VOLUME XI No. 4 DECEMBER 1959

SYNTHESE

AN INTERNATIONAL QUARTERLY FOR THE LOGICAL AND THE PSYCHOLOGICAL STUDY OF THE FOUNDATIONS OF SCIENCE

REPRINT

D. REIDEL PUBLISHING COMPANY / DORDRECHT - HOLLAND

SYNTHESE

An International Quarterly for the Logical and the Psychological Study of the Foundations of Science

Comprising
Communications of the
INSTITUTE FOR THE UNITY OF SCIENCE
INTERNATIONAL SOCIETY FOR SIGNIFICS
INSTITUT FRANÇAIS DE SOCIOLOGIE
French and Dutch Sections of the
INTERNATIONAL ASSOCIATION FOR LOGIC AND THE PHILOSOPHY OF SCIENCE

SYNTHESE aims at publishing contributions to the study of the foundations and the methodology of science, the analysis of its linguistic aspects and to the study of its psychological and sociological implications.

EDITORIAL OFFICE: 29 Cornelis Krusemanstraat, Amsterdam-Z. Communications for the Editors, Manuscripts, typewritten in English or French, and Books for Review should be addressed to the Secretary of the General Editorial Committee, Postbox No. 7017, 40 Courbetstraat, Amsterdam-Z, Netherlands.

J. HEMELRIJK

DAVID VAN DANTZIG'S STATISTICAL WORK

1. HISTORICAL NOTES

Originally a pure mathematician, professor at the Technical University, Delft, David van Dantzig (1900–1959) started his work in the field of probability theory about 1940 with his first publication on this subject [1]. Afterwards, during the war, being dismissed by the Germans from his professorship, he continued to study probability and statistics. He at once understood the eminent importance of this branch of applied mathematics for other sciences, industry and society in general and his conscientious nature drove him to further study and to thinking about ways and means to propagate and develop research and application of mathematics in general and statistics in particular after the war.

Before the war Holland was decidedly an underdeveloped country as far as mathematical statistics was concerned. Good work had already been done by several pioneers in the application of statistics and at the Agricultural University of Wageningen, where courses in mathematical statistics were given by Prof. M. J. van Uven, but these activities were isolated and uncoordinated. Some more names and details can be found in [28].

What was really needed was a bridge between mathematicians on one hand and other scientists, engineers and technicians on the other hand. Also study and research in the domain of applied mathematics (including physical mathematics and statistics) would have to be stimulated and coordinated with similar activities in the field of pure mathematics and numerical methods.

All this and especially the possibility of consultation on a large scale could not be realized by the universities alone and this led to the idea of establishing a mathematical organization, harbouring all of the activities mentioned – and some more – in one institution. This idea, shared by Prof. J. G. van der Corput and Prof. J. F. Koksma reached its definite form in the *Mathematical Centre*, founded in Amsterdam in February 1946 by the triumvirate mentioned. The importance of this joint initiative can hardly be overestimated. The Mathematical Centre flourished and grew rapidly and is now a unique institution helping to give drive to the development and application of mathematics in wide fields of scientific research, technics and industry.

Having studied probability and statistics during the war, Van Dantzig was then – as he was when he died – our foremost statistician. He set himself the task to realise his ideas on consultation and research as head of the Department of Mathematical Statistics of the Mathematical Centre, a task which could exceedingly well be combined with his new professorship (in the 'Theory of Collective Phenomena', a chair created especially for him) at the municipal University of Amsterdam.

From then on he has been more than busy with courses and statistical consultation, working with pupils, visiting congresses, delivering lectures, working in committees and boards and, besides all these activities, producing papers and working out new theories up till the very day of his untimely death. His powers of thought and work were amazing and it was largely due to his stimulating activities that Holland reached an international level in statistics after a few years.

His theoretical work on probability and statistics also found international recognition. His personal relations with many outstanding mathematicians and statisticians were excellent and he was a well known visitor of and speaker at international congresses. In 1950 he lectured for half a year at Berkeley, U.S.A., as visiting professor of California University. He was elected member of the *International Statistical Institute*, fellow of the *Royal Statistical Society*, the *Institute of Mathematical Statistics* and the *American Statistical Association.* He also was a member of the Dutch Royal Academy of Sciences *(Koninklijke Nederlandse Akademie van Wetenschappen)* and one of the first members of the Dutch Statistical Society *(Vereniging voor Statistiek)*. His five lectures for the yearly congress of this society are rightly famous for their wit and their clarity of thought and expression.

During all these activities a strong nostalgia for pure mathematics never left him. '*Applied mathematics seems to be like wine: it becomes pure just in course of time*' [1]) is one of many of his expressions of this nostalgia; his (unfinished) endeavour to find a new basis for probability in topology (one of his 'old loves') is another. He did not succeed, however, in

[1]) '*The function of Mathematics in Modern Society and its consequences for the teaching of Mathematics*', Euclides *31* (1955) 88-102.

disentangling himself from his applied work. Some of the subjects on which he worked were of too great importance for that. The most important one was the statistical and hydrodynamical research concerning the high water levels at the Dutch coast and the movements of the North Sea, on which he worked for years after the disastrous flood of 1953. This subject is too large to be treated extensively here. A final version of the report of the Departments of Mathematical Statistics and Applied Mathematics of the Mathematical Centre, written under his supervision and for a large part by himself for the Delta Commission of the Dutch Government, was lying on his desk on the day of his death. For weeks he had been putting the final touches to this large document, which will appear as a book – one volume of a very voluminous report of the Delta Commission to the Government – and which will constitute a monument for his tireless efforts to further the interests of mathematics and its application to subjects important for society.

2. HIS VIEWS ON FOUNDATIONS OF PROBABILITY AND STATISTICS

Van Dantzig's extensive interest in and knowledge of significs destined him to take part in the discussions on the foundations of his new domain of research. It is well known that several very divergent views exist on the foundations of probability. Although the theory of probability may be seen – following Kolmogorov – as an axiomatic mathematical theory, part of measure theory, the application in the form of statistics and the practical interpretation of a numerical probability have given and still give rise to controversies between scientists working in this field. Two main branches of interpretation may be distinguished: the objective (frequentist) and the subjective (or personal) interpretation.

The frequentists see a probability as an approximation of a frequency in the long run and this was also van Dantzig's standpoint. He complemented this interpretation – as others have done – by the observation that the practical behaviour of mankind indicates that possible events with a very small probability are not expected to occur in isolated observations and that these probabilities are therefore treated in the same way as zero probability. His further analysis of this interpretation, of the rôle of the mathematical model in a scientific investigation and of the method of its choice, its 'switching on' and 'switching off' constitute a real contribution to a better comprehension of the mechanism of the statistical method, leading to a sharper and more critical distinction of its different phases and improving the exactness of its applications.

Although not quite satisfied with the frequency interpretation he could never agree with the ideas of the subjectivists, who interpret probability as a degree of belief, a degree of confidence or something like this. His criticism of these ideas was sharp and to the point. In one of his papers on this subject [31], '*Statistical priesthood (Savage on personal probabilities)*', he launched a heavy attack on L. J. Savage's book '*The foundations of statistics*', 1954. The following is a quotation from this paper which neatly summarizes his views:

(1) '*Statistical work has value only insofar as its results are independent of the preferences of the individual statistician who performs it. Although such an independence in any absolute sense cannot be reached, it can be obtained to a practically sufficient degree, which is not essentially less than one obtainable in other sciences.*

(2) *Strictly speaking statistics needs as a mathematical tool no calculus of probabilities, but only a calculus of (finite) frequency quotients. The concepts of probability and of infinity are introduced for mathematical convenience only.*

(3) *Statistics uses the empirical hypothesis that apparatus ('lotteries') exist, admitting random choices of one among any given number of elements. Such apparatus do not exist in absolute perfection and their degree of perfection can only be defined after development of their theory. Their rôle is analogous to that of rigid bodies in euclidean geometry and of perfect clocks in dynamics. Empirical interpretation of probability statements is only possible with reference to such random apparatus or to natural phenomena empirically found to behave statistically sufficiently like these.*

(4) *Because of imperfection of random apparatus and of simplifying mathematical assumptions probability statements of very great precision have no empirical correlate. In particular the distinction between very small probabilities and zero has none. In accordance with this, actual human behaviour is only understandable on the assumption that possible events having theoretically extremely small probabilities are actually neglected.*

(5) *Subjective expectations, valuations and preferences and their changes from person to person or in the course of time can and should be investigated by means of 'objective' statistical methods. Trying to use them as a*

basis of statistics is like trying to gauge a fever thermometer by means of the patient's shivers.'

This statement of his views contains in a nutshell the basis of his attack on subjectivistic theory, which is worked out further in the rest of the paper.

A second paper called '*Statistical priesthood II (Sir Ronald on fiducial inference)*' [33] is a critique of R. A. Fisher's book '*Statistical methods and scientific inference*', 1956. This paper also combines sharp criticism with excellent wit and it contains an analysis of Fisher's method of fiducial inference. Although searching for positive points in this method, Van Dantzig finally finds that he must reject it. The method of his analysis may be illustrated by means of a very simple example, mentioned in the first part of the paper.

Let x be a random variable, normally distributed with unknown mean μ. Then everybody agrees that

$$P(x < \mu) = \tfrac{1}{2}. \tag{1}$$

Fisher now writes

$$P(\mu > x) = \tfrac{1}{2}, \tag{2}$$

which is equivalent to (1) and if an observation of x yields the value 1.37 he substitutes this value into (2), getting

$$P(\mu > 1.37) = \tfrac{1}{2}. \tag{3}$$

This statement is – for a frequentist – meaningless, μ not being a random variable but an unknown constant. If $\mu > 1.37$ is in fact true, then the value of $P(\mu > 1.37)$ is 1 and if $\mu < 1.37$ then its value is 0. Now Fisher himself adheres to the frequency interpretation and he does not consider (3) as a probability statement but as a new kind of statement, derived from a 'fiducial distribution' attributed to μ; this fiducial distribution is also normal, with the same variance as the probability distribution of x and with x itself as mean. Considering (2) as a fiducial statement derived from this fiducial distribution and then substituting $x = 1.37$ (3) is obtained.

It is of course a quite legitimate procedure to introduce a new notion called 'fiducial distribution', with substitution of a number for x admitted.

It is, on the other hand, very confusing to use the same symbol P for this distribution. And furthermore, increasing the confusion to the point of incorrectness, Fisher insists that (2) allows of a frequency interpretation. This is true as long as P means 'probability' and as long as no substitution for x is allowed. Van Dantzig makes this clear by rewriting (2) in the form

$$\text{(2a)} \qquad \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\mu} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx = \tfrac{1}{2}.$$

It is clear that in this formula, which is equivalent to (2) and the datum that x is normally distributed, no substitution of 1.37 (or any other number) for x can be allowed. Therefore (2), considered as a fiducial statement with substitution admitted, cannot be equivalent to (2a). But then the meaning of the fiducial statement is totally in the dark.

Van Dantzig himself never wrote x for a random variable, but he indicated the random character of a variable by *underlining* its symbol. The distinction between random variables and algebraic ones – neglected by Fisher – was of course made consistently at the time by several authors, but they used different symbolisms, e.g. capitals for random variables and small symbols for algebraic ones or greek versus roman type. The method of underlining is, however, a very practical one which in our opinion is to be preferred. In this notation (1) becomes

$$\text{(1')} \qquad P(\underline{x} < \mu) = \tfrac{1}{2}$$

and (2), in its probability sense, becomes

$$\text{(2')} \qquad P(\mu > \underline{x}) = \tfrac{1}{2},$$

clearly equivalent to (1′) .But in its fiducial sense it seems to be something like

$$\text{(2'')} \qquad P(\underline{\mu} > x) = \tfrac{1}{2},$$

at least if one insists on a frequency interpretation with substitution for x admitted. For in this form x is an algebraic variable and substitution *is* possible. However, (2″) does not in any way follow from (2′) by means of the axioms of probability theory. It should therefore be based on a definition of $\underline{\mu}$, which Fisher does not give explicitly. Defining

$$(3') \qquad \underline{\mu} \overset{def}{=} \mu - \underline{x} + x, \qquad [1]$$

where μ is the unknown value of the mean of $\underline{x}$ and x the value assumed by $\underline{x}$ in the observation, (2″) can indeed be proved.
It is not clear whether Fisher would agree with this view on his method. The further analysis, along these lines, of more complicated cases led to rather disappointing conclusions, the most important one being that the Behrens-Fisher two sample test cannot be justified in this way and is incorrect from the frequentist point of view.
This analysis clearly shows the importance and the power of a systematic and clear notation.

3. HIS WORK IN STATISTICAL CONSULTATION

During the first years of the Mathematical Centre Van Dantzig did most of the consultational work himself. Most instructing were his talks with research workers in other fields of science, who were always compelled in the course of the interview to sharply define the means and the aim of their investigation. He always started the discussion by asking for a formulation of the aim of the experiments and the discussion of this point alone was always of great value to the experimenters already and often led to a change in the experimental design, at least if the experiments had not already been performed. This made him an ardent propagandist for consultation before the actual start of the experiments. He also insisted on building up an adequate mathematical model together with the consultees and, if necessary, on an adaptation of the experimental design in order to bring it closer to the mathematical model. Another point, which in experiments is too often overlooked, is always to investigate the accuracy (or lack of it) of the observations. Then the statistical methods to be used in the analysis had to be chosen beforehand. Time and money should be allotted to this phase of the project, a phase which he emphatically included in the investigation as a whole. The analysis

[1]) $\overset{def}{=}$ indicates that the left hand member is defined by the right hand one.

should not be considered a trifling work to be executed by statistical sourcerers after the experiment has been completed.

As for the conclusions to be drawn from the analysis, these were obtained by 'switching off' the mathematical model and they were painstakingly formulated in an exact statistical language. Their final practical or scientific interpretation was then discussed with the consultees and help was given to 'translate' them in the language pertinent to the particular domain of the experiment, but the responsibility of these final interpretations always remained with the experimenters themselves.

In this way he not only succeeded in forming a group of statistical pupils who were well schooled in statistical consultation but he also taught many scientists in other fields to think more clearly and to realise the importance of paying due attention to the theoretical preparations and the possibility of well founded statistical analysis which are necessary for experiments on a high scientific level.

4. HIS THEORY OF COLLECTIVE MARKS

Space does not allow us to give a complete survey of Van Dantzig's research work in probability and statistics. Interested readers should consult his own papers, listed as references at the end of this paper. We have selected one topic, the theory of collective marks, which was his pet theoretical subject. This theory in itself is already too large for complete description [4] [5] [22] [27] [34] [37]. Only the principles are stated and two examples, not extensively published before, are given.

The basic idea of collective marks is extremely simple; it can be demonstrated with the following example.

On December 31st, 1939, the number of registrated inhabitants of Amsterdam was 757386; 365512 were men and 391874 women. A very simple way to summarize these data is

$$757386\,A = 365512\,M + 391874\,W, \tag{4}$$

where $A \equiv$ 'inhabitant of Amsterdam', $M \equiv$ 'man' and $W \equiv$ 'woman'. Regarding (4) as an equation it may also be written

$$A = 0.4826\,M + 0.5174\,W; \tag{5}$$

the coefficients are now relative frequencies.

Generally, let Γ be an arbitrary finite collection with n elements and let $A_1,\ldots,A_k$ be a set of properties such that each element of Γ has exactly one of these properties. Let the frequencies of $A_1,\ldots,A_k$ be $n_1,\ldots,n_k$ $(\sum n_i = n)$ and let

(6) $$f_i = n_i/n,$$

then

(7) $$C \overset{def}{=} \sum f_i A_i$$

is, by definition, the *collective mark* of Γ with respect to the 'category' of properties (or 'marks') $A_1\ldots,A_k$.

Regarding $A_i\,(i = 1,\ldots k,)$ in (7) as abstract symbols C becomes an abstract symbol too.

Several kinds of substitutions may now be considered. If e.g. E is an arbitrary property on Γ, with relative frequency a_i on the subset of Γ consisting of all elements having property A_i $(i = 1,\ldots,k)$, then

(8) $$c = \sum f_i a_i$$

is the relative frequency of E on Γ. Taking e.g. for E the property of having red hair and for m and w respectively the fractions of red-haired men and women in Amsterdam, (5) gives

$$a = 0.4826\,m + 0.5174\,w,$$

the fraction of red-haired inhabitants of Amsterdam at the end of 1939.

This notion of a collective mark applies analogously to a *probability distribution*. We consider only the discrete numerical case of a random variable $\underline{x}$ assuming values $x_1,x_2,\ldots$ with probabilities $p_1,p_2,\ldots$ $(\sum p_i = 1)$. The extension to other cases is evident. The collective mark of the probability distribution is then defined by

(9) $$C \overset{def}{=} \sum p_i A_i,$$

where $A_i\,(i = 1,2,\ldots)$ and C are abstract symbols.

If now E is an arbitrary possible event and $P(\bar{E}|x_i)$ is the conditional probability, given $\underline{x} = x_i$, that E does *not* occur, then substituting $P(\bar{E}|x_i)$ for A_i gives

(10) $$P(\bar{E}) = \sum p_i P(\bar{E}|x_i),$$

the unconditional probability of $\bar{E}$. This is the *fundamental property* of a collective mark.
A collective mark is a *functional:* substituting a function $A(x_i)$ for A_i, C becomes a number. Functionals like this have been introduced by several authors, but Van Dantzig's method contains new elements which enlarge its scope beyond previous methods.
The most interesting substitution for A_i is A^i, giving

(11) $$C(A) \overset{def}{=} \sum p_i A^i,$$

a polynomial in the abstract symbol A. The probabilities p_i are still uniquely determined by means of the formal differentiation

(12) $$p_i = \frac{1}{i\,!}\left[\frac{d^i C(A)}{dA^i}\right] A = 0.$$

Taking for A a complex number, $C(A)$ is the *generating function;* with $A_i = e^{tx_i}$ the *moment-generating function* is obtained and $A_j = e^{itx_j}$ with real t gives the *characteristic function*. All these cases have been considered extensively by Laplace, De Moivre, Cauchy, Lévy a.o.
Van Dantzig gives a new turn to his generalisation of these functions by introducing an imaginary event E ('a catastrophe; e.g. the ignition of a red light') which may or may not take place together with a realisation of the random variable $\underline{x}$. With every possible value x_i of $\underline{x}$ is associated a probability

(13) $$P(\bar{E}|x_i)$$

that E will *not* occur. According to (10) substitution of this probability for A_i transforms C into $P(\bar{E})$, the probability that a realisation of $\underline{x}$ will not result in the catastrophe E. Sometimes an adroit choice of the chance mechanism ('lottery') represented by (13) makes it possible to compute $P(\bar{E})$ directly; and to derive the collective mark, the characteristic function etc. at once from $P(\bar{E})$. This leads to beautiful short cut derivations illustrated in the following sections.
For the further elaboration of the theory itself, e.g. the generalisation to more dimensions, we refer to the original papers.

5. DERIVATION OF THE BINOMIAL DISTRIBUTION

Consider n independent experiments each having probability p of a success and $q = 1-p$ of a failure. Let $\underline{x}$ be the total number of successes, then the well known probability distribution of $\underline{x}$ may be derived as follows. The collective mark (11) of $\underline{x}$ is given by

$$C(A) = \sum_{x=0}^{n} p_x A^x \tag{14}$$

and for (13) we now use a lottery L from which x tickets are drawn independently and with replacement if $\underline{x}$ assumes the value x. The lottery L is such that in each drawing the probability that the ticket indicates E (the catastrophe) is equal to $1-A$ $(0 < A < 1)$. Thus $P(\bar{E}|x) = A^x$ and according to (10) and (14)

$$C(A) = P(\bar{E}). \tag{15}$$

The x drawings from L may be executed one after each success. The probability that one experiment, with a drawing from L if the experiment gives a success, does not lead to E is then:

$$pA + q$$

and for the series of n experiments this gives

$$C(A) = P(\bar{E}) = (pA + q)^n. \tag{16}$$

Expansion of the right hand member according to Newton's binomium leads together with (14) to

$$p_x = P(\underline{x} = x) = \binom{n}{x} p^x q^{n-x}. \tag{17}$$

6. APPLICATION TO RANK CORRELATION [34]

Let

$$\left.\begin{array}{ll} \underline{x}_{1,1}, \ldots, \underline{x}_{1,n_1} & S_1 \\ \underline{x}_{2,1}, \ldots, \underline{x}_{2,n_2} & S_2 \\ \cdots\cdots\cdots\cdots & \cdots \\ \underline{x}_{m,1}, \ldots, \underline{x}_{m,n_m} & S_m \end{array}\right\} \sum_{\lambda} n_\lambda = n \tag{18}$$

be m independent random samples $S_1,\ldots,S_m$ from continuous probability distributions, then the following statistic $\underline{T}$ is fundamental in rank correlation theory:

(19) $$\underline{T} \overset{def}{=} \textit{the number of pairs } (\underline{x}_{\lambda,i}, \underline{x}_{\mu,j}) \textit{ with}$$
$$\lambda < \mu \textit{ and } \underline{x}_{\lambda,i} > \underline{x}_{\mu,j}$$
$$(\lambda, \mu = 1,\ldots,m;\ i = 1,\ldots,n_\lambda;\ j = 1,\ldots,n_\mu).$$

(In words: the number of times a member of an 'earlier' sample exceeds a member of a 'later' sample.)
It is clear that $\underline{T}$ can assume integral values between 0 and

$$N = \sum_{\lambda<\mu} n_\lambda n_\mu$$

and that small values tend to indicate the presence of an increasing trend of the samples in the given order and large values of $\underline{T}$ a decreasing trend. For purposes of statistical testing it is important to know the probability distribution of $\underline{T}$ under the hypothesis H_0 that all samples come from the same population, i.e. that all $\underline{x}_{\lambda,i}$ $(\lambda = 1,\ldots,m;\ i = 1,\ldots,n_\lambda)$ are independent with the same probability distribution. (Van Dantzig expressed this by: all $\underline{x}_{\lambda,i}$ are '*isomorous*'.)
This probability distribution can be derived by means of recurrence methods and its characteristic function can be found and proved by complete induction. Far more simple and elegant, however, is the derivation of the characteristic function by means of the method of collective marks.
Let h marbles, numbered $1,\ldots,h$, roll along the real (x–) axis starting from $+\infty$ and coming to rest independently according to the probability distribution of $\underline{x}$. Each endpoint is thus a realisation of $\underline{x}$ and these are independent. The probability that two marbles come to rest at the same point is zero. The result of this experiment is a permutation of the numbers $1,\ldots,h$ as read from the marbles along the real axis. It is easy to see that all $h!$ permutations have the same probability $(h!)^{-1}$ and that for any given subgroup of marbles all permutations of their numbers are equally probable and stochastically independent from the permutations within other subgroups not containing any marbles of the former.

Also all possible situations of two subgroups (without common elements) with respect to each other are equally probable.
The marbles are rolled one at a time in the order of their numbers, no two being in motion at the same time. When a rolling marble on its way passes another one (which has been rolled already) a ticket is drawn from a lottery L with a probability $1-A$ of a catastrophe. The ticket is then put back for the following independent random drawing. Consider now the conduct of the k^{th} marble. The $k!$ permutations of the first k marbles all have equal probability, thus the probability that the k^{th} one passes $0,1,\ldots k-1$ marbles is k^{-1} for each of these possibilities, independently of the history of the first $k-1$ marbles. This means that the probability that the k^{th} marble does not cause a catastrophe is

$$(20)\quad k^{-1}(1 + A + A^2 + \ldots + A^{k-1}) = (1 - A^k)/k(1 - A).$$

This holds for every k $(= 1,\ldots,h)$ independently of past history and thus we have for all h marbles

$$(21)\quad P_h(\bar{E}) = (1 - \mathrm{A})^{-h} \prod_{k=1}^{h} (1 - A^k)/k = \{h!(1 - A)^h\}^{-1} \prod_{k=1}^{h} (1 - A^k).$$

But, as in (15), this is also the collective mark of the total number of passings during the whole process:

$$(22)\qquad P_h(\bar{E}) = C_h(A).$$

Now rolling one marble for every $\underline{x}_{\lambda,i}$ of (18) starting with n_1 marbles for S_1, then n_2 for S_2, etc., (21) and (22) hold for $h = n = \sum n_\lambda$ with regard to the total number of passings. These may, however, be split up into two classes:
1. 'Same sample passings', occurring when a marble passes another (earlier) marble from the same sample;
2. 'Different sample passings', occurring when a marble passes another one from an earlier sample.
The statistic $\underline{T}$ is exactly the total number of different sample passings.
For the same sample passings of S_λ we have

$$(23)\qquad P_{\lambda,same}(\bar{E}) = C_{n_\lambda}(A) \quad (\lambda = 1,\ldots,m)$$

and, the numbers of same sample passings being stochastically independent for different samples, the probability that the same sample passings all together do not cause the catastrophe E to happen is

$$(24) \qquad P_{same}(\bar{E}) = \prod_{\lambda=1}^{m} C_{n_\lambda}(A).$$

The total number of passings is equal to the sum of the numbers of passings of both kinds and these numbers are again independent. Thus we have

$$(25) \qquad P_{total}(\bar{E}) = P_{same}(\bar{E}) \,.\, P_{different}(\bar{E}),$$

or according to (22) and (24)

$$(26) \qquad C_n(A) = \prod_{\lambda=1}^{m} C_{n_\lambda}(A) .\, C_T(A),$$

where

$$(27) \qquad C_T(A) = \sum_{T=0}^{N} p_T \,.\, A^T = P_{different}(\bar{E})$$

is the collective mark of $\underline{T}$. Substituting (21) the result is

$$(28) \qquad \begin{cases} C_T(A) = \sum_{T=0}^{N} p_T \,.\, A^T = C_n(A) \Big/ \prod_{\lambda=1}^{m} C_{n_\lambda}(A) = \\ \qquad = \dfrac{n_1! \ldots n_m!}{n!} \dfrac{\prod_{k=1}^{n} (1 - A^k)}{\prod_{\lambda=1}^{m} \prod_{j=1}^{n_\lambda} (1 - A^j)} . \end{cases}$$

Substituting $A = e^{it}$ the characteristic function is obtained. For $n_\lambda = 1$ ($\lambda = 1, \ldots, m$), i.e. for samples of 1, $2\underline{T} - m$ is equal to Kendall's ranking statistic $\underline{S}$; the characteristic function then reduces to

$$(29) \qquad \mathscr{E}\, e^{it\underline{S}} = \prod_{k=1}^{m} \frac{\sin kt}{k \sin t},$$

a formula earlier derived by Kendall by means of recurrence relations. The method of collective marks has also been successfully applied to more complicated problems by Van Dantzig himself and by some of his pupils. It certainly deserves further consideration and development.

7. BIBLIOGRAPHY OF VAN DANTZIG'S STATISTICAL PAPERS

For a complete bibliography of all his mathematical papers see: Statistica Neerlandica *13,* (1959), 415–432.

[1] *Mathematical and empirical foundations of probability theory,* (in Dutch), Ned. Tijdschr. voor Nat. *8* (1941) 70–93.

[2] *Correspondence with B. de Finetti (Punti di Vista),* Statistica 1941, 229–242 (the Italian journal 'Statistica').

[3] *Probability theory,* (in Dutch; mimeographed course), Mathematisch Centrum, Amsterdam, 1947.

[4] *Mathematical Statistics,* (in Dutch; mimeographed course), Mathematisch Centrum, Amsterdam, 1947–48.

[5] *Sur la méthode des fonctions génératrices,* Le Calcul des Probabilités, Coll. Int. du Centre Nat. de la Rech. Sc. *13* (1949) 29–46.

[6] *Blaise Pascal and the meaning of mathematical thinking for the study of human society,* (in Dutch; inaugural address), Euclides *25* (1948) 203–232.

[7] *Sur l'analyse logique des relations entre le Calcul des Probabilités et ses applications,* Congrès Intern. de Phil. des Sciences, 1–8 Juillet 1949, Vol. 4 'Calcul des Probabilités'; Actualités Scientif. et Industr. Nr. 1146.

[8] *Proposals for the standardization of symbols in mathematical statistics and biometrics,* Statistica *4* (1950) 80–86.

[9] *Laplace, probabiliste et statisticien et ses précurseurs,* Archives Intern. d'Histoire des Sciences *8,* no. 30 (1955) 27–37.

[10] *Une nouvelle généralisation de l'inégalité de Bienaymé,* Ann. de l'Inst. Henri Poincaré *12* (1951) 31–43.

[11] *Some historical relations between mathematical and descriptive statistics,* (in Dutch), Statistica *4* (1950) 233–248.

[12] *On the consistency and the power of Wilcoxon's two-sample test,* Proc. Kon. Ned. Ak. v. Wet. A*54*; Indag. Math. *13* (1951) 1–8.

[13] *Les problèmes que pose l'application du Calcul des Probabilités,* Collection de Logique mathématique, serie B, *1* (1952) 53–65.

[14] *Time-discrete stochastic processes in arbitrary sets, with applications to processes with absorbing regions and to the problem of loops in Markov chains,* mimeographed, Mathematisch Centrum, Amsterdam, 1952.

[15] *Carnap's foundation of probability theory,* Synthese *8* (1953) 459–470.

[16] *Mathematical consultation for medical, biological and other research,* (in Dutch), Proc. Kon. Ned. Ak. v. Wet. *61*; Indag. Math. *14* (1952) 13–18.
[17] *Nature as opponent,* (in Dutch), Statistica *5* (1951), 149–159.
[18] *Utilité d'une distribution de probabilités, ou Distribution des probabilités d'une utilité?,* Coll. d'Econométrie du Centre Nat. de Recherche Sc., Paris, 1952.
[19] *Another form of the weak law of large numbers,* Nieuw Archief v. Wiskunde (3), *1* (1953) 129–145.
[20] *Prediction and prophecy,* (in Dutch), Statistica *6* (1952) 195–204.
[21] *Statistical methods based on few assumptions,* (with J. Hemelrijk), Bull. of the I.S.I. *34,* II (1954) 3–31.
[22] *On arbitrary hereditary time-discrete stochastic processes, considered as stationary Markov chains, and the corresponding general form of Wald's fundamental identity,* (with C. Scheffer), Proc. Kon. Ned. Ak. v. Wet. A *57*; Indag. Math. *16* (1954) 377–388.
[23] *The responsabilities of the statistician,* (in Dutch), Statistica *7* (1954), 199–208.
[24] *Mathematical consultation in practice,* (in Dutch), Euclides *30* (1954) 53–67.
[25] *Mathematical problems raised by the flood disaster 1953,* Proc. of the Int. Math. Congress (Amsterdam 1954), I, 218–239.
[26] *Sur les ensembles de confiance généraux et les méthodes dites non paramétriques,* Coll. sur l'Analyse Statistique, Bruxelles 1954, 73–91 (édition du Centre Belge de Recherches Mathématiques).
[27] *Chaînes de Markov dans les ensembles abstraits et applications aux processus avec régions absorbantes et au problème des boucles,* Ann. de l'Inst. Henri Poincaré XIV, fasc. III (1955) 145–199.
[28] *Ten years mathematical statistics,* (in Dutch), Statistica *9* (1955) 233–242.
[29] *Econometric decision problems for flood prevention,* Econometrica *24* (1956) 276–287.
[30] *A course in Markov chains,* (in Dutch; edited by G. Zoutendijk and J. Hirsch), mimeographed, Mathematisch Centrum, Amsterdam, 1956.
[31] *Statistical Priesthood (Savage on personal probabilities),* Statistica Neerlandica *11* (1957) 1–16.

[32] *From 'Rekeningh in Spelen van Geluck' to decision theory,* (in Dutch), Yearbook II of the University of Amsterdam (1956–'57) 39–50.
[33] *Statistical Priesthood II (Sir Ronald on Scientific Inference),* Statistica Neerlandica *11* (1957) 185–200.
[34] *Les fonctions génératrices liées à quelques tests non-paramétriques,* mimeographed, Mathematisch Centrum, Amsterdam, 1957.
[35] *Extrapolation of the frequency-line of high water levels at Hoek van Holland by means of selected storms,* (in Dutch; with J. Hemelrijk), Report for the Delta Commission, 1959.
[36] *The econometric decision problem concerning the prevention of floods in Holland,* (in Dutch), Report for the Delta Commission, 1959.
[37] *Itérations markoviennes dans les ensembles abstraits,* (with G. Zoutendijk), J. de Math. pure et appl. *38* (1959) 183–200.
[38] *Sur quelques questions de la théorie mathématique du choix pondéré,* Coll. 'Sur la Décision' Paris, mai 1959, to appear.
[39] *Prediction and Prophecy,* to appear.

Technical University, Delft; Mathematical Centre, Amsterdam.

SYNTHESE

An International Quarterly for the Logical and the Psychological Study of the Foundations of Science

Subscriptions: The annual subscription of the journal, which is published quarterly, is $ 7.50, 56 s. or Hfl 28.—. Orders for service of less than a full year will be charged at the singlecopy rate of $ 2.50, 20 s. or Hfl 10.—. Reserve-stock permitting, the publishers will replace free of charge numbers lost in transit.

Reprints: Each contributor will receive 25 reprints.

SYNTHESE LIBRARY

A series of monographs on the recent development of

SYMBOLIC LOGIC, SIGNIFICS, SOCIOLOGY OF LANGUAGE, SOCIOLOGY OF SCIENCE AND OF KNOWLEDGE, STATISTICS OF LANGUAGE

and related fields

J. M. BOCHENSKI

A PRECIS OF MATHEMATICAL LOGIC

X + 100 pages, Hfl 13.75 ($ 3.75, 27½ s.)

PIERRE GUIRAUD

PROBLÈMES ET MÉTHODES DE LA STATISTIQUE LINGUISTIQUE

VIII + 146 pages, Hfl 13.75 ($ 3.75, 27½ s.)

The following volumes are in active preparation:

E. W. BETH

MATHEMATICAL THOUGHT

An Introduction to the Philosophy of Mathematics

E. W. BETH

FORMAL METHODS

An Introduction to Symbolic Logic

SELECTIONS FROM THE WORK OF GERRIT MANNOURY

D. REIDEL PUBLISHING COMPANY

DORDRECHT - HOLLAND