Revista Colombiana de Matemáticas
Vol. XIX (1985), pág. 43 - 57

FACTUAL AND COGNITIVE PROBABILITY

Rolando Chuaqui*

ABSTRACT. The paper presents a modification of the definition of probability presented in earlier papers of the author (e.g. *A semantical definition of probability*, in Non-Classical Logics, Model Theory and Computability, North Holland, pp.135-167). This modification separates the two aspects of probability: probability as a part of physical theories (factual), and as a basis for statistical inference (cognitive). Factual probability is represented by probability structures as in the earlier papers, but now built independently of the language. Cognitive probability is interpreted as a form of "partial truth". The paper also contains a discussion of the Principle of Insufficient Reason and of Bayesian and classical statistical methods, in the light of the new definition.

This paper presents a modification of the semantical definition of probability introduced in Chuaqui 1977 and 1980. The new definition presented here brings forth the two aspects of probability: as a basis for statistical inference and as a part of physical theories.

The main modification introduced is making independent of the language the definition of the group of transformations that preserve the laws of the phenomenon. Thus, the determination of the probability measure for the simple probability structures of Chuaqui 1977 becomes independent of linguistic elements, and the simple probability interpretations $\langle \mathbf{K}, \mathcal{B}, \mu \rangle$ of Chuaqui 1980 may be considered as models of reality.

The connection with cognitive elements is established via the concept of probability as degree of partial truth, introduced in earlier papers.

The first section analyzes the different uses of probability, while in the second, I give a brief account and a classification of theories on probability. Section 3 introduces the modification of the definition in Chuaqui 1977 and 1980 that permits to consider probability structures as models of reality.

---

* This work was partially supported by the Organization of American States through its Regional Scientific and Technological Development Program.

The fourth section justifies the cognitive uses of probability by considering it as a logical concept, while the fifth applies these remarks to an analysis of the *Principle of Insufficient Reason*. The last section complements the study in Chuaqui 1980 of classical and Bayesian statistical methods.

I would like to thank Professor William Reinhardt, Newton C.A.da Costa, and Leopoldo Bertossi for many useful comments.

1. USES OF PROBABILITY.

Probability is used as a basis for statistical inference and as a part of scientific theories. In its first use, I shall distinguish two different applications. One application is to use probability as the very guide of life in the face of uncertainty. Thus, probability is the basis of decision theory in its classical and Bayesian forms. The need for statistical inference arises from our uncertainty as to how we ought to behave under circumstances where we are ignorant concerning the state of the world. As Kyburg 1974 says, "we attempt to develop rules of behavior which we may follow, whatever the state of the world, in the expectation that we are following rules whose characteristics are generally (1) desirable, and (2) attainable". The most desirable rule is one that tells us how to discover the true state of the world; the most attainable and simplest is to forget about the arithmetic and act as we feel like it. A compromise between these two extremes is to follow what we may know about the probabilities of the different possible states of the world, i.e. about a measure of the degree of truth of each possible state of the world.

The second application of statistical inference is the evaluation of scientific hypotheses. This application can be thought of as a case of a decision whether to accept or not a scientific hypothesis, and thus it can be assimilated to processes of the first kind. However, I believe with R.A. Fischer that "... such processes have a logical basis very different from those of a scientist engaged in gaining from his observations an improved understading of reality". (Fisher 1956 p.5).

Besides these statistical uses, probability statements appear as part of physical theories such as Statistical Mechanics or Quantum Mechanics. Also, most of the theory of stochastic processes serves as basis for scientific theories of particular phenomena, such as Brownian motion and radioactive decay.

Although the word "probability" itself might not occur in a sientific theory, probability concepts are present as general statements expressing stochastic relations among random quantities. For instance, there may be functions expressing distributions or densities of certain quantities under certain circumstances. Anyway, probabilities of events are obtained from them and used in applications.

Intuitively, there are other evidences of probabilities as independent of our knowledge or belief. For instance, if we toss a coin 100 times and in 60 obtain "heads", it seems natural to believe that this is a property of the coin or rather of the coin together with the mechanism for tossing it.

Since the statistical applications stem from our ignorance of the true state of the world, I shall call them *cognitive uses*. On the other hand, the other uses will be the *factual uses*.

2. COGNITIVE VERSUS FACTUAL INTERPRETATIONS.

The interpretations of the concept of probability which have been offered stress either the cognitive or the factual uses of probability. Among the first, I would put the subjectivist and logical views. Among the second, the frequentist and propensity views. The subjectivist do not attempt to explain how we get our probabilities while the holders of the logical views do. On the other side, the propensity theories do not attempt to define probabilities but only to measure them, while the frequentists build models of reality where probability is defined in terms of other concepts.

Probability, however, has both cognitive and factual aspects. Thus, any interpretation should give an account of both. In order to do this, many scholars hold a dual view: there is an interpretation of probability as *degree of belief* or credence (cognitive) and another as *chance or propensity* (factual). The connection between the two should be given by a principle of the following form (a similar principle was formulated in Lewis 1980): let X be the proposition that the chance, at time t, of A holding equals x, where x is a real number of the unit interval. Let C_X be any reasonable "degree of belief" function of a person that believes X at time t. Then $C_X(A) = x$.

In this principle, X is supposed to contain the statement about the factual probability of A. $C_X(A)$ is the cognitive probability of A. The two are supposed to be connected by the principle.

I believe that any interpretation of probability should explain both the factual and cognitive uses of probability and justify a principle such as the one above. I shall attempt to provide such an account by modifying the views espoused in Chuaqui 1977 and 1980.

3. PROBABILITY STRUCTURES AS MODELS OF CHANCE.

3.1. Simple probability models. In Chuaqui 1977, the theory of simple probability structures was presented. From the simple probability structures $\mathbf{K}$, there were obtained probability interpretations of languages, constituted by triples $\langle \mathbf{K}, \boldsymbol{B}, \mu \rangle$, where $\boldsymbol{B}$ is a field of subsets of $\mathbf{K}$ and μ a probability measure.

We want to consider these probability structures as models of reality which determine the probability measure μ. However, in my original presentation, the definition of μ depended heavily on the language used. I believe this to be a feature that precludes their use as models of reality. I shall now offer a modification of the concept of simple probability structures that will give us a definition of μ, independent of the language.

There are two elements in $\langle \mathbf{K}, \boldsymbol{B}, \mu \rangle$ that are language dependent in Chuaqui 1977. The first is the definition of B as the field of sets generated by all the sets $Mod_{\mathbf{K}}(\phi)$ where ϕ is a sentence of the appropriate language (recall that $Mod_{\mathbf{K}}(\phi)$ is the set of models of $\mathbf{K}$ in which ϕ holds). However, B is the algebra of events and there is no need of choosing it in this way, it can be taken to be just the algebra of events that are necessary for an adequate description of the situation. In case $\mathbf{K}$ is finite for instance, B will generally be the power set of $\mathbf{K}$. A more detailed explanation of B will be given in Section 4.

The other linguistic element is the definition of the group $G_{\mathbf{K}}$, namely conditions (1)(a) and (1)(b). This we have to get rid of.

In order to make the situation clear, I shall begin with the same example as in Chuaqui 1977, the choosing of a sample S of size m from a finite population of balls P. When we say "S has n red balls" we mean that one of the properties of the outcome was that the sample had n red balls. The same outcome has many different properties. We can think of an ideal approximation of an outcome, namely a relational system that represents a possible model of the situation involved. In the case we are looking at, we can schematize the possible outcomes as systems $\mathfrak{A}_S = \langle P, R_0, \ldots, R_{m-1}, S \rangle$, where P is the finite set of balls, $R_0, \ldots, R_{m-1}$ are fixed subsets of P that represent the properties we are interested in (for instance, red), and S is any subset of P of m members (the sample). For each subset S of m members there is a corresponding system $\mathfrak{A}_S$; the set of possible outcomes $\mathbf{K}$, consist of all models $\mathfrak{A}_S$ of the form described above.

Let us analyze a possible outcome $\mathfrak{A}_S = \langle P, R_0, \ldots, R_{m-1}, S \rangle$. The properties $R_0, \ldots, R_{m-1}$ are intrinsic properties of the balls in P. That is, when we move the balls around or choose a sample, their properties remain. Also, these properties are fixed in all $\mathfrak{A}_S$. We may thus call $\langle P, R_0, \ldots, R_{m-1} \rangle$ the *intrinsic part of* $\mathbf{K}$. On the other hand, $\langle P, S \rangle$ gives the structure of the experiment, and is variable in each $\mathfrak{A}_S$. $\langle P, S \rangle$ is called the *structural* part of $\mathfrak{A}_S$ and denoted $\mathfrak{A}_{S,st}$. We may have different experiments performed on the same set of balls P. The probability structures for these different experiments might have the same intrinsic part but different structure. For instance, if the experiment consists of the choosing of a sample with two R_0-balls, the outcome will be of the form $\langle P, R_0, \ldots, R_{m-1}, Q \rangle$ where Q is a subset of P with two R_0-balls.

The second example I shall give is a modification of Example 3 of Chuaqui 1977. Suppose we have a circular roulette with a point for each real number.

For simplicity, a fixed force is applied but the roulette starts from a variable position. Each outcome results from beginning at a particular position. The systems in $\mathbf{K}$ may be taken to be of the form

$$\mathfrak{A}_I = \langle C, r, f, I\rangle$$

where C is the set of points in the circle, r represents translations in the circle (a ternary operation, $r(a,b,c)$ is c rotated by the angle from a to b), f is the continuous unary function that associates each initial position with a final position, and I is the set containing the initial position (I contains one element of C). Here, $\langle C,f\rangle$ constitutes the intrinsic part of $\mathbf{K}$, because it is an intrinsic property of each point x in C that it yield a final position $f(x)$. With this f we can express the asymmetry of the roulette, if it is asymmetric. We can model symmetric roulettes without this f. The structure of the experiment is given by $\langle C,r,I\rangle$. I is variable in the different $\mathfrak{A}_I$. However r is fixed. Thus, we cannot distinguish between the intrinsic part and the structure by just looking at the variable part of $\mathbf{K}$. Notice that C is not enough for defining a circle. It is necessary to add an operation between the elements of C. For instance, I have chosen in this paper r. This operation r should be part of the structure of the experiment, because it really consists of a rotation of the circle. If we had just $\langle C,I\rangle$ in the structure, the experiment would be the choosing of a point in a set C and not in a circle $\langle C,r\rangle$.

From these two examples, we see that in order to describe the simple probability structure we need to specify, besides the class of possible outcomes $\mathbf{K}$, its intrinsic part. Thus, we define a simple probability structure as a pair $K = \langle \mathbf{K}, \mathfrak{A}\rangle$ where $\mathbf{K}$ is a set of relational structures of a fixed similarity type (called the *set of outcomes* of K) and $\mathfrak{A}$ is a relational structure (called the *intrinsic part* of K) such that $\mathfrak{L}\restriction J = \mathfrak{A}$ for every $\mathfrak{L}\in \mathbf{K}$, where J is the index set of the similarity type of $\mathfrak{A}$ (thus, all structure $\mathfrak{L}\in \mathbf{K}$, have the same universe say A).

The group of functions $G_{\mathbf{K}}$ is now determined by K and B without reference to the language. For $\mathfrak{L}\in \mathbf{K}$, let $\mathfrak{L}_{st} = \mathfrak{L}\restriction(I-J)$ where I is the index set of the similarity type of $\mathbf{K}$. That, is $\mathfrak{L}_{st}$ represents the structure of the experiment. Also, for $\mathbf{B}\subseteq \mathbf{K}$, $\mathbf{B}_{st} = \{\mathfrak{L}_{st} : \mathfrak{L}\in \mathbf{B}\}$.

The group of functions $G_{\mathbf{K}}$ that preserve the "laws of the phenomenon" contains all permutations f of the universe A such that

(1) For any $\mathfrak{L}\in \mathbf{K}$, $f^*(\mathfrak{L}_{st}) \in \mathbf{K}_{st}$ and $f^{-1*}(\mathfrak{L}_{st}) \in \mathbf{K}_{st}$

(2) For any $\mathbf{B}\in B$, $\mathbf{B}^f \in B$ and $\mathbf{B}^{f^{-1}} \in B$, where $\mathbf{B}^f$ is the unique $\mathbf{C}\subseteq \mathbf{K}$ such that $\mathbf{C}_{st} = \{f^*\mathfrak{L} : \mathfrak{L}\in \mathbf{B}_{st}\}$.

Condition (1) can be expressed simply by $\mathbf{K}^f \in \mathbf{K}$ and $\mathbf{K}^{f^{-1}} \in \mathbf{K}$. The measure μ is a measure invariant under $G_{\mathbf{K}}$.

In our first example $G_{\mathbf{K}}$ consists of all permutations of the universe P. In the second, $G_{\mathbf{K}}$ contains all automorphisms of $<C,r>$. In general, if $<A,P_0,..,P_{n-1}>$ is constant in all elements of **K** and none of the relations $P_0,\ldots,P_{n-1}$ are in the intrinsic part of **K**, then $G_{\mathbf{K}}$ is a subgroup of the automorphisms of $<A,P_0,\ldots,P_{n-1}>$.

In order to extend the situation to compound structures, we define a symmetry relation between subsets A, B of **K**:

$$\mathbf{A} \underset{\mathbf{K}}{\sim} \mathbf{B} \text{ iff } \mathbf{A}^f = \mathbf{B} \text{ for a certain } f \in G_{\mathbf{K}}.$$

The measure μ is now invariant under $\sim_{\mathbf{K}}$, i.e. $\mathbf{A} \underset{\mathbf{K}}{\sim} \mathbf{B}$ implies $\mu(\mathbf{A}) = \mu(\mathbf{B})$.

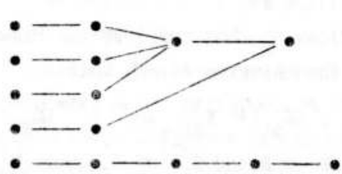
Summarizing, we have obtained a model of reality $<\boldsymbol{K},\boldsymbol{B}>$ that gives a probability measure μ invariant under $G_{\mathbf{K}}$. If this μ were the unique invariant measure, then $<\boldsymbol{K},B>$ would be a sufficient description of the model. Uniqueness is not a rare phenomenon. As a matter of fact, all situations I have analyzed yield a unique measure. However, the general conditions for uniqueness that I know of are rather technical and, hence, they are not natural to add as requirements for $<K,B>$. Thus, the *simple probability models* have to be specifed by the triple $\boldsymbol{J} = <K,B,\mu>$ where $K = <\mathbf{K},\mathcal{A}>$ is a simple probability structure, B a σ-algebra of subsets of **K**, and μ a probability measure invariant under $G_{\mathbf{K}}$; **K** is called the *set of outcomes of* $\boldsymbol{J}$.

3.2. Compound probability models. The compound probability structures introduced in Chuaqui 1980 need only minor modifications.

However, it may be noted that with the new definition of simple probability structures, the compound structures seem much more natural, since now the symmetries are independent of the language.

An outline of the definition of compound probability structure will help to understand the situation. These structures are determined by three elements.

1. The causal structure. The basic elements of the causal structure are the (causal) trees $<T,\leqslant_T>$. T is a set and $\leqslant_T$ is a well founded partial ordering on T such that the successors of any $t \in T$ are countable and well ordered by $\leqslant_T$. A graphical representation of an example of a tree is:



where $\leqslant_T$ is in the horizontal direction from left to right.

A tree is a generalization of the notion of causal dependence. Thus, what happens at $t \in T$ influences what happens at a succeding s, i.e. at $s_T \geqslant t$. If

s and t are not related by $\leqslant_T$, then they are independent moments. The elements of T can be considered in most cases as time moments, but this is not necessarily so.

A causal structure **F** consists of a family of trees that includes all its subtrees. ($\langle T,\leqslant_T\rangle$ is a subtree of $\langle S,\leqslant_S\rangle$ if $\leqslant_T \subseteq \leqslant_S$, and $t \in T$, $s \in S$ and $s \leqslant_S t$ imply $s \in T$.) In most cases, it is enough to consider all the subtrees of a given tree $\langle T,\leqslant_T\rangle$.

2. ***The set of outcomes.*** The outcomes are functions f with domain a certain $T \in \mathbf{F}$. For any $t \in T$, $f(t)$ is what happens at t, and is a member of the set of outcomes $\mathbf{H}(f,t)$ of a simple probability structure $\boldsymbol{H}(f,t)$. This simple probability model is determined by the preceding values of f, i.e. f restricted to $T_t = \{s : s \leqslant_T t \text{ and } s \neq t\}$. Calling $\mathbf{H}_T$ the set of all outcomes with domain T, the events are subsets of $\mathbf{H}_T$ for $T \in \mathbf{F}$. In fact, $\mathbf{H}(f,t) = \{j(t) : j \in \mathbf{H}_T$ and $j \restriction T_t = f \restriction T_t\}$ is the sure event determined by $f \restriction T_t$, i.e. the set of outcomes that are possible if $f \restriction T_t$ has occurred.

3. ***The symmetry relation.*** On each simple probability model $\boldsymbol{H}(f,t)$ we obtain a symmetry relation $\sim_{f,t}$ as explained above. From these relations, the symmetry relation $\sim$ between compound events is obtained. Since now $\sim_{f,t}$ is independent of the language, $\sim$ will also be. The relation $\sim$ is defined in several stages.

a) We first define isomorphism between two simple probability models $\mathcal{J} = \langle\langle \mathbf{K},\mathcal{O}\!\mathcal{l}\rangle,\mathcal{B},\mu\rangle$ and $\mathcal{J}' = \langle\langle \mathbf{K}',\mathcal{O}\!\mathcal{l}'\rangle,\mathcal{B}',\mu'\rangle$. Suppose A and A' are the universes of $\mathbf{K}$ and $\mathbf{K}'$ respectively.

$\mathcal{J} \cong_g \mathcal{J}'$ if and only if g is a one to one function from A onto A' satisfying:

(i) $\mathbf{K}' = \mathbf{K}^g$ (i.e. $\mathbf{K}'_{st} = \{g^*\mathfrak{L} : \mathfrak{L} \in \mathbf{K}_{st}\}$).

(ii) For any $\mathbf{B} \in \mathcal{B}$, $\mathbf{B}^g \in \mathcal{B}'$.

(iii) For any $\mathbf{B} \in \mathcal{B}$, $\mu(\mathbf{B}) = \mu'(\mathbf{B}^g)$.

In case μ is the unique $G_{\mathbf{K}}$-invariant measure, condition (iii) is implied by the other two, since then μ' is also the unique $G_{\mathbf{K}'}$-invariant measure. This is so because the groups $G_{\mathbf{K}}$ and $G_{\mathbf{K}'}$ are related by the isomorphism g as follows:

$$f \in G_{\mathbf{K}} \text{ iff } g \circ f \circ g^{-1} \in G_{\mathbf{K}'}$$

Let $\mathbf{B} \in \mathcal{B}$, $\mathbf{C} \in \mathcal{B}'$, then

$$\mathbf{B} \approx_g \mathbf{C} \text{ iff } \mathcal{J} \cong_g \mathcal{J}' \text{ and } \mathbf{B}^g \sim_{\mathbf{K}'} \mathbf{C}.$$

This definition of isomorphism constitutes the only difference with the definition in Chuaqui 1980, Section 5, of compound probability structures. We could use the same definition as there, but I believe that the new one is an improvement of the former; in particular because

$$\mathcal{J} \cong_g \mathcal{J} \text{ (i.e. with } \mathcal{J}' = \mathcal{J}) \text{ iff } g \in G_{\mathbf{K}}.$$

b) We now introduce the notion of isomorphism between sets $\mathbf{H}_T$ and $\mathbf{H}_{T'}$, for T, $T' \in \mathbf{F}$. These isomorphisms are pairs of functions $\langle h,k\rangle$ such that h is an isomorphism between $\langle T, \leqslant_T\rangle$ and $\langle T', \leqslant_{T'}\rangle$, and k is a one to one function from $\mathbf{H}_T$ onto $\mathbf{H}_{T'}$, with the following properties.

(i) If $t \in T$, f, $f' \in \mathbf{H}_T$ and $f \upharpoonright T_t = f' \upharpoonright T_t$ (i.e. what occurs before t is the same for f as for f'), then $k(f) \upharpoonright T'_{h(t)} = k(f') \upharpoonright T'_{h(t)}$ (i.e. what occurs before $h(t) \in T'$ is the same for $k(f)$ as for $k(f')$).

(ii) $H(f,t)$ and $H(k(f),h(t))$ are isomorphic (in the sense of (a)) for $f \in \mathbf{H}_T$ and $t \in T$. In fact, the corresponding isomorphism $g_{f,t}$ is such that if $\mathcal{L} = j(t)$ for a certain $j \in \mathbf{H}_T$ with $j \upharpoonright T_t = f \upharpoonright T_t$, then $g^*_{f,t} \mathcal{L}_{\text{st}} = (k(j)(h(t)))_{\text{st}}$. Notice that $j(t)$ is an outcome in $\mathbf{H}(f,t)$ and $k(j)(h(t))$, an outcome in $\mathbf{H}(k(f),h(t))$.

c) Now let $\mathbf{A} \subseteq \mathbf{H}_T$ and $\mathbf{B} \subseteq \mathbf{H}_{T'}$ for T, $T' \in \mathbf{F}$. We say that $\mathbf{A} \sim \mathbf{B}$ iff there is an isomorphism h of $\langle T, \leqslant_T\rangle$ onto $\langle T', \leqslant_{T'}\rangle$ and there are S, S' and k such that,

(i) $\langle S, \leqslant_S\rangle$ is a subtree of $\langle T, \leqslant_T\rangle$ and $S' = h^*S$.
(ii) $\mathbf{H}_S$ is isomorphic to $\mathbf{H}_{S'}$ by $\langle h \upharpoonright S, k\rangle$.
(iii) For every $t \in S$ and $f \in \mathbf{H}_S$ the corresponding parts of $\mathbf{A}$ and $\mathbf{B}$ by $\langle h,k\rangle$ are equivalent, i.e. if we define $\mathbf{A}(f,t) = \{j(t)\colon j \in \mathbf{A} \text{ and } j \upharpoonright T_t = f \upharpoonright T_t\}$ then $\mathbf{A}(f,t) \approx_{g_{f,t}} \mathbf{B}(k(f),h(t))$.
(iv) For $t \in T\text{-}S$ and $f \in \mathbf{H}_T$, $\mathbf{A}(f,t)$ is equivalent to the sure event at its level, i.e. $\mathbf{A}(f,t) \sim_{f,t} \mathbf{H}(f,t)$.
(v) Similarly, for $t' \in T'\text{-}S'$ and $f' \in \mathbf{H}_{T'}$, $\mathbf{B}(f',t') \sim_{f't'} \mathbf{H}(f',t')$.

The measure μ on the compound events should be invariant under $\sim$. In Chuaqui 1980, a procedure for defining such a measure is given. The algebra of compound events consists just of the measurable sets with respect to this measure.

As a first example of a compound probability structure, I shall take the case of independent trials of the same experiment. Assume that the experiment is modeled by a simple probability model $\mathcal{J} = \langle K, B, \mu\rangle$ and that there are n independent trials where $n \leqslant \omega$. We assume a symmetry relation is defined on $\mathcal{J}$, say $\sim_K$.

The causal structure for the compound model is constituted by the subtrees of $\langle n, =\rangle$. Since the trials are independent, there is no causal relation between them; thus, we take = for the partial ordering.

The compound set of outcomes is $^n\mathbf{K}$, where $\mathbf{K} = \langle \mathbf{K}, \mathcal{O}\!\ell\rangle$. For every $f \in {}^n\mathbf{K}$ and $t \in n$, $H(f,t) = \mathcal{J}$. Since $\langle n\rangle$ is a very simple tree, all $\langle m, =\rangle$ with $m \subseteq n$

are subtrees. Also, the only condition for the isomorphism of two such subtrees is that they have the same cardinality, and any isomorphism can be extended to an automorphism of $\langle n, =\rangle$, i.e. a permutation of n.

Events are subsets of $\mathbf{H}_m = {}^m\mathbf{K}$ for $m \subseteq n$. For any $\mathbf{A} \subseteq \mathbf{H}_m$, $f \in \mathbf{H}$ and $t \in m$ $\mathbf{A}(f,t) = \{j(t) : j \in \mathbf{A}\}$; thus $\mathbf{A}(f,t)$ is independent of f and $\mathbf{H}(f,t) = \mathbf{K}$.

Suppose that $\mathbf{H}_m$ and $\mathbf{H}_{m'}$ are given and that h is a one to one function from m onto m'. Let us analyze which k are such that $\langle h,k\rangle$ is an isomorphism of $\mathbf{H}_m$ onto $\mathbf{H}_{m'}$. k should be a one to one function from $\mathbf{H}_m$ onto $\mathbf{H}_{m'}$. Since for every $t \in m$, $T_t = \emptyset$, condition b(i) is always satisfied. Condition b(ii) requires that there be an isomorphism g_t of $\mathcal{J}$ onto itself (i.e. $g_t \in G_{\mathbf{K}}$) such that $g_t^*(j(t)_{st}) = (k(j)(h(t)))_{st}$ for every $j \in \mathbf{H}_m$. Thus, if $\mathbf{B} \subseteq \mathbf{H}_m$ and $\mathbf{C} \subseteq \mathbf{H}_{m'}$, then $\mathbf{B} \sim \mathbf{C}$ iff $\mathbf{B}(f,t) \sim_{\mathbf{K}} \mathbf{B}(f,h(t))$ for every $t \in m$.

Therefore, the compound measure is given by the product measure, and the compound events are the measurable sets according to this product measure.

As a second example, take $T = \{t_0, t_1\}$, a set of two elements, with $t_0 \leqslant_T t_1$, and its subtrees, as $\mathbf{F}$. For t_0, we have the simple probability model $\mathcal{J}_0 = \langle\langle \mathbf{K}_0, \mathcal{O}\!\ell_0\rangle, \mathcal{B}_0, \mu_0\rangle$. For each $\mathfrak{L} \in \mathbf{K}_0$, we have the simple probability model $\mathcal{J}_{\mathfrak{L}} = \langle\langle \mathbf{K}_{\mathfrak{L}}, \mathcal{O}\!\ell\rangle, \mathcal{B}_{\mathfrak{L}}, \mu_{\mathfrak{L}}\rangle$ The compound outcomes are the functions f with $D_0 f = T$ and such that $f(t_0) \in \mathbf{K}_0$ and $f(t_0) \in \mathbf{K}_{f(t_0)}$. Suppose, further, that $K_{\mathfrak{L}}$ is not isomorphic to $K_{\mathfrak{L}'}$, for $\mathfrak{L} \neq \mathfrak{L}'$.

The only subtree of $\langle T, \leqslant_T\rangle$ is $\langle\{t_0\}, \leqslant_T\rangle$. There is only one automorphism of $\langle T, \leqslant_T\rangle$, namely the identity.

It is clear that $H(f,t_0) = \mathcal{J}_0$ and $H(f,t_1) = \mathcal{J}_{f(t_0)}$ for every outcome f. Let us see which are the k's for which the pair $\langle$identity, $k\rangle$ is an isomorphism of $\mathbf{H}_T$ onto $\mathbf{H}_T$. k must be a one to one function from $\mathbf{H}_T$ onto $\mathbf{H}_T$; b(i) can be expressed by:

(1) if $f(t_0) = f'(t_0)$, then $k(f)(t_0) = k(f')(t_0)$,

(b-ii) adds two conditions:

(2) $g_{f,t_0} \in G_{\mathbf{K}}$

(3) $g_{f,t} \in G_{\mathbf{K}_{f(t_0)}}$.

Thus, if $\mathbf{B}, \mathbf{C} \subseteq \mathbf{H}_T$, then $\mathbf{B} \sim \mathbf{C}$ if and only if (i) and (ii), or (i) and (iii) are satisfied, where

(i) $\mathbf{B}(f,t_0) \sim_{\mathbf{K}_0} \mathbf{C}(f,t_0)$

(ii) For every $\mathfrak{L} \in \mathbf{K}_0$, we have $\mathbf{B}(\mathfrak{L}) \sim_{\mathbf{K}_{\mathfrak{L}}} \mathbf{C}(\mathfrak{L})$, where $\mathbf{B}(\mathfrak{L}) = \{f(t_1) : f \in \mathbf{B}, f(t_0) = \mathfrak{L}\}$.

(iii) For every $\mathfrak{L} \in \mathbf{K}_0$, $\mathbf{B}(\mathfrak{L}) \sim_{\mathbf{K}_{\mathfrak{L}}} \mathbf{K}_{\mathfrak{L}}$.

The compound measure μ is defined by

$$\mu(\mathbf{B}) = \int_{\mathbf{B}(f,t_0)} \mu(\mathbf{B}(\mathfrak{L}))\, d_{\mu_0}.$$

As an instance of this last example, assume that $\mathbf{K}_0 = \{\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3\}$ where $\mathcal{L}_i$ represents the choosing of urn i (i.e. there are three urns 1,2,3). In urn i there are n_i balls with m_i white balls. $\mathbf{K}_{\mathcal{L}_i}$ represents the choosing of a ball from urn i. The event "a white ball is chosen" is represented by the set of outcomes f, where $f(0) \in \mathbf{K}_0$ and $f(1) \in \mathbf{K}_{f(0)}$ is a model where a white ball is selected. Let us call this event W. Suppose that $\mu_0(\mathcal{L}_1) = \mu_0(\mathcal{L}_2) = \mu_0(\mathcal{L}_3)$. Then

$$\mu(W) = \int_{\mathbf{K}_0} \mu_{\mathcal{L}_i}(W(\mathcal{L}_i))\,d\mu_0$$
$$= (\mu_{\mathcal{L}_1}(W(\mathcal{L}_1)) + \mu_{\mathcal{L}_2}(W(\mathcal{L}_2)) + \mu_{\mathcal{L}_3}(W(\mathcal{L}_3))) \cdot \frac{1}{3}\ .$$

4. PROBABILITY AS A LOGICAL RELATION.

The justification of the Connecting Principle is given via the conception of probability as partial truth that was developed in Chuaqui 1977 and 1980. This is a logical conception of probability. We, thus, need a language $\mathcal{L}$.

A probability model $\mathcal{J} = \langle K, B, \mu \rangle$ is *appropriate for* $\mathcal{L}$ if the following two conditions are satisfied:

(i) The similarity type of the structures in K is the same as that of $\mathcal{L}$.

(ii) B includes the set $\{Mod_{\mathbf{K}}(\phi) : \phi$ a sentence of $\mathcal{L}\}$, where $Mod_{\mathbf{K}}(\phi) = \{\mathcal{L} \in \mathbf{K} : \mathcal{L} \models \phi\}$.

If $\mathcal{J}$ is an appropriate probability model for $\mathcal{L}$, we define, as in the earlier papers,

$$P_{\mathcal{J}}(\phi) = \mu(Mod_{\mathbf{K}}(\phi)),$$

for all sentences ϕ of $\mathcal{L}$.

P provides a logical interpretation of probability as a relation between a sentence and a probability structure $\mathcal{J}$, which is considered as an interpretation of the sentence. $P_{\mathcal{J}}(\phi)$ represents a measure of the "degree of partial truth of ϕ under the interpretation $\mathcal{J}$". Notice that $\mathcal{J}$ has a dual role. On one hand, it is a model of reality. On the other hand, it serves as an interpretation of the language. The usual relational structures (or possible models) of logic can also be seen in this dual role. But in this latter case, reality is completely specified and, hence, every sentence is either true or false.

This logical interpretation of probability can serve to justify the *Connecting Principle*, which is now reformulated as follows:

Let C_X be any reasonable "degree of belief" function of a person who accepts the proposition X that the probability structure $\mathcal{J}$ is an adequate description of the situation involved, and that $P_{\mathcal{J}}(\phi) = r$, for a certain

$r \in [0,1]$. Then

$$C_X(\phi) = r.$$

The person p should believe ϕ to the degree $P_J(\phi)$, because he believes ϕ to be true to this degree. Thus, the connection between factual probability and degree of belief is obtained via truth. This is very natural because we believe what we believe to be true. Thus, we should believe ϕ with degree α when we believe ϕ to be true with degree α. This is similar to the relation between usual logic and belief. If somebody believes ϕ and that ϕ logically implies ψ then he should believe ψ.

Degrees of belief should be applied to propositions instead of sentences. Thus, a more accurate description should involve Intensional Logic (such as that of Reinhardt 1980).

Notice that I put $C_X(\phi)$ and not $C(\phi|X)$. This is so, because I believe the acceptance of J does not change C by conditionalization (see Kyburg 1980 and Chuaqui 1980, Section 2).

My main differences with Bayesians (at least with strict Bayesians) are two. In the first place, I do not believe that probabilities (or degrees of belief) can be assigned to all events. Only given a well-defined situation in which the possible outcomes are determined, it may be possible to assign probabilities. In the second place, as it was mentioned before, I do not believe that the only changes in the probability (or degree of belief) function proceed by conditionalization. The discussion of these matters would take us too far, so they will be left for another paper.

5. A NOTE ON THE PRINCIPLE OF INSUFFICIENT REASON.

As an example of the application of the ideas given previously, I shall analyse the *Principle of Insufficient Reason* or *Principle of Indifference*:

"The Principle of Indifference asserts that if there is no *known* reason for predicating of our subject one rather than another of several alternatives, then relatively to such knowledge the assertions of each of these alternatives have an *equal* probability" (version of Keynes 1921, p.42).

This principle may be interpreted in two different ways: cognitive and factual. We can say that the equal probability is established when one *knows* of no reason or when *there are* in fact no reasons. The Principle as stated by Keynes (and also as stated by J. Bernoulli and Laplace) gives the first interpretation. In this form, it is indeed contradictory, as the well-known paradoxes show. However, a factual interpretation is also possible and may not be contradictory. In fact, a careful reading of Laplace 1820, leads me to believe that his intention was factual, although the actual wording is clearly cognitive.

I believe that my symmetry relation for events in probability structures is such a factual interpretation. In order to illustrate these ideas I shall give an analysis of Bertrand's paradox (Bertrand 1889). It is more convenient to introduce first the notion of random variable. Let $\mathbf{H}$ be the set of outcomes of a probability structure (simple or compound). Then a function $X: \mathbf{H} \to \mathbf{R}$ ($\mathbf{R}$ the set of real numbers) is a random variable if and only if $X^{-1*}A \in \boldsymbol{B}$ (the algebra of events) for every Borel set $A \subseteq \mathbf{R}$.

The experiment for Bertrand's paradox is the choosing of a chord at random in a circle and determining the probability distribution of the lengths of the chords. If the experiment is not further specified, we would have a simple probability structure $(\mathbf{K}, \langle A \rangle)$ where A is the set of chords and $\mathbf{K}$ contains all models of the form $\langle A,S \rangle$ where S consists of one element of A (the chord selected). We define the random variable X: $\mathbf{K} \to \mathbf{R}$ by $X(\langle A,S \rangle)$ = the length of the chord in S. The σ-algebra of events $\boldsymbol{B}$ should be $\{X^{-1*}A$: A Borel, $A \subseteq \mathbf{R}\}$. $G_{\mathbf{K}}$ is the set of all permutations of A that transform B into $\boldsymbol{B}$. In a sense, they are a counterpart of Borelian functions on $\mathbf{R}$. It is not difficult to prove that there is no measure invariant under $G_{\mathbf{K}}$.

So we should specify the experiment further. In fact, the origin of Bertrand's paradox arises from the fact that we can specify this experiment in several different ways which yield different distributions. Each of these ways can be put into the framework of my probability structures. I shall indicate how this can be done for two of these specifications.

(a) Choose two points on the circle and draw the chord between them. We have a compound probability structure with a causal structure consisting of the subtrees of the tree $\langle \{t_0,t_1\}, = \rangle$, i.e. $T = \{t_0,t_1\}$ consists of two independet elements. Let $K = \langle \mathbf{K}, \mathfrak{A} \rangle$ where $\mathbf{K}$ consists of all structures $\mathfrak{L}_0 = \langle C,r,O \rangle$ with C the points in the circle, r rotations, and O the selected point; $\mathfrak{A} = \langle C \rangle$. The experiment in question is modeled by $\mathbf{K}$ with the random variable X: ${}^T\mathbf{K} \to \mathbf{R}$ where $X(f)$ = the distance between the point selected at $f(t_0)$ and that selected at $f(t_1)$; $\boldsymbol{B} = \{X^{-1*}(A)$: A a Borel subset of $\mathbf{R}\}$. The group $G_{\mathbf{K}}$ corresponds to the rotations of the circle. Thus, there is an invariant measure.

(b) Select first a point on the circle and then a point on the radius through the first point. The chord chosen is the perpendicular to the radius through the second point. We are again in front of a compound probability structure consisting of the subtrees of $\langle T, \leqslant_T \rangle$ where $T = \{t_0,t_1\}$ has two elements and $t_0 \leqslant_T t_1$. At t_0 we have the same $\boldsymbol{K}$ as in (a). For each $\mathfrak{L}_0 \in \mathbf{K}$ we have $\boldsymbol{K} = \langle \mathbf{K}_{\mathfrak{L}_0}, \mathfrak{C} \rangle$ where $\mathbf{K}_{\mathfrak{L}_0}$ is the set of structures of the form $\langle D,t,I \rangle$, D contains the points on the radius passing through the point O, t the translations of the line modulus the radius, and I the point selected on the radius; $\mathfrak{C} = \langle D \rangle$. In order to complete the model we need a random variable Y: $\mathbf{H} \to \mathbf{R}$, where $\mathbf{H} = \{f: D_0 f = \{t_0, t_1\}, f(t_0) \in \mathbf{K}$ and $f(t_0) \in \mathbf{K}_{f(t_0)}\}$ is the set of outcomes, and $Y(f)$ is the length of the chord perpendicular to the radius selected at $f(t_0)$ and through the point selected at $f(t_1)$. The algebra of events is $\{Y^{-1*}A$: A a Borel subset of $\mathbf{R}\}$.

The group of $\mathbf{K}\mathfrak{L}_0$ is the group of translations; so again there is an invariant measure. By the usual analysis (see, e.g. Parzen 1960) we can obtain the distributions that are expected.

6. CLASSICAL AND BAYESIAN STATISTICAL INFERENCE REVISITED.

The purpose of this section is to improve the analysis of classical and Bayesian statistical inference given in Sections 3 and 4 of Chuaqui 1980, in the light of the modifications introduced so far.

First, I shall analyze classical inference. Let us assume that we have an experiment which can be repeated and we propose a simple probability model $\boldsymbol{J} = \langle\langle \mathbf{K}, \mathcal{O}\!\mathcal{l} \rangle, \boldsymbol{B}, \mu\rangle$ for it. This is now a factual model, so we can assume it as a scientific hypothesis. We then repeat the experiment n times for a large n and obtain a sequence of results. The probability of events consisting of sequences of results is computed by building the compound probability structure $^{\omega}\mathbf{K}$ and proceeding as in Chuaqui 1980. Recall that here we have a compound probability structure, with causal structure $\mathbf{F}$ composed of the subtrees of $\langle \omega, = \rangle$, i.e. all moments in ω are independent. Events are subsets of $^{\omega}\mathbf{K}$, the set of outcomes. If a compound event $\mathbf{A}$ occurs which has a low probability according to $^{\omega}\mathbf{K}$ and high according to another structure $^{\omega}\mathbf{K}'$, then we reject the original hypothesis that $\boldsymbol{J} = \langle\langle \mathbf{K}, \mathcal{O}\!\mathcal{l} \rangle, B, \mu\rangle$ is the adequate simple probability model.

This account is the same as that of Chuaqui 1980, and can be completed as there. What I would like to make precise is the type of simple probability models $\boldsymbol{J}$ that can be taken as hypothetical models. Suppose, first, that the experiment is that of tossing a coin. In this case, we have a complete physical explanation of the phenomenon and the model $\boldsymbol{J}$ can be built accordingly. However, there are many cases where the only known facts are frequencies observed in sequences of repeated experiments. Thus, the only natural $\boldsymbol{J}$ is one that just mimics the choosing of elements of a set. We have to assume that there is an unknown physical explanation for this way of choosing.

For instance, suppose that we observe that the relative frequency of the event is about 1/3. Then we should assume a $\mathbf{K}$ with models of the form $\mathcal{O}\!\mathcal{l}_0 = \langle A, E, O\rangle$ where A contains three elements, E (the event considered) contains one fixed element of A, and O (the chosen element) contains one element of A, different for each model. The intrisic part is $\langle A, E\rangle$ and the structural part $\langle A, O\rangle$.

If the relative frequency tends to $\sqrt{2}/2$, for instance, then our structures in $\mathbf{K}$ should be like that of a symmetrical roulette $<C,r,f,E,O>$, with E a fixed interval of length $\sqrt{2}/2$ and O the chosen element.

However, these types of models without any physical explanation are not completely satisfactory, because we do not know the mechanism for choosing O.

Thus, in a sense, the assumptions are not as well substantiated in the purely statistical case as when we have physical models. A case in point is the observed relation between cigarette smoking and cancer. The evidence is almost purely statistical, since there are no generally accepted physiological models for this phenomenon. I believe this is one of the reasons for the difficulties in accepting this connection as proven. The statistical evidence had to be overwhelming for the general public to accept that cigarrete smoking increases the chances of getting cancer.

Something similar is true for Bayesian inference. Here we have the causal tree $<\{t_0,t_1\} \leqslant_T>$ with $t_0 \leqslant_T t_1$, a simple probability model $\mathcal{J}_0 = <<\mathbf{K},\mathcal{O}_0>, B_0,\mu_0>$, and for each $\mathfrak{L} \in \mathbf{K}_0$, another simple probability model $\mathcal{J}_{\mathfrak{L}}$. We may assume as hypothesis any $\mathcal{J}_0$. The trouble here is that it is often the case that we have no evidence for $\mathcal{J}_0$ ($\mathcal{J}_0$ determines what are usually called "a priori" probabilities). Thus, these models might be less justified than the classical ones. Also, we might have evidence for a simple probability structure that admits no invariant measure. In this case, Bayesian methods cannot be used. Only if we have good evidence for a $\mathcal{J}_0$ that admits a probability measure, the method is perfectly adequate.

Given such probability models as the $\mathcal{J}_0$'s for the prior probabilities or the $\mathcal{J}$'s based only on frequencies, it is one of the aims of science to replace them by probability models based on physical laws.

REFERENCES.

J. Bertrand, 1889, *Calcul de Probabilitês*, París.

R. Chuaqui, 1977, *A semantical definition of probability*, Non-Classical Logics, Model Theory and Computability (Arruda, da Costa and Chuaqui eds.) North-Holland, Amsterdam, pp.135-167.

————————, 1980, *Foundations of statistical methods using a semantical definition of probability*, Mathematical logic in Latin America (Arruda, Chuaqui and da Costa eds.), Norht-Holland, Amsterdam, pp.103-120.

R.A. Fisher, 1956, *Statistical Methods and Scientific Inference*, New York.

J.M. Keynes, 1921, *A Treatise on Probability*, New York and London.

H.E. Kyburg, 1974, *Logical Foundations of Statistical Inference*, Reidel, Dordrecht-Holland.

————————, 1980, *Conditionalization*, The Journal of Philosophy, pp.98-114.

P.S. de Laplace, 1820, *Essai philosophique sur les probabilités*, París (English translation in Dover Publications, New York, 1951).

D. Lewis, 1980, *A subjectivist's guide to objective chance*, Studies in Inductive Logic and Probability, Vol. II (R.C. Jeffrey Ed.), University of California Press, Berkeley and Los Angeles, pp. 263-293.

E. Parzen, 1960, *Modern Probability Theory and its Applications*, Wiley, New York.

W.N. Reinhardt, 1980, *Satisfaction definitions and axioms of infinity in a theory of properties with necessity operator*, in Mathematical Logic in Latin America (Arruda, Chuaqui and da Costa eds.), North-Holland, Amsterdam, pp. 267-303.

***

Facultad de Matemáticas
Universidad Católica de Chile
Santiago, Chile.