AXIOMATIC PROBABILITY AND POINT SETS

The axioms of Kolmogorov. Let S denote a sample space with a probability measure P defined over it, such that probability of any event $A \subset S$ is given by $P(A)$. Then, the probability measure obeys the following axioms:

(1) $P(A) \geq 0$,

(2) $P(S) = 1$,

(3) If $\{A_1, A_2, \ldots A_j, \ldots\}$ is a sequence of mutually exclusive events such that $A_i \cap A_j = \emptyset$ for all i, j, then $P(A_1 \cup A_2 \cup \cdots \cup A_j \cup \cdots) = P(A_1) + P(A_2) + \cdots + P(A_j) + \cdots$.

The axioms are supplemented by two definitions:

(4) The conditional probability of A given B is defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

(5) The events A, B are said to be statistically independent if

$$P(A \cap B) = P(A)P(B).$$

This set of axioms was provided by Kolmogorov in 1936.

Operations on Sets. The axioms of probability concern sets of events. In order to employ these axioms, it is necessary to invoke the rules of Boolean algebra, which are associated with a pair of binary operations. First, we must define these operations together with some special sets.

A binary operation of union, *denoted by the symbol $\cup$, may be defined relative to any two sets A and B. The operation generates the set*

$$A \cup B = \{x;\ x \in A \text{ or } x \in B\}.$$

Here the word "or" is used in the inclusive sense to imply that x is either in A or in B or in both. For example, if S is the set of all vertebrates, A is the characteristic of having fur and B is the characteristic of laying eggs, then $A \cup B$ certainly has the duck-bill platypus amongst its elements as well as foxes and geese.

A binary operation of intersection, *denoted by the symbol $\cap$, may be defined relative to any two sets A and B. The operation generates the set*

$$A \cap B = \{x;\ x \in A \text{ and } x \in B\}.$$

In terms of the previous example, $A \cap B$ (unless I am mistaken) has only the duck-bill platypus and the spiny ant eater as its two elements.

Two sets A and B are said to be disjoint if their intersection is the empty set $A \cap B = \emptyset$.

If A is the set of vertebrate fish and B is the set of mammals, then, according to modern usage, their intersection is the empty set. However, as recently as Victorian times, whales, which are mammals, were liable to be described as fish.

Let $A \subset S$. Then the complement of A in S, denoted by A^c, is the set of all the elements of S which do not belong to A: $A^c = \{x;\ x \notin A\}$.

The rules of Boolean Algebra. The binary operations of union $\cup$ and intersection $\cap$ are roughly analogous, respectively, to the arithmetic operations of addition $+$ and multiplication $\times$, and they obey a similar set of laws. In fact, the laws of Boolean algebra are virtually symmetric with respect to the two operations in the sense that, in any of the statements of the laws that are listed below, the symbols can be interchanged without affecting their truth. This in not the case in arithmetic. The laws are as follows:

Commutative law: $A \cup B = B \cup A$,
$A \cap B = B \cap A$,

Associative law: $(A \cup B) \cup C = A \cup (B \cup C)$,
$(A \cap B) \cap C = A \cap (B \cap C)$,

Distributive law: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$,
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$,

Idempotency law: $A \cup A = A$,
$A \cap A = A$.

These various laws have the status of axioms. These axioms ae accompnies by three definitions:

There is a universal set S, containing all other sets, such that, for any $A \subset S$, there is

$$A \cup S = S, \qquad A \cap S = A.$$

There is a null set or empty set $\emptyset$ such that, for any $A \subset S$, there is

$$A \cup \emptyset = A, \qquad A \cap \emptyset = \emptyset.$$

For any $A \subset S$, there is exists a unique complementary set A^c such that

$$A \cup A^c = S, \qquad A \cap A^c = \emptyset.$$

There are several useful identities that are deducible from the axioms and from the definitions. Thus, *De Morgan's Rules* state that

$$(A \cup B)^c = A^c \cap B^c \qquad \text{and} \qquad (A \cap B)^c = A^c \cup B^c.$$

With all of these these rules in hand, we may proceed to the business of proving some simple lemmas of probability:

Lemma: the probability of the null event. Axiom 3 implies that $P(S \cup \emptyset) = P(S) + P(\emptyset)$, since S and $\emptyset$ are disjoint sets by definition, i.e. $S \cap \emptyset = \emptyset$. But also $S \cup \emptyset = S$, so $P(S \cup \emptyset) = P(S) = 1$, where the second equality is from axiom 2. Therefore, $P(S \cup \emptyset) = P(S) + P(\emptyset) = P(S)$, so $P(\emptyset) = 0$.

Lemma: the probability of the complementary event. If A and A^c are complementary events, then there is $A \cup A^c = S$ and $A \cap A^c = \emptyset$. Therefore, $P(A \cup A^c) = P(A) + P(A^c) = 1$, since $P(A \cup A^c) = P(S) = 1$, whence $P(A^c) = 1 - P(A)$.

Theorem: independence and the complementary event. If A, B are statistically independent such that $P(A \cap B) = P(A)P(B)$ then A, B^c are also statistically independent such that $P(A \cap B^c) = P(A)P(B^c)$.

Proof. Consider

$$A = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c).$$

The final expression denotes the union of disjoint sets, so there is

$$P(A) = P(A \cap B) + P(A \cap B^c).$$

Since, by assumption, there is $P(A \cap B) = P(A)P(B)$, it follows that

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A \cap B) = P(A) - P(A)P(B) \\ &= P(A)\{1 - P(B)\} = P(A)P(B^c). \end{aligned}$$

Theorem: the union of of events. *The probability that either A or B will happen or that both will happen is the probability of A happening plus the probability of B happening less the probability of the joint occurrence of A and B:*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Proof. There is $A \cup (B \cap A^c) = (A \cup B) \cap (A \cup A^c) = A \cup B$, which is to say that $A \cup B$ can be expressed as the union of two disjoint sets. Therefore, according to axiom 3, there is

$$P(A \cup B) = P(A) + P(B \cap A^c).$$

But $B = B \cap (A \cup A^c) = (B \cap A) \cup (B \cap A^c)$ is also the union of two disjoint sets, so there is also

$$P(B) = P(B \cap A) + P(B \cap A^c) \Longrightarrow P(B \cap A^c) = P(B) - P(B \cap A).$$

Substituting the latter expression into the one above gives

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

Bayes' Theorem

Observe that the formula for conditional probability implies that

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A),$$

whence we

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

This is the basis of Bayes' law of inverse probabilities, which provides an idealised model of a process by which we might adapt our our probabilistic beliefs or hypotheses concerning the state of the world in view of the empirical evidence that accumulates.

Consider a set $\Omega = \{H_1, H_2, \ldots, H_n\}$, wherein $H_i \cap H_j = \emptyset; i \neq j$, which comprises all possible explanations of an event E. Under some circumstances, it is possible to define a probability measure over Ω that indicates the relative likelihoods of the alternative hypotheses therein. In the absence of the evidence of E, they are described as *prior* likelihoods or probabilities.

The evidence given by the event E will cast some light upon the likelihoods of the hypotheses, which is to say that we can define a set of modified *posterior* likelihoods over the set Ω. The posterior likelihood of an hypothesis H_i in the light of the event E is given by

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{P(E)}, \quad \text{where}$$

$$P(E) = \sum_i P(E \cap H_i) = \sum_i P(E|H_i)P(H_i).$$

We use the following terminology:

$P(H_i)$ is the prior likelihood of the ith hypothesis H_i,

$P(H_i|E)$ is the posterior likelihood of the ith hypothesis,

$P(E|H_i)$ is the conditional probability of the event E under the hypothesis H_i,

$P(E)$ is the unconditional probability of the event E.

Example. The Manager of Ffyfes, who import bananas to the U.K. from many sources, has discovered an unmarked crate, and he wishes to determine its origin. 40 percent of the crates in stock come from Guatemala and 60 percent from Cuba. On average, 1/2 the Guatemalan bananas are bad and 1/6 of the Cuban bananas are bad. The manager opens the crate and pulls out a banana that happens to be bad. In the light of this evidence, what it is the most likely origin of the crate?

Answer. Let H_1 denote the hypothesis that the crate if from Cuba and let H_2 denote the hypothesis that it is from Guatemala. Let E be the event of discovering a rotten banana. There are the following items of information:

$$P(H_1) = \frac{60}{100} = \frac{3}{5} \qquad P(H_2) = \frac{20}{100} = \frac{2}{5},$$

$$P(E|H_1) = \frac{1}{6} \qquad P(E|H_2) = \frac{1}{2},$$

$$P(E|H_1)P(H_1) = \frac{1}{6} \times \frac{3}{5} = \frac{1}{10} \simeq P(H_1|E),$$

$$P(E|H_2)P(H_2) = \frac{1}{2} \times \frac{2}{5} = \frac{1}{5} \simeq P(H_2|E).$$

In the light of the evidence, it seems twice as likely that Guatemala is where the crate is from.