Foundations of a New System of Probability Theory

Jürgen Humburg

The present work represents a summary of my book 'Grundzüge zu einem neuen Aufbau der Wahrscheinlichkeitstheorie' [5]. For this reason, I have frequently dispensed with providing proof and in this connection refer the interested reader to my book.

ABSTRACT. The aim of my book is to explain the content of the different notions of probability.

Based on a concept of *logical probability*, which is modified as compared with Carnap, we succeed by means of the mathematical results of de Finetti in defining the concept of *statistical probability*.

The starting point is the fundamental concept that certain phenomena are of the same kind, that certain occurrences can be repeated, that certain experiments are 'identical'. We introduce for this idea the notion: *concept K of similarity*. From concept K of similarity we derive 'logically' some probability-theoretic conclusions:

If the events $E(\lambda)$ are similar – of the same kind – on the basis of such a concept K, it holds good that intersections of n of these events are equiprobable on the basis of K; in formulae:

$E(\lambda_1) \dots E(\lambda_n) \sim_K E(\lambda'_1) \dots E(\lambda'_n)$, $\lambda_i \neq \lambda_j$, $\lambda'_i \neq \lambda'_j$ for $i \neq j$.

On the basis of some further axioms a partial *comparative probability structure* results from K, which forms the starting point of our further investigations and which we call *logical probability on the basis of K*.

We investigate a metrisation of this partial comparative structure, i.e. normed σ-additive functions m_K, which are compatible with this structure; we call these functions m_K *measure-functions* in relation to K.

The measure-functions may be interpreted as subjective probabilities of individuals, who accept the concept K.

Now it holds good: For each measure-function there exists with measure one the limit of relative frequencies in a sequence of the $E(\lambda)$.

In such an event, where all measure-functions coincide, we speak of a *quantitative logical probability*, which is the common measure of this event. In formulae we have:

$$l_K(h_n \to \lim h_n) = 1,$$

in words: There is the quantitative logical probability one that the limit of the relative frequencies exists. Another way of saying this is that the event $\Omega^* := (h_n \to \lim h_n)$ is a *maximal element* in the comparative structure resulting from K.

Therefore we are entitled to introduce this limit and call it *statistical probability* P.

With the aid of the measure-functions it is possible to calculate the velocity of this convergence. The analog of the Bernoulli inequation holds true:

$$m_K(|h_n - \mathrm{P}| \leqslant \epsilon) \geqslant 1 - 1/4n\epsilon^2.$$

It is further possible in the work to obtain relationships for the concept of statistical independence which are expressed in terms of the comparative probability.

The theory has a special significance for quantum mechanics: The similarity of the phenomena in the domain of quantum mechanics explains the statistical behaviour of the phenomena.

The usual mathematical statistics are explained in my book. But it seems more expedient on the basis of this new theory to use besides the notion of statistical probability also the notion of logical probability; the notion of subjective probability has only a heuristic function in my system.

The following dualism is to be noted: The statistical behaviour of similar phenomena may be described on the one hand according to the model of the classical probability theory by means of a figure called statistical probability, on the other hand we may express all formulae by means of a function, called statistical probability function. This function is defined as the limit of the relative frequencies depending on the respective state ω of the universe. The statistical probability function is the primary notion, the notion of statistical probability is derived from it; it is defined as the value of the statistical probability function for the true unknown state $\hat{\omega}$ of the universe.

As far as the Hume problem, the problem of inductive inference, is concerned, the book seems to give an example of how to solve it.

The developed notions such as concept, measure-function, logical probability, etc. seem to be important beyond the concept of similarity.

SYMBOLS

$\emptyset$ denotes the empty set;

Ω denotes the starting set taken as a basis, ω the respective variable for the elements of Ω;

$A \dot{+} B$ denotes the union of the sets A, B;

$A \,.\, B$ denotes the intersection of A, B;

$\bar{A}$ denotes the complement of A with respect to the starting set Ω;

$\Sigma^{\cdot} A_i$ denotes the union of A_i;

$\Pi^{\cdot} A_i$ denotes the intersection of A_i;

Topoi 5 (1986), 39–50.

$A + B := A \dot{+} B$, if $A . B = \emptyset$,

$\Sigma A_i := \Sigma^{\cdot} A_i$, if $A_i . A_j = \emptyset$ for all i, j with $i \neq j$.

Let $\mathscr{G}$ be a subset of the power set $\mathscr{P}(\Omega)$ of Ω; then $^B\mathscr{G}$ denotes the Borellian extension of $\mathscr{G}$, i.e. the smallest σ-field which contains $\mathscr{G}$.

The measurable space taken generally as a basis is designated by $(\Omega, \mathscr{F})$. We write a $\mathscr{F}$-measurable function (random variable) in the following manner, $f(\omega), h(\omega), p(\omega)$, etc.

To the actual true normally unknown element $\hat{\omega}$ of Ω we introduce for such a function the following symbol (bold face): $\mathbf{f} := f(\hat{\omega})$.

1. The concept of logical probability

(1.1.) We start with a measurable space $(\Omega, \mathscr{F})$ of events.

(1.2.) *The simple logical probability*

From the logical structure of certain events A, B the following expression results:

$A \lesssim B$; in words: for *logical reasons* A is *at most as probable* as B.

This expression can be clarified as follows: a bet on A is for logical reasons at most as good as a bet on B; the bet on A is to have the same stakes as the bet on B; for a more exact explanation of the concept of a bet see (1.6.).

Example. $A \lesssim A \dot{+} B$, $A . B \lesssim A$.

These examples are a special case of the following axiom:

AXIOM 0. $(\vdash A \subset B) \Rightarrow A \lesssim B$.

The substantiation of this axiom is very simple: If we can deduce from the occurrence of A the occurrence of B if we bet on A the analogous bet on B is also won. Consequently, we shall not prefer the bet on A to the bet on B.

We do not intend investigating here whether further axioms exist for the above notion of logical probability but turn to the term developed below of the logical probability on the basis of a concept.

(1.3.) *The logical probability on the basis of a concept*

By *concept* K we understand theoretical ideas T and information I which we combine to $K = T \wedge I$.

As you will see the following proposition results from the logical structure of K and of certain events A, B:

$A \lesssim_K B$; in words: on the basis of K B is at least as probable as A.

We thus take the logical probability '$\lesssim_K$' to be a comparative evaluation of certain events on the basis of a concept K; the expression $A \lesssim_K B$ results solely from the logical structure of K, A, B in accordance with the axioms to be formulated.

Firstly, the following extension of axiom 0 may be formulated:

AXIOM 1. If K implies logically that A is contained in B, in formulae $\vdash K \rightarrow A \subset B$, then $A \lesssim_K B$ holds true.

The most essential axiom in this connection is obtained with a concept K, on the basis of which certain events $E(\lambda), \lambda \in \Lambda$, are similar – of the same kind. This we explain as follows: The characteristics by which the events differ are without influence on the occurrence of these events. A more precise definition of the notion similar is obtained by formalizing K and the $E(\lambda)$ in formal language: The irrelevance of the different characterisations may be reflected in the formal structure of K and $E(\lambda)$. Firstly a

DEFINITION. $A \sim_K B :\Leftrightarrow A \lesssim_K B$ and $B \lesssim_K A$; in words: A, B are equiprobable on the basis of K.

AXIOM 2. If the events $E(\lambda)$ are similar on the basis of K we have $E(\lambda_1) \ldots E(\lambda_n) \sim_K E(\lambda_1') \ldots E(\lambda_n')$, $\lambda_i \neq \lambda_j$, $\lambda_i' \neq \lambda_j'$ for $i \neq j$.

Such a partial comparative structure resulting from a concept K is called *logical probability on the basis of K*.

An example of the concept of similar events is a sequence of polarized photons which hit a filter; $E(i) :=$ the ith photon passes through the filter. On the basis of the ideas T of quantum mechanics the chronological order of the photons is of no importance regarding the passage through the filter.

Another example is the repeated throw of a die, given the additional information that the die may not change in the course of the repeated throws. The theoretical ideas T in this case are the ideas of Newton's mechanics. An example for a non-similar ensemble is the throw of a die, given the additional information that the die contains a clock, which changes the centre of mass of the die.

(1.4.) *The measure-functions*

As already mentioned in (1.1.) we assume that the events considered originate from a σ-field $\mathscr{F}$ over Ω. By the above axioms the comparative structure '$\lesssim_K$' is defined for certain ordered pairs of events (A, B); the entirety of these (A, B) is designated $\mathscr{H}_K$.

DEFINITION. m_K is a measure-function to the given comparative structure '$\lesssim_K$' when:

(a) m_K is a measure on $\mathscr{F}$.

(b) $m_K(\Omega) = 1$.

(c) $m_K(A) \leqslant m_K(B)$ for all (A, B) of $\mathscr{H}_K$.

We confine our interest in the following to the concept of similarity.

(1.5.) *The form of the measure-functions*

Let the event field $\mathscr{F}$ with the basic set Ω be the Borellian extension of the events similar on the basis of K, $E(\lambda)$, $\lambda \in \Lambda$, i.e.: $\mathscr{F} := {}^B\{E(\lambda): \lambda \in \Lambda\}$.

PROPOSITION. A measure-function m_K on $\mathscr{F}$ is clearly defined by the following numbers $w_n := m_K(E(\lambda_1) \ldots E(\lambda_n))$, all λ_i being different; according to Axiom 2 $m_K(E(\lambda_1) \ldots E(\lambda_n))$ depends only on n but not on the specific parameters λ_i. For proof see [5].

The following extension of the de Finetti theorem applies:

DE FINETTI THEOREM. Presupposition: $|\Lambda| \geqslant \aleph_0$.

(a) If a normed measure m fulfils on the σ-field $\mathscr{F}$ produced by events $E(\lambda)$, $\lambda \in \Lambda$, the condition $m(E(\lambda_1) \ldots E(\lambda_n)) \doteq w_n$ for $\lambda_i \neq \lambda_j$ for $i \neq j$, then there exists a normed measure ϕ on the Borellian sets of the unity interval with the property $w_n = \int x^n \, d\phi$; for a general event $A \in \mathscr{F}$ $m(A) = \int p_x(A) \, d\phi$ applies; $p_x(A)$ is the 'probability' of A formed by the usual probability calculation, assuming that the $E(\lambda)$ are independent with the fixed 'probability' $x \in [0, 1]$.

(b) Conversely, if ϕ represents a normed measure on the Borellian sets of the unity interval, the definition $m(A) := \int p_x(A) \, d\phi$ leads to a measure-function on $\mathscr{F}$ which fulfils the conditions of (a).

For proof see [5].

The family of measure-functions on $\mathscr{F}$ is thus given by the set of all normed measures over the unity interval.

(1.6.) *The subjective interpretation of the measure-functions as fair betting odds*

We shall consider an individual $\mathscr{I}$ who accepts the concept K and concludes 'bets' on events according to the following

DEFINITION. A bet W on the event A is an agreement between $\mathscr{I}$ and a 'bank' which states that $\mathscr{I}$ on the occurrence of A will receive the sum of money $\bar{s} \geqslant 0$ from the bank whilst $\mathscr{I}$ on the occurrence of $\bar{A}$ will have to pay the amount of money $s \geqslant 0$ to the bank. $s, \bar{s}$ denote stakes and $g := s + \bar{s}$ is the total stake; it is always assumed that $g > 0$; $q := s/g$ is called the betting odds of this bet.

We shall consider systems $(W_1, \ldots, W_r)$ of bets, $r = 1, 2, 3, \ldots$ or $r = \infty$, and assume that $\mathscr{I}$ for certain systems $\mathscr{W}$, $\mathscr{W}'$ knows whether he prefers the system $\mathscr{W}$ to the system $\mathscr{W}'$, in formulae $\mathscr{W} \succ \mathscr{W}'$, or whether he considers the two systems to be equally good, in formulae $\mathscr{W} \equiv \mathscr{W}'$.

This *preference structure* of $\mathscr{I}$ for bets is to obey a series of conditions which are set forth in [6] so that we may speak of a rational preference structure.

We need the following

DEFINITION. $\bar{W}$ denotes the inverse bet to W when $\bar{W}$ is obtained by interchanging the roles of gambler and bank, i.e. for $\bar{W}$: $\mathscr{I}$ obtains on the occurrence of $\bar{A}$ the amount s whilst on the occurrence of A he must pay the amount $\bar{s}$. $\bar{W}$ has the betting odds $\bar{q} = 1 - q$.

DEFINITION. $\mathscr{I}$ considers a bet W to be fair when: $W \equiv \bar{W}$.

On the basis of the simple axioms set forth in [6] on the preference structure the following propositions apply:

PROPOSITION 1. For any event A there exists exactly one fair betting odds $q(A)$ dependently of $\mathscr{I}$.

PROPOSITION 2. $0 = q(\emptyset) \leqslant q(A) \leqslant q(\Omega) = 1$.

PROPOSITION 3. $q(\sum_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} q(A_n)$.

The fair betting odds $q(A)$ represent a measure for the subjective probability of $\mathscr{I}$. The measure-functions may therefore be taken to be the fair betting odds, the subjective probability, of an individual who accepts the concept K.

For a more detailed explanation see [5], [6].

(1.7.) *The completion of the comparative structure*

From the interpretation of the measure-functions as possible subjective probability results the following

AXIOM 3. If for all measure-functions of the concept K: $m_K(E) \leqslant m_K(E')$, then let $E \lesssim_K E'$.

This gives the following propositions:

PROPOSITION 1. The structure, $\lesssim_K$, is transitive.

PROPOSITION 2. If for all $i \in N$: $A_i \lesssim_K B_i$, then applies $\Sigma A_i \lesssim_K \Sigma B_i$.

(1.8.) *The quantitative logical probability*

Let A be an event with the property that for a given concept K each measure-function to this concept for A has the same value. We then introduce for the event A the following designation: $l_K(A) := m_K(A)$; $l_K(A)$ is called the *quantitative logical probability* of A with the concept K.

The function $l_K(A)$ is thus defined only for specific events; apart from the trivial examples $A = \emptyset$ and $A = \Omega$, in (3.1.) we shall meet an example of this.

It is easily verified that the domain of l_K is a σ-field $\mathscr{L}_K$ and l_K is a measure on $\mathscr{L}_K$; more exactly, $(\Omega, \mathscr{L}_K, l_K)$ is a probability field in the usual mathematical sense.

The definition of the quantitative logical probability corresponds to the example of deductive logic: a proposition is logically true if for every interpretation it is given the value 'true'. Accordingly, A has the quantitative logical probability $l_K(A)$ if for every possible subjective probability assessment A is given the same value $l_K(A)$.

From the considerations under (1.6.) the following results:

APPLICATION PRINCIPLE 1. Let the event A have a quantitative logical probability $l_K(A)$. Let W be a bet on A. Then $l_K(A)$ represents for any individual accepting the concept K the betting odds at which the bet W is fair on the basis of K.

(1.9.) *The quantitative logical upper (lower) probability*

In (3.4) we shall meet an event A having the property that for every measure-function m_K to K there holds: $m_K(A) \geqslant s > 0$, although the values $m_K(A)$ are all different and thus a quantitative logical probability for A does not exist. We therefore introduce the following notions:

$$\bar{l}_K(A) := \sup_{\substack{\text{all } m_K \\ \text{to } K}} m_K(A),$$

$$\underline{l}_K(A) := \inf_{\substack{\text{all } m_K \\ \text{to } K}} m_K(A);$$

$\bar{l}_K$ (or $\underline{l}_K$) are called *quantitative logical upper (or lower) probability* respectively.

According to the considerations under (1.6.) the situation can be interpreted as follows:

APPLICATION PRINCIPLE 2. Let W be a bet on A. If the betting odds of W are smaller than $\underline{l}_K(A)$ then the bet W is to be preferred to the bet $\bar{W}$. If the betting odds are greater than $\bar{l}_K(A)$ then conversely the bet $\bar{W}$ is to be preferred to the bet W, assuming of course that the concept K is accepted.

(1.10.) *The conditional concepts*

Assuming that to the concept K the additional information $\hat{A}$ is added which states that the event $A \in \mathscr{F}$ has occurred. We shall first assume that on the basis of the information $\hat{A}$ we pass to the concept $K' := K \wedge \hat{A}$. Subsequently, in (5.2.) we shall come across the case where we consider the concept K to be broken down by the information $\hat{A}$ and pass to a completely new concept K_{new}. However, it will be assumed here that the concept K' corresponds in its theoretical ideas to K and differs from K only by the additional information $\hat{A}$.

MULTIPLICATION PRINCIPLE. Let $q(E)$ be the fair betting odds of the individual $\mathscr{I}$ for the concept K, $q'(E)$ for the concept $K' = K \wedge \hat{A}$. Then for any $E \in \mathscr{F}$ we have:

$$q'(E) \,.\, q(A) = q(E \,.\, A).$$

For the proof of this principle see [6].

AXIOM 4. The concept $K' = K \wedge \hat{A}$ has the following measure-functions:

$$m_{K'}(E) = m_K(A \,.\, E)/m_K(A).$$

Measure-functions with the property $m_K(A) = 0$ do not make any contribution on the transition to K'. For on the basis of the concept K' only events $E \subset A$ play a part; a measure-function with $m_K(A) = 0$ says, however, nothing about events $E \subset A$ because for all these events without distinction $m_K(E) = 0$ applies. Thus, such measure-functions make no contribution on the transition to K'.

The comparative structure following from K' is defined by Axiom 3:

PROPOSITION. Let $K' = K \wedge \hat{A}$; then the comparative structure $\lesssim_{K'}$ results as follows from $\lesssim_K$:

$$E_1 \lesssim_{K'} E_2 \Leftrightarrow E_1 \,.\, A \lesssim_K E_2 \,.\, A.$$

2. The concept of the experiment and statistical ensemble

(2.1.) *Experiment and experimental schema*

It is a fundamental principle of every science that a certain experiment can be repeated, i.e. that certain phenomena are of the same type.

We shall attempt to understand this principle as follows: A class of similar experiments is given by an experiment instruction B within the language L which each of these experiments must obey. To be more exact: We shall assume that the description B contains a parameter λ whose specialisation defines an experiment; let there be no two experiments which satisfy the description $B(\lambda)$.

Accordingly, we call

$B(\lambda)$ a *concrete experiment*,

B $= \{B(\lambda): \lambda \in \Lambda\}$ an *experimental schema.*

As an example we shall consider the experiment $M(x, t)$ carried out by Michelson; the parameters (x, t) give the place and time for the performance of the experiment.

The experiment description B includes the statement of which events are possible. We shall assume here that each of the experiments $B(\lambda)$ has exactly k atomic possible events $E_1(\lambda), \ldots, E_k(\lambda)$, i.e. the events $E_i(\lambda)$ are to form a complete logical disjunction. E is called an event schema of B when:

$$\text{E} = \{E(\lambda): \lambda \in \Lambda\} \quad \text{with} \quad E(\lambda) = E_{i_1}(\lambda) + \cdots + E_{i_r}(\lambda).$$

(2.2.) *Generalization of the concept of similarity*

The notion that the experiments $B(\lambda)$ are of the same type is based with us on a concept K. In (3.1.) it is shown how the notion of similarity is to be understood on the basis of a concept K. We shall now extend this notion to the ensemble B $= \{B(\lambda): \lambda \in \Lambda\}$: The experiments $B(\lambda)$ are called similar on the basis of K if on the basis of K the parameter λ is irrelevant to the occurrence of the events $E_i(\lambda)$. Analogous to (3.1.) we have:

AXIOM 5. From the similarity of the experiments $B(\lambda)$ on the basis of a concept K we have:

$$E_{i_1}(\lambda_1) \ldots E_{i_n}(\lambda_n) \sim_K E_{i_1}(\mu_1) \ldots E_{i_n}(\mu_n), \quad \text{where } \lambda_i \neq \lambda_j, \mu_i \neq \mu_j \text{ for } i \neq j.$$

PROPOSITION. Let B $= \{B(\lambda): \lambda \in \Lambda\}$ be similar on the basis of K. Let E $= \{E(\lambda): \lambda \in \Lambda\}$ be any event schema of B, i.e. $E(\lambda) = E_{i_1}(\lambda) + \cdots + E_{i_r}(\lambda)$. Then: The $E(\lambda)$ are similar on the basis of K, i.e.:

$$E(\lambda_1) \ldots E(\lambda_n) \sim_K E(\mu_1) \ldots E(\mu_n) \quad \text{where } \lambda_i \neq \lambda_j, \mu_i \neq \mu_j \text{ for } i \neq j.$$

PROPOSITION. Let $\text{A}_i = \{A_i(\lambda): \lambda \in \Lambda\}$ be any event schemata to the ensemble B similar on the basis of K. It then holds that the events $A_i(\lambda)$ are similar, i.e.:

$$A_{i_1}(\lambda_1) \ldots A_{i_n}(\lambda_n) \sim_K A_{i_1}(\mu_1) \ldots A_{i_n}(\mu_n), \ \lambda_i \neq \lambda_j, \mu_i \neq \mu_j \text{ for } i \neq j.$$

DEFINITION. If from the concept K the similarity of the schema B follows according to Axiom 5, we call the class B a *statistical ensemble* on the basis of K.

The measure-functions. The measure-functions to this more general structure are defined as in (1.4.). In addition, a de Finetti theorem of more general character applies which however we shall not require in the following and consequently will not set forth there. For more details see [5].

3. The statistical behaviour of similar phenomena: The notion of statistical probability

(3.1.) *The convergence of relative frequencies*

The starting point of our observations is formed by a statistical ensemble B $= \{B(\lambda): \lambda \in \Lambda\}$ on the basis of a concept K and the σ-field $\mathcal{F} = {}^{B}\{E_i(\lambda): 1 \leqslant i \leqslant k, \lambda \in \Lambda\}$ over the basic set Ω, and any event schema E of B.

We now wish to prove the following: On the basis of the concept K with quantitative logical probability one any realization sequence of B has the same limit of the relative frequencies for the occurrence of the events $E(\lambda)$ in this sequence.

To prove this we shall consider a fixedly predetermined sequence $B(\lambda_i)$. $h_n(\mathsf{E}, \omega)$ be the relative frequencies of the event schema E in this sequence. We then define the following function:

$$p(\mathsf{E}, \omega) := \begin{cases} \lim_{n\to\infty} h_n(\mathsf{E}, \omega), & \text{if this limit exists} \\ 0 & \text{, if the limit does not exist.} \end{cases}$$

DEFINITION. $p(\mathsf{E}, \omega)$ is called the *statistical probability function* of E.

Note. According to its definition $p(\mathsf{E}, \omega)$ depends on the specific sequence $\{\lambda_i\}$; it is shown in Proposition 2 that $p(\mathsf{E}, \omega)$ except for an l_K null set is independent of the sequence $\{\lambda_i\}$.

Let $\Omega = \Omega^* + \Omega'$ with $\Omega^* := \{\omega : \lim_{n\to\infty} h_n(\mathsf{E}, \omega) = p(\mathsf{E}, \omega)\}$. Then:

PROPOSITION 1. $l_K(\Omega^*) = 1$ or more exactly $l_K(\{\omega : h_n(\mathsf{E}, \omega) \to p(\mathsf{E}, \omega)\}) = 1$.

Proof. The proof results from the usual strong law of large numbers in which the de Finetti theorem of (1.5.) is applied. For according to this law, with the symbols of (1.5.) the following equation holds true:

$$p_x(\{\omega : h_n(\mathsf{E}, \omega) \to x\}) = 1.$$

Since $\{\omega : h_n(\mathsf{E}, \omega) \to x\} \subset \{\omega : h_n(\mathsf{E}, \omega) \to p(\mathsf{E}, \omega)\}$ there follows $p_x(\{\omega : h_n(\mathsf{E}, \omega) \to p(\mathsf{E}, \omega)\}) = 1$; the integration of this equation with respect to ϕ gives: $m_K(\{\omega : h_n(\mathsf{E}, \omega) \to p(\mathsf{E}, \omega)\}) = 1$. This holds true for every m_K; it follows from this that the proposition is true.

Considered mathematically, Proposition 1 is nothing other than the law of strong convergence of the large numbers for symmetrical probability functions and is found in principle already in de Finetti. The expression $l_K(\{\omega : h_n(\mathsf{E}, \omega) \to p(\mathsf{E}, \omega)\}) = 1$ is intended only to make obvious that convergence of the relative frequencies represents with practical certainty an objective consequence of the concept of similarity. The statement $l_K(\Omega^*) = 1$ can also be expressed as follows: $\Omega^* \sim_K \Omega$. This means that the event of the convergence of the relative frequencies is a maximal element, equally probable with the logical certainty, in the comparative structure resulting from the concept K. In this form the fact that this statement is an objective consequence from the concept similarity is particularly clear.

PROPOSITION 2. Every two realization sequences have with quantitative logical probability one the same limit of the relative frequencies: $l_K(\{\lim_{n\to\infty} h_n = \lim_{n\to\infty} h'_n\}) = 1$; in this let $h_n = h_n(\mathsf{E}, \omega)$ be the relative frequency of E in the sequence $B(\lambda_i)$ and h'_n in a sequence $B(\lambda'_i)$.

For proof see [5].

(3.2.) *The notion of statistical probability*

To reconstrue the classical probability calculation on the basis of our statement we must associate with each event schema E of a statistical ensemble B a number, the statistical probability of E. Following Mises (see [8]) we define the statistical probability as limit of the actually occurring relative frequencies; this limit exists with quantitative logical probability one.

More exactly. Let $\hat{\omega}$ be the true result from Ω, normally unknown to us, and actually existing in nature.

We then call the following number

$$\mathsf{P}(\mathsf{E}) := p(\mathsf{E}, \hat{\omega})$$

the statistical probability of E.

Each event A of $\mathscr{F}$ corresponds to a statement $\hat{A}$, which states that A has occurred. We denote the functions $m_K, l_K, \underline{l}_K, \overline{l}_K$ transferred to the statements with $\mathsf{m}_\kappa, \mathsf{l}_\kappa$, etc.: $\mathsf{m}_\kappa(\hat{A}) := m_K(A)$, $\mathsf{l}_\kappa(\hat{A}) := l_K(A)$, etc.; it should be noted that logically equivalent statements are given the same values $\mathsf{m}_\kappa, \mathsf{l}_\kappa$. The statement $\hat{A}$ is logically equivalent to the statement $\hat{\omega} \in A$. Consequently, from (3.1.) the proposition $\mathsf{l}_\kappa(\hat{\omega} \in \Omega^*) = 1$ results. The statement $\hat{\omega} \in \Omega^*$ is logically equivalent to the statement $\mathsf{h}_n(\mathsf{E}) \to \mathsf{P}(\mathsf{E})$, wherein $\mathsf{h}_n(\mathsf{E}) := h_n(\mathsf{E}, \hat{\omega})$. Consequently:

$$\boxed{\mathsf{l}_\kappa(\mathsf{h}_n(\mathsf{E}) \to \mathsf{P}(\mathsf{E})) = 1}$$

The dualism of the formulae for the statistical probability and the statistical probability function applies for all the formulae derived by us below.

Note. Whereas the notion of the quantitative logical probability related to a concrete event the notion of the statistical probability or function relates to an event schema E and generally says nothing about a specific $E(\lambda_0)$ of E. For in general a specific $E(\lambda_0)$ does not define the event schema E clearly. For more details see [5].

(3.3.) *The addition principle for the statistical probability*

Let E, E′ be two disjunctive event schemata of B. Then the following is true:

$l_K(\{\omega: p(E+E',\omega) = p(E,\omega) + p(E',\omega)\}) = 1$ or

$$\boxed{l_K(P(E+E') = P(E) + P(E')) = 1}$$

The proof results from (3.1.) and the additivity of the relative frequencies.

(3.4.) *The velocity of the convergence*

The above results only have a practical significance if we know something about the velocity of the convergence of the relative frequencies. Now, this velocity can be estimated with quantitative logical probability. The analog of the Bernoulli inequation holds true:

(a) $l_K(\{\omega: | h_n(E,\omega) - p(E,\omega)| \leqslant \epsilon\}) \geqslant 1 - 1/4n\epsilon^2$

$$\boxed{l_K(|h_n(E) - P(E)| \leqslant \epsilon) \geqslant 1 - 1/4n\epsilon^2}$$

(b) $l_K(\{\omega: | h_n(E,\omega) - h_r(E,\omega)| \leqslant \epsilon\}) \geqslant$

$$1 - \left|\frac{1}{r} - \frac{1}{n}\right| \frac{1}{4\epsilon^2}$$

$$\boxed{l_K(|h_n(E) - h_r(E)| \leqslant \epsilon) \geqslant 1 - \left|\frac{1}{r} - \frac{1}{n}\right| \frac{1}{4\epsilon^2}}$$

For proof see [5].

(3.5.) *The notion of the subconcept*

DEFINITION. $\hat{K}$ is called a *subconcept* of K if the measure-functions permitted on the basis of $\hat{K}$ form a subset of the measure-functions permitted on the basis of K.

From this follows according to Axiom 3:

$A \lesssim_K B \Rightarrow A \lesssim_{\hat{K}} B.$

(3.6.) *The concept K_x*

Let E be any event schema of B; $x \in [0,1]$. We then define in accordance with (1.10.):

$K_x := K \wedge \hat{A}$ with $A := \{\omega: p(E,\omega) = x\}$;

$\hat{A} \Leftrightarrow P(E) = x.$

As a brief consideration will show, (1.10.) gives exactly the function p_x (see (1.5.)) as a measure-function with respect to K_x. Thus, in particular the following applies: K_x is a subconcept of K.

The following also applies:

$$l_{K_x}(E(\lambda)) = p_x(E(\lambda)) = x = P(E) \overset{l_{K_x}-\text{almost sure}}{=} p(E,\omega).$$

The quantitative logical probability of $E(\lambda)$ is on the basis of K_x equal to the statistical probability, or l_{K_x}-almost sure is equal to the statistical probability function. This shows why we also call these quantities probability, as was originally conceived in the sense of fair betting odds.

An example for the concept K_x is a coin tossed repeatedly whose symmetry has been proved by precise measurement; then we may accept the concept $K_{1/2}$.

(3.7.) *The concepts K^**

DEFINITION. Let E be any event schema of B. The concept K^* is then characterized as follows: Only such measure-functions m_{K^*} are permitted such that for the event schema E the following holds true: The measure ϕ^* associated in accordance with (1.5.) has a continuous positive density on the unity interval: $d\phi^* = f(x)dx$; f continuous and positive. K^* is thus a subconcept of K.

A concept K^* is given exactly whenever the existing theoretical ideas and information do not make obvious any probability-theoretical singularities with regard to the nature of P(E).

Let $B(\lambda_i)$ be any realization sequence of B. After n attempts let $E(\lambda)$ occur r_n times and $\bar{E}(\lambda)$ occur $(n - r_n)$ times. The following then holds for the measure-functions associated with the concept $K^*_{n,r_n} := K^* \wedge (h_n = r_n/n)^1$ in accordance with the multiplication rule (1.10.) provided $\mu_i \in \Lambda - \{\lambda_1, \ldots, \lambda_n\}$

$$m_{K^*_{n,r_n}}(E(\mu_1) \ldots E(\mu_s)) = \int x^s \, d\phi^*_{n,r_n}$$

with

$$d\phi^*_{n,r_n} := \frac{x^{r_n}(1-x)^{n-r_n}}{\int x^{r_n}(1-x)^{n-r_n}\, d\phi^*}\, d\phi^*.$$

PROPOSITION 1. Let $\Phi^*_{n,r_n}(x) = m_{K^*_{n,r_n}}(\{\omega: p(E,\omega) \leqslant x\})$ be the distribution function associated with the measure ϕ^*_{n,r_n}. If the relative frequencies r_n/n converge towards p when n tends to infinity, the distribution function Φ^*_{n,r_n} converges in the sense of the distribution convergence towards $D(x-p)$; $D(x)$ represents the Dirichlet function:

$$D(x) := \begin{Bmatrix} 0 & \text{if} & x < 0 \\ 1 & & x \geqslant 0 \end{Bmatrix}.$$ For proof see [5].

Thus we have from (3.1.):

PROPOSITION 2. Let $\Phi^*_{n,\,r_n}(\omega)$ denote the dependence of the above distribution function on ω, since of course r_n depends on ω. Then:

$$l_{K^*}(\{\omega : \phi^*_{n,\,r_{n(\omega)}} \to D(x - p(\mathsf{E}, \omega))\}) = 1.$$

This proposition can be described as follows: The concept K^* converges with increasing information with quantitative logical probability one towards the concept $K_{\mathsf{P(E)}}$ defined in accordance with (3.6.); in formulae:

$$K^*_{\mathsf{n}} := K^*_{n,\,r_n}(\hat{\omega}),$$

$$\boxed{l_{K^*}(K^*_{\mathsf{n}} \to K_{\mathsf{P(E)}}) = 1}$$

PROPOSITION 3. $m^*_n(\omega) := m_{K^*_{n,\,r_n(\omega)}}(E(\lambda_{n+1}))$ is the conditional subjective probability of $E(\lambda_{n+1})$ on the basis of the condition $\mathsf{h_n} = r_n/n$. The following then holds good: $l_{K^*}(\{\omega : m^*_n(\omega) \to p(\mathsf{E}, \omega)\}) = 1$.

$\mathsf{m}^*_{\mathsf{n}} := m^*_n(\hat{\omega})$ is the conditional subjective probability actually occurring in the course of the observation. The following holds true:

$$\boxed{l_{K^*}(\mathsf{m}^*_{\mathsf{n}} \to \mathsf{P(E)}) = 1}$$

This proposition again illustrates the reasons why we brought $p(\mathsf{E}, \omega)$ and $\mathsf{P(E)}$ into connection with the word probability. According to the above formula $\mathsf{P(E)}$ is the limit value, independent of any information and subjective arbitrariness, of the conditional probability of every possible subjective probability assessment of the concept K^*.

4. The notion of statistical independence

(4.1.) *Definition of the statistical independence on the basis of K*

Two events schemata $\mathsf{A_1}$ and $\mathsf{A_2}$ of a statistical ensemble B are called on the basis of K *statistically independent* if: $l_K(\{\omega : p(\mathsf{A_1} \,.\, \mathsf{A_2}, \omega) = p(\mathsf{A_1}, \omega) \,.\, p(\mathsf{A_2}, \omega)\}) = 1$, or

$$\boxed{l_K(\mathsf{P}(\mathsf{A_1} \,.\, \mathsf{A_2}) = \mathsf{P}(\mathsf{A_1}) \,.\, \mathsf{P}(\mathsf{A_2})) = 1}$$

(4.2.) *Equivalent conditions for the statistical independence*

$\mathsf{A_1}$ and $\mathsf{A_2}$ are statistically independent on the basis of K exactly if:

$$A_1(\lambda_1) \ldots A_1(\lambda_n) \,.\, A_2(\rho_1) \ldots A_2(\rho_r)$$
$$\sim_K A_1(\mu_1) \ldots A_1(\mu_n) \,.\, A_2(\delta_1) \ldots A_2(\delta_r);$$

where it must be assumed that $\lambda_i \neq \lambda_j$, $\rho_i \neq \rho_j$, $\mu_i \neq \mu_j$, $\delta_i \neq \delta_j$ for $i \neq j$; it is, however, admissible that certain of the λ_i coincide with certain of the ρ_j.

Proof see [5].

5. The verification of concepts

(5.1.) The concept of similarity of experiments results in the convergence of the relative frequencies with quantitative logical probability one. If we observe in nature a statistical behavior – practical constancy of the relative frequencies for a large number of conducted experiments – this concept will be confirmed. The concept K will break down if the relative frequencies do not exhibit practical constancy. If a concept $\bar{K}$ becomes apparent which is 'more' in harmony with experience than the concept K then the concept K will be rejected in favor of the concept $\bar{K}$. If in the choice between K and $\bar{K}$ only the constancy of the relative frequencies plays a part, from the 'Bernoulli inequation' derived under (3.4b) logical error probabilities for the rejection of K can be calculated.

However, in general the situation on comparing K and $\bar{K}$ is more complicated. For more exact details see [5].

(5.2.) *Example for the abandonment of the concept of similarity*

If we observe that in successive throwing of a dice after every ninth throw the six appears whilst otherwise the one occurs we will abandon the concept K of the similarity of these throws and adopt a concept K_{new} which states that the throw experiment contains a mechanism which provides the observed regularity. Within our formalism the concept K_{new} is characterized in that it contains only the measure-function m_{new} which assigns to the above regularity the measure one.

It should be noted with regard to this example that the concept K_{new} also results in the convergence of the relative frequencies so that the conditions formulated in (5.1.) for the rejection of K are not present. For more details on the formal comprehension of this example see [5].

6. The application of the theory to quantum mechanic

In quantum mechanics the concept of similarity appears with particular stringency: Let us consider the example given at the beginning of a polarized sequence of photons which impinge on a polarization filter. The impingement of the individual polarized photons on the filter represents chronologically different processes which on the basis of the ideas of quantum mechanics are similar because no further distinction can be made between them by measurements: No measurement exists on a polarized photon which would provide any information on whether this photon is more likely to pass through the filter than another photon polarized in the same manner. Let us further assume that the individual photons impinge on the filter at long time intervals apart, about an hour; it is then clear that the passage of n photons in this sequence is logically equally probable to the passage of n other photons.

If the polarized photons impinge on the filter closely adjacent to each other chronologically, it is admittedly a priori not clear that the passage of a photon through the filter is not influenced by the passage of the other photons and to be exact would first have to be empirically confirmed. If, however, we assume that the impinging of the photons on the filter represents isolated physical processes it follows from this idea that the passage of n photons is logically equally probable to the passage of n other photons. Consequently, according to the ideas T of quantum mechanics outlined in this manner the passage of polarized photons through a filter represents a class of similar processes which fulfil the Axiom 2 of (1.3.). Consequently, since these phenomena are similar, they must exhibit a statistical behavior. The above theory thus establishes a logical relationship between the theoretical structure of the similarity of the phenomena given in quantum mechanics and the empirical observation of their statistical behavior:

The statistical behavior of quantum-mechanical phenomena is explained by their similarity.

Conversely, the statistical behavior of the phenomena permits the concept formulation existing in quantum mechanics of physical similarity of these phenomena.

7. Comparison with other probability systems

(7.1.) *Comparison with de Finetti's system*

The above system is similar in its mathematical structure to the system of de Finetti [4].

The difference between our system and that of de Finetti is:

(1) The σ-additivity of the measure-functions proved in [6].

(2) That the notion of subjective probability is eliminated and replaced by the notion of logical probability on the basis of a concept.

(3) For the symmetry of the measure-functions in the concept K of similarity an objective basis has been created.

(4) The possibility taken into account theoretically in (1.10.) and illustrated by an example in (5.2.) of abandoning the concept of similarity in the case of corresponding information.

Quite generally, we criticize de Finetti's system as follows: Let us consider for example radioactive decay. A scientist $\mathscr{I}$ will then assume the concept K of similarity and the laws derived by us which hold true for every possible 'subjective' interpretation of the comparative structure. However, in general he will not be able to define a specific subjective assessment as binding for him. For on what grounds should he select a specific subjective assessment? We therefore maintain that de Finetti's system is only made applicable by the construction made by us in which only statements are considered which hold good for every 'subjective' assessment without it being maintained that real individuals exist who made such a subjective assessment.

(7.2.) *Comparison with Carnap's system*

The above work is similar in the notion of logical probability to the system of Carnap [1].

As the above work shows, Carnap's axioms imply the statistical behavior of the phenomena. These axioms thus have an empirical content and therefore cannot be interpreted as a priori characteristic of a formal language system but must be based on a scientific concept with empirical content.

As in de Finetti's system it is also impossible in Carnap's system to abandon the symmetry of the measure-functions on corresponding observation. As explained in (7.1.) we consider this attitude to be untenable.

(7.3.) *Comparison with Mises' system*

The above structure is similar to the system of Mises [8] insofar as the statistical probability is also introduced as limit of the relative frequencies. In contrast to Mises, in our system Mises' axioms do not hold strictly true but only with quantitative probability one, or to be more exact Mises' axioms freed from their original trivial contradiction.

We criticize Mises' system as follows:

(a) In Mises' system the phenomenon of statistical regularities is only described but not explained as in our system as a consequence of the concept of similarity. Mises gives no condition, such as the concept of similarity, for deciding when an experiment sequence has the characteristic of convergence in infinity.

(b) The limit of the relative frequencies existing in Mises' system does not necessarily result, as it does in our system (see (3.6.)), in the meaning of fair betting odds for each terms of the sequence. However, for the practical use of the probability theory this interpretation is essential. This defect could be remedied by adding to Mises' system a corresponding axiom.

(c) Mises' system provides no information on the velocity of the convergence of the relative frequencies and consequently has no practical relevance whatsoever because only the behavior in the finite is of practical significance. We note that of course in Mises' system as well the Bernoulli inequation can also be derived. However, in this system this equation also only provides information on the behavior at infinity of a collective derived from the starting collective and provides no information whatever on the behavior in the finite. As remarked under (b), only when we add to Mises' system the axiom of the interpretation of the limit as fair betting odds does the Bernoulli inequation in Mises' system permit interpretation which leads to a statement on the velocity of the convergence.

(7.4.) *Comparison with Kolmogoroff's system*

The system of Kolmogoroff [7] introduces the notion of probability as an undefined basic concept which obeys certain axioms. The relationship with the relative frequencies results for instance from the Bernoulli inequation: $p(|h_n(E)-p(E)|\geqslant\epsilon)\leqslant 1/4n\epsilon^2$.

This system is consistent in itself and also represents a good description of the phenomenon of statistical regularities. However, this system contains an epistemological circle: The axioms required are motivated by the probability having to represent the 'limit value' of the relative frequencies. Thus, it can only be said that Kolmogoroff's theory describes the phenomenon of statistical regularities; it provides no explanation of this regularity and attempts to comprehend it in its axioms.

The circularity of Kolmogoroff's system is particularly clearly apparent in the Bernoulli inequation: The probability $p(E)$ to be explained is approximated in the interior of the formula by the relative frequencies; however, the measure of the approximation is also represented in the exterior of the formula by the p which itself has to be explained. In contrast, in our system the degree of approximation of the statistical probability by the relative frequencies is expressed with the quantitative logical lower probability l_K which has already been defined previously independently of the notion of statistical probability.

8. Conclusion

The basic concept of the present work was the notion of the concept K. All further fundamental concepts were obtained by successive 'derivation' from the concept K. The result of the present work resided substantially in the following

THESIS. From the concept K of similarity there follows the statistical behavior of the phenomena; in particular, on the basis of K a concept of the statistical probability P can be defined. However, there is no logical certainty on the basis of K but only a sureness which is practically equivalent and can be called absolute that P has all the properties usually required of the concept of statistical probability, for instance that P is additive, or that the relative frequencies converge toward P, etc.

The concept K is therefore not absolute but a scientific statement, an overhypothesis, which can be verified by observations.

Below we give a brief review of the questions which could still be dealt with using the above concept:

(8.1.) *The σ-additivity of* P

Investigations of the statement

$$l_K(\text{P is a measure on } \mathscr{E}) = 1,$$

for the case that the domain of P, the σ-field of the possible event schemata in relation to the experimental schema B, is not finite.

The above statement is equivalent to the following statement:

$$l_K(\{\omega: p(\Sigma A_i, \omega) = \Sigma p(A_i, \omega) \text{ for all } A_i \in \mathscr{E}\}) = 1.$$

The investigations of this question appear difficult. It would, however, be clear that the measure property of P is given only under certain regularity conditions applied to $\mathscr{E}$.

The following statement is always valid:

$$l_K(\{\omega: p(\Sigma A_i, \omega) = \Sigma p(A_i, \omega)\}) = 1.$$

(8.2.) *Investigations with regard to non-similar concepts*

Such as investigations on stationary sequences, Markoff chains, concepts which fulfil with quantitative logical probability one the Mises axioms, etc.

(8.3.) *The converse law of statistical probability*

The converse of the thesis set forth above in the sense that a statistical probability can only be defined if the concept of similarity is assumed. For more details see [5].

(8.4.) *Special investigations on the Carnap theory*

Special investigations on the concepts K_C of the confirmation and measure-functions worked out by Carnap ([1]–[3]); for instance determination of $\mathscr{L}_{K_C}, l_{K_C}, \underline{l}_{K_C}, \overline{l}_{K_C}$.

(8.5.) *Special investigations on quantum mechanics*

Special developments of the above theory for quantum mechanics, the peculiarity of quantum mechanics being the event field $\mathscr{V}(\lambda)$ belonging to the experiment $B(\lambda)$ is a lattice and not a σ-algebra.

For the validity already maintained in this work of our theory for the photon example considered see [5].

(8.6.) *Investigations of the Hume problem*

In reference to an example we intend illustrating how, within the framework of our concept formation, a certain solution of the Hume problem is obtained, i.e. the conclusion from the observed to the unobserved, not stringently but nevertheless with any desired high probability.

For an event ensemble E let the concept K of similarity be presumed. In addition, only 'deterministic' observations occur, i.e. it always hold that $h_n(E) = 1$.

It then results from the analogon to the Bernoulli inequation:

$$\underline{l}_K(|1 - P(E)| \leqslant \epsilon) \geqslant 1 - 1/4n\epsilon^2.$$

From this, by an adequate number n of observations any desired certainty can be obtained so that: $P(E) \geqslant 1 - \epsilon$.

Thus, on the basis of such observations we can go over to the concept $K_{1-\epsilon} := K \wedge (P(E) \geqslant 1 - \epsilon)$ according to (1.10.).

A brief thought will show that on the basis of the concept $K_{1-\epsilon}$ the following is true: $\underline{l}_{K_{1-\epsilon}}(E(\lambda_0)) \geqslant 1 - \epsilon$; $E(\lambda_0)$ is any unobserved event of E.

Summarizing, it may therefore be stated: On the basis of the concept of similarity by a sufficiently great number of observations any desired certainty can be achieved for prognosticating the unobserved event $E(\lambda_0)$ if all observed $E(\lambda)$ have occurred.

This provides a certain solution for the Hume problem. This solution is not absolute because the concept of similarity is not absolute but a scientific statement which can also be abandoned.

The concept K represents a type of uniformity principle on the basis of which the Hume problem can be solved. For the position to be assigned to the 'uniformity principle' K, see [5].

Note

1 $h_n = h_n(E)$, see (3.2.).

Literature

[1] Carnap, R. and W. Stegmüller: *Induktive Logik und Wahrscheinlichkeit*, Wien, 1958.
[2] Carnap, R. and C. R. Jeffrey: *Studies in Inductive Logic and Probability*, Vol. I, Berkeley, Los Angeles, London, 1971.
[3] Jeffrey, C. R.: *Studies in Inductive Logic and Probability*, Vol. II, Berkeley, Los Angeles, 1980.
[4] Finetti, B. de: *La prévision: ses lois logiques, ses sources subjectives*, Vol. VII, 1, Institut H. Poincaré, Paris, 1936.
[5] Humburg, J.: *Grundzüge zu einem neuen Aufbau der Wahrscheinlichkeitstheorie*, Europäische Hochschulschriften, Verlag P. Lang, Frankfurt, Bern, 1981.
[6] Humburg, J: *Wahrscheinlichkeits- und Wettbegriff*, Vol. 19, Fasc. 1, Metrika, 1972.

[7] Kolmogoroff, A. N.: *Foundations of the Theory of Probability*, New York, 1951.
[8] Mises, R. v.: *Wahrscheinlichkeit, Statistik und Wahrheit*, Wien, 1928, 1951.

Schleissheimerstrasse 173,
München, B.R.D.