Contents

Unit 1. BASICS OF PROBABILITIES

Probability is a mathematical language for quantifying uncertainty. Probability theory can be applied to a variety of problems, from coin tossing to analyzing computer algorithms. The starting point is the definition of the space of elementary events, as well as the set of elementary outcomes — the set of possible results, which obtaining leads to the occurrence of event *A*.

The space of elementary events Ω is a set of possible outcomes of the experiment. The values of ω in Ω are called sample outcomes or realizations. Events ω are subsets of Ω.

Example 1. Find the sample space for the sex of 3 children in a family, if we are interested in knowing specifically the sex of the 1st, 2nd and 3rd children (i.e. order matters). We are going to use a tree diagram (Fig. 1). A tree diagram is a schematic with branches emanating from a starting point showing all possible outcomes of a probability experiment.

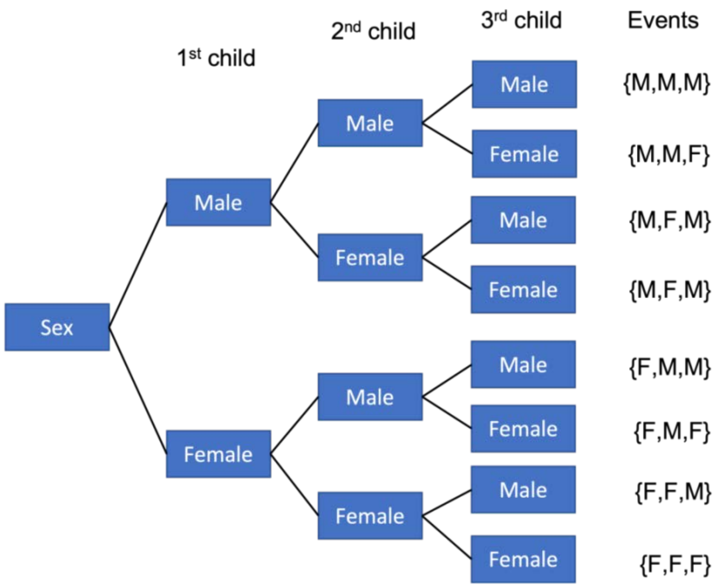

Fig. 1. The tree diagram containing all possible permutations

A *probability experiment* is a chance process that leads to well-defined results, called outcomes. An *outcome* is the result of a *single trial* of a probability experiment. An *event* is some specified outcomes that may or may not occur when a probability experiment is performed (can have multiple outcomes — boy or girl). We can define possible events using a *sample space*. A sample space lists the set of all possible outcomes of a probability experiment. *Probability* is the chance of a *particular* event occurring and is the basis of inferential statistics.

Example 2. If we toss a coin twice, we get the following outcomes ω = {*HT*, *TT*, *HH*, *TH*}. The event when we get the head on the first toss equals the fallowing set of events *A* = {*TH*, *HH*}.

Let's model this situation in Python (Fig. 2).

```
from random import sample
Omega=[1,2] #пусть 1 - Орел, а 2 - решка
k=0
for i in range(100):
    a=sample(Omega,1)
    b=sample(Omega,1)
    if a==[1]:
        k+=1
print(k)
```

```
46
```

Fig. 2. A coin toss modelling to calculate the number of the "Head" outcomes

After carrying out 100 trials, we get 46 outcomes, in which, by tossing a coin twice, we get the Head on the first toss of two.

Example 3. Suppose we need to measure the temperature of some object. All event space equals $\Omega = R = (-\infty; \infty)$.

Let the event correspond to the temperature of the object greater than 10 degrees, but less or equal to 23. So $A = (10, 23]$. This shows that the event may have a non-discrete value.

Example 4. Let E be the event in which the Head falls on the third toss of the coin.

$$E = \{(\omega_1, \omega_2, \omega_3, \ldots), \omega_1 = T, \omega_2 = T, \omega_3 = H, \omega_i = \{T, H\}, \text{for } i > 3\}.$$

The complement of A is denoted as $A^C = \{\omega \in \Omega, \omega \notin A\}$.

It is clear that the complement Ω is equal to the {}-empty set.

We also need the union of the events $A \cup B = \{\omega \in \Omega, \omega \in A$ or $\omega \in B$ or $\omega \in$ both sets}.

The intersection of events $A \cap B = \{\omega \in \Omega; \omega \in A$ and $\omega \in B\}$. The intersection of events can be denoted by $A \cap B$ or AB.

If $A_i \cap A_j = \emptyset$ provided that $i<>j$, then these sets are called disjoint.

For example, $A_1 = [1, 2)$, $A_2 = [2, 3)$, $A_3 = [3, 5)\ldots$ $A_1 \cap A_2 \cap A_3 = \emptyset$, which indicates non-overlapping events, because the intersection is equal to an empty set. The sequence of events (subsets) A_1, A_2, A_3 is called monotonically increasing if $A_1 \subset A_2 \subset A_3 \ldots$ (Fig. 3, *a*), which we can define as $\lim_{n\to\infty} A_n = \cup^{\infty}_{i=1} A_i$ and monotonically decreasing, if $A_1 \supset A_2 \supset A_3 \ldots$ (Fig. 3, *b*), for $\lim_{n\to\infty} A_n = \cap^{\infty}_{i=1} A_i$.

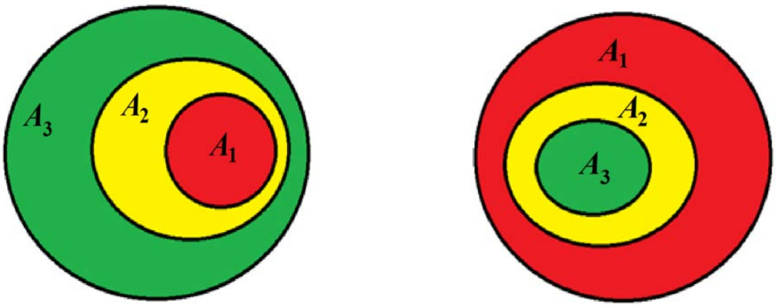

Fig. 3. *a* — $A_1 \subset A_2 \subset A_3$ — (monotonically increasing sets);
b — $A_1 \supset A_2 \supset A_3$ (monotonically decreasing sets)

Each event A can be assigned some real number $P(A)$. For example, how many times a certain number occurs in the sample (Fig. 4).

```
import seaborn as sns
import matplotlib.pyplot as plt
a=[]
for i in range(1,7):
    for j in range(1,7):
        a.append(i+j)
sns.histplot(a,bins=11)
plt.show()
```

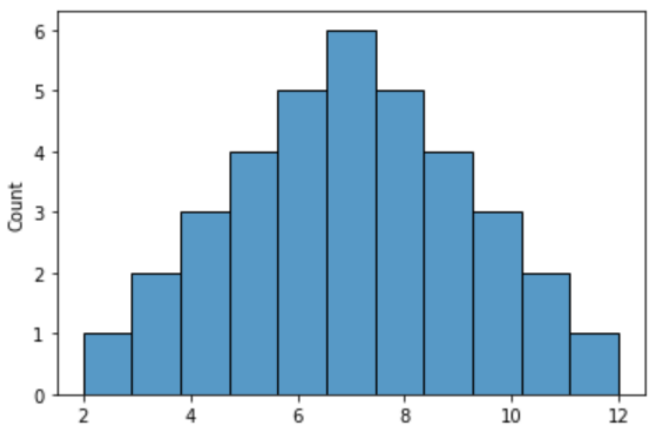

Fig. 4. Dice rolling modelling in Python

Suppose we roll two dice. The above histogram shows the distribution of dice roll outcomes in an ideal scenario, i.e. with the number of experiments tending to infinity. We see that the sum of face values equal to seven occurs as many as 6 times, and 2 and 12 — once each. Accordingly, you can set a distribution series, and it can be perfectly built in Python using a dictionary. {2: 1, 3: 2, 4: 3, 5: 4, 6: 5, 7: 6, 8: 5, 9: 4, 10: 3, 11: 2, 12: 1}.

When we roll a die twice, 36 equally likely outcomes are possible. Note that here the order of the outcome matters. When determining probability that the sum is 11 (denoted as: $P(\text{sum} = 11)$), this can be arrived at two ways, 5 + 6 and 6 + 5. $P(\text{sum} = 11)$ = (ways in which 11 can happen) / N = = 2 / 36 = 0,056 = 5,6 %.

If each random variable is associated with the probability of a given sum of faces: 2 corresponds to 2,78 %; 3 to 5,56 %; 4 to 8,33 %; 5 to 11,11 %; 6 to 13,89 %; 7 to 16,67 %; 8 to 13,89 %; 9 to 11,11 %; 10 to 8,33 %; 11 to 5,56 %; 12 to 2,78 %.

The sum of all random values equals 100 %.

Let's test this by simulating a toss of two coins (Fig. 5) with the sample function and plotting histograms of 100, 1000, 10 000 outcomes.

```
from random import sample
Omega = [1,2,3,4,5,6]
alist=[]
b=[100,1000,10000]
for j in b:
    alist=[]
    for i in range(j):
        a=sample(Omega,1)
        b=sample(Omega,1)
        alist.append(a[0]+b[0])
    sns.histplot(alist,bins=11)
    plt.show()
```

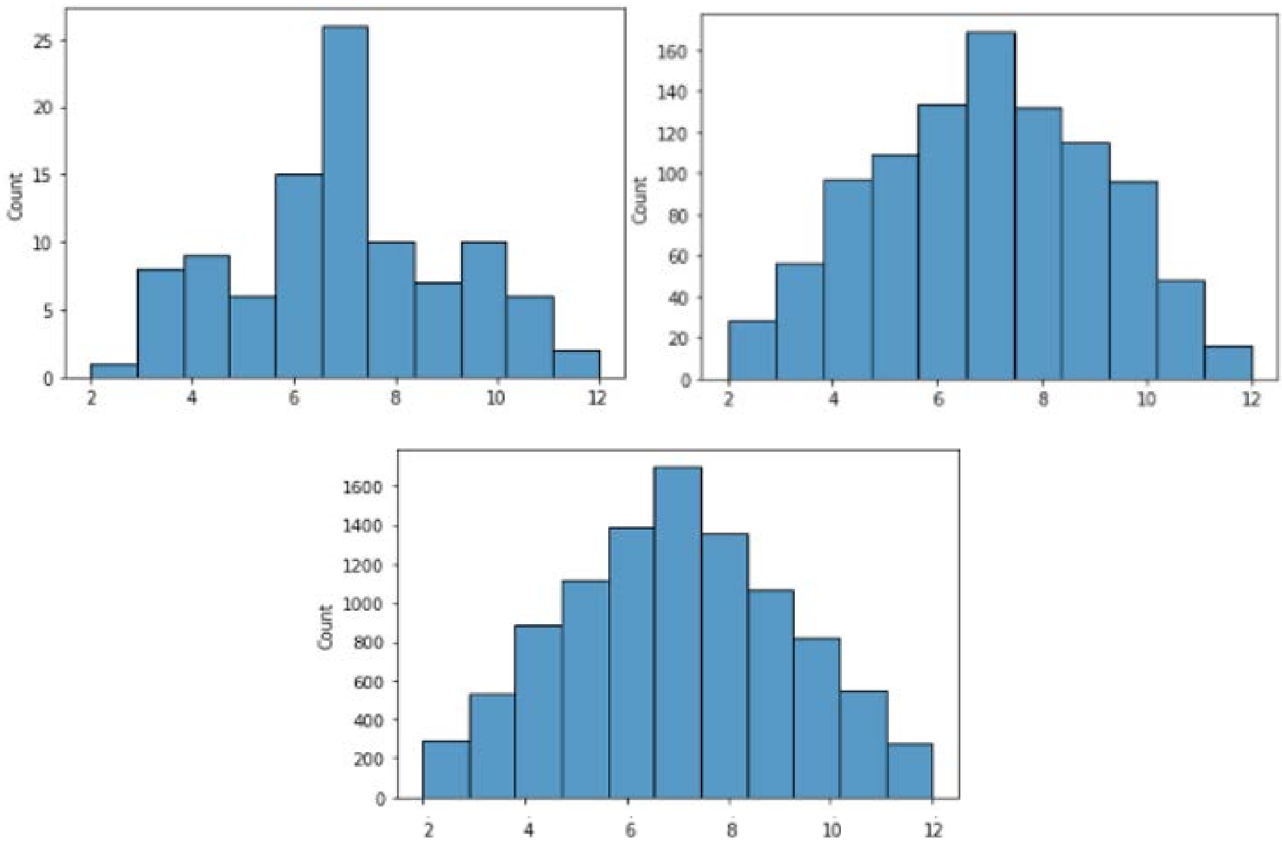

Fig. 5. The dice rolling simulation confirms that more experiments are conducted more their results are consistent with the theoretical ones

As we can see, the more experiments, the more their outcomes approach the distribution law of a discrete random variable, i.e. distribution series for a given set of elementary events.

Now let's imagine that one of the dice is made so that it doesn't roll a 6 (Fig. 6) when rolled. Now imagine that the first dice has only odd faces, and the second one has only even ones.

We can determine whether it is a fraudulent dice by comparing the given law of distribution of a random variable and the distribution of outcomes of elementary events resulting from experiments. In our examples it is clear from the gaps that the dice is/are not correct.

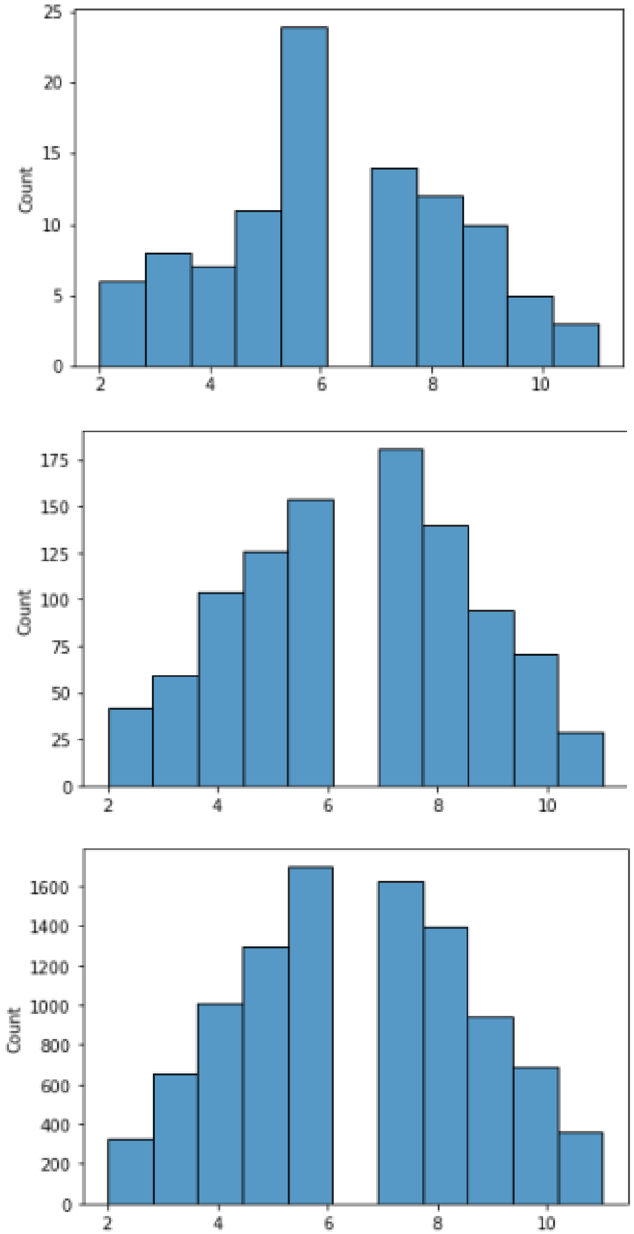

Fig. 6. Simulation of a dice rolling fraudulent distribution

Properties of probabilities

Property 1: the probability of an event is always between 0 and 1, inclusive.

Property 2: the probability of an event that cannot occur is 0. An event that cannot occur is called an *impossible event*.

Property 3: the probability of an event that must occur is 1. An event that must occur is called a *certain event*.

Property 4: the sum of probabilities of all possible outcomes in the sample space is 1.

How do we count?

This might seem like a silly subtitle, but is important to cover formally when we think about probability. We utilize counting rules to quantify the number of ways an event (outcome) can happen. We can define it using the equation:

$$\text{Probability} = \frac{\text{\#of outcomes of interest in an event}}{\text{total \# of all possible outcomes in the sample space}}.$$

Counting rules

The *fundamental counting rule* is to find the total number of outcomes in a sequence of events, multiply the number of outcomes from each event.

Example 1. Two exercise programs and three diet plans for subjects with diabetes: exercise programs = 2 outcomes; diet plans = 3 outcomes. Total number of possible strategies = 2 · 3 = 6.

The fundamental counting rule applies when repetitions are permitted, i.e. the number of outcomes per event does not change.

Example 2. The numbers 1–9 are to be used in a 6 digit student ID card. How many unique cards are possible if repetitions of the same individual number are permitted?

$$9 \cdot 9 \cdot 9 \cdot 9 \cdot 9 \cdot 9 = 9^6 = 531\ 441,$$

where each “9” is the number of options (digits) we have to pick from.

If repetitions are not allowed, then the number of outcomes is reduced by 1 per event. Thus, if we redid the example above, but repetition was not allowed: 9 · 8 · 7 · 6 · 5 · 4 = 6048.

The distribution law of a continuous random variable is conveniently specified using the so-called probability density function $f(x)$ (Fig. 7). The probability $P(a < X < b)$ that the value accepted by the random variable X falls into the interval $(a{:}b)$ is determined by the equality:

$$P(a < X < b) = \int_a^b f(x)dx.$$

The graph of the function $f(x)$ is called the distribution curve.

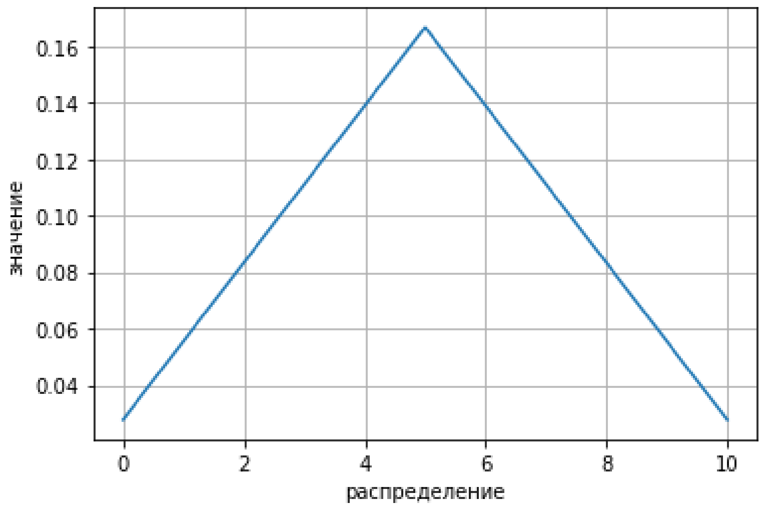

Fig. 7. In this example we observe the probability density function (PDF) of a double dice roll

In this example, we observe the probability density function (PDF) of the roll of two dice (Fig. 8).

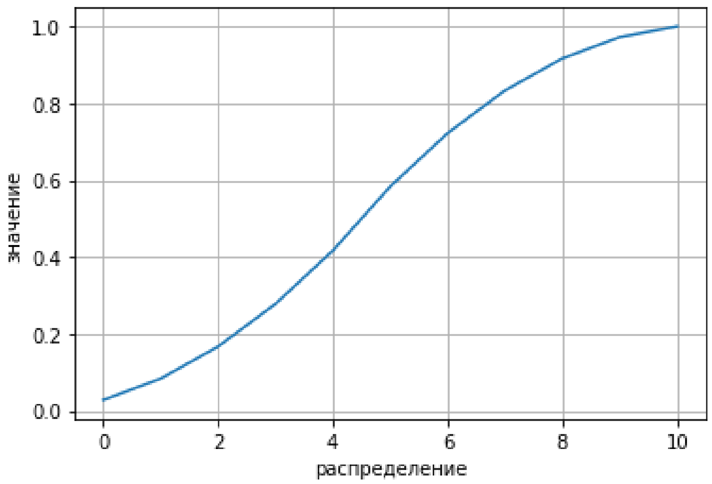

Fig. 8. Dice rolling cumulative distribution function (CDF)

This is the cumulative distribution function (CDF) for our example of rolling two dice. And here is the code for both functions (Fig. 9).

```
y=Z1[:,1]
y=y/100
plt.plot(y)
plt.grid()
plt.xlabel('распределение')
plt.ylabel('значение')
plt.show()
print(np.sum(y))
alist=[]
for i in range(len(y)):
    alist.append(np.sum(y[:i+1]))
plt.plot(alist)
plt.grid()
plt.xlabel('распределение')
plt.ylabel('значение')
plt.show()
```

Fig. 9. The PDF’s and CDF’s listings

If we consider elementary events from the point of view of sets, then we need to specify the following properties:

1. Events A_1 and A_2 are called incompatible or non-overlapping if the occurrence of one of these events excludes the possibility of the other, in other words, A_1 and A_2 cannot occur simultaneously.

2. The union or sum of events A_1 and A_2 is the event A, which means the implementation of at least one of the events A_1, A_2: $A = A_1 \cup A_2$, where ∪ is a special union symbol. The union of events A_1, A_2, ... is defined similarly, denoted as $A = \cup_k A_k$.

3. The intersection or product of events A_1 and A_2 is the event A, which means the implementation of both the event A_1 and the event A_2: $A = A_1 \cap A_2$, where ∩ is a special intersection symbol. The product of events A_1A_2 ... is defined similarly, denoted as $A = \cap_k A_k$.

4. The difference between events A_1 и A_2 is the event A, which means that event A_1 occurs, but event A_2 does not occur: $A_2 : A = A_1/A_2$.

5. The event $\bar{A}$ is called complementary to the event A which means that the event A does not occur: $\bar{A} = \Omega / A$.

6. A function P that assigns a random variable $P(A)$ to each elementary event A is a probability distribution or a probability measure if it satisfies the 3 axioms:

1) $P(A) \geq 0$ for every A;

2) $P(\Omega) = 1$;

3) If $A_1 \ldots A_n$ are not joint events, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

From these axioms, we can learn many properties of P:

1) $P(\emptyset) = 0$;

2) $A \subset B \to P(A) \leq P(B)$;

3) $0 \leq P(A) \leq 1$;

4) $P(A^c) = 1 - P(A)$;

5) $A \cap B = \emptyset \to P(A \cup B) = P(A) + P(B)$.

Statement: for any A and B, $P(A \cup B) = P(A) + P(B) - P(AB)$.

Proof (Fig. 10):

$$P(A \cup B) = P((AB^C) \cup P(AB) \cup P(A^CB)) = P(AB^C) + P(AB) + P(A^CB) =$$

$$= P(AB^C) + P(AB) + P(A^CB) + P(AB) - P(AB) = P((AB^C) \cup (AB)) + P((A^CB) \cup (AB)) - P(AB) =$$

$$= P(A) + P(B) - P(AB).$$

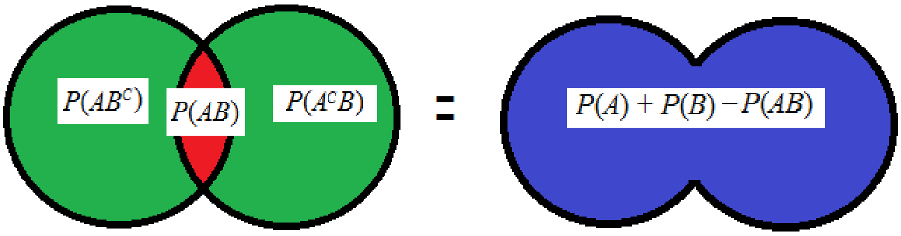

Fig. 10. Calculating the probability of the union of two sets

Example 1. We toss two coins. Let H_1 be the event when Head is the outcome on the first roll and let H_2 be the event when heads come up on the second roll. With equal outcomes, we get:

$$P(\{H_1, H_2\}) = P(\{H_1, T_2\}) = P(\{T_1, H_2\}) = P(\{T_1, T_2\}) = 1/4,$$

then:

$$P(H_1) + P(H_2) - P(H_1 H_2) = 1/2 + 1/2 - 1/4 = 3/4.$$

Example 2. The patient came to the doctor with a slightly elevated temperature and sore throat. According to the doctor, the patient may have a bacterial or viral infection, or even both infections:

A — the patient has a bacterial infection;

B — the patient has a viral infection.

Let: $P(A) = 0{,}7$; $P(B) = 0{,}4$.

What is the probability that the patient has both infections?

Hint $P(A) + P(B) > 1$, when can this situation happen?

Solution:

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

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

We know that the patient has either a virus or bacteria:

$$P(A \cup B) = 1;$$

$$P(AB) = 0{,}7 + 0{,}4 - 1 = 0{,}1.$$

Example 3. What if you were asked the probability of getting a club or a heart when drawing a single card from the deck, what would you say?

P(club or heart) = *P*(club) + *P*(heart) – *P*(club and heart) = 13/52 + 13/52 – 0 = 2/4 = 0,5 = 50 %. We get a zero for the intersection *P*(club and heart). What happens if you have 3 events that are not mutually exclusive? You can use the Venn diagram to help answer this question (Fig. 11).

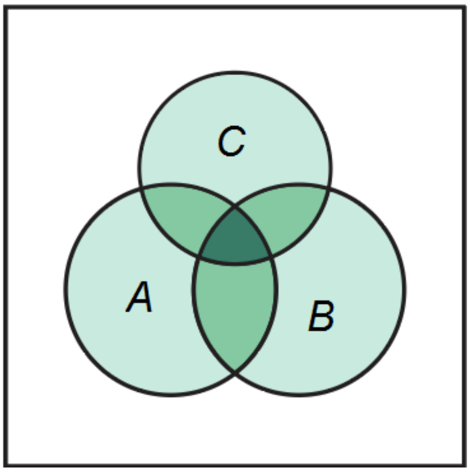

Fig. 11. The intersection of three sets: *A*, *B* and *C*

$$P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C) – P(A \text{ and } B) – P(B \text{ and } C) − P(A \text{ and } C) – P(A \text{ and } B \text{ and } C).$$

Example 4. At a particular school with 200 female students, 58 play volleyball, 40 play basketball, and 8 play both. What is the probability that a randomly selected female student plays neither sport?

V — volleyball; *B* — basketball.

As certain what the probabilities we’ve been provided with. Then, to find the complement:

$$P(V) = \frac{58}{200};$$

$$P(B) = \frac{40}{200};$$

$$P(V \text{ and } B) = \frac{8}{200};$$

$$P((V \text{ or } B)^c) = 1 - P(F \text{ or } B);$$

$$P(V \text{ or } B) = P(F) + P(B) - P(F \text{ and } B) = \frac{58}{200} + \frac{40}{200} - \frac{8}{200} = 0{,}45;$$

$$P((V \text{ or } B)^c) = 1 - 0{,}45 = 0{,}55 = 55\ \%.$$

The theorem of addition of probabilities. The probability of the sum of two incompatible events is equal to the sum of the probabilities of these events:

$$P(A + B) = P(A) + P(B).$$

This theorem is generalized to the case of an arbitrary number of pairwise incompatible events:

$$P(\sum_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i).$$

Probability multiplication theorem. The probability of the product of two events is equal to the product of the probability of one of them by the conditional probability of the other, calculated under the condition that the first takes place:

$$P(AB) = P(A)\ P(B / A) \text{ or } P(AB) = P(B)\ P(A / B).$$

The probability of the product of several events is equal to the product of the probabilities of these events, and the probability of each next event in order is calculated under the condition that all previous ones took place:

$$P(A_1 A_2 \ldots\ A_k) = P(\prod_{i=1}^{k} A_i) = P(A_1)P\left(\frac{A_2}{A_1}\right)P\left(\frac{A_3}{(A_1 A_2 \ldots A_{(k-1)})}\right).$$

In the case of independent events, the formula is valid:

$$P(\prod_{i=1}^{k} A_i) = \prod_{i=1}^{k} P(A_i).$$

Example 1. An urn contains 10 white, 15 black, 20 blue and 25 red balls.

One ball is taken out. Find the probability that the drawn ball is: white; black; blue; red; white or black; blue or red.

We have:

$$n = 10 + 15 + 20 + 25 = 70;$$

$$P(W) = 10 / 70 = 1 / 7\ P(B) = 15 / 70 = 3 / 14;$$

$$P(\text{Blue}) = 20 / 70 = 2 / 7;$$

$$P(R) = 25 / 70 = 5 / 14.$$

Applying the probability addition theorem, we get:

$$P(W + B) = P(W) + P(B) = 1 / 7 + 3 / 14 = 5 / 14;$$

$$P(\text{Blue} + R) = P(\text{Blue}) + P(R) = 2 / 7 + 5 / 14 = 9 / 14;$$

$$P(W + B + \text{Blue}) = 1 - P(R) = 1 - 5 / 14 = 9 / 14.$$

The first box contains 2 white and 10 black balls; the second box contains 8 white and 4 black balls. A ball was taken from each box. What is the probability that both balls are white?

In this case, we are talking about the combination of events *A* and *B*, where event *A* is the appearance of a white ball from the first box, event *B* is the appearance of a white ball from the second box. Moreover, *A* and *B* are independent events. We have:

$$P(A) = 2 / 12 = 1 / 6;$$

$$P(B) = 8 / 12 = 2 / 3.$$

Applying the probability multiplication theorem, we find:

$$P(AB) = P(A)\ P(B) = 1 / 6 \cdot 2 / 3 = 1 / 9.$$

Example 2. Three shooters independently shoot at the target. The probability of hitting the target for the first shooter is 0,75, for the second — 0,8, for the third — 0,9. Determine the probability that all three shooters hit the target at the same time.

We have:

– $P(A) = 0{,}75$;

– $P(B) = 0{,}8$;

– $P(C) = 0{,}9$,

then

$$P(ABC) = P(A)\ P(B)\ P(C) = 0{,}75 \cdot 0{,}8 \cdot 0{,}9 = 0{,}54.$$

Example 3. In the conditions of the previous problem, determine the probability that at least one shooter hits the target.

Here:

– $P(A^c) = 1 - 0{,}75 = 0{,}25$ (probability for the first shooter to miss);

– $P(B^c) = 1 - 0{,}8 = 0{,}2$ (probability for the second shooter to miss);

– $P(C^c) = 1 - 0{,}9 = 0{,}1$ (probability for the third shooter to miss),

then $P(A^cB^cC^c)$ is the probability of simultaneous miss of all three shooters which is defined as follows:

$$P(A^cB^cC^c) = P\ (A^c\ )\ P(B^c\ )\ P(C^c) = 0{,}25 \cdot 0{,}2 \cdot 0{,}1 = 0{,}005.$$

But the event opposite to the event $A^cB^cC^c$ consists in hitting the target by at least one shooter. Therefore, the desired probability $P = 1 - P(A^cB^cC^c)$ i.e. $P = 1 - 0{,}005 = 0{,}995$.