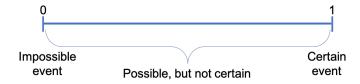


Statistics

Probability

Luca Pennella September 12th, 2024

Probability helps prevent randomness from being masked and perceived as non-randomness.



Probability is related to uncertain events:

- Ball landing on black on the roulette wheel
- The weather tomorrow in Trieste being sunny
- Inter soccer team winning the Italian league
- The guilt of a defendant

Intro Rules Ex1 Conditional probability Independen

Introduction to probability: subjective interpretation

Bruno de Finetti in "Theory of probability" (1970)

1.3.1. Meanwhile, for those who are not aware of it, it is necessary to mention that in the conception we follow and sustain here only **subjective probabilities** exist – that is, the degree of belief in the occurrence of an event attributed by a given person at a given instant and with a given set of information. This is in contrast to other conceptions that limit themselves to special types of cases in which they attribute meaning to 'objective probabilities' (for instance, cases of symmetry as for dice etc., 'statistical' cases of 'repeatable' events, etc.).

The idea of probability is empirical: it is based on observation rather than theorizing.

Probability models are used to study the variation in observed data so that inferences about the underlying process can be developed.

- ▶ De Finetti's idea emphasizes that probability is a measure of individual's degree of belief about an event occurring.
- ► Psychologists have shown experimentally that heuristics, which are essentially mental shortcuts, often lead to errors when assessing uncertainty
- Individuals may rely on these mental shortcuts, which can be influenced by biases and cognitive errors, when assigning probabilities subjectively
- ► I toss a coin six times and record the outcome of Heads or Tails. Which result is more likely?

HHHHTTTT or HTTHHTHH

Both sequences have an equal probability of occurring in a fair coin toss!

- By saying that probability is the 'degree of belief' we have given a good definition, but this is not of help when it comes to numerically determining this degree of belief.
- ▶ What is the probability that it will rain tomorrow?
 - By guessing?
 - Average monthly rainfall: 28%, meaning P(rain) = 0.28
 - OSMER: 5%, meaning P(rain) = 0.05

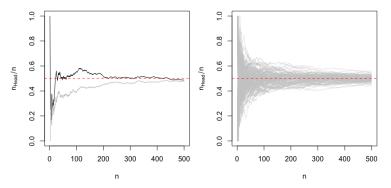
- ► All evaluations are valid: they are opinions, each legitimate, perhaps some more reasonable or more informed than others.
- ► For example, OSMER, for a relatively trivial event like 'rain tomorrow,' uses complex meteorological models and combines a lot of information.
- ► For now, we will look at some cases where the evaluation is simple and intuitive. In such situations, it will be easy to understand some rules for combining probabilities using the four operations and a bit of logic.

Frequentist interpretation

- ▶ Probability is not something directly observable
- ► To have a way of determining it, we can try to link it to something observable
- lackbox Let's consider an event E as E = 'getting Heads (H) in a coin toss'
- ► This event is repeatable, meaning we can toss a coin many times. Let's do it, or imagine doing it, and at each toss, calculate the percentage of heads observed up to that point.

Frequentist interpretation

Here are the outcomes from 500 tosses: chance behaviour is unpredictable in the short run but has a regular and predictable pattern in the long run



The **probability** of any outcome of a random phenomenon is the proportion of times the outcome would occur in a very long series of repetitions

Random experiment

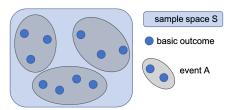
A random experiment/phenomenon is a process leading to two or more possible outcomes, without knowing exactly which one will occur

Sample space

The **sample space** S of a random experiment is the set of all the possible **basic outcomes**, that is, outcomes that can not occur together

Event

An **event** A is an outcome or a set of possible outcomes



Relative frequency probability

The **relative frequency probability** is the limit of the proportion of times that an event A occurs in a large number of trials, n,

$$P(A) = \frac{n_A}{n}$$

where n_A is the number of A outcomes and n is the total number of trials or outcomes.

Classical probability

The **classical probability** is the proportion of times that an event will occur, assuming that all outcomes in a sample space are equally likely to occur

$$P(A) = \frac{N_A}{N}$$

where N_A is the number of outcomes that satisfy A, and N is the total number of outcomes in the sample space.

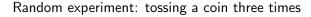
Sample space: $S = \{1, 2, 3, 4, 5, 6\}$

Event:
$$A = \text{get a } 6$$

$$P(A) = \frac{N_A}{N} = \frac{1}{6}$$

Event: $B = get an even number = \{2, 4, 6\}$

$$P(B) = \frac{N_B}{N} = \frac{3}{6} = \frac{1}{2}$$



$$S = \{HHH, HHT, HTT, TTT, TTH, THH, THT, HTH\}$$

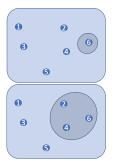
$$A = \text{two heads} = \{HHT, THH, HTH\}$$

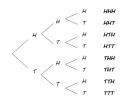
$$P(A) = \frac{N_A}{N} = \frac{3}{8}$$

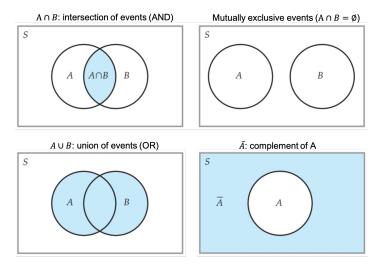
 $B = \text{more than one head} = \{HHT, THH, HTH, HHH\}$

$$P(B) = \frac{N_B}{N} = \frac{1}{2}$$

Elementary events can be combined to define more complex events, that is, **compound events**. There are rules for combining the probabilities of elementary events to calculate the robability of compound events.







Collectively exhaustive: given K events E_1, E_2, \ldots, E_K in the sample space S, if $E_1 \cup E_2 \cup \ldots \cup E_K = S$, the K events are said to be collectively exhaustive

Roulette:
$$S = \{0,1,..., 36\}$$

$$A= first column = \{1,4,7,\ldots, 34\}$$

$$\mathsf{B} {=}\ \mathsf{red} = \{1,\!3,\!5,\!7,\!8,\!12,\!14,\!16,\!18,\!\dots\}$$

$$C = \{0\}$$

$$A \cap B =$$
first column AND red
= $\{1, 7, 16, 19, 25, 34\}$

$$A \cup B$$
 = first column OR red
= $\{1, 3, 4, 5, 7, 9, 10, 12, ...\}$
= $A + B - (A \cap B)$

$$A \cap C = first column AND \{0\} = \emptyset$$

$$A \cup C$$
 = first column OR $\{0\} = \{0, 1, 4, 7, \dots, 34\}$
= $A + C$

$$\bar{C} = \{1, 2, 3, \dots, 36\} = S - C$$



The set of all basic outcomes contained in a sample space is mutually exclusive and collectively exhaustive

Postulates

- $ightharpoonup 0 \le P(A) \le 1$: any probability of an event is a number between 0 and 1
- $P(A) = \sum_{\Delta} P(\text{basic outcome}_i) = P(O_1) + P(O_2) + \dots$ basic outcomes are mutually exclusive
- ightharpoonup P(S) = 1: all possible basic outcomes together must have probability 1, they are collectively exhaustive
- ▶ Complement rule: $P(\bar{A}) = 1 P(A)$, note that A and \bar{A} are mutually exclusive and collectively exhaustive
- Addition rule of probability: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Complement rule

$$P(\bar{A}) = 1 - P(A)$$

Note that A and \bar{A} are collectively exhaustive

$$A \cup \bar{A} = S$$

and that A and \bar{A} are mutually exclusive:

$$P(A \cup \bar{A}) = P(A) + P(\bar{A})$$

then

$$P(A) + P(\bar{A}) = P(A \cup \bar{A}) = P(S) = 1$$

Also follows that $P(\emptyset) = 0$

Addition rule of probability

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

Note that $A \cup B = A \cup (\bar{A} \cap B)$, A and $\bar{A} \cap B$ are mutually exclusive, then

$$P(A \cup B) = P(A) + P(\bar{A} \cap B)$$

Moreover, we can write $B = (A \cap B) \cup (\bar{A} \cap B)$ with $(A \cap B)$ and $(\bar{A} \cap B)$ mutually exclusive

$$P(B) = P(A \cap B) + P(\bar{A} \cap B) \rightarrow P(\bar{A} \cap B) = P(B) - P(A \cap B)$$

Then,

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

Random experiment: rolling a dice

Sample space: $S = \{1, 2, 3, 4, 5, 6\}$

Event: A = get a 6

Event: B = get an even number $= \{2, 4, 6\}$

$$P(A) = \frac{1}{6}$$

$$P(B) = P(\text{get a 2}) + P(\text{get a 4}) + P(\text{get a 6}) = \frac{3}{6} = \frac{1}{2}$$

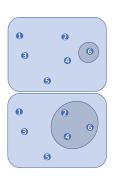
$$P(S) = P(\text{get a } 1) + P(\text{get a } 2) + ... + P(\text{get a } 6) = 1$$

$$P(\bar{A}) = 1 - \frac{1}{6} = \frac{5}{6}$$

$$P(A \cap B) = P(\text{get a 6}) = \frac{1}{6}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{6} + \frac{1}{2} - \frac{1}{6} = \frac{1}{2}$$

Note that A implies B ($A \subset B$), then $P(A) \leq P(B)$



Roulette:
$$S = \{0,1,..., 36\}$$

$$A= first \ column= \{1,4,7,\ldots,\ 34\}$$

$$\mathsf{B} = \mathsf{red} = \{1,3,5,7,8,12,14,16,18,\dots\}$$

$$C = \{0\}$$

$$P(A \cap B) = P(\text{first column AND red}) = \frac{6}{37}$$

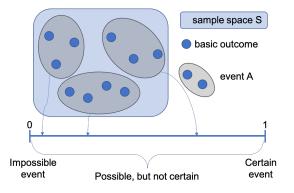
$$P(A \cup B) = P(\text{first column OR red}) = \frac{24}{37}$$

$$P(A \cap C) = P(\text{first column AND}\{0\}) = 1$$

$$P(A \cup C) = P(\text{first column OR}\{0\}) = \frac{13}{37}$$

$$P(\bar{C}) = P(S) - P(C) = 1 - \frac{1}{37} = \frac{36}{37}$$





Probability model

A **probability model** is a mathematical description of a random experiment consisting of a sample space and a way of assigning probabilities to events

It is a convenient way to describe the distribution of an experiment's outcomes and involves listing all possible outcomes and their associated probabilities.

Canada has two official languages, English and French. Choose a Canadian randomly and ask, "What is your mother tongue?". Here is the distribution of responses, combining many separate languages from the broad Asian/Pacific region:

Language	English	French	Asian/Pacific	Other
Probability	0.59	0.23	0.07	?

- a) Complete the table
- b) What is the probability that a Canadian's mother tongue is not English?

$$S = \{English, French, Asian/Pacific, Other\}$$

Canada has two official languages, English and French. Choose a Canadian randomly and ask, "What is your mother tongue?". Here is the distribution of responses, combining many separate languages from the broad Asian/Pacific region:

Language	English	French	Asian/Pacific	Other
Probability	0.59	0.23	0.07	?

- a) Complete the table
- b) What is the probability that a Canadian's mother tongue is not English?

$$S = \{English, French, Asian/Pacific, Other\}$$

$$P(\overline{\textit{English}}) = P(\textit{French}) + P(\textit{Asian/Pacific}) + P(\textit{Other})$$

equivalently

$$P(\overline{\textit{English}}) = 1 - P(\textit{English}) = .41$$

The sales manager wants to estimate the probability that a car will be returned for service during the warranty period. The following table shows a manager's probability assessment for the number of returns

Number of returns	0	1	2	3	4
Probability	0.28	0.36	0.23	0.09	0.04

Let A be the event "the number of returns will be more than two", and let B be the event "the number of returns will be less than four".

- a) Find the probability of event A
- b) Find the probability of event B
- c) Find the probability of the complement of event A
- d) Find the probability of the intersection of events A and B
- e) Find the probability of the union of events A and B

Number of returns	0	1	2	3	4
Probability	0.28	0.36	0.23	0.09	0.04

Let A be the event "the number of returns will be more than two", and let B be the event "the number of returns will be less than four".

a) Find the probability of event A

Number of returns	0	1	2	3	4
Probability	0.28	0.36	0.23	0.09	0.04

Let A be the event "the number of returns will be more than two", and let B be the event "the number of returns will be less than four".

a) Find the probability of event A

$$P(A) = P(\text{returns more than 2}) = P(\text{returns} > 2)$$

= $P(\text{returns} = 3) + P(\text{returns} = 4)$

b) Find the probability of event B

Number of returns 0 1 2 3 4 Probability 0.28 0.36 0.23 0.09 0.04

Let A be the event "the number of returns will be more than two", and let B be the event "the number of returns will be less than four".

a) Find the probability of event A

$$P(A) = P(\text{returns more than 2}) = P(\text{returns} > 2)$$

= $P(\text{returns} = 3) + P(\text{returns} = 4)$

b) Find the probability of event B

$$P(B) = P(\text{returns less than 4}) = P(\text{returns} < 4)$$

= $P(\text{returns} = 0) + P(\text{returns} = 1) + P(\text{returns} = 2) + P(\text{returns} = 3)$

or

Number of returns	0	1	2	3	4
Probability	0.28	0.36	0.23	0.09	0.04

Let *A* be the event "the number of returns will be more than two", and let *B* be the event "the number of returns will be less than four".

a) Find the probability of event A

$$P(A) = P(\text{returns more than 2}) = P(\text{returns} > 2)$$

= $P(\text{returns} = 3) + P(\text{returns} = 4)$

b) Find the probability of event B

$$P(B) = P(\text{returns less than 4}) = P(\text{returns} < 4)$$

= $P(\text{returns} = 0) + P(\text{returns} = 1) + P(\text{returns} = 2) + P(\text{returns} = 3)$

or

$$P(B) = P(\text{returns less than 4}) = 1 - P(\text{returns} = 4)$$

Number of returns	0	1	2	3	4
Probability	0.28	0.36	0.23	0.09	0.04

Let A be the event "the number of returns will be more than two", and let B be the event "the number of returns will be less than four".

c) Find the probability of the complement of event A

Number of returns	0	1	2	3	4
Probability	0.28	0.36	0.23	0.09	0.04

Let A be the event "the number of returns will be more than two", and let B be the event "the number of returns will be less than four".

c) Find the probability of the complement of event A

$$P(\bar{A}) = 1 - P(A) = 1 - P(\text{returns more than 2})$$

d) Find the probability of the intersection of events A and B

Number of returns 0 1 2 3 4 Probability 0.28 0.36 0.23 0.09 0.04

Let A be the event "the number of returns will be more than two", and let B be the event "the number of returns will be less than four".

c) Find the probability of the complement of event A

$$P(\bar{A}) = 1 - P(A) = 1 - P(\text{returns more than 2})$$

d) Find the probability of the intersection of events A and B

$$P(A \cap B) = P(\text{returns more than 2 AND returns less than 4}) = P(\text{returns} = 3)$$

d) Find the probability of the union of events A and B

Number of returns	0	1	2	3	4
Probability	0.28	0.36	0.23	0.09	0.04

Let A be the event "the number of returns will be more than two", and let B be the event "the number of returns will be less than four".

c) Find the probability of the complement of event A

$$P(\bar{A}) = 1 - P(A) = 1 - P(\text{returns more than 2})$$

d) Find the probability of the intersection of events A and B

$$P(A \cap B) = P(\text{returns more than 2 AND returns less than 4}) = P(\text{returns} = 3)$$

d) Find the probability of the union of events A and B

$$P(A \cup B) = P(\text{returns more than 2 OR returns less than 4}) = 1$$

A cell phone company found that 75% of customers want text messaging on their phone, 80% want photo capability, and 65% want both.

What is the probability that a customer will want at least one of these?

Define the events:

A =costumer wants text messaging

B =costumer wants photo capability

 $A \cap B =$ customer wants text messaging AND photo capability

We know that P(A) = .75, P(B) = .8, and $P(A \cap B) = .65$

A cell phone company found that 75% of customers want text messaging on their phone, 80% want photo capability, and 65% want both.

What is the probability that a customer will want at least one of these?

Define the events:

A =costumer wants text messaging

B =costumer wants photo capability

 $A \cap B =$ customer wants text messaging AND photo capability

We know that P(A) = .75, P(B) = .8, and $P(A \cap B) = .65$

We need $P(A \cup B) =$ probability that a customer wants text messaging OR photo capability

A cell phone company found that 75% of customers want text messaging on their phone, 80% want photo capability, and 65% want both.

What is the probability that a customer will want at least one of these?

Define the events:

A =costumer wants text messaging

B =costumer wants photo capability

 $A \cap B =$ customer wants text messaging AND photo capability

We know that P(A) = .75, P(B) = .8, and $P(A \cap B) = .65$

We need $P(A \cup B)$ = probability that a customer wants text messaging OR photo capability

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = .75 + .8 - .65$$

Motor vehicles sold in the US are classified as either cars or light trucks and as either domestic or imported

Last year, 80% of the new vehicles sold to individuals were domestic, 54% were light trucks, and 47% were domestic light trucks

Choose a vehicle sale at random and compute:

 $P(\mathsf{domestic} \cup \mathsf{light} \; \mathsf{truck})$

Motor vehicles sold in the US are classified as either cars or light trucks and as either domestic or imported

Last year, 80% of the new vehicles sold to individuals were domestic, 54% were light trucks, and 47% were domestic light trucks

Choose a vehicle sale at random and compute:

$$P(\mathsf{domestic} \cup \mathsf{light} \; \mathsf{truck})$$

$$P(\text{domestic}) = 0.80$$

 $P(\text{truck}) = 0.54$
 $P(\text{domestic} \cap \text{light truck}) = 0.47$
Then,

$$P(\text{domestic} \cup \text{light truck}) = P(\text{domestic}) + P(\text{light truck})$$

$$- P(\text{domestic} \cap \text{light truck})$$

$$= 0.8 + 0.54 - 0.47$$

Conditional probability

P(domestic) = 0.80

P(truck) = 0.54

 $P(domestic \cap truck) = 0.47$

 $P(domestic \cup truck) = 0.87$

	Domestic	Imported	
Light truck			
Car			

P(domestic) = 0.80 P(truck) = 0.54 $P(\text{domestic} \cap \text{truck}) = 0.47$ $P(\text{domestic} \cup \text{truck}) = 0.87$

	Domestic	Imported	
Light truck			
Car			

	Domestic	Imported	
Light truck	0.47		0.54
Car			
	0.80		

The total row/column can be obtained from the joint probabilities by the addition rule $P(\text{truck}) = P(\text{truck} \cap \text{domestic}) + P(\text{truck} \cap \text{imported})$

	Domestic	Imported	
Light truck	0.47	0.07	0.54
Car	0.33	0.13	0.46
	0.80	0.20	

The probability that a vehicle is a truck is

$$P(\text{truck}) = P(\text{truck} \cap \text{domestic}) + P(\text{truck} \cap \text{imported}) = 0.47 + 0.07 = 0.54$$

Does knowing that the chosen vehicle is imported, change the probability that it is a truck?

	Domestic	Imported	
Light truck	0.47	0.07	0.54
Car	0.33	0.13	0.46
	0.80	0.20	

Conditional probability

The probability that a vehicle is a truck is

$$P(\text{truck}) = P(\text{truck} \cap \text{domestic}) + P(\text{truck} \cap \text{imported}) = 0.47 + 0.07 = 0.54$$

Does knowing that the chosen vehicle is imported, change the probability that it is a truck?

$$P(\text{truck}|\text{imported}) = \frac{P(\text{truck} \cap \text{imported})}{P(\text{imported})} = \frac{0.07}{0.20} = .35$$

The probability of an event can change if we know that some other event has occurred

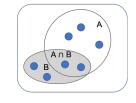
Conditional probability

Let A and B be two events, the **conditional probability** of event A, given that event B has occurred is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
, provided that $P(B) > 0$

Similarly

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$
, provided that $P(A) > 0$



Note that:

if
$$A \subset B$$
, then $P(B|A) = 1$

if A and B are mutually exclusive, then P(A|B) = 0



Multiplication Rule

From the definition of conditional probability follows the **multiplication** rule

Given two events A and B, the probability of their intersection is

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

Namely, for both events to occur, the first one must occur, i.e., P(A), and then, given that the first occurred, the second must occur, i.e. P(B|A).

Statistical Independence

Two events A and B are said to be **statistically independent** if and only if

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

From the multiplication rule, we can also derive

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

and

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

that is, the information about the occurrence of B is of no value in determining P(A) (same for A and B)

Independence

A jar contains 3 red balls and 7 black balls

$$A = \text{draw a red ball}$$

 $P(A) = \frac{3}{10} = 0.3$

B = the second ball drawn is red (replacing the first ball)



$$P(B|A) = P(B|\bar{A}) = \frac{3}{10}$$

A and B are independent

B = the second ball drawn is red (not replacing the first ball)





$$P(B|A) = \frac{2}{9} \text{ and } P(B|\bar{A}) = \frac{3}{9}$$

A and B are not independent

A jar contains 3 red balls and 7 black balls

$$A = \text{draw a red ball}$$

 $P(A) = \frac{3}{10} = 0.3$

B = the second ball drawn is red (replacing the first ball)



$$P(B|A) = P(B|\bar{A}) = \frac{3}{10}$$

A and B are independent

B = the second ball drawn is red (not replacing the first ball)





$$P(B|A) = \frac{2}{9} \text{ and } P(B|\bar{A}) = \frac{3}{9}$$

A and B are not independent

We have to prove that P(B|A) = P(B) in the first scenario and $P(B|A) \neq P(B)$ in the second one

Given the events A and B with 0 < P(B) < 1, then

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(B)P(A|B) + P(\bar{B})P(A|\bar{B})$$

- B and $ar{B}$ are mutually exclusive and collectively exhaustive, $P(B \cup ar{B}) = 1$
- P(B) and $P(\bar{B})$ act as weights in considering the conditional probabilities

Law of total probability

Given the events A and B with 0 < P(B) < 1, then

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(B)P(A|B) + P(\bar{B})P(A|\bar{B})$$

- B and $ar{B}$ are mutually exclusive and collectively exhaustive, $P(B \cup ar{B}) = 1$
- P(B) and $P(\bar{B})$ act as weights in considering the conditional probabilities

In general,

Given the events A and $B_1, B_2, ...$ with $B_i \cap B_j = \emptyset$, $i \neq j$ and collectively exhaustive then

$$P(A) = \sum_{i} P(A \cap Bj) = P(B_j)P(A|B_j)$$

Independence

A jar contains 3 red balls and 7 black balls

$$A = \text{draw a red ball}$$

 $P(A) = \frac{3}{10} = 0.3$

B =the second ball drawn is red (replacing the first ball)



$$P(B|A) = P(B|\bar{A}) = \frac{3}{10}$$

A and B are independent

$$P(B) = P(A)P(B|A) + P(A\bar{)}P(B|\bar{A})$$
$$= 0.3\frac{3}{10} + 0.7\frac{3}{10} = \frac{3}{10}$$

B = the second ball drawn is red (not replacing the first ball)





$$P(B|A) = \frac{2}{9} \text{ and } P(B|\bar{A}) = \frac{3}{9}$$

A and B are not independent

$$P(B) = P(A)P(B|A) + P(\bar{A})P(B|\bar{A})$$
$$= 0.3\frac{2}{9} + 0.7\frac{3}{9} = \frac{3}{10}$$

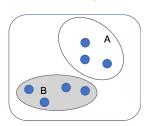
Multiplication rule

If A and B are independent then $P(A \cap B) = P(A)P(B)$, not otherwise

Addition rule

If A and B are mutually exclusive then $P(A \cup B) = P(A) + P(B)$, not otherwise

Mutually exclusive events \neq independent events



$$P(A|B)=0$$

A and B are not independent

Independence cannot be depicted in the Venn diagram because it involves the probabilities of the events rather than the outcomes

Complement rule

$$P(\bar{A})=1-P(A)$$

Addition rule

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

if the events are mutually exclusive $P(A \cup B) = P(A) + P(B)$

Conditional probability

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

Multiplication Rule

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

if the events are independent: P(A|B) = P(A) and $P(A \cap B) = P(B)P(A) = P(A)P(B)$

Law of total probability

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(B)P(A|B) + P(\bar{B})P(A|\bar{B})$$

Independence