

CHAPTER 1

Conjunction Fallacy

The probability of the joint realization of two events, say A and B , cannot be larger than the probability of either of the two events, considered on its own. Denoting the probabilities of the events by $\mathbb{P}(A)$, $\mathbb{P}(B)$, and $\mathbb{P}(A \cap B)$, the basic law of probability law takes the form,

$$\mathbb{P}(A \cap B) \leq \mathbb{P}(A) \quad \text{or} \quad \mathbb{P}(A \cap B) \leq \mathbb{P}(B).$$

Binary Relations

Definition A *Binary Relation*, given sets X and Y , is a set R such that

$$R \subseteq X \times Y.$$

A binary relation is a set of ordered pairs $xy \in X \times Y$, where xy is an abbreviation for (x, y) . If $X = Y$, then R is said to be a binary relation on X . A binary relations R is a *quasi order* on a set X if it is *reflexive* and *transitive*. That is, for all x, y , and z in X ,

$$\begin{array}{ll} xRx & \text{(reflexivity)} \\ xRy \ \& \ yRz \implies xRz & \text{(transitivity)} \end{array}$$

A binary relation R is an *equivalence relation* on a set X if it is *reflexive*, *transitive*, and *symmetric* on X . for all x, y , and z in X ,

$$xRy \iff yRx.$$

Partitions

The family $\mathcal{X} = \{[x] | x \in X\}$ of subsets is called a *partition* of X induced by \sim . Any partition \mathcal{X} of X satisfies the following three properties,

1. $Y \in \mathcal{X}$ implies $Y \neq \emptyset$;
2. $Y, Z \in \mathcal{X}$ and $Y \neq Z$ imply $Y \cap Z = \emptyset$;
3. $\cup \mathcal{X} = X$.

Conversely, any family \mathcal{X} of subsets of a set X satisfying [1], [2], and [3] is called a partition of X .

CHAPTER 2: Sample Spaces

The Sample Space

What is critical is that all feasible outcomes have a description in the sample space. It is of no importance that the sample space contains outcomes that never happen in practice: in the framework of a probabilistic model, such outcomes may be assigned zero probability.

A sample space which is either **finite** or **countable** is called *discrete*.

Finite vs. Countable

Finite: A definite number. Not infinite. In other words it could be measured, or given a value.

Countable: Either finite or a countably infinite (implying that the cardinality of a countable set is a subset of \mathbb{N}). Elements of a countable set can always be counted one at a time and, although the counting may never finish, every element of the set is associated with a unique natural number.

Concept of an Event

Fundamentally, an *Event* is a subset of the sample space. In general, each event pertaining to a particular experiment may be identified with a subset of the relevant sample space. However, with respect to the converse, the general assumption that all subsets of the sample space are events to which a probability could be assigned would create mathematical difficulties.

Indicator Functions

Definition: An *Indicator Function* is a function defined on a set Ω that indicates membership of an element in a subset of X of Ω . Let Ω be a non-empty set and X a subset of Ω . The indicator function on X is a function $I_X : \Omega \rightarrow \{0, 1\}$ by,

$$I_X(w) = \begin{cases} 0 & w \in X \\ 1 & w \in \bar{X} \text{ (or } w \notin X) \end{cases}$$

Properties of Indicator Functions:

- For subsets X, Y of Ω , $I_{X \cap Y} = I_X \cdot I_Y$, i.e. for every $w \in \Omega$, $I_{X \cap Y}(w) = I_X(w) \cdot I_Y(w)$.
- $I_{X \cup Y} = I_X + I_Y - I_{X \cap Y}$ (if X, Y are disjoint, $I_{X \cup Y} = I_X + I_Y$)
- $I_{A^c} = 1 - I_A$
- If $X \subseteq Y$, then $I_X \subseteq I_Y$, i.e. $I_X(w) \leq I_Y(w)$ for every $w \in \Omega$.

CHAPTER 3: Probability and Area

Axioms for a Field

Definition: Let Ω be a set. Let \mathcal{F} be a nonempty collection of subsets of Ω satisfying the axioms

[F1] For all $A, B \in \mathcal{F}$, $A \cup B \in \mathcal{F}$.

[F2] For each $A \in \mathcal{F}$, $\bar{A} \in \mathcal{F}$ also.

Theorem 3.2: Let \mathcal{F} be a field of subsets of Ω . Then, .

(i) $\Omega \in \mathcal{F}$

(ii) $\emptyset \in \mathcal{F}$

(iii) For all $A, B \in \mathcal{F}$, $A \setminus B = A\bar{B} \in \mathcal{F}$

(iv) $\bigcup_{i=1}^n A_i \in \mathcal{F}$ } \mathcal{F} is closed under finite union **and** finite intersection
(v) $\bigcap_{i=1}^n A_i \in \mathcal{F}$ }

Infinite Sample Spaces and σ -Fields

Definition: A field (Ω, \mathcal{F}) is called a σ -field if it is *closed under countable union*, that is, if for any countable collection A_1, \dots, A_n, \dots of sets in \mathcal{F} , we have $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$. Note that for any sets A_1, \dots, A_n, \dots , we have

$$\bigcap_{n=1}^{\infty} A_n = \overline{\bigcup_{n=1}^{\infty} \bar{A}_n}$$

As a σ -field, \mathcal{F} is closed under both countable union and complementation. We conclude that \mathcal{F} is also closed under countable intersection.

Borel Fields

Recall that a set S of real numbers is an *open* set of \mathbb{R} if for any $x \in S$, there exists some $\delta > 0$ such that whenever $|x - y| < \delta$, then $y \in S$. The standard family of events for \mathbb{R} is a distinguished σ -field \mathcal{B} containing all the open sets of \mathbb{R} . In fact, \mathcal{B} is the ‘smallest’ σ -field containing these open sets. The collection \mathcal{B} is called the *Borel Field* of \mathbb{R} , and the events in \mathcal{B} are referred to as the *Borel sets* of \mathbb{R} . These definitions extend naturally to the case where all the sample space is \mathbb{R}^n .

Set Theory Formulas	Probability Theory
$A \in \mathcal{F}$	A is an event
$A, B \in \mathcal{F}, \quad A \cap B = \emptyset$	The events A and B are <i>incompatible</i>
$A = \emptyset$	The event A is <i>impossible</i>
$A = \Omega$	The event A is <i>certain</i>
$A_i \in \mathcal{F}, \text{ for all } i \in I, \quad \cup_{i \in I} A_i = B$	The event B is realized when at least one of the events A_i is realized
$A_i \in \mathcal{F}, \text{ for all } i \in I, \quad \cap_{i \in I} A_i = B$	The event B occurs when all of the events A_i are realized
$A, B \in \mathcal{F}, \quad A \subseteq B$	If event A occurs, then event B necessarily occurs
$A \in \mathcal{F}, \mathbb{P}(A) = 0$	The event A has measure zero

General Strategy

We start with a field of sets \mathcal{F} on a set Ω . Then, we introduce the concept of a probability measure \mathbb{P} as a function assigning a number $\mathbb{P}(A)$, with $0 \leq \mathbb{P}(A) \leq 1$, to every event A in \mathcal{F} . This function $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ will be assumed to satisfy a number of axioms (given in Ch. 4). The probability measure \mathbb{P} is not assumed in general to be defined on all subsets of the sample space due to the case of an uncountable sample space leading to a contradiction.

CHAPTER 4: Probability Measures

We shall define a probability measure \mathbb{P} as a function assigning a number $\mathbb{P}(A)$ to each event in A in a field \mathcal{F} . For all events A , \mathbb{P} will satisfy the following conditions,

- (i) $\mathbb{P}(\Omega) = 1$;
- (ii) $0 \leq \mathbb{P}(A) \leq 1$;
- (iii) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ for incompatible events A, B ;
- (iv) $\mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$ for pairwise incompatible events A_1, \dots, A_n (*finite additivity*);

By requiring that if A_1, \dots, A_i, \dots is a countably infinite sequence of pairwise incompatible events, then $\bigcup_{i=1}^{\infty} A_i$ is an event. Moreover,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i). \quad (1)$$

Finitely Additive Probability Space

Let \mathcal{F} be a field of sets on Ω . Then, the triple $(\Omega, \mathcal{F}, \mathbb{P})$ is a *finitely additive probability space* iff \mathbb{P} is a real valued function on \mathcal{F} satisfying the following three conditions: For all $A, B \in \mathcal{F}$,

- [K1] $\mathbb{P}(\Omega) = 1$;
- [K2] $\mathbb{P}(A) \geq 0$;
- [K3] if $A \cap B = \emptyset$, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$;

The function \mathbb{P} is called a *probability measure*. In the special case where \mathcal{F} is a σ -field,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) \quad (2)$$

holds for any countable family $\{A_i | i \in \mathbb{N}\}$ of pairwise disjoint events, then $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space*.

Probability and Counting Measure

Suppose that Ω is a finite set, and let \mathcal{F} be a field of its subsets. For any event $A \in \mathcal{F}$, we define

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$$

The value of the probability measure on \mathbb{P} for a particular event A is the ratio of the number of elements of A to the total number of elements in the sample space Ω . The probability measure is sometimes called the *counting measure* and is appropriate when it makes sense to attribute an equal weight or likelihood to each point of the sample space.

Probability Distribution

Definition: Let Ω be a finite or countable sample space. Then p is a *probability distribution* on Ω iff p is a real valued function on Ω satisfying

$$\begin{aligned} p(x) &\geq 0 \text{ for all } x \in \Omega; \\ \sum_{x \in \Omega} p(x) &= 1. \end{aligned}$$

Theorem: Let \mathcal{F} be a field of sets on a finite or countable sample space Ω , and let p be a probability distribution on Ω . Define a function \mathbb{P} on \mathcal{F} by

$$\mathbb{P}(A) = \sum_{x \in A} p(x), \quad (3)$$

for all $A \in \mathcal{F}$. Then $(\Omega, \mathcal{F}, \mathbb{P})$ is a *probability space*. When $(A_i)_{i \in I}$ is a sequence of pairwise events, with the index set I being finite or countable, and each of the events A_i being finite or countable,

$$\begin{aligned}
\sum_{i \in I} \mathbb{P}(A_i) &= \sum_{i \in I} \sum_{x \in A_i} p(x) \\
&= \sum_{x \in \bigcup_{i \in I} A_i} p(x) \\
&= \mathbb{P}\left(\bigcup_{i \in I} A_i\right)
\end{aligned}$$

Remarks

The material in this chapter suggests that the notions of a field of events and of a probability measure are only considered with finite or countable sample spaces because the probability of any subset of the sample space could be computed from the probability distribution. With uncountable sample spaces, the notion of a probability distribution is useless. It turns out that ‘probability density functions’ is a conceptual notion related to that of probability distributions.

CHAPTER 5: Basic Rules of Probability Calculus

Theorem: Let Ω be a sample space, and $(\Omega, \mathcal{F}, \mathbb{P})$ a finite additive probability space. We do not assume that Ω or \mathcal{F} are finite. Then, for any events $A, B, C \in \mathcal{F}$,

$$(i) \quad \mathbb{P}(A) + \mathbb{P}(\bar{A}) = 1$$

$$(ii) \quad \mathbb{P}(\emptyset) = 0$$

(iii) If A_1, \dots, A_n are incompatible events in \mathcal{F} , then

$$\mathbb{P}\left(\sum_{j=1}^n A_j\right) = \sum_{j=1}^n \mathbb{P}(A_j)$$

(iv) If $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$

$$(v) \quad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

(vi)

$$\begin{aligned}
\mathbb{P}(A \cup B \cup C) &= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) \\
&\quad - \mathbb{P}(AB) - \mathbb{P}(AC) - \mathbb{P}(BC) \\
&\quad + \mathbb{P}(ABC)
\end{aligned}$$

$$(vii) \quad \mathbb{P}(A\bar{B}) = \mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(AB)$$

Poincaré's Identity

We write $S(k, n)$ for the collection of subsets of the set $\{1, \dots, n\}$ containing exactly k elements, with $1 \leq k \leq n$. For any finite collection A_1, \dots, A_n of events, we have

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=0}^{n-1} (-1)^i \sum_{J \in S(i_1, n)} \mathbb{P}\left(\bigcap_{k \in J} A_k\right)$$

CHAPTER 6

Sampling with Replacement and with Ordering

Let A_1, A_2, \dots, A_n be n finite sets containing $m_1, m_2, \dots, m_n > 0$ points, respectively. Then,

$$|A_1, A_2, \dots, A_n| = m_1 \cdot m_2 \dots \cdot m_n \quad (4)$$

In a set containing $n > 0$ elements, there are exactly n^m ordered samples of size $m \geq 1$, with replacement.

Sampling without Replacement and with Ordering

The number of ordered samples of size m without replacement, in a set of $n \geq m > 1$ elements, is equal to

$$(n)_m = n(n-1) \cdot \dots \cdot (n-m+1) = \frac{n!}{(n-m)!}$$

The first equality defines the notation $(n)_m$. We have thus $(n)_n = n!$.

Stirling's Formula

For any positive integer n , we have

$$n! \sim (2\pi)^{\frac{1}{2}} n^{n+\frac{1}{2}} e^{-n}. \quad (5)$$

CHAPTER 7

Binomial Coefficient

In a set of size $n \geq 0$, the number of subsets of size $m \geq 0$ is equal to

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}.$$

The coefficient $\binom{n}{m}$ is called the binomial coefficient and is used in the *binomial theorem*. It represents the number of ways of splitting a set of n elements into two subsets containing $n-m$ and m elements with $0 \leq m \leq n$.

Binomial Theorem

Let a and b be any two real number, and let n be a positive integer. Then,

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}.$$

We can generalize this theorem to satisfy the following when $0 \leq m \leq n$,

$$\begin{aligned} \binom{n}{m} &= \binom{n}{n-m}; \\ \binom{n}{m} &= \binom{n-1}{m-1} + \binom{n-1}{m}; \\ 2^n &= \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = \sum_{k=0}^n \binom{n}{k}. \end{aligned}$$

Multinomial Coefficient

In attempts to generalize the binomial coefficient, we define the *multinomial coefficient*. It will later be used in the definition of the *multinomial distribution*. There are exactly

$$\binom{n}{m_1 m_2 \dots m_k} = \frac{n!}{m_1! \cdot m_2! \cdot \dots \cdot m_k!}$$

ways of splitting a set containing $n \geq 0$ elements into $k \geq 0$ subsets containing m_1, m_2, \dots, m_k elements with $m_1 + m_2 + \dots + m_k = n$ and $m_i \geq 0$ for $i = 1, 2, \dots, k$.

Multinomial Theorem

Let a_1, a_2, \dots, a_k be any k real numbers, and let n be a positive integer. Then,

$$(a_1 + a_2 + \dots + a_k)^n = \sum_{(m_1, m_2, \dots, m_k)} \binom{n}{m_1 m_2 \dots m_k} a_1^{m_1} a_2^{m_2} \cdots a_k^{m_k},$$

where the summation runs over all k -tuples (m_1, m_2, \dots, m_k) of non-negative integers m_i , $0 \leq i \leq k$, satisfying $\sum_{i=1}^k m_i = n$.

CHAPTER 8: Discrete Distributions

Five cases of discrete distributions are considered in this chapter. In the first three cases, the sample space is finite.

Bernoulli Distribution

Recall that a *probability distribution* is a real valued function p defined on a finite or countable sample space Ω and satisfying the two conditions:

1. $p(x) \geq 0$ for all $x \in \Omega$;
2. $\sum_{x \in \Omega} p(x) = 1$.

Let $\Omega = \{0, 1\}$ be a sample space, and let α be any real number, with $0 \leq \alpha \leq 1$. The probability distribution p defined on Ω by

$$p(x) = \begin{cases} \alpha & \text{if } x = 1 \\ 1 - \alpha & \text{if } x = 0 \end{cases}$$

is called a *Bernoulli distribution* with parameter α . Any probability distribution on the sample space Ω is a Bernoulli distribution because there is always some parameter $\alpha \geq 0$ satisfying the above equation.

Bernoulli Distribution: Empirical Situation

An experimenter is selecting a single ball from an urn containing only black and white balls in proportions α and $1 - \alpha$, respectively. When a black ball is selected, 1 is recorded; otherwise, 0 is recorded.

Binomial Distribution

Let $\Omega = \{0, 1, \dots, n\}$ be a sample space, with n a positive integer, and let α be any real number, with $0 \leq \alpha \leq 1$. The probability distribution p defined on Ω by

$$p(m) = \binom{n}{m} \alpha^m (1 - \alpha)^{n-m}$$

is called a *binomial distribution* with parameters α and n .

Binomial Distribution: Empirical Situation

An experimenter is selecting n balls, with replacement, from an urn containing only black and white balls. The probability of getting a black ball if a single ball is selected is equal to α . The number m of black balls obtained is recorded.

Multinomial Distribution

Let n and k be two integers with $n \geq k > 0$. Let the sample space Ω be the set of all k -tuples (m_1, m_2, \dots, m_k) of non-negative integer satisfying $\sum_{i=1}^k m_i = n$. Let $\alpha_1, \dots, \alpha_k$ be non-negative real numbers satisfying $\sum_{i=1}^k \alpha_i = 1$. The probability distribution p defined on Ω by the equation

$$p(m_1, m_2, \dots, m_k) = \binom{n}{m_1 m_2 \dots m_k} a_1^{m_1} a_2^{m_2} \dots a_k^{m_k}$$

is called the *multinomial distribution* with parameters $\alpha_1, \dots, \alpha_k$ and n .

Multinomial Distribution: Empirical Situation

An experimenter is sampling balls with replacement from an urn containing k different types of balls, numbered $1 \dots k$. If a single ball is selected from the urn, the probability that it is a ball of type i ($1 \leq i \leq k$) is equal to α_i ; thus, ($0 \leq i \leq k$). Suppose n balls are sampled, and that the experimenter records the number m_i of balls of each type.

Geometric Distribution

Let $\Omega = \{1, 2, \dots, n, \dots\}$ be the sample space, and let α be a real number with $0 < \alpha < 1$. The function p defined on Ω by the equation

$$p(n) = \alpha(1 - \alpha)^{n-1}$$

is called the *geometric distribution* with parameter α . The function p is clearly non-negative.

Geometric Distribution: Empirical Situation

An experimenter is drawing balls with replacement from an urn containing black and white balls. We suppose that the probability of drawing a black ball is a constant α . The drawing is continued until the first black ball is drawn. The experimenter only records the number of the particular trial where this occurs. Thus, if the first black ball is drawn on trial 5, the experimenter records the number 5 as the outcome of the experiment.

Poisson Distribution

Let $\Omega = \{0, 1, 2, \dots, n, \dots\}$ be the sample space, and let λ be a positive real number. The probability distribution p defined on Ω by the equation

$$p(k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

is called a *Poisson distribution* with parameter λ .

Poisson Distribution: Empirical Situation

An experimenter is watching a display for the appearance of signals of some kind. We suppose that these signals have the following characteristics: they are punctual; the occurrence of a signal has no effect on the occurrence of the next one; and the occurrence of the signals is random but homogeneous, in the sense that the signal is just as likely to occur within one interval of time of a given length within some other interval of the same length.

CHAPTER 9: Conditional Probability

The conditional probability of some event A ‘given’ an event B is the probability that A occurs when we know for sure that B must also occur (or has occurred). The usual notation is $\mathbb{P}(A | B)$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a finite additive probability space. For all events A, B such that $\mathbb{P}(B) \neq 0$, we define

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

Some Consequences of Conditional Probability

Theorem: Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a finitely additive probability space. For all events A, B , and C , such that $\mathbb{P}(C) \neq 0$, we have

- (i) $\mathbb{P}(\Omega | C) = 1$;
- (ii) $\mathbb{P}(A | C) \geq 0$;
- (iii) if $A \cap B \neq \emptyset$, then $\mathbb{P}(A \cup B | C) = \mathbb{P}(A | C) + \mathbb{P}(B | C)$;
- (iv) if $A \cap B = \emptyset$, then $\mathbb{P}(A | C) = 0$;
- (v) if $\mathbb{P}(A) \neq 0$, then $\mathbb{P}(A | C) = \frac{\mathbb{P}(A)\mathbb{P}(C | A)}{\mathbb{P}(C)}$.

Theorem: Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a finitely additive probability space. Let C be some event such that $\mathbb{P}(C) \neq 0$. For any $A \in \mathcal{F}$, define $\mathbb{P}_C(A) = \mathbb{P}(A | C)$. Then $(\Omega, \mathcal{F}, \mathbb{P}_C)$ is a finitely additive probability space.

When we condition by some event $C \in \mathcal{F}$, with $C \neq \Omega$, it is almost as if we were restricting consideration to a smaller sample space $C \subseteq \Omega$. For example, suppose that $A \cap C$ and $B \cap C$ are two incompatible events, and that $\mathbb{P}(C) \neq 0$. It is not difficult to show that

$$\mathbb{P}(A \cup B | C) = \mathbb{P}(A | C) + \mathbb{P}(B | C),$$

which is essentially the defining additivity property of probabilities holding for smaller sample space C .

Theorem of Total Probabilities

The idea here is that the probability of an event A can in many cases be decomposed additively in terms of the probabilities of other events that *cover* Ω . The idea is to decompose the probability of A through the sum

$$\mathbb{P}(A) = \mathbb{P}(A \cup H_1) + \mathbb{P}(A \cup H_2) + \dots + \mathbb{P}(A \cup H_n).$$

Note that $\mathbb{P}(A | H_i)\mathbb{P}(H_i) = \mathbb{P}(A \cap H_i)$ for $1 \leq i \leq n$. For events E_1, E_2, \dots, E_n that are pairwise incompatible, the notation $\sum_{i=1}^n E_i$ has the same meaning as the notation $\bigcup_{i=1}^n E_i$.

Thus, we can more formally define the *Theorem of Total Probabilities*.

Theorem: Let $(H_i)_{1 \leq i \leq n}$ be a family of pairwise incompatible events in a finitely additive probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that $\sum_{i=1}^n H_i = \Omega$. Then, for all $A \in \mathcal{F}$,

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A \cap H_i).$$

If $\mathbb{P}(H_i) \neq 0$, for all i , $1 \leq i \leq n$, then

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A | H_i) \mathbb{P}(H_i).$$

Remarks

The theorem presented in this section represents results based on a finite set of events. Versions of this theorem also hold in the cases of countable and uncountable collections of events H_i , but the probabilities of joint events $A \cap E_i$ are replaced by a joint density function.

CHAPTER 10: Independence & Bayes' Theorem

In a finite additive probability space $(\Omega, \mathcal{F}, \mathbb{P})$, two events A, B are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B).$$

An event A is *independent* of some event B when knowing that B is realized does not affect the probability of A . It is not required that $\mathbb{P}(A), \mathbb{P}(B) > 0$. However, if $\mathbb{P}(A) = 0$, then $\mathbb{P}(A \cap B) = 0$, since $A \cap B \subseteq A$, and $\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B) = 0$. An event of probability zero is thus independent of any other event. Suppose that A and B are two independent events in a finite additive probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then,

- (i) \bar{A} and B are independent;
- (ii) if $\mathbb{P}(B) \neq 0$, then $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Bayes Theorem

Let $(H_i)_{1 \leq i \leq n}$ be a family of pairwise incompatible events in a finitely additive probability space $(\Omega, \mathcal{F}, \mathbb{P})$, satisfying $\bigcup_{i=1}^n H_i = \Omega$ and $\mathbb{P}(H_i) \neq 0$, for all $i, 1 \leq i \leq n$. If $D \in \mathcal{F}$ is any event such that $\mathbb{P}(D) \neq 0$, then, for $1 \leq i \leq n$,

$$\mathbb{P}(H_i|D) = \frac{\mathbb{P}(D|H_i) \mathbb{P}(H_i)}{\sum_{j=1}^n \mathbb{P}(D|H_j) \mathbb{P}(H_j)}.$$

Remarks

The events H_i are called *hypotheses* and $\mathbb{P}(H_i)$ is called *a priori probability* of hypothesis H_i . The conditional probability $\mathbb{P}(H_i \mid D)$ is referred to as the *a posteriori probability* provided by the realization of the event D . Suppose that some experiment has been performed, and some ‘data’ (the event D) have been collected. A number of alternative ‘theories’ are considered. An intuitively appealing question is: what are the ‘respective probabilities’ of the various theories, given the data? Bayes Theorem suggests that such ‘probabilities’ might perhaps be recomputed from the a priori probabilities of these hypotheses via the system of equations for $1 \leq i \leq n$. For instance, in what sense can we talk about the ‘probability’ of a theory?