ABSTRACT. This presents a characterization of random sequences, in the sense of Kolmogorov, Chaitin and Martin-Löf, using the notion of hypotheses tests developed in Chuaqui (1991).

This paper originated in a challenge made to me by Patrick Suppes to include random sequences in my interpretation of probability. In 1984, when I was a visiting scholar at IMSSS, Stanford University, I started writing my book *Truth, Possibility and Probability*, (Chuaqui, 1991). I have visited the Institute many times since, collaborating with Pat and profiting from his brilliant ideas; in particular, I worked on a project there during 1986–9. During this time and later, I had many opportunities to discuss my ideas with Pat and, although I do not think he agrees with many of them, he made many useful observations that helped me to improve greatly the final version of the book. One such observation, whose solution did not get into the book, was that, for my theory to be complete, I needed to give an account of the complexity based definition of random sequences, as developed by Kolmogorov, Chaitin and others. After the book was sent to the publisher, I obtained a characterization of random sequences, which I presented on March 1991, in a preliminary version, at a conference on random sequences organized by Pat. This is the characterization discussed in this paper.

The first section summarizes my ideas on probability, which are developed in full in Chuaqui (1991). Section 2 gives my version of hypotheses tests, and the last section gives the characterization of random sequences.

1. TRUTH, POSSIBILITY AND PROBABILITY

I begin with a short informal survey of the main ideas behind the interpretation of probability that is fully developed in (Chuaqui 1991). The basis for the interpretation of probability is based on three elements:
(1) *Probability is based on an objective notion.* Although I believe that there is room for a subjective notion of probability, especially in decision theory, I find the basic notion to be that of chance or factual probability. Factual probability is determined by the real possibilities, which I accept as properties of the objects and the experimental conditions. The set of possibilities for a given situation is represented mathematically by a model $K$, which is a set with a certain structure.

(2) *Probability is the degree of possibility of truth.* In addition to the objective notion of chance, I believe that there are two epistemic notions: degree of support and degree of belief. I consider the epistemic probability of a proposition $\varphi$ primarily as the degree of support of $\varphi$ given that we accept a model $K$ for the phenomenon involved, and only secondarily, as the degree of belief. The degree of support of a proposition $\varphi$ given that $K$ is an adequate model of the situation is the measure of the set of possibilities (i.e., members of $K$) where the proposition is true.

(3) *The classical definition of probability is essentially correct.* The measure on the different sets of possibilities (i.e., on subsets of $K$) is based on a notion of equiprobability, just as the classical definition is. The main advances over the classical definition are that a precise definition of equiprobability is given, and that models are constructed for most situations where probability is applied. My conception of equiprobability is based on symmetries given by groups of transformations.

Chance is a property of *chance setups*, using the terminology of (Hacking, 1965). The only essential characteristic for a setup to be subject to chance is that it must have a unique result that is a member of a fixed set of really possible results. In essence, a chance setup consists of a set of objects plus a set of conditions under which these objects have certain real possibilities.

An *outcome* of a chance setup includes, in my theory, a description of the chance setup plus a particular result – one of the factually possible results of the setup. Thus, an outcome codes the mechanism of the setup and the specific result that occurs in the outcome. A chance setup can then be identified with its set of possible outcomes – one possible outcome for each possible result.

Events are sets of outcomes. An event occurs if one of its outcome elements obtains. It is natural to think that an event has a greater chance
of occurrence if it contains 'more' outcomes. The chance of an event is then the measure of the set of possible outcomes that constitute the event. Each possible outcome in an event represents an indivisible real possibility for the event to happen, with each outcome having the same possibility as any other. Each real possibility has, so to speak, a certain propensity to occur and we assume that the ultimate possibilities represented by the outcomes have the same force. In other words, each outcome has the same tendency to occur.

A feature of my approach is that chance applies to single events, not necessarily as a part of a sequence of similar experiments. It is possible to look at this definition of factual probability as an explanation for single-case propensities. Propensity is usually taken to be a real kind of power, force, or tendency. Each possibility for an event to occur represents added power for its occurrence. Thus, if there are more possibilities, there is more power; and equal possibilities determine equal power.

Because of the incorporation of the mechanism of the setup in each outcome, the set of possible outcomes is all we need to know about a chance setup for deciding which probability space is appropriate for it. Thus, the set of possible outcomes for a particular setup should determine the family of events and the probability measure that are adequate for the setup.

Although, for our purpose in this paper, the method for determining the probabilities is irrelevant, for the sake of completeness I shall give a brief account. A complete account appears in Chuaqui (1991). The first step for the determination of probabilities is a detailed analysis of what an outcome is, and the delimitation of the properties and relations that are relevant for the setup. Outcomes can have many different properties. Looking at the example of the roll of a die, we are usually interested in the number that obtains on the upper face, but not in the distance to a certain wall or its color. In formalizing the set of possible outcomes, we include just the relevant properties. Some properties, however, may turn out to be relevant in my account, although in the usual probabilistic formulations they are not considered explicitly. For instance, in the roll of a biased die, the distribution of the weights in the die is relevant. It is possible to include the relevant properties by representing an outcome by a set-theoretical model, in the sense of the logical theory of models.

Another element that should be included in the description of the setup, and hence in the structure that represents the outcomes, is the
relations of stochastic dependence and independence among the different parts of the results, when these results are compounded of simpler ones. Suppose, for instance, that the setup consists of the choosing of an urn at random, and then the choosing of a ball from this urn. The choosing of the urn determines the range of possibilities for the choosing of the balls. So we see that dependence is dependence on the range of possibilities. On the other hand, when we toss a coin twice, the range of possibilities for the second toss does not depend on the result of the first toss; in this case we have independence. These notions of dependence and independence are represented by partial orderings.

The description of a chance setup is given by a set of systems, each representing a possible outcome. The probability measure is defined to be the measure (if it exists and is unique) that satisfies the following two conditions. In the first place, it should be invariant under the transformations that preserve the setup. These transformations are the permutations of the objects involved in the setup that transform a possible outcome into another possible outcome. Invariance under these transformations is a precise rendering of an objective principle of symmetry. This principle is objective, because the group of transformations that determines the symmetries is obtained from the set of possible outcomes, which, in its turn, depends only on the chance setup, which is objective.

In the second place, the measure should preserve the relations of dependence and independence included in the outcomes. This preservation of dependence is, in fact, another invariance principle: Two chance setups which have the same form should assign the same probabilities to corresponding events. Two chance setups $K_1$ and $K_2$ are called homomorphic, i.e., they have the same 'form', if there is a function (the homomorphism) that transforms $K_1$ into $K_2$, preserving the dependence and internal invariance structure of the setups. If an event $A$ of $K_1$ contains all the outcomes whose images by the homomorphism form an event $B$ of $K_2$ (i.e., $A$ is the inverse image by the homomorphism of $B$), then the probability of $B$ in $K_2$ should be the same as the probability of $A$ in $K_1$.

The family of events is also determined by the set of systems representing the possible outcomes of the setup. Events are those sets of possible outcomes whose measure is determined by the group of symmetries and the relation of homomorphism.
As I show in my book, most (possibly all) setups where a probability measure has been obtained and used in the past, can be modeled by these methods. A general representation theorem is also proved: *any stochastic process can be approximated, up to an infinitesimal, by one of my models.*

Another side of probability is the degree of support that a proposition $\varphi$ has relative to a certain model, given by a set of possible outcomes $K$, which I write $\Pr_K(\varphi)$. This degree of support, $\Pr_K(\varphi)$, is understood as the *degree of possibility of truth* (or, more informally of partial truth) of $\varphi$ in $K$. The probability $\Pr_K(\varphi)$ is defined to be the (invariant and homomorphism-preserving) measure of the set of outcomes in $K$ that makes $\varphi$ true; i.e., the degree of support of $\varphi$ given $K$ is the same as the measure of the chance of the event consisting of the outcomes in $K$ where $\varphi$ is true. If $\varphi$ is true whenever an outcome in $K$ occurs, we say that $\varphi$ is true in $K$, the set of outcomes that make $\varphi$ true is $K$ itself, and hence $\Pr_K(\varphi) = 1$. When $\varphi$ is false whenever an outcome in $K$ occurs we say that $\varphi$ is false in $K$, the set of outcomes that make $\varphi$ true is empty, and hence $\Pr_K(\varphi) = 0$. For the cases in between, $\Pr_K(\varphi)$ is a real number between 0 and 1. In this way, the degree of support of $\varphi$ given $K$ is rendered as the degree of ‘possibility of truth’ of $\varphi$ in $K$. So a proposition $\varphi$ is ‘more possible to be true’ in $K$ than another proposition $\psi$, if the set of possibilities where $\varphi$ is true is larger (in the sense of the measure) than the set of possibilities where $\psi$ is true.

My theory also makes sense of the following principle:

**Principle of direct inference.** *Let $A$ be an event. Then the degree of support of $A$’s holding, given that the chance of $A$ is $r$, is also $r$. Hence, the degree of belief of a person who believes that the chance of $A$ is $r$, should be also $r.*

This principle can now easily be justified with my account. The degree of possibility of truth of $\varphi$ in $K$ is the same as the chance of the event $A$ that consists of the outcomes in $K$ where $\varphi$ is true. Since we may take $\varphi$, in this case, as the proposition that $A$ holds, we have that the degree of support of $A$’s holding (i.e., of $\varphi$) given that the chance of $A$ is $r$ (this chance is obtained from $K$), is $r$. One can also justify the second part of the principle: if one is rational, one should believe that $\varphi$ in the same measure that one has support for $\varphi$ being true.
When studying pure probability (what may be called probabilistic logic), we relativize probability to different sets $X$ of possible outcomes. Each $X$ is a possible model, which, instead of determining the truth-value of a proposition, spans its possibilities, and, in so doing, defines its probability. In order to apply the theory, it must be assumed that $X$ is an accurate description of the factual possibilities in the real world. That is, we must accept $X$ as true to the world. Just as for the models that determine the truth or falsehood of all relevant propositions, there are rules for rejection and acceptance of $X$ as a representation of the real world, which will be discussed later in the paper.

We now analyze decision theory and statistical inference. Decision theory and statistical inference have different bases. In decision theory we use probabilities for deciding on an action, while in statistical inference proper, we use probability in order to accept or reject hypotheses.

For decision theory, with which we will not be concerned here, I accept essentially a Bayesian account. In this case, we accept a certain probabilistic model for the situation involved and we assign the probabilities according to this model. The model itself is not in question.

I believe, on the other hand, that statistical inference proper is inference, that is, it contains rules that allow us to pass from propositions accepted as true to other propositions that are to be accepted as true. The main difference with classical logical inference is that instead of arriving at a definitive acceptance of a proposition, we can only accept propositions provisionally. Our acceptance is always subject to future revision.

Classical statistical inference, in fact, gives rules for the provisional rejection or acceptance of hypotheses. These rules, however, are not based on the probabilities of the hypotheses themselves, but on the probability that is assigned according to the different hypotheses to other propositions. I argue in Chuaqui (1991), in agreement with classical statistical inference, that it is impossible to accept or reject hypotheses, even provisionally, based on their probabilities.

2. RULES FOR REJECTION OF HYPOTHESES

I shall analyze briefly the main ideas for rules of rejection. For the justification of these rules, we only need, of the theory I have summarized, the notion that probability is a measure of the degree of possibility of
truth. The method for determining probabilities that was explained in the previous section is irrelevant.

A probabilistic hypothesis is the proposition that a set of possible outcomes, say $K$, is an adequate model of reality, i.e., of the chance setup that is in question. For briefness of expression, I shall call this hypothesis just $K$. This hypothesis $K$ determines a probability distribution over the propositions (or, simply, over the events) we are interested in. It is clear that if a proposition false in $K$ were true in reality, we should reject $K$. But usually propositions false in $K$ do not have too much content for probabilistic hypotheses, and hence they are hard to find true in reality. So a natural rule of rejection would be to reject $K$, if the next best thing to a proposition false in $K$ obtained in reality, namely, an approximately false proposition.

Since I consider probability as a measure of the possibility of truth, a proposition $\phi$ is approximately false in $K$, if $\phi$ has low probability in $K$, i.e., if its degree of support given $K$ is low. Hence, we would assume that if an event for which $K$ determined a low probability value occurred, then we should reject $K$. This rule, however, does not work. Suppose that the chance setup for which $K$ is a model of a lottery. The drawing of any particular ticket has very low probability according to this model $K$, but such a drawing does not constitute reasons at all for rejecting $K$. The reason for this failure is that, in order to reject $K$, we must have some grounds to question it; namely, we must have some alternative models in mind for the chance setup. We are forced to deal with classes of alternative hypotheses $K_i$, for each $i$ in a certain index set $I$. All these hypotheses are compatible with what we accept as the laws of the phenomenon in question. How to decide which is the right class of alternative hypotheses is not determined, in general, by statistical or probabilistic considerations.

Now we can improve the definition of an approximately false proposition. A proposition $\phi$ is approximately false in $K_j$, relative to the class $K_i$, for $i$ in $I$, if $\phi$ has much lower probability in $K_j$, than in a certain $K_i$, with $i$ in $I$ and $i \neq j$. This definition, however, as we shall see later, does not quite work, so we only accept it provisionally, as a basis for discussion.

In order to declare $K_j$ false, however, just the occurrence of an unlikely event is not enough. We may, in case an unlikely event occurred, provisionally reject $K_j$, but we should be able to revise this judgement. Thus, we must have at hand a sequence of unlikely events, whose prob-
abilities approach zero. If we have such a sequence, then, in principle, we could get, with enough effort or time, the occurrence of an event with a probability as low as we want, and we would also be able to revise our judgement. But, as before, we also need to consider the alternatives. Hence, we need to be able to find a sequence \( \varphi_n \) of propositions such that the probabilities of \( \varphi_n \) in \( K_j \) tend to zero, and on an alternative model \( K_i \) tend to one. This sequence of propositions should be in principle decidable. That is, there should be a way of deciding whether each \( \varphi_n \) is true in reality or not. Thus, we should have at hand a sequence of experiments or, simply, observations, say \( E_n \), which determine the truth or falsity of \( \varphi_n \) in reality. What must be fixed in advance is the sequence of experiments \( E_n \). The exact sequence of propositions \( \varphi_n \) might change in the course of the study. Thus, the important aspect of the test is the sequence of experiments \( E_n \).

The point in dealing with sequences of probabilities approximating zero is not that we can get in this way propositions with low probabilities, but with arbitrarily low probabilities. That is, if we wait long enough or spend enough effort, we can get a proposition with as low a probability as we want. In this way, we can set in motion a dialectical process: one sets the low probability at which one would be satisfied, and the opponent (which may be the same person) tests the right proposition. In order to have a real dialectical process, we must have an effective sequence of propositions whose probabilities in \( K_j \) tend to zero, and in an alternative \( K_m \), tend to one. That is, given an \( n \), we should be able to calculate the proposition \( \varphi_n \) of the sequence, and to indicate the experiment \( E_n \) which decides \( \varphi_n \) in reality. In actual practice one is able only to get down to a certain level, i.e., one can perform \( E_n \) only up to a certain \( n \). This accounts for the fact that, although we may reject a hypothesis and consider it false, this rejection is provisional. If new propositions obtain with a high probability according to the hypothesis and low, according to an alternative, we may be forced to accept it again.

We proceed now to introduce the notion of an unlikely result. The following example is a good introduction to our notion of an unlikely result. Suppose that we are tossing a coin ten times and recording the frequency of heads. Given that the coin is unbiased, what is an ‘unlikely result of the experiment’? The result of exactly five heads has low probability, but we would not call it unlikely. Suppose that eight heads are obtained. We call this result unlikely (with respect to an unbiased coin), if the event of obtaining at least eight heads or at most
two, has low probability. That is, the event of eight heads or worse (i.e., less likely according to the hypothesis that the coin is unbiased) has low probability. We shall use this idea about an unlikely event for our rule of rejection.

We proceed now to delimit more precisely the notion of an unlikely result. In order to do this, we need to introduce two notions: the notion of ‘evidential equivalence’, and the notion of ‘worse result’. We assume that the alternative hypotheses are $K_i$, for $i \in I$. We begin with the first of these notions. We notice the following fact. Suppose that there are two possible results of $E_n$, $a$ and $b$, satisfying the condition: there is a $c > 0$ such that for every $i \in I$

\[ (*) \quad Pr_{K_i}[E_n = a] = c \cdot Pr_{K_i}[E_n = b]. \]

This means that, for any $i, j \in I$,

\[
\frac{Pr_{K_i}[E_n = a]}{Pr_{K_j}[E_n = a]} = \frac{Pr_{K_i}[E_n = b]}{Pr_{K_j}[E_n = b]},
\]

provided that $Pr_{K_j}[E_n = a] \neq 0$ and $Pr_{K_j}[E_n = b] \neq 0$.

An example of this situation is the following. Suppose that we are tossing a coin 60 times and that the hypotheses are that the coins have different biases. Two different sequences with the same number of heads, have the same probability under any of the hypotheses. Thus, (*) is true for this case with $c = 1$.

A natural way of measuring the degree to which one probability is smaller than another is their ratio. Thus, that

\[
r = \frac{Pr_{K_j}(\varphi)}{Pr_{K_i}(\varphi)}
\]

is low for a certain $i$ in $I$ gives evidence against $K_j$. It is not necessary to require that $Pr_{K_j}(\varphi)$ be low and $Pr_{K_i}(\varphi)$ be high, but only that their ratio be low. For instance, if we had that $\varphi$ was impossible under $K_j$, but possible under $K_i$, we would reject $K_j$, although the probability under $K_i$ may be very low.

Thus, in a situation such as (*), $a$ and $b$ do not discriminate between the different models. That is, the evidence they give with respect to the models is the same. We say, in case (*) is satisfied, that $a$ and $b$ are evidentially equivalent with respect to $E_n$ and the systems $K_i$, for
In the example of the coin, then, any two sequences with the same number of heads are evidentially equivalent. It is clear that the relation of evidential equivalence is an equivalence relation between possible results. We write the equivalence class of $a$ as $[a]$. Since all elements of an equivalence class have the same evidential import, we have to consider them together. Thus, we could replace the experiment $E_n$ by another $E'_n$ such that $E'_n = [a]$ instead of $E_n = a$. In the case of the coin, again, the new experiment, instead of having as its result the actual sequences, would have the number of heads in the sequence. Notice that $[E'_n = [a]]$ can also be written as $[E_n \in [a]]$. In statistical practice, this replacement of $E$ by $E'$ is usually done.

We turn now to the second notion mentioned above. We call a possible result $b$ as least as bad for $K_j$ at $E_n$ as a possible result $a$, in symbols $b \leq a$, if $Pr_{K_j}[E_n \in [b]] \leq Pr_{K_j}[E_n \in [a]]$ and there is an $i \in I$ where the inequality is reversed. Then, for rejection of $K_j$, a partial experiment $E_n$ must have a result $a$ such that the probability under $K_j$ of $E_n$ having the value $a$ or any result as least as bad as $a$ is small, say less than a certain $\alpha$. Thus, we arrive at the notion of “unlikely result”: a result, $a$, of an experiment $E_n$ is unlikely for $K_j$, if the probability of the event of $E_n$ having value $a$ or any value at least as bad for $K_j$ as $a$, is low under $K_j$. I think this is a natural characterization of unlikely results.

3. HYPOTHESES TESTS

In what follows, we shall consider our statistical models as just probability distributions. Thus, we assume as alternative hypotheses a set $\Omega$ of probability distributions. How the distributions are obtained is not our concern for the problems of this paper.

In hypotheses tests, we test a hypothesis $H_0$, called the null or working hypothesis, considering at the same time an alternative hypothesis $H_1$. The working hypothesis $H_0$ and the alternative hypothesis $H_1$ serve to partition the space $\Omega$. Under $H_0$, $\theta$ lies in a subspace $\Omega'$; under $H_1$, $\theta$ lies in the complementary subspace $\Omega - \Omega'$.

$$H_0 : \theta \in \Omega' \quad \text{and} \quad H_1 : \theta \in \Omega - \Omega'.$$

The purpose of a hypothesis test is to determine whether $H_0$ or $H_1$ is consistent with the data. Thus, accepting $H_0$ means simply that we
are not in a position of rejecting $H_0$, i.e., that $H_0$ is consistent with
the data. Similarly, rejecting $H_0$, and hence, accepting $H_1$, means that
$H_0$ is inconsistent with the data, and hence, with respect to $H_1$, it only
means that $H_1$ is consistent with the data.

I now begin with the justification of hypotheses tests, according to
the view which I espoused in Chuaqui (1991). Although I consider
hypotheses tests to be valid, my justification is different from that of
the developers of these tests, Neyman and Pearson.

The most important feature of hypotheses tests, according to my point
of view, is that the test statistics must be a function of a discriminating
experiment, which we now begin to define. For simplicity, we write
$\Omega_0 = \Omega'$ and $\Omega_1 = \Omega - \Omega'$. We assume that $E = \{E_n\}_{n=1}^\infty$ is a
sequence of random variables on each of the models in $H_i$, $i = 0, 1$. We
begin with the notion of evidentially equivalent results, which is just
the notion already discussed, adapted to the new situation. Let $r$ and $s$
be possible results of $E_n$. Then $r \sim_n s$ if there is a $c > 0$ such that for
every $\theta \in \Omega$

$$Pr_\theta[E_n = r] = c \cdot Pr_\theta[E_n = s].$$

As before, $[r]$ is the equivalence class of $r$.

We also need the notion of worse results, which we discussed informally above. We say that $r$ is at least as bad as $s$, for $\theta \in \Omega_i$ at $E_n$, in
symbols $r \preceq_{\theta_n} s$, if

$$Pr_\theta[E_n \in [r]] \leq Pr_\theta[E_n \in [s]].$$

and there is $\theta \in \Omega_j$, with $j \neq i$, such that the inequality is reversed.

The rejection set for $\theta$, with $\theta \in \Omega$, determined by the result $r$ of $E_n$
is then

$$R_{n\theta r} = \{s \mid s \preceq_{\theta_n} r\}.$$  

The probability $Pr_\theta[E_n \in R_{n\theta r}]$ is the $p$-value of the test for $\theta$ with
result $r$.

In order to shorten some of the definitions, we introduce the following
expression. Let $E_n$, for $n \in N$, be a sequence of random variables, and
let $A$ be a set of infinite sequences of possible results of $\{E_n\}_{n=1}^\infty$. We
say that $\{E_n\}_{n=1}^\infty$ is almost surely (a.s.) eventually in $A$, if almost surely
there is a sequence $\{r_n\}_{n=1}^\infty \in A$, such that $E_n = r_n$, for all $n \in N$. 


That is, the set of possible outcomes where there is no such sequence \( \{ r_n \}_{n=1}^{\infty} \) in \( A \) is a null set:

\[
\Pr \left[ \{ E_n \}_{n=1}^{\infty} \in A \right] = 1.
\]

We also say that \( \{ E_n \}_{n=1}^{\infty} \) is *almost surely eventually not in \( A \), if almost surely for no sequence of possible results, \( \{ r_n \}_{n=1}^{\infty} \in A \), we have that \( E_n = r_n \), for all \( n \in N \). That is

\[
\Pr \left[ \{ E_n \}_{n=1}^{\infty} \not\in A \right] = 1.
\]

We are now ready to define our experiments. We say that the system \( E = \{ E_n \}_{n=1}^{\infty} \) is a *discriminating experiment (d.e.) for \( H_0 \) against \( H_1 \) if

1. The sequence \( E_n, n \in N \), is a sequence of random variables over the \( n \)-product space of the spaces in \( H_0 \) and \( H_1 \).
2. For each \( \theta \in \Omega_1 \), there is a set \( A_\theta \) of sequences of results such that \( \{ E_n \}_{n=1}^{\infty} \) is almost surely eventually in \( A_\theta \) according to \( \theta \), and \( \{ E_n \}_{n=1}^{\infty} \) is almost surely eventually not in \( A_\theta \) according to \( \theta' \), for every \( \theta' \in \Omega_0 \), and such that for any \( \{ r_n \}_{n=1}^{\infty} \) in \( A_\theta \), and any \( \alpha > 0 \) there is an \( n \) such that for every \( \theta' \in \Omega_0 \) and every \( m \geq n \)

\[
\Pr_{\theta'} [ E_m \in R_{m\theta'r_m} ] \leq \alpha.
\]

That is, the \( p \)-value for \( \Omega_0 \) of the test can be made as small as one wishes.

3. For each \( \theta \in \Omega_0 \) there is a set \( B \) of sequences of results such that \( \{ E_n \}_{n=1}^{\infty} \) is a.s. eventually in \( B \), according to \( \theta \), and a.s. eventually not in \( B \), according to \( \theta' \), for all \( \theta' \in \Omega_1 \), and such that for any \( \{ r_n \}_{n=1}^{\infty} \) in \( B \), any \( \theta' \in \Omega_1 \) and any \( \alpha > 0 \), there is an \( n \) such that for every \( m \geq n \)

\[
\Pr_{\theta'} [ E_m \in R_{m\theta'r_m} ] \leq \alpha.
\]

The rule of rejection for this case, which is dialectical in nature, is now:

**Dialectical rule of rejection of \( H_0 \) against \( H_1 \).** Let \( \Omega \) be a set of possible probability models for a setup, and suppose that we are considering the null hypothesis \( H_0 : \theta \in \Omega' \) against \( H_1 : \theta \in \Omega - \Omega' \). We say that \( H_0 \) should be provisionally rejected at level of significance \( \alpha \), if there is a discriminating experiment \( \{ E_n \}_{n=1}^{\infty} \) for \( H_0 \) against \( H_1 \), such that

\[
\Pr_{\theta'} [ E_m \in R_{m\theta'r_m} ] < \alpha.
\]
(1) \[ E_n = a \] obtains, for some \( a \) and \( n \in \mathbb{N} \).

(2) \( \Pr_\theta [E_n \in R_{n\theta a}] \leq \alpha \), for every \( \theta \in \Omega' \), i.e., the p-value must be uniformly small for every \( \theta \in \Omega' \).

4. RANDOM SEQUENCES

As might be expected from the preceding remarks, I do not agree with the frequentist view of probability. Inverting von Mises' dictum, my position may be summarized as: 'first the probability and then the random sequences'. A quote from my book will make my ideas clear (Chuaqui, 1991, p. 66; see also Chuaqui, 1991, p. 284):

It seems to me that there is an objective notion of randomness. When we say that we are choosing a ball at random from an urn, we are stating a property of the chance setup consisting of the selection of a ball. It means that it is just as likely for one ball to be chosen as for any other. In general, an outcome is random, if it is an outcome of a setup with equiprobable outcomes, i.e., outcomes which are symmetric with respect to the group of symmetries of the setup. Since the group of invariance is determined by objective properties of the setup, randomness is also objective. In a similar way, a sequence is random, if it is obtained in a chance setup with sequences as outcomes and where all outcomes are equiprobable, i.e., symmetric under the group of symmetries.

Thus, for me, a sequence is random if it is produced by a random process, and a process is random if it is a chance setup which assigns the same probability to each possible outcome.

I believe, however, that the Kolmogorov-complexity definition is interesting in its own right. The main purpose of the present paper is to give an equivalent definition of Kolmogorov-complexity random sequences in terms of hypotheses tests, in the sense that these tests were defined in the previous section. This will give sense to random sequences according to my account of probability, solving, at least partially, Pat Suppes' problem mentioned at the beginning of the paper. My account is related to the characterization in Martin-Löf (1966), but I use only one type of test, which is a hypotheses test, and not all sequential tests, which are more similar to significance tests.

I shall mainly use the notation and exposition of Kolmogorov-complexity random sequences contained in van Lambalgen (1987, Ch. 5). As in this reference, we write \( 2^n \) for the set of binary sequences of length \( n \), \( 2^\omega \), for the set of binary infinite sequences, and \( 2^{<\omega} \), for the set of finite sequences.
We begin with the Kolmogorov–Chaitin definition of random sequences, see van Lambalgen (1987, pp. 118–122). We follow the definitions given first in Chaitin (1975). Turing machines are assumed to have worktapes, a read-only input tape and a write-only output tape. Furthermore, we constrain the reading head to read the input in one direction only and we do not allow blanks as endmarkers. These are called *prefix algorithms*. Let \( U \) be a universal prefix machine. For \( x \in 2^n \) we write

\[
I(w) = \min\{|p| \mid U(p) = w\},
\]

where \(|p|\) is the length of \( p \).

We also consider algorithms that produce infinite sequences, for which we also take a universal machine, \( U_\infty \), with a domain that is prefix-free. In this case, the machine scans the input tape, and after the last bit of the input \( p \), it starts again. For \( x \in 2^\omega \), we define

\[
I_\infty(x) = \begin{cases} 
\infty & \text{if there is no } p \text{ such that } U_\infty(p) = x, \\
\min\{p \mid U_\infty(p) = x\}, & \text{otherwise.}
\end{cases}
\]

If we have an input \( p \) that computes \( x \in 2^\omega \), then knowing \( p \) plus \( n \), one can compute \( x(n) = (x_1, x_2, \ldots, x_n) \). Thus

\[
I(x(n)) \leq I_\infty(x) + I(n) + d,
\]

for a certain constant \( d \). We write \( \omega_p \) for the infinite sequence generated by input \( p \) on \( U_\infty \).

We shall need the following counting lemma. The proof of (1) and (2) can be found in Chaitin (1975, p. 337). The proof of (3) is easy and can be found in van Lambalgen (1987, pp. 121, 122).

**LEMMA 1.** (1) For some constant \( c \), and for every \( k \leq n \)

\[
\#\{w \in 2^n \mid I(w) \leq k + I(n)\} \leq 2^k \cdot c.
\]

(2) For some constant \( c \), and for every \( k \leq n \)

\[
\#\{w \in 2^n \mid I(w) > k + I(n)\} > 2^n \left(1 - \frac{2^k}{2^n} \cdot c\right).
\]
We have
\[ \sum_{w \in 2^\omega} 2^{-I(w)} \leq 1. \]

We also need (van Lambalgen 1987, Lemma 5.1.2.9):

**Lemma 2.** Let \( f : \omega \to \omega \) be a total recursive function. Then

1. If \( \sum_n 2^{-f(n)} = \infty \), then for all \( m \), there is an \( n \geq m \), such that \( I(n) > f(n) + m \).
2. If \( \sum_n 2^{-f(n)} < \infty \), then there exists an \( m \), such that for all \( n \), \( I(n) \leq f(n) + m \).

We have the following corollary:

**Corollary 3.** If \( a > 1 \) (and computable), then for some \( c \), \( I(n) \leq \log_2 n + c \). Hence, for every \( m \), there is an \( n_0 \), such that for every \( n \geq n_0 \), \( n - m \geq I(n) \).

We shall be mainly interested in the product measure, \( \lambda \), generated by the uniform measure defined on \( \{0, 1\} \). When \( w \) is a finite sequence and \( \mu \) a measure on infinite sequences we shall write \( \mu(w) = \mu([w]) \), where, in this context, \( [w] \) is the cylinder generated by \( w \). We denote by \( R(\mu) \) the set of infinite sequence that are random according to the probability measure on \( 2^\omega \), \( \mu \), in the sense of Martin-Löf (1966). We only need the following characterization, which could be adopted as a definition.

**Theorem 4.** Let \( x \in 2^\omega \) and \( \mu \) be a computable measure. Then \( x \in R(\mu) \) if and only if there is an \( m \), such that, for all \( n \), \( I(x(n)) > -\log_2(\mu(x(n))) - m \). In particular, \( x \in R(\lambda) \) if and only if there is an \( m \), such that, for all \( n \), \( I(x(n)) > n - m \).

This characterization has been credited to Schnorr and Solovay. The first published proof, according to van Lambalgen (1987), appears to be in Dies (1976). A proof also appears in van Lambalgen (1987, p. 139). We have that, if \( \mu \) is an ergodic measure (for instance, if \( \mu \) is the product measure corresponding to \( \mu(0) = p \), for a certain \( p \in [0, 1] \),
then \(- \log_2(\mu(x(n)))\) is of order \(n\), for almost all \(x\). On the other hand, by Corollary 3, \(I(n)\) is \(o(n)\). Using these facts it is not difficult to prove that, if \(x \in R(\mu)\), then \(x \neq \omega_p\), for all finite inputs \(p\). However, I have not been able to prove the converse, which I conjecture is not true.

In the case of \(\lambda\), we have a characterization of \(R(\lambda)\) using hypotheses tests. We shall test the hypothesis \(H_0\) against \(H_k\), where \(H_0\) is the hypothesis that the infinite sequence \(x\) is produced by a random process with the uniform probability \(\lambda\), and \(H_k\), for \(k = 1\) or \(2\) or \(\ldots\) is the hypothesis that the infinite sequence \(x\) is produced by \(U_\infty\) with an input \(p\) of length \(|p| \leq k\). A similar, but much simpler case with only one sequence produced by a machine, was discussed in Chuaqui (1991, pp. 104, 105).

Recall that \(\omega_p\) is the sequence produced with input \(p\), and we call \(Pr_p\) the probability induced over finite and infinite sequences by input \(p\). This probability is obviously determined for \(w \in 2^n\), by \(Pr_p(w) = 1\) if and only if \(w = \omega_p(n)\). Thus, \(H_k\) consists of all \(Pr_p\), for input \(p\), with \(|p| \leq k\). The hypothesis \(H_0\) is simple, consisting of only one probability distribution, \(Pr_{H_0}\), which is equal to \(\lambda\), the uniform distribution. We write, for any of these probabilities, \(Pr[E_n \in A] = Pr A\), where \(A \subseteq 2^n\) and \(E_n\) is a random variable on \(2^n\). Similarly, \(Pr[E \in B] = Pr B\), if \(B \subseteq 2^\omega\) and \(E\) is a random variable on \(2^\omega\).

As we noticed above, we have:

**Proposition 5.** For a certain constant \(d\), if \(|p| \leq k\), then \(I(\omega_p(n)) \leq k + I(n) + d\), and, if \(|p| \leq k\) and \(I(x(n)) > k + I(n) + d\), then \(x(n) \neq \omega_p(n)\).

**Discriminating Experiment**

The experiment that we need is the sequence of \(E_n\) for \(n \in \mathbb{N}\), where \(E_n\) is the binary sequence obtained up to \(n\). We write \(E_\Omega\) for the infinite sequence which is the union of the \(E_n\); thus, \(E_n\) is an element of \(2^n\), \(E_\Omega\), of \(2^\omega\), and \(E_\Omega(n) = E_n\).

It is not difficult to prove directly from the definition of \(R(\lambda)\) with sequential tests (see van Lambalgen 1987, p. 60) that

\[
Pr_{H_0}[E_\Omega \in R(\lambda)] = 1.
\]
Since $\lambda$ is nonatomic

$$\Pr_{H_0}[E_{\Omega} \in \{\omega_p\}] = 0,$$

for every input $p$. On the other hand, by Theorem 4 and Proposition 5

$$\Pr_p[E_{\Omega} \in R(\lambda)] = 0,$$

and, obviously

$$\Pr_p[E_{\Omega} \in \{\omega_p\}] = 1.$$

The set $B$ in clause (3) of the definition of a discriminating experiment is, thus, $R(\lambda)$, and the set $A_p$, for $Pr_p \in H_k$ in clause (2) of this definition, is $\{\omega_p\}$. In order to verify that $E$ satisfies these two conditions for discriminating experiments (2) and (3), we first need the notions of evidential equivalence and worse results.

**Evidential Equivalence**

If $x(n), x'(n) > k + I(n) + d$, then

$$x(n) \sim x'(n),$$

because

$$\Pr_{H_0}[E_n = x(n)] = \lambda(x(n)) = \lambda(x'(n)) = \Pr_{H_0}[E_n = x'(n)],$$

and

$$\Pr_p[E_n = x(n)] = 0 = \Pr_p[E_n = x'(n)],$$

for any input $p$ with $|p| \leq k$.

Besides, if $I(z(n)) \leq k + I(n) + d$, then, by Proposition 5, $z(n) = \omega_p(n)$, for a certain $p$ with $|p| \leq k$. Thus

$$\Pr_p[E_n = z(n)] = 1,$$

and

$$\Pr_p[E_n = z'(n)] = 0,$$

for any $z'(n) \neq z(n)$. So that $z(n)$, with $I(z(n)) \leq k + I(n) + d$, is not evidentially equivalent to any other sequence in $2^n$. 
We have two cases:

Case 1. $I(x(n)) > k + I(n) + d$. Then, since $[x(n)] = \{w \in 2^n \mid I(w) > k + I(n) + d\}$, and $\lambda(x(n)) = 1/2^n$, by Lemma 1, part (2)

$$\Pr_{H_0}[E_n \in [x(n)]] > 2^n \left(1 - \frac{2^k}{2^n} c\right) \lambda(x(n))$$

$$= 2^n \left(1 - \frac{2^k}{2^n} c\right) 2^{-n} = 1 - \frac{2^k}{2^n} c,$$

for some constant $c$, independent of $n$ and $k$ (the constant $c$ absorbs the constant $d$). So, since $R_{nH_0x(n)} \supseteq [x(n)]$

$$(**) \quad \Pr_{H_0}[E_n \in R_{nH_0x(n)}] > 1 - \frac{2^k}{2^n} c.$$

Case 2. $I(x(n)) \leq k + I(n) + d$. We have

$$\Pr_{H_0}[E_n = x'(n)] = \lambda(x'(n)) = 2^{-n},$$

for all $x'$, in particular for those with $I(x'(n)) \leq k + I(n) + d$. Thus, since $[x(n)] = \{x(n)\}$, we have, for sufficiently large $n$

$$\Pr_{H_0}[E_n \in [x(n)]] = \Pr_{H_0}[E_n = x(n)]$$

$$= 2^{-n} < \Pr_{H_0}[E_n \in [z(n)]],$$

for $I(z(n)) > k + I(n) + d$. Hence, $z(n)$ is not as least as bad as $x(n)$ for $H_0$. Therefore

$$(***) \quad R_{nH_0x(n)} \subseteq \{w \in 2^n \mid I(w) \leq k + I(n) + d\}.$$  

We now verify conditions (2) and (3) of the definition of a discriminating experiment.

(2) The hypotheses in $H_k$ are of the form $Pr_p$, for $|p| \leq k$. As we mentioned above, $A_p = \{\omega_p\}$. If $x(n) = \omega_p(n)$, then, by (***)

$$\Pr_{H_0}[E_n \in R_{nH_0x(n)}] \leq \Pr\{w \in 2^n \mid I(w) \leq k + I(n) + d\}.$$  

By Lemma 1, part (1) this is: $\leq 2^k c \lambda(x(n)) = 2^{k-n} c$, for a certain constant $c$, which absorbs the constant $d$. Thus, the $H_0$-probability of the rejection set can be made as small as one wants by increasing $n$. 


(3) Now consider \( H_0 \). Take \( B = R(\lambda) \). Let \(|p| \leq k, x \in B \), and take \( m \) such that, for all \( n \), \( I(x(n)) > n - m \). For any \( n \), we have \( I(\omega_p(n)) \leq k + I(n) + d \). Let \( n_0 \) be so large that \( n - m > k + I(n) + d \), for all \( n \geq n_0 \) (Corollary 3). Then, for any \( n \geq n_0 \)

\[
I(\omega_p(n)) \leq n - m < I(x(n)),
\]

and, hence, \( x(n) \neq \omega_p(n) \). Therefore

\[
\Pr_p[E_n \in [x(n)]] = 0 < \alpha,
\]

for any \( \alpha > 0 \).

Finally, we have the following main theorem:

**Theorem 6.** For every \( x \in 2^\omega \), \( x \in R(\lambda) \) if and only if there is an \( m \) such that for every \( n \), \( [E_n = x(n)] \) does not reject \( H_0 \) against \( H_{n-m-I(n)-d} \) at level \( 2^{-m} \).

**Proof.** Let \( x \in 2^\omega \). We use Theorem 4: \( x \in R(\lambda) \) if and only if there is an \( m \), such that for all \( n \), we have \( I(x(n)) > n - m \).

Suppose, first, that \( x \in R(\lambda) \). Find \( m \) large enough so that for \( n \), \( I(x(n)) > n - m \) and \( 1 - (2^k/2^n)c > 2^{-m} \), with \( k = n - m - I(n) - d \). Then, for all \( n \), \( I(x(n)) > k + I(n) + d \), and by \((**)\)

\[
\Pr_{H_0}[E_n \in r_{nH_0}x(n)] \geq 1 - \frac{2^k}{2^n} c \geq 2^{-m}.
\]

That is, \( x(n) \) does not reject \( H_0 \) against \( H_{n-m-I(n)-d} \) at level \( 2^{-m} \).

Suppose, now, that \( x \notin R(\lambda) \). Then, for every \( m \), there is an \( n \) (which must satisfy \( n \geq m \)), such that \( I(x(n)) \leq n - m \). Let \( m \) be given. Let \( n \geq m \) be such that \( 1 - (2^k/2^n)c > 2^{-m} \), where \( k = n - m - I(n) - d \). Then, for sufficiently large \( n \), by \((***)\)

\[
\Pr_{H_0}[E_n \in r_{nH_0}x(n)] \leq \lambda\{w \in 2^n \mid I(w) \leq k + I(n) + d\} \\
\leq \sum\{\lambda(w) \mid w \in 2^n, I(w) \leq k + I(n) + d\} \\
\leq \sum\{2^{-n} \mid w \in 2^n, I(w) \leq k + I(n) + d\}.
\]

Since \( I(w) \leq n - m(= k + I(n) + d) \) if and only if \( 2^{-n} \leq 2^{-m}2^{-I(w)} \), we have

\[
\leq \sum\{2^{-m}2^{-I(w)} \mid w \in 2^n, I(w) \leq k + I(n) + d\}
\]
\[
= 2^{-m} \sum \{2^{-I(w)} \mid w \in 2^n, I(w) \leq k + I(n) + d\}
\]
\[
\leq 2^{-m} \sum_{w \in 2^{<\omega}} 2^{-I(w)},
\]
and, by Lemma 1, part (3)
\[
\leq 2^{-m}.
\]

Thus, for all \(m\), there is an \(n_0\), such that, for all \(n \geq n_0\), \(x(n)\) rejects \(H_0\) against \(H_{n-m-I(n)-d}\) at level \(2^{-m}\). \(\square\)

ACKNOWLEDGEMENT

Research supported by DIUC Grant No. 89041.

Facultad de Matemáticas,
Pontificia Universidad Católica de Chile,
Casilla 306, Santiago 22, Chile

REFERENCES


COMMENTS BY PATRICK SUPPES

I have had so many discussions in recent years about the foundations of probability with Rolando that I feel my reading of his paper and the comments I have written about it are just a continuation of what has gone before, although it is true that recently we have talked much more about
infinitesimal analysis than about probability. To begin with, I want to say that I think his book *Truth, Possibility and Probability* (1991) is the best piece of work yet done on the logical foundations of probability, broadly conceived. The details of his theory have moved a long way from those of Keynes, Carnap and others working this line of territory. I am too much of a Bayesian to buy the whole story, but a great many of the things he has worked out are appealing. Of particular importance is his use of groups of symmetries to revive the classical Laplacean definition of probability. The formal way in which he uses groups of symmetries is one that Bayesians could accept, and in practice do accept in standard cases. The only real point of difference is not in the formal derivation of the prior distribution, when groups of symmetries dominate the situation, but rather in the epistemic status of these symmetries. For Rolando, the symmetries are evidently objective, but for Bayesians they would be a matter of belief. However, this difference is not as great as it seems, for under many conceptions, the Bayesian beliefs about symmetries would have the same status as beliefs about Newton’s laws or other physical laws.

The point of substantive difference is that Bayesians like myself are skeptical about being able to find enough symmetries to really believe in to fix the prior distribution uniquely. The world, for me, is just too irregular and too crotchety in its detailed features to permit a realistic analysis in terms of equal probabilities for a great variety of practical problems.

On the other hand, Rolando’s view that statistical inference proper is inference – a point he stresses at the end of Section 1 – is very much along the same line of argument that Bayesians use. It is a standard Bayesian view that given a prior distribution, as new evidence comes forth everything is just a matter of computation given the new premises embodying the new evidence. For the Bayesian, in a strict sense there are not new inductive principles, but just direct applications of Bayes’ theorem, which is a deductive principle of inference.

Of course, as Rolando comes to Section 2 on the analysis of statistical hypotheses, including rules for their acceptance or rejection, he points out that the specific method for determining probabilities is essentially irrelevant, a position that is increasingly clear from the great agreement on many aspects of testing hypotheses by Bayesians and objectivists or frequentists. This does not mean that there are not fine points of difference, but if one examines the use of hypothesis testing in the
scientific literature, it is fairly hard in most cases to drive a serious wedge between the different foundational views of probability as to what hypotheses are to be rejected or accepted.

I do feel that Rolando needs to incorporate a great deal more of the standard theory and practice of statistics into his theory of hypothesis testing. At this late date, it is surprising, for example, to read Rolando’s treatment of the subject with so little discussion of such important ideas as those of likelihood ratio of hypotheses, when the ideas he develops are very close to the standard doctrine. I have, however, put this matter too simply. In the final pages of his book Rolando rejects the maximum likelihood principle (1991, pp. 422–424). He does not do so hastily and although I cannot agree with him on this matter, the discussion is too intricate to begin afresh in this context.

For different reasons I, too, have been wary of the likelihood principle, but the reason is not the likelihood principle itself, but its consequences when only one or two simple models are tested. A typical example is in looking at sequences of heads and tails. If we apply the maximum likelihood principle to estimate the probability $p$ of a head on the hypothesis that we are observing a Bernoulli sequence with unknown $p$, the way we analyze the data does not permit us to take any look at dependencies from trial to trial or changes in $p$ over time, both matters of great concern in real data. So my criticism is that we are often led by the likelihood principle to look at data in too summary a way and not at all the pluralistic features that, from a scientific standpoint, we should well be concerned with. In fact, for almost all complex data sets, overall tests of acceptance or rejection of hypotheses are not usually of great importance or taken too seriously. What is important are particular features, and of course these particular features may be tested in many cases by something like a maximum likelihood principle.

**Random Sequences.** Although Rolando seems to be standing von Mises’s and others’ ideas on their head in his dictum that “first the probability and then the random sequences”, this really only applies to von Mises and no one in practice takes very seriously that probability is going to be constructed from random sequences. Moreover, von Mises’ characterization is not generally regarded as sufficiently strong. What is needed is the stronger Kolmogorov complexity definition which implies von Mises’ characterization but not conversely.
The last section of Rolando’s paper is a very nice development within his framework of a theory of random sequences that is essentially equivalent to Kolmogorov’s. For this purpose he has been able to use some of the results in Lambalgen’s 1987 thesis and subsequent monograph on random sequences. (I do want to remark that Lambalgen’s own viewpoint is much more sympathetic to von Mises than Rolando’s.)

My one serious comment about random sequences in the present context is that the real problem for statistical practice is how to think about the randomness of finite sequences. It is obvious enough in the massive literature on randomness that this is fundamentally a completely different topic from that of randomness in infinite sequences, where standard limiting operations can be used. It is especially in the theory of finite random sequences that, in my own judgment, the Bayesian view is much more likely to be realistic and therefore correct, than an objectivist view like Rolando’s. The reason is easy enough to state without trying to expand on the technical considerations. The tests of randomness in finite sequences must necessarily be partial in some sense for if we ran all possible tests, clearly the only finite sequences that would pass all tests would be those that had extreme probabilities of 0 or 1. Thus we can never be sure that by some tricky procedure a sequence has not been generated that has a relatively short program for its generation, but one that is unlikely to be discovered in usual or even unusual tests of randomness in finite sequences. It is the Bayesian, who takes account of where the sequence came from, how it was generated, etc., that is most likely to end up without egg on his face, or at least so I think.

In other words, I think that objectivists and Bayesians can come out with very close agreement on randomness in the idealized case of infinite sequences. There need be no real differences of opinion. The serious differences only arise as we turn to finite sequences. This is undoubtedly a topic that Rolando and I will devote a couple of seminars to sometime in the future.

REFERENCE