RANDOM VARIABLES FROM THE POINT OF VIEW OF A GENERAL THEORY OF VARIABLES

KARL MENGER
ILLINOIS INSTITUTE OF TECHNOLOGY

1. Introduction

In his great book *Sequential Analysis*, Wald defines (see p. 5 in [1]) a random variable as a variable x such that "for any given number c a definite probability can be ascribed to the event that x will take a value less than c." As a first example of a random variable, Wald mentions the outcome x of the experiment of weighing an object selected at random from a lot of n known objects. He calls x a random variable "since a probability can be ascribed to the event that x will take a value less than c, for any given c." If n_c is the number of objects in the lot whose weight is less than c, that probability is n_c/n. On page 11, Wald says that "statistical problems arise when the distribution function of a random variable is not known and we want to draw some inference concerning the unknown distribution function on the basis of a limited number of observations." He then mentions, as an example, the random variable x assuming the value 0 if a unit selected from a completely unknown lot of products is nondefective, and the value 1 if the unit is defective.

In 1947, I submitted to Wald the following two observations: (1) the concept "variable" on which the notion of random variable is based (see p. 5 in [1]) does not appear to be that of a numerical variable, the only one then clearly defined; (2) the statement and example on page 11 seem to be at variance with the definition of random variables on page 5.

I believe that I carry out Wald's intentions by saying that he fully agreed with both remarks and expressed the hope to clarify the statistical concept of random variables at a later occasion. His untimely death in 1950, after the completion of his fundamental book on statistical decision functions (in which he essentially retained the treatment of random variables of *Sequential Analysis*) prevented him from carrying out this plan.

For the past few years I have tried to analyze the ideas behind the general term "variable"—a term that, in spite of its frequent and heretofore indiscriminate use, has never been introduced by a comprehensive definition (either explicitly, in terms of other concepts, or implicitly, by postulates). As a result of these studies [2], [3], [4], and especially [5], it appears that there is not one comprehensive concept of variable. The underlying material has been resolved into an extensive spectrum of concepts. That array begins in mathematical logic; it traverses algebra, analysis, the various types of geometry, and physical science; it touches social science, and it ends in statistics. Some of those concepts have only one common bond—the name variable. In content, they differ about as much as do the tangent of an angle in trigonometry and the tangent to a curve in geometry. But whereas no one has ever confused the latter two ideas because of a flimsy equivocation, the equivocal use of the term "variable" has indeed resulted in confusion. Some of those ambiguities account for the obscurity in the foundations of pure analysis, others for the lack of articulate rules concerning certain applications of analysis to science. "Variables" will be discussed in sections 2–5.

The term "random," which is widely used in statistics and the theory of probability, is in a condition that very much resembles that of "variable." It will be analyzed in section 6.

In sections 7–11, these results are applied to random variables. Various types must be distinguished even within the realm of these "variables." In the last section, the relation between random variables in statistics and in the theory of probability will be shown in a new light.

2. Logico-mathematical variables

In pure mathematics, the only concepts of variable that possess clear traditional definitions are of the type of the so-called *numerical variables*. The latter concept may be illustrated as follows. The formulas

$$(1) \qquad \frac{2}{3^2-1}=\frac{1}{3-1}-\frac{1}{3+1}, \qquad \frac{2}{e^2-1}=\frac{1}{e-1}-\frac{1}{e+1}, \ldots,$$

and countless other formulas can be synthesized in one formula containing a letter and accompanied by, as it were, a legend with directions concerning the use of the letter. Such a general statement is

$$(2) \qquad \frac{2}{x^2-1}=\frac{1}{x-1}-\frac{1}{x+1},$$

where x may be replaced by the designation of any number $\neq 1$ and -1. Usually, the legend is abbreviated to "for any number x $\neq$ 1 and -1." The class of all numbers $\neq 1$, -1 is called the *scope* of the numerical variable x in (2). If x is replaced by 3, e, and other numbers, specific statements result, namely, (1) and other formulas.

The statement

$$(3) \qquad (x+3y)^2 = x^2+6xy+9y^2 \qquad \text{for any x and any y}$$

involves two numerical variables. The meaning of (2) and (3) is not changed if x is replaced by any other letter (except e), or x and y by any two nonidentical letters, say, by a and b or by a and y, and even if they are interchanged as in

$$(4) \qquad (y+3x)^2 = y^2+6yx+9x^2 \qquad \text{for any x and any y}.$$

If the scope of a numerical variable consists of a single number, then the letter designates a specific number as, for instance, does e.

In this paper, all numbers and numerical variables are printed in roman type, while italics are reserved for functions, function variables to be defined presently, and concepts introduced in subsequent sections.

Analysis abounds in laws of the following type:

$$(5) \qquad \mathbf{D}(f\cdot g) = f\cdot\mathbf{D}g+g\cdot\mathbf{D}f,$$

where *f* and *g* may be replaced by the designations of any two differentiable functions, such as *log*, *cos*, etc. The letters *f* and *g* in (5) are called *function variables* each having the class of all differentiable functions as its scope. The following statement,[1] which is equivalent to (5), involves numerical as well as functional variables:

$$\mathbf{D}\,(f \cdot g)\,\mathrm{x} = f\mathrm{x} \cdot \mathbf{D}\,g\,\mathrm{x} + g\,\mathrm{x} \cdot \mathbf{D}\,f\mathrm{x} \tag{6}$$

for any f, g, and any x for which f, g, and $f \cdot g$ have derivatives.

More generally, if a formula contains a letter and is accompanied by a legend concerning the replacement of the letter by the designations of elements of a certain class, then logicians and mathematicians refer to the letter as a *variable*, and call the said class the *scope* of the variable.

In examples (2) to (6), each replacement results in a statement. In such a case, I shall say that the variable is used *indicatively*. One of the numerous other uses of variables is illustrated in the following definition.

(7) Let *log* be the class of all pairs (x, *log* x) for any $\mathrm{x} > 0$.

If, in the formal part of (7), x is replaced, say, by 3, the result is an element of the class defined by (7), namely, the pair (3, *log* 3), and not a statement. I shall refer to this use of x as *conjunctive*.

3. Consistent classes of quantities

The transition from logico-mathematical variables to the objects of statistics (the height, the weight, etc.) and to what scientists call variables (the mass of radium, the time, etc.) is a step into a different world. This contrast is what I had in mind when comparing "variables" in various fields with "tangents" in the fields of trigonometry and geometry.

A simple scheme that is useful in the mathematical treatment of scientific and statistical material is supplied by the following concept of *quantity:* an ordered pair in which the second member (the *value* of the quantity) is a number, whereas the first member (the *object* of the quantity) may be anything. I shall call two quantities *consistent* unless their objects are equal, and their values are unequal. For instance, if A is a certain resident of Chicago, then (A, 69) is a quantity. Here 69 may be A's height in inches. If B is another resident of Chicago, (B, 71) and (B, 169), where 71 may be B's height, 169 his weight in pounds, are two inconsistent quantities, though either is consistent with the quantity (A, 69).

Consider the class of all pairs (C, *h* C) for any resident C of Chicago, where *h* C denotes C's height in inches. I shall designate this class by *h*. It is what statisticians study under the name of the height in inches within the population of Chicago. Any two quantities belonging to *h* are consistent. I shall call a class with this property a *consistent class of quantities*—briefly, c.c.q.

In the definition of *h*, the letter C is a logico-mathematical variable (used conjunctively) whose scope is the population of Chicago. Replacing C by A or B, one obtains the quantities (A, *h* A) = (A, 69) and (B, *h* B) = (B, 71) belonging to *h*. But the class *h*, as a whole, is totally different from a numerical variable. It is not a symbol that may be replaced by designations of specific numbers. It is a class of pairs whose

[1] The typographical convention mentioned above (that is, the use of roman type for numbers and numerical variables, and of italics for functions and function variables) makes it unnecessary to say in the legend to (6): for any "functions" *f*, *g* and any "number" x. Moreover, without any danger of confusion, *f*(x) can be abbreviated to *f*x.

second elements are numbers; and it is a specific class, just as 3 is a specific number and *log* is a specific function. That *h* and, in fact, *any object of statistical studies is totally different from numerical variables* is the **first point** that I wish to stress in this paper.

Another example of a consistent class of quantities is *w*, the weight in pounds of the population of San Francisco—the class of all pairs (D, *w* D) for any resident D of San Francisco, where *w* D denotes the weight in pounds of D. Also the union or set-theoretical sum of the classes *h* and *w* is a consistent class of quantities, though it is not likely ever to be practically significant.

Among the highly significant notions covered by the concept "consistent classes of quantities" are all those to which scientists refer as "variables" and "variable quantities." (In this connection, I shall avoid the former term to forestall confusion with logico-mathematical variables.) That *this vast material actually comes under the heading of consistent classes of quantities* is the **second point** that I wish to emphasize. Since those concepts are not usually defined as classes of ordered pairs, the preceding remark will be illustrated by two examples.

Consider decaying radioactive substances and let *m* be the mass in grams. This mass is a consistent class of quantities whose objects are instantaneous specimens of the substance or substances under consideration, that is, pieces of the material at a definite instant. If $m\ \mu$ denotes the mass in grams of the specimen μ, then *m* may be defined as the class of all pairs $(\mu, m\ \mu)$ for any instantaneous specimen μ.

If *P* is a specific swinging pendulum, denote by *a* the angle in radians between the pendulum and a vertical line. The variable quantity *a* is the class of all pairs $(\sigma, a\ \sigma)$ for any state σ of P, where $a\ \sigma$ is the said angle in the state σ.

The class of all objects—all specimens, all states, etc.—is called the *domain* of the variable quantity; the class of all values of a variable quantity is referred to as its *range*.

A c.c.q. whose domain consists of numbers (that is, a class of ordered pairs of numbers) is called a *function*—by mathematicians as well as scientists. An example is the logarithmic function or the function *log* as defined in (7). More generally, for any positive integer p, a consistent class of quantities whose objects are ordered p-tuples of numbers is universally called a p-*place function*. Examples include the maximum, the sum, the average, and weighted means of ordered p-tuples of numbers.

While some mathematicians propose to call any c.c.q. a function—even the mass *m*, a function whose domain is the class of specimens—all physicists and many mathematicians refer to *m* (by itself) as a variable quantity and not as a function—a term that they reserve for *log*, the exponential function, the sum, etc. The difference is purely terminological and, therefore, utterly unimportant.

However, even mathematicians who refer to the height *h* and the mass *m* as functions need a specific term for what every scientist calls functions. They may refer to *log*, to the exponential function, etc., as "functions in the narrow sense of the word" or in some other cumbersome way. But they must refer to them *somehow*, because—and this has nothing to do with terminology, but is a fact and, I believe, a rather significant fact, and the **third point** that I wish to stress in this paper—*there is an important difference between functions (such as log and exp) and other consistent classes of quantities*. Not only are functions the only c.c.q.'s for which derivatives and integrals can be defined—even the most general theories of integration are inapplicable to the height *h* and the mass *m*—but *functions are the only c.c.q.'s into which other c.c.q.'s can be substituted*. One can define the logarithm of the mass and the logarithm of the cosine, but there is no mass of the height nor a mass of the logarithm.

In discussions, algebraists usually deny the significance of the last point. The point was questioned also in the discussion following the presentation of this paper. Opponents claim that even the mass lends itself to substitutions, namely, of "functions" whose values are instantaneous specimens. And indeed, if with each block β of pitchblende one associates the specimen $\mu(\beta)$ of radium that can be extracted from β, then he can substitute this "function" into m and thereby define the class of all pairs $[\beta, m\, \mu(\beta)]$ as the mass of radium in blocks of pitchblende. But he has not substituted a consistent class of quantities, since the values of a c.c.q. are numbers and not specimens of radium. Moreover, since Galileo, most quantitative laws of science are in terms of substitutions such as that of m into *log*. These laws express the connections between scientific c.c.q.'s by means of functions, for example, the connection of the time with the mass by the function *log*.

The applicability of numbers and functions to the most diverse quantities and c.c.q.'s is the very reason for the omnipresence of the former in science, whereas specimens of radium and m occur only in certain branches of physics, and inhabitants of Chicago and h are studied only in sociology. What some modern algebraists, in a spirit of hyperformalism, seem to overlook is the specific role of mathematics as a universal tool—a role which is epitomized in the characterization of functions among consistent classes of quantities.

4. Constant variable quantities. Observations. Consistent classes of pairs

In contrast to numbers, variable quantities can be divided into those that are constant and those that are not. A partial realization of this distinction is probably what has prompted some analysts to contrast "constants" and "variables," even though in the absence of an explicit definition of variable quantities they usually express that distinction in an obscure way.

I shall call a c.c.q. *constant* if its range consists of one single element, and nonconstant otherwise. For instance, the mileage as well as the speed of a specific car is constant while the car is parked. At least the former is nonconstant while the car is moving. In a parked car, the range of the speed consists of the single number 0; that of the mileage consists of the value attained when the car was parked.

In particular, of course, there are constant and nonconstant functions. The constant function *3* (designated by an italic) of the value 3 (in roman type) is the class of all pairs (x, 3) for any number x.

Useful in a mathematical treatment of statistical and probabilistic material is also the following concept of *observation:* a pair in which the first member is an *act* of observation, and the second member is the *result* of that act. Essentially, this concept is due to von Mises who refers to an act as "Beobachtung" and to the result as "Merkmal" (characteristic).

If the result is a number, for example, a scale mark, then the observation is a quantity. But there are acts of observation the results of which are p-tuples of numbers, for example, the p numbers of points observed when p dice have been rolled. There are also acts with altogether nonnumerical results, such as head and tail observed when a coin has been tossed.

The domain of m is also the domain of other variable quantities: an instantaneous specimen μ has also a volume, a temperature, etc. In contrast, an act of observation has, in general, only one result. (If the mass of μ is being observed on scales calibrated in different units, then the results are different, but so also are the acts.) Hence any two observations are consistent.

With most scientific variable quantities, such as m and a, there correspond classes of observations. For instance, there is a class m^* of *mass observations* in grams consisting of all pairs $(\beta, m^*\beta)$ for any act β directed to a scale (calibrated in grams) on which a specimen of a decaying substance is being weighed, where $m^*\beta$ denotes the result of the act β.

Various acts of observation $\beta_1, \beta_2, \cdots$ may be directed to the same instantaneous specimen μ; for example, acts of various observers or, if the decay is a quasi-stationary process, successive acts of the same observer. The results $m^*\beta_1, m^*\beta_2, \cdots$ may well be unequal, in which case the class of all quantities (μ, r) for any specimen and any result r of an act directed to μ is not consistent. In fact, each specimen μ gives rise to a variable quantity m^*_μ of all mass observations directed to μ, and the value $m\mu$ of m is somehow derived from the variable quantity m^*_μ.

As Dr. M. A. Woodbury pointed out in the discussion of the present paper, the preceding remarks apply to a certain extent also to functions. For instance, with the function *log* there corresponds a variable quantity *logarithmic computation:* the class of all pairs each consisting of an act of computing the logarithm of a number and the result of that act.

Clearly, the *actual domain* of a class of observations consists of a finite number of acts, the *potential domain* (including the acts that may still be carried out) may be either finite or what I shall call *indefinite*. Similarly, one has to distinguish the *actual range* (that is, the class of all actual results) and the *potential range* (including the possible results of acts belonging to the domain). In a class of at most five observations of a die (and in some larger classes) the range consists of at most five numbers, the potential range of six. The potential range is what von Mises called "Merkmalmenge" (mistranslated into "sample space"—neither are the elements samples nor is, in general, the range a space). The actual range of the class of all observations directed to the gravitational acceleration at a certain place consists essentially of one number. The acceleration at the particular place is constant.

Classes of observations are special *consistent classes of pairs*. A c.c.p. is a class of pairs (of any kind) not containing two pairs whose first members are equal and whose second members are unequal.

5. C.C.P. variables

Just as (3) is valid for any two numbers, and (5) for any two functions, the following statement is valid for any two consistent classes of quantities of a certain kind:

$$\text{If } w = log\ u, \text{ then } \frac{\mathrm{d}w}{\mathrm{d}u} = \frac{1}{u} \tag{8}$$

for any two c.c.q.'s w and u, if the range of u consists of positive numbers.

The letters u and w in (8) are symbols accompanied by a legend according to which u and w may be replaced by the designations of specific c.c.q.'s. For instance, one may replace u by m, the mass present in a chemical reaction, and w by the time; or u by r, the distance from a certain charge, and w by p, the potential due to the charge. One thus obtains the statements: If $t = log\ m$, then $dt/dm = 1/m$; and if $p = log\ r$, then $dp/dr = 1/r$. In other words, u and w are logico-mathematical variables—but variables whose scopes consist of c.c.q.'s. I therefore shall call u and w in (8) *c.c.q. variables.* (Such variables must not be replaced by numbers. If, in (8), u were replaced by 1, and w by 0, the implication would be nonsensical even though the antecedent would be valid.)

The confusion in the literature is epitomized in the traditional misstatement that the following formulas (often presented without legends) have the same meaning:

(9) $$\frac{d\, log\, u}{du} = \frac{1}{u}$$ for any c.c.q. u whose values are positive ;

(10) $$\mathbf{D}\ log\, \mathrm{x} = \frac{1}{\mathrm{x}}$$ for any number $\mathrm{x} > 0$.

Actually, (9) deals with the rate of change of the c.c.q. variable $log\ u$ with u, and (10) with the value for x of the derivative of the function *log*. (If w is connected with u by the function f, then the rate of change of w with u is connected with u by $\mathbf{D}f$, the derivative of f.)

It is clear how c.c.p. variables are to be defined.

The development of the idea of c.c.q. variables is the **fourth point** that I wish to emphasize in concluding the discussion of "variables" in analysis and science. Traditionally, the term "variable" has been used indiscriminately in the sense of

(1) numerical variable;

(2) consistent class of quantities (in particular, for scientific variable quantity);

(3) c.c.q. variable.

6. Various meanings of the term "random"

The following descriptions include five of the most important uses of the term "random" in statistics.

(A) Randomness is often attributed to *samples* from a population, that is, to subclasses of a class. Analysis reveals that the subclasses cannot be divided into those that are, and those that are not, random samples or even potential random samples, since in many cases every subclass is a possible random sample. Randomness in connection with samples is a property of selections of subclasses from the population rather than a property of subclasses. Its rigorous treatment (of which no example is known to me) would present exceedingly difficult problems, since a sound theory would have to be formulated in terms of (a) acts of selection, (b) aims of selections, (c) information available—as shown by the following simple examples. One cannot regard the selection of those who are over six feet tall from the people passing a busy corner as a random selection in a study of height or even weight, though one might regard it as such for other purposes, for example, possibly in a study of income. Neither would the former purposes be served by the selection of every one-hundredth passer-by if one should know that arrangements have been made as a result of which the 100th, 200th, 300th,··· passer-by would be over six feet tall. Without such knowledge, however, one might regard the choice of every one-hundredth as a random selection.

(B) The word "random" is used with regard to certain *sequences*, namely, to irregular sequences of the type of von Mises' collectives (a concept made precise and consistent by Wald's relativization to a definite set of principles of selection) and to *sample numbers*.

(C) Numerous references are made to random *events* and, in a related sense, to random *experiments*. Random and nonrandom events in the sense in which these words are used by a man on the street will be contrasted in a simple example. Suppose that a man were offered a game based on the following understanding. Unobserved by the man, two coins are tossed. A reliable friend of the man inspects the outcome. If no tail turns up, the friend will say "zero"; if at least one tail appears, he will, at his discretion, say either "one" or "two." The man on the street will attribute random character to the two events "zero" and *either* "one" *or* "two." But he will not refer to any of the four events "one," "two," *either* "zero" *or* "one," and *either* "zero" *or* "two" as random events. The formulation of general principles guiding the man on the street in attributing random character to some events and not to others is an important (and, to my knowledge, unsolved) problem—important, because it might suggest postulates for the implicit definition of random events.

(D) Randomness is attributed to *processes*, such as Brownian motion and diffusion. The partly obscure relations of the meanings (A), (B), (C) with each other and with (D) will not here be investigated. But agreement seems to exist in the literature as to the relation between (D) and the following type (E). Random processes are generally considered as classes or families of random variables.

(E) Thus the consideration of the various uses of the terms "variable" and "random" has finally led to the concept discussed in the introduction to this paper. In the literature, *random variables* (or, briefly, r.v.'s), in contrast to the types (A) and (C) of randomness, are introduced by definitions. Besides the one quoted in the introduction, there is the well-known measure-theoretical definition of a r.v. But are these two definitions equivalent? And do they cover all that statisticians actually study under the name of r.v.'s? For reasons discussed in the following sections, I do not believe that these questions can be answered affirmatively.

A r.v. in either sense will be seen to be a consistent class of pairs, and consequently general statements that are valid for *any* r.v. may conveniently be expressed in terms of c.c.p. variables. With numerical variables, r.v.'s have, notwithstanding frequent remarks to the contrary [6], [7], nothing whatever to do.[2]

7. Probabilistic and statistical random variables

The **fifth point** here to be stressed is the fact that, *in the literature, two different concepts have been studied under the name of "random variable."* A r.v. as defined by Kolmogorov and others in the *theory of probability* (that is, a measurable function on what Neyman calls a fundamental probability set, briefly, f.p.s.) is not identical with a r.v. as envisaged by Wald and others in *statistics* (that is, a consistent class of pairs for which a distribution function is known). A r.v. of either type is a consistent class of pairs that is related to a fundamental probability set, but with the basic difference that a probabilis-

[2] This point was clearly recognized by Halmos who, in his excellent treatment of probability, says, "A random variable is a function, a function whose numerical values are determined by chance...in other words, a function attached to an experiment" (see p. 188 in [8]). However, I fail to see the basis of Halmos' factual statement, "Ever since rigor has come to be demanded in mathematical definitions, it has been recognized that the word 'variable,' particularly a variable whose values are 'determined' somehow or other, means in precise language a function" (see p. 187 in [8]). On the contrary, the classical treatises by de la Vallée Poussin, and G. H. Hardy, as well as recent books by R. Courant and A. A. Albert, take positions that are the direct opposite of what Halmos calls the recognized point of view.

tic r.v. has a f.p.s. as its *domain*, whereas a statistical r.v. has a f.p.s. as its *range*. More precisely, the two concepts can be defined as follows:

A probabilistic r.v. is a triple $\mathfrak{R}$, p, v consisting of

(1) a σ-field, $\mathfrak{R}$, of sets (called *random events*) including a maximal set R_{max}, that is, a set that contains each element of $\mathfrak{R}$ as a subset;

(2) a σ-additive function p (called *probability*) whose domain is $\mathfrak{R}$ and which assumes the value 1 for R_{max};

(3) a consistent class of quantities v whose domain is R_{max} and which has the following property: If $R_{v,x}$ denotes the class of all elements α of R_{max} such that $v\alpha < x$, then $R_{v,x}$ belongs to $\mathfrak{R}$, for any number x. (This property is also expressed by saying that the event that v assumes a value less than x is a random event.)

The function p_v assuming the value $p_v(x) = p(R_{v,x})$ for any number x is called the *probabilistic distribution function* of the random variable $\mathfrak{R}$, p, v.

A statistical r.v. is a triple w, $\mathfrak{S}$, q consisting of

(1) a consistent class of pairs w;

(2) a σ-field $\mathfrak{S}$ of subsets of Ran w (the potential range of w) that includes Ran w itself (the elements of $\mathfrak{S}$ are called *statistical random events*);

(3) a σ-additive function q (called *statistical probability*) whose domain is $\mathfrak{S}$ and which assumes the value 1 for Ran w.

An example of a probabilistic r.v. is the triple $\mathfrak{T}$, r, f ,where $\mathfrak{T}$ is the field consisting of the four sets: the empty set, $\{0\}$, $\{1, 2\}$, and $\{0, 1, 2\}$ (the maximal set); r is the additive function whose values for the afore-mentioned random events are 0, 1/4, 3/4, and 1, respectively; and f is the consistent class of quantities (0, 0), (1, 1), and (2, 1), that is, the function assuming the value $x \cdot (3 - x)$ for any x in $\{0, 1, 2\}$.

If r is replaced by the function r' assuming the values 0, 1/3, 2/3, and 1, one obtains a different probabilistic r.v. A third is $\mathfrak{T}_1$, r_1, f, where $\mathfrak{T}_1$ consists of all eight subsets of $\{0, 1, 2\}$ and r_1 is the additive function assuming the values 1/4, 1/2, and 1/4 for the sets $\{0\}$, $\{1\}$, $\{2\}$, respectively.

If j is the c.c.q. consisting of the three pairs (0, 0), (1, 1) and (2, 2), that is, the identity function on $\{0, 1, 2\}$, then neither $\mathfrak{T}$, r, j nor $\mathfrak{T}$, r', j is a probabilistic r.v. Indeed, the event that j assumes a value $<3/2$ is the class $\{0, 1\}$, which does not belong to $\mathfrak{T}$. However, the triple $\mathfrak{T}_1$, r_1, j is a probabilistic r.v.

An example of a statistical r.v. is the triple a, $\mathfrak{T}$, r, where a is the outcome of the game described in section 6 (C), that is, the class of all pairs (α, $a\,\alpha$) for any act α of the man listening to the announcements "zero," "one," and "two" of his friend, where $a\,\alpha$ is the result of α, and the announcements are labeled 0, 1, and 2, respectively. The triples a, $\mathfrak{T}$, r' and a, $\mathfrak{T}_1$, r_1 are different statistical r.v.'s.

Let h be the function assuming the values 0, 1, and 2 for all numbers between $-\pi$ and $-\pi/3$, between $-\pi/3$ and $\pi/3$, and between $\pi/3$ and π, respectively. If $\mathfrak{U}$ is the class of all Borel subsets of the sum of the three open intervals, and t is the Borel-Lebesgue measure, then h, $\mathfrak{U}$, t is a statistical r.v., while $\mathfrak{T}$, r, h is a probabilistic r.v. (just as a, $\mathfrak{T}$, r is a statistical r.v., while $\mathfrak{T}$, r, f is a probabilistic r.v.).

On a roulette that is divided into three sectors each subtending $2\pi/3$ radians, one can play a game such that for its outcome, g, the triple g, $\mathfrak{U}$, r' (where r' has its previous meaning) is a statistical r.v. Any game of dice or cards in conjunction with a class of random outcomes and their probabilities gives rise to a statistical r.v.

8. Qualitative random variables

Neyman and other outstanding statisticians, after defining what in the preceding section has been called probabilistic random variables, emphasize that the most interesting problems concerning r.v.'s arise in so-called hypothetical cases, where the probability and distribution functions are unknown. For the same reasons that prompted observation (2) mentioned in the introduction, I cannot quite accept this point of view. How can one, after having defined r.v. relative to a σ-field of sets as well as to a σ-additive function p (even if p enters only into the definition of the distribution function of the random variable) discuss random variables for which no function p is known?

One must not answer that even where p is unknown, this function exists. This answer would be mathematically significant only if the assumption of the mere existence of a σ-additive function on a σ-field $\mathfrak{R}$ imposed restrictions on $\mathfrak{R}$. This, however, the assumption does not do. There exist σ-additive functions (assuming the value 1 for R_{max}) on every σ-field with a nonvacuous maximal set. An example is, if s denotes any element of R_{max}, the function assuming for a set R in $\mathfrak{R}$ the value 1 or 0 according to whether R does or does not contain s. Hence no restriction whatever is imposed on $\mathfrak{R}$ by assuming that a probability function exists.[3]

It rather appears—and this is the **sixth point** that I wish to emphasize—that *statisticians frequently study* (especially in problems to which they attach particular interest) *a more rudimentary concept than random variables*, namely, mere pairs $\mathfrak{R},u$ or $w,\mathfrak{S}$ rather than triples including also a σ-additive function on the σ-field of sets. I shall refer to such pairs as *qualitative random variables*. Particularly important are statistical qualitative r.v.'s.

Let g be the class of all pairs $(\alpha, g\,\alpha)$, for any act α of reading the points that turn up in a game of dice, where $g\,\alpha$ is the result of the act α. Let $\mathfrak{S}$ be the class of all subsets of Ran g, for instance, if the game is with one die, the class of the 2^6 subsets of $\{1, 2, 3, 4, 5, 6\}$. If I am told that the die is loaded but not given any other information (in particular, no records of past outcomes), then I may well decide to consider g, $\mathfrak{S}$ a qualitative statistical r.v.

If a, $\mathfrak{T}$ and j have the same meaning as in section 7, and $\mathfrak{T}'$ denotes the class of all subsets of $\{0, 1, 2\}$, then $a,\mathfrak{T}'$ and $\mathfrak{T},j$ are not even qualitative r.v.'s.

9. R.V. variables

After what has been said in section 5, two examples will sufficiently illustrate the concept of r.v. variables.

If $\mathfrak{R}$ is the σ-field of all subsets of Dom u, then $\mathfrak{R}$, u is a qualitative probabilistic r.v. for any c.c.q. u. If Dom u is finite, and p is the additive function assuming equal values for any two subsets consisting of a single element of Dom u ("elementary events"), then $\mathfrak{R}$, p, u is a probabilistic r.v. Similarly, if $\mathfrak{S}$ is the σ-field of all subsets of Ran w, then w, $\mathfrak{S}$ is a qualitative statistical r.v. In the preceding statements, u and w are c.c.p. variables.

Also the following somewhat curious general statement deals with concepts that belong to the category of r.v. variables. *The probability of denumerably many incidents all*

[3] This point was brought out in a discussion with Dr. R. Seall and Mr. M. McKiernan. The assumption that a probability function *of a certain kind* exists may impose restrictions of $\mathfrak{R}$. For instance, the existence of a function p assuming infinitely many values obviously presupposes that $\mathfrak{R}$ contains infinitely many elements.

of which give rise to random events is itself a probabilistic random variable. More precisely, assume that $\mathfrak{R}$ includes all subsets of R_{max}, and that R_{max} is denumerable. Then, for any σ-additive function p on $\mathfrak{R}$, one can define a function p^* on R_{max} assuming for any "incident" (that is, for any element r of R_{max}) the same value that p assumes for the corresponding "elementary event" (that is, the set $\{r\}$). The triple $\mathfrak{R}$, p, p^* is a probabilistic r.v.

10. Four types of statistical studies concerning consistent classes of pairs

The definition and the treatment of probabilistic r.v.'s in the literature have been considerably more precise and lucid than those of statistical r.v.'s. The reason is that, in the case of the former, nothing conceptually different has to be added to the ideas of σ-field and σ-additive function on which mathematicians have concentrated for a long time. In contrast, statistical r.v.'s have domains whose elements (such as acts of observation or physical objects) are, as it were, one step closer to reality, and concerning elements of this kind pure and even applied mathematics have been rather inarticulate. Yet an analysis reveals that most statistical studies actually deal with those domains and their elements.

(1) In the case of some variable quantities, all a safety engineer wishes to ascertain is the maximum and/or minimum value. (The difference between these two numbers is what most statisticians call the "range" of the variable quantity.) Such studies do not require references to the domains of the variable quantities.

(2) The typical statistical investigations of individual c.c.p.'s, which are concerned with frequencies, cannot be confined to the ranges and do require references to the domains of the c.c.p.'s. For instance, all that can be said about the range R of a class of observations in relation to an element r of its potential range is either that R does, or that R does not, contain r. (In any class or set, any element has, as it were, the frequency 1 or 0.) It cannot be correctly said (although it sometimes *is* said) that R contains r, say, 3 times. The frequency 3 is associated with r if and only if the domain of the class of observations contains exactly three acts with the result r. Thus all significant processes of counting (and, in more complicated cases, measuring) in statistical investigations are performed within the *domains* of c.c.p.'s. Subsequently, the numerical results of the processes are paired with the elements of the *ranges*, whereby the latter are transformed into what might be called *weighted classes*. Oddly enough, no traditional term (such as *weighted* and *relatively weighted* class) exists for the concept of a class with a frequency or relative frequency defined for the elements, in spite of the paramount importance of this concept in statistics and, incidentally, in algebra.

(3) Typical order statistics (the theory of runs, etc.) are sequences of observations or quantities. It was in his study of sequences of a certain kind ("collectives") that von Mises introduced the distinction between acts and results. Just for sequences of pairs, however, frequency studies do not really require references to the first members of the pairs. Consider, for instance, the sequences of observations

$$(a_1, r_1), (a_2, r_2), \cdots, (a_n, r_n). \tag{11}$$

All frequency studies concerning the sequence (11) can be based on the sequence r_1, r_2, $\cdots$, r_n of the results. One associates the frequency 3 with the result r if there are exactly three indices, say, 2, 3, and 7, such that $r_2 = r_3 = r_7 = r$. Thus the indices (which constitute, as it were, the domain of the *sequence* and are attached to the results as well as to the acts) may take over the role of the acts (which constitute the domain of the *class of pairs* belonging to the sequence).

(4) In investigations into the correlation of two c.c.p.'s and the regression curve of a variable quantity w on a variable quantity v, references to the domains of the c.c.p.'s and to Dom v and Dom w are absolutely essential. What one primarily pairs are elements of those domains. Subsequently one studies pairs of elements of the ranges and pairs of values of v and w, induced by that primary pairing.

Summarizing one can say that few statistical investigations—essentially only those in order statistics—can dispense with references to the domains of c.c.p.'s. *In frequency studies concerning nonsequential c.c.p.'s, for instance, concerning one variable quantity as well as connections between two variable quantities, references to the domains are indispensable.* This is the **seventh point** that I wish to emphasize.

The remarks about correlation and regression curves bring out an important **eighth point.** *The questions as to what is the correlation coefficient or the regression line of w on v are unanswerable since they are incomplete.* In problems of this kind, references not only to the elements of Dom v and Dom w but to specific pairs of elements (one element belonging to Dom v, and one to Dom w) are indispensable. *A definite pairing of the domains (or of subclasses of the domains) must be given, and only relative to that pairing can the afore-mentioned questions be investigated.*

Pairings are mentioned by statisticians. To physicists, since Galileo and Boyle, pairing of *simultaneous* acts of observation has become second nature and is tacitly understood. If m^* is weight observation, say, with regard to a piece of radium that in 1900 weighed 1 gram, and t^* is time observation, then Dom m^* consists of acts of scale readings, and Dom t^* of acts of calendar readings. In the formula

$$t^* = 1900 - 2.3 \cdot 10^3 \, log \; m^*, \tag{12}$$

it is perfectly, if implicitly, understood that the values of t^* and m^* for simultaneous scale and calendar readings are being connected. Formula (12) is an abbreviation for

$$t^*\tau = 1900 - 2.3 \cdot 10^3 \, log \; m^*\beta \, , \tag{13}$$

for any pair (β, τ) of simultaneous acts of scale and calendar readings.

Formula (12) does not connect *any* value of t^* and *any* value of m^*. Hence it is completely misleading when some mathematicians claim [9] that laws such as (12) deal only with the values of quantities, and not with quantities themselves; that t^* and m^* are numerical variables, and the like. If m^* and t^* were numerical variables, they could be interchanged as can x and y in (3).

But if P and P_1 are two pendulums, and a and a_1 are their angles with a vertical line (compare section 4), then the relation between a and a_1 relative to the class Σ of all pairs (σ, σ_1) of simultaneous states may not be of particular significance. One may be more interested in the class Σ' of pairs (σ, σ_1'), where σ_1' is the state of P_1, one quarter of a period after σ_1. The results of pairing a and a_1 relative to Σ' and Σ are quite different. It may well be that $a_1 = -a$ relative to Σ', while a_1 is not at all a function of a relative to Σ and a different function of a relative to another pairing Σ'' of the domains.

The omission of references to the pairing of the domains and even to the domains altogether accounts for the lack of articulate rules concerning the application of analysis to science.

The need for a pairing of the domains *as one of the data* becomes perfectly obvious on the level of c.c.p. variables. For, what pairing of Dom v and Dom w—of one abstract set with another abstract set—is "natural"? To be applicable to radioactivity as well as to sociology, general statistical statements must contain not only c.c.p. variables, such as v and w, that may be replaced, say, by m^* and t^* in one case, and both by the height of men in Chicago, in the other. They must contain also a pairing variable Π that may be replaced by the pairing, say, according to the simultaneity in one case, and according to the father-son relation, in the other. That this procedure has been neglected in the literature is due to the lack of an explicit definition and of a clear treatment of c.c.q. variables.

11. What are statistical random events?

Modern statisticians have derived ingenious methods for the translation of a frequency record concerning the domain of a qualitative statistical random variable w, $\mathfrak{S}$ into a definition of a probability function q on $\mathfrak{S}$. Theories guide the statistician in choosing q.

Less attention seems to have been paid to the problem of defining $\mathfrak{S}$, even though guiding principles in this respect would be equally desirable.

Suppose the man on the street plays a long series of games with his friend, as described in section 7 (C). If, at the outset, he is told that the coins are unbiased, then, on the basis of past experience with unbiased coins, he will assume the probabilities 1/4 and 3/4 for 0 and 1 or 2, respectively. If he is told that the coins are somehow biased without being given further information, he will set up a frequency record on the basis of which a statistician will advise him. But concerning the outcomes 1 and 2 which, according to his understanding with his friend, are entirely left to the latter's discretion, no frequency record will be significant. Even if the friend should seem to speak the truth (which would give the probabilities 1/2 and 1/4 to 1 and 2, respectively) or if he should use random sample numbers to simulate probabilities 3/8 for both 1 and 2, the man on the street will not make predictions even on averages of future outcomes since he knows that his friend may change his policy at any time.

If the friend plays the same game also with a second person, who was told (rightly or wrongly) that the friend would always speak the truth, then the same frequency record that was insignificant to the first man will appear to be significant to the second.

12. The connection between probability and statistics

Many authors have likened the relation between probability theory and statistics to that between postulational and physical geometry. Postulational geometry and probability are based on unproven assumptions concerning undefined concepts, whereas physical geometry and statistics deal with phases of reality that approximately satisfy the postulates. If chalk dots and chalk streaks on a blackboard are called "points" and "lines," then Euclidean geometry, dealing with undefined points and lines is approximately valid. If the outcome of rolling a die is called "incident" and the relative frequency of an outcome in a long series of trials "probability" of the incident, then the theory of a certain probabilistic random variable (whose domain consists of six undefined elements called "incidents") is approximately valid.

Besides postulational and physical geometry, there exists a third theory dealing with points and lines, namely, pure analytic geometry. There (in contrast to physical geometry), points, lines, etc. are *defined*, but (in contrast to physical geometry) they are defined without any reference to reality, namely, as ordered pairs of numbers, as classes of such pairs satisfying linear equations, etc. As a concluding **ninth point,** I wish to emphasize that, besides the postulational probability theory and statistics, *there exists a third theory dealing with incidents, events, probability, etc. In it* (in contrast to the postulational theory) *events and probabilities are defined;* but (in contrast to statistics) *they are defined without any reference to reality.*

I shall illustrate this idea in the case of the weak law of large numbers: The probability is close to 1 that, in a sufficiently large sequence of independent trials of the same kind, the relative frequency of success is very close to the a priori probability of success in each trial. The purely mathematical background of the (more colorful than lucid) traditional formulation of this important law is the following:

FIRST COMBINATORIC LAW OF LARGE NUMBERS. Let B be a finite set, A a subset of B, and let a and b denote the numbers of elements in A and B. For any sequence Γ_k of k elements belonging to B, let $a(\Gamma_k)$ denote the relative frequency in Γ_k of elements belonging to A, that is to say, the number of those elements in Γ_k divided by k. For any two numbers, c and c′, call $F_k(c, c')$ the number of all sequences Γ_k for which

(14) $$c \leqq a(\Gamma_k) \leqq c',$$

so that, in particular,

(15) $$F_k(0, 1) = b^k \text{ for any } k = 1, 2, \cdots.$$

Then, for any pair of positive numbers x and y (no matter how small), there exists a number $N(x, y)$ such that

(16) $$k > N(x, y) \text{ implies } F_k\left(\frac{a}{b} - x, \frac{a}{b} + x\right) > (1 - y) \cdot b^k.$$

An example of such a number $N(x, y)$ is Cantelli's number

(17) $$\frac{2}{x^2} \log \frac{4}{x^2 y} + 2.$$

In words: *If* k *is sufficiently large, then among the* b^k *sequences of* k *elements of* B *those in which the relative frequency of the elements belonging to* A *differs from* a/b *by less than* x *have a relative frequency* $> 1 - y$.

To the chalk dots in physical geometry and the undefined elements called points in postulational geometry, there corresponds in analytic geometry the purely mathematical definition of a point as an ordered pair of numbers. Similarly, to the repeated trials and observed frequencies in statistics and the undefined elements called events and probability in the postulational probability theory, there correspond in the combinatoric probability theory the following purely mathematical definitions.

A *trial* is an element of a finite set B. A *success* is an element of a subset A of B. The *a priori probability* of success in a trial is the ratio of the numbers of elements in A to that of the elements in B; a sequence of independent trials is a sequence of elements of B. An *event* is a class of such sequences having the same length. The *probability* of an event is the ratio of the number of sequences in the class to the total number of sequences of the same length. Clearly, any two sequences of the same length are considered as equally probable.

That the first combinatoric theorem differs from the traditional law is clear not only from the absence in the former of references to reality but from the fact that the combinatoric theorem covers only cases with *rational a priori probability*, whereas in the traditional law the a priori probability might well be $1/\sqrt{2}$ or $1/e$. However, one can obtain a Second Combinatoric Theorem covering any (rational or irrational) a priori probability p. For this purpose a Poisson setup is needed; that is to say, a sequence of finite sets $B_1, B_2, \cdots$ and a sequence of respective subsets $A_1, A_2, \cdots$ must be given with a_n/b_n converging to p.

Clearly, the combinatoric method can be extended to the multinomial case, in which each set B_n has k^m mutually disjoint and jointly exhaustive subsets $A_n^1, \cdots, A_n^m$—the same m for any n. The method can further be extended to sampling theory, confidence limits and, to a certain extent, to continuous distributions.

REFERENCES

[1] A. Wald, *Sequential Analysis*, New York, John Wiley and Sons, 1947.

[2] Karl Menger, "The ideas of variable and function," *Proc. Nat. Acad. Sci.*, Vol. 39 (1953), pp. 956–961.

[3] ———, "On variables in mathematics and in natural sciences," *Brit. Jour. Phil. Sci.*, Vol. 5 (1954), pp. 134–142.

[4] ———, "Variables de diverses natures," *Bull. des Sci. Math.*, Vol. 78 (1954), pp. 229–234.

[5] ———, *Calculus. A Modern Approach*, Chicago, Illinois Institute of Technology Bookstore, 1953; final edition, Boston, Ginn and Co., 1955.

[6] J. F. Kenney, *Mathematics of Statistics*, Part I, 2d ed., New York, D. Van Nostrand Company, 1947.

[7] M. G. Kendall, *The Advanced Theory of Statistics*, Vol. I, 4th ed., London, Griffin, 1948.

[8] P. R. Halmos, *Measure Theory*, New York, D. Van Nostrand Company, 1950.

[9] R. Courant, *Differential and Integral Calculus*, Vol. I, New York, Interscience, 1951.