Therefore, rule (B) leads to the two values $1/S_1$ and $1/S_2$ of $m_B(h')$, which are different and hence incompatible.]

However, each of the two rules becomes consistent if restricted to any one division. It seems natural to take here for each system L the strongest division possible in the system, which is that formed by the *Q*'s. The individual distributions for all the individuals of the system with respect to the *Q*'s are the state-descriptions. The corresponding statistical descriptions state merely the *Q*-numbers N_i; we call them *structure-descriptions*.[45] (In our systems L, a structure-description for given *Q*-numbers N_i may be formulated as a disjunction of those state-descriptions for which the given numbers N_i hold.) Thus the two modified rules are as follows:

(A′) State-descriptions have equal m-values,
(B′) Structure-descriptions have equal m-values.

45. [I], §§27, 34.

Keynes[46] discusses the whole problem and mentions that C. S. Peirce favored rule (A). Keynes himself does, too, but proposes a modification similar to (A'). Ludwig Wittgenstein[47] proposes a rule which is essentially the same as (A'). Since rule (A') defines m t, the authors mentioned accept the function ct, which is the same as e_m. We have seen that this method leads to unacceptable results. «'e shall see later that rule (B') leads to our function c* (§ 15).

§ 14. The Second Extreme Method: X = 0; the Straight Rule

The functions characterized by X = 0 are G_o, c_o, mo, and t_o. According to (11-1) and (11-4) to (11-7), we have for s = 0:

(14-1) $$G_0(K,s,s_i),$$

(14-2) $c_0(h_i,e_i)$
(14-3) $e_0(\mathrm{rf},Q_i,K,e_i)$ } $= s_i/s,$

(14-4) $c_0(h_M,e_M)$
(14-5) $e_0(\mathrm{rf},M,K,e_M)$ } $= s_M/s.$

Numerical examples for (14-4) and (14-5) are given in table (12-19), column λ = 0.

The values stated here hold only for $s > 0$. In the case of $s = 0$ the functions yield 0/0 and hence are inapplicable; this case will be discussed later.

Theorem (14-4) says that in the case of the singular predictive inference c_0 is equal to the observed rf. We call this the *straight rule of confirmation*. (14-5) says that the estimate e_0 of rf in an unobserved class is equal to the observed rf. We call this the *straight rule of estimation*. The method of the straight rule assigns the weight λ = 0 to the logical factor w/K; in other words, it disregards this factor; it takes as c_0 or e_0 in the above cases simply the value of the empirical factor s_M/s.

In the case of the tautological evidence we have $s = s_M = 0$. In this case the results for c_0 and e_0 stated above are inapplicable. Those authors who accept the straight rule in either form do indeed usually reject tautological evidence. However, our limit convention (11-3a) enables us to apply the method with λ = 0 also to tautological evidence, in other words,

46. *Op. cit*., pp. 49 ff., especially p. 56 at bottom.
47. *Tractatus logico-philosophicus* (London, 1922), *5. 15.

to establish a function m_0. We found earlier that the results (11-8) to (11-11) hold also for $\lambda = 0$. Hence:

(14-6) $m_0(h_i) = c_0(h_i,t)$ } $= 1/\kappa$,
(14-7) $e_0(\text{rf},Q_i,K,t)$
(14-8) $m_0(h_M) = c_0(h_M,t)$ } $= w/\kappa$.
(14-9) $e_0(\text{rf},M,K,t)$

Let k be a state-description with the Q-numbers N; (i = 1 to K). Application of the limit convention to (12-1) yields:

(14-10) $$m_0(k) = \lim_{\lambda\to 0} \frac{\prod_i [\lambda/\kappa(\lambda/\kappa+1)...(\lambda/\kappa+N_i-1)]}{\lambda(\lambda+1)(\lambda+2)...(\lambda+N-1)},$$

where the product runs through those; i for which $N_i > 0$, in other words, for which Q_i is nonempty in k. Let p be the number of the nonempty Q's in k ($1 \le p \le \kappa$); thus the above product contains p factors. Hence:

(14-11) $$m_0(k) = \lim_{\lambda\to 0}\left[\frac{(\lambda/\kappa)^p}{\lambda}\right]\cdot\frac{\prod_i (N_i-1)}{(N-1)!}.$$

We distinguish two cases. *Case 1*: Let $p = 1$. In this case all individuals have the same Q and hence are completely alike (with respect to the properties expressible in the given system). We say in this case that k is a *homogeneous stale-description*. There is only one positive N_i, and this is $= N$. Hence the last quotient in (14-11) becomes $(N-1)!/(N-1)!$ and drops out. The quotient under the limit becomes $1/\kappa$. Hence:

(**14-12**) If k is a homogeneous state-description, $m_0(k) = 1/\kappa$.

Case 2: $p > 1$. The quotient under the limit in (14-11) becomes λ^{p-1}/κ^p hence the limit is 0. Therefore, since the last quotient in (14-11) is always finite:

(**14-13**) If k is a nonhomogeneous state-description, $m_0(k) = 0$.

This result leads (with (5-5)) to the consequence that for many factual sentences $m_0 = 0$. Therefore, we can, in general, not determine $c_0(h,e)$, by $m_0(e\ .\ h)/\ m_0(e)$ (as in (5-6)). We have instead to apply the limit convention directly to c_0:

(14-14) $$c_0(h,e) = \lim_{\lambda\to 0} c_\lambda(h,e) = \lim_{\lambda\to 0}\frac{m_\lambda(e.h)}{m_\lambda(e)}$$

For every non-L-false e and every $\lambda > 0$, $m_\lambda(e) > 0$ (§ 12); hence the quotient has a definite value. It can be shown that these values converge toward a limit for $\lambda \rightarrow 0$. Therefore, for every pair of sentences h,e (e not L-false) in any system L^{π}_{n}, $c_0(h,e)$ has a definite value, determined by (14-14). The application of this general procedure to the special case of h_M and e_M leads to the values stated in (14-4) for $s > 0$ and in (14-5) for $s = 0$.

Those m-functions which fulfil the conditions C1-C5 (§ 4) and have a positive value for every state-description are called *regular*;[48] likewise, the c-functions based upon regular m-functions. I believe that only regular c-functions represent adequate inductive methods; therefore, only these functions are dealt within [I]. It can be shown on the basis of (12-1) that, for any finite positive λ, m_λ is regular, and hence also c_λ. The same holds for m_∞, according to (13-7), and hence for c_∞. On the other hand, we see from (14-13) that m_0 is not regular: We call m_0 and c_0 *quasi-regular* functions, which means that, although they are not regular, they can be represented as limits of regular functions. c_0 can be represented approximately by a suitable regular function c_λ with $\lambda > 0$; by taking a sufficiently small, but positive, value of λ, any desired degree of approximation can be achieved.

The method of the straight rule leads in some cases to values of c_0 or e_0 which seem to me not adequate. First, the result that $m_0 = 0$ for every nonhomogeneous state-description seems unsatisfactory because these state-descriptions obviously represent possible cases. This remark, however, is directed only to those who accept m-functions, not to the authors who propose the method of the straight rule, because they usually reject tautological evidence and hence any function m. Let us therefore now consider cases with factual evidence. For $s > 0$ and $s_M = 0$, the straight rule (14-4) yields $c_0 = 0$. The hypothesis predicts in this case that an unobserved individual has the property M, for which so far no instances have been observed. Suppose, e.g., that M is $P_1 . P_2$, that we have observed some instances of P_1 and some of P_2 but none of $P_1 . P_2$. Then the probability (in the logical sense) that an unobserved individual is $P_1 . P_2$ may be low, but it cannot be regarded as 0 because it is, after all, a possible case, i.e., one logically compatible with the given evidence. Further, consider the case $s_M = s > 0$, i.e., all observed individuals are found to be M. Then $c_0 = 1$ not only for the singular hypothesis h that the next individual is M, but even for the universal law that all individuals are M. This result seems hardly acceptable, the more so since

48. [I], §55A.

it holds for every $s > 0$ however small, even for $s = 1$. Thus, if all we know about the universe is that the only observed thing is *M*, then the method of the straight rule leads us to assume with inductive certainty (i.e., $c_0 = 1$) that all things are *M* and to take the estimate of the rf of *M* in the universe to be 1. The strangeness of these results becomes especially striking when we remember that both a value of c for the singular predictive inference and a value of an estimate of rf may be interpreted as stating a fair betting quotient.[49] Thus, on the basis of the observation of just one thing, which was found to be a black dog, the method of the straight rule declares a betting quotient of 1 to be fair for a bet on the prediction that the next thing will again be a black dog and likewise for a bet on the prediction that all things without a single exception will turn out to be black dogs. Now a betting quotient of *u* means odds (ratio of stakes) of u: $(1 - it)$; hence a betting quotient of 1 means odds of 1:0. Thus this method tells you that if you bet in the case described any amount on either of the two predictions mentioned, while the stake of your partner is 0, then this is a fair bet. I wonder whether there is any sane man, whether scientist, businessman, or gambler, who would actually regard a bet of this kind as fair. I think that these examples show that the method of the straight rule, although not totally inadequate like c_∞ is not quite satisfactory in application to very small samples. We shall later point out the weakness of the straight rule of estimation from another point of view, with respect to its mean square error in the sampling distribution (§§ 23, 24),[50]

On the other hand, there is no need to reject outright every application of the straight rule in practical work. In cases where s and s_M are sufficiently large, it does not make a great difference in the values of $c_\lambda(h_M,e_M)$ and $e_\lambda(\text{rf},M,K,e_M)$, whether $\lambda = 0$ is chosen or a not too large positive λ, e.g., $\lambda = 2$, as in the modified Laplace method (§ 12). Therefore, in cases of this kind the straight rule may be used as a convenient approximation for the values of c or e supplied by more adequate nonextreme methods. The convenience of the straight rule is due to its great simplicity; but it is bought at the price of neglecting the logical factor w/K.

The straight rule, especially that of estimation, is implied by various historically given methods. R. A. Fisher's *method of maximum likelihood* is a general method for the estimation of magnitudes characteristic of a popu-

49. [I], § 41B, C, D.

50. The method of the straight rule will be examined more thoroughly and its disadvantages explained in greater detail in [II].

lation, on the basis of a given sample from the population.[51] If it is applied to the estimation of rf, it leads to the straight rule of estimation.[52]

Many statisticians today seem to accept, explicitly or implicitly, the principle that, if a choice among available estimate-functions is to be made, a so-called "*unbiased*" one is preferable. This concept will be explained later (§ 19), and it will be shown that e_0 is the only unbiased function among the functions e_λ. Thus the principle mentioned. leads to the acceptance of the straight rule of estimation.

One gets the impression that, in the problem of estimating the rf of *M* in a population (or, in the terminology of statistics, the "probability" of *M*), there is general agreement among the various schools of statisticians: *the* solution consists in taking as the estimate the observed rf in the sample, hence what we call the straight rule of estimation. Kendall, after explaining the various approaches, sums up the situation: "In this case, therefore, all the approaches lead to the same conclusion (a happy state of affairs which ... does not always exist)."[53]

Reichenbach lays down the following *rule of induction*: "If an initial section of is elements of a sequence x_i is given, resulting in the frequency *f*", and if, furthermore, nothing is known about the probability of the second level for the occurrence of a certain limit *p*, we posit that the frequency f^i $(i > n)$ will approach a limit *p* within $f^i \pm \delta$ when the sequence is continued."[54] Although the term 'estimation' is not used, the explanations and examples given by the author seem to show that the rule is meant as a rule of estimation which might be formulated in our terminology as follows: "Let the evidence *e* say that in an initial section containing *s* individuals of an infinite sequence there are s_M with the property *M* and let *e* not contain any further factual information; then take s_M/s as the estimate for the limit of the rf of *M* in the sequence." If this interpretation is correct, then the rule of induction is essentially the same as the straight rule of estimation. Reichenbach does not have a method of confirmation, because he rejects the concept of degree of confirmation.

§ 15. Inductive Methods of the Second Kind: λ Is Dependent upon *κ*

We assign to the second kind of inductive methods those for which the λ-functions, determined from their *G*-functions by (9-4), are dependent

51. R. A. Fisher, "On the mathematical foundations . . .", p. 323; Cramér, *op. cit*., pp. 498 ff.; Kendall, *op. cit*., I, 178 ff., II, 12 ff.

52 Fisher, "On the mathematical foundations ...," p. 324; Kendall, *op. cit*., I, 199.

53. Kendall, *op. cit*., I, 200.

54. Reichenbach, *op. cit*., p. 446.

upon κ. Since all λ-functions for the methods which we here take into consideration are independent of s and s_i (C11, § 9), κ is here the only argument; thus we write a λ -function here in the form '$\lambda(\kappa)$'. That a function (κ) is actually dependent upon κ, in other words, that κ is not a vacuous argument, means that there are at least two distinct values of κ, say κ_1 and κ_2, such that $\lambda(\kappa_1) \neq \lambda(\kappa_2)$

It seems that the only inductive method of the second kind which has been proposed or even considered so far is that represented by the function c*, which I defined in earlier publications.[55] First, m* is defined as that m-function which is symmetrical (C7, § 4) and has equal values for all structure-descriptions (§ 13). Then c* is defined as based upon m* by (5-6), and e* as based upon c* by (6-2). These definitions yield:

(15-1) $c^*(h_i,e_i)$

(15-2) $e^*(\mathrm{rf},Q_i,\kappa,e_i)$

$$\left.\begin{array}{l} c^*(h_i,e_i) \\ e^*(\mathrm{rf},Q_i,\kappa,e_i) \end{array}\right\} = \frac{s_i+1}{s+\kappa}$$

and hence, with (6-9) and (6-4)

(15-3) $c^*(h_M,e_M)$

(15-4) $e^*(\mathrm{rf},M,\kappa,e_M)$

$$\left.\begin{array}{l} c^*(h_M,e_M) \\ e^*(\mathrm{rf},M,\kappa,e_M) \end{array}\right\} = \frac{s_M+w}{s+\kappa}$$

According to (15-1), this method is characterized by the following *G*-function, *G**:

(15-5) $$G^*(\kappa,s,s_i) = \frac{s_i+1}{s+\kappa}.$$

and hence, according to (9-4), by the following λ-function, λ*:

(15-6) $$\lambda^*(\kappa) = \kappa.$$

We shall find later (see,(15-16)) that this λ -function leads indeed to the result that all structure-descriptions have the same m*-value. For numerical examples of (15-3) and (15-4) see table (12-19).

We shall restrict our present investigation to a certain class of methods of the second kind obtained by a simple generalization of the only method of this kind known so far, which is characterized by $\lambda^*(\kappa) = \kappa$. This class contains those λ -functions which are multiples of κ:

(15-7) $\lambda(\kappa) = C\kappa$, with a constant coefficient *C*.

C may be any finite positive number. (For $C = 0$ and $C = \infty$ we obtain $\lambda = 0$ and $\lambda = \infty$, respectively, which are methods of the first kind.)

55. The definition of c* and some theorems (without proofs) were given in "On inductive logic," *Philos. of Science*, 12 (1945), 72-97, and in [I], Appendix, § 110. The full theory of c* will be developed in [II].

Let us briefly examine a more comprehensive class, that of the linear functions of κ:

(15-8) $\lambda(\kappa) = C_0 + C\kappa$.

We might regard this as a general schema for λ-functions of both kinds. With $C = 0$ it would yield the functions of the first kind, with $C > 0$ a more comprehensive class of functions of the second kind. Let m be the m-function corresponding to the λ-function $C_0 + C\kappa$ with given constants C_0 and C. Then we find the value of tit for a state-description k by substituting in (10-7) '$C_0 + C\kappa$' for 'λ' (which there stands for '$\lambda(\kappa)$'). We see that the result is rather complicated. But it becomes much simpler in the cast $C_0 = 0$, $\lambda(\kappa) = C\kappa$; hence $\lambda(\kappa)/\kappa = C$:
(C(C-}-1)(C+2)...(C-f-N;-1)1

(15-9) $$\mathrm{m}(k) = \frac{\prod_i [C(C+1)(C+2)...(C+N_i-1)]}{C\kappa(C\kappa+1)(C\kappa+2)...(C\kappa+N-1)},$$

Therefore, we shall restrict bur considerations to λ -functions of the simple form (15-7). For this form, we obtain from (9-6) to (9-9):

(15-10) $\mathrm{c}(h_i,e_i)$

(15-11) $\mathrm{e}(\mathrm{rf},Q_i,K,e_i)$

$$= \frac{s_i + C}{s + C\kappa},$$

(15-12) $\mathrm{c}(h_M,e_M)$

(15-13) $\mathrm{e}(\mathrm{rf},M,K,e_M)$

$$= \frac{s_M + Cw}{s + C\kappa}.$$

The functions stated in these theorems become especially simple if C is a positive integer. In this case, (15-9) becomes:

(15-14) $$\mathrm{m}(k) = \frac{(C\kappa-1)!}{C\kappa+N-1)!}\prod_i \frac{(C+N_i-1)!}{(C-1)!}.$$

Let us consider two examples.

1. $C = 1$, hence $\lambda(\kappa) = \kappa$. This is the method of c* mentioned earlier. From (15-14):

(15-15) $$\mathrm{m}^*(k) = \frac{(\kappa-1)!}{N+\kappa-1)!}\prod_i N_i!.$$

Thus, c* represents the simplest method of the second kind.

Incidentally, every state-description which is isomorphic to k in the sense of having the same Q-numbers N_i has the same m*-value. The corresponding structure-description k' is the disjunction of these state-descriptions (§ 13). Therefore, according to the special addition principle

(C4, § 4) m*(k') is obtained by multiplying m*(k) with the number of the isomorphic state-descriptions, which is N!/ $\prod_i N_i!$.[56] Hence:

(15-16) $$\mathrm{m}^*(k') = \frac{(\kappa-1)!N!}{(N+\kappa-1)!} = \frac{1}{\binom{N+\kappa-1}{\kappa-1}}.$$

Since this is independent of the Q-numbers N_i; every structure-description has the same m*-value.

2. $C = 2$, hence $\lambda(\kappa) = 2\kappa$. From (15-14):

(15-17) $$\mathrm{m}(k) = \frac{(2\kappa-1)!}{N+2\kappa-1)!}\prod_i (N_i+1)!$$

From (15-10) to (15-13):

(15-18) $\mathrm{c}(h_i,e_i)$

(15-19) $\mathrm{e}(\mathrm{rf},Q_i,K,e_i)$ $\Bigg\}$ $= \dfrac{s_i+2}{s+2\kappa}$,

(15-20) $\mathrm{c}(h_M,e_M)$

(15-21) $\mathrm{e}(\mathrm{rf},M,K,e_M)$ $\Bigg\}$ $= \dfrac{s_M+2w}{s+2\kappa}$.

The numerical values stated in table (12-19) for λ = 4, 8, 16, 32, hold for this method in a system with κ = 2, 4, 8, 16, respectively.

One might, of course, consider a much wider class of methods of the second kind, including also nonlinear functions of κ, either all of them or certain kinds. Our present restriction to linear functions and even to those of the simple form (15-7) is motivated by the fact that our investigations of methods of the second kind have so far not led to any practical reasons for going beyond this simple form. But it may, of course, very well be that this situation will change in the future; and then it will be time to abandon the present restrictions.

§ 16. The Difference between the Two Kinds of Inductive Methods

Since we restrict the investigation of methods of the second kind to the λ -functions of the form CK (15-7), we should, strictly speaking, regard C as the characteristic parameter of these methods. It is, however, more convenient to use a common way of speaking for methods of both kinds by referring to λ as a *parameter* in both cases (instead of using the term 'λ -function' in the second case). In a method of the first kind a value of the parameter λ is characteristic of an inductive method in itself, irrespective of language-systems, while for a method of the second kind a value of the

56. [I], § 35, T35-4.

parameter λ is characteristic of the given method with respect to a given language-system or, in-other words, a given value of K. It seems to me that the difference between the two kinds of methods is not so fundamental and Important as it might appear at first glance. One might perhaps think that in practical work with a method of the second kind, in contradistinction. to one of the first kind, it would be necessary to change the value of λ all the time. This, however, is actually not the case. If an investigation concerning a given universe of discourse (say, the universe of atoms in the physical world or that of observable physical events or that of human beings) studies inductive inferences with various bodies of possible evidence and various forms of hypotheses or estimates of rf, then this whole investigation refers to one fined set of relevant properties of the individuals in question and thus to one language-system, or at least to one number K, while the number N of individuals may be disregarded for most questions (it becomes relevant only for hypotheses containing quantifiers, e.g., universal laws). "Thus for an investigation of this sort λ has a fixed value which remains the same for all the various questions studied within that investigation. A changes only with a change of λ, hence only in the transition from one universe to another universe with a different multiplicity of properties. For any given universe there is only one adequate language-system or at least only one adequate value of K. (Strictly speaking, different language-systems may be adequate for a given universe, but only if they are intertranslatable and differ merely in the selection of primitive properties; in this case the systems are in a certain sense equivalent and have the same K, i.e.. the same number of the strongest factual properties expressible.)

Let its consider, as an example, the method of the second kind characterized by the λ-function $\lambda(K) = 2K$. For a system L^{π}_{n} , (with any N), $K = 2^3 = 8$; hence $\lambda(K) = 16$. Thus, in application to this system, we may simply write '$\lambda = 16$', because in this domain the given method of the second kind coincides with the method of the first kind characterized by $\lambda = 16$. For the given system, both methods are represented by the same functions c_{16}, m_{16}, and e_{16}.

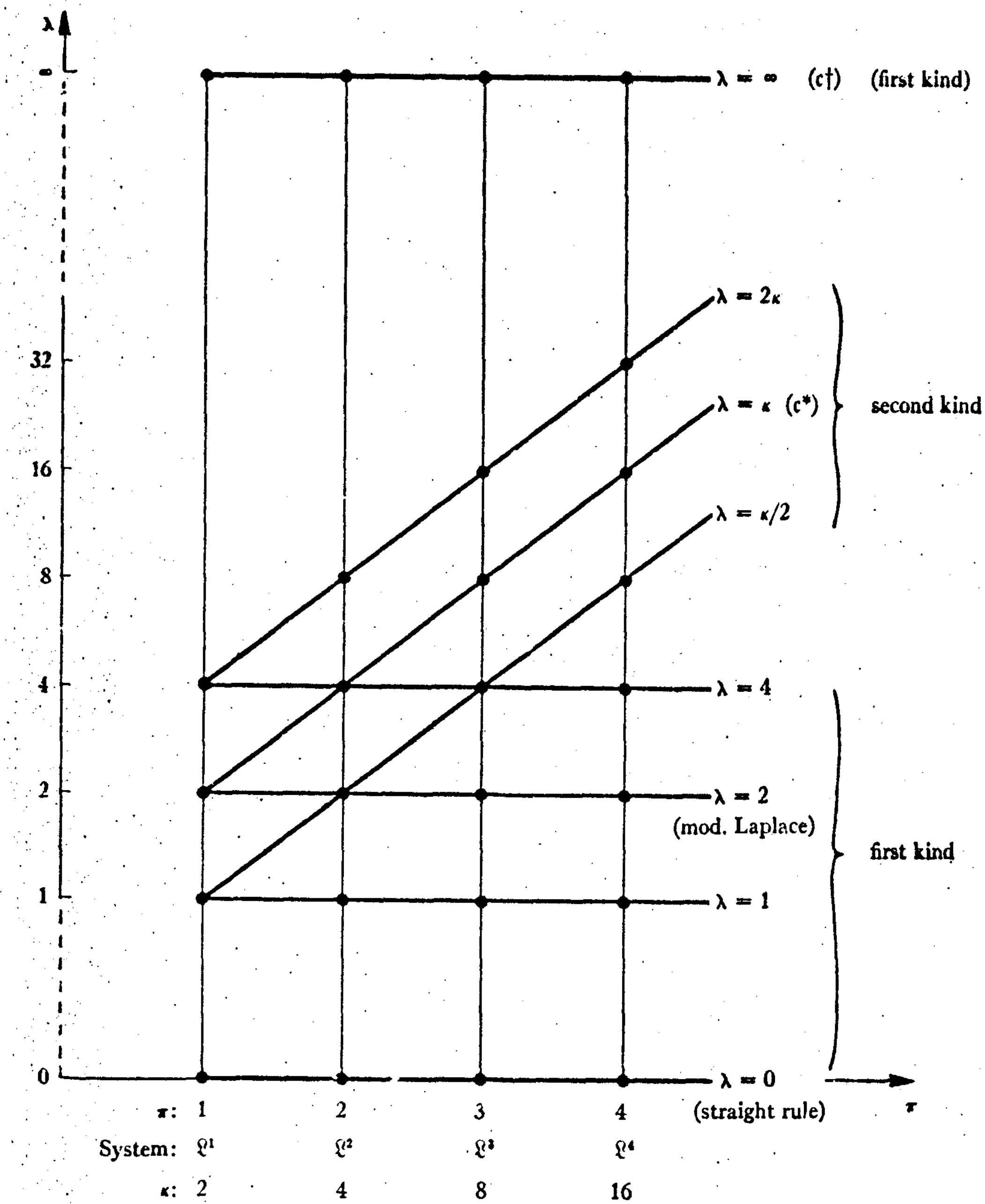
In certain cases the methods of the first kind have indeed the advantage of greater simplicity in comparison with those of the second kind. Suppose that X wishes to study some questions concerning a few properties M_1, M_2, etc., which are defined as molecular properties in terms of a few primitive properties, say. P_1, P_2, and P_3; the latter must be (deductively) independent of each other.[57] It may be that X is aware that there are still

57. [I], § 18B.

other primitive properties P_4, P_5, etc., in the universe in question, but he would like to disregard them, if possible, because they do not happen to appear in the definitions of the molecular properties under investigation. And, above all, he would like to avoid the bothersome problem of ascertaining the total number π of primitive properties in the universe in question, and the number K of the strongest factual properties in the universe. He cannot escape this problem if he has chosen a method of the second kind, because in this case values of c and e like those stated in (15-10) to (15-13) depend upon **a** and hence cannot be established without a knowledge of K. On the other hand, if *X* has chosen a method of the first kind, he does not need any knowledge of K but merely the definitions of the molecular properties under investigation. As we see from (11-6) and (11-7), K enters those values of c_λ and e_λ only through w/K; now the relative width w/K of *M* is uniquely determined by the definition of *M* in terms of primitive properties. [For example, the relative width of P_1 is ½, that of $P_1 \vee {\sim}P_2$ is ¾ that of $P_1 \vee (P_2 . {\sim}P_3)$ is ⅝, irrespective of the number of additional primitive properties.]

The difference between the methods of the first and the second kind can be easily recognized from the accompanying *diagram*. On the left-hand side is the scale for λ. For the sake of convenience, a logarithmic scale form is used for finite positive values (this has the effect that also the methods of the second kind are represented by straight lines). The values 0 and ∞ on the scale are separated from the others by dotted lines in order to indicate that they are at infinite distances. The other vertical lines represent the four simplest language-systems, characterized by values of π (1, 2, 3, 4) and K (2, 4, 8, 16). The heavy lines represent some examples of inductive methods by indicating their λ -values for the four systems. A method of the first kind has the same λ for all systems; therefore, it is represented by a horizontal line (given for λ = 0, l, 2, 4, ∞). A method of the second kind has λ-values which increase with increasing a and is therefore represented by a slanted line. The examples are for $\lambda = K/2$, K(c*), and $2K$. The diagram shows that, with respect to a given language-system, the numerical value of λ is all that matters, and that an inductive method which is general, i.e., applicable to various systems, is simply a set of λ-values for the systems, equal values in the case of a method of the first kind, different values in the case of a method of the second kind.

There are any number of procedures for characterizing inductive methods by sets of parameters. The procedure of the λ-parameter here developed has the advantage of great simplicity, since, with respect to a



THE TWO KINDS OF INDUCTIVE METHODS

given language-system, each method (of, the class here investigated) is completely characterized by the numerical value of one parameter. The same holds, of course, for any other parameter which is determined from λ by any univalued function *f*(λ) whose inverse is likewise univalued (i.e:, *f* is such that for $\lambda_1 \neq \lambda_2$ $f(\lambda_1) \neq f(\lambda_2)$). Here are three simple examples: the parameter λ′ defined by 1/λ; λ″ = 1/(λ + 1); λ‴ = λ/(λ + 1). When λ runs from 0 to ∞, λ runs from ∞ to 0, λ″ from 1 to 0, and λ‴ from 0 to 1. λ″ and λ‴ have the advantage that all their values are finite. λ has the advantage of a simple interpretation: the weight given to the logical factor is the same as that given to the observation of λ individuals. However, aside from these considerations of convenience, it makes, of course, no essential difference which one among the parameters mentioned or other equivalent ones is chosen.

§17. Complete Inductive Methods

A method of confirmation for the systems L and its function c may be regarded as complete if c has a value for any pair of sentences *h*,*e* (*e* not L-false) in any system L^{π}_{n}. A method of estimation for rf in the systems L and its function e may be regarded as complete, if e has a value e(rf,*M*,*K*,*e*) for any triple *M*,*K*,*e* in any system L^{π}_{n}, provided that *M* is a molecular property, *K* a nonnull class of individuals, and *e* a non-L-false sentence. It is easily seen that a function e based upon a complete function c is itself complete. A *complete inductive method* for the systems L is a combination of a complete c and a complete e. The elements of our λ-system are complete inductive methods in this sense. Each consists of a complete c and an e based upon c. According to the earlier explanations, any method of this kind in the λ-system possesses a characteristic function *G* such that (6-15) and (6-16) hold for any h_i, e_i, *M*, and *K* of the kinds previously specified, and a λ-function determined by (9-4).

The theories concerning one or the other aspect of inductive reasoning (e.g., theories of probability in the inductive sense, of statistical inference, or of estimation) which have been constructed up to now do not usually supply a complete inductive method in the sense just explained. (The method represented by c* and e* is perhaps the only exception.) However, the incomplete methods available so far can usually be supplemented so as to form a complete inductive method in the λ-system. We shall now discuss the procedure of supplementation; dividing the possible cases into three kinds: (1) only a method of confirmation is given, (2) only a method of estimation for rf is given, but no method of confirmation, (3) both kinds of methods are given. Historically given methods are here considered only to the extent of their application to the language systems L.

1. Suppose that an author gives only a method of confirmation, represented by a function c, but no method of estimation for rf. Usually c is incomplete. For instance, many authors exclude L-true sentences as evidence (in other words, an empty sample, $s = 0$); some of them exclude, in addition, samples of very small size (e.g., $s = 1$; the required minimum size is sometimes not exactly specified). If, however, a method of confirmation in our sense is given at all, that is, if rules determining numerical values. c are given at least for some nontrivial cases, their these comprise usually, or perhaps always, cases of the singular predictive inference. If so, these rules determine immediately the characteristic function G of the method and its λ-function. Then we can supplement c to a complete function c′ by determining the values of c′ for the remaining cases with the help of the λ-function, as explained earlier (§ 10). Finally, we add to c′ the complete function e′ based upon it; it has the same λ-function as c′.

2. Suppose that an author proposes or discusses a method of estimation for rf represented by a function e, but no method of confirmation. This is frequently the case, e.g., for Reichenbach and the majority of contemporary workers in mathematical statistics. Usually a method of this kind is incomplete; here again cases of empty or even very small samples are often excluded. But at least some nontrivial values are assigned to e, and they comprise usually cases of the predictive estimation for rf, M, K, e_M, as previously specified. If so, these values will immediately supply the characteristic function G by (6-17) and hence the λ-function. Then we transform e into the complete inductive method consisting of those complete functions c′ and e′ which are characterized by the resulting λ-function. Thus we have seen, for example, how various methods of estimation of rf which lead to the straight rule of estimation but are represented by an incomplete function e, excluding an empty sample and maybe also very small samples (e.g., the maximum likelihood method), are transformed into the complete function c_0, to which the complete function c_0 is then added (§ 14). In this case the limit-convention plays an essential role in the procedure of supplementation.

3. Suppose that an author presents or investigates a method which involves both a method of confirmation and a method of estimation for rf, represented by the functions c and e, respectively. So far, very few authors have studied combined methods of this kind. If such a method is given, we can establish the characteristic function for c, say G_1, and separately that for e, say G_2. Then G_1 yields a λ-function λ_1, and G_2 yields λ_2. The resulting functions determine whether the two parts of the method fit together or not. They do if and only if $\lambda_1$and $\lambda_2$are the same function. It

seems that in the few historically given cases this condition is fulfilled. Then the method is extended so as to supply complete functions c′ and e′ determined by the common λ-function. If it is found, however, that λ_1 and λ_2 are not the same, then c and e must be regarded not as parts of one inductive method in our sense but as belonging to two distinct methods. Each of the two parts could then be transformed separately with the help of, its λ-function into a complete inductive method.

§ 18. The Choice of an inductive Method

The λ-system represents an infinite number of inductive methods, each characterized by a function or number λ. Thus the question arises as to which of the available methods a man *X* ought to choose if he wants to determine degrees of confirmation and estimates on the basis of his observational results. This is fundamentally not a theoretical question. A possible answer to a theoretical question is an assertion; as such it can be judged as true or false, and, if it is true; it demands the assent of all. Here, however, the answer consists in a practical decision to be made by *X*. A decision cannot be judged as true or fare but only as more or less adequate, that is, suitable for given purposes. However, the adequacy of the choice depends, of course, on many theoretical results concerning the properties of the various inductive methods; and therefore the theoretical results may influence the decision. Nevertheless, the decision itself still remains a practical matter, a matter of *X* making up his mind, like choosing an instrument for a certain kind of work. First, *X* has to decide whether to choose a method of the first or the second kind. As we have seen, a method of the first kind has the advantage that only those primitive properties which actually occur in *e* and *h* have to be taken into consideration for the calculation of c(*h*,*e*). Suppose that *X*, for this or other reasons, prefers a method of the first kind. Then he has to choose as a value of the parameter λ any positive real number or 0 or ∞. He learns about the properties of the method with $\lambda = \infty$, which were stated earlier (§ 13), especially the fact that c_∞ and e_∞ are independent of *e* (which is a theoretical result). Suppose he regards these properties as unsatisfactory (which is a practical evaluation) and hence rejects this method (which is a practical decision). Suppose he rejects also those methods for which λ is very large, though finite, because they have the same unsatisfactory properties, though to a lesser degree. What might be *X*'s attitude toward the other extreme method, the straight rule, characterized by $\lambda = 0$? He may perhaps accept it, following the majority of statistic fans and Reichenbach, or he may not. Let us suppose that he rejects it in view of those properties.

which seem to me disadvantageous (of which only a few are mentioned in this monograph). He may. then likewise reject those related methods for which λ is near to 0 (a small positive. fraction), because they have the same disadvantageous properties, though to a lesser degree. This would still leave *X* the choice from an infinite number of nonextreme methods in the medium range of the λ-scale. There are various points of view which might have more or less weight for his decision. He may have the feeling (as I do) that small λ-values seem to lead to more adequate values of c and e. Integers lead to simpler formulas (see the denominator in (12-1)), and especially powers of 2 (since λ/κ occurs frequently and κ is always a power of 2). Thus he might consider (as I would) in the first place λ = 2, the modified Laplace method, or λ = 1 (§ 12); in the second place 4 or 8; further, 16, etc.; also ½, ¼, etc.

Now let us suppose that *X* prefers a method of the second kind, with $\lambda(\kappa) = C\kappa$. If he chooses a positive integer for the coefficient *C*, the formulas become much simpler than in any nonextreme method of the first kind (see (15-14)). He might take into consideration, in the first place, *C* = 1 (as I would), hence $\lambda(\kappa) = \kappa$, i.e., the function c*. The definition of this function c* is of great simplicity and elegance; it is simpler than that of any other c-function that comes into consideration (excluding c† as entirely inadequate). In the second place, he might regard *C* = 2 or 3; and further perhaps a few more subsequent integers, but not many. With larger values of *C*, λ would be rather large for systems with many primitive properties, and then the adequacy of the resulting values of c and e would become dubious. (For example, even with *C* = 6, for a system with a- = 8; hence $\kappa = 2^8 = 256$, we have λ = 1,536, which might be regarded as too high for adequacy.) Thus only a small number of methods of the second kind seem advantageous.

Suppose that *X* has chosen a certain inductive method and used it during a certain period for the inductive problems which occurred. If, in view of the services this method has given him, he is not satisfied with it, he may at any time abandon it and adopt another method which seems to him preferable. This is not the same as a change in method from problem to problem. Once he adopts the new method, he will apply it to all inductive problems, problems of confirmation for all kinds of hypotheses; of estimation for all kinds of situations (and, ultimately, for all kinds of magnitudes, according to the general concept of estimation explained in § 6); of choosing a practical decision; etc. One inductive method is here envisaged as covering all inductive problems. How can *X* go over from one inductive method to another? It is not easy to change a belief at will;

good theoretical reasons are required. It is psychologically difficult to change a faith supported by strong emotional factors (e.g., a religious or political creed). The adoption of an inductive method is neither an expression of belief nor an act of faith, though either or both may come in as motivating factors. An inductive method is rather an instrument for the task of constructing a picture of the world on the basis of observational data and especially of forming expectations of future events as a guidance for practical conduct. *X* may change this instrument just as he changes a saw or an automobile, and for similar reasons. If *X*, after using his car for some time, is no longer satisfied with it, he will consider taking another one, provided that he finds one that seems to him preferable. Relevant points of view for his preference might be: performance, economy, aesthetic satisfaction, and others. Similarly, after working with a particular inductive method for a time, he may not be quite satisfied and therefore look around for another method. He will take into consideration the performance of a method, that is, the values it supplies and their relation to later empirical results, e.g., the truth-frequency of predictions and the error of estimates; further, the economy in use, measured by the simplicity of the calculations required; maybe also aesthetic features, like the logical elegance of the definitions and rules involved. The λ-system makes it easy to look for another inductive method because it offers an inexhaustible stock of ready-made methods systematically ordered on a scale. If *X* feels that the method he has used so far does not give sufficient weight to the empirical factor in comparison to the logical factor, lie will choose a method with a smaller λ—a little smaller or much smaller, according to his wishes. On the other hand, if he wishes to give more influence to the logical factor and less to the empirical factor, he will move up his mark on the λ-scale. Here, as anywhere else, life is a process of never ending adjustment; there are no absolutes, neither absolutely certain knowledge about the world nor absolutely perfect methods of working in the world.

PART II

COMPARISON OF THE SUCCESS OF INDUCTIVE METHODS

§ 19. Sampling Distributions

In the first part, the λ-system of inductive methods was constructed, in which each method is characterized by its λ. In this part we shall develop a procedure for measuring the success of a given inductive method within a given state-description. This procedure will be based on the concept of the mean square error of an estimate-f unction in the sampling distribution of a state-description. It will then be shown how for any given state-description the optimum inductive method can be determined. The comparative study of inductive methods carried out in this part is not intended to show that one particular method is superior to all others. The purpose is rather to clarify the situation by exhibiting the comparative properties of the various methods. This will supply to anyone who wishes to choose an inductive method a rational foundation for his choice.

In preparation for the analysis of success, we shall define in this section some well-known elementary statistical concepts concerning sampling distributions and estimate-functions and list some formulas involving these concepts. These results will be applied to the λ-methods in subsequent sections.

We shall mention, first, a few general statistical concepts. Let *u* be a variate (e.g., a measurable magnitude or the cardinal number of a class) which has the values $u_1, u_2, \ldots, u_n$ in a given set of n cases (the values are not necessarily different). We define:

(19-1) The *mean* of *u*: $\bar{u} = \prod_{i=1}^{n} u_i/n$.

(19-2) The *mean square deviation* of it with respect to a fixed value *a*:

$$\overline{(u-a)^2} = \sum_i (u_i - a)^2 / n \,.$$

(19-3) The *variance* of *u* ($\sigma^2(u)$) is the mean square deviation of *u* with respect to $\bar{u}$:

$$\sigma^2(u) = \sum_i (u_i - \bar{u})^2/n.$$

The variance is the smallest value of the mean square deviation. For any reference point a different from $\bar{u}$ we have:

(19-4) $$\overline{(u-a)^2} = \sigma^2(u) + (\bar{u} - a)^2.$$

Hence, for $a = 0$:

(19-5) $$\overline{u^2} = \sigma^2(u) + \bar{u}^2.$$

If the mean and the variance of u are known, we can easily find the mean and the variance of a linear function of u, with any constants b and c:

(19-6) $$\overline{bu} = b\bar{u},$$
(19-7) $$\overline{u+c} = \bar{u} + \mathrm{c},$$
(19-8) $$\overline{bu+c} = b\bar{u} + \mathrm{c}.$$
(19-9) $$\sigma^2(u + c) = \sigma^2(u)$$
(19-10) $$\sigma^2(bu) = \sigma^2(bu + c) = b^2\sigma^2(u)$$

Let a system L_n^π be given and in it a state-description k. Let e_M, (as in § 4) describe a sample of size s ($s > 0$) in k containing s_M individuals with the property M. Let K be the unobserved part of the universe (or population), i.e., the class of the $N - s$ individuals not mentioned in e_M. We assume that N is very large in relation to s, so that we can identify the unobserved part with the whole universe with sufficient approximation. Let N_M be the number of individuals in k which have the property M. Then the rf of M in the universe as described in k is $r = N_M/N$; this we shall take (approximately) also as the rf of M in K. Let e be an estimate-function for rf; we shall study its values e(rf,M,K,e_M) in various cases.

We consider the totality of possible samples of the fixed size s from the given population as described in k, that is, the set of all those subclasses of the universe which contain s individuals. Let S be their number. Any method of estimation can be applied to each of these samples, and we consider the distribution of the resulting values among the S samples. This is called the *sampling distribution* of the values in question. We shall study the sampling distributions of the estimates and of their errors.

We shall now apply the statistical concepts defined above to the sampling distribution. The number s_M of individuals with M in the samples of size s varies from 0 to s. For any given value of s_M the number of samples for which this value holds can easily be determined. The results are as follows:

(**19-11**) The number of all samples of size s is $S = \binom{N}{s}$.

(19-12) The exact number of samples with a given value of s_M is $S_M = \binom{N_M}{s_M}\binom{N - N_M}{s - s_M}$.

Hence the proportion of these samples among all samples of the size s is S_M/S. If N is very large in relation to s, as is presupposed in our discussion, then the following holds approximately:

(**19-13**) The proportion of samples with a given s_M (and hence the statistical probability that a random sample has a given value of s_M) is

$$\binom{s}{s_M} r^{s_M}(1-r)^{s-s_M}.$$

This is known as the binomial law.[1] On this basis the mean and the variance of s_M for the samples of the fixed size s can be ascertained; the results are well known:[2]

(**19-14**) $$\overline{s_M} = sr,$$

(**19-15**) $$\sigma^2(s_M) = sr(1 - \mathrm{r}).$$

Hence with (19-6) and (19-10):

(**19-16**) $$\overline{s_M / s} = r,$$

(**19-17**) $$\sigma^2(s_M/\mathrm{s}) = r(1 - \mathrm{r})/s.$$

The *error* v(e,x,e) of any estimate e(x,e) of a magnitude x on evidence e is the difference between the estimate and x itself[3] (the latter means either the actual value of the magnitude or its assumed value in a given state-description):

(**19-18**) $$\mathrm{v}(\mathrm{e},x,e) = \mathrm{e}(x,e) - x.$$

(In the following we shall often write simply 'v' and 'e', omitting the arguments when they are obvious from the context.) The mean of the error is, according to (19-7):

(**19-19**) $$\bar{\mathrm{v}} = \bar{\mathrm{e}} - x.$$

1. [I], §95, T95-1c. (This theorem states the value of c in the direct inference. This value is equal to the proportion of samples of the kind in question.)

2. M. G. Kendall, *Advanced theory of statistics*, I (London, 1943), 117, 197 ff.; H. Cramér, *Mathematical methods of statistics* (Princeton, 1946), 1).193.

3. [I], § 102, D102-1.

The variance of the error is, according to (19-9), equal to the variance of the estimate:

(19-20) $$\sigma^2(\mathrm{v}) = \sigma^2(\mathrm{e}).$$

In the terminology of statistics, an estimate-function e for some magnitude x is said to be *unbiased* if, for any population and any sample size s, the mean of the estimate is equal to the actual value of the magnitude in the population.[4] If the function e is not unbiased, the difference $\overline{\mathrm{e}(x,e)} - x$. is called the bias; according to (19-19) it is equal to the mean error $\bar{\mathrm{v}}$.

Of special interest is the *mean square error*, which we denote by '$1\mathrm{v}^2(e,x)$' (often briefly '$1\mathrm{v}^2$').[5] It is defined as the mean of 'v^2', the square of the error of the estimate e, in the sampling distribution:

(19-21) $$1\mathrm{v}^2(e,x) = \overline{\mathrm{v}^2},$$

Hence with (19-5):

(19-22) $$1\mathrm{v}^2 = \sigma^2(\mathrm{v}) + \overline{\mathrm{v}^2},$$

or, according to (19-20):

(19-23) $$1\mathrm{v}^2 = \sigma^2(\mathrm{e}) + \overline{\mathrm{v}^2}.$$

§ 20. The Mean Square Error as a Measure of Success

A given inductive method can be studied in two different respects. On the one hand, its internal logical character may be analyzed. On the other hand, we may confront it with a given series of events or a whole world, either the actual universe or an assumed one described in a given state-description, and examine how well it performs if it is applied to various parts of the world in order to obtain degrees of confirmation or estimates concerning other parts. We shall study problems of this kind in this and subsequent sections. Our discussions will always refer to a given state-description, not to the actual state of the universe. Therefore, our problems are of a purely logical nature. Questions concerning the success of a given inductive method in the actual world would be of a factual, nonlogical nature. And if they concerned not merely that part of the world which is known to us by past observations but also a part or the whole of the future, then the answer could be given with certainty only after all observation reports were in, if that were ever possible. And if our question

4. Cramér, *op. cit*., pp. 351, 478; Kendall, *op. cit*., I, 200; II, 3 ff.

5. The mean square error $1\mathrm{v}^2$ must not be confused with the estimated square error f^2 ([I], § 103, D103-1a). The farmer is based on the *actual* square error v^2 and therefore can be determined only if the actual value x is known (or assumed in a given state-description). The latter, on the other hand, is the estimate of v^2 with respect to given evidence e and hence depends merely on e.

concerned not the actual success but the probability of success or an estimate of success, then it would make sense only on the basis of a chosen inductive method. The purpose of the intended study, however, is to examine the various inductive methods on a neutral basis without presupposing the acceptance of one of them. Therefore, we must relativize the problem with respect to state-descriptions. This has the advantage that our results hold with deductive certainty and hence must be accepted by anybody independently of the preferences he may have with respect to inductive methods. On the other hand, by framing the problem as a logical question, our investigation must necessarily abstain from making any judgment concerning the success of an inductive method in the total actual world. A judgment of the latter kind is obviously impossible from an inductively neutral standpoint.

Our problem involves the logical concept of the success of a given inductive method within a given state-description k. We may perhaps have a rough idea of what we mean by this measure of success, but not yet an explication[6] for it, that is, a way of representing in the form of an exactly defined concept what we vaguely have in mind. Thus this explication will be our first task. There are various ways of solving it. We shall first outline, without adopting it, a possible explication involving bets based on c-functions. Then we shall choose another explication based on e-functions. This will later be used in our analysis of success.

Suppose we wish to compare the success of two given inductive methods, represented by the functions c and c′, respectively, in a given state-description k with N individuals. We might do this by considering a comprehensive system of bets between X, using c, and X', using c′. We take a sample in k of size s, described in e_M, involving a property M. h_M attributes M to an individual outside the sample. We assume that e_M represents the available evidence for both X and X'. Then X regards $c(h_M,e_M)$ as a fair betting quotient for a bet on h_M, and $1 - c(h_M,e_M)$ for a bet on non- h_M; analogously for X' with $c'(h_M,e_M)$ and $1 - c'(h_M,e_M)$. Suppose that $c(h_M,e_M) > c'(h_M,e_M)$. Let q be the arithmetic mean of these two values. Then we let X bet on h_M with the betting quotient q (which is less than his value $c(h_M,e_M)$);. thus X' bets on non-h_M, with $1 - q$ (which is less than his value $1 - c'(h_M,e_M)$). We stipulate that the sum of the two stakes is 1. Hence the stake of X is q and that of X', $1 - q$. Simultaneous bets of this hind are made for each of the $N - s$ individuals outside the sample. Since the state-description is given to us (though, of course, unknown to the two bettors), we can calculate how many of these bets are

6. [I], chap. i.

won by X and how many by X', and thus we can set up the balance. Then we do the same for every other sample of the fixed size s. This whole procedure is carried out not only for some one property M, but for a suitable set of properties, say, for all primitive properties or for all Q's. Then we regard the over-all balance of X for the total set of bets as the measure of the relative success of c in comparison with c′ for the given state-description k. If X has a positive balance and hence X' a negative one, c is regarded as more successful in k than c′.

If we wish to have a measure of success for one function c alone, we might consider a similar system of bets between X, who uses c, and a fictitious semi-omniscient being Y who knows, not the state-description and hence all single facts, like an omniscient being, but all the frequencies. Y takes as his betting quotient for a singular prediction h_M involving M the rf of M in the part of the universe not covered by e_M. We take again as the measure of success of X in k his balance for the total system of bets; in this case, his balance is either negative or zero.

The measure of success which we shall actually use is based, not on bets, but on the errors of estimations. Thus it involves e(rf,M,K, e_M), not c(h_M,e_M); but from our point of view this males no essential difference because of the equality of these two values if e is based upon c.

Now let us see how the concepts defined in the preceding section may be used for the purpose of explicating the concept of a measure of success of a method of estimation. The mean $\bar{u}$ of a variate u represents an average, a central location of the scattered values u_i. The mean square deviation with respect to a represents a measure of the dispersion or scattering of the values around a. And, in particular, the variance represents a measure of the dispersion of the values u_i around their mean $\bar{u}$. Now an estimate-function e for a magnitude x is not good if its values, based on samples of the fixed size s in the state-description k, are dispersed too widely. Therefore, it seems desirable for e to be such that the mean square deviation of its values in the sampling distribution be not too large. But which reference point should be taken here? The mean $\bar{\text{e}}$ as reference point leads to the variance σ^2(e). This measures merely the closeness *to each other* of the various estimates, based on various samples. Other things being equal, this concentration of the values is also a desirable property of e; but it cannot be regarded as decisive. The closeness of the estimates *to the actual value* x is more suitable as a measure of success; hence, in our problem, the closeness of the estimates to the value x given by k. Therefore, we shall regard the function e as more successful in k, the smaller the mean square deviation of the values of e with respect to x, in other words; the mean square error $1v^2(e,x)$ in the sampling distribution in k. [To measure the adequacy of an estimate-function e by its variance is justified only if e is such that its mean $\bar{e}$ and the actual value always coincide, in other words, if e is unbiased (§ 19). This is the case for most of those estimate-functions applied by statisticians; hence their customary reference for estimate-functions with minimum variance (the so-called most-efficient estimate-functions[7]). The functions e_λ in our system, however, are mostly not unbiased, as we shall see soon; therefore, it would not be justified here to require minimum variance.]

We return now to our estimate-functions e_λ. We had (11-7):

(20-1) $$e_\lambda(\text{rf},M,K,e_M) = \frac{s_M + (w/\kappa)\lambda}{s+\lambda}$$

We refer now again to the given state-description k, in which the rf of M is r. We find the mean and the variance of the estimates supplied by a particular function e_λ with a fixed λ for all samples of the fixed size s in k from (20-1), (19-14), and (19-15) with the help of (19-8) and (19-10):

(20-2) $$\overline{e_\lambda} = \frac{sr + (w/\kappa)\lambda}{s+\lambda},$$

(20-3) $$\sigma^2(e_\lambda) = \frac{sr(1-r)}{(s+\lambda)^2}.$$

A function e_λ is unbiased if and only if always $\overline{e_\lambda} = r$ (§ 19). We see from (20-2) that this condition is fulfilled only for $\lambda = 0$. Thus the function e_0, representing the straight rule of estimation, is the only unbiased estimate-function in our λ-system. We shall later come back to this point.

In our present discussions, an estimate is based on a sample of size s in a state-description k which describes a universe with a finite number N of individuals. It is an estimate of the rf of M in the class K of those $N - s$ individuals which do not belong to the sample. We suppose that N is very large in relation to s; therefore, we take the rf of M in K as approximately equal to that in the whole universe. Let us digress for a moment to consider the estimate for an *infinite universe*; that is, the estimate $e(\text{rf},M,K_\infty,e_M)$ of the limit of the rf of M in an infinite sequence K_∞. It seems natural to take as this estimate the limit of $e(\text{rf},M,K_m,e_M)$ for $m \to \infty$, where K_m is a class of m individuals not mentioned in e_M. Now, for any estimate-function e_λ of the λ-system, $e_\lambda(\text{rf},M,K_m,e_M)$ always has, according to (11-7), the value $(s_M + (w/\kappa)\lambda)/(s+\lambda)$, independently of m; hence $e_\lambda(\text{rf},M,K_\infty,e_M)$ has the same value. Let us assume that there is a limit of the rf of M in the infinite sequence K_∞; let this limit

7. Kendall, *op. cit.*, II, 5 f.

be *r*. Let us consider a sequence of samples with increasing size *s*; the sample with the size *s* contains the first *s* individuals of the infinite sequence. Reichenbach[8] shows that his rule of induction (see above, §14), which is essentially the same as the straight rule of estimation, is *self-correcting* in the following sense. If we choose a positive δ, however small, then there is a finite *n* such that, for every sample in the sequence of samples with $s > n$, the estimate based on this sample does not deviate from *r* by more than δ. This follows immediately from the definition of *r* as the limit of s_M/s for $s \to \infty$. The values of any e_λ with positive finite a converge for $s \to \infty$ toward those of e_0, as is seen from the value of e_λ stated above. Therefore, these functions are likewise self-correcting. On the other hand, e_∞ is not self-correcting, as is seen immediately from (13-5). Closely related to the concept of self-correction is that of consistence[9] (defined by the "convergence in probability" of the estimates toward the actual value with $s \to \infty$). This is likewise a property of every function e_λ except e_∞.

Now we return to the discussion of the finite state-description *k*. We find the following results concerning the error v of the estimate $e_\lambda(\text{rf},M,K,e_M)$, and the mean $\bar{\text{v}}$ and the variance $\sigma^2(\text{v})$ of this error in the distribution of samples of size *s* in *k* from (20-1), (20-2), and (20-3) with the help of (19-18), (19-19), and (19-20):

(20-4) $$\text{v} = \frac{s_M + (w/\kappa)\lambda}{s+\lambda} - r,$$

(20-5) $$\bar{\text{v}} = \frac{(w/\kappa) - r}{s+\lambda},$$

(20-6) $$\sigma^2(\text{v})\ \frac{sr(1-r)}{(s+\lambda)^2}.$$

For the mean square error 1v^2 in the sampling distribution in *k* we obtain from (20-6) and (20-5) according to (19-22):

(20-7) $$1\text{v}^2(e_\lambda,M,k,s) = \frac{sr(1-r) + (w/\kappa - r)^2\lambda^2}{(s+\lambda)^2},$$
$$= \frac{(\lambda^2 - s)r^2 + (s - 2\lambda^2 w/\kappa)r + \lambda^2(w/\kappa)^2}{(s+\lambda)^2}.$$

The second form shows that 1v^2 as a function of *r* (for fixed *w*, *κ*, λ, and *s*) is, in general, represented by a parabola, which is convex upward if $\lambda^2 < s$

8. *The theory of probability* (Berkeley, 1949), pp. 445 ff.

9. This concept was introduced by R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Philos. Transactions of the Royal Society*, Ser. A, Vol. 222 (1922}; see also Cramér, *op. cit.*, p. 351; Kendall, *op. cit.*, II, 3.

and convex downward if $\lambda^2 > s$. If $\lambda^2 = s$, the curve is a straight line; in this case

(20-8) $$1\mathrm{v}^2 = \frac{(1-2w/K)r+(w/K)^2}{(\sqrt{s}+1)^2}$$

This line is, in general, inclined toward the r-axis. However, in the case that $w/K = ½$, and only in this case, the line is parallel to the r-axis and hence $1\mathrm{v}^2$ is independent of r. In this case

(20-9) $$1\mathrm{v}^2 = 1/[4(\sqrt{s}+1)^2].$$

The case $\lambda^2 = s$ just discussed is here to be understood in the sense that the numerical value chosen for λ (chosen either generally or for a particular language-system) and the size s of the samples under investigation happen to stand in the relation $\lambda^2 = s$. However, the results mentioned hold also for a λ-function defined as follows:

(20-10) $$\lambda(K,s,s_i) = \underset{+}{\sqrt{s}}\,.$$

This λ-function belongs to our λ-system; but it does not belong to the class of those λ-functions which we selected for closer consideration, since it is dependent upon s and hence does not fulfil condition C11 (§ 9). For this function and for a property M with $w/K = ½$, $1\mathrm{v}^2$ has the value stated in (20-9), independent of r.

We consider now, for any λ and any s, the case that M is a property with $w/K = ½$, e.g., a primitive property. (The value of K itself does not matter, because 'K' occurs in these formulas only in the context of 'w/K'.) In this case, the formulas (20-1) to (20-7) yield the following results:

(20-11) $$\mathrm{e}_\lambda(\mathrm{rf},M,K,e_M) = \frac{s_M+\lambda/2}{s+\lambda}.$$

(20-12) $$\overline{\mathrm{e}}_\lambda = \frac{sr+\lambda/2}{s+\lambda}.$$

(20-13) $$\sigma^2(\mathrm{e}_\lambda) = \sigma^2(\mathrm{v}) = \frac{sr(1-r)}{(s+\lambda)^2}.$$ as earlier.

(20-14) $$\mathrm{v} = \frac{s_M+\lambda/2}{s+\lambda} - r.$$

(20-15) $$\overline{\mathrm{v}} = \frac{(1/2-r)\lambda}{s+\lambda}.$$

(20-16) $$1\mathrm{v}^2(\mathrm{e}_\lambda) = \frac{sr(1-r)+(1/2-r)^2\lambda^2}{(s+\lambda)^2},$$
$$= \frac{(\lambda^2-s)r^2-(\lambda^2-s)r+\lambda^2/4}{(s+\lambda)^2}.$$

The curve for $1v^2$ is, in general, a parabola, as above. However, if $\lambda^2 = s$, we have the case discussed earlier, in which $1v^2$ has the value (20-9) independent of r.

§ 21. The Mean Square Error with Respect to All *Q*-Properties

We have determined in (20-7) the mean square error $1v^2$ of the estimates supplied by the function e_λ for the rf of the particular property M in the state-description k, based on all samples of the size s. This value, however, is obviously not suitable as a measure for the success of e_λ in k in general. For this purpose it would be necessary to take into consideration also the success of e_λ for other properties than just M. This we shall do by determining the mean square error of the estimates for all Q-properties together, which we denote by '$1v_Q^2$'. The Q-properties are fundamental; any other factual property is a disjunction of Q-properties, and hence its rf is the sum of the rf's of the corresponding Q's. Therefore, $1v_Q^2$ represents, in a sense; a measure for all properties. e_Q describes a sample of s individuals of which s_i ($i = 1$ to κ) have the property Q_i. According to (11-5) and (G-12), $e_\lambda(\text{rf}, Q_i, K, e_Q) = (s_i + \lambda/\kappa)/(s + \lambda)$. N_i is the cardinal number of Q_i in the state-description k. $r_i = N_i/N$ is the rf of Q_i in k.

From (20-7), for Q_i, hence with $w = 1$:

(21-1)
$$1v^2(e_\lambda, Q_i, k, s) = \frac{sr_i(1-r_i)+(1/\kappa - r_i)^2\lambda^2}{(s+\lambda)^2},$$
$$= \frac{(\lambda^2 - s)r_i^2 + (s - 2\lambda^2/\kappa)r_i + \lambda^2/\kappa^2}{(s+\lambda)^2}.$$

$1v_Q^2(e_\lambda, k, s)$ is the mean square error of κS estimates $e_\lambda(\text{rf}, Q_i, K, e_Q)$, one estimate for each of the S samples of size s in k and each of the κ Q's. Therefore, it is the mean of the κ values of $1v^2$ for Q_i, with $i = 1$ to κ:

(21-2)
$$1v_Q^2(e_\lambda, k, s) = \frac{1}{k}\sum 1v^2(e_\lambda, Q_i, k, s).$$

('$\sum$' denotes here and in the subsequent sections always summation over i from 1 to κ, unless indicated otherwise.) Hence with (21-1), noting that $\sum r_i = 1$:

(21-3)
$$1v_Q^2(e_\lambda, k, s) = \frac{s - \lambda^2/\kappa + (\lambda^2 - s)\sum r_i^2}{\kappa(s+\lambda)^2}$$

We shall regard the success of e_λ, in k (for samples of size s) as indicated by the smallness of $1v_Q^2(e_\lambda, k, s)$.

In the preceding discussion reference was made to the numbers N and κ characterizing the given language-system, and to the Q-numbers N_i ($i = 1$

to K) of the given state-description k. Nothing else concerning the given state-description was used; thus the result could just as well be formulated for a structure-description. Now we see from (21-3) that for $1v_Q^2$ the values N and N_i themselves are not relevant (except for the assumption that N must be very large in relation to s), but only $N_i/N = r_i$. And the values $r_1, r_2, \ldots, r_K$ enter the result only through $\sum r_i^2$; thus this sum is the only magnitude concerning the state-description k which is relevant for $1v_Q^2$.

This sum is a very important characteristic of the state-description. It has its maximum value 1, if one r_i is 1 and all others 0, in other words, if one N_i is N and all others 0. A state-description of this kind was called *homogeneous* (§ 14). It may also be regarded as having the maximum degree of order or uniformity or regularity. The concept of the uniformity of the universe has often been used by philosophers but has never been clearly defined. It seems to me that it should be explicated as a qualitative concept, a *degree of order*. We shall not try here to give a definition; for our present discussion it will suffice to make a few remarks in comparative form.[10] The uniformity of the world was often discussed by earlier authors in connection with the problem of the validity of inductive reasoning. It was understood in the sense of the existence of regularities in the world, expressible by universal laws. Now it can easily be shown that, the mere Q's are empty, the more universal laws hold,[11] and thus the higher is the degree of order. In a homogeneous state-description all Q's except ore are empty; thus the degree of order reaches its maximum.

On the other hand, consider a state-description in which all N_i-values are equal. (Strictly speaking, in order to make this possible, we must assume that N is divisible by K; but, since we presuppose N to be very large, the following remarks hold with sufficient approximation also if N is not divisible by K and the N_i-values are not exactly equal but nearly so.) Then, for every i, $N_i = N/K$; hence $r_i = 1/K$, and $r_i^2 = 1/K^2$. In this case $\sum r_i^2$ has its minimum value $1/K$. Here the degree of order reaches its minimum.

$\sum r_i^2$ varies from $1/K$ to 1. We shall find it convenient for the formulation of several of our results to make use of the differences between a given value of $\sum r_i^2$ and those two extreme values; therefore we define:

(21-4)

$$D_1 = 1 - \sum r_i^2 \, .$$
$$D_2 = \sum r_i^2 - 1/K$$

10. The concept of degree of order will be discussed in detail in a forthcoming article, and a tentative quantitative explication for it will be given.

11. [I], § 37.

Then we can reformulate (21-3) as follows:

(21-5) $$1v^2_Q(e_\lambda,k,s) = \frac{D_1 s + D_2\lambda^2}{\kappa(s+\lambda)^2}.$$

For the two extreme cases with respect to $\sum r_i^2$, the following holds:

(21-6) Let $\sum r_i^2 = 1$; maximum degree of order.

(a) $D_1 = 0$,

(b) $D_2 = (K - 1)/K$,

(c) $$1v^2_Q(e_\lambda,k,s) = \frac{(\kappa-1)\lambda^2}{\kappa^2(s+\lambda)^2}.$$

(21-7) Let $\sum r_i^2 = 1/K$; minimum degree of order.

(a) $D_1 = (K - 1)/K$

(b) $D_2 = 0$

(c) $$1v^2_Q(e_\lambda,k,s) = \frac{(\kappa-1)s}{\kappa^2(s+\lambda)^2}.$$

We consider, now, a system of the simplest kind, with *M* as *the only primitive property*. Let *r* be the rf of *M* in *k*, as before. Here we have $K = 2$; the two *Q*'s are *M* and non-*M*; hence $r_1 = r$ and $r_2 = 1 - r$. The previous results (20-11) to (20-16) hold here. Further we have:

(21-8) $$\sum r_i^2 = r^2 + (1-r)^2 = 1 - 2r(1-r).$$

Hence with (21-4)

(21-9) $$D_1 = 2_r(1-r),$$
(21-10) $$D_2 = 1/2 - D_1 = 2(1/2 - r)^2.$$

Hence from (21-5) with $K = 2$:

(21-11) $$1v^2_Q(e_\lambda,k,s) = \frac{sr(1-r)+(1/2-r)^2\lambda^2}{(s+\lambda)^2}, = \frac{(\lambda^2-s)r^2-(\lambda^2-s)r+\lambda^2/4}{(s+\lambda)^2}.$$

This is the same as the value of $1v^2$ for *M* stated in (20-16), because here $1v^2$ has the same value for each of the two *Q*'s, viz., *M* and non-*M*.

Let us look at some numerical examples for (21-11) with samples of size $s = 10$. The subsequent table shows, for a few selected methods characterized by their λ, the mean square error $1v^2_Q$ as a function of *r*. In the table, *r* runs only from 0 to 0.5; $1v^2_Q$ has the same value for $1 - r$ as for *r*, as (21-11) shows. On each line, the smallest value of $1v^2_Q$ is indicated by boldface type. The results in the table suggest that, the nearer *r* comes to

0.5, the higher is the λ for which $1v_Q^2$ has its minimum. We shall soon show that this is indeed the case.

(21-12) *The Mean Square Error* $1v_Q^2$ *as a Function of* λ *and r*

(For $\kappa = 2$, $s = 10$)

r	λ =						
	0	1	2	4	8	16	∞
	0.0	0.0	0.0	0.0	0.0	0.0	
0	000	021	069	204	493	946	0.2500
0.1	090	088	107	177	343	619	0.1600
0.2	160	140	136	155	227	364	0.0900
0.3	210	177	157	140	144	182	0.0400
0.4	240	199	170	131	094	073	0.0100
0.5	250	207	174	128	077	037	0.0000

Let us again consider the case that $\lambda^2 = s$. This case occurs if either the numerical values of λ and *s* happen to stand in this relation or the λ-function, defined as $\sqrt{s}$ (20-10) has been adopted. Then, from (21-3):

(21-13) $$1v_Q^2(e_\lambda,k,s) = \frac{\kappa - 1}{\kappa^2(\sqrt{s}+1)^2}.$$

This is independent of the r_i-values; hence it holds in every state-description. Thus, with respect to samples of the size *s*, the estimate-function e_λ with $\lambda = \sqrt{s}$ has equal success in all state-descriptions.

§ 22. The Optimum Inductive Method for a Given State-Description

The λ-system orders the inductive methods in a continuum. The value of the parameter λ for a given inductive method determines the position of the method within the continuum. We can now investigate the continuous changes in an inductive method brought about by a continuous change in λ. This makes it possible to find that inductive method which shows some characteristic to the highest degree with respect to given conditions. Thus we shall now study the question how to determine that inductive method which is most successful in a given possible world represented by a state-description, in other words, that method which has the smallest mean square error.

Let a language-system and a state-description *k* be given; then κ and $\sum r_i^2$ are fixed. We choose a sample size *s* and consider all the samples of this size in *k*. We have seen that, for any given λ, the mean square error $1v_Q^2(e_\lambda,k,s)$is determined by (21-3) or (21-5). We shall now study the change of $1v_Q^2(e_\lambda,k,s)$ when λ varies. We assume that the given value of $\sum r_i^2$ is not one of the two extremes ($1/\kappa < \sum r_i^2 < 1$, hence $D_1 > 0$, $D_2 > 0$), leaving the two extreme values aside for the moment. We find the following results (by partial differentiation of the function stated in (21-5) with respect to λ). With increasing λ, $1v_Q^2$ first decreases, takes on a minimum value, and then increases again (compare table (21-12)). Let λ^Δ be that value of a for which $1v_Q^2$ has its minimum. We find:

(22-1) $$\lambda^\Delta = \frac{D_1}{D_2} = \frac{1-\sum r_i^2}{\sum r_i^2 - 1/\kappa}.$$

The minimum value of $1v_Q^2$, which it reaches at λ^Δ, is:

(22-2) $$\text{Min}_\lambda\,(1v_Q^2) = \frac{D_1 s + D_2 \lambda^{\Delta 2}}{\kappa(s+\lambda^\Delta)^2} = \frac{D_1 D_2}{\kappa(D_1 + sD_2)}.$$

Now we consider the two extreme values for $\sum r_i^2$. The following results (22-3) and (22-4) are obtained from (21-6) and (21-7), respectively.

(22-3) Let $\sum r_i^2 = 1$; hence $D_1 = 0$.

(*a*) $\lambda^\Delta = 0$.

(*b*) $1v_Q^2(e_0) = 0$

In this case, $1v_Q^2$ has its minimum value 0 for $\lambda = 0$; from here on it increases always with increasing λ.

(22-4) Let $\sum r_i^2 = 1/\kappa$; hence $D_2 = 0$.

(*a*) $1v_Q^2$ decreases always with increasing λ; its limit for $\lambda \to \infty$. is smaller than any of its values for finite λ; hence: $\lambda^\Delta = \infty$.

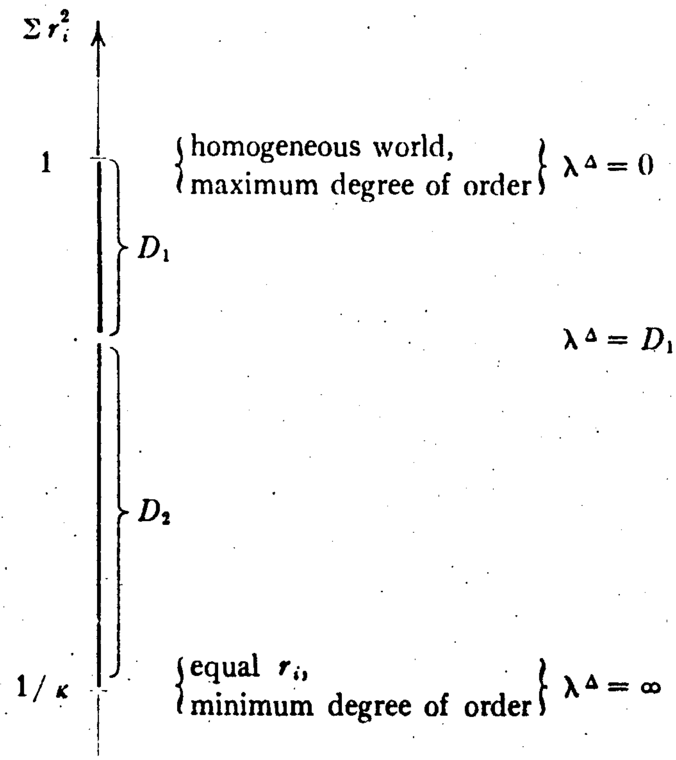
(*b*) $1v_Q^2(e_\infty) = 0$ (i.e., $\lim_{\lambda\to\infty} 1v_Q^2(e_\lambda) = 0$.

The formulas (22-1) and (22-2) were first stated only for nonextreme values of $\sum r_i^2$. However, the value of λ^Δ stated in (22-3) (a) is the same as that resulting from (22-1) for $\sum r_i^2 = 1$. And the value $\lambda^\Delta = \infty$ stated in (22-4) (a) is in accordance with the limit of the function stated in (22-1) for $\sum r_i^2 \to 1/K$. Thus (22-1) holds for all values of $\sum r_i^2$, including the two extreme ones. The same is true for (22-2), as is seen from (22-3)(*b*) and (22-4)(*b*).

We call λ^Δ the optimum λ-value, and the inductive method characterized by λ^Δ ° the *optimum method* for the given state-description, because this method has a smaller mean square error than any other one. The e-function characterized by λ^Δ is denoted by 'e^Δ' and called the optimum e-function.

The formula (22-1) for λ^Δ is surprisingly simple. (This gives additional support to the view, originally based on other reasons not mentioned in this monograph, that $1v_Q^2$ is a suitable measure for the success of an inductive method.) In particular, λ^Δ is independent of *s*, although $1v_Q^2$ was still dependent upon *s*. Thus λ^Δ, as determined by (22-1) for a given state-description *k*, characterizes one method which is most successful in application to any sample size *s* in *k* (provided that *s* is small in relation to *N*), and which hence may be regarded as *the* optimum method for *k*.

The situation with respect to optimum methods is as follows; it is illustrated in the accompanying diagram. We consider various possible



λ^Δ as a function of $\sum r_i^2$.

worlds, represented by state-descriptions. We begin with a world of maximum degree of order; it is homogeneous, i.e., all individuals are alike, hence $\sum r_i^2 = 1$. We find here $\lambda^\Delta = 0$; this means that the optimum method is the straight rule. This is quite plausible in view of the fact that the method of the straight rule implicitly regards they homogeneous state-descriptions as the only possible ones (see (14-12) and (14-1â)). When we go to state-descriptions with smaller degrees of order, $\sum r_i^2$ moves from 1 down toward $1/K$, and λ^Δ increases more and more. As long as $\sum r_i^2$ has an intermediate value, λ^Δ is always positive and finite; it is just the quotient of the distances of the value of $\sum r_i^2$ from the two extremes. Finally, for the state-description with minimum degree of order, i.e., the one in which all Q's have the same rf, and hence $\sum r_i^2$ has its minimum value $1/K$, we have $\lambda^\Delta = \infty$.This means that the optimum method is represented by e_∞. This is plausible because the value of the rf of every Q in this state-description, viz., $1/K$, is the value which e_∞ assigns to the estimate of the rf of any Q with respect to any evidence (see (13-3)).

The result that, for any given state-description k, there is a unique inductive-method which is the most successful method in k and that this optimum method can easily be specified by its λ-parameter according to (22-1) is of great theoretical interest. However, it is not of immediate. practical usefulness. The practical knowledge situation for any human being at any time is such that he knows only a relatively small part of the universe, never the whole; it is just this fact that makes the use of inductive methods necessary. Therefore, he is never in a position where lie could directly apply the rule which tells him how to find the optimum inductive method for his universe from a certain magnitude concerning the universe as a whole (vie., $\sum r_i^2$). We shall, however, see later (§ 2-1) that it is still possible to make some weaker statements about the optimum method for the unknown universe.

Theoretically interesting is likewise the inverse result: given any inductive method characterized by a value of λ, there is always a state-description for which the given method is the optimum method.

Strictly speaking, for a given system with a fixed N, there is only a finite number (vie., $N + 1$) of possible values of N_i and of r_i. However, since we suppose N to be very large, the difference between consecutive values of r_i is very small (viz., $1/N$). Therefore, it is possible with a good approximation to proceed as if r_i could vary continuously, running through all real numbers from 0 to l, and likewise $\sum r_i^2$, running from $1/K$ to 1. Thus we consider λ^Δ as a continuous function of $\sum r_i^2$, given by (22-1), while, strictly speaking, it has only a finite number of possible values. Now we consider the inverse function; we obtain from (22-1):

(22-5)
$$\sum r_i^2 = \frac{\lambda^\Delta / \kappa + 1}{\lambda^\Delta + 1}$$

Thus, if we choose any nonnegative real number as λ^Δ (strictly speaking, any of those values of λ^Δ which are possible for the given N according to (22-1)), then (22-5) determines uniquely a corresponding value of $\sum r_i^2$ such that, for any state-description with this $\sum r_i^2$ the chosen λ^Δ is the optimum λ. We see from (22-5) that, for $\lambda^\Delta = 0$, $\sum r_i^2 = 1$; thus, as we already know, the method of the straight rule is the optimum method for any homogeneous state-description. The limit of the function stated in (22-5) for $\lambda^{\Delta} \rightarrow \infty$ is $1/K$; hence the method with $\lambda = \infty$ is, as we know, the optimum method for the state-description with $\sum r_i^2 = 1/K$, that is, the one with equal values of r_i. The situation will be illustrated by the subsequent numerical example (22-8).

The foregoing general results will now be applied to a system with only *one primitive property M*, hence $K = 2$. From (22-1) with (21-9) and (21-10):

(22-6) For $K = 2$,
$$\lambda^{\Delta} = \frac{D_1}{D_2} = \frac{2D_1}{1-2D_1} = \frac{4r(1-r)}{1-4r(1-r)} = \frac{r(1-r)}{(1/2-r)^2}.$$

From (22-2) we find the value of $1v_Q^2$ for λ^{Δ}:

(22-7) For $K = 2$,
$$\mathrm{Min}_{\lambda}(1v_Q^2) = \frac{D_1(1-2D_1)}{4D_1 + 2s(1-2D_1)} = \frac{r(1-r)[1-4r(1-r)]}{4r(1-r)+s[1-4r(1-r)]}.$$

The following numerical results are calculated by (22-6). λ^{Δ} is given for certain values of r up to 0.5; the curve of λ^{Δ} is symmetrical around $r = 0.5$, that is, λ^{Δ} has the same value for r and $1 - r$. λ^{Δ} is the optimum a for a state-description in which the rf of the primitive property is r.

(22-8)

r	λ^{Δ}	r	λ^{Δ}
0	0	0.3	5.250
0.05	0.235	0.35	10.111
0.1	0.562	0.4	24.000
0.15	1.041	0.45	99.000
0.2	1.778	0.5	∞
0.25	3.000		

From (22-5) we have here:

(22-9) For $K = 2$,
$$\sum r_i^2 = \frac{\lambda^{\Delta}+2}{2\lambda^{\Delta}+2}$$

Hence with (21-8):

(22-10)
$$r = \frac{1}{2}\left[1 \pm -\frac{1}{\sqrt{\lambda^{\Delta}+1}}\right]$$

If we choose arbitrarily a value λ^{Δ}, then the method characterized by this λ is the optimum method for those state-descriptions in which the rf of the primitive property *M* has one of the two values stated in (22-10) (and hence the rf of non-*M* has the other value). (22-10) yields the following numerical results for some: values of λ^{Δ}; *r* is here the smaller one of the two values in (22-10).

(22-11)

Method	λ^{Δ}	r
Straight rule	0	0
	.	.
	.	.
	¼	0.053
	½	0.092
	1	0.146
c* and mod. Laplace	2	0.211
	4	0.276
	8	0.333
	16	0.379
	32	0.413
	64	0.438
	128	0.456
	.	.
	.	.
c†	∞	0.500

§ 23. Are Unbiased Estimate-Functions Preferable?

Many contemporary statisticians seem to regard unbiased estimates as preferable, those for which the mean $\overline{e(x,e)}$ in the sampling distribution is always equal to the actual value *x* of the magnitude in question. As far as I am aware, no rational reasons for this preference have been offered. It seems a nice situation if the grouping around *x* of the various estimates of *x* in the totality of samples shows this kind of symmetry. But it is clear that the decisive criterion for the success of an estimate-function cannot be this condition, which does not say anything about the closeness of the estimates to *x*. This closeness must be measured by a function of the errors, e.g., the mean square error or something similar. We shall show in this section that the straight rule, which is the one unbiased estimate-function within our system concerning estimation of rf, is definitely inferior to other functions which are not unbiased.

In order to obtain a more concrete picture of the situation, let us compare in a *numerical example* the success of two particular methods of estimation with respect to a given state-description *k* in a system with $x = 2$ and *M* as the only primitive property. The first method is that of the straight rule, characterized by $\lambda = 0$ and represented by the estimate-function e_0. The second is the optimum method for *k*, with λ^{Δ}; let its e-function be e^{Δ}. Let the rf of *M* in *k* be $r = 0.3$; we consider samples of size $s = 10$. Thus the data of our ????????:

(23-1) $K = 2$; $w/K = ½$; ?????? $s = 10$.

For these data we obtain, from (21-9) and (21-10):

(23-2) $D_1 = 0.42$; $D_2 = 0.08$;

hence with (22-6)

(23-3) $\lambda^\Delta = 5.25$.

The value of s_M, the number of individuals with M in a sample, runs from 0 to 10. [According to the binomial law (19-13), the proportion of samples with a particular value s_M among the totality of samples of the fixed size 10 for $r = 0.3$ is $\binom{10}{s_M} \times 0.3^{s_M} \times 0.7^{10-s_M}$.] The following results hold for any λ. From (20-11) to (20-16):

(23-4) $e_\lambda(\mathrm{rf},M,K,e_M) = \dfrac{s_M + \lambda/2}{10+\lambda}$,

(23-5) $\overline{e_\lambda} = \dfrac{3+\lambda/2}{10+\lambda}$,

(23-6) $\sigma^2(e_\lambda) = \sigma^2(v) = \dfrac{2.1}{(10+\lambda)^2}$,

(23-7) $v = e_\lambda - 0.3$,

(23-8) $\bar{v} = \dfrac{0.2\lambda^2}{10+\lambda}$.

From (21-11):

(23-9) $1v_Q^2(e_\lambda) = \dfrac{2.1+0.04\lambda^2}{(10+\lambda)^2}$.

The straight rule: $\lambda = 0$. Here $e_0(\mathrm{rf},M,K,e_M) = s_M/10$. These values are listed in the subsequent table (23-13). $\sigma^2(e_0) = \sigma^2(v) = 0.0210$. $v = s_M/10 - 0.3$; the values of v^2 are listed in the table $\overline{e_0} = 0.3 = r$; hence e_0 is unbiased, as we know already. $\bar{v} = 0$. Therefore, according to (19-23), $1v^2$ for M is simply the same as the variance of the estimate; and the same holds for $1v_Q^2$ (see the remark on (21-11); the result can also be obtained from (23-9))
= 0.0210 .

(23-10) $1v_Q^2(e_0) = \sigma^2(e_0) = 0.0210$.

The optimum method: $\lambda^\Delta = 5.25$. $e^\Delta(\mathrm{rf},M,K,e_M) = (s_M + 2.625)/15.25$. These values are listed in the table; also those of the square error $v^2 == (e^\Delta - 0.3)^2$. $\overline{e^\Delta} = 5.625/15.25 = 0.3689$. Hence e^Δ is not unbiased; the difference between $\overline{e^\Delta}$ and r (which is equal to $\bar{v}$, see (19-19)) is 0.0689.

(23-11) $\sigma^2(e^\Delta) = \sigma^2(v) = 0.00901$.

According to (19-23), we must add to this $\overline{v^2}$ = 0.00475 in order to obtain $1v^2$ for *M*, which is here equal to $1v_Q^2$; hence:

(23-12) $$1v_Q^2(e^\Delta) = 0.013\ 76\ .$$

(23-13) *Estimates and Square Errors for* $\lambda = 0$ and $\lambda^\Delta = 5.25$ *as Functions of* s_M (for $K = 2$, $r = 0.3$, $s - 10$)

	$\lambda = 0$		$\lambda^\Delta = 5.25$	
s_M	e_0	v^2	e^Δ	v^2
0	0	0.09	0.1722	0.01636
1	0.1	0.04	0.2378	0.00387
2	0.2	0.01	0.3025	0.00001
3	0.3	0	0.3688	0.00475
4	0.4	0.01	0.4345	0.01806
5	0.5	0.04	0.5000	0.04006
6	0.6	0.09	0.5656	0.07051
7	0.7	0.16	0.6312	0.10963
8	0.8	0.25	0.6968	0.15737
9	0.9	0.36	0.7623	0.2136
10	1.0	0.49	0.8280	0.2787
	$\sigma^2(e_0) = 0.0210$		$\sigma^2(e^\Delta) = 0.0091$	
	$1v_Q^2(e_0) = 0.0210$		$1v_Q^2(e^\Delta) = 0.0138$	

Comparison. We see that not only the variance of the estimate is smaller for e^Δ than for e_0, but also the mean square error. In other words, the values of e^Δ, in comparison with those of e_0, show a closer concentration not only around their mean but also around the actual value r = 0.3. And this is the case in spite of the fact that e_0 is unbiased while e^Δ in not. The same comparative result holds for any other value of *r*, except 0 and 1, in other words, for any other state-description, except the two homogeneous ones. These results seem to me to show that the widespread preference for the method of the straight rule e_0, in the form of either the principle of maximum likelihood or the principle of unbiased estimation, is not justified.

§ 24. The Problem of the Success in the Actual Universe.

The discussions in the preceding sections referred always to a given state-description *k*. The successes of various methods *within k* were compared, and the optimum method for *k*, characterized by λ^Δ, was determined. As mentioned earlier, the practical situation of any observer *X* is, however, such that he knows no description of the whole universe but only a sample constituting a small part of the universe. Therefore, he is not in a position to determine λ^Δ for his actual universe. Nevertheless, he can find out certain things about this unknown λ^Δ; in particular, he can determine a lower bound for λ^Δ. This makes it possible to specify a method of which it is certain that it is more successful in the unknown total universe than the method of the straight rule. All this presupposes merely that *X* knows that the universe is not homogeneous, i.e., that it contains at least two individuals not completely alike. This will now be shown.

Let a language system be given with given primitive properties and, based upon them, κ *Q*-properties, and with a number *N* of individuals. e_Q describes a sample of *s* individuals, of which s_i ($i = 1$ to κ) are Q_i. We assume that the sample is nonhomogeneous, i.e., that at least two distinct *Q*'s occur in it. Then, obviously, the universe cannot possibly be homogeneous. Let k_T be the unknown true state-description, the one which ascribes to each individual that *Q* which it actually has. Let the *Q*-numbers in k_T be N_{iT}, and let $r_{iT}, = N_{iT}/N$. Since *X* knows only the sample, he cannot state k_T; nevertheless, he can, on the basis of his evidence e_Q, ascertain certain things about the universe and its description k_T. Some things can be established inductively and hence without complete certainty. But there are other things about k_T which *X* can state with deductive certainty. The following discussion concerns these latter things. For example, it follows from e_Q that, for any *i*, $N_{iT} \geq s_i$, and hence $r_{iT} \geq s_i/N$. We put $\sum_T = \sum r_i^2{}_T$. We choose Q_m arbitrarily as one of those *Q*'s which have the highest s_i-value occurring in e_Q. We construct the state-description *k*′ in such away that it contains e_Q and ascribes to all individuals not occurring in e_Q the property Q_m. For every *i*, let N'_i be the cardinal number of Q_i in *k*′, and $r_i' = N'_i/N$. Then, for any $i \neq$ m, $N'_i = s_i$; and $N'_m = N - s + s_m$. We put $\sum' = \sum r_i^2$. Since e_Q is nonhomogeneous, *k*′ is likewise; hence:

(24-1) $$\sum' < 1 .$$

k_T cannot possibly be nearer to homogeneity than *k*′; in exact terms:

(24-2) $$\sum_T \leq \sum' < 1 .$$

The value $\sum'$ is thus a known upper bound for the unknown $\sum_T$; it is the maximum of $\sum r_i^2$ for all those state-descriptions which are compatible with e_Q. Let λ_T^Δ be the optimum λ for k_T, and λ′ the optimum λ for *k*′. λ′ is known, but λ_T^Δ is not. According to (22-1):

(24-3) $$\lambda_T^\Delta = \frac{1-\sum_T}{\sum_T - 1/\kappa}$$

(24-4) $$\lambda' = \frac{1-\sum'}{\sum' - 1/\kappa}$$

Hence with (24-1):

(24-5) $\lambda' > 0$.

With (24-2)

(24-6) $\lambda_T^\Delta \geq \lambda' > 0$..

Thus the known value λ' represents a lower bound for the unknown λ_T^Δ. It is the minimum of the optimum λ-values for all those state-descriptions which are compatible with e_Q.

Let e^Δ and e′ be the estimate-functions characterized by λ_T^Δ and λ' , respectively. e′ is known, e^Δ is not. In any state-description *k* in which $\sum r_i^2 < 1$, $1v_Q^2(e_\lambda,k,s)$ always decreases when λ varies from 0 to λ^Δ. [For $\sum r_i^2 < 1/K$, this follows from the explanations preceding (22-1); for $\sum r_i^2 = 1/K$, from (22-4)(a).) k_T fulfils the condition (from 124-2)); hence:

(24-7) $1v_Q^2(e^\Delta,k_T,s) \leq 1v_Q^2(e',k_T,s) < 1v_Q^2(e_0,k_T,s)$.

Thus, although k_T itself is unknown, it can be stated with certainty that in k_T the known function e′ is more successful in the sense of having a lesser mean square error for the estimation of all,*Q*'s with respect to all samples of the size *s* than e_0, the straight rule.

We shall now illustrate the foregoing general discussion by a *numerical example*. Let the universe be a bag with $N = 1{,}000$ balls. Each ball is either black (*B*) or white (non-*B*, *W*), but otherwise they are alike; hence we have $K = 2$. Let e_M describe a sample of $s = 20$ balls, among them $s_1 = 7$ *B*, $s_2 == 13$ *W*. The true description k_T of all thousand balls is unknown; but we know that $N_{1T}. \geq 7$, hence $r_{1T} \geq 0.007$, and $N_{2T} \geq 13$, hence $r_{2T} \geq 0.013$. k' contains e_M and, in addition, describes the 980 unobserved balls as *W*. Thus $N'_1 = 7$, $N'_2 = 993$; hence $r'_1 = 0.007$, $r'_2 = 0.993$. Then we find the following results by (21-8), (21-9), (21-10), (22-1): 0.986098; $D'_1 = 0.013902$; $D'_2 = 0.486098$; $\lambda' = 0.02860$. Thus we can specify an estimate-function e′, viz., that characterized by $\lambda' = 0.0286$, which is more successful in the unknown universe of the thousand balls in the sense of having a smaller $1v_Q^2$ than the function e_0 of the straight rule.

If we were to talk in terms of probability, either in an intuitive, presystematic way or on the basis of any adequate method of confirmation, we might say that in the above example it is very improbable that the actual universe is similar to that described in k' with the rf of *B* being $r'_1 = 0.007$. Since 35 per cent of the balls in the observed sample are *B*, there is a very great probability that r_{1T}, the actual rf of *B*, is, say, between $\frac{1}{6}$ and $\frac{5}{6}$ and hence that $\lambda_T^\Delta > 1$, according to (22-8). If this is the case, there is an e-function, namely, e^{Δ} characterized by $\lambda_T^{\Delta} > 1$, which is considerably more successful in the actual universe than e_0. Since $\lambda' = 0.0286$ is near to 0, the difference between the results of the functions e′ and e_0 is only small. The reason why we referred to λ′ rather than to any $\lambda > 1$ is the following. The statement that $\lambda_T^{\Delta} > 1$ and hence that there is a method which is very much more successful than to is not certain, but only highly probable. It is possible, though very improbable, that no method with $\lambda > 1$ is more successful than e_0, e.g., if $r_{1T} = 0.007$. On the other hand, the statement that e′ is more successful than e_0 in the actual universe does not involve any question of probability but holds with deductive certainty.

[In the terminology of Wald's theory of decision functions (see the Appendix), the result can be formulated as follows. If any nonhomogeneous sample has been observed and $1v_Q^2$ is taken as the risk function, e′ has a lower risk than e_0 in every distribution (i.e., state-description) compatible with the sample. In other words, with respect to this class of distributions, e′ is uniformly better than e_0, and therefore e_0 is not an admissible estimate-function.)

It is important to keep clearly in mind the meaning of $1v_Q^2$ in order to interpret correctly the result (24-7). $1v_Q^2(e', k_T, s)$ is the mean square error of the estimates of the rf's of all *Q*'s, supplied by the function *e′* and based on *all samples* of the fixed size *s* in k_T. If *X*, after observing the sample described in e_Q, goes on observing more and more individuals without ever forgetting any result of an observation once made, then the samples with which he will be concerned in the future form a particular sequence of samples of increasing size, each containing the preceding one and hence all containing the one described in e_Q. The result (24-7) refers, " not to this sequence of the samples of *X*, but to all samples of size *s*. Therefore, *X* cannot infer from this result that he will be more successful in his future estimations if he uses e′ than if he uses e_0. Whether or not this is the case depends upon which particular sequence of individuals will happen to come his way. What *X* learns from the result is something which concerns, not his own course of life in particular, but rather the universe as a whole and hence, so to speak, the average observer.

The following formulation of our result does not refer to an observed sample and hence precludes any possibility of the misinterpretation just discussed:

(24-8) Let a language-system be given, and hence fixed values of *N* and *K*. Let the function *e′* be characterized by $\lambda' = 2K(N-1)/[(K-1)N^2 - 2K(N-1)]$ (which is > 0). Then, for any nonhomogeneous state-description k in the given system and any s (small in relation to N), $1v_Q^2(e',k_T,s) < 1v_Q^2(e_0,k_T,s)$.

[The specified value of λ′ is found as follows: The maximum value ∑′ of $\sum r_i^2$ for nonhomogeneous state-descriptions holds for those cases in which one Q-number is $N - 1$, another one 1, and all others are 0. Hence $\sum' = (N^2 - 2N + 2)/N^2$. The value of λ′ is then determined by, (24-4).]

It follows from (24-8) that, *if* the true state-description k_T is nonhomogeneous, then e' has a smaller $1v_Q^2$ in k_T than e_0. Note that the condition says here merely that k_T is nonhomogeneous, in other words, that at least two distinct Q's occur in k_T; information to the effect that two specified Q's occur (as would be given by the description of a nonhomogeneous sample, e.g., by e_Q in the earlier discussion) is not necessary for the conclusion that e' is more successful on the whole than e_0.

To sum up, we have found that in any universe which contains at least two unlike individuals there is a specifiable estimate-function e' with $\lambda' > 0$ such that, with deductive certainty, the mean square error of λ′ for all Q's through the whole universe is less than that of the straight rule $e_0$with $\lambda = 0$. e_0 is unbiased, λ′ is not. It seems to me that this result shows a very serious disadvantage of the principle of preferring unbiased estimate-functions and of the straight rule.

APPENDIX

§25. Wald's Theory of Decision Functions and the Minimax Principle

One of the most interesting recent developments in the field of mathematical statistics is the general theory of statistical decision functions constructed by A. Wald.[1] A decision function δ of simplest form for a given problem situation is a function or general rule which assigns a decision to every observational result possible in the situation. The observational result or evidence may, for example, refer to a sample taken from a population. The decision may concern, for example, the acceptance or rejection of a hypothesis or an estimate with respect to a distribution in :the population. The theory has been developed in a very general form going far beyond the simple questions just mentioned. Wald and his collaborators have found many results which are not only interesting from a theoretical point of view but also fruitful in practical applications, e.g., in the so-called sequential analysis for the purpose of quality control in industrial mass production.

We shall later examine the consequences of Wald's theory for one of the problems discussed in this monograph, viz., the estimation of rf. For this purpose we shall now indicate the basic concepts of the theory in their simplest forms and specialized for the problem mentioned. We illustrate the concepts with the help of the following example. Suppose the observer *X* takes a random sample of $s = 50$ units from a lot *K* of $N = 1{,}000$ units produced by a certain manufacturing process. By inspecting the sample, he finds the number s_M of defectives (property *M*) in it. The. decision he has to make is the acceptance of an estimate r' of the rf r of *M* in the lot *K*. Thus the decision function δ is in this case an estimate-function e(rf,*M*,*K*,e_M). *X*'s choice of a decision function is influenced, on the one hand, by the observational result (in our example, the number s_M) and, oil the other hand, by the possible losses which *X* would suffer in the various cases in which his estimate r' differs from the actual value r. *X* is supposed to know the loss which would result in each of the possible cases. The losses are determined by the economic conditions of the situation or by the stipulated rules of a game of chance, or the like. To give an example for our case, let us suppose that there is a rule to the effect that,

1. A. Wald, *Statistical decision functions* (New York, 1950).

if X's estimate is not correct (i.e., $r' \neq r$), he has to pay a fine equal to the square error of his estimate, $v^2 = (r' - r)^2$. Suppose that the value r and a certain function e is given. Then the following sampling distributions (§ 19) for samples of the fixed size s can be calculated: first, the distribution of s_M (this is given by the binomial law (19-13)); then the distribution of the estimate r' determined by s_M, and, finally, the distribution of the loss determined by r and r'. The mean (expectation-value) of the loss in the sampling distribution for a given state of the population and a given decision function, i.e., in our case, for a given r and a given function e, is called the risk $R(r,e)$. Although X does not know the actual value of r, he can determine the risk-function $R(r,e)$ and hence calculate the risk for any value of r and any function e. If, for example, the loss is assumed to be equal to the square error v^2 (this particular assumption is repeatedly mentioned by Wald as an example, e.g., pp. 22, 140,142), then the risk in a given state-description k with a given r is the mean square error $1v_Q^2(e,M,k,s)$. In Wald's theory, the essential characteristic of a decision function consists in its risks in the various possible states of the population; "it seems reasonable to judge the merit of any given decision function ... for purposes of inductive behavior entirely on the basis of the risk function . . . associated with it" (*op. cit*., p. 12). The decision problem consists in the choice of a decision function in a given problem situation.

We have indicated the fundamental ideas of Wald's theory in their simplest form as far as they are relevant for our problem, the estimation of rf. From the point of view of our conception of inductive logic, no objections are raised against these basic parts of the theory. The doubts which we shall express later concern Wald's minimax principle, which lies outside the part of the theory explained so far and is not implied by it.

One of the most important but also most difficult problems in the field of decision functions is the search for a general principle that would tell us which decision function to choose in any given problem situation. For example, in the situation discussed above, such a general principle would determine an estimate-function e for rf; but the same principle would be applicable also to problem situations of entirely different kinds. A general principle that would determine for every problem situation the most suitable decision function would obviously be extremely valuable.

Wald discusses in detail two general principles, which he calls the Bayes principle and the minimax principle. The *Bayes principle* makes use of an a priori probability distribution ξ for the possible states of the population. In our example, the possible states of a lot are characterized by the possible values of *r*, the rf of *M*. Suppose that the manufacturing process has remained unchanged for years and that the observer *X* has frequently made inspections of whole lots of a thousand units each. Then he may feel that he has some knowledge concerning the proportion; in the long run, of those lots which show a certain value of *r*. For example, he might assume the statement that, on the average, one lot among fifty has exactly 3 defectives ($r = 0.003$), and analogous other statements concerning other values of *r*. This proportion (in the example, 1/50) is called the a priori probability for the possible state in question of a lot; the distribution ξ specifies these probabilities for all possible states of a lot. If such a distribution ξ exists and is known to *X*, he can determine for any decision function δ its "average risk" $R^*(\xi,\delta)$ with respect to ξ, i.e., the weighted mean of the risks of δ for the various possible states of the population, with the probabilities of these states as weights. In our example, $R^*(\xi,e)$ is the weighted mean of the risk $R(r,e)$, with the ξ-probabilities of the possible values of *r* as weights. Now, if a distribution ξ exists and is known to *X*, then the Bayes principle advises him to choose that decision function δ (called the "Bayes solution") for which the average risk $R^*(\xi,\delta)$ with respect to ξ takes its minimum value. In the case of a game of chance, *X* may well know the distribution ξ in question (e.g., the probabilities of the various possible hands in a card game). On the other hand, with respect to the results of a manufacturing process, as in the above example, it is obviously not easy for *X* to obtain reliable knowledge concerning the probability distribution. And in most situations which occur in practical life, say, in business, in political, or in social affairs, the task seems hopeless. Thus Wald is certainly right when he says: "In many statistical problems the existence of an a priori [probability] distribution cannot be postulated; and, in those cases where the existence of an a priori distribution can he assumed, it is usually unknown to the experimenter and therefore the Bayes solution cannot be determined" (*ibid*., p. 16).

What Wald here calls "a priori probability", is, of course, probability in the statistical sense; in our terminology, it is probability$_2$, not probability$_1$ or degree of confirmation, This is clearly shown by Wald's remark that these probabilities are often unknown. The values of probability$_1$ cannot be unknown, at least not in the same sense as those of probability$_2$ (compare [I], pp. 174 f.). Like most contemporary statisticians, Wald does not wish to use any concept of degree of confirmation. However, his theory of decision functions could easily be combined with an inductive logic based on a concept of degree of confirmation c. Once the latter concept is introduced, Wald's "Bayes method" can be applied in a modified form, with the values of c taking the place of those of statistical probability. In this modification Wald's mathematical apparatus would remain unchanged in its essential features, only the interpretation of the distribution ξ would be changed and, consequently, the procedure for establishing the values of ξ. While the values of probability$_2$ are found empirically, those of c are determined in a purely logical way. In my view, this modified Bayes method would be preferable both to the Bayes method in Wald's form and to his minimax method. The disadvantages of the Bayes method have clearly been pointed out by Wald himself. We shall now turn to the minimax method.

Since Wald finds the a priori probability distribution unworkable in most cases and, on the other hand, is unwilling to accept a concept of degree of confirmation, he is compelled to look for a general principle not involving either of these concepts. He has proposed a principle of this kind, the so-called *minimax-principle*. After determining the risk-function, *X* can find out, for any decision function δ, the maximum value of its risks in the various possible states of the population. The minimax principle advises *X* to choose that decision function for which this maximum risk is as small as possible. This decision function is called the "minimax solution". In our example, the minimax solution is that estimate-function e for rf, for which $\mathrm{Max}_1 R(r,\mathrm{e})$ takes its minimum value. To make the problem specific, let us assume, as Wald does in this context, that the risk is given by the mean square error $1v^2$.

Hodges and Lehmann[2] have stated an estimate-function applicable to rf and have shown that it is the only minimax solution in the sense just explained. They mention that this result was previously found by H. Rubin, but apparently not published. Wald reports the result in his book (*op. cit*., pp. 142 f.). The function is stated in a general form for the estimation of the parameter of any binomial distribution of a random variable. Now the sampling distribution for the number s_M of individuals with the property *M* in a sample of the fixed size *s* has the binomial form $\binom{s}{s_M} r^{s_M} (1-r)??^{s_M}$ where the parameter *r* is the rf of *M* in the population (see (19-13)). Thus the function stated by Hodges and Lehmann can be used for the estimation of the rf of *M*, based on a sample characterized by *s* and s_M. Since the binomial distribution holds for *any* property *M*, the function in question is an estimate-function for the rf of *any arbitrary property M*. We denote this function by 'e_W' because it is the one to

2. J. L. Hodges, Jr., and E. L. Lehmann, "Minimax point estimation," *Annals of Math. Statistics*, 21 (1950). 182-97.

which Wald's minimax principle leads. Its definition, transcribed in our notation, is as follows:

(25-1) $$\mathrm{e}_W(\mathrm{rf}, M, K, e_M) = \frac{1}{1+\sqrt{s}}\left(\frac{s_M}{\sqrt{s}} + \frac{1}{2}\right).$$

We notice at once that this function, as is customary in mathematical statistics, does not refer to the width w of M but only to the values of s and s_M. However, from the point of view of our conception, w is an essential factor in the situation. Therefore, we may well wonder, even before examining the form of the function, whether it will not lead to undesirable results. This suspicion is soon confirmed.

A. We find easily that the function e_W, does not fulfil the principle of additivity (6-18). Let M_1 and M_2 be L-exclusive properties with the cardinal numbers s_1 and s_2, respectively, in the sample. Then the cardinal number of their disjunction $M_{1,2}$ is $s_1 + s_2$. Therefore:

(25-2) $$\mathrm{e}_W(\mathrm{rf}, M_{1,2}) = \frac{1}{1+\sqrt{s}}\left(\frac{s_1+s_2}{\sqrt{s}} + \frac{1}{2}\right).$$

On the other hand,

(25-3) $$\mathrm{e}_W(\mathrm{rf}, M_1) + \mathrm{e}_W(\mathrm{rf}, M_2) = \frac{1}{1+\sqrt{s}}\left(\frac{s_1+s_2}{\sqrt{s}} + 1\right) > \mathrm{e}_W(\mathrm{rf}, M_{1,2}).$$

Hodges and Lehmann state that the general minimax estimate is not always additive; they add: "This is a definite disadvantage of the minimax principle" (*op. cit.*, p. 189). From the point of view of our theory, any estimate-function for rf which violates the principle of additivity is entirely unacceptable. The indispensability of this principle within our theory follows from the parallelism between e(rf) and c for a singular prediction (see (6-3)). It is obvious from (25-3) and (6-18) that e_W cannot belong to the λ-system.

B. Consider the case that every individual in the sample is either M_1 or M_2; hence $s_1 + s_2 = s$. Here we obtain from (25-3) and (25-2):

(25-4) $$\mathrm{e}_W(\mathrm{rf}, M_1) + \mathrm{e}_W(\mathrm{rf}, M_2) = 1\ ,$$
(25-5) $$\mathrm{e}_W(\mathrm{rf}, M_{1,2}) < 1\ .$$

Let us look at these two results from the point of view of our conception. It is clear that we cannot accept both, since we require additivity. Which one should we then reject? If we examine this question, we find that the, answer depends upon the logical nature of $M_{1,2}$. We have to distinguish two cases. (1) Suppose that $M_{1,2}$ is a factual property (hence its relative width $w/K < 1$). In this case, (25-5) seems not only plausible but even required. The analogous condition is fulfilled for every function e_λ with positive, finite λ and with $\lambda = \infty$ (see (12-6) and (13-5)). It is not fulfilled for the straight rule e_0, (as shown by (14-5)); but this fact was just one of our reasons for rejecting this method. Thus in the present case, we would reject (25-4). (2) Suppose that $M_{1,2}$ is L-universal (hence $w/K = 1$). This is, for instance, the case if M_2 is non-M_1. In this case, (25-4) is required and (25-5) to be rejected. The strangeness of (25-5) in this case is clearly seen when we notice that here, with logical necessity, rf $(M_{1,2}) = 1$. Thus (25-5) violates the following very general requirement for estimate-function which seems to me indispensable:

(25-6) Let the evidence *e* be such that one and only one value *u* of a certain magnitude is compatible with *e*; in other words, the statement that the magnitude has the value *u* is a logical consequence of *e*. Then any estimate of the magnitude on the basis of *e* must be equal to *u*.

This requirement is obviously fulfilled by any estimate-function which is defined as a c-mean estimate (i.e., in the form (6-1)). Hence it is fulfilled, in particular, by all e-functions for rf in the λ-system.

Our discussion of the two cases (1) and (2) shows that the judgment on the acceptability of the results (25-4) and (25 5) must depend upon the logical nature of $M_{1,2}$. The fact that the function e_W ignores this logical nature is thus recognized as its basic defect.

C. The function e_W violates also the following requirement (25-7). (Its nonadditivity mentioned under (A) is a consequence of this fact.)

(25-7) The sum of the estimates of the rf's of several properties which are L-exclusive in pairs, in the same class *K* and on the basis of the same evidence, must not exceed 1.

Suppose that the *s* individuals of the sample have *s* distinct properties $M_1, M_2, \ldots, M_s$, which are L-exclusive in pairs. Then, for every *n* from ?? to s, $s_{.,,n} = 1$. Hence from (25-1):

(25-8) $$\sum_n e_W(\mathrm{rf}, M_n) = \frac{s}{1+\sqrt{s}}\left(\frac{1}{\sqrt{s}} + \frac{1}{2}\right) = \frac{\sqrt{s}}{2} \cdot \frac{\sqrt{s}+2}{\sqrt{s}+1} > \frac{\sqrt{s}}{2}.$$

For every $s \geq 3$, the sum is > 1. For $s = 100$, it is >5; for $s = 10{,}000$, it is >50. These results are also incompatible with our conception that estimates of rf represent fair betting quotients.[3]

The features of the estimate-function e_W just discussed seem to me to make this function entirely unacceptable. Since the minimax principle

3. See [I], § 41 D (6).

leads with necessity to this function, it cannot, in my view, be regarded as a valid *general* principle for the choice of decision functions. This does not, of course, exclude the possibility that the principle may lead to suitable solutions in ocher problem situations.

In the main body of this monograph, which was written before I became acquainted with the above-mentioned publications by Wald and by Hodges and Lehmann, I repeatedly made remarks to the effect that all historically known estimate-functions for rf fulfil certain conditions. Some of these remarks must now be restricted, because e_W represents an exception. However, since we found e_W unacceptable from the point of view of inductive logic, the mere historical fact that this function has actually been proposed will not be sufficient to alter the considerations which guided the construction of the λ-system.

Our previous examination of the Laplace method (§ 12) led to results similar in some respects to those now obtained for e_W. We found that Laplace's rule (12-7) likewise yields unacceptable results clue to the fact that it is formulated in a too general form so as to apply to *all* properties without regard to their logical nature. We eliminated these unacceptable results by restricting the rule to properties with relative width $w/K = ½$, including the primitive properties. Then we constructed a general method, applicable to all properties, by choosing a simple λ-function such that, for the special case of $w/K = ½$, it leads to the same values as Laplace's rule. We called the method characterized by this λ-function (vie., λ = 2) the modified Laplace method.

We shall now proceed analogously in the case of e_W, in order to find a function which is similar to e_W without possessing its undesirable features. More specifically, we try to find a function e′ fulfilling the following two conditions, if that is possible:

(i) The function e′ belongs to the λ-system; hence it is characterized by a λ-function of the λ-system, say λ′.

(ii) For any property M with $w/K = ½$, e′ has the same values as e_W; hence in this case:

$$e'(\text{rf},M,K,e_M) = \frac{1}{1+\sqrt{s}}\left(\frac{s_M}{\sqrt{s}}+\frac{1}{2}\right).$$

Furthermore, we would prefer as simple a function λ′ as possible. Let us therefore tentatively add the following condition of simplicity; we shall find that there is indeed a function fulfilling all three conditions:

(iii) λ′ is independent of s_i; we shall accordingly omit s_i as an argument and write simply '$\lambda'(K,s)$'.

If (iii) is fulfilled, the following must hold (derived from (9-3), (4-5), (4-1), and (6-4), in analogy to (9-9)):

(iv) For every M,

$$\text{e}(\text{rf}, M, K, e_M) = \frac{s_M + (w/\kappa)\lambda'(\kappa, s)}{s + \lambda'(\kappa, s)};$$

hence:

(v) For every M,

$$\lambda'(K,s) = \frac{s_M + \text{e}'(\text{rf}, M, K, e_M)}{\text{e}'(\text{rf}, M, K, e_M) - w/\kappa}.$$

Now condition (ii) is to hold for a special kind of M; hence from (v):

(vi) For any M with $w/K = ½$,

$$\lambda'(K,s) = \frac{s_M - \dfrac{1}{1+\sqrt{s}}\left(\dfrac{s_M}{\sqrt{s}} + \dfrac{1}{2}\right)}{\dfrac{1}{1+\sqrt{s}}\left(\dfrac{s_M}{\sqrt{s}} + \dfrac{1}{2}\right) - \dfrac{1}{2}} = \sqrt{s}.$$

We choose now as λ′ that function which is *always* equal to $\sqrt{s}$, not only in the special case referred to in (vi); hence we define:

(25-9) $$\lambda'(K,s) = \sqrt{s}$$

We take as e′ the e-function characterized by this function λ′. Hence, according to (iv):

(**25-10**) For any property M,

$$\text{e}'(\text{rf}, M, K, e_M) = \frac{s_M + (w/\kappa)\sqrt{s}}{s + \sqrt{s}} = \frac{1}{1+\sqrt{s}}\left(\frac{s_M}{\sqrt{s}} + \frac{w}{\kappa}\right).$$

Comparing this with (25-1), we see that, in the case $w/K = ½$, e′ has indeed the same values as e_W. We propose this estimate-function e′ as a modification of e_W. It does not, of course, fulfil that special purpose for which e_W was originally constructed; it is not in accord with the minimax principle. However, e′ is similar to e_W. but is free. of the defects of this function. e′ is based upon a c-function c′ (see below, (25-11)) which fulfils our earlier conditions C1-C10; therefore, c′ and e′ belong to the λ-system. Furthermore, e′ fulfils the requirements of additivity (6-18), of the unique value (25-6), and of the upper bound 1 for the sum of estimates (25-7), which are all violated by e_W.

The method of estimation represented by e′ and characterized $\lambda' = \sqrt{s}$ is the same as one mentioned earlier in §§ 20 and 21 in the context of our discussion of the mean square error $1v_Q^2$ for all Q's (see the definition (20-10) and the result (21-13)). We did not examine it in detail, because this λ-function, though it belongs to the λ-system, does not belong to that narrower class to which we restricted most of our investigation. This class contains only those λ-functions which fulfil the condition C11 (§ 9) of being independent of s and s_i. However, the method of e′ may well deserve closer investigation. As explained earlier (§ 18), those values of λ which are neither too low (zero or small fractions) nor too high seem to lead to plausible values of e and c. Now, for small positive values of s, λ′ is likewise small, but not too small, since it is ≥ 1. [For s = 0, i.e., tautological evidence, we have λ′ = 0; here the value of e(rf,M,K,t) is, of course, w/K, like that of any other e-function of the λ-system; see (11-11) and (14-9).] For large values of s, λ′ becomes rather large. However, for large samples, the differences of the numerical values of the various e-functions are not large (as long as λ is neither ∞ nor extremely large). [Example. For s = 10,000 and M with w/K = ½, the value of e′ (and also that of e_W) is $(s_M + 50)/10{,}100$. Compare this, e.g., with the modified Laplace method e_2, which yields $(s_M + 1)/10{,}002$. The ratio of these two values is considerable when s_M is small. The difference between the two values, however, is only about 0.005 for any s_M.] .

We found earlier that, in the case that $\lambda^2 = s$, hence for the method of λ′ and e′, the mean square error $1v^2$(e′,M,k,s) in the state-description k is independent of r (the rf of M in k) if and only if w/K = ½; in this case:, $1v^2 = 1/[4(\sqrt{s} + 1)^2]$ (see (20-9)). Hodges and Lehmann show (*op. cit.*, p. 190) that, for their function e_W, $1v^2$ has always the value just stated, independently of r (e_W is a "constant risk estimate", where $1v^2$ is taken as the risk-function); there is no restriction with respect to M. For e′, however, the result is restricted, as mentioned above, to the case w/K = ½. That the result for e_W, holds for any M must not be regarded as an advantage of e_W; it is, on the contrary, a symptom for the too general character of the definition of e_W, which, as we have seen, leads to undesirable consequences. For no estimate-function of the λ-system is it possible that $1v^2$ be independent of r for all properties M; it is not even possible for any single property M with $w/K \neq$ ½. Further, according to (21-13), for the function e′ the mean square error for all Q's together in a state-description k ($1v^2$(e′,k,s)) has the same value in all state-descriptions: This may perhaps also be regarded as an advantage of e′.

We shall now briefly consider also the functions c′ and m′ corresponding to e′ and the characteristic function G' for c′ and e′. From the point of view of our theory, c′ and e′ belong together as parts of one complete inductive method. We obtain from (25-10) with (6-4):

(25-11) $$\mathrm{c}'(h_M,e_M) = \frac{s_M/\sqrt{s} + w/\kappa}{\sqrt{s}+1}.$$

Hence with (4-6)

(25-12) $$G'(\kappa,s,s_i) = \mathrm{c}'(h_i,e_i) = \frac{s_i/\sqrt{s} + 1/\kappa}{\sqrt{s}+1}.$$

Hence with (5-4):

(25-13) For any state-description k with the Q-numbers N_i,

$$\mathrm{m}'(k) = \prod_i \prod_{p=1}^{N_i} \frac{(p-1)/\sqrt{m_i+p-1} + 1/\kappa}{\sqrt{m_i+p-1}+1},$$

where the first product runs through those i for which $N_i > 0$.

Although the function given in (25-11) for the singular predictive inference is relativeiy simple, that given in (25-13) for a state-description is rather complicated. [In distinction to (12-1), the factors here do not increase by 1; therefore the products cannot be simply stated in terms of the Γ-function, let alone in terms of the factorial.] Consequently, it might be that theorems on c′ are, in general, rather complicated. Nevertheless, it would be worth while to study both functions c′ and e′ more in detail.

INDEX