THE FOUNDATIONS OF FINITE CONDITIONAL PROBABILITY *

By

Roberto S. Mariano

1. **Introduction:** The axiomatization of the theory of probability is not a new subject in mathematics. Many have tried, and successfully so, to set forth in an axiom system the theory of probability. However, in the majority of these attempts, the concept of the probability of an event has been invariably considered as one of the primitive undefined notions of the axiom system.

In this paper, we shall try to establish the axiomatic foundation of the theory of probability concerning finite sample spaces on a more general notion — that of the conditional probability of an event. More general in that from the theory developed with this concept as a primitive undefined notion, we can easily derive the elementary theory of probability ordinarily established by considering the probability of an event as one of the primitive undefined notions. The simple derivation will be shown in the later part of the paper.

Also, as is natural for studies of this nature, some metamathematical considerations of the adopted axiom system will be given.

2. **Primitive Notion.** We consider the following as the primitive undefined notions of our axiom system:

1. The sample space, which we shall designate by a finite non-empty set S.
2. A set of possible events, represented by F , a family of subsets of S.

---

* This paper is written in partial fulfillment of an A.B. Honors Course degree in Mathematics at the Ateneo de Manila University, Loyola Heights, Quezon City.

3. A real-valued function $p(A_i, A_j)$ defined on $F \times F$ for A_i and A_j belonging to F . This real-valued function is what we call the conditional probability of A_i given A_j. It must be noted that $p(A_i, A_j)$ is not necessarily equal to $p(A_j, A_i)$.

3. **Axioms.** If S is the sample space, F a family of subsets of S and p a real-valued function defined on $F \times F$, a set function structure $X = \langle S, F, p \rangle$ is a finite conditional probability space if and only if:

1. F is a field. A family of sets, F is a field if and only if:

a. For A in F , the complement of A, $\ddot{A}$ is also in F.

b. For n sets belonging to F , the intersection of any number of these sets belongs to F.

c. For n sets belonging to F , the union of any number of these sets belongs to F .

2. For any non-empty set A in F , $p(A, A) = 1$.

3. For any two sets A_i and A_j in F, $p(A_i, A_j) \geq 0$.

4. $p(A_i, A_j) \leq 1$.

5. For two mutually exclusive sets, A_i and A_j in F,

$$p(A_i \cup A_j, A) = p(A_i, A) + p(A_j, A), \quad \text{for } A \text{ in } F.$$

4. **Consistency of the Axioms.** To prove the consistency of the axioms that we have just set up, it is sufficient to show that a model can be constructed where all these axioms are simultaneously satisfied.

Denoting the sample space by S, let $S = \{ a, b \}$,

that is, since as we have said before, we consider the sample space as a set of elements, we let a and b be the elements of the set S. Denoting by F , a family of subsets of S, let

$F = \{A_1, A_2, A_3, A_4\}$, where $A_1 = \{a\}$, $A_2 = \{b\}$, $A_3 = \{a,b\}$, $A_4 = \phi$. Furthermore, for A_i and A_j belonging to F, let a real-valued function $p(A_i, A_j)$ be defined on F x F such that:

$p(A_1,A_1) = 1$	$p(A_3,A_1) = 1$
$p(A_1,A_2) = 0$	$p(A_3,A_2) = 1$
$p(A_1,A_3) = 1/2$	$p(A_3,A_3) = 1$
$p(A_1,A_4) = 0$	$p(A_3,A_4) = 0$
$p(A_2,A_1) = 0$	$p(A_4,A_1) = 0$
$p(A_2,A_2) = 1$	$p(A_4,A_2) = 0$
$p(A_2,A_3) = 1/2$	$p(A_4,A_3) = 0$
$p(A_2,A_4) = 0$	$p(A_4,A_4) = 0$

F , the family of subsets of S, with the function $p(A_i, A_j)$ defined on F x F , is our model.

The first four axioms are evidently satisfied by our model. To show that axiom 5 is satisfied by this model, we have to show that the formula in axiom five holds for the following pairs of subsets of S, which are the only mutually exclusive subsets of S: A_1 and A_2, A_1 and A_4, A_2 and A_4, A_3 and A_4, A_4 and A_4.

For $A_1 \cup A_2$, since $A_1 \cup A_2 = A_3$,

$$p(A_1 \cup A_2, A_1) = p(A_3, A_1) = 1 = 1 + 0 = p(A_1, A_1) + p(A_2, A_1).$$

$$p(A_1 \cup A_2, A_2) = p(A_3, A_2) = 1 = 0 + 1 = p(A_1, A_2) + p(A_2, A_2).$$

$$p(A_1 \cup A_2, A_3) = p(A_3, A_3) = 1 = 1/2 + 1/2 = p(A_1, A_3) + p(A_2, A_3).$$

$$p(A_1 \cup A_2, A_4) = p(A_3, A_4) = 0 = 0 + 0 = p(A_1, A_4) + p(A_2, A_4).$$

For $A_i \cup A_4$, $i = 1, 2, 3, 4$, since $A_i \cup A_4 = A_i$ and $p(A_4, A_i) = 0$,

$$p(A_i \cup A_4, A_1) = p(A_i, A_1) = p(A_i, A_1) + p(A_4, A_1).$$

$$p(A_i \cup A_4, A_2) = p(A_i, A_2) = p(A_i, A_2) + p(A_4, A_2).$$

$$p(A_i \cup A_4, A_3) = p(A_i, A_3) = p(A_i, A_3) + p(A_4, A_3).$$

$$p(A_i \cup A_4, A_4) = p(A_i, A_4) = p(A_i, A_4) + p(A_4, A_4).$$

From these above equalities we see that the formula in axiom 5 holds for the above mentioned pairs of mutually exclusive subsets of S. Since commutativity holds for the union of two sets, that is, for two sets A and B, $A \cup B = B \cup A$, axiom 5 also holds evidently for the commuted forms of the unions of the above mentioned pairs.

5. **Independence of the Axioms.** In order to establish the independence of a particular axiom in our system we have to construct a model which simultaneous satisfies the negation of that particular axiom and also the four other axioms. We shall do that now for each of our five axioms.

To show the independence of axiom 1 in our system, let our model be the following:

1. $F_1 = \{A_1, A_2, A_3\}$, with A_1, A_2, and A_3 as defined before,

2. A function $f(A_i, A_j)$ defined on $F_1 \times F_1$ such that:

$f(A_1, A_1) = 1$ $\quad$ $f(A_2, A_3) = 1/2$

$f(A_1, A_2) = 0$ $\quad$ $f(A_3, A_3) = 1$

$f(A_1, A_3) = 1/2$ $\quad$ $f(A_3, A_2) = 1$

$f(A_2, A_1) = 0$ $\quad$ $f(A_3, A_1) = 1$

Evidently, the second, third, and fourth axioms and the negation of the first are satisfied by this model. The satisfaction of the fifth axiom is verified in the pertinent equalities contained in the later part of the previous section.

To verify the independence of axiom 2, let our model in this case be the following:

1. F , a family of subsets of S as defined in the previous section, and

2. A real-valued function defined on F x F , denoted by $g(A_i, A_j)$, for A_i and A_j in F , whose values are the same as those of $p(A_i, A_j)$ but for the following:

$g(A_1, A_1) = 2/3$ $\quad$ $g(A_3, A_2) = 2/3$

$g(A_2, A_2) = 2/3$ $\quad$ $g(A_1, A_3) = 1/3$

$g(A_3, A_3) = 2/3$ $\quad$ $g(A_2, A_3) = 1/3$

$g(A_3, A_1) = 2/3$

Again it is evident that the first, third, and fourth axioms and the negation of the second are satisfied by this model. noting that

$$g(A_3,A_1) = g(A_1 \cup A_2,A_1) = 2/3 = 2/3 + 0$$
$$= g(A_1,A_1) + g(A_2,A_1),$$
$$g(A_3,A_2) = g(A_1 \cup A_2,A_2) = 2/3 = 2/3 + 0$$
$$= g(A_2,A_2) + g(A_1,A_2),$$
$$g(A_3,A_3) = g(A_1 \cup A_2,A_3) = 2/3 = 1/3 + 1/3$$
$$= g(A_1,A_3) + g(A_2,A_3),$$

and referring to the pertinent equalities in the previous section, we see that this model also satisfies the fifth axiom.

In the case of the third axiom, let our model be the following:

1. F , a family of subsets of S whose elements are as defined before.

2. A real valued function defined on F x F , $h(A_i, A_j)$, for A_i and A_j in F , whose values are the same as those of $p(A_i, A_j)$ but for the following:

$$h(A_1,A_4) = -1$$

$$h(A_2,A_4) = -1$$

$$h(A_3,A_4) = -2.$$

Evidently, the first, second, and fourth axioms and the negation of the third are sastified by this model. The pertinent equalities in the previous section and the fact that

$$h(A_1 \cup A_2, A_4) = h(A_3, A_4) = -2 = (-1) + (-1)$$
$$= h(A_1, A_4) + h(A_2, A_4),$$

show that the fifth axiom is satisfied by this model. These facts establish the independence of the third axiom in our system.

To show the independence of axiom 4 in our system, let our model be the following:

1. F , a field of subsets of S as defined before.

2. A real-valued function defined on F x F , $u(A_i, A_j)$, for A_i and A_j in F , whose values are same as those of $p(A_i, A_j)$ but for the following:

   $u(A_1, A_4) = 2$

   $u(A_2, A_4) = 2$

   $u(A_3, A_4) = 4.$

By inspection, we see that this model satisfies the first three axioms and the negation of the fourth. Noting the equalities in the previous section and the fact that

$$u(A_1 \cup A_2, A_4) = u(A_3, A_4) = 4 = 2 + 2$$
$$= u(A_1, A_4) + u(A_2, A_4),$$

we also see that the fifth axiom is satisfied.

To prove the independence of axiom 5 in our system, let the following be our model:

1. The same F and S as in the previous cases.

2. A real-valued function $v(A_i, A_j)$ defined on F x F for any A_i and A_j in F, whose values are the same as those of $p(A_i, A_j)$ except for the following:

$$v(A_3, A_4) = 1.$$

This model satisfies the first four axioms, as can be verified by an inspection of the values of the fuction defined. Furthermore, since

$$v(A_1 \cup A_2, A_4) = v(A_3, A_4) = 1$$

$$\neq [v(A_1, A_4) + v(A_2, A_4) = 0],$$

the negation of the fifth axiom is satisfied by the model and hence the independence of the fifth axiom in our axiom system is established.

6. **Consequences of the Axioms.** Let us now consider the theorems that we can infer from the axioms. Unless otherwise stated, we shall presuppose as given in all these theorems a finite non-empty sample space S, a field of subsets of S, denoted by F, and a real valued function $p(A_i, A_j)$ defined on F x F for and A_i and A_j in F.

THEOREM 1. For any A_i in F,

$$p(\phi, A_i) = 0 \quad \text{and} \quad p(A_i, S) + p(\bar{A}_i, S) = 1$$

Proof:

a. Since for any A_j in F, $A_j \cup \phi = A_j$ and $A_j \cap \phi = \phi$, then, by axiom 5

$$p(A_j \cup \phi, A_i) = p(A_j, A_i) = p(A_j, A_i) + p(\phi, A_i)$$

which implies that $p(\phi, A_i) = 0$, for the last equality to hold. Q.E.D.

b. Since $A_i \cup \bar{A}_i = S$ and $A_i \cap \bar{A}_i = \phi$,

$$p(A_i \cup \bar{A}_i, S) = p(S,S) = p(A_i, S) + p(\bar{A}_i, S),$$

by axiom 5. But by axiom 2, $p(S, S) = 1$, which implies that

$$p(\bar{A}_i, S) + p(A_i, S) = 1. \quad \text{Q.E.D.}$$

THEOREM 2. Let $A = \{x_1, x_2, \ldots, x_n\}$, $E_i = \{x_i\}$, for $i = 1, \ldots, n$ for A and E_i belonging to F. Then, for any A_j in F,

$$p(A, A_j) = \sum_{i=1}^{n} p(E_i, A_j).$$

Proof:

$A = E_1 \cup E_2 \cup \ldots \cup E_n$, where the E_i's are pair-wise disjoint.

Since $E_1 \cap (E_2 \cup \ldots \cup E_n) = \phi$, by axiom 5

$p(A,A_j) = p(E_1,A_j) + p(\bigcup_{i=2}^{n} E_i,A_j)$.

Since $E_2 \cap (E_3 \cup \ldots \cup E_n) = \phi$, it follows again by axiom 5 that

$p(A,A_j) = p(E_1,A_j) + p(E_2,A_j) + p(\bigcup_{i=3}^{n} E_i,A_j)$.

The same line of reasoning can be used repeatedly until we arrive at

$$p(A,A_j) = p(E_1,A_j) + \ldots + p(E_n,A_j)$$

$$= \sum_{i=1}^{n} p(E_i,A_j). \quad \text{Q.E.D.}$$

COROLLARY 1. If events or sets A_i's in F are pair-wise disjoint, for $i = 1,2,\ldots,n$, then for A_j in F, $p(\bigcup_{i=1}^{n} A_i,A_j) = \sum_{i=1}^{n} p(A_i,A_j)$.

Proof:

In theorem 2, $A = \bigcup_{i=1}^{n} A_i$, $A_j = A_j$, $E_i = A_i$, for $i = 1,2,\ldots,n$. By direct substitution the corollary holds. Q.E.D.

COROLLARY 2. Given A in F and sets A_i's, $i = 1,2,\ldots,n$, which are pairwise disjoint and exhaustive subsets of A. Then

$$\sum_{i=1}^{n} p(A_i, A) = 1.$$

Proof:

This evidently holds, by corollary 1 and axiom 2.

THEOREM 3. If $A_i \subset A_j$, for non-empty sets A_i and A_j in F, $p(A_j, A_i) = 1$.

Proof:

$A_j = A_i \cup (A_j \cap \bar{A}_i)$ which implies that
$p(A_j, A_i) = p[A_i \cup (A_j \cap \bar{A}_i), A_i]$.
But $A_i \cap (A_j \cap \bar{A}_i) = \phi$ which implies, by axiom 5 $p(A_j, A_i) = p(A_i, A_i) + p(A_j \cap \bar{A}_i, A_i)$
$= 1 + p(A_j \cap \bar{A}_i, A_i)$,

by axiom 2. But by axiom 4, $p(A_j, A_i) \leqq 1$ and by axiom 3, $p(A_j \cap \bar{A}_i, A_i) \geqq 0$ which implies that $p(A_j, A_i) = 1$. Q.E.D.

THEOREM 4. For any sets A, B, and C in F,

$$p(A \cup B, C) = p(A,C) + p(B,C) - p(A \cap B, C).$$

Proof:

$$A = (A \cap B) \cup (A \cap \bar{B}), \quad \text{where} \quad A \cap B \cap \bar{B} \cap A = \phi$$

$$B = (A \cap B) \cup (\bar{A} \cap B), \quad \text{where} \quad A \cap B \cap \bar{A} \cap B = \phi$$

which implies by axiom 5 that,

$$p(A \cap \bar{B}, C) = p(A,C) - p(A \cap B, C) \quad \text{and}$$

$$p(\bar{A} \cap B, C) = p(B,C) - p(A \cap B, C).$$

But $A \cup B = (A \cap B) \cup (A \cap \bar{B}) \cup (\bar{A} \cap B)$, and these sets are also elements of F which are pair-wise disjoint. This implies, by corollary 1

$$p(A \cup B, C) = p(A \cap B, C) + p(A \cap \bar{B}, C) + p(\bar{A} \cap B, C)$$

$$= p(A,C) + p(B,C) - p(A \cap B, C).$$

Q.E.D.

THEOREM 5. For any B and A_i's belonging to F, $i = 1, \ldots, n$

$$p\left(\bigcup_{i=1}^{n} A_i, B\right) = \sum_{i=1}^{n} p(A_i, B) - \sum_{\substack{i \neq j \\ 1 \leq i, j \leq n}} p(A_i \cap A_j, B)$$

$$+ \sum_{\substack{i \neq j \neq k \\ 1 \leq i,j,k \leq n}} p(A_i \cap A_j \cap A_k, B) + \ldots +$$

$$(-1)^n \sum_{\substack{i_1 \neq i_2 \neq \ldots \neq i_{n-1} \\ 1 \leq i_i\text{'s} \leq n}} p(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_{n-1}}, B)$$

$$+ (-1)^{n+1} p(A_1 \cap A_2 \cap \ldots \cap A_n, B).$$

Proof: (By Mathematical Induction)

1. The formula vacuously holds for $n = 1$.

2. Suppose the formula holds for $n = m$, then

$$p(K,B) = p(\bigcup_{i=1}^{m+1} A_i, B) = p\left[(\bigcup_{i=1}^{m} A_i) \cup A_{m+1}, B\right]$$

$$= p(\bigcup_{i=1}^{m} A_i, B) + p(A_{m+1}, B)$$

$$- p\left[(\bigcup_{i=1}^{m} A_i) \cap A_{m+1}, B\right],$$

by theorem 4.

3. Since the last term of the last equality involves a union of m terms, our hypothesis of induction applies, namely:

$$p\left[(\bigcup_{i=1}^{m} A_i) \cap A_{m+1}, B\right] = \sum_{i=1}^{m} p(A_i \cap A_{m+1}, B)$$

$$- \sum_{\substack{i \neq j \\ 1 \leq i, j \leq m}} p(A_i \cap A_j \cap A_{m+1}, B) + \ldots +$$

$$(-1)^m \sum_{\substack{i_1 \neq i_2 \neq \ldots \neq i_{m-1} \\ 1 \leq i_k\text{'s} \leq m}} p(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_{m-1}} \ldots$$

4. Substituting in (2), rearranging, and making use of our hypothesis of induction.

$$p(K,B) = \sum_{i=1}^{m} p(A_i, B) + p(A_{m+1}, B)$$

$$- \sum_{\substack{i=j \\ 1 \leq 1, j \leq m}} p(A_i \cap A_j, B) - \sum_{i=1}^{m} p(A_i \cap A_{m+1}, B)$$

$$+ \ldots + (-1)^{m+1} p(A_1 \cap A_2 \cap \ldots \cap A_m, B)$$

$$+ (-1)^{m+1} \sum_{\substack{i_1 \neq i_2 \neq \ldots \neq i_{m-1} \\ 1 \leq i_k\text{'s} \leq m}} P(A_{i_1} \cap A_{i_2} \cap \ldots$$

$$\cap A_{i_{m-1}} \cap A_{m+1}, B)$$

$$+ (-1)^{m+2} p(A_1 \cap A_2 \cap \ldots \cap A_{m+1}, B).$$

5. By combining like terms under the same summation sign in this formula, we shall have derived the equation in our theorem, for $n = m + 1$ and hence we shall have proven that our formula holds for $n = m + 1$ if we suppose that it holds for $n = m$.

Q.E.D.

DEFINITION 1. Given a sample space S, a field F of subests of S, and an event A belonging to F . The probability of the event A, denoted by p(A) is its conditional probability given S. In formula notation, we have:

$$p(A) = p(A,S).$$

From this definition, the elementary theory of probability, ordinarily derived by considering the probability of an event as a primitive undefined notion, can be established by a simple specialization of the axioms and the theorems that we have derived sa far. This speciaization is done by considering the conditional probability of an event A given S, the whole sample space, whenever a theorem or an axiom is applicable to this case. Thus, following this procedure, we have the following theorems which make up the ordinary elementary theory of probability:

THEOREM 6. $p(S) = 1$

$0 \leqq p(A) \leqq 1$, for any A in F.

$p(A \quad B) = p(A) + p(B)$, for any two disjoint events A and B belonging to F.

Proof:

These immediately follow from definition 1 and Axioms 2 — 4.

We will not give a strict proof for the succeeding theorems. These theorems evidently and logically follow from the theorems previously proven and the given definition of the probability of an event.

THEOREM 7. For any A in F, $p(A) + p(\bar{A}) = 1$.

$p(\phi) = 0$.

THEOREM 8. Let $A = \{x_1, x_2, \ldots, x_n\}$, $E_i = \{x_i\}$, for $i = 1,2,\ldots,n$ for A and E_i belonging to F. Then $p(A) = \sum_{i=1}^{n} p(E_i)$.

COROLLARY 3. If events A_i's in F are pairwise disjoint for $i = 1,2,\ldots, n$, then

$$p\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} p(A_i).$$

COROLLARY 4. If events A_i's in F are pairwise disjoint, and exhaustive subsets of S, for $i = 1,2,\ldots,n$, then $\sum_{i=1}^{n} p(A_i) = 1$.

THEOREM 9. For any sets A and B in F,

$$p(A \cup B) = p(A) + p(B) - p(A \cap B).$$

THEOREM 10. For any A_i's belonging to F, for $i = 1, 2, \ldots, n$,

$$p\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} p(A_i) - \sum_{\substack{i \neq j \\ 1 \leq i, j \leq n}} p(A_i \cap A_j)$$

$$+ \ldots + (-1)^{n+1} p(A_1 \cap A_2 \cap \ldots \cap A_n).$$

7. **Conclusion.** It would seem premature to end at this juncture but we are constrained to do so due to lack of time.

Up to this point, we have set up our axiom system, shown the feasibility of adopting such a system by establishing its consistency and the independence of the particular axioms taken, and furthermore, we have derived the more important theorems consequent upon our axioms.

However, the work is still far from being complete. In the academic point of view, further investigation is still to be carried out as to whether other theorems can be derived, and most probably there are still others. And also among other things, we have to consider independent events, ordinary probability distribution and conditional probability distribution on the elements of the sample space. And in the practical point of view, it is but logical to study the applicability of this theory that we are trying to develop.