

Probability I

Lecture 2

Davide La Vecchia

Spring Semester 2020

Definition

We define probability a set function with values in $[0, 1]$, which satisfies the following axioms:

- (i) $P(A) \geq 0$, for every event A
- (ii) $P(S) = 1$
- (iii) If A_1, A_2, \dots is a sequence of mutually exclusive events (namely $A_i \cap A_j = \emptyset$, for $i \neq j$, and $i, j = 1, 2, \dots$), and such that $A = \bigcup_{i=1}^{\infty} A_i$, then

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

Properties of $P(\cdot)$

One can make use of the three axioms to build more sophisticated statements. For instance,

Theorem

$$P(\emptyset) = 0.$$

Proof.

Take $A_1 = A_2 = A_3 = \dots = \emptyset$. Then by (1) in Axiom (ii) we have

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

which is true only if it is an infinite sum of zeros. Thus

$$P(\emptyset) = 0.$$



Properties of $P(\cdot)$ (cont'd)

Theorem

If A_1, A_2, \dots are mutually exclusive events, then

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

Proof.

Let $A_{n+1} = A_{n+2} = \dots = \emptyset$, then $\bigcup_{i=1}^n A_i = \bigcup_{i=1}^{\infty} A_i$, and, from (1) (see Axiom (iii)) it follows

$$P\left(\bigcup_{i=1}^n A_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^n P(A_i) + \underbrace{\sum_{i=n+1}^{\infty} P(A_i)}_{\equiv 0}.$$



Properties of $P(\cdot)$ (cont'd)

Theorem

If A is an event, then $P(A^c) = 1 - P(A)$.

Proof.

Let $A \cup A^c = S$ and $A \cap A^c = \emptyset$, so and

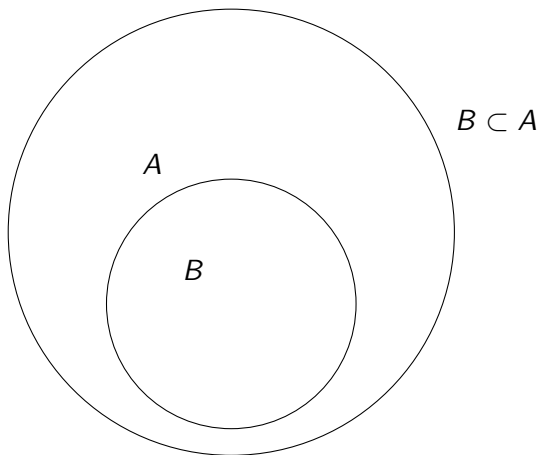
$$P(S) = P(A \cup A^c) = P(A) + P(A^c).$$

By Axiom (i) Axiom (ii) we have $P(S) = 1$, so the desired result follows from:

$$1 = P(A) + P(A^c).$$



Properties of $P(\cdot)$ (cont'd)



Properties of $P(\cdot)$ (cont'd)

Theorem

For any two events A and B , such that $B \subset A$, we have

$$P(A) \geq P(B).$$

Proof.

Let us write

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

and notice that $B \cap (B^c \cap A) = \phi$, so that

$$\begin{aligned} P(A) &= P\{B \cup (B^c \cap A)\} \\ &= P(B) + P(B^c \cap A) \end{aligned}$$

which implies (since $P(B^c \cap A) \geq 0$) that

$$P(A) \geq P(B).$$



Properties of $P(\cdot)$ (cont'd)

Theorem (Boole's inequality)

For the events A_1, A_2, \dots, A_n ,

$$P(A_1 \cup A_2 \cup \dots \cup A_n) \leq \sum_{i=1}^n P(A_i).$$

For instance, let us consider $n = 2$. Then we have:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \leq P(A_1) + P(A_2)$$

since $P(A_1 \cap A_2) \geq 0$ by definition.

Remark

It is worth notice that if $A_j \cap A_i = \emptyset$, for every i and j , with $i \neq j$, then $P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i)$, as stated in (2).

Properties of $P(\cdot)$ (cont'd)

Theorem

For any two events A and B then

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

Proof.

Consider that $A \cup B = A \cup (A^c \cap B)$, and $A \cap (A^c \cap B) = \phi$. Now remember^a that $A^c \cap B = B - (A \cap B)$, so,

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



^aSee Lecture 1 for the meaning of set difference.

How to use these properties?

Example (Real-life example I)

If we flip a balanced coin twice, what is the probability of getting at least one head?

The sample space is $\{HH, HT, TH, TT\}$

Since the coin is balanced, these outcomes are equally likely and we assign to each sample point probability $\frac{1}{4}$

A=event that we get at least one H = $\{HH, HT, TH\}$

$$\begin{aligned}\Pr\{A\} &= \Pr\{HH \cup HT \cup TH\} = \Pr\{HH\} + \Pr\{HT\} + \Pr\{TH\} \\ &= 1/4 + 1/4 + 1/4 = 3/4\end{aligned}$$

How to use these properties?

Example (Real-life example II)

Shopper TRK is an electronic device designed to count the number of shoppers entering a shopping centre

When two shoppers enter the shopping centre together, one walking in front of the other the following probabilities apply

- 1) there is a 0.98 probability that the first shopper is detected
- 2) there is a 0.94 probability that the second shopper is detected
- 3) there is a 0.93 probability that both shoppers are detected

What is the probability that the device will detect at least one of the two shoppers entering?

How to use these properties?

Example (Real-life example II, cont'd)

Sample space $\{DD, DU, UD, UU\}$

$$1) \Pr\{DD \cup DU\} = 0.98$$

$$2) \Pr\{DD \cup UD\} = 0.94$$

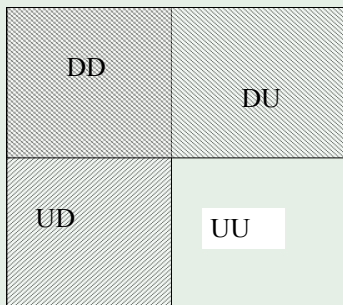
$$3) \Pr\{DD\} = 0.93$$

$$\Pr\{DD \cup UD \cup DU\} = ?$$

How to use these properties?

Example (Real-life example II, cont'd)

$$\begin{aligned}\Pr\{DDUUDUDU\} &= \Pr\{(DDUUD) \cup (DDUDU)\} \\ &= \Pr\{DDUUD\} + \Pr\{DDUDU\} - \Pr\{(DDUUD) \cap (DDUDU)\}\end{aligned}$$



$$(DDUUD) \cap (DDUDU) = DD$$

How to use these properties?

Example (Real-life example II, cont'd)

So the desired probability is

$$\begin{aligned}\Pr\{DD\cup UD\cup DU\} &= \Pr\{DD\cup UD\} + \Pr\{DD\cup DU\} - \Pr\{DD\} \\ &= 0.98 + 0.94 - 0.93 = 0.99\end{aligned}$$

How to use these properties?

Example (De Morgan's law)

Given $P(A \cup B) = 0.7$ and $P(A \cup B^c) = 0.9$, find $P(A)$.

By De Morgan's law,

$$P(A^c \cap B^c) = P((A \cup B)^c) = 1 - P(A \cup B) = 1 - 0.7 = 0.3$$

and similarly

$$P(A^c \cap B) = 1 - P(A \cup B^c) = 1 - 0.9 = 0.1.$$

Thus,

$$P(A^c) = P(A^c \cap B^c) + P(A^c \cap B) = 0.3 + 0.1 = 0.4,$$

so

$$P(A) = 1 - 0.4 = 0.6.$$

How to use these properties?

Example (Probability, union, and complement)

John is taking two books along on his holiday vacation. With probability 0.5, he will like the first book; with probability 0.4, he will like the second book; and with probability 0.3, he will like both books. What is the probability that he likes neither book?

Let A_i be the event that John likes book i , for $i = 1, 2$. Then the probability that he likes at least one book is (remember the short hand notation $A_1 \cap A_2 = A_1 A_2$)

$$\begin{aligned} P\left(\bigcup_{i=1}^2 A_i\right) &= P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 A_2) \\ &= 0.5 + 0.4 - 0.3 = 0.6. \end{aligned}$$

Because the event the John likes neither books is the complement of the event that he likes at least one of them (namely $A_1 \cup A_2$), we have

$$P(A_1^c \cap A_2^c) = P((A_1 \cup A_2)^c) = 1 - P(A_1 \cup A_2) = 0.4.$$

How to use these properties?

Example (Real-life example III)

Let A denote the event that the midtown temperature in Los Angeles (LA) is 70 F, and let B denote the event that the midtown temperature in New York (NY) is 70 F. Also, let C denote the event that the maximum of midtown temperatures in NY and LA is 70 F. If

$$P(A) = 0.3, \quad P(B) = 0.4 \quad \text{and} \quad P(C) = 0.2,$$

find the probability that the minimum of the two midtown temperatures is 70 F.

How to use these properties?

Example (cont'd)

Let D denote the event that the minimum temperature is 70 F. Then

$$P(A \cup B) = P(A) + P(B) - P(AB) = 0.7 - P(AB)$$

$$P(C \cup D) = P(C) + P(D) - P(CD) = 0.2 - P(D) - P(DC).$$

Since

$$A \cup B = C \cup D \quad \text{and} \quad AB = CD,$$

subtracting one of the preceding equations from the other we get

$$\begin{aligned} P(A \cup B) - P(C \cup D) &= 0.7 - P(AB) - [0.2 - P(D) - P(DC)] \\ &= 0.5 - P(D) = 0, \end{aligned}$$

thus $P(D) = 0.5$.

Conditional probability

As a measure of uncertainty, the probability depends on the information available.

Example

Suppose you have two dice and throw them; the possible outcomes are:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(6,5)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Conditional probability

Example (cont'd)

Let us define $A = \text{getting } 5$, or equivalently $A = \{5\}$. What is $P(A)$, namely, the probability of getting 5?

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(6,5)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Conditional probability

Example (cont'd)

The dice are fair so we can get 36 events with equal probability $1/36$.
Namely:

$$Pr(i, j) = \frac{1}{36}, \quad \text{for } i, j = 1, \dots, 6$$

Thus, we can make use of the blue guys as

$$\begin{aligned} P(5) &= Pr\{(1, 4) \cup (2, 3) \cup (3, 2) \cup (4, 1)\} \\ &= Pr\{(1, 4)\} + Pr\{(2, 3)\} + Pr\{(3, 2)\} + Pr\{(4, 1)\} \\ &= 1/36 + 1/36 + 1/36 + 1/36 \\ &= 4/36 \\ &= 1/9. \end{aligned}$$

Conditional probability

Example (cont'd)

Now, suppose that we throw the die first and then we get 2.

Q. What is the probability of getting 5 given that we have observed 2 in the first throw?

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

A.

$$\Pr\{\text{getting 5 given 2 in the first throw}\} = \Pr\{\text{getting 5 in the second throw}\} = 1/6$$

Conditional probability

Remark

- *Information changes the probability*
- *By knowing that we got 2 in the first throw, we have changed the sample space:*

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(6,5)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

- *Probability can change drastically — e.g., suppose that in our example we have 6 in the first throw \Rightarrow the probability of observing 5 in two draws ... is zero !!!*

Definition

Let A and B be two events. The conditional probability of event A given event B , denoted by $P(A|B)$, is defined by:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{if } P(B) > 0,$$

and it is left undefined if $P(B) = 0$.

Conditional probability

Example (A check)

Let us define the set B as:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(6,5)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

So we have

$$\begin{aligned}P(B) &= \Pr\{(2,1) \cup (2,2) \cup (2,3) \cup (2,4) \cup (2,5) \cup (2,6)\} \\&= \Pr(2,1) + \Pr(2,2) + \Pr(2,3) + \Pr(2,4) + \Pr(2,5) + \Pr(2,6) \\&= 6/36 = 1/6\end{aligned}$$

Conditional probability

Example (A check)

and consider $A \cap B$

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

So we have $P(A \cap B) = Pr(2, 3) = 1/36$ so,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/36}{1/6} = \frac{1}{6}.$$

Definition

Two events A and B are independent if the occurrence of one event has no effect on the probability of occurrence of the other event. Thus,

$$P(A|B) = P(A)$$

or equivalently

$$P(B|A) = P(B)$$

Clearly, if $P(A|B) \neq P(A)$, then A and B are *dependent*.

Independence – another characterization

Two events A and B are independent if

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

now by definition of conditional probability we know that

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

so we have

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

and rearranging the terms, we find that two events are independent iff

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

Example

A coin is tossed three times and the eight possible outcomes
HHH, HHT, HTH, THH, HTT, THT, TTH, TTT
are assumed to be equally likely with probability $1/8$

Define

A = an H occurs on each of the first two tosses

B = T occurs in the third toss

D = two Ts occur in the three tosses

Q1: Are A and B independent?

Q2: Are B and D independent?

Example (cont'd)

We have

$$A = \{HHH, HHT\}$$

$$\Pr\{A\} = 2/8 = 1/4$$

$$B = \{HHT, HTT, THT, TTT\}$$

$$\Pr\{B\} = 4/8 = 1/2$$

$$D = \{HTT, THT, TTH\}$$

$$\Pr\{D\} = 3/8$$

$$A \cap B = \{HHT\}$$

$$\Pr\{A \cap B\} = 1/8$$

$$B \cap D = \{HTT, THT\}$$

$$\Pr\{B \cap D\} = 2/8 = 1/4$$

$$\Pr\{A\}\Pr\{B\} = \frac{1}{4} \times \frac{1}{2} = \frac{1}{8} \quad \Pr\{A \cap B\} = \frac{1}{8} \Rightarrow A, B \text{ independent}$$

$$\Pr\{B\}\Pr\{D\} = \frac{1}{2} \times \frac{3}{8} = \frac{3}{16} \quad \Pr\{B \cap D\} = \frac{1}{4} \Rightarrow B \text{ and } D \text{ dependent}$$

Theorem I

Theorem (Th. of total probabilities)

Let $B_1, B_2, \dots, B_k, \dots, B_n$ be mutually disjoint events, satisfying $S = \cup_{i=1}^n B_i$, and $P(B_i) > 0$, for every $i = 1, 2, \dots, n$ then for every A we have that:

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i). \quad (3)$$

Proof.

Write $A = A \cap S = A \cap (\cup_{i=1}^n B_i) = \cup_{i=1}^n (A \cap B_i)$. Since the $\{B_i \cap A\}$ are mutually disjoint, we have

$$P(A) = P(\cup_{i=1}^n (A \cap B_i)) = \sum_{i=1}^n P(A \cap B_i) = \sum_{i=1}^n P(A|B_i)P(B_i).$$



Theorem I (cont'd)

Remark. The theorem remains valid even if $n = \infty$ in Eq. (3). (Double check, and re-do the proof using $n = \infty$.)

Corollary

Let B satisfy $0 < P(B) < 1$; then for every event A :

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

Proof.

Exercise [Hint: $S = B \cup B^c$].



Theorem II

Theorem I can be applied to derive the well-celebrated Bayes' Theorem.

Theorem (Bayes' Theorem)

Let $B_1, B_2, \dots, B_k, \dots, B_n$ be mutually disjoint events, satisfying

$$S = \cup_{i=1}^n B_i,$$

and $P(B_i) > 0$, for every $i = 1, 2, \dots, n$. Then for every event A for which $P(A) > 0$, we have that

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^n P(A|B_i)P(B_i)}. \quad (4)$$

Theorem II (cont'd)

Proof.

Let us write

$$\begin{aligned}P(B_k|A) &= \frac{P(A \cap B_k)}{P(A)} \\&= \frac{P(A \cap B_k)}{\sum_{i=1}^n P(A|B_i)P(B_i)} \\&= \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^n P(A|B_i)P(B_i)}\end{aligned}$$

That concludes the proof. □

Theorem II (cont'd)

Example

Let us consider a special case, where we have only two events A and B .

From the definition of conditional probability

$$\Pr\{A|B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}} \qquad \Pr\{B|A\} = \frac{\Pr\{A \cap B\}}{\Pr\{A\}}$$

This can be written as

$$\Pr\{A \cap B\} = \Pr\{A|B\} \Pr\{B\} \qquad \Pr\{A \cap B\} = \Pr\{B|A\} \Pr\{A\}$$

That is

$$\Pr\{A|B\} \Pr\{B\} = \Pr\{B|A\} \Pr\{A\}$$

Rearranging this we have

$$\Pr\{A|B\} = \frac{\Pr\{A\}}{\Pr\{B\}} \times \Pr\{B|A\} \quad \Leftarrow \text{Bayes' Theorem}$$

... so thanks to Bayes' Theorem we can reverse the role of $A|B$ and $B|A$.

Example (Guessing in a multiple choice exam)

Example 1.13 In answering a question on a multiple choice test a student either knows the answer or he guesses. Let p be the probability that she knows the answer and $1 - p$ the probability that she guesses. Assume that a student who guesses at the answer will be correct with probability $1/m$, where m is the number of multiple-choice alternatives. What is the conditional probability that a student knew the answer to a question given that she answered it correctly?

Solution: Let C and K denote respectively the event that the student answers the question correctly and the event that she actually knows the answer. Now

$$\begin{aligned}P(K|C) &= \frac{P(KC)}{P(C)} = \frac{P(C|K)P(K)}{P(C|K)P(K) + P(C|K^c)P(K^c)} \\&= \frac{p}{p + (1/m)(1 - p)} \\&= \frac{mp}{1 + (m - 1)p}\end{aligned}$$

Thus, for example, if $m = 5$, $p = \frac{1}{2}$, then the probability that a student knew the answer to a question she correctly answered is $\frac{5}{6}$. ♦

Another real-life example

Example

Members of a consulting firm in Geneva rent cars from two rental agencies:

- 60% from AVIS
- 40% from Mobility

Now consider that

- 9% of the cars from AVIS need a tune-up
- 20% of the cars from Mobility need a tune-up

If a car delivered to the consulting firm needs a tune-up, what is the probability that the care came from AVIS?

Aim

Let us set: $A := \{\text{car rented from AVIS}\}$ and $B := \{\text{car needs a tune-up}\}$. We know $P(B|A)$ and we look for $P(A|B) \Rightarrow$ Bayes' theorem!!

A real-life example (cont'd)

Example (cont'd)

$$\Pr\{A\}=0.6$$

$$\Pr\{B|A\}=0.09$$

$$\Pr\{B|A^C\}=0.2$$

$$\Pr\{B\}=\Pr\{(B\cap A)\cup(B\cap A^C)\}$$

$$=\Pr\{B\cap A\}+\Pr\{B\cap A^C\}$$

$$=\Pr\{B|A\}\Pr\{A\}+\Pr\{B|A^C\}\Pr\{A^C\}$$

$$=0.09\times 0.6+0.2\times 0.4$$

$$=0.134$$

$$\Pr\{A|B\}=\frac{\Pr\{A\}}{\Pr\{B\}}\times\Pr\{B|A\}=\frac{0.6}{0.134}0.09=0.402985$$