THE THEORY OF PROBABILITY FROM THE POINT OF VIEW OF ADMISSIBLE NUMBERS

By Arthur H. Copeland

I. INTRODUCTION

The definition of the word probability has never been agreed upon. Before we decide on a definition, let us first consider what use we hope to make of the theory of probability. It is reasonable to demand of this theory that we shall be able to apply it, and that, by means of it, we shall be able to make predictions.

If we say that the probability is .9 that a given event will occur under certain circumstances, then are we making some prediction about the success (i.e. occurrence) of the event? Let us suppose that the circumstances are presented. We may observe that the event succeeds or we may observe that it fails. Whichever the case may be, the result of the experiment cannot be interpreted in terms of the number, .9. This is always the case. We can never interpret the result of a single trial of an event in terms of the probability of that event.

Next let us assume that n trials are made of an event whose probability is .9, and that, as a result of this experiment, r successes and $n-r$ failures are obtained. If n is large, we should expect the ratio, r/n, to be approximately .9, that is, approximately nine-tenths of the trials to be successful. We shall call the number, r/n, the success ratio.

We have not even now obtained a satisfactory interpretation for the number, .9. We have not specified any limit to the discrepancy between the numbers, r/n and .9, and we have not specified the magnitude of n. Thus, if r/n differs from .9 by a small amount, it also differs from .899 by a small amount. Are we to be satisfied with the statement that the event in question has a multiplicity of probabilities including the numbers, .9 and .899?

We can make the above statement more exact as follows:

Given any positive number, $\mathcal{E}$, we can find a number, n, such that the discrepancy between r/n and .9 is less than $\mathcal{E}$. After the number, $\mathcal{E}$, has been chosen, it is at least conceivable that a sufficient number of trials can be made so that r/n will differ from .9 by less than $\mathcal{E}$. If we make this interpretation of the probability, .9, and if we wish to make the statement that .9 is *the* probability of the event, then we are assuming that .9 is the only number that has this property. We are therefore assuming that the ratio, r/n, approaches .9 as n becomes infinite.

So far as I know, no one has ever given an alternative concept of probability which is capable of being interpreted in terms of the result either of a single trial or of a sequence of trials. Unless and until such a concept is given, we are compelled to assume that probability is the limit of the success ratio, if we wish to include an emperical interpretation. Since this paper is being presented to a group of statisticians, I think it will be safe to assume that we are agreed that probability is concerned with the results of trials of events.

It may be that we arrived at the probability, .9, by means of the following reasoning. There are 9 possible ways in which the event can succeed and 1 in which it can fail. All 10 possibilities are equally likely and mutually exclusive.

When we make a trial of the event, one and only one of the possibilities succeeds. The words, equally likely, have no interpretation in terms of the result of a single trial. The reader will have little difficulty in continuing the analysis of these words in a manner similar to that of the concept of probability. In fact the concept, being equally likely, is identical with the concept, having the same probability. We shall, therefore, reject the concept of equal likelihood as a basis for a definition of probability.

There is one other objection to this method of finding the probability of an event. Namely, there is good reason to believe that it never gives the correct result. In making this statement we are assuming, of course, that probability is defined as the limit of the success ratio. In order that the 10 possibilities may be equally likely, it is necessary that there be perfect symmetry between these possibilities. We cannot, therefore, have any mark to distinguish the one unfavorable possibility from the other nine favorable possibilities. Experiment indicates that such distinguishing marks are sufficient to make noticeable differences in the probabilities. For example, the dots on the faces of a die cause differences in the frequencies with which the respective faces turn up.

In spite of these objections, the above method of finding the probability of an event, gives very good approximations in most of the cases where it is applied. There is no method which gives exact values for probabilities. It seems wise not to reject this method, but rather to discard any illusions which we may have concerning the exactness of its results.

We have seen that we must assume the probability of an event to be the limit of the success ratio, if we are agreed that probability is concerned with the results of trials. Let us express this assumption in terms of the Cauchy criterion for the existence of a limit.

Given a positive number, $\mathcal{E}$, there exists a number, N, such that $|r/n - r'/n'| < \mathcal{E}$ whenever $n \geqq N$ and $n' \geqq N$, where r is the number of successes in n trials and r' is the number of successes in n' trials. Physical experiment seems to indicate that this condition is satisfied. Furthermore, if we reject this assumption we deny the possibility of experimental verification of probabilities. On the other hand, it can be proved that the number, N, can never be known. This situation is unsatisfactory for a mathematical theory.

To avoid this difficulty we shall construct an imaginary idealized universe in much the same manner as is done in the case of geometry. This universe will contain sequences of successes and failures which are formed in accordance with mathematical laws. These sequences will satisfy the fundamental assumptions of probability and hence will be infinite. We make the assumption that the physical universe is an approximation to this idealized universe.

II. THE ALGEBRA OF EVENTS

We shall show how the elements of the theory of probability can be treated from the point of view which we have described. Consider first the following physical example. A coin is flipped ten times and the event in question is the occurrence of a head. The following is a record of the successes and failures,

$$1,\ 1,\ 0,\ 1,\ 0,\ 0,\ 0,\ 1,\ 0,\ 0$$

where the 1's stand for successes and the 0's for failures. The ratio, $4/10$, of the number of successes to the number of trials, is obtained by adding all of the ten numbers and dividing by ten. If we had made a much larger number of trials of the event, we should expect that the corresponding success ratio would have been much closer to the probability, one-half.

The above sequence of 1's and 0's can be interpreted as a number written in the binary scale. Let us write

$$.110,100,010,0$$

This number has the value. $1/2+1/4+0/8+1/16+0/32+0/64+0/128+1/256+0/512+0/1024=209/256$. We should not, however, think of this number as ending with the tenth digit. In fact we could compute as many more of the digits as we desired by continuing the experiment. The computation of the values of these numbers will not be important for our purposes. The above computation was inserted merely to aid in the understanding of the notation which we shall describe.

We shall now consider the construction of our idealized universe. The sequence of successes and failures of a given imaginary event can be represented by a number, $x=.x^{(1)}x^{(2)}x^{(3)}\ldots x^{(k)}\ldots$, written in the binary scale, the kth digit, $x^{(k)}$, of x being 1 or 0 according as the event succeeds or fails on the kth trial. We shall denote the success ratio for the first n trials of this event by $\rho_n(x)$.

Then

$$(1) \qquad \rho_n(x)=\sum_{k=1}^{n} x^{(k)}/n$$

We shall denote the probability of the event, x, by $p(x)$ and we shall define $p(x)$ by means of the equation

$$(2) \qquad p(x) = \lim_{n \to \infty} p_n(x).$$

We are, of course, assuming that this limit exists.

Most of the important questions in the theory of probability involve relations between different events. We shall therefore construct an algebra which is especially adapted to the discussion of related events. If $x = .x^{(1)}x^{(2)}x^{(3)}\dots$ and $y = .y^{(1)}y^{(2)}y^{(3)}\dots$ are any two events, then the event, x and y, will be denoted by $x \cdot y$. We have the equation,

$$(3) \qquad x \cdot y = .(x^{(1)} \cdot y^{(1)}), (x^{(2)} \cdot y^{(2)}), (x^{(3)} \cdot y^{(3)}) \dots .$$

The first digit of $x \cdot y$ is 1 if and only if the first digits of x and y are both 1. That is, the event, $x \cdot y$, succeeds on the first trial if and only if x and y both succeed on the first trial. Similarly for the second and third trials etc. The expressions inside the parentheses are understood to be ordinary algebraic products. The expression, $x \cdot y$, is a symbolic product.

The event, x or y or both, is denoted by $x \vee y$. We have the equation

$$(4) \qquad x \vee y = .(x^{(1)} + y^{(1)} - x^{(1)} \cdot y^{(1)}), (x^{(2)} + y^{(2)} - x^{(2)} \cdot y^{(2)}), \dots$$

It will be observed that the first digit, $(x^{(1)} + y^{(1)} - x^{(1)} \cdot y^{(1)})$ of $x \vee y$ is 1 if $x^{(1)} = y^{(1)} = 1$ or if $x^{(1)} = 1$, $y^{(1)} = 0$ or if $x^{(1)} = 0$, $y^{(1)} = 1$, but that this digit is 0 if $x^{(1)} = y^{(1)} = 0$. Thus the event, $x \vee y$, succeeds on the first trial if x succeeds on its first trial or y succeeds on its first trial or both x and y succeed on their first trials. Similarly for the second and third trials etc.

We shall use the symbol, $\sim x$, to denote the event, not x. It is easily seen that $\sim x$ is given by the equation,

$$(5) \qquad \sim x = .\,(1-x^{(1)}),\,(1-x^{(2)}),\,(1-x^{(3)}),\,\cdots\,.$$

Let us denote the event, y if x, by $y \subset x$.* Before attempting to give a formula for $y \subset x$ let us first consider the expression, $m \cdot p_m(x)$. This expression is equal to the number of successes of the event, x, in its first m trials. Thus if m_n is the number of the trial on which the nth success of x occurs, then

$$(6) \qquad m_n \cdot p_{m_n}(x) = n.$$

We can write

$$(7) \qquad y \subset x = .\,y^{(m_1)}\,y^{(m_2)}\,y^{(m_3)}\cdots$$

Thus we consider those trials of y for which the event, x, occurs. In other words we consider a given trial of y if (and only if) x occurs on that trial. Hence equation (7) gives us the correct expression for the event, y if x.

[*The operators, $\cdot$, $\vee$ and $\sim$ are also used in symbolic logic with similar interpretations. See Whitehead and Russell, *Principia Mathematica*, vol. 1. The symbol, $\subset$, is an inverted implication sign. The expression, $y \subset x$, could be read, y is implied by x, or, y if x. For the benefit of those who are familiar with *Principia Mathematica*, it may be added that the symbols, x, y, etc. are propositional functions rather than propositions. Each x is associated with a sequence of events, and each is a propositional function of the form, the kth event will succeed, k being a free variable. The probability is a property of the set of propositions rather than of any given proposition. Thus we should speak of the probability of a propositional function rather than of the probability of a proposition.]

PROBLEMS

In problems, 1 to 3, assume that x and y have the following values

$$x = .110,100,010,011,101,000,10\cdots$$
$$y = .110,111,011,000,100,010,11\cdots$$

1. Compute $p_{15}(x)$ and $p_{20}(y)$.
2. Compute the first 20 digits of (a) $x\cdot y$, (b) $x\vee y$, (c) $\sim x$, (d) $y\cdot\sim x$.
3. Compute as many digits as possible of $y\subset x$ and $x\subset y$.
4. Prove the following identities:

(a) $x\cdot y = y\cdot x$	(g) $\sim\sim x = x$
(b) $x\cdot(y\cdot z) = (x\cdot y)\cdot z$	(h) $\sim(x\cdot y) = \sim x\vee\sim y$
(c) $x\vee y = y\vee x$	(i) $\sim(x\vee y) = \sim x\cdot\sim y$
(d) $x\vee(y\vee z) = (x\vee y)\vee z$	(j) $x\vee\sim x = 1$
(e) $x\cdot(y\vee z) = (x\cdot y)\vee(x\cdot z)$	(k) $(x\cdot y)\vee(x\cdot\sim y) = x$
(f) $x\vee(y\cdot z) = (x\vee y)\cdot(x\vee z)$	

5. Prove that $p_n(x\vee y) = p_n(x) + p_n(y) - p_n(x\cdot y)$
6. Prove that $p(x\vee y) = p(x) + p(y) - p(x\cdot y)$
7. Prove that $p(\sim x) \leq 1 - p(x)$
8. Prove that $p[y\cdot\sim x] = p(y) - p(x\cdot y)$
9. Prove that if $x\cdot\sim(y\vee z\vee w) = 0$ then $x = (x\cdot y)\vee(x\cdot z)\vee(x\cdot w)$

III. THE COMPUTATION OF PROBABILITIES

We shall say that two events, x and y, are mutually exclusive provided x fails whenever y occurs and y fails whenever x occurs. It is easily seen that x and y are mutually exclusive if and only if $x \cdot y = 0$. It follows from problem (6) that

$$(8) \qquad p(x \vee y) = p(x) + p(y) \quad \text{if} \quad x \cdot y = 0.$$

If we have three events, x, y, and z, which are mutually exclusive, then $x \cdot y = y \cdot z = z \cdot x = 0$. Hence

$$p(x \vee y \vee z) = p(x \vee y) + p(z) = p(x) + p(y) + p(z).$$

We have the following theorem.

Theorem 1. If the events, $x_1, x_2, x_3, \ldots x_n$, are mutually exclusive then

$$p(x_1 \vee x_2 \vee \cdots \vee x_n) = p(x_1) + p(x_2) + p(x_3) + \cdots + p(x_n).$$

Suppose we have a set of events, $x_1, x_2, \cdots x_n$, such that at least one of the events must occur. Then $x_1 \vee x_2 \vee x_3 \vee \cdots \vee x_n = 1$. Suppose further that these events are mutually exclusive and that their probabilities are equal. Then $p(x_1) + p(x_2) + \cdots + p(x_n) = 1$ and therefore $p(x_1) = p(x_2) = \cdots = p(x_n) = 1/n$. This principle is very useful in the computation of probabilities.

Example 1. From a pack of 52 cards 1 card is drawn. What is the probability that this card is the ace of spades? It is reasonable to assume that the probability of drawing any one of the 52 cards, is the same as that of drawing any other card. Thus we have 52 events which have the same probabilities. Moreover these events are mutually exclusive and it is a certainty that at least one of the events will occur. Hence the desired probability is $1/52$.

Example 2. From a pack of 52 cards, 13 cards are drawn. What is the probability that these cards are all spades? We assume that any combination of 13 cards has the same probability as any other combination of 13 cards. Since there are ${}_{52}C_{13}$ such combinations, the probability is $1/{}_{52}C_{13} = 1/635{,}013{,}559{,}600$.

We shall now compute the probability of the event, $y \subset x$.

We have the equations,

$$(9)\quad p_n(y \subset x) = \frac{\sum_{i=1}^{n} y^{m_i}}{n} = \frac{\sum_{k=1}^{m_n} \frac{x^k y^k}{m_n}}{\frac{n}{m_n}} = \frac{p_{m_n}(x\cdot y)}{p_{m_n}(x)}$$

where $m_n \cdot p_{m_n}(x) = n$.
If we allow n to become infinite we get

$$(10)\qquad p(y \subset x) = p(x\cdot y)/p(x).$$

Multiplying both sides of equation (10) by $p(x)$ we get

$$(11)\qquad p(x)\cdot p(y\subset x) = p(y\cdot x).$$

Example 3. A pack of 52 cards is divided into 4 piles of 13 cards each. One pile contains just 1 heart and the other 3 piles contain 4 hearts each. A pile is selected at random and a card is drawn from this pile. What is the probability that the pile selected will be the one containing just the one heart and that the card selected from this pile will be the heart? Let y represent the drawing of a heart and x represent the drawing of the pile containing just one heart. Then $p(x) = 1/4$ and $p(y \subset x) = 1/13$. Hence $p(y\cdot x) = p(x)\cdot p(y\subset x) = 1/52$. This is the desired probability.

We shall say that an event, y, is independent on an event, x, provided the probability that y will occur is the same whether x occurs or not. If we express this condition for independence in terms of our symbols we will get

$$(12)\qquad p(y\subset x) = p(y \subset \sim x).$$

Hence

$$(13)\qquad \frac{p(y\cdot x)}{p(x)} = \frac{p(y\cdot \sim x)}{p(\sim x)} = \frac{p(y) - p(y\cdot x)}{1 - p(x)}$$

Therefore

$$(14)\qquad p(x\cdot y) = p(x)\cdot p(y)$$

It is a simple matter to reverse our steps and start with equation (14) and obtain equation (12). Moreover, from the symmetry of equation (14) it is easily seen that if y is independent of x, then x is independent of y. We have now proved the following theorem.

Theorem 2. A necessary and sufficient condition that two events, x and y, be independent, is that $p(x\cdot y)=p(x)\cdot p(y)$.

Example 4. A coin and a die are thrown together. What is the probability that the coin will turn up a head and the die will turn up a 3? Let x represent the occurrence of a head and y represent the occurrence of a 3. Then $p(x)=1/2$ and $p(y)=1/6$. Since the events are independent it follows that $p(x\cdot y)=1/12$.

It should be observed that equation (11) is always true but that equation (14) can only be used when the two events are independent. The term, contingent, is used to apply to events which are not independent. If x and y are two contingent events we must use equation (11) to compute $p(x\cdot y)$.

In order that three events, x, y, z, may be independent, it is necessary and sufficient that $p(x\cdot y)=p(x)p(y)$,

$$p(y\cdot z)=p(y)p(z),\ p(z\cdot x)=p(z)p(x),\ p(x\cdot y\cdot z)$$
$$=p(x)p(yz)=p(y)p(z\cdot x)=p(z)p(x\cdot y).$$

This definition is easily generalized to the case of n events.

It is generally assumed that the trials of an event are independent. What does this assumption mean? Suppose, for example, that we wish to say that the first trial of an event is independent of the second. The first trial constitutes an event, x_1, and the second trial constitutes an event, x_2, but we have only defined one trial of x_1 and one trial of x_2. Independence is defined in terms of probabilities, and probabilities can be given meaning only in terms of sequences of trials.

We can get around the difficulty in the following manner. Suppose we wish to consider the independence of n trials of an event, x. We will consider n events, $x_1, x_2, x_3, \ldots x_n$. The first trial ~~of~~ x_1 ~~will be the~~ first trial of x, the first trial of x_2 will be the second trial of x, the first trial of x_3 will be the third trial of x, etc. The 2nd trial of x_1 will be the $(n+1)st$ trial of x, the 2nd trial of x_2 will be the $(n+2)nd$ trial of x, etc. In general, the digits of the number, x_r, are selected from the digits of the number, x. The digits selected are, the rth, the $(r+n)th$, the $(r+2n)th$, $(r+3n)th$, etc. That is

$$x_r = .\, x^{(r)} x^{(r+n)} x^{(r+2n)} x^{(r+3n)} \ldots \tag{15}$$

We can now speak of the independence of the numbers, x_1, x_2, $\ldots x_n$.

It will be observed that

$$\frac{2^{-r}}{1-2^{-n}} = 2^{-r} + 2^{-r-n} + 2^{-r-2n} + 2^{-r-3n} + \ldots \tag{16}$$

and hence we can write

$$x_r = x\, c\, \frac{2^{-r}}{1-2^{-n}} \tag{17}$$

We shall abbreviate this notation still further and write

$$(r/n)x = x\, c\, \frac{2^{-r}}{1-2^{-n}}\,. \tag{18}$$

It is natural to assume that $p[(r/n)x] = p(x)$ for every pair of numbers, r and n, such that $0 < r \leqq n$. If we assume this, and if we assume that the numbers, $(1/n)x, (2/n)x, (3/n)x \ldots (n/n)x$, are independent, then x must satisfy the following equations.

$$p[(r_1/n)x \cdot (r_2/n)x \cdots (r_k/n)x] = [p(x)]^k \tag{19}$$

for every n and for every set of integers, $r_1, r_2, r_3, \ldots r_k$, such that $0 < r_1 < r_2 < \cdots < r_k \leqq n$.

Any number, x, which satisfies equations (19) is called an admissible number. It can be proved that there exist admissible numbers.* It is clear that an admissible number, x, characterizes the behavior which we should expect from a sequence of trials of an event with probability, $p(x)$.

[*See the author's article, *Admissible numbers in the theory of probability*, American Journal of Mathematics, Vol. L, No. 4, Oct. 1929].

Example 5. An event, x, has the probability, $p(x)$. What is the probability of obtaining precisely two successes in three trials of the event? It is required to find

$$p\{[(1/3)x\cdot(2/3)x\cdot\sim(3/3)x]\vee[(2/3)x\cdot(3/3)x\cdot\sim(1/3)x]$$
$$\vee[(3/3)x\cdot(1/3)x\cdot\sim(2/3)x]\}.$$

Each of the square brackets contains three independent numbers. Thus for each square bracket we have the probability, $[p(x)]^2 p(\sim x)$. The square brackets themselves constitute three mutually exclusive events. Hence the desired probability is $3[p(x)]^2 p(\sim x)$.

Let us find the probability of r successes and $n-r$ failures in n trials of an event. Let $p(x)=p$ and $p(\sim x)=q$. The probability that a given set of r trials will all be successful, is p^r, and the probability that the remaining $n-r$ trials will all be failures, is q^{n-r}. The r successful trials can be chosen in ${}_nC_r$ ways. Since all of these ways are mutually exclusive, the desired probability is ${}_nC_r\, p^r q^{n-r}$.

Consider the following problem. Let $x_1, x_2, \ldots x_n$ be a set of mutually exclusive events whose probabilities are known. We shall call these events causes. Let y be an event which can occur only as a result of one of the causes. The probabilities of y if x_1, y if x_2, etc. are also known. An experiment is performed and it is observed that y occurs. What is the probability that this occurrence is a result of kth cause? The answer to this question is given by the following theorem.

Theorem 3. If $x_1, x_2, x_3, \ldots x_n$ is a set of mutually exclusive events, and if y is such that $y\cdot\sim(x_1\vee x_2\vee\cdots\vee x_n)=0$, then

$$p(x_k\subset y)=\frac{p(x_k)\cdot p(y\subset x_k)}{\sum_{i=1}^{n} p(x_i)\cdot p(y\subset x_i)}.$$

Since $y\cdot\sim(x_1\vee x_2\vee\cdots\vee x_n)=0$ it follows that

$$y=(y\cdot x_1)\vee(y\cdot x_2)\vee\cdots\vee(y\cdot x_n).$$

Hence $p(y)=p(y\cdot x_1)+p(y\cdot x_2)+\cdots+p(y\cdot x_n)$.

Therefore

$$p(y) = p(x_1)\,p(y \subset x_1) + p(x_2)\,p(y \subset x_2) + \cdots + p(x_n)\,p(y \subset x_n).$$

To complete the proof of the theorem it is only necessary to substitute this value of $p(y)$ in the equation, $p(x_k \subset y) = p(x_k \cdot y)/p(y)$, and then substitute $p(x_k) \cdot p(y \subset x_k)$ for $p(x_k \cdot y)$.

Theorem 3 is known as Bayes' principle. The probabilities, $p(x_1)$, $p(x_2)$, . . . $p(x_n)$, are called *a priori* probabilities, whereas the probabilities, $p(x_1 \subset y)$, $p(x_2 \subset y)$, . . . $p(x_n \subset y)$, are called *a postiori*.

Example 6. There are four urns, U_0, U_1, U_2, U_3. The urn, U_0, contains three black balls, U_1 contains one white ball and two black balls, U_2 contains two white and one black, and U_3 contains three white balls. An urn is selected at random and a ball is drawn from it and found to be white. What is the probability that the ball came from U_2? Let x_0, x_1, x_2, x_3 represent respectively the drawing of U_0, U_1, U_2, U_3, and let y represent the drawing of a white ball from the urn selected. Then

$$p(x_0) = p(x_1) = p(x_2) = p(x_3) = 1/4,$$

and $p(y \subset x_0) = 0$, $p(y \subset x_1) = 1/3$, $p(y \subset x_2) = 2/3$, $p(y \subset x_3) = 3/3$.

Hence $$p(x_2 \subset y) = \frac{\frac{1}{4} \cdot \frac{2}{3}}{\frac{1}{4} \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{2}{3} + \frac{1}{4} \cdot \frac{3}{3}} = 1/3.$$

Example 7. Two people, A and B, make the same statement independently. Let this event be denoted by y. Let x denote the event that the statement is true. Then y can be the result of two causes, $x_1 = x$ and $x_2 = \sim x$. It is given that the probabilities of A and B speaking the truth, are respectively a and b. What is the *a postiori* probability that the statement is true? We know that

$p(y \subset x_1) = ab$ and $p(y \subset x_2) = (1-a)(1-b)$. Hence

$$p(x \subset y) = \frac{a.b \cdot p(x)}{a \cdot b \cdot p(x) + (1-a)(1-b) \cdot p(\sim x)}$$

It might be added by way of warning that it is easy to state a problem of this kind, which is without meaning.

Let us consider the problem of finding the probability that an event, x, will preceed an event, y, a tie being excluded. We have the four possible situations, $x \cdot y$, $(\sim x)\cdot(\sim y)$, $x\cdot(\sim y)$, $(\sim x)\cdot y$. The first situation represents the tie, and this situation we have excluded. In the second situation neither x nor y succeeds, and this situation should also be excluded. The event, x, will preceed y provided x succeeds and y fails if either of the last two situations occurs. Hence the desired probability is

$$p[(x\cdot\sim y)\subset\{(x\cdot\sim y)\vee(y\cdot\sim x)\}]=\frac{p(x\cdot\sim y)}{p(x\cdot\sim y)+p(y\cdot\sim x)}.$$

When x and y are mutually exclusive this last expression takes the following form:

$$\frac{p(x)}{p(x)+p(y)}.$$

IV. CONCLUSION

The above examples illustrate how the theory of probability can be developed in terms of our idealized universe. By this method we can construct a consistent mathematical theory, and one which admits the possibility of experimental verification.