2. PROBABILITY THEORY

Study time: 70 minutes

Aim - you will be able to

- characterize probability theory
- explain general notions of probability theory
- explain and use general relations between events
- explain a notion of probability
- define probability by basic axioms
- define properties of probability function
- use a conditional probability
- explain theorem of total probability and Bayes theorem

2.1. General notions

Study time: 20 minutes

Aim

- characterize probability theory
- explain general notions of probability theory

Explication

Probability Theory is the deductive part of statistics. Its purpose is to give a precise mathematical definition or structure to what has been thus far an intuitive notion of randomness. Making randomness more precise will allow us to make exact probability statements. For example when discussing association, we could only make rough statements in terms of tendencies.
Mathematically, probability is a set function. That is, it is a function defined on some domain of sets. Therefore, we begin this discussion my considering the fundamental nature of sets and the basic operations performed on sets, the elements of the domain of our probability function.

- **General notions of the probability theory**

Definition of Set - set A is a collection of elements. Elements are basic intuitive mathematically undefined entities. To define a set, it is necessary to be able to determine whether any element is included or not included in the set. The notion of inclusion is also an intuitive undefined concept.

Definition of Elementary Events - In the case of probability theory, the elements of sets on which probability measures are defined are called elementary events. In practice, these elementary events may be measurement units, cases, or sample points.

Example:

{reverse, obverse} –when tossing the coin

{1,2,3,4,5,6} – when tossing the dice

We denote a set of all results Ω. This set we call **sample space** (of the elementary events). The elementary event $\{\omega\}$ is a subset of the Ω set which contains one element ω from Ω set, $\omega \in \Omega$.
Then the event A will be an arbitrary subset of Ω, $A \subset \Omega$.

From statistics data we can easily determine that share of boys born in particular years with respect to all born children is moving around 51,5%. Despite the fact that in individual cases we can't foretell a sex of a child we can relatively exactly guess how many boys we find among 10 000 born children.

From this example imply that relative frequencies of some events are stabilized with increase repetition number on certain values. We shall call this phenomenon *a stability of the relative frequencies*. This stability of relative frequencies is an empiric basis of the probability theory. **Relative frequency** is number n(A)/n where n is a total number of experiments and n(A) is a number of experiment realizations in which event A became.

Summary of notions

Probability theory is mathematical branch whose logical structure is created axiomatically.
Mathematic statistics is a science which is concerned with questions of data mining data analyzing and results forming.
Random experiment is every finite process whose result is not given in advance by conditions upon whose is runed.
Sample space Ω is a set of all possibly results of the experiment.
Relative frequencies of some events with increase repetition number show certain **stability**.

2.2. Operations with the elementary events

Study time: 20 minutes

Aim

- types of the elementary events
- general relations between events

Explication

What are types of the elementary events?

If the elementary event $\omega \in \Omega$ ($\omega \in A$) came then we can say that an event A came with the experiment realization. We denote this result $\omega \in A$ as **result favourable to the event A**.

Certain event
- is the event which become with every realization of the experiment. It is equivalent with the Ω set.
Certain event is for example: *we toss one of these numbers 1,2,3,4,5,6* (while tossing a dice)

Impossible event
- is the event which can never become in the experiment. We will denote it as $\varnothing$.
Impossible event is for example: *we toss number 8* (while tossing a dice).

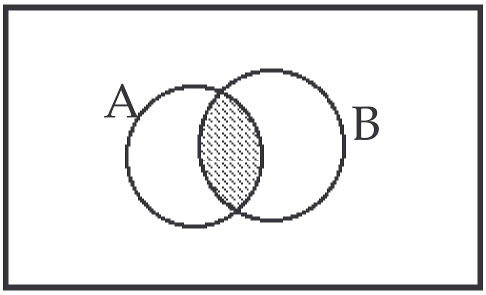
What are relations between events?

Operations on Sets - The operations of union, intersection, complementation (negation), subtraction, the concept of subset, and the null set and universal set or sample space make up the algebra of sets.

Intersection $A \cap B$
The set of all elements that are both in A and in B.

Graphic example:

$$A \bigcap B = \{\omega | \ \omega \in A \wedge \omega \in \mathrm{B}\}$$

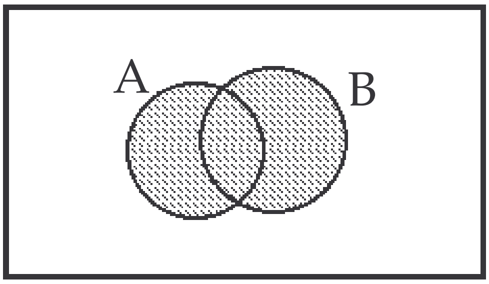


Example – tossing a dice: event A – we toss a number 2 ,3 or 4 and event B – we toss a even number. It is obvious that $A \cap B = \{2,4\}$.

Union $A \cup B$
The set of all elements that are either in A or in B.

Graphic example:

$$A \bigcup B = \{\omega \mid \omega \in A \vee \omega \in B\}$$



Example – tossing a dice: Event $A = \{1,3,4\}$ and event B is when we toss even number. It's obvious that $A \cup B = \{1,2,3,4,6\}$.

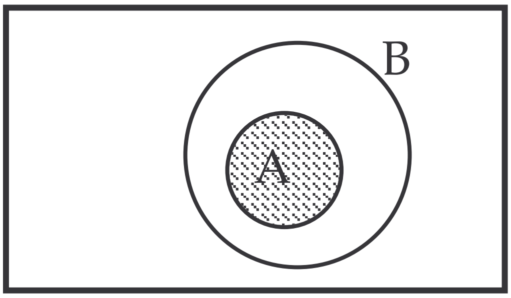
Disjoint events $A \cap B = \varnothing$
Two events A and B can't become together. They have none common result.

Example – tossing a dice: Event A – we toss even number and even B – we toss odd number. These events never have a same result. If event A become than event B can't become.

Subsets (Subevent) $A \subset B$
A is a subset of B if every element of A is also an element of B. It's mean if event A become than event B become too.

Graphic example:

$$A \subset B \Leftrightarrow \{\omega \in \mathrm{A} \Rightarrow \omega \in B\}$$



Example – tossing a dice: Event A – we toss number 2 and event B – we toss even number. The event A is subevent of the event B.

Events A and B are equivalent A = B if $A \subset B$ and at the same time $B \subset A$.

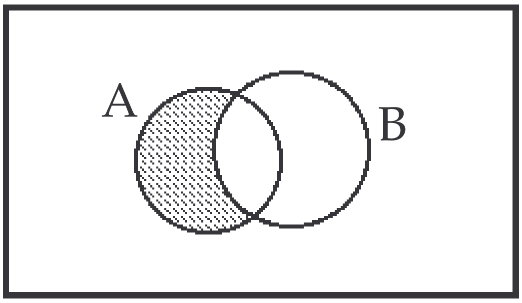
Example – tossing a dice: Event A – we toss even number, event B – we toss number what is divide of number 2. These events are equivalent.

Subtraction A-B
The set of all elements that are in A but not in B

$$A - B = A \cap \overline{B}$$

$$A - B = \{\omega | \omega \in A \wedge \omega \notin B\}$$

Graphic example:

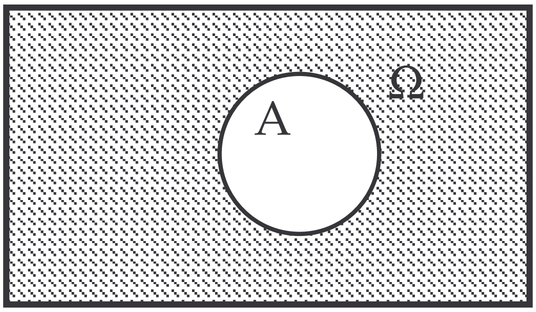


Example – tossing a dice: Event A – we toss a number greater than two and event B – we toss an even number. Subtraction of the events A and B is an event A – B ={3,5}.

Complement of the event A (opposite event)
The set of all elements that are not in A.

$$\overline{A} = \{\omega | \omega \notin A\}$$

Graphic example:



Example – tossing a dice: Event A – we toss an even number, then an event $\overline{A}$ - we toss an odd number.

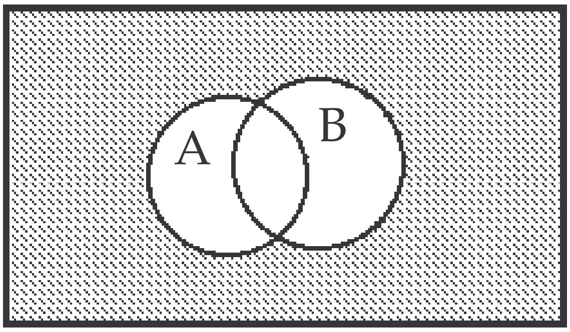
DeMorgan's Laws

- DeMorgan's Laws are logical consequences of the fundamental concepts and basic operations of sets

1. law

The set of all elements that are neither in A nor in B.

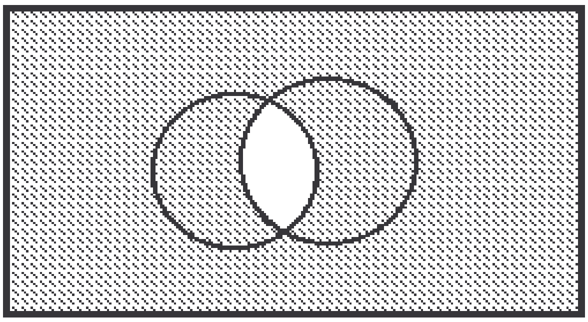
$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$



2. law

The set of all elements that are either not in A or not in B.

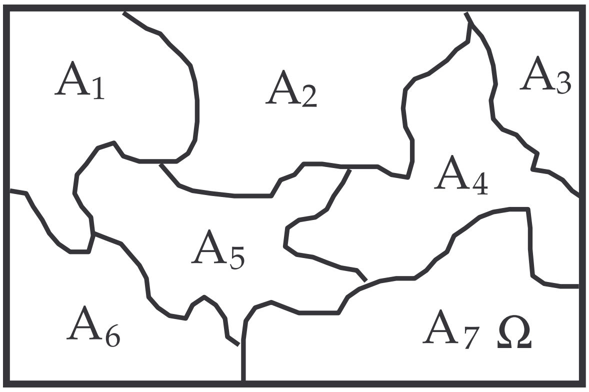
$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$



Mutually disjoint sets and partitioning the sample space

The collection of sets $\{A_1, A_2, A_3, \ldots\}$ partition the sample space Ω: $A_i \bigcap A_j = \varnothing$ for $i \neq j$

$$\Omega = \bigcup_{i=1}^{n} A_i$$



2.3. Probability theory

Study time: 30 minutes

Aim

- notion of probability
- basic theorems and axioms of probability
- types of probability
- conditional probability
- theorem of total probability and Bayes theorem

Explication

Notion of probability

Probability of the event A is a number P(A) which has a property that a relative frequency of the event A with increase realizations number is approaching to the number P(A).

This probability definition is known as *classic probability definition.*

Now we introduce axiomatic probability definition.

Axiomatic probability definition

Probability space is a triad (Ω, S, P) where

(i) Ω is sample space (elements of Ω are elementary events)

(ii) S is a set of subsets of Ω that it holds:

a) $\Omega \in S$;

b) if $A \in S$ then $\overline{A} = \Omega - A \in S$;

c) if $A_1, A_2, A_3, \ldots \in S$ then $\bigcup_{i=1}^{\infty} A_i \in S$

Elements of S we denote as **events**.

(iii) P is function from S to $< 0,1 >$ such that it holds:

a) $P(\Omega) = 1$- probabilities are scaled to lie in the interval [0,1];

b) $P(\overline{A}) = 1 - P(A)$ for every $A \in S$;

c) For a collection of mutually disjoint sets, the probability of their union is equal to the sum of their probabilities.

$$\text{If } A \cap B = \varnothing, \text{ then}$$
$$P\{A \bigcup B\} = P\{A\} + P\{B\}$$

In general,

$$A_i \cap A_j = \varnothing, \ \forall 1 \le i, j \le \infty; i \ne j,$$
$$P\{\bigcup_{i=1}^{\infty} A_i\} = \sum_{i=1}^{\infty} P\{A_i\}$$

Function P is called **probability measure** or shorter **probability**.

Example – tossing a dice:
$\Omega = \{1,2,3,4,5,6,\}$,
S is a set of subsets of Ω (sometimes we denote $S = \exp\ \Omega$) and probability is defined by $P(A)=\dfrac{\text{card}A}{6}$ where card A is number of set A elements.

Basic Theorems of Probability

The following theorems are the logical consequences of the three basic probability axioms we have postulated.

1. For disjoint events A and B hold:
   $A \cap B = \varnothing$ then
   $P\{A \bigcup B\} = P\{A\} + P\{B\}$

2. If for two events A,B hold:
   $B \subset A$ then $P\{B\} \le P\{A\}$
   - note that A is partitioned by B and its complement, and hence P{A} is sum of these two parts

3. For every event A holds: $P\{\overline{A}\} = 1 - P\{A\}$
   - the union of the two sets is the sample space, the intersection is the null sets

4. It holds: $P\{\varnothing\} = 0$

5. It hold: $P\{B - A\} = P\{B\} - P\{B \bigcap A\}$
   - note that B-A and B intersection A are two disjoint sets whose union is B

6. Especially if $A \subset B$ then $P\{B - A\} = P\{B\} - P\{A\}$

7. For arbitrary events A,B hold:
   $P\{A \cup B\} = P\{A\} + P\{B\} - P\{A \cap B\}$

8. Follows from de Morgan's laws

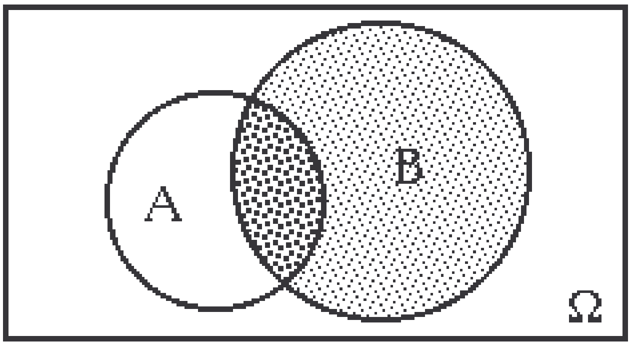
$$
\begin{aligned}
P\{A \cup B\} &= 1 - P\{\overline{A \cup B}\} \\
&= 1 - P\{\overline{A} \cap \overline{B}\}
\end{aligned}
$$

Definition of Conditional Probability

The definition of conditional probability determines how probabilities adjust to changing conditions. When we say that the condition B applies, we mean that the set B is known to have occurred and therefore the rest of the sample space in the complement of B has zero probability. Under these new circumstances, the revised probability of any other event, A, can be determined from the following definition of conditional probability:

$$P\{A|B\} = \frac{P\{A \cap B\}}{P\{B\}}$$

By this formula, the probability of that part of the event A which is in B or intersects with B is revised upwards to reflect the condition that B has occurred and becomes the new probability of A. It is assumed that the probability of B is not zero.



$P\{A|B\}$ - *probability of the event A conditional by the event B*

Conditional Probability Definition of Independence

If the condition that B has occurred does not affect the probability of A, then we say that A is independent of B.

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

From the definition of conditional probability, this implies

$$P\{A\} = \frac{P\{A \cap B\}}{P\{B\}}$$

and hence,

$$P\{A \cap B\} = P\{A\} \cdot P\{B\}$$

It is clear from this demonstration that if A is independent of B, then B is also independent of A.

Example – tossing a dice:
For events A - "we toss 1 in the first toss" and B - "we toss 1 in the second toss" and event C = A $\cap$ B - "we toss 1 in the both tosses" then it holds:

$$P\{C\} = P\{A \cap B\} = P\{A\}.P\{B\} = \frac{1}{6}\cdot\frac{1}{6} = \frac{1}{36}$$

Theorem of Total Probability

If a collection of sets $\{B_1, B_2, B_3, \ldots, B_n\}$ partition the sample space Ω, that is,

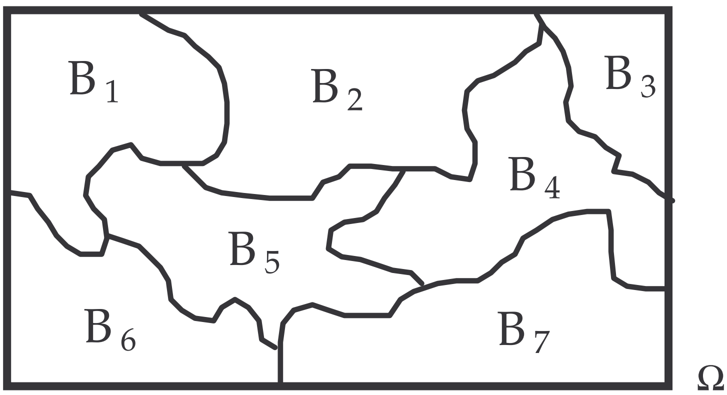
$$B_i \bigcap B_j = \varnothing; \forall\ i \neq j$$

$$\bigcup_{i=1}^{n} B_i = \Omega$$

then for any set A ($P\{A\} \neq 0$) in the sample space Ω,

$$P\{A\} = \sum_{i=1}^{n} P\{A|B_i\} \cdot P\{B_i\}$$

n=7



Proof: Since the collection of sets $\{B_1, B_2, B_3, \ldots, B_n\}$ partitions the sample space Ω,

$$P\{A\} = \sum_{i=1}^{n} P\{A \bigcap B_i\}$$

From the definition of conditional probability

$$P\{A \bigcap B_i\} = P\{A|B_i\} P\{B_i\}$$

Bayes Theorem

If the collection of sets $\{B_1, B_2, B_3, \ldots, B_n\}$ partitions the sample space Ω, then

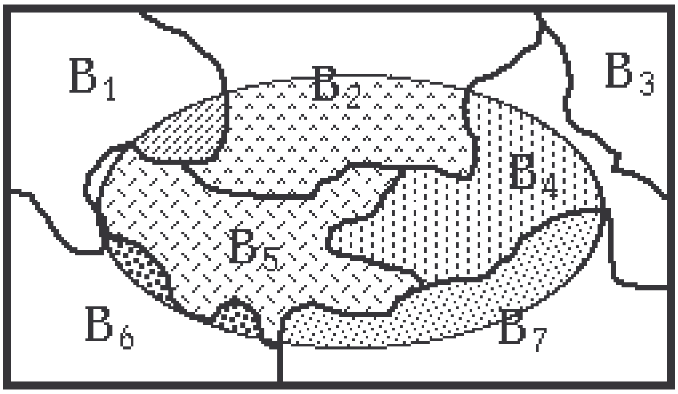
$$P\{B_k|A\} = \frac{P\{A|B_k\}\,P\{B_k\}}{\sum_{i=1}^{n} P\{A|B_i\}\,P\{B_i\}}$$

Proof: From the definition of conditional probability,

$$P\{B_k|A\} = \frac{P\{B_k \cap A\}}{P\{A\}} = \frac{P\{A\}P\{B_k\}}{P\{A\}}$$

Substituting for P{A} from the Theorem of Total Probability, the proof follows.

Graphical representation of Bayes theorem (vyšrafovaná plocha znázorňuje jev A):



PROBABILITY THEORY – SOLVED EXAMPLES

Solved example

Probability of failing of the extinguishing system is 20%. Probability that alarming system fails is 10% and probability that both systems fail is 4%. What is a probability that:
a) at least one system will be working?
b) both systems will be working?

Solution:

We denote: H ... extinguishing system works
S ... alarming system woks

We know that: $P(\overline{H}) = 0{,}20$
$P(\overline{S}) = 0{,}10$
$P(\overline{H} \cap \overline{S}) = 0{,}04$

We must find:

ada) $P(H \cup S)$

We have two possibilities for solving:

By the definition: Events H and S are not the disjoint events and hence:

$$P(H \cup S) = P(H) + P(S) - P(H \cap S),$$

but would be a problem determine a $P(H \cap S)$

By the opposite event: From de Morgan's laws we can write:

$$P(H \cup S) = 1 - P(\overline{H \cup S}) = 1 - P(\overline{H} \cap \overline{S}),$$

$$\underline{\underline{P(H \cup S)}} = 1 - 0{,}04 = \underline{\underline{0{,}96}}$$

The probability (that at least one system will be working) is 96%.

adb) $P(H \cap S)$

We can't solve it by the definition**:**

$$(P(H \cap S) = P(H|S) \cdot P(S) = P(S|H) \cdot P(H)),$$

because we have no information about dependency of the failures of the individual systems. Hence we try to use the opposite event:

$$P(H\cap S)=1-P(\overline{H\cap S})=1-P(\overline{H}\cup\overline{S})=1-\left[P(\overline{H})+P(\overline{S})-P(\overline{H}\cap\overline{S})\right],$$

$$\underline{\underline{P(H\cap S)}}=1-\left[P(\overline{H})+P(\overline{S})-P(\overline{H}\cap\overline{S})\right]=1-[0{,}20+0{,}10-0{,}04]=\underline{\underline{0{,}74}}$$

The probability (that both systems will be working) is 74%.

Solved example

120 students passed mathematics and physics exams. 30 of them failed to pass both exams. 8 failed to pass only math exam and 5 failed to pass only physics exam. What is probability that random student:

a) passed math exam if we know that he failed to pass physics exam
b) passed physics exam if we know that he failed to pass math exam
c) passed math exam if we know that he passed physics exam

Solution:

We denote: M ... he passed mathematics exam
F ... he passed physics exam

We know that:

$$P(\overline{M}\cap\overline{F})=\frac{30}{120}$$
$$P(\overline{M}\cap F)=\frac{8}{120}$$
$$P(M\cap\overline{F})=\frac{5}{120}$$

We must find:

ada) $P(M|\overline{F})$

by the definition of conditional probability:

$$P(M|\overline{F})=\frac{P(M\cap\overline{F})}{P(\overline{F})}=\frac{P(M\cap\overline{F})}{P(M\cap\overline{F})+P(\overline{M}\cap\overline{F})},$$

$$\underline{\underline{P(M|\overline{F})}}=\frac{P(M\cap\overline{F})}{P(M\cap\overline{F})+P(\overline{M}\cap\overline{F})}=\frac{\frac{5}{120}}{\frac{5}{120}+\frac{30}{120}}=\frac{5}{35}=\frac{1}{7}\cong\underline{\underline{0{,}14}}$$

The probability (that he passed math exam if we know that he failed to pass physics exam) is 14%.

adb) $P(F|\overline{M})$

the same way as ada):

$$P(F|\overline{M})=\frac{P(F\cap\overline{M})}{P(\overline{M})}=\frac{P(F\cap\overline{M})}{P(F\cap\overline{M})+P(\overline{F}\cap\overline{M})},$$

$$\underline{\underline{P(F|\overline{M})}}=\frac{P(F\cap\overline{M})}{P(F\cap\overline{M})+P(\overline{F}\cap\overline{M})}=\frac{\frac{8}{120}}{\frac{8}{120}+\frac{30}{120}}=\frac{8}{38}=\frac{4}{19}\cong\underline{\underline{0{,}21}}$$

The probability (that he passed physics exam if we know that he failed to pass math exam) is 21%.

adc) $P(M|F)$

from the definition:

$$P(M|F)=\frac{P(M\cap F)}{P(F)},$$

we have two possibilities:

1)

$$\underline{\underline{P(M|F)}}=\frac{P(M\cap F)}{P(F)}=\frac{1-P(\overline{M\cap F})}{1-P(\overline{F})}=\frac{1-P(\overline{M}\cup\overline{F})}{1-[P(\overline{F}\cap M)+P(\overline{F}\cap\overline{M})]}=\frac{1-[P(\overline{F})+P(\overline{M})-P(\overline{F}\cap\overline{M})]}{1-[P(\overline{F}\cap M)+P(\overline{F}\cap\overline{M})]}=$$

$$=\frac{1-[[P(\overline{F}\cap M)+P(\overline{F}\cap\overline{M})]+[P(F\cap\overline{M})+P(\overline{F}\cap\overline{M})]-P(\overline{F}\cap\overline{M})]}{1-[P(\overline{F}\cap M)+P(\overline{F}\cap\overline{M})]}=$$

$$=\frac{1-[P(\overline{F}\cap M)+P(F\cap\overline{M})+P(\overline{F}\cap\overline{M})]}{1-[P(\overline{F}\cap M)+P(\overline{F}\cap\overline{M})]}=\frac{1-\left[\frac{5}{120}+\frac{8}{120}+\frac{30}{120}\right]}{1-\left[\frac{5}{120}+\frac{30}{120}\right]}=\frac{\frac{77}{120}}{\frac{85}{120}}=\frac{77}{85}\cong\underline{\underline{0{,}91}}$$

2)
We write given data into the table:

	They passed math exam	They failed to pass math exam	Total
They passed physics exam		8	
They failed to pass physics exam	5	30	35
Total		38	**120**

and we calculate remaining data:

How much students passed physics exam? It is total number(120) minus number of students who failed to pass physics exam (35) and that is 85. Analogously for number of students who passed math exam: 120 – 38 = 82. And for number of students who passed both exams: 82 – 5 = 77.

	They passed math exam	**They failed to pass math exam**	**Total**
They passed physics exam	77	8	85
They failed to pass physics exam	5	30	35
Total	82	38	**120**

Finding probabilities are:

$$P(M \cap F) = \frac{77}{120}; \quad P(F) = \frac{85}{120},$$

from that imply:

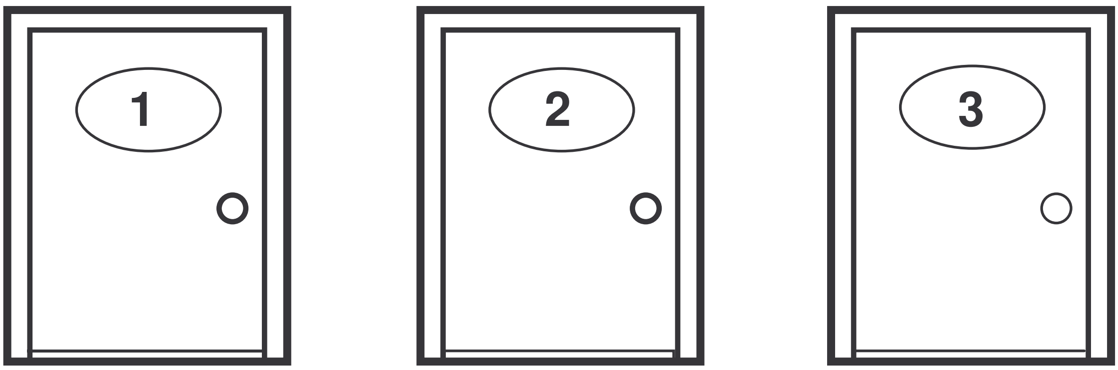
$$\underline{\underline{P(M|F)}} = \frac{P(M \cap F)}{P(F)} = \frac{\frac{77}{120}}{\frac{85}{120}} = \frac{77}{85} \cong \underline{\underline{0{,}91}}$$

The probability (that he passed math exam if we know that he passed physics exam) is 91%.

Solved example (Example of an Application of Bayes's Theorem)

In a famous television game show, the winner of the preliminary round is given the opportunity to enhance his winnings. The contestant is presented with three closed doors and told that behind one of the doors is a new automobile while behind the other two doors are goats. If the contestant correctly selects the door which conceals the automobile, he will win the automobile.

The game show host asks the contestant to make a preliminary selection, after which the host opens one of the other two doors to reveal a goat. The contestant is then given the option of switching his choice to the other door which remains closed. Should he change his choice?



Solution:

The sample space consists of three possible arrangements {AGG, GAG, GGA}.

Assume that each of the three arrangements have the following probabilities:

$p_1 = P\{AGG\}$ $p_2 = P\{GAG\}$ $p_3 = P\{GGA\}$

where $p_1 + p_2 + p_3 = 1$.

Assume without loss of generality that the contestant's preliminary choice is Door #1 and the host opens Door #3 to reveal a goat. One the basis of this information we must revise our probability assessments. It is clear that the host cannot open Door #3 if it conceals the automobile.

$$P\{\text{Door \#3} \mid \text{GGA}\} = 0$$

Also, the host must open Door #3 if Door #2 conceals the automobile since he cannot open Door #1, the contestant's choice.

$$P\{\text{Door \#3} \mid \text{GAG}\} = 1$$

Finally if the automobile is behind the contestant's first choice, Door #1, the host can choose to open either Door #2 or Door #3. Suppose he chooses to open Door #3 with some probability q.

$$P\{\text{Door \#3} \mid \text{AGG}\} = q$$

Then by Bayes' Theorem, we can compute the revised probability that the automobile is behind Door #2 as

$$P\{\text{GAG} \mid \text{Door \#3}\} = \frac{P\{\text{Door \#3} \mid \text{GAG}\} \cdot P\{\text{GAG}\}}{P\{\text{Door \#3}\}}$$

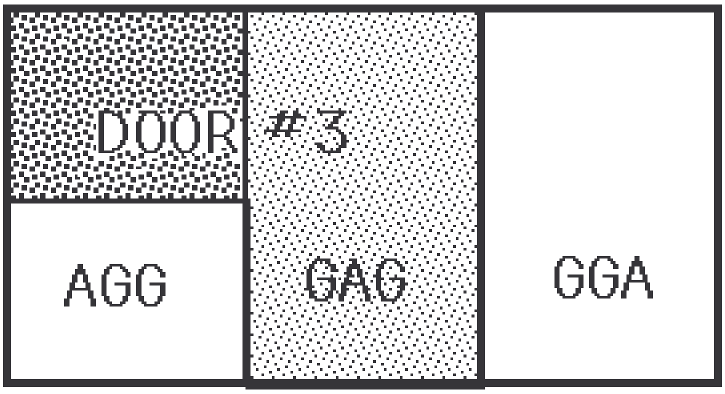
Substituting known values into this equation we obtain,

$$P\{\text{GAG} \mid \text{Door \#3}\} = \frac{1 \times p_2}{(q \times p_1) + (1 \times p_2) + (0 \times p_3)} = \frac{p_2}{qp_1 + p_2}$$

Thus the probability that the automobile is behind Door #2 after the host has opened Door #3 is greater than one half if,

$$qp_1 < p_2 \cdot$$

In this case, the contestant should change his choice. In the normal case where the original probabilities of the three arrangements, p_i, are equal and the host chooses randomly between Door #2 and Door #3, the revised probability of Door #2 concealing the automobile will be greater than one half. Therefore, unless the contestant has a strong a priori belief that Door #1 conceals the automobile, and/or believes that the host will prefer to open Door #3 before Door #2, he should switch his choice.



As the above diagram illustrates, if the original probabilities of all three arrangements are equal and the host chooses randomly which door to open, then of the one half of the sample space covered by opening Door #3, two thirds falls in the region occupied by arrangement GAG. Therefore, if the host opens Door #3, Door #2 becomes twice as likely as Door #1 to conceal the automobile.

Σ Summary of notions

Random experiment is every finite process whose result is not determined in advance by conditions upon whose it runs and which is at least theoretically infinitely repeatable.
Possible results of random experiment are called **elementary events.**
A set of all elementary events we call **a sample space**.
Probability measure is real function defined upon subset system of the sample space which is non-negative normed and σ-aditive.
Conditional probability is a probability of event with conditional that some other (not impossible) event happened.
A and B events are **independent** if intersection probability of these two events is equal to a product of individual event probabilities.
Total probability theorem gives us a way how to determine probability of some event A while presuming that complete set of mutual disjoint events is given.
Bayes's theorem allows us to determine conditional probabilities of individual events in this complete set while presuming that A event happened.

Questions

1. How we determine probability of two events union?
2. How we determine probability of two events intersection?
3. When are two events independent?

Problems

Example 1: Suppose that a man and a woman each have a pack of 52 playing cards. Each draws a card from his/her pack. Find the probability that they each draw the ace of clubs.

{*Answer:* independent events - 0.00037}

Example 2: A glass jar contains 6 red, 5 green, 8 blue and 3 yellow marbles. If a single marble is chosen at random from the jar, what is the probability of choosing a red marble? a green marble? a blue marble? a yellow marble?

{*Answer:* P(red)=3/11, P(green)=5/22, P(blue)=4/11, P(yellow)=3/22}

Example 3: Suppose there are two bowls full of cookies. Bowl #1 has 10 chocolate chip cookies and 30 plain cookies, while bowl #2 has 20 of each. Fred picks a bowl at random, and then picks a cookie at random. We may assume there is no reason to believe Fred treats one bowl differently from another, likewise for the cookies. The cookie turns out to be a plain one. How probable is it that Fred picked it out of bowl #1?

{*Answer*: Conditional probability - 0.6}

Example 4: Suppose a certain drug test is 99% accurate, that is, the test will correctly identify a drug user as testing positive 99% of the time, and will correctly identify a non-user as testing negative 99% of the time. This would seem to be a relatively accurate test, but Bayes's theorem will reveal a potential flaw. Let's assume a corporation decides to test its employees for opium use, and 0.5% of the employees use the drug. We want to know the probability that, given a positive drug test, an employee is actually a drug user.

{*Answer:* Bayes's theorem - 0.3322}