§1. Probability Theory Topics

1. Definitions of Probability

Classical definition: Probability is the ratio of the number of favorable outcomes to the total number of possible outcomes in a random experiment, assuming all outcomes are equally likely. Consider the finite or countable set (the set of all possible outcomes) called the sample space denoted by Ω.

$$\Omega = (\omega_1, ..., \omega_n),\ A \subset \Omega \Rightarrow P(A) = \frac{|A|}{|\Omega|},$$

where $|..|$ denote the power of sets (number of elements).
Example. Let's say we have the following fruits: 6 apples, 5 oranges, and 3 kiwis, and we choose 2 fruits from them. What is the probability that among the selected fruits:
A) There will be 2 apples;
B) There will be no kiwis;
C) Both fruits will be different.
Solution. The total number of ways to choose 2 fruits from 14 is:

$$|\Omega| = C(14, 2) = \frac{14 \times 13}{2} = 91$$

Now, let's solve each part:
a) Probability of selecting 2 apples:
The number of ways to choose 2 apples from 6 is:

$$|A| = C(6, 2) = \frac{6 \times 5}{2} = 15$$

Therefore, the probability of selecting 2 apples is:

$$P(A) = P(2 \text{ apples}) = \frac{C(6, 2)}{C(14, 2)} = \frac{15}{91}$$

b) Probability of selecting no kiwis:
If no kiwis are selected, we are choosing from apples and oranges, i.e., from $6 + 5 = 11$ fruits. The number of ways to choose 2 fruits from 11 is:

$$|B| = C(11, 2) = \frac{11 \times 10}{2} = 55$$

Thus, the probability of selecting no kiwis is:

$$P(B) = P(\text{no kiwis}) = \frac{C(11, 2)}{C(14, 2)} = \frac{55}{91}$$

c) Probability that both fruits will be different:
To select two different fruits, we can choose one pair from apples, oranges, and kiwis. The number of such combinations is:
1 apple and 1 orange: $C(6, 1) \times C(5, 1) = 6 \times 5 = 30$
1 apple and 1 kiwi: $C(6, 1) \times C(3, 1) = 6 \times 3 = 18$
1 orange and 1 kiwi: $C(5, 1) \times C(3, 1) = 5 \times 3 = 15$

Thus, the total number of ways to choose two different fruits is:

$$|C| = 30 + 18 + 15 = 63$$

So, the probability that both fruits will be different is:

$$P(C) = P(\text{different fruits}) = \frac{63}{91}$$

Now we remember other ways to calculate the power of a set.

2. Elements of Combinatorics

Combinatorics is the branch of mathematics that studies the counting, arrangement, and combination of objects. In probability theory, combinatorics is used to count the number of possible outcomes in random experiments.

- **Permutations (transpositions):** The number of ways to arrange n objects. The number of permutations of n objects is $n!$.

$$P_n = n! = 1 \cdot 2 \cdot 3 \cdot ... \cdot n$$

Example: The number of ways to arrange 3 books on a shelf is $3! = 6$.

- **Arrangements (or placements):** The number of ways to select k objects from n, where the order matters. The number of placements is given by:

$$A_n^k = A(n,k) = \frac{n!}{(n-k)!} = n \cdot (n-1) \cdot (n-2) \cdot ... \cdot (n-k+1)$$

Example: The number of ways to place 3 people in 3 seats is $A(3,3) = 6$.

- **Combinations:** The number of ways to select k objects from n without considering the order. The number of combinations is given by:

$$\binom{n}{k} = C_n^k = C(n,k) = \frac{n!}{k!(n-k)!}, n \geq k.$$

Example: The number of ways to choose 2 people from 5 is $C(5,2) = 10$.

The Number of Arrangements with Repetitions:
The number of arrangements with repetitions is the number of ways to choose r objects from n available objects, where objects may repeat. The formula for the number of arrangements with repetitions is as follows:

$$A(n,r) = n^r,$$

where: n is the number of objects, r is the number of objects being selected.

Addition Rule (for disjoint events): If two events A and B cannot happen at the same time (they are disjoint), the probability of either event occurring is the sum of the probabilities of the individual events:

$$|A \cup B| = |A| + |B|$$

If the events are disjoint, then $A \cap B = \emptyset$.

Problems.

Let Ω be the set of numbers from 10000 to 99999. These are all five-digit numbers. We need to find the probability that a randomly selected number will be:

a) Divisible by 2,
b) Divisible by 5,
c) Divisible by 10,
d) Divisible by 3,
e) Divisible by 9.

Solution. Obviously that $|\Omega| = 90000$, it is the number of the 5-digit numbers.

a) Divisible by 2:

A number is divisible by 2 if its last digit is even (0, 2, 4, 6, 8). There are 5 options for the last digit. For the other 4 digits, we can choose any number (0-9 for the second, third, and fourth digits, and 1-9 for the first digit).

The number of numbers divisible by 2 is:

$$N_2 = 9 \times 10 \times 10 \times 10 \times 5 = 45000$$

The probability that the number is divisible by 2 is:

$$P(\text{divisible by 2}) = \frac{45000}{90000} = \frac{1}{2}$$

b) Divisible by 5:

A number is divisible by 5 if its last digit is 0 or 5. There are 2 options for the last digit. For the other 4 digits, we can choose any number as before.

The number of numbers divisible by 5 is:

$$N_5 = 9 \times 10 \times 10 \times 10 \times 2 = 18000$$

The probability that the number is divisible by 5 is:

$$P(\text{divisible by 5}) = \frac{18000}{90000} = \frac{1}{5}$$

c) Divisible by 10:

A number is divisible by 10 if its last digit is 0. For the other 4 digits, we can choose any number.

The number of numbers divisible by 10 is:

$$N_{10} = 9 \times 10 \times 10 \times 10 \times 1 = 9000$$

The probability that the number is divisible by 10 is:

$$P(\text{divisible by 10}) = \frac{9000}{90000} = \frac{1}{10}$$

d) Divisible by 3: A number is divisible by 3 if the sum of its digits is divisible by 3. Fix the last 4 digits and consider the remainders of their sum modulo 3, which can be 0, 1, or 2. Note that if the remainder is 0, then for the entire number to be divisible by 3, the first digit must be 3, 6, or 9. Similarly, if the remainder is 1, the first digit must be 2, 5, or 8, and if the remainder is 2, the first digit must be 1, 4, or 7. Than the number of numbers divisible by 3 is:

$$N_3 = 3 \times 10 \times 10 \times 10 \times 10 = 30000$$

The probability that the number is divisible by 3 is:

$$P(\text{divisible by 3}) = \frac{30000}{90000} = \frac{1}{3}$$

e) Divisible by 9:

Similarly to the case of divisibility by 3, for divisibility by 9, we note that when the last 4 digits are chosen arbitrarily, their sum can have 9 different remainders when divided by 9 (0, 1, 2, ..., 9), and for each such remainder, only 1 digit will be suitable for the first position in order for the entire number to be divisible by 9.

The number of numbers divisible by 9 is:

$$N_9 = 1 \times 10 \times 10 \times 10 \times 10 = 10000$$

The probability that the number is divisible by 9 is:

$$P(\text{divisible by 9}) = \frac{10000}{90000} = \frac{1}{9}$$

3. Geometry probability.

If the probability space is a geometric object (Ω is infinite, for example $[0,1], (0,1), \mathbb{R}, \mathbb{R}^2$), the probability of an event A is usually calculated as the ratio of the geometric measure (such as area or volume) of the event A to the total geometric measure of the space. Formally, for a space with measure μ, the probability of event A can be expressed as:

$$P(A) = \frac{\mu(A)}{\mu(\Omega)}$$

where $\mu(A)$ is the measure (such as length, square or volume) of the set of favorable outcomes, and $\mu(\Omega)$ is the measure of the entire sample space of possible outcomes.

Problem 1.

Let the set $\Omega = [0,1]$. Find the probability of the following events:

1. $A = \left[0, \frac{1}{2}\right)$
2. $B = \left\{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 0, 1\right\}$
3. $C = \left\{\frac{1}{n} \mid n \in \mathbb{N}\right\}$
4. $D = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{3}}{2}\right)$
5. $E = \left(\frac{\sqrt{2}}{2}, \sqrt{3}\right)$

Solution. Let $\Omega = [0,1]$ be the sample space. The probability of an event A is calculated as the length of the interval or set A divided by the total length of the sample space Ω, which is 1.

a) $A = [0, \frac{1}{2})$

This is the interval from 0 to $\frac{1}{2}$, not including $\frac{1}{2}$. The length of the interval A is $\frac{1}{2} - 0 = \frac{1}{2}$.

$$P(A) = \frac{\text{Length of } A}{\text{Length of } \Omega} = \frac{\frac{1}{2}}{1} = \frac{1}{2}$$

b) $B = \left\{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 0, 1\right\}$

This is a finite set of individual points. In classical probability theory, the probability that a random point lands on any particular point is 0. Therefore:

$$P(B) = 0$$

c) $C = \left\{ \frac{1}{n} \mid n \in \mathbb{N} \right\}$

This is a countable set of points of the form $\frac{1}{n}$, where $n \in \mathbb{N}$. For countable sets, the probability of the event is also 0:

$$P(C) = 0$$

d) $D = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{3}}{2} \right)$

This is an open interval between $\frac{\sqrt{2}}{2}$ and $\frac{\sqrt{3}}{2}$. The length of the interval is:

$$\text{Length of } D = \frac{\sqrt{3}}{2} - \frac{\sqrt{2}}{2}$$

Approximating:

$$\frac{\sqrt{3}}{2} \approx 0.866 \quad \text{and} \quad \frac{\sqrt{2}}{2} \approx 0.707$$

So the length of the interval is:

$$0.866 - 0.707 = 0.159$$

Thus:

$$P(D) = 0.159$$

e) $E = \left(\frac{\sqrt{2}}{2}, \sqrt{3} \right)$

This is an open interval between $\frac{\sqrt{2}}{2}$ and $\sqrt{3}$. However, $\sqrt{3} \approx 1.732$, which is greater than 1. Since the sample space is $[0, 1]$, we must consider the intersection of E with Ω, which is the interval $\left(\frac{\sqrt{2}}{2}, 1 \right)$.

The length of this interval is:

$$\text{Length of } \left(\frac{\sqrt{2}}{2}, 1 \right) = 1 - \frac{\sqrt{2}}{2}$$

Approximating:

$$1 - \frac{\sqrt{2}}{2} \approx 1 - 0.707 = 0.293$$

Thus:

$$P(E) = 0.293$$

Problem 2.

What is the probability of landing inside a circle with radius 1/2 and center at $(0, 0)$, if $\Omega = [0, 1] \times [0, 1]$?

1. The area of the square $\Omega = [0, 1] \times [0, 1]$ is:

$$S_\Omega = 1 \times 1 = 1.$$

2. The area of the circle with radius 1 and center at $(0, 0)$ is:

$$S_{\text{circle}} = \pi \times \left(\frac{1}{2} \right)^2 = \frac{\pi}{4}.$$

However, the circle does not completely fit within the square, so we must calculate the portion of the circle that lies inside the square, it is obvious that the square of the portion is $\frac{1}{4}$ of S_{circle}.

Thus, the probability of landing inside the circle is the area of the circle inside the square divided by the area of the square $= \frac{\pi}{16}$.

Problem 3

What is the probability of landing inside a circle with radius 1 and center at $(0,0)$, if $\Omega = [-1,1] \times [-1,1]$?

1. The area of the square $\Omega = [-1,1] \times [-1,1]$ is:

$$S_\Omega = 2 \times 2 = 4.$$

2. The area of the circle with radius 1 and center at $(0,0)$ is:

$$S_{\text{circle}} = \pi \times 1^2 = \pi.$$

The circle completely fits inside the square, since the square has side length 2, and the circle's radius is 1.

Thus, the probability of landing inside the circle is the ratio of the area of the circle to the area of the square:

$$P = \frac{S_{\text{circle}}}{S_\Omega} = \frac{\pi}{4}.$$

Problem 4.

We select two numbers from the interval $[0,10]$. What is the probability that:

a) The first number is greater than the second;

b) One of the numbers is twice the other;

c) Their sum is greater than 9;

d) Their product is less than 4.

Solution

Let X and Y be the two numbers selected from the interval $[0,10]$.

We assume that the selection is uniformly distributed, so the joint probability density function (pdf) is uniform over the square $[0,10] \times [0,10]$.

a) The first number is greater than the second

We want to find $P(X > Y)$. Geometrically, this corresponds to the area of the region in the square $[0,10] \times [0,10]$ where $X > Y$. The area where $X > Y$ is half of the total area of the square, which is 100. Thus:

$$P(X > Y) = \frac{50}{100} = 0.5$$

b) One of the numbers is twice the other

We want to find $P(X > 2Y \text{ or } Y > 2X)$. The area where either $X > 2Y$ or $Y > 2X$ is two triangles within the square. Each triangle has an area of 25. Thus, the total area is 50, and the probability is:

$$P(X > 2Y \text{ or } Y > 2X) = \frac{50}{100} = 0.5$$

c) Their sum is greater than 9

We want to find $P(X + Y > 9)$. The inequality $X + Y = 9$ defines a line in the square. The area above this line corresponds to the region where $X + Y > 9$. The area of the region where $X + Y \leq 9$ is a triangle with base and height of 9, so its area is 40.5. Therefore:

$$P(X + Y > 9) = \frac{59.5}{100} = 0.595$$

d) Their product is less than 4

We want to find $P(X \times Y < 4)$. This is the region under the curve $X \times Y = 4$, which is a hyperbola. Using integration, we approximate the probability as:

$$P(X \times Y < 4) = \frac{100 - \int\limits_{\frac{2}{5}}^{10} (10 - \frac{4}{x})dx}{100} = \frac{4 + 8\ln(5)}{100} \approx 0.1688$$

4. Bertrand's paradox.

Introduction.

Bertrand's paradox is a problem in probability theory that demonstrates how the definition of a random variable (in this case, a random chord in a circle) can lead to different results depending on the method used to define the randomness. This paradox shows that the concept of a "random chord" is not well-defined unless we specify how the chord is chosen. The problem arises from the fact that different methods of defining a random chord lead to different probabilities for the length of the chord compared to the radius of the circle. These different methods lead to differing results, which is why this is a paradox.

The paradox involves several different ways to define a "random" chord, each leading to a different probability for the length of the chord compared to the radius of the circle. These different methods lead to differing results, which is why this is a paradox.

Statement of the Problem

Consider a circle of radius R. A **random** chord is chosen in one of the following ways:
The random endpoints method: Choose two random points on the circumference of the circle and draw the chord joining them. To calculate the probability in question imagine the triangle rotated so its vertex coincides with one of the chord endpoints. Observe that if the other chord endpoint lies on the arc between the endpoints of the triangle side opposite the first point, the chord is longer than a side of the triangle. The length of the arc is one third of the circumference of the circle, therefore the probability that a random chord is longer than a side of the inscribed triangle is $\frac{1}{3}$.

The random radial point method: Choose a radius of the circle, choose a point on the radius and construct the chord through this point and perpendicular to the radius. To calculate the probability in question imagine the triangle rotated so a side is perpendicular to the radius. The chord is longer than a side of the triangle if the chosen point is nearer the center of the circle than the point where the side of the triangle intersects the radius. The side of the triangle bisects the radius, therefore the probability a random chord is longer than a side of the inscribed triangle is $\frac{1}{2}$.

The random midpoint method. Choose a point anywhere within the circle and construct a chord with the chosen point as its midpoint. The chord is longer than a side of the inscribed triangle if the chosen point falls within a concentric circle of radius$\frac{1}{2}$ the radius of the larger circle. The area of the smaller circle is one fourth the area of the larger circle, therefore the probability a random chord is longer than a side of the inscribed triangle is $\frac{1}{4}$.

Conclusion

This paradox demonstrates that the probability of an event can depend on how the random variable is defined. In this case, the definition of a "random chord" leads to different probabilities for the length of the chord relative to the radius of the circle. This illustrates that the concept of "randomness" needs to be carefully defined in probability problems.

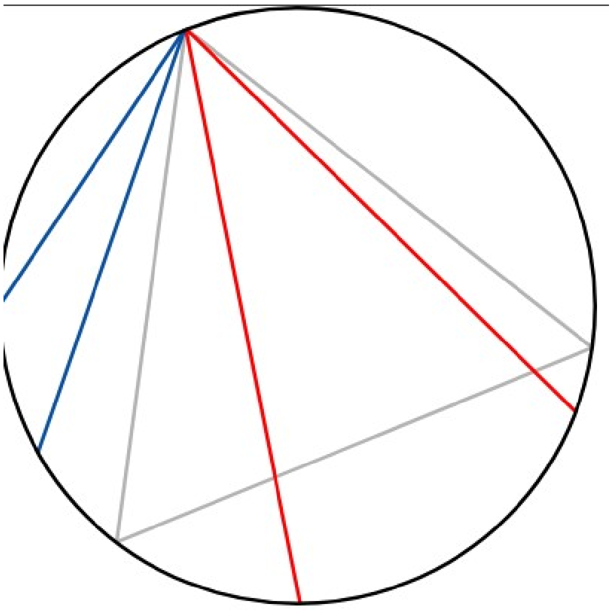

Figure 1: The "random endpoints" method

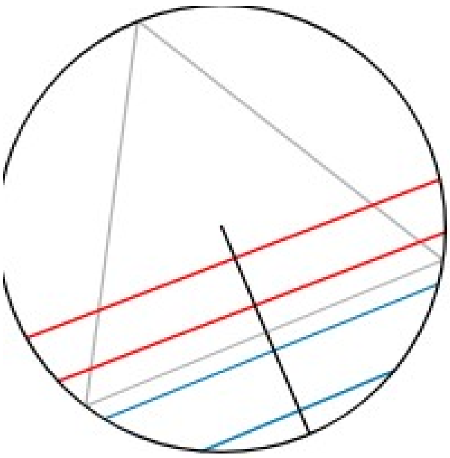

Figure 2: The "random radial point" method

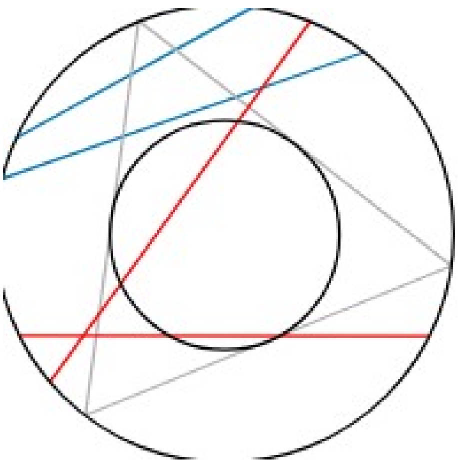

Figure 3: The "random midpoint" method.

§2. Probability Space

A probability space is a mathematical structure that consists of three components: $(\Omega, \mathcal{F}, P)$, where:

- Ω is the sample space (the set of all possible outcomes),
- $\mathcal{F}$ is a sigma-algebra (a collection of subsets of Ω that includes the empty set and is closed under complement and countable unions),
- P is the probability measure, which assigns probabilities to events in $\mathcal{F}$.

A probability space is not just a set of outcomes, but a structure that allows us to perform probability calculations using functions and axioms.

1. Properties of Probability Measure

The probability measure P has several key properties:

- **Non-negativity:** For any event A, $P(A) \geq 0$,
- **Normalization:** $P(\Omega) = 1$, the probability of the entire sample space is 1,
- **Countable additivity:** If the events $A_1, A_2, A_3, \ldots$ are disjoint (pairwise non-intersecting), then:

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

2. Kolmogorov's Axiomatic Framework

As mentioned earlier, Kolmogorov's axiomatic approach provides a formal definition of probability using a system of axioms. The three main axioms are:

1. $\forall A \in \mathfrak{F} \quad P(A) \geq 0$,
2. $P(\Omega) = 1$,
3. $\forall A \in \mathfrak{F}, \quad \overline{A} = \Omega \setminus A, \quad P(\overline{A}) = 1 - P(A)$.
4. For any countable collection of disjoint events $A_1, A_2, A_3, \ldots$, $(\forall i \neq j \quad A_i \cap A_j = \emptyset)$ we have:

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

These axioms form the foundation for all further probability theory.

3. Property of probability.

1. $\forall A, B \in \mathfrak{F},\ A \subset B, \Rightarrow P(A) \leq P(B)$.
Proof. Consider A and $B \setminus A$, which are disjoint, then apply axioms 1 and 4.
$P(B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A) \geq P(A)$. 2. $\forall A, B\mathfrak{F},\ A \subset B\ P(B \setminus A) = P(B) - P(A)$.
Proof. From the property 1 we have $P(B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A)$ then $P(B \setminus A) = P(B) - P(A)$.
3. $\forall A, B \in \mathfrak{F} \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
Proof Consider B and $A \setminus (A \cap B)$, which are disjoint, then apply the 4-th axiom and the 2-nd property.

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

4. $\forall A, B, C \quad P(A \cup B \cup C) = P(A) + P(B) = P(C) - P(A \cap B) - P(A \cap C) - P(C \cap B) + P(A \cap B \cap C)$.
Proof

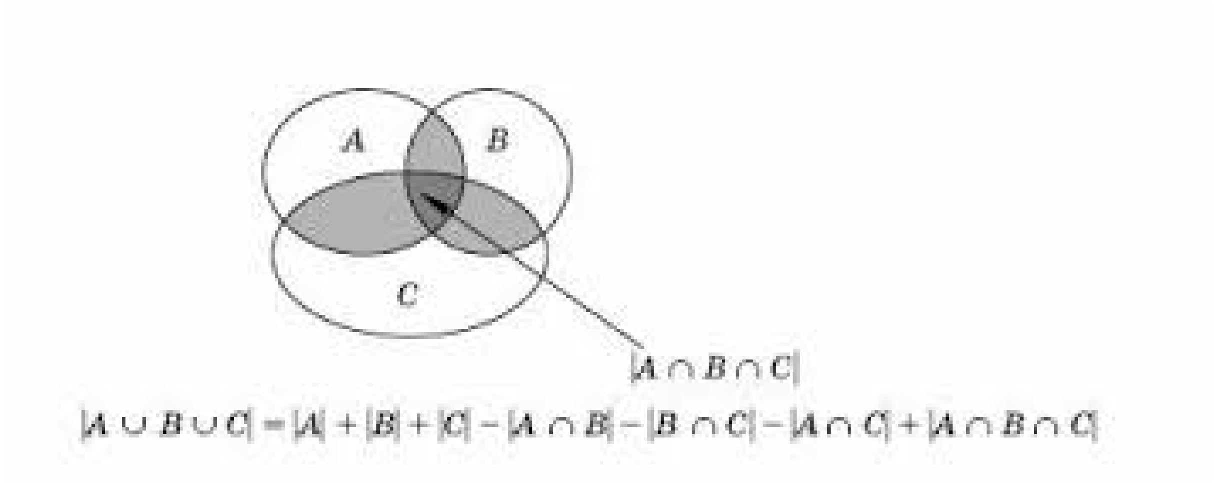

Figure 4: Union of 3 sets

5. **The inclusion-exclusion principle**.
Let $A_1, A_2, \ldots, A_n \in \mathfrak{F}$ then:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{1\le i<j\le n} P(A_i \cap A_j) + \sum_{1\le i<j<k\le n} P(A_i \cap A_j \cap A_k) - \cdots +$$

$$+(-1)^{k+1} \sum_{1\le i_1<i_2<\cdots<i_k\le n} P\left(\bigcap_{j=1}^{k} A_{i_j}\right) + \cdots + (-1)^{n+1} P\left(\bigcap_{i=1}^{n} A_i\right)$$

Problem.

Let there be a standard deck of 52 cards. We are drawing 5 cards. What is the probability that:

a) exactly 2 aces and exactly 3 black cards are drawn. b) at least 3 aces are drawn, exactly 3 red cards, exactly 1 diamond, exactly 1 club, and exactly 1 spade are drawn.
Let the deck of cards consist of 52 cards, with 4 aces, 26 black cards (spades and clubs), 26 red cards (hearts and diamonds), and 13 cards from each suit.

Solution.

The total number of ways to choose 5 cards from 52 is:

$$\binom{52}{5}$$

a) Probability of drawing exactly 2 aces and exactly 3 black cards.

We need to choose 3 black cards from 26, 2 red cards from 26 and 2 aces from 4.

We break the event into three mutually exclusive sub-events:
1. Two red aces and three non-ace black cards.
2. Two black aces, two non-ace red cards, and one non-ace black card.
3. Two aces of different colors, two non-ace black cards, and one non-ace red card.
Now we calculate the favorable outcomes for each sub-event:

1. Two red aces and three non-ace black cards.

There are 2 red aces in the deck, and we need to choose 2 red aces from 2:

$$\binom{2}{2}$$

There are 26 black cards in the deck, and we need to choose 3 non-ace black cards from 24:

$$\binom{24}{3}$$

Thus, the favorable outcomes for this sub-event are:

$$\binom{2}{2} \cdot \binom{24}{3}$$

2. Two black aces, two non-ace red cards, and one non-ace black card.

We need to choose 2 black aces from 2:

$$\binom{2}{2}$$

There are 26 red cards, and we need to choose 2 non-ace red cards from 24:

$$\binom{24}{2}$$

We also need to choose 1 non-ace black card from 24:

$$\binom{24}{1}$$

Thus, the favorable outcomes for this sub-event are:

$$\binom{2}{2} \cdot \binom{24}{2} \cdot \binom{24}{1}$$

3. Two aces of different colors, two non-ace black cards, and one non-ace red card.

We need to choose 1 red ace and 1 black ace:

$$\binom{2}{1} \cdot \binom{2}{1}$$

We need to choose 2 non-ace black cards from 24:

$$\binom{24}{2}$$

We need to choose 1 non-ace red card from 24:

$$\binom{24}{1}$$

Thus, the favorable outcomes for this sub-event are:

$$\binom{2}{1} \cdot \binom{2}{1} \cdot \binom{24}{2} \cdot \binom{24}{1}$$

Total favorable outcomes

The total number of favorable outcomes is:

$$\binom{2}{2} \cdot \binom{24}{3} + \binom{2}{2} \cdot \binom{24}{2} \cdot \binom{24}{1} + \binom{2}{1} \cdot \binom{2}{1} \cdot \binom{24}{2} \cdot \binom{24}{1}$$

Probability

The probability is:

$$P = \frac{\binom{2}{2} \cdot \binom{24}{3} + \binom{2}{2} \cdot \binom{24}{2} \cdot \binom{24}{1} + \binom{2}{1} \cdot \binom{2}{1} \cdot \binom{24}{2} \cdot \binom{24}{1}}{\binom{52}{5}}$$

b)

Probability of drawing at least 3 aces, exactly 3 red cards, exactly 1 diamond, exactly 1 club, and exactly 1 spade.
We break the event "drawing at least 3 aces, exactly 3 red cards, exactly 1 diamond, exactly 1 club, and exactly 1 spade" into the following mutually exclusive sub-events:
1. Exactly 4 aces and 1 non-ace heart card;
2. Exactly 3 aces (2 red and 1 black), 1 non-ace heart cards and 1 black non-ace card other color that the black ace.
3.Exactly 3 aces (2 black and 1 red), 2 red non-ace cards (1 heart and 1 diamond or 2 heart)

1. **Exactly 4 aces and 1 non-ace heart card:**

$$\binom{4}{4} \times \binom{12}{1} = 1 \times 12 = 12$$

2. **Exactly 3 aces (2 red and 1 black), 1 non-ace heart card, and 1 black non-ace card of a color different from the black ace:**

$$\binom{2}{2} \times \binom{2}{1} \times \binom{12}{1} \times \binom{12}{1} = 1 \times 2 \times 12 \times 12 = 288$$

3. **Exactly 3 aces (2 black and 1 red), 2 red non-ace cards (1 heart and 1 diamond or 2 hearts):**

$$\binom{2}{2} \times \binom{1}{1} \times (\binom{12}{1} \times \binom{12}{1} + \binom{12}{2})$$

Thus, the probability of this event is:

$$P = \frac{\binom{4}{4} \times \binom{12}{1} + \binom{2}{2} \times \binom{2}{1} \times \binom{12}{1} \times \binom{12}{1} + \binom{2}{2} \times \binom{1}{1} \times (\binom{12}{1} \times \binom{12}{1} + \binom{12}{2})}{\binom{52}{5}} \approx 0.000196$$

Problem 1. What is the probability that a randomly selected seven-digit number will have all different digits?
solution. $|\Omega| = 9 \cdot 10^6$ or 10^7;
$|A| = 9 * 9 * 8 * 7 * 6 * 5 * 4 = A_1^7 0 - A_9^6$;
$P = \frac{9*9*8*7*6*5*4}{9*10^6}$.
Problem 2. There are 9 candies and 5 bags. Candies are randomly distributed into the bags. What is the probability that none of the bags will be empty?
Solution. To divide the set into 5 parts, we need 4 dividers. For this, we add 4 more candies of a different type (e.g., blue ones). Then $|Omega| = C(13, 4)$ and $|A| = C(8, 4)$, then

$P = \frac{5*6*7*8}{10*11*12*13} \approx 0.0979.$

Problem 3. Entering the restaurant, n guests left their hats with the doorman, and upon leaving, they received their hats back. The doorman handed out the hats randomly. What is the probability that each guest will receive someone else's hat? For $n = 2, 3, 4, 5$.

Solution: Let the set A_i denote the event where the i-th guest leaves with their own hat. Than $|derangement| = |\overline{A_1 \cup A_2 \cup ... \cup A_n}| = n! - |A_1 \cup A_2 \cup ... \cup A_n|$.

Formula for the probability of a derangement:

For n objects, the probability that no object remains in its original position is:

$$P(\text{derangement}) = \frac{D_n}{n!}$$

where D_n is the number of derangements for n objects, and $n!$ is the total number of possible permutations.

Number of derangements D_n:

The number of derangements for n objects is given by the formula:

$$D_n = n! \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^n}{n!}\right)$$

Now, let's compute the number of derangements and the probability for $n = 2, 3, 4, 5$.

For $n = 2$:

$$D_2 = 2! \left(1 - \frac{1}{1!} + \frac{(-1)^2}{2!}\right) = 2\left(1 - 1 + \frac{1}{2}\right) = 1$$

The total number of permutations for two objects: $2! = 2$.

Thus, the probability of a derangement for $n = 2$ is:

$$P(\text{derangement for } n = 2) = \frac{D_2}{2!} = \frac{1}{2}$$

For $n = 3$:

$$D_3 = 3! \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!}\right) = 6\left(1 - 1 + \frac{1}{2} - \frac{1}{6}\right) = 6 \times \frac{1}{3} = 2$$

The total number of permutations for three objects: $3! = 6$.

Thus, the probability of a derangement for $n = 3$ is:

$$P(\text{derangement for } n = 3) = \frac{D_3}{3!} = \frac{2}{6} = \frac{1}{3}$$

For $n = 4$:

$$D_4 = 4! \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!}\right) = 24\left(1 - 1 + \frac{1}{2} - \frac{1}{6} + \frac{1}{24}\right) = 24 \times \frac{9}{24} = 9$$

The total number of permutations for four objects: $4! = 24$.

Thus, the probability of a derangement for $n = 4$ is:

$$P(\text{derangement for } n = 4) = \frac{D_4}{4!} = \frac{9}{24} = \frac{3}{8}$$

For $n = 5$:

$$D_5 = 5! \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!}\right) = 120\left(1 - 1 + \frac{1}{2} - \frac{1}{6} + \frac{1}{24} - \frac{1}{120}\right) = 120 \times \frac{44}{120} = 44$$

The total number of permutations for five objects: $5! = 120$.

Thus, the probability of a derangement for $n = 5$ is:

$$P(\text{derangement for } n = 5) = \frac{D_5}{5!} = \frac{44}{120} = \frac{11}{30}$$

Final Answer:

- For $n = 2$: $\frac{1}{2}$
- For $n = 3$: $\frac{1}{3}$
- For $n = 4$: $\frac{3}{8}$
- For $n = 5$: $\frac{11}{30}$

§3.Conditional probability

Conditional probability is the probability of event A, given that event B has occurred. It is denoted as $P(A \mid B)$ and is defined by the following formula:

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

Here:

- $P(A \mid B)$ is the conditional probability of event A, given that event B has occurred.
- $P(A \cap B)$ is the probability that both events A and B occur.
- $P(B)$ is the probability that event B occurs.

If $P(B) = 0$, then the conditional probability $P(A \mid B)$ is undefined because there is no situation where event B could occur.

Table 1: Property of conditional probability $P(B) > 0$

probability	conditional probability
$P(A) \in [0, 1]$	$P(A\|B) \in [0, 1]$
$P(\Omega) = 1, P(\emptyset) = 0$	$P(\Omega\|B) = 1, P(\emptyset\|B) = 0$
$A \subset C, P(A) \leq P(C)$	$P(A\|B) \leq P(C\|B)$
$P(\overline{A}) = 1 - P(A)$	$P(\overline{A}\|B) = 1 - P(A\|B)$
$P(A + C) = P(A) + P(C) - P(AC)$	$P(A + C\|B) = P(A\|B) + P(C\|B) - P(AC\|B)$
$A \subset C, P(C - A) = P(C) - P(A)$	$P(C - A\|B) = P(C\|B) - P(A\|B)$

Probability multiplication formula $P(AB) = P(B)P(A|B) = P(A)P(B|A)$
For $A_1, ..., A_n$:

$$P(A_1 * A_2 * ... * A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 * A_2) * ... * P(A_n|A_1 * ... * A_{n-1}).$$

Problems.

1. In a box, there are 6 white balls and 9 black balls. Two balls are randomly drawn at once. What is the probability that both are white, given that the first ball is white?
answer
$\frac{5}{14}$
2. A dice is rolled twice. What is the probability of rolling at least one six, given that all rolled numbers are different?
answer.
$\frac{1}{3}$
3. The letters of the word "mathematics" are written on cards. Cards are drawn randomly, one after another.

What is the probability of obtaining the words "math" and "mama"?

answer.

a) $\frac{2}{11} \cdot \frac{2}{10} \cdot \frac{2}{9} \cdot \frac{1}{8}$

b) $\frac{2}{11} \cdot \frac{2}{10} \cdot \frac{1}{9} \cdot \frac{1}{8}$.

3. Two numbers are randomly selected from the interval [0,1]. What is the probability that:

a) The sum is greater than 1/2, given that the second number is greater than the first?

b) One number is twice the other, given that their product is greater than 1/20?

answer.

a)$\frac{7}{8}$

b)$\frac{9-\ln(10)}{20-2\ln(10)}$

4. Two numbers are randomly selected: the first from the interval [0,5] and the second from [-2,2]. What is the probability that:

a) Their sum is less than 2 and the first number is greater than the second?

b) Their sum is greater than 2 and their product is greater than -1?

c) Their sum is less than 1 and their product is at least -2?

answers.

a)$\frac{7}{20}$

b) $\frac{\frac{19}{2}-\sqrt{2}+\ln(5)-\ln(1+\sqrt{2})}{20}$

c) $\frac{10+2\ln\left(\frac{5}{2}\right)}{20}$

Problem 3. The letters of the word "mathematics" are written on cards. Cards are drawn randomly, one after another. What is the probability of obtaining the words "math" and "mama"?

answer.

a) $\frac{2}{11} \cdot \frac{2}{10} \cdot \frac{2}{9} \cdot \frac{1}{8}$

b) $\frac{2}{11} \cdot \frac{2}{10} \cdot \frac{1}{9} \cdot \frac{1}{8}$.

3. Two numbers are randomly selected from the interval [0,1]. What is the probability that:

a) The sum is greater than 1/2, given that the second number is greater than the first?

b) One number is twice the other, given that their product is greater than 1/20?

answer.

a)$\frac{7}{8}$

b)$\frac{9-\ln(10)}{20-\ln(20)}$

4. Two numbers are randomly selected: the first from the interval [0,5] and the second from [-2,2]. What is the probability that:

a) Their sum is less than 2 and the first number is greater than the second?

b) Their sum is less than 1 and their product is less than -2?

answers.

a)$\frac{7}{20}$

b) $\frac{5-4\ln 2}{40}$

Independence.

The event A is independent from B, if the conditional probability of A given B is equal to the probability of A:

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

Similarly, $P(B \mid A) = P(B)$.

In probability theory, **independent events** are events whose occurrence or non-occurrence does not affect the probability of the other event. Formally, two events A and B are independent if and only if their joint probability equals the product of their individual probabilities:

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

where $P(A \cap B)$ is the probability that both events A and B occur simultaneously, and $P(A)$ and $P(B)$ are the probabilities of events A and B occurring individually.

Properties of Independent Events

Independence and Unions of Events: If events A and B are independent, the probability of their union is given by:

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

Since $P(A \cap B) = P(A) \times P(B)$, this simplifies to:

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

Independence and Intersections of Multiple Events:

- If events $A_1, A_2, \ldots, A_n$ are independent, the probability of their simultaneous occurrence is the product of their individual probabilities:

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \times P(A_2) \times \cdots \times P(A_n)$$

Definition: The events $A_1, A_2, \ldots, A_n$ are independent collectively if for any k, such that $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, the following holds:

$$P(A_{i_1} A_{i_2} \ldots A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \ldots P(A_{i_k})$$

They are pairwise independent if for any i, j such that $1 \leq i < j \leq n$,

$$P(A_i \cap A_j) = P(A_i) P(A_j)$$

Note

It is important to distinguish between independent and mutually exclusive events. Mutually exclusive events cannot occur at the same time (e.g., flipping a coin and getting both heads and tails simultaneously), whereas independent events can occur together, but the occurrence of one does not affect the probability of the other.

Understanding the concept of independent events is fundamental in probability theory and statistics, as it underlies many statistical methods and models.

Problems.

0. Counterexample.

Consider a regular pyramid (where the probability of each face is $\frac{1}{4}$)), painted in three colors, for example, red (R), blue (B), and green (G), in the following way: each of three faces is entirely painted in one color, and the fourth face is painted in all three colors at once (RGB). Will the following events R, G, and B be independent?

1.

Prove or disprove.

a) If A and B are independent, then $\overline{A}$ and B are also independent. ✓

b) If A, B, and C are independent jointly, then

A and $B + C$ are independent. ✓

Solution. $P(A(B+C)) = P(AB + AC) = P(AB) + P(AC) - P(ABC) = P(A)P(B) + P(A)P(C) - P(A)P(BC) = P(A)(P(B) + P(C) - P(BC)) = P(A)P(B+C)$.

A and (B-C) are independent.✓
Solution. Note $B - C = B \cap \overline{C}$.

$$P(A(B-C)) = P(AB\overline{C}) = P(A)P(B)P(\overline{C}) = P(A)P(B\overline{C}) = P(A)P(B-C).$$

c) $P(A|B) + P(A|\overline{B}) = 1$?
solution. It is not true. Construct a counterexample: let A,B be independent, then $P(A|B) = P(A|\overline{B}) = P(A)$. Then our sum will be equal to 2 P(A), and we can choose such A that it won't be 1.
d)Will be $P(A) > P(B)$ if $P(A|B) > P(B|A)$ and $P(A), P(B) > 0$– Homework.
2.
Consider a deck of 36 cards. One card is drawn. Now, consider the following events:
A – the drawn card is a spade.
B – the drawn card is a king. Will these events be independent?
Will something change if we add 2 jokers which have only colour.
solution. The first part: $P(A) = \frac{9}{36} = \frac{1}{4}; P(B) = \frac{4}{36} = \frac{1}{9}; P(AB) = \frac{1}{36} = P(A) \cdot P(B)$, hence A,B–independent. The second part, when we add 2 jokers (red and black), $|\Omega| = 38$
$P(A) = \frac{9}{38}, P(B) = \frac{4}{38}, P(AB) = \frac{1}{38} \neq P(A) \cdot P(B)$, hence A,B aren't independent. But if consider the following events
C– the drawn card a joker, D– the drawn cark is red, than

$$P(C) = \frac{2}{38};\ P(D) = \frac{19}{38},\ P(CD) = \frac{1}{38} = P(C) \cdot P(D),$$

hence these events are independent
3.
A dice is rolled twice. Let X be the result of the first roll and Y be the result of the second roll. Will the following events be independent together:

$$A = \{X + Y < 9\}, \quad B = \{X + Y \text{ is divisible by } 3\}, \quad C = \{X \text{ is divisible by } 2, Y \text{ is divisible by } 3\}?$$

Solution. $P(A) = \frac{26}{36}$, $P(B) = \frac{1}{3}$, $P(C) = \frac{1}{6}$
$P(AB) = \frac{7}{36} \neq P(A)P(B)$,
$P(AC) = \frac{1}{12} \neq P(A)P(C)$,
$P(BC) = \frac{1}{18} = P(B)P(C)$. Hence only B,C are independent.
4. Let $P(A) = 0.2$, $P(B) = 0.4$. A,B are independent, but are not mutually exclusive. Find $P(AB \cup \overline{B}|A)$
Solution. $P(AB \cup \overline{B}|A) = \frac{P((AB \cup \overline{B})A)}{P(A)} = \frac{P(A)}{P(A)} = 1$.

Total Probability Formula and Bayes' Theorem

Total Probability Formula (TPF):

Let $\{H_i\}, i = 1..n$ be a **complete system of events**, i.e.,
$\forall i \neq j \quad H_i \cap H_j = \varnothing; \quad \sum\limits_{i=1}^{n} H_i = \Omega, \quad P(H_i) > 0$. For any A,

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

Proof: $P(A) = P(A \cdot \Omega) = P(A \cdot (\cup_i H_i)) =$ since the events AH_i are mutually exclusive, because their intersection is obviously empty for $i \neq j$,

$$\sum_i P(AH_i) = \sum_i P(H_i)P(A|H_i).$$

Bayes' Theorem:
Let $\{H_i\}, i = 1..n$ be a **complete system of events** and some event A. Then, for all $k = 1..n$,

$$P(H_k|A) = \frac{P(H_k)P(A|H_k)}{\sum\limits_{i=1}^{n} P(H_i)P(A|H_i)}.$$

Note: 1.$P(H_k|A) = \frac{P(H_k \cdot A)}{P(A)}$, after which the numerator is expanded using the multiplication rule of probabilities and the denominator using TPF.
2. $\sum\limits_{k=1}^{n} P(H_k|A) = P(\Omega|A) = 1$. 3. The following formula is correct too:

$$P(A|C) = \sum_{k=1}^{n} P(A|H_k\, C)P(H_k|C),$$

where $P(H_iC) > 0, H_i \cap H_j = \emptyset, i \neq j, AC \subset \cup_{k=1}^{n} H_k$
Proof.

$$P(A|C) = \frac{P(AC)}{P(C)} = \frac{P(AC(\cup_{k=1}^{n} H_k))}{P(C)} = \frac{P(\cup_{k=1}^{n}(ACH_k))}{P(C)} = \frac{\sum\limits_{k=1}^{n} P(AH_kC)}{P(C)} =$$

$$= \frac{\sum\limits_{k=1}^{n} P(A|H_kC)P(H_kC)}{P(C)} = P(A|C) = \sum_{k=1}^{n} P(A|H_k\, C)P(H_k|C).$$

Problems

1. In the first box, there are 3 white and 8 black balls, and in the second, there are 6 white and 5 black balls. One ball is transferred from the first box to the second at random. What is the probability of drawing a black ball from the first box now? What is the probability of drawing a white ball from the second box now?
Answer.
P(drawing black ball from the first box)=$\frac{8}{11}$
P(drawing a white ball from the second box)=$\frac{23}{43}$
2. In a box, there are three new balls and three already used ones. For the first game, two balls are randomly selected, and after the game, they are placed back. What is the probability of drawing two new balls in the second game?
Answer. $P = \frac{2}{25}$
3. There are three factories producing parts. The first factory produces 45% of the total production, and the second factory produces 20%. The probability of a defective part from the first factory is 2%, from the second factory is 4%, and from the third factory is 3%.
After being sent to the store, a defective part was found. What is the probability that it was produced by the first, second, or third factory?
Answer.
P(defect)=0.0275. P(defect on the 1-st factory)=$\frac{18}{55}$,
P(defect on the 2-nd factory)=$\frac{16}{55}$,
P(defect on the 3-rd factory)=$\frac{21}{55}$.
4. Two hunters simultaneously and independently shoot at a wild boar. It is known that the first hunter hits with probability 0.8, and the second with probability 0.4. The boar is shot, and there is 1 bullet in it. How should the boar be divided? **Answer.** The boar should be divided between the first and the second in a ratio of 6 to 1.

§4. Bernoulli Trials (Scheme).

Problem Statement. Let a certain probability $p \in (0,1)$ be given, which represents the probability of some event $A \subset \Omega$. We define success as the occurrence of event A, and failure otherwise. Thus, p is the probability of success, and $q = 1-p$ is the probability of failure. In this case, we consider a simple experiment with two possible outcomes. This experiment is repeated independently n times.
How many successes can occur in n experiments, and what is the probability of this?
Example.

a) What is the probability that event A occurs twice in three trials?

$$3p^2(1-p) = c_3^2 p^2(1-p)$$

b) What is the probability that event A occurs once or twice in four trials?

$$c_4^2 p^2(1-p)^2$$

c) What is the probability that event A occurs at least three times in four trials?

$$c_4^3 p^3(1-p) + c_4^4 p^4$$

d) What is the probability that event A occurs at least twice and exactly three times in five trials?

$$1-(1-p)^5 - 5p(1-p)^4 \quad \textbf{and} \quad c_5^3 p^3(1-p)^2 = 10p^3(1-p)^2$$

e) What is the probability that event A occurs exactly k times in n trials? $c_n^k p^k(1-p)^{n-k}$

Define the probability of obtaining k successes in n trials:

$$P_n(k) = \mathbb{P}(k \text{ successes in } n \text{ experiments}) = \binom{n}{k} p^k(1-p)^{n-k}.$$

Note: The probability of a fixed sequence of successes and failures:

$$P = p^k(1-p)^{n-k}.$$

Is the obtained value $P_n(k)$ truly a probability?
1. It is obvious that $P_n(k) > 0$. 2. Consider whether $\sum\limits_{k=0}^{n} P_n(k)$ equals 1:

$$\sum_{k=0}^{n} P_n(k) = \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k} = (p+(1-p))^n = 1.$$

Lemma. For a fixed n, the sequence $P_n(0), P_n(1), ..., P_n(n)$ is unimodal, i.e.,

$$\begin{cases} P_n(k+1) \geq P_n(k), & \text{for } k \leq np-1+p \\ P_n(k+1) < P_n(k), & \text{for } k > np-1+p. \end{cases}$$

Proof. Consider the ratio of consecutive elements:

$$\frac{P_n(k+1)}{P_n(k)} = \frac{n-k}{k+1} \cdot \frac{p}{1-p},$$

and analyze when this is < 1 or ≥ 1, obtaining the desired result.

Corollary. The maximum of $P_n(k)$ is attained for $k \in [np - 1 + p, np + p]$.
Proof.
Let m be a maximum point, then by the lemma:

- $P_n(m+1) \leq P_n(m)$, which holds if $m \geq np - 1 + p \iff m \geq np - q$.
- $P_n(m-1) \geq P_n(m)$, which holds if $m - 1 \leq np - 1 + p \iff m \leq np + p$.

Thus, $np - q \leq m \leq np + p$, and the length of this interval is $p + q = 1$.
Note.

1. If the boundaries are non-integer, the maximum is unique; if they are integers, there are two maxima.
2. Such a number k (the number of successes in n trials) is called the most probable outcome.

Now, define the probability that the number of successes is at most m_2 and at least m_1 in n trials:

$$P_n(m_1, m_2) = \sum_{k=m_1}^{m_2} P_n(k) = \sum_{k=m_1}^{m_2} \binom{n}{k} p^k (1-p)^{n-k}.$$

Alternatively,

$$P_n(m_1, m_2) = 1 - \sum_{k=0}^{m_1-1} \binom{n}{k} p^k (1-p)^{n-k} - \sum_{k=m_2+1}^{n} \binom{n}{k} p^k (1-p)^{n-k}.$$

Problems.

1. The probability of hitting a target in a single shot is $\frac{1}{3}$. Six shots are fired. What is the probability of exactly 2 hits? Exactly 3 hits? At least 2 hits? At least 1 hit? What is the most probable number of hits?
Answers.

$$P_6(2) = C_6^2 p^2 (1-p)^4 = 15\frac{2^4}{3^6} = \frac{240}{729};$$

$$P_6(3) = c_6^3 p^3 (1-p)^3 = 20\frac{8}{729} = \frac{160}{729};$$

$$P_6(\geq 2) = 1 - P_6(0) - P_6(1) = 1 - \left(\frac{2}{3}\right)^6 - c_6^1 \frac{2^5}{729} = \frac{729 - 64 - 192}{729} = \frac{473}{729};$$

$$P_6(\geq 1) = 1 - P_6(0) = 1 - \frac{64}{729} = \frac{665}{729};$$

the most probable number of hits (m) $\in [np - q, np + p] = [2 - \frac{2}{3}, 2 + \frac{1}{3}] \Rightarrow$ m=2.

2. The probability of hitting a target at least once in three shots is 0.875. What is the probability of hitting the target in a single shot?
Answer. The probability of hitting the target in a single shot p=0.5

Problem.
3. A plane with 150 seats has sold 153 tickets. Find the probability of overbooking if the probability of an individual passenger not showing up is 0.01.
Answer.n=153,p=0.01 then
P(overbooking)=$P_{153}(\leq 2) = P_{153}(0) + P_{153}(1) + P_{153}(2) \approx 0.8$.
if n=152, p=0.01 P(overbooking)=$P_{152}(\leq 1) \approx 0.55$.
4. Two players alternately toss a coin. The winner is the first to obtain heads. Find the probability that the

game ends on the k-th toss. How much more likely is the first player to win than the second?
Answer. P(the game ends on the k-th toss)=$\frac{1}{2^k}$;
P(I-winner)= $\sum\limits_{i=1}^{\infty} \frac{1}{2^{2i-1}} = \frac{2}{3}$;
P(II-winner)=$\sum\limits_{i=1}^{\infty} \frac{1}{2^{2i}} = \frac{1}{3}$.
5. A mathematician carries two matchboxes, each originally containing N matches. When he needs a match, he randomly selects one of the boxes. Find the probability that when he first finds an empty box, the other contains exactly r matches.
answer.

$$P = c_{2N-r-1}^{N-1} 2^{-(2N-r-1)}$$

Poisson's Theorem.

Let $np = \lambda = \text{const}$. Then for any m and constant λ:

$$\lim_{n\to\infty, np=\lambda} P_n(m) = \frac{\lambda^m}{m!} e^{-\lambda}.$$

Note.
Let

$$p_\lambda(m) = \frac{\lambda^m}{m!} e^{-\lambda},$$

then the following property holds:

$$\sum_{m=0}^{\infty} \frac{\lambda^m}{m!} e^{-\lambda} = 1.$$

Proof.

$$P_n(m) = C_n^m p^m (1-p)^{n-m} = \frac{n!}{m!(n-m)!} \left(\frac{\lambda}{n}\right)^m \left(1 - \frac{\lambda}{n}\right)^{n-m} =$$

$$= \frac{1}{m!} n(n-1)\dots(n-m+1) \frac{\lambda^m}{n^m} \left(1 - \frac{\lambda}{n}\right)^{n-m} =$$

$$= \frac{\lambda^m}{m!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{m-1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^{-m}.$$

Since for any fixed s, the following convergence holds:

$$1 - \frac{s}{n} \to 1 \quad \text{as} \quad n \to \infty,$$

then

$$P_n(m) \sim \frac{\lambda^m}{m!} \left(1 - \frac{\lambda}{n}\right)^n.$$

Now, using the *"remarkable limit"*

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda},$$

Problem.

0.. Apply the Poisson theorem to the 3-rd problem of the previous lists of problems.
n=153, p=0.01 => $\lambda = np = 1.53$
P(overbooking)=$P_{153}(\leq 2) = P_{153}(0) + P_{153}(1) + P_{153}(2) \approx$
$\approx Pois_{\lambda=1.53}(0) + Pois_{\lambda=1.53}(1) + Pois_{\lambda=1.53}(2) = e^{-1.53}\left(1 + 1.53 + \frac{1.53^2}{2}\right) \approx 0.801$
1.A machine contains 2000 components. The probability of failure for a single component within a year is 0.001. What is the probability of exactly 2 failures in a year? Exactly 1? None? Exactly 3? What is the probability of at least 2 failures in a year?
Answer. $\lambda = np = 2$ then
$P(2) \approx Pois_{\lambda=2}(2) = 2e^{-2} \approx 0.271;$
$P(1) \approx Pois_{\lambda=2}(2) = 2e^{-2} \approx 0.271;$
$P(0) \approx Pois_{\lambda=2}(2) = e^{-2} \approx 0.135;$
$P(3) \approx Pois_{\lambda=2}(2) = \frac{4}{3}e^{-2} \approx 0.18;$
$P(\geq 2) \approx Pois_{\lambda=2}(\geq 2) = 1 - Pois(0) - Pois(1) = 1 - 3e^{-2} \approx 0.594;$

2. In a village, there are 1,000 houses, each insured for a year against fire for a sum of 100,000 yuan. The annual insurance premium is 300 yuan. The probability of a house catching fire within a year is estimated to be 0.002. What is the probability that the insurance company will incur losses over the year?
Answer. $\lambda = np = 2$ then P(the insurance company will incur losses over the year)$\approx Pois_{\lambda=2}(\geq 3)$ or $\approx Pois_{\lambda=2}(> 3)$.
Consider only the first probability (the second will be almost the same)

$$Pois_{\lambda=2}(\geq 3) = 1 - Pois_{\lambda=2}(0) - Pois_{\lambda=2}(1) - Pois_{\lambda=3} = 1 - 5e^{-2} \approx 0.323.$$

3. To estimate the number of fish in a lake, 1,000 fish are caught, marked, and released back into the lake. Then, 150 fish are caught one by one and released again. For what total number of fish in the lake will the probability of catching exactly 10 marked fish among the 150 be the highest?
Answer. Let N be the number of fish then p(N)=$\frac{1000}{N}$ and we need find N:

$$max_N\, P_{150}(10) = max_N\, C_{150}^{10}(p(N))^{10}\,(1 - p(N))^{140} \Rightarrow N = 15000$$

Problem

A box contains 500 tomatoes. The probability that a tomato starts to rot within one week is 0.04. What is the probability that among 200 tomatoes, exactly 5, exactly 8, exactly 10, exactly 15, exactly 20, exactly 50, and exactly 100 will be rotten?
Answer. $\lambda = 0.04 \cdot 200 = 8$. Then
$P_{200}(5) \approx Pois_8(5) = \frac{8^5}{|}5!e^{-8} \approx 0.092;$
$P_{200}(8) \approx Pois_8(8) = \frac{8^8}{|}8!e^{-8} \approx 0.1396;$
$P_{200}(10) \approx Pois_8(10) = \frac{8^10}{|}10!e^{-8} \approx 0.099;$
$P_{200}(15) \approx Pois_8(15) = \frac{8^{15}}{|}15!e^{-8} \approx 0.009;$
$P_{200}(20) \approx Pois_8(20) = \frac{8^{20}}{|}20!e^{-8} \approx 0.000159;$
$P_{200}(50) \approx Pois_8(50) = \frac{8^{50}}{|}50!e^{-8} \approx 1.57 \cdot 10^{-23};$
$P_{200}(100) \approx Pois_8(100) = \frac{8^{100}}{100!}e^{-8} \approx 7.3 \cdot 10^{-72};$

the local and interval Moivre-Laplace theorem

Now, define the probability that the number of successes is at most m_2 and at least m_1 in n trials:

$$P_n(m_1, m_2) = \sum_{k=m_1}^{m_2} P_n(k) = \sum_{k=m_1}^{m_2} \binom{n}{k} p^k (1-p)^{n-k}.$$

Alternatively,

$$P_n(m_1, m_2) = 1 - \sum_{k=0}^{m_1-1} \binom{n}{k} p^k (1-p)^{n-k} - \sum_{k=m_2+1}^{n} \binom{n}{k} p^k (1-p)^{n-k}.$$

Moivre-Laplace Local Limit Theorem.
Let us denote $x_n = \frac{m-np}{\sqrt{npq}}$. Suppose that $m \to \infty, n \to \infty$ and the values x_n are bounded (sufficiently $(n^{\frac{1}{6}-\epsilon})$). Then

$$\sqrt{npq} P_n(m) \sim \frac{1}{\sqrt{2\pi}} e^{-\frac{x_n^2}{2}}.$$

Remark. Let us define

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

this function is called the Gaussian function. Then, using the new notation, the probability from the theorem can be expressed as follows:

$$P_n(m) \sim \frac{1}{\sqrt{npq}} \phi(x).$$

Proof.
textbfProof. Since $n - m = nq - x_n\sqrt{npq}$ (see, for example, (9.2) below), due to the boundedness of x_n, the difference $n - m$ tends to ∞ along with n and m. We use Stirling's formula:

$$k! \sim \sqrt{2\pi k} \left(\frac{k}{e}\right)^k$$

for $k = n$, $k = m$, and $k = n - m$. Then we obtain

$$\sqrt{npq} P_n(m) = \sqrt{npq} \binom{n}{m} p^m q^{n-m} = \frac{\sqrt{npq} n!}{m!(n-m)!} p^m q^{n-m} \sim$$

$$\sim \frac{\sqrt{npq}\,\sqrt{n}\, n^n}{\sqrt{2\pi}\,\sqrt{m}\, m^m \sqrt{n-m}\,(n-m)^{n-m}}\, p^m\, q^{n-m} = \frac{1}{\sqrt{2\pi}} \left(\frac{np}{m}\right)^m \left(\frac{nq}{n-m}\right)^{n-m} \sqrt{\frac{np}{m}\frac{nq}{n-m}}.$$

Since, by definition of x_n, we have

$$m = np + x_n\sqrt{npq}, \tag{1}$$

$$n - m = n - np - x_n\sqrt{npq} = nq - x_n\sqrt{npq}, \tag{2}$$

it follows that

$$\frac{m}{np} = 1 + \frac{x_n\sqrt{q}}{\sqrt{np}} \to 1 \quad as \quad n \to \infty, \tag{3}$$

and

$$\frac{n-m}{nq} = 1 - \frac{x_n\sqrt{p}}{\sqrt{nq}} as \quad n \to \infty. \tag{4}$$

Therefore, for sufficiently large n and using $x^a = \exp(a \cdot \ln(x))$

$$\sqrt{npq} P_n(m) \sim \frac{1}{\sqrt{2\pi}} \exp\left(-m \ln\left(1 + \frac{x_n\sqrt{q}}{\sqrt{np}}\right) - (n-m)\ln\left(1 - \frac{x_n\sqrt{p}}{\sqrt{nq}}\right)\right).$$

Next, we use the following asymptotic expression:

$$\ln(1+z) = z - \frac{z^2}{2}(1+o(1)), \quad \text{as } z \to 0.$$

Then

$$\sqrt{npq}P_n(m) \sim \frac{1}{\sqrt{2\pi}} \exp\left(-m\left(\frac{x_n\sqrt{q}}{\sqrt{np}} - \frac{x_n^2 q}{2np}(1+o(1))\right) - \right.$$
$$\left.(n-m)\left(-\frac{x_n\sqrt{p}}{\sqrt{nq}} - \frac{x_n^2 p}{2nq}(1+o(1))\right)\right). \tag{5}$$

Using (1), (2),

$$x_n\left(\frac{(n-m)\sqrt{p}}{\sqrt{nq}} - \frac{m\sqrt{q}}{\sqrt{np}}\right) = \frac{x_n}{\sqrt{npq}}(nqp - x_n\sqrt{npq}p - npq - x_n\sqrt{npq}q) =$$
$$-x_n^2(p+q) = -x_n^2.$$

Applying (3), (4), we obtain that

$$\frac{x_n^2}{2}\left(\frac{(n-m)p}{nq}(1+o(1)) + \frac{mq}{np}(1+o(1))\right) \sim \frac{x_n^2}{2}\left(p(1+o(1)) + q(1+o(1))\right) = \frac{x_n^2}{2}.$$

Substituting the obtained expressions into (5), we get

$$\sqrt{npq}P_n(m) \sim \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_n^2}{2}(1+o(1))\right) \sim \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_n^2}{2}\right).$$

Problem.
Find the probability that in 150 shots, the target will be hit exactly 70 times if the probability of hitting the target in one shot is 0.4.
b) 60 times?

Moivre-Laplace Integral Limit Theorem.
Let us assume $a_n = \frac{m_1 - np}{\sqrt{npq}}, \quad b_n = \frac{m_2 - np}{\sqrt{npq}}$. Suppose that $m_1 \to \infty, n \to \infty$ and the values x_n are bounded. Then

$$P_n(m_1, m_2) = P_n(m_1 \le m \le m_2) \approx \frac{1}{\sqrt{2\pi}} \int_{a_n}^{b_n} e^{-\frac{x^2}{2}} dx.$$

Remark.
1. Let us define

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt, \Phi(-\infty) = 0, \quad \Phi(\infty) = 1, \Phi(0) = \frac{1}{2}, \Phi(-x) = 1 - \Phi(x),$$

This function is called the cumulative distribution function of the normal distribution, and its values can be easily found in a table. Then

$$P_n(m_1, m_2) \approx \Phi(b_n) - \Phi(a_n).$$

2. Let us define

$$\Phi_0(x) = \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^x e^{-\frac{t^2}{2}} dt, \quad \Phi_0(0) = 0, \Phi_0(\infty) = 1, \Phi_0(-x) = -\Phi_0(x)$$

This function is called the Laplace function, and its values can be found in tables or computed using mathematical software. Then the probability from the integral theorem can be rewritten as follows:

$$P_n(m_1, m_2) \approx \frac{1}{2}\left(\Phi_0(b_n) - \Phi_0(a_n)\right)$$

Proof. According to the local Moivre-Laplace theorem,

$$P_n(m) = \frac{1}{\sqrt{2\pi}\sqrt{npq}} e^{-x_n^2(m)/2}(1 + o_m(1)),$$

where $x_n(m) = \frac{m-np}{\sqrt{npq}}$, $\quad m_1 \leqslant m \leqslant m_2$, and $o_m(1)$ is a quantity that tends to zero as $n \to \infty$. The quantity $o_m(1)$ depends on m. Note $a_n = x_n(m_1)$; $\quad b_n = x_n(m_2)$

When proving Theorem 2, one could establish that the uniform asymptotic relation holds:

$$\lim_{n\to\infty} \sup_{m_1 \leqslant m \leqslant m_2} |o_m(1)| = 0.$$

This would lead to a significant complication of calculations, so we accept this relation without proof.

According to the definition:

$$P_n(m_1, m_2) = \sum_{m=m_1}^{m_2} P_n(m) = \sum_{m=m_1}^{m_2} \frac{1}{\sqrt{2\pi}\sqrt{npq}} e^{-x_n^2(m)/2}(1 + o_m(1)).$$

Let us denote $\Delta x_n(m) = x_n(m+1) - x_n(m)$. Then $\Delta x_n(m) = \frac{1}{\sqrt{npq}}$, and the probability $P_n(m_1, m_2)$ can be expressed as

$$P_n(m_1, m_2) = \sum_{m=m_1}^{m_2} P_n(m) = \sum_{m=m_1}^{m_2} \frac{1}{\sqrt{2\pi}\sqrt{npq}} e^{-x_n^2(m)/2}(1 + o_m(1)) \sim \frac{1}{\sqrt{2\pi}} \int\limits_{x_n(m_1)}^{x_n(m_2)} e^{-x^2/2} \Delta x_n(m)(1 + o_m(1)).$$

This follows from the definition of the integral and the following expression

$$\left|P_n(m_1, m_2) - \frac{1}{\sqrt{2\pi}} \sum_{m_1}^{m_2} e^{-\frac{x_n(m)^2}{2}} \Delta x_n(m)\right| \leq \sup_{m_1 \leq m \leq m_2} |o_m(1)| \frac{1}{\sqrt{2\pi}} \sum_{m=m_1}^{m_2} e^{-\frac{x_n(m)^2}{2}} \Delta x_n(m);$$

$$\frac{1}{\sqrt{2\pi}} \sum_{m=m_1}^{m_2} e^{-\frac{x_n(m)^2}{2}} \Delta x_n(m) = \frac{1}{\sqrt{2\pi}} \int\limits_{a_n}^{b_n} e^{-\frac{x^2}{2}} dx + o(1).$$

Then

$$\left|P_n(m_1, m_2) - \frac{1}{\sqrt{2\pi}} \int\limits_{a_n}^{b_n} e^{-\frac{x^2}{2}} dx - o(1)\right| \leq o(1) \left(\frac{1}{\sqrt{2\pi}} \int\limits_{a_n}^{b_n} e^{-\frac{x^2}{2}} dx + o(1) \right).$$

Problems.

1. A fair coin is flipped 400 times. What is the probability that heads will appear at least 180 times? Between 190 and 210 times? No more than 200 times?

Answer

$P_{400}(\geq 180) \approx \Phi(20) - \Phi(-2) = \Phi(2) \approx 0.977$;
$P_{400}(190, 210) \approx 2\Phi(1) - 1 \approx 0.683$;
$P_{400}(\leq 200) \approx \Phi(0) - \Phi(-20) \approx 0.5$.

2. Out of 24,000 city lamps, each will remain functional for a year with a probability of 0.6. What is the probability that by the end of the year, between 14,300 and 14,500 lamps will still be functioning?
Answer

$$P_{24000}(14300, 14500) \approx 2\,\Phi\left(\frac{100}{24\sqrt{10}}\right) - 1 \approx 0.812.$$

3. A small town is visited by 100 tourists daily, who go out for lunch. Each tourist chooses one of two restaurants with equal probability and independently of each other. The owner of one of the restaurants wants to ensure that with a probability of approximately 0.99, all visitors to his restaurant can dine there simultaneously. How many seats should be available in the restaurant?
Answer. At least 62.
4. The probability that a single component fails within a year is 0.2. What is the probability that exactly 15 out of 100 components will fail within the year?
Answer. $P_100(15) \approx 0.046$.
5. A fair four-sided dice (numbered 1 to 4) is rolled 500 times. a) What is the probability that an even number appears exactly 225 times? 250 times? That the number of even outcomes falls between 200 and 300? At least 220 times? At most 240 times?
b) What is the probability that a multiple of 3 appears exactly 120 times? 125 times? 130 times? That it appears between 100 and 150 times? At least 110 times? At most 140 times?
answers.
a)

$$\begin{aligned}
&P_{500}(225) \approx 0.003;\\
&P_{500}(250) \approx 0.036;\\
&P_{500}(200, 300) \approx 1,\\
&P_{500}(\geq 220) \approx 0.9963,\\
&P_{500}(\leq 240) \approx 0.1867.
\end{aligned}$$

b)

$$\begin{aligned}
&P_{500}(120) \approx 0.0385\\
&P_{500}(125) \approx 0.0414\\
&P(_{500}(130) \approx 0.0385\\
&P_{500}(100, 150) \approx 0.9902\\
&P_{500}(\geq 110) \approx 0.9394\\
&P_{500}(\leq 140) \approx 0.9394
\end{aligned}$$

Parametrization of the Bernoulli trials.

Let's consider $X_i = 1$ if event A occurs in the i-th trial, and $X_i = 0$ otherwise. Then, the number of successes is given by

$$S_n = \sum_{i=1}^{n} X_i.$$

Definition. The value

$$\mu = \frac{S_n}{n}$$

is called the relative frequency.

Lemma. Show that for all $\epsilon > 0$,

$$P(|\mu - p| \geq \epsilon) \to 0 \quad \text{as} \quad n \to \infty.$$

Proof (Lemma).
I. Note

$$P(|\mu - p| \geq \epsilon) = 1 - P(|\mu - p| < \epsilon) \Rightarrow P(|\mu - p| \geq \epsilon) \to 0 \Leftrightarrow P(|\mu - p| < \epsilon) \to 1. \quad \text{as} \quad n \to \infty.$$

Consider $P(|\mu - p| < \epsilon)$ be the interval M-L theorem we get

$$P(|\mu - p| < \epsilon) = P(S_n \in [n(p-\epsilon), n(p+\epsilon)]) \approx \Phi\left(\frac{n(p+\epsilon) - np}{\sqrt{npq}}\right) - \Phi\left(\frac{n(p-\epsilon) - np}{\sqrt{npq}}\right) \approx 2\Phi\left(\frac{n\epsilon}{\sqrt{npq}}\right) - 1 \to 1$$

as $\quad n \to \infty$, because $\frac{n\epsilon}{\sqrt{npq}} \to +\infty$ as $n \to +\infty$ ans $\Phi(+\infty) = 1$.

The second proof.
The firstly we estimate the probability, that the number of successes wil be more than r:

$$P(S_n \geq r) = \sum_{k=r}^{n} P_n(k) \overset{i=k-r}{=} \sum_{i=0}^{n-r} P_n(r+i) = \sum_{i=0}^{\infty} P_n(r+i),$$

because $P_n(r+i) = 0$, for $i > n - r$.
let's assume $r > np$. Now, remember the lemma about the most probable number:

$$\frac{P_n(r+1)}{P_n(r)} = \frac{(n-r)p}{(r+1)q} \leq \frac{(n-r)p}{rq} = 1 - \frac{(r-np)}{rq} =: d$$

so under the assumption $r > np$: $P_n(r+i) \searrow$ and $P_n(i+r) \leq d^i P_n(r)$, then

$$P(S_n \geq r) = \sum_{i=0}^{\infty} P_n(r+i) \leq \sum_{i=0}^{\infty} d^i P_n(r) = P_n(r)\frac{rq}{r-np} \leq \frac{rq}{(r-np)^2}, \quad r > np.$$

The last inequality follow from there exist more than $r - np$ integer k, such that $m \leq k \leq r$ $(np - q \leq m \leq np + p < np + 1 \leq r)$. The sum of the corresponding terms of the binomial probability less than one, and each of them is at least $P_n(r)$. Therefore, $P_n(r)$ is not greater than $(r - np)^{-1}$.
Similarly

$$P(S_n \leq r) \leq \frac{(n-r)p}{(np-r)^2}, \quad \text{for} \quad r < np.$$

Indeed, the event that no more than r successes have occurred is equivalent to at least nr trials ending in failure. Applying the previous inequality to failures, we obtain the desired result.
Now proof the Law of Large number by using the inequalities.

$$P(|\mu - p| \geq \epsilon) = P(S_n \geq n(p+\epsilon)) + P(S_n \leq n(p-\epsilon)) \leq \frac{2}{n\epsilon^2} \to 0 \quad \text{as} \quad n \to \infty.$$

Problem.

1. For an experimental verification of the Law of Large Numbers, Buffon conducted the following experiment: A coin was tossed 4040 times, and heads appeared 2048 times.
a) Find the probability that, if the experiment is repeated, the absolute deviation of the relative frequency from

the probability 1/2 does not exceed the obtained result.
b) Find the practical upper bound of the possible deviation of the relative frequency from the probability, assuming that an event with a probability of 0.999 is practically certain.
Answer. a) The relative frequency of the first experiment $\mu_I = \frac{2048}{4040}$ and the difference between μ_I and p $= \frac{2048}{4040} - \frac{1}{2} = \frac{28}{4040}$ then we need to calculate

$$P\left(|\mu - p| \leq \frac{28}{4040}\right) = P(S_n \in [1992, 2048]) = P_n(1992, 2048) \approx 0.622.$$

b) Find k such as $P(|\mu - p| \leq k) = 0.999$.

$$P(|\mu-p| \leq k) = P(|S_n-2020| \leq 4040{\cdot}k) = P(S_n \in [2020-4040k, 2020+4040k]) = P_n(2020-4040k, 2020+4040k)$$

$$\approx 2\Phi\left(\frac{4040k}{\sqrt{1010}}\right) - 1 = 0.999 \Leftrightarrow \frac{4040k}{\sqrt{1010}} = \Phi^{-1}(0.9995) \approx 3.29 \Leftrightarrow k \approx 0.026; \quad 4040k \approx 104.56$$

That means $0.999 = P(S_n \in [1915.44; 2124.56]) \leq P(S_n \in [1915, 2125])$.

2. A dice is rolled 80 times. Find the approximate bounds within which the number of times a six appears will be contained with a probability of 0.9973.
3. In a batch of 768 watermelons, each watermelon is unripe with a probability of 1/4. Find the probability that the number of ripe watermelons falls within the range of 564 to 600.
4. The probability of finding a white mushroom among others is 1/4. What is the probability that out of 300 mushrooms, 75 will be white?
5. A factory shipped 5000 light bulbs to a store. The probability that a bulb will break during transportation is 0.0002. Find the probability that no more than three broken bulbs arrive at the store.

§5. The concept of a random variable.

A probability space is a mathematical structure that consists of three components: $(\Omega, \mathcal{F}, P)$, where:

- Ω is the sample space (the set of all possible outcomes),
- $\mathcal{F}$ is a sigma-algebra (a collection of subsets of Ω that contains the empty set and is closed under complement and countable unions),
- $\mathbb{P}$ is the probability measure, which assigns probabilities to events in $\mathcal{F}$.

Reminder.

Definition. A collection of subsets $\mathfrak{F}$ of a set Ω is a σ-algebra, if:
1. $\Omega \in \mathfrak{F}$ and $\emptyset \in \mathfrak{F}$;
2. If $A \in \mathfrak{F}$ then its complement $\overline{A} \in \mathfrak{F}$ as well;
3. If $A_1, A_2, A_3 ... \in \mathfrak{F}$ then $\cup_{i=1}^{\infty} A_i \in \mathfrak{F}$. These three conditions are sufficient to define the structure of a σ-algebra.
Examples.
1. $\mathfrak{F} = \{\emptyset, \Omega\}$;
2. $\mathfrak{F}_A = \{\emptyset, \Omega, A, \overline{A}\}$;
3. $\mathfrak{F}_{A,B} =?$; $A \cap B = \emptyset$;
4. $\mathfrak{F}_{A,B} =?$; $A \cap B \neq \emptyset$;

Definition. A random variable is a function that assigns a real number to an outcome.

$$X : (\Omega, \mathfrak{F}) \to (\mathbb{R}, \mathbb{B}),$$

which is measurable, i.e., $\forall B \in \mathbb{B} \quad X^{-1}(B) = \{\omega \mid X(\omega) \in B\} \subset \mathfrak{F}$, where $\mathbb{B}$ is the Borel σ-algebra.
Example. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $\xi(\omega) = \omega, \quad \eta(\omega) = \omega^2$.
Will these random variables be measurable under the following σ-algebras?
a) The maximal σ-algebra = the set of all possible subsets;
b) The trivial σ-algebra $=\{\Omega, \emptyset\}$;
c) The σ-algebra$=\{\Omega, \emptyset, \{1, 3, 5\}, \{2, 4, 6\}\}$.
Which random variables can be measurable with respect to the σ-algebras from (b) and (c).

How to choose the best σ-algebra for a given random variable?
Definition The sigma-algebra generated by a random variable X is the smallest σ-algebra with respect to which X is a measurable function or the smallest σ-algebra that contains the preimage of every Borel set under X. Formally, it is defined as:

$$\sigma(X) = \{X^{-1}(B) \mid B \in \mathcal{B}(\mathbb{R})\}$$

where $X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\}$ is the preimage of a Borel set B under the random variable X.
Example. Let $\Omega = [-1, 1], \mathfrak{F} = \mathbb{B}([-1, 1]), \mathbb{P} = \frac{1}{2}$. Construct the σ-algebra generated by random variables $\xi(\omega) = \omega, \quad \eta(\omega) = |\omega|, \quad \zeta(\omega) = \max(0, |\omega - \frac{1}{4}| - \frac{1}{2})$.

The distribution and the distribution function of the random variables.

Definition. The distribution of a random variable X is the probability measure P_X on $(\mathbb{R}, \mathbb{B})$, which can be given as

$$P_X(B) = \mathbb{P}(\{\omega : \ X(\omega) \in B\}).$$

Definition. Random variables X,Y are identically distributed ($X \overset{d}{=} Y$) if $P_X = P_Y$.
Note. X and Y can be given on the different probability spaces (Ω and Ω').
Distribution function or cumulative distribution function.

$$F_\xi(x) = P_\xi((-\infty, x)) = \mathbb{P}(\{\omega : \xi(\omega) < x\}) = P(\xi < x).$$

Continuity of probability.
Let there be a system of events such that $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$
Then,

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

And if $B_1 \supseteq B_2 \supseteq B_3 \supseteq \dots$, then

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} B_i\right) = \lim_{i\to\infty} \mathbb{P}(B_i).$$

We represent A_n as a union of disjoint sets:

$$A_n = (A_n \setminus A_{n-1}) \cup (A_{n-1} \setminus A_{n-2}) \cup \dots \cup (A_2 \setminus A_1) \cup A_1$$

Thus,

$$\mathbb{P}(A_n) = \mathbb{P}(A_n \setminus A_{n-1}) + \cdots + \mathbb{P}(A_1).$$

$$\bigcup_{i=1}^{\infty} A_i = \left(\bigcup_{j=2}^{\infty}(A_j \setminus A_{j-1})\right) \cup A_1$$

$$= \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(A_i \setminus A_{i-1}) = \lim_{n\to\infty} A_n,$$

where

$$A_0 = \emptyset.$$

To prove the second part, we transition to complements:

$$B_i = \overline{A_i}$$

$$\bigcap_{i=1}^{\infty} B_i = \overline{\bigcup_i \overline{B_i}}.$$

Note. The bijection function exists between $F_X \leftrightarrow P_x$ on $(\mathbb{R}, \mathbb{B})$.
1. $\bullet P_X([a,b)] = F_x(b) - F_x(a)$.
2. $\bullet P_X([a,b]) = P_X\left(\cap_{c>b}[a,c)\right) = \lim\limits_{c\to b+0} P_x([a,c)) = F_x(b+0) - F_x(a)$.
3. $\bullet P_X((a,b)) = F_x(b) - F_x(a-0)$;
4. $\bullet P_X((a,b]) = F_x(b+0) - F_x(a-0)$.

Properties of a distribution function.

1. $F_\xi(\cdot) \nearrow$
2. F_ξ is left-continuous and it has the right limit.
3. $\lim\limits_{x\to+\infty} F_\xi(x) = 1 \qquad \lim\limits_{x\to-\infty} F_\xi(x) = 0.$

From the second property $\lim\limits_{x\to x_0+0} F_\xi(x) = P(\xi \leq x_0) = F_\xi(x) + P(\xi = x_0)$.

Discrete distribution

Ω–finite or countable $=\{\omega_1, \omega_2, \omega_3, ...\}$.
Let ξ be a random variable with the values $\{x_i\}_i \in \mathbb{R}$, $p_i = P(\xi = x_i)$, $\quad \sum_i p_i = 1$.
The distribution function will be the following

ξ	x_1	x_2	...	x_k
$\mathbb{P}$	p_1	p_2	...	p_k

$$F_\xi(x) = \mathbb{P}(\xi < x) = \sum_{x_i<x} \mathbb{P}(\xi = x_i) = \sum_{x_i<x} p_i.$$

Example. A fair standard 4-sided die {1,...,4} is rolled 4 times. Let X be the random variable representing the number of values divisible by 3. Construct the probability distribution and the distribution function.

Characterisitics of the discrete random variable.

Definition: The mathematical expectation of a discrete random variable X is defined as the number:

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i \cdot P(X = x_i) = \sum_{i=1}^{n} x_i \cdot p_i.$$

We define the second and the k-th moment of a discrete random variable X as follows:

$$\mathbb{E}(X^2) = \sum_{i=1}^{n} x_i^2 \cdot p_i, \quad m_k = \mathbb{E}(X^k) = \sum_{i=1}^{n} x_i^k \cdot p_i.$$

Definition: The variance of a discrete random variable X is defined as the number:

$$\mathbb{D}(X) = \sum_{i=1}^{n} (x_i - \mathbb{E}(X))^2 \cdot p_i = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = m_2 - (m_1)^2 \geq 0.$$

Remark: Variance is always a non-negative quantity, which follows from Hölder's inequality.
Reminder. For any sequences of numbers a_i and b_i and positive numbers p, q such that $\frac{1}{p} + \frac{1}{q} = 1$, the following inequality holds:

$$\sum_{i=1}^{n} |a_i b_i| \leq \left(\sum_{i=1}^{n} |a_i|^p\right)^{\frac{1}{p}} \cdot \left(\sum_{i=1}^{n} |b_i|^q\right)^{\frac{1}{q}}.$$

Problem Find the E(X) and D(X) from the upper example.

The most famous Discrete distributions.

Let's find the E(X) and D(X) for the following distributions.

1. The Bernoulli distribution (Ber(p)).
$X \in \{0, 1\}, \quad P(X = 0) = q; \quad P(X = 1) = p$

$E(X) = p; D(X) = pq.$

2. The Binomial distribution (Bin(n,p)).
$X \in \{0, 1, 2, ..., n\}, \quad P(X = k) = c(n, k)p^k(1 - p)^{n-k}, k = 0, 1, 2, ...n.$

$EX = np, \quad DX = npq.$

3. The Poisson distribution (Pois(λ)).
$X \in \mathbb{N} \cup \{0\}, \quad P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, 2, ...$

$E(X) = \lambda, \quad D(X) = \lambda.$

4. The discrete Uniform distribution (Unif{1,...,n}).
$X \in \{1, ..., n\}, p = n^{-1}.$

$E(X) = \frac{n+1}{2}, \quad D(X) = \frac{n^2+1}{12};$

5. The geometry distribution of the I type ($Geom_1(p)$).

X is the number of trials until the first success (including it).
$X \in \mathbb{N}, \quad P(X = k) = p \cdot (1-p)^{k-1}, k = 1, 2, 3...$
$E(X) = \frac{1}{p}, \quad D(X) = \frac{q}{p^2}.$

6. The geometry distribution of the II type.($Geom_2(p)$)
X is the number of failures before the first success (i.e., not including it).
$X \in \mathbb{N} \cup \{0\}, \quad P(X = k) = p \cdot (1-p)^{k}, k = 0, 1, 2, 3...$

$E(X) = \frac{q}{p}, \quad D(X) = \frac{q}{p^2}.$

Problems.

A die is rolled n times.
a) Compute the expected number of points and variance for one roll.
b) For two rolls.
c) For 100 rolls.
d) Find the expected number of fives and their variance (quadratic deviation from the mean) for 1200 rolls.

Absolutely continuous distribution

Ω is continuous, for example $\mathbb{R}$; $[0,1]$; $[a,b]$...
Now $P(\xi = k) = 0$

Definition. Probability density function

of a random variable ξ is a non-negative function $f(y)$ such that:

$$F_\xi(x) = \int_{-\infty}^{x} f(y)dy.$$

Properties of the probability density function:

1. $\int_{-\infty}^{\infty} f(y)dy = 1;$
2. $f(x) = F'(x)$ if the cumulative distribution function is differentiable;
3. $\mathbb{P}(a \le \xi < b) = \int_a^b f(y)dy = F(b) - F(a).$

Expectation and Variance

$$\mathbb{E}(\xi) = \int_{-\infty}^{\infty} x \cdot f(x)dx; \quad \mathbb{D}(\xi) = \mathbb{E}(\xi - \mathbb{E}(\xi))^2,$$

where $\mathbb{E}(\xi^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx.$

Remark. For a function of a random variable $g(\xi)$:

$$\mathbb{E}(g(\xi)) = \int_{-\infty}^{\infty} g(x) \cdot f(x)dx$$

k-th moment: $g(x) = x^k \quad m_k = \mathbb{E}(X^k) = \int\limits_{-\infty}^{\infty} x^k \cdot f(x)dx$

Examples of Continuous Distributions

1. **Uniform Distribution on [0,1] = Unif[0,1]**

$$f(x) = 1 \cdot \mathrm{Ind}_x([0,1]);$$

Find the cumulative distribution function and compute $E(X)(=\frac{1}{2})$, $D(X)(=\frac{1}{12})$. For $f(x) = \frac{1}{b-a} \cdot \mathrm{Ind}_x([a,b])$, the density of the uniform law Unif[a,b], find the same values.
$F(X) = 0, x < 0; \qquad \frac{x-a}{b-a}, x \in [a,b]; \qquad 1, x > b.$
$E(X) = \frac{a+b}{2}; D(X) = \frac{(b-a)^2}{12}.$

2. **Standard Normal Distribution Norm(0,1)**

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}, \quad F(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

Find expectation and variance.

3. **Exponential Distribution** $Exp(\lambda)$ Given the cumulative distribution function:

$$F(x) = 1 - e^{-\lambda \cdot x}, \quad x \geq 0, \quad \lambda > 0.$$

Find its probability density function and expectation.

$$f(x) = F'(x) = \lambda e^{-\lambda \cdot x}; \quad E(X) = \int\limits_{0}^{\infty} x f(x)dx = \frac{1}{\lambda}, \quad D(X) = \frac{1}{\lambda^2}.$$

4. **Standard Cauchy Distribution Cauchy(0,1)** The probability density function is given by:

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}.$$

Find the cumulative distribution function.

Problem 1. A random variable X has the probability density function:

$$f(x) = \begin{cases} 2x - 4, & \text{if } 2 \leq x \leq 3, \\ 0, & \text{otherwise} \end{cases}$$

Find the cumulative distribution function, compute $P(2.5 < x \leq 3.5)$, and evaluate $E(X)$, $D(X)$, and $E(X^3)$.

Problem 2. Given the probability density function:

$$f(x) = \begin{cases} 0, & \text{if } x < 0, \\ ax^2, & \text{if } 0 \leq x \leq 2, \\ 0, & \text{if } x > 2 \end{cases}$$

Find a, the cumulative distribution function $F(X)$, the expectation, and the following probabilities:

$$P(X \geq 1); \quad P(0.5 < X \leq 1.5); \quad P(\frac{1}{3} \leq X < 4); \quad P(-2 \leq X \leq \frac{3}{4})$$

Problem 3. A random variable X has the probability density function:

$$f(x) = \begin{cases} 0, & \text{if } x < 0, \\ ax, & \text{if } 0 \leq x \leq 1, \\ \frac{b}{x^2}, & \text{if } 1 \leq x \leq 2, \\ 0, & \text{if } x > 2 \end{cases}$$

Find a, b, if
a) $F(1) = \frac{1}{2}$,
b) E(X)=1.
Find the cdf F(X).

§6. The function of a random variable and its distribution.

Let $(\Omega, \mathfrak{F}, \mathbb{P})$ - a probability space, and ξ is a random varible on this space. Consider the new variable $\eta := g(\xi)$, where $g(\cdot) : \mathbb{R} \to \mathbb{R}$.
Try to answer on the following questions:
1. Will be η a random variable?
2. If it is a random variable, how to find its distribution?

Consider the following example.
Let ξ be a random variable with the following distribution:

ξ	0	1	2	3	4
$\mathbb{P}$	0.1	0.15	p	0.2	0.3

Find p and the following variables $2\xi - 3$, $\xi^2 - 2$, $(\xi - 1)^2$ and $(\xi - 2)^2$.

Return to our questions and give an answer on the first.
Theorem 1. Let ξ be a random variable, and let $g : \mathbb{R} \to \mathbb{R}$ be a Borel (Borel-measurable) function, i.e., such that for every Borel set B, its preimage $g^{-1}(B)$ is again a Borel set. Then $g(\xi)$ is a random variable.
Proof. We check that the preimage of any Borel set under the mapping $g(\xi) : \Omega \to \mathbb{R}$ is an event. Take an arbitrary $B \in \mathcal{B}(\mathbb{R})$ and define $B_1 = g^{-1}(B)$. The set B_1 is Borel since the function g is Borel-measurable. We find $(g(\xi))^{-1}(B)$:

$$\{\omega \mid g(\xi(\omega)) \in B\} = \{\omega \mid \xi(\omega) \in g^{-1}(B)\} = \xi^{-1}(B_1) \in \mathcal{F}$$

since $B_1 \in \mathcal{B}(\mathbb{R})$ and ξ is a random variable. □
Remark. All commonly used functions are Borel-measurable.
II. Let the random variable ξ have a cumulative distribution function $F_\xi(x)$ and a probability density function $f_\xi(x)$. Using a Borel function $g : \mathbb{R} \to \mathbb{R}$, we define a new random variable $\eta = g(\xi)$. The task is to determine the cumulative distribution function of η and, if it exists, the probability density function of η.
Remark/ The probability density function of the random variable $\eta = g(\xi)$ does not always exist for arbitrary functions g. For example, if g is a piecewise constant function, then η has a discrete distribution, and its probability density function does not exist.
Exercise. Provide an example of a probability density function of a random variable ξ and a continuous function g such that $\eta = g(\xi)$ has:

- a discrete distribution;
- a non-degenerate discrete distribution.

Theorem 2. Let ξ have a cumulative distribution function $F_\xi(x)$ and a probability density function $f_\xi(x)$, with a constant $a \neq 0$. Then the random variable $\eta = a\xi + b$ has the probability density function

$$f_\eta(x) = \frac{1}{|a|} f_\xi\left(\frac{x-b}{a}\right).$$

Proof. First, consider the case when $a > 0$:

$$F_\eta(x) = P(a\xi + b < x) = P\left(\xi < \frac{x-b}{a}\right) = F_\xi\left(\frac{x-b}{a}\right) = \int\limits_{-\infty}^{(x-b)/a} f_\xi(t)\,dt.$$

We perform a change of variable in the last integral. Let t be replaced by the new variable y defined as $t = (y-b)/a$. Then $dt = dy/a$, the upper limit of integration $t = (x-b)/a$ transforms into $y = x$, and the lower limit $t = -\infty$ transforms into $y = -\infty$. This yields

$$F_\eta(x) = \int\limits_{-\infty}^{x} \frac{1}{a} f_\xi\left(\frac{y-b}{a}\right) dy.$$

The function inside the integral is the probability density function $f_\eta(y)$ of the random variable $\eta = a\xi + b$ for $a > 0$.

Now, consider the case when $a < 0$:

$$F_\eta(x) = P(a\xi + b < x) = P\left(\xi > \frac{x-b}{a}\right) = \int\limits_{(x-b)/a}^{+\infty} f_\xi(t)\,dt.$$

We apply the same change of variable $t = (y-b)/a$, where $y = at + b$. However, now the upper limit $t = +\infty$ transforms into $y = -\infty$ since $a < 0$. This results in

$$F_\eta(x) = \int\limits_{x}^{+\infty} \frac{1}{|a|} f_\xi\left(\frac{y-b}{a}\right) dy = \int\limits_{-\infty}^{x} \frac{1}{|a|} f_\xi\left(\frac{y-b}{a}\right) dy.$$

The function inside the integral is the probability density function $f_\eta(y)$ of the random variable η.

Theorem 3. Let ξ have a probability density function $f_\xi(x)$, and let $g : \mathbb{R} \to \mathbb{R}$ be a monotonic function. Then the random variable $\eta = g(\xi)$ has a probability density function given by

$$f_\eta(x) = \left|\left(g^{-1}(x)\right)'\right| f_\xi\left(g^{-1}(x)\right).$$

Here, g^{-1} is the inverse function of g, and $(g^{-1}(x))'$ is its derivative.

Proof. Similarly to the proof of the second theorem.

Remark.

1. If the function g is a piecewise monotonic function, for example, x^2 for $x \in [-1, 1]$, then the formula from Theorem 3 can be written as follows:

$$f_\eta(y) = \sum_{i=1}^{n} f_\xi(g_i^{-1}(y)) \cdot |(g_i^{-1}(y))'| = \sum_{i=1}^{n} \frac{f_\xi(g_i^{-1}(y))}{|g'(g_i^{-1}(y))|}$$

where $g_i^{-1}(y)$ are the values corresponding to a given y, that is, the points x_i such that $g(x_i) = y$.
2. If $suppr(f_\xi) = A$ then $f_\eta(x) = ..., x \in g(A)$.

From Theorem 2, we obtain the following familiar results:
Corollary 1. If $\xi \sim \mathcal{N}orm(0, 1)$, then $\eta = \sigma\xi + a \sim \mathcal{N}orm(a, \sigma^2)$.
Proof. Indeed,

$$f_\eta(x) = \frac{1}{\sigma} f_\xi\left(\frac{x-a}{\sigma}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Corollary 2. If $\xi \sim \mathcal{N}orm(a, \sigma^2)$, then $\frac{\xi - a}{\sigma} \sim \mathcal{N}orm(0, 1)$.
Corollary 3. If $\xi \sim Unif[0, 1]$, then $a\xi + b \sim Unif[b, a + b]$ for $a > 0$.
Corollary 4. If $\xi \sim Exp(\alpha)$, then $\alpha\xi \sim Exp(1)$.

Problems.

Find the distribution of the following random variables:
a) $Y = \exp(-X), \quad X \sim Exp(\lambda)$;
b) $Y = X^{-1}, \quad X \sim Cauchy(0, 1)$;
c) $Y = -\ln(1 - X), \quad X \sim Unif[0, 1]$;
d)$Y = X^2, \quad X \sim Unif[-1, 1]$;
e) $Y = (X - 1)^2, \quad X \sim Unif[0, 3]$;
f)* $Y = F_\xi(X)$, F_ξ is a continuous distribution function of a random variable ξ.

Quantile Transformation.

Theorem 1. *Let the cumulative distribution function (CDF) $F(x) = F_\xi(x)$ be continuous. Then the random variable $\eta = F(\xi)$ has a uniform distribution on the interval $[0, 1]$.*

Proof. Let's find the CDF of the random variable η. Note that $0 \le \eta \le 1$. Suppose first that the function F is strictly increasing. Then it has an inverse, and therefore:

$$F_\eta(x) = P(F(\xi) < x) = \begin{cases} 0, & \text{if } x \le 0, \\ P(\xi < F^{-1}(x)), & \text{if } 0 < x < 1, \\ 1, & \text{if } x \ge 1. \end{cases}$$

However, we have $P(\xi < F^{-1}(x)) = F(F^{-1}(x)) = x$, i.e., $\eta \sim U_{0,1}$.

If the function F is not strictly increasing, it may have flat segments. In this case, we denote $F^{-1}(x)$ as the leftmost point in the closed set $\{t \mid F(t) = x\}$, the preimage of the point $x \in (0, 1)$. With this definition of the *inverse* function, the expressions remain valid along with the equality $P(\xi < F^{-1}(x)) = F(F^{-1}(x)) = x$ for any $x \in (0, 1)$.

Theorem 1 can be used to generate random variables with a specified distribution using a uniformly distributed random variable (e.g., from a random number generator). The following statement holds not only for continuous functions but also for any distribution function F. Let $F^{-1}(x)$ be defined as the infimum of the set of points t for which $F(t) \ge x$:

$$F^{-1}(x) = \inf\{t \mid F(t) \ge x\}.$$

For a continuous function F, this definition of the *inverse function* coincides with the one introduced in the proof of Theorem 1.

Theorem 2. *Let $\eta \sim U_{0,1}$, and let F be an arbitrary distribution function. Then the random variable $\xi = F^{-1}(\eta)$ (quantile transformation applied to η) has the distribution function F.*

Corollary 1. *Let $\eta \sim U_{0,1}$. The following relations hold:*

$$-\frac{1}{a}\ln(1-\eta) \in E_a,$$

$$a + \sigma\tan(\pi\eta - \pi/2) \in C_{a,\sigma},$$

$$\Phi^{-1}_{\sigma,1}(\eta) \in N_{\sigma,1}.$$

Problems.
How to construct a random variable with the following df:
a)

$$F(x) = \frac{e^{x-a}}{1+e^{x-a}}, a \in \mathbb{R}, x \in \mathbb{R}.$$

b)

$$F(X) = x^k, k > 1, x \in [0,1].$$

Characterisitcs of random variables.

1. **Raw moment of order k**
The raw moment of order k of a random variable X is defined as the expected value of its k-th power:

$$a_k = \mathbb{E}[X^k].$$

For a discrete random variable with probability mass function $p(x)$:

$$a_k = \sum_x x^k p(x).$$

For a continuous random variable with probability density function $f(x)$:

$$a_k = \int_{-\infty}^{\infty} x^k f(x)dx.$$

2. **Central moment of order k**
The central moment of order k of a random variable X is defined as:

$$\mu_k = \mathbb{E}[(X - \mathbb{E}[X])^k].$$

For a discrete random variable:

$$\mu_k = \sum_x (x - \mathbb{E}[X])^k p(x).$$

For a continuous random variable:

$$\mu_k = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^k f(x)dx.$$

- $\sigma = \sqrt{D(X)}$– standart deviation.

3. **Mode**
The mode of a random variable X is the value at which its probability mass function (for discrete variables) or probability density function (for continuous variables) reaches a maximum:

$$\text{mode}(X) = \arg\max_x P(X = x) \quad \text{(discrete)},$$

$$\text{mode}(X) = \arg\max_x f(x) \quad \text{(continuous)}.$$

4. **Quantile of level p**

The quantile of level p of a random variable X is defined as:

$$Q_p = F^{-1}(p) = \inf\{x \mid F(x) \geq p\},$$

where $F(x)$ is the cumulative distribution function (CDF) of X.

5. **Skewness coefficient**

The skewness coefficient measures the asymmetry of a probability distribution:

$$\gamma_1 = \frac{\mathbb{E}[(X - \mathbb{E}[X])^3]}{\sigma^3} = \frac{\mu_3}{\sigma^3}.$$

- If $\gamma_1 > 0$, the distribution is skewed to the right (longer right tail).
- If $\gamma_1 < 0$, the distribution is skewed to the left (longer left tail).

6. **Excess kurtosis**

The excess kurtosis measures the peakedness of a probability distribution:

$$\gamma_2 = \frac{\mathbb{E}[(X - \mathbb{E}[X])^4]}{\sigma^4} - 3 = \frac{\mu_4}{\sigma^4} - 3.$$

- If $\gamma_2 = 0$, the distribution has the same kurtosis as the normal distribution (mesokurtic).
- If $\gamma_2 > 0$, the distribution is more peaked (leptokurtic).
- If $\gamma_2 < 0$, the distribution is flatter (platykurtic).

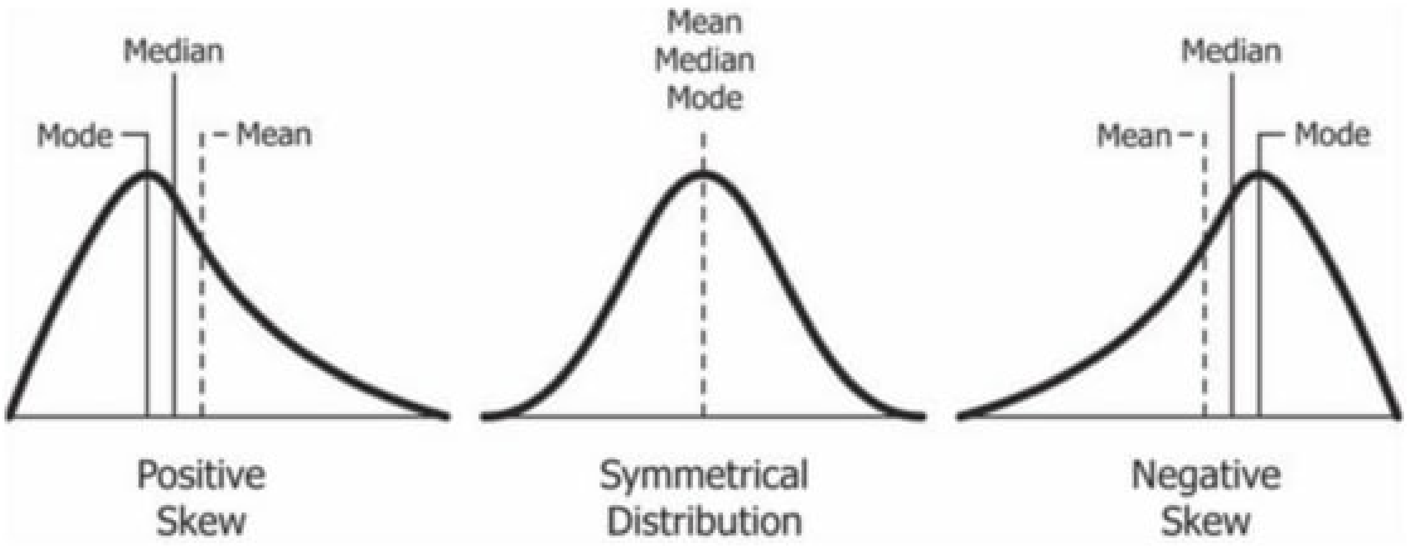

Figure 5: Skewness

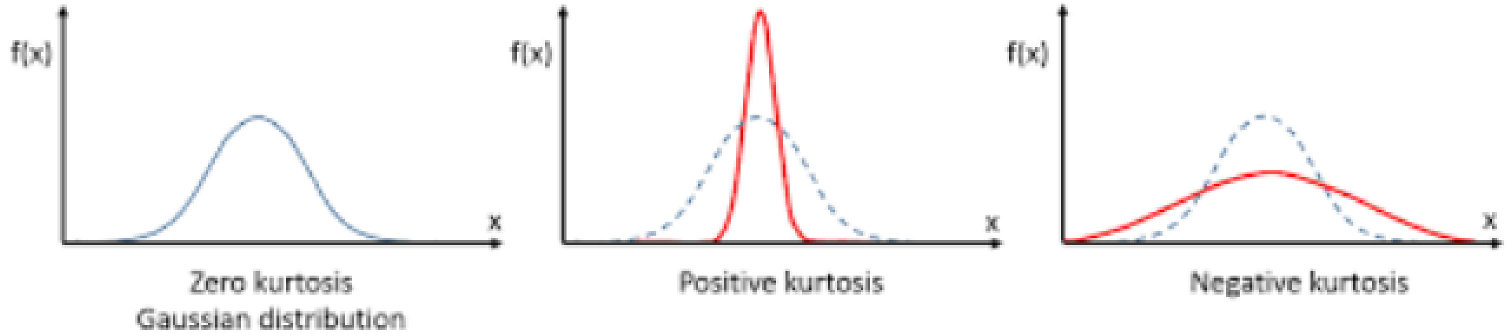

Figure 6: Kurtosis

7. Probability-generating function

The probability-generating function for a discrete random variable X taking non-negative integer values is defined as:

$$G_X(s) = \mathbb{E}[s^X] = \sum_{k=0}^{\infty} P(X = k)s^k, \quad \text{for } s \in [0, 1].$$

This function is useful for finding probabilities, moments, and analyzing the properties of the distribution.
Properties.
1. $G_x(1) = 1, \quad G_x(0) = P(X = 0)$.
2. $G_x(s) \nearrow s \in [0, 1]$.
3. Convex.
4. if $E(X) < \infty \Leftrightarrow G_x'(1) < \infty$ and $E(X) = G_x'(1)$. (w/p)

Problems. Find this characteristics for the following distributions: $Ber(p), Bin(n, p), DUnif(1, ..., n), Pois(\lambda)$.

8. Characteristic function

The characteristic function of a random variable X is defined as the expectation of the complex exponential:

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} f_X(x)dx, \quad t \in \mathbb{R}.$$

It is used for studying the distribution of a random variable and plays a key role in probability theory, especially in the central limit theorem.
definition. Complex random variables are $X = X_1 + i\,X_2 : \Omega \to \mathbb{C}$, where X_1, X_2- are real-number random variables.
properies.
1. $t \in \mathbb{R}$, $X \in \mathbb{R} \Rightarrow |e^{itX}| < 1 \Rightarrow$ mathematical expectation is exist. So the characteristic function is defined for all $t \in \mathbb{R}$.
2. $\phi_x(0) = 1, |\phi_x(t)| \leq 1$.

3. $\phi_x(-t) = \overline{\phi_x(t)}$.
4. $\phi_x(t) \in \mathbb{R}\, \forall t \Leftrightarrow$ Distribution of X is symmetric $P_X = P_{-X}$. (w/p)
5. $\phi_{aX+b}(t) = e^{itb}\phi_x(at)$.
6. Uniform continuity of $\varphi_X(t)$ on $\mathbb{R}$

$$\forall \varepsilon > 0 \quad \exists \delta > 0 : \quad |h| < \delta \implies |\varphi(t + h) - \varphi(t)| < \varepsilon \quad \forall t.$$

or

$$\sup_{t\in\mathbb{R}} |\varphi(t+h) - \varphi(t)| \to 0.$$

7. **Non-negative definiteness.**

$$\forall n \in \mathbb{N} \quad \forall t_1, \dots, t_n \quad \forall c_1, \dots, c_n \in \mathbb{C}$$

$$\sum_{j,k=1}^{n} c_j \varphi(t_j - t_k) \overline{c_k} \geq 0.$$

Proof. To prove this, let's expand the sum:

$$\sum_{j,k=1}^{n} c_j \varphi(t_j - t_k) \overline{c_k} = \sum_{j,k=1}^{n} c_j \cdot \mathbb{E} e^{it_jX - it_kX} \cdot \overline{c_k} =$$

$$= \mathbb{E} \sum_{j,k=1}^{n} c_j \cdot e^{it_jX} \cdot \overline{c_k e^{it_kX}}.$$

By the linearity of expectation,

$$\mathbb{E} \sum_{j,k=1}^{n} c_j \cdot e^{it_jX} \cdot \overline{c_k e^{it_kX}} = \mathbb{E} \left(\sum_{j=1}^{n} c_j \cdot e^{it_jX} \right) \overline{\left(\sum_{k=1}^{n} c_k \cdot e^{it_kX} \right)} =$$

$$= \mathbb{E} \left| \sum_{j=1}^{n} c_j \cdot e^{it_jX} \right|^2 \geq 0.$$

Theorem. Bochner-Khinchin

Any continuous **non-negative definite** function $\varphi : \mathbb{R} \to \mathbb{C}$, such that

$$\varphi(0) = 1,$$

is the characteristic function of some random variable.

Problems. Find this characteristics for Unif[a,b], Unif[-a,a], Pois(λ) Exp(1), Exp(λ) and Norm(0,1).

Characteristic function and the moments.

If we consider the first derivatives of a characteristic function, we can see:
$\phi'(0) = iE(X)$;
$\phi''(0) = -E(X^2)$;
....
$\phi^{(k)}(t) = i^k E(X^k e^{itX})$; $\qquad \phi^{(k)}(0) = i^k E(X^k)$.
Remark. This relationship holds as long as the n-th moment $E(X^n)$ exists.
Proposition. Let the first n-th moments exist, i.e. $E(|X|^n) < \infty, \quad n \in \mathbb{N}$. Then

$$\forall k = 1, ..., n \qquad \phi^{(k)}(t) = i^k E\left(X^k \, e^{itX}\right);$$

and

$$E(X^k) = (-i)^k \phi^{(k)}(0).$$

Proof. by Induction.
Base is obvious.
Inductive Step:
Assume it holds for $k < n$. Let us check it for $k+1$. We find the $(k+1)$-th derivative:

$$\frac{\varphi_X^{(k)}(t+h) - \varphi_X^{(k)}(t)}{h} = \frac{i^k}{h}\left(\mathbb{E}X^k e^{i(t+h)X} - \mathbb{E}X^k e^{itX}\right) = \frac{i^k}{h}\mathbb{E}\left(X^k\left(e^{i(t+h)X} - e^{itX}\right)\right)$$

We consider the part of this

$$\left|X^k e^{itX}(e^{ihX} - 1)\right| \leq |X|^{k+1} \cdot |h|;$$

It follows from $y \in \mathbb{R}, k \in \mathbb{N} \cup \{0\}$

$$\left|e^{iy} - 1 - iy - \frac{y^2}{2} + \frac{iy^3}{6} - ... + \frac{i^k y^k}{k!}\right| \leq \frac{|y|^{k+1}}{(k+1)!}.$$

This implies the expression is integrable, since $k+1 \leq n$, i.e., it's finite. Hence, we can apply the Lebesgue Dominated Convergence Theorem.

Thus we get:

$$\frac{i^k}{h}\mathbb{E}X^k e^{itX}(e^{ihX} - 1) \xrightarrow[h \to 0]{} i^k \mathbb{E}X^{k+1} e^{itX}$$

So the formula also holds for $k+1$.

The converse direction (if the derivative exists, then the moment exists) is also true, but the conditions there are more subtle.

Proposition (Existence of Derivative Without Existence of Moments).
Suppose for some $k \in \mathbb{N}$, the derivative $\varphi_X^{(2k)}(0)$ exists, where

$$\varphi_X(t) = \mathbb{E}e^{itX}$$

Then

$$\mathbb{E}X^{2k} < \infty \quad \text{and} \quad \mathbb{E}X^{2k} = (-1)^k \varphi_X^{(2k)}(0).$$

(w/p) **Remark.** We do not write the absolute value in the statement since the power is even.
The Idea of Proof. Induction on k. Base case: $k = 1$

$$\mathbb{E}X^2 = \int x^2 P_X(dx) = \int \lim_{t \to 0} \frac{\sin^2(tx)}{t^2} P_X(dx) \quad \text{(by Fatou's lemma)} \leq \liminf_{t \to 0} \int \frac{\sin^2(tx)}{t^2} P_X(dx)$$

$$= \liminf_{t \to 0} \int \frac{e^{2itx} - 2 + e^{-2itx}}{-4t^2} P_X(dx) = \liminf_{t \to 0} \frac{\varphi_X(2t) - 2 + \varphi_X(-2t)}{-4t^2}$$

Recall that, by assumption, our function has a second derivative. Let's expand it:

$$\varphi_X(t) = 1 + \varphi_X'(0)t + \varphi_X''(0) \cdot \frac{t^2}{2} + o(t^2), \quad t \to 0.$$

Substitute this into the chain of equalities above:

$$\frac{\varphi_X(2t) - 2 + \varphi_X(-2t)}{-4t^2} = \frac{\left(\frac{(2t)^2}{2} + \frac{(-2t)^2}{2}\right)\varphi_X''(0) + o(t^2)}{-4t^2} = -\varphi_X''(0).$$

Therefore, the moment exists:

$$\mathbb{E}X^2 < \infty.$$

Referring to the previous theorem:

$$\mathbb{E}X^2 = -\varphi_X''(0).$$

Remark. For an odd order, the statement is not true; that is, the existence of a derivative of odd order does not imply the existence of a moment of the same order, but it does imply the existence of a moment of one order lower.

Reconstruction of a distribution from its characteristic function.

Theorem. Let X be a random variable, $\phi_x(t) = Ee^{itX}$, $P(X = a) = P(X = b) = 0, a < b$. Then

$$P(X \in [a, b]) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-iat} - e^{-ibt}}{it} \cdot \phi_x(t)dt.$$

9. Laplace transform function

The Laplace transform of a random variable X is defined as:

$$\mathcal{L}_X(s) = \mathbb{E}[e^{-sX}] = \int_0^\infty e^{-sx} f_X(x)dx, \quad s > 0.$$

It is often applied in analyzing distributions with exponential decay, in queueing theory, and reliability studies.
Properties.
1. $L_{a\,X+b}(s) = e^{-b\,s} L_x(a\,s), \quad a, b \in \mathbb{R}$.
2. $L_x(0) = 1, \quad L_x(s) \le 1$.
3. $sign\left(L_x^{(k)}(s)\right) = (-1)^k, \quad (X \ne 0)$.
4. $L_x^{(k)}(0) = (-1)^k \mathbb{E}(X^k)$.
5. If $X \in \mathbb{Z}_{\ge 0}$, then $G_x(e^{-u}) = L_x(u), \quad u \ge 0$.

10. Moment-generating function

The moment-generating function of a random variable X is given by:

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^\infty e^{tx} f_X(x)dx, \quad \text{if it converges.}$$

If $M_X(t)$ exists in some neighborhood of zero, all moments of X can be found by differentiation:

$$\mathbb{E}[X^n] = M_X^{(n)}(0).$$

This function is useful for obtaining moments of a distribution and is applied in parameter estimation methods.

Problems.

a) Find this characteristics for Unif[0,1], Exp(1), Norm(0,1), Ber(p).

Problem.

On one shelf, there are 6 red books and 4 black books, and on the other shelf, there are 5 red books and 6 black books. We randomly choose a shelf and take 2 books from it. Let X be the number of red books taken, and Y be the number of black books left on the second shelf.
Construct their distribution.

§7. Random vectors.

Formal Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A **random vector** of dimension n is a mapping:

$$\mathbf{X} : \Omega \to \mathbb{R}^n,$$

where

$$\mathbf{X}(\omega) = (X_1(\omega), \ldots, X_n(\omega))^T,$$

and each component X_i is a random variable, i.e., a measurable function $X_i : \Omega \to \mathbb{R}$.
Distributions of a random vector are divided into four types: discrete, absolutely continuous, singular continuous (singular), and their mixtures.
Definition. Singular continuous distribution.

$$\exists A \subseteq \mathbb{R}^n : \quad m_{leb}(A) = 0, \quad P_X(A) = 1\,(> 0)$$

. **Example.** Consider a random vector (X,Y) that is concentrated on a certain curve γ in $\mathbb{R}^2$.

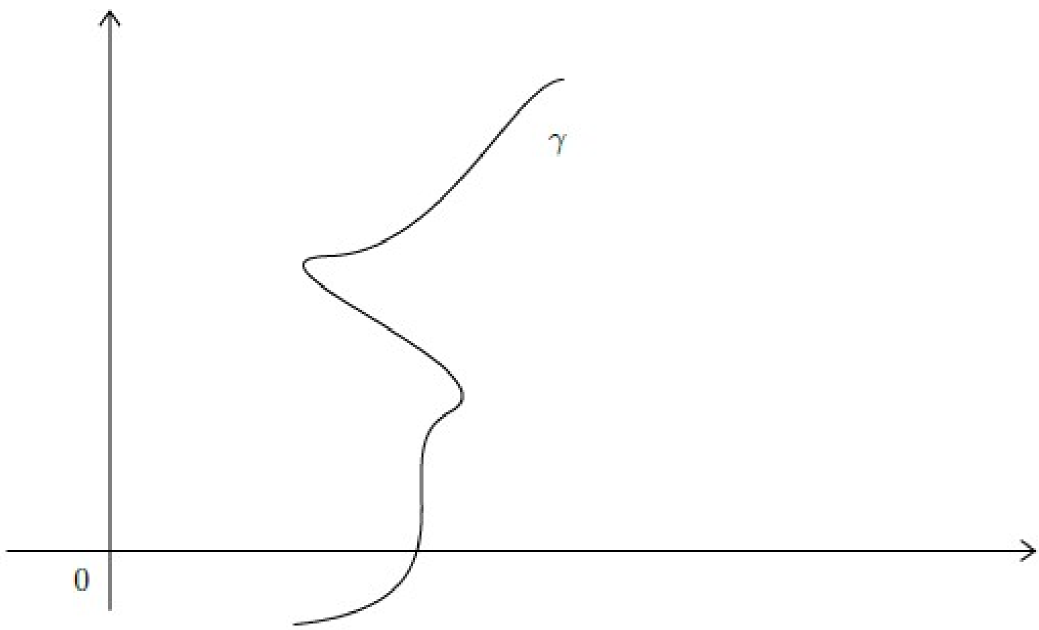

Figure 7: Singular distribution

Joint Distribution

Distribution Function (CDF)

The joint distribution function of the random vector $\mathbf{X} = (X_1, \ldots, X_n)^T$ is:

$$F_{\mathbf{X}}(x_1, \ldots, x_n) = \mathbb{P}(X_1 < x_1, \ldots, X_n < x_n)$$

Dicrete case.

Joint Probability Mass Function Instead of a density, we define a **probability mass function (PMF)**:

$$p_{\mathbf{X}}(x_1, \ldots, x_n) = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)$$

with normalization:

$$\sum_{(x_1,\ldots,x_n)\in\mathcal{X}} p_{\mathbf{X}}(x_1, \ldots, x_n) = 1$$

Abs. continuous case.

Probability Density Function (PDF) If there exists a function $f_{\mathbf{X}}(x_1, \ldots, x_n)$ such that:

$$F_{\mathbf{X}}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\mathbf{X}}(t_1, \ldots, t_n)\, dt_1 \ldots dt_n,$$

then $\mathbf{X}$ has a density function.
Property.
1.$f_{\mathbf{X}}(\cdot) \geq 0$;
2.

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(t_1, \ldots, t_n)\, dt_1 \ldots dt_n = 1;$$

3.

$$f_{\mathbb{X}}(x_1, ..., x_n) = \frac{\partial^n}{\partial x_1 ... \partial x_n} F(x_1, ..., x_n)$$

Marginal Distributions

The distribution of a single component or a subset of components is called the **marginal distribution**. **discrete case.**
It is obtained by summation over the remaining variables:

$$p_{X_1}(x_1) = \sum_{(x_2,\ldots,x_n)} p_{\mathbf{X}}(x_1, x_2, \ldots, x_n)$$

Abs. continuous case.
It is obtained by integrating over the remaining variables:

$$f_{X_1}(x_1) = \int_{\mathbb{R}^{n-1}} f_{\mathbf{X}}(x_1, x_2, \ldots, x_n)\, dx_2 \ldots dx_n$$

Moments.

Discrete case

$$E(g(X, Y)) = \sum_{i=1,\ldots,\infty;\quad j=1,\ldots,\infty} g(x_i, y_j) P(X = x_i, Y = y_j).$$

absolute continuous case

$$E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{x,y}(x, y) dx dy.$$

Independence.

Random variables $X_1, ..., X_n$ are independent (jointly), if

$$\forall B_1, ..., B_d \in \mathbb{B}(\mathbb{R}) \qquad P(X_1 \in B_1, ..., X_d \in B_d) = P(X_1 \in B_1) \cdot ... \cdot P(X_d \in B_d).$$

discrete case.
The components $X_1, \ldots, X_n$ are independent if:

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i)$$

abs. continuous case.
Components X_1,...,X_j are independent if

$$f_{X_1,...,X_n}(x_1, ..., x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

Problems.

1. A random vector is given by the following probability density function:
a)

$$f(x, y) = \begin{cases} cx, & \text{if } x, y \in [0, 1], \\ 0, & \text{otherwise.} \end{cases}$$

Find the constant c, construct the cumulative distribution function, and check for independence.

b)

$$f(x, y) = \begin{cases} c, & \text{if } x, y \in (0, 1], \text{ and } y > x, \\ 0, & \text{otherwise.} \end{cases}$$

Find the constant c, construct the cumulative distribution function, and check for independence.
2. Joint distribution function of a random vector (X,Y) is equal $F(x, y) = 1 - 2^(-x^2) - 3^(-y^2) + 2^(-x^2)3^(-y^2)$
Find the probability density function of a random vector (X,Y) and $P(X \in [1, 2[, Y \in [0, 1]]])$.
3. The joint probability density function of two random variables (X, Y) is given by:

$$f(x, y) = \frac{1}{2} \sin(x + y), \quad \text{for } 0 \le x \le \frac{\pi}{2}, \ 0 \le y \le \frac{\pi}{2}.$$

For other values of the arguments, it is assumed to be zero. Determine the joint cumulative distribution function of the random variables X and Y.
4. Determine the distribution function of a random variable equal to the maximum of two independent random variables X and Y, with distribution functions $F_X(x)$ and $F_Y(x)$, respectively.
Determine the distribution function of a random variable equal to the minimum of two independent random variables X and Y, with distribution functions $F_X(x)$ and $F_Y(x)$, respectively.

Property of joint Distribution Function

Property of DF. Let us list the properties of a joint distribution function. For simplicity, we consider the random vector (ξ_1, ξ_2) of two components.

(F0) For any x_1, x_2, the following inequality holds: $0 \le F_{\xi_1,\xi_2}(x_1, x_2) \le 1$.

(F1) $F_{\xi_1,\xi_2}(x_1, x_2)$ is non-decreasing in each coordinate of the vector (x_1, x_2).

(F2) For each $i = 1, 2$, we have $\lim_{x_i \to -\infty} F_{\xi_1,\xi_2}(x_1, x_2) = 0$ (or in general case $\lim\limits_{x_i \to -\infty} F_{\xi_1,...,\xi_n}(x_1, ..., x_n) = 0, \quad \forall i$). The double limit

$$\lim_{x_1 \to +\infty} \lim_{x_2 \to +\infty} F_{\xi_1,\xi_2}(x_1, x_2) = 1$$

also exists.

(F3) The function $F_{\xi_1,\xi_2}(x_1, x_2)$ is left-continuous in each coordinate of the vector (x_1, x_2).

(F4) To recover the marginal distribution functions of ξ_1 and ξ_2 from the joint distribution function, one should send the other variable to infinity:

$$\lim_{x_1 \to +\infty} F_{\xi_1,\xi_2}(x_1, x_2) = F_{\xi_2}(x_2), \quad \lim_{x_2 \to +\infty} F_{\xi_1,\xi_2}(x_1, x_2) = F_{\xi_1}(x_1).$$

Problem. Prove that the function

$$F(x_1, x_2) = \begin{cases} 0, & \text{if } x_1 \leq 0 \text{ or } x_2 \leq 0 \text{ or } x_1 + x_2 \leq 1; \\ 1, & \text{if simultaneously } x_1 > 0, x_2 > 0, x_1 + x_2 > 1 \end{cases}$$

satisfies all the properties (F0)—(F3), but is not a distribution function of any vector (ξ_1, ξ_2), at least because there exists a vector and a rectangle $[a_1, b_1) \times [a_2, b_2)$ such that the "probability" of falling into it (calculated using this "distribution function") is negative:

$$P(a_1 < \xi_1 < b_1, \quad a_2 < \xi_2 < b_2) < 0.$$

It is easy to prove (and easy to verify) that for any $a_1 < b_1, a_2 < b_2$, the following identity holds:

$$P(a_1 < \xi_1 < b_1, \quad a_2 < \xi_2 < b_2) = F_{\xi_1,\xi_2}(b_1, b_2) + F_{\xi_1,\xi_2}(a_1, a_2) - F_{\xi_1,\xi_2}(a_1, b_2) - F_{\xi_1,\xi_2}(b_1, a_2).$$

In addition to properties (F0)—(F3), the function F must satisfy the non-negativity of this expression (for any $a_1 < b_1, a_2 < b_2$).

Decision of problem 4.

$$F_{\max}(t) = P(\max(X, Y) < t) = F_x(t) \cdot F_y(t), \quad t \in \mathbb{R};$$

$$F_{\min}(t) = P(\min(X, Y) < t) = 1 - (1 - F_x(t)) \cdot (1 - F_y(t)), \quad t \in \mathbb{R}.$$

Problems.

1. Find the joint distribution of $(min(X, Y); max(X, Y))$.
2. Let X_i- be independent identically distributed random variablex with df $F_x(\cdot)$, consider the new sequence $(sort(X_i)): \quad X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq ... \leq X_{(n)}$. Find the distribution of $X_{(i)}, \forall i \in 1, ..., n$.

Answer.

1.

$$F_{\text{min,max}}(s, t) = F^2(t) - (F(t) - F(s))^2, s < t; \quad F^2(t), s \geq t.$$

2.

$$F_i(t) = P(X_{(i)} < t) = \sum_{j=i}^{n} c(n, j) F^j(t) (1 - F(t))^{n-j}.$$

Conditional Distribution

For discrete random variable:

$$\mathbb{P}(X_1 = x_1 \mid X_2 = x_2) = \frac{\mathbb{P}(X_1 = x_1, X_2 = x_2)}{\mathbb{P}(X_2 = x_2)}$$

For absolute continuous:
Conditional density: if $f_{\mathbf{X}}$ exists, then the conditional density of X_1 given X_2 is:

$$f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)}$$

This is a key concept in Bayesian statistics and information theory.

Covariation and Correlation

Covariance:

$$cov(X, Y) = \mathbb{E}\left((X - \mathbb{E}X) \cdot (Y - \mathbb{E}Y)\right) = \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y);$$

Deginition. X,Y are uncorrelated if cov(X,Y)=0, i.e. $E(XY) = E(X) \cdot E(Y)$.
Property.
0. $cov(X, Y) = E(XY) - E(X) \cdot E(Y)$;
1. Symmetry: $cov(X, Y) = cov(Y, X)$;
2. $cov(X, X) = D(X)$;
3. Linearity: $cov(aX + b, Y) = acov(X, Y)$;
4. if X,Y– independent $\Rightarrow cov(X, Y) = 0$;
$\Leftarrow$ is incorrect.
counterexample: $X \sim Unif[-1, 1], Y = X^2$.
5. $D(X \pm Y) = D(X) + D(Y) \pm 2cov(X, Y)$;
6. $cov(X, Y) \leq \sqrt{D(X) \cdot D(Y)}$.
Correlation:

$$corr(X, Y) = \rho_{x,y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{D}(X)\text{D}(Y)}}.$$

Correlation measures the strength of linear dependence between variables.
property:
0. $corr(X, Y) = corr(Y, X)$;
1. $corr(\cdot, \cdot) \in [-1, 1]$;
2. $corr(aX + b, Y) = sgn(a) \cdot corr(X, Y)$;
3. X,Y are uncorrelated $\Leftrightarrow \rho_{x,y} = 0$.
4. **Lemma. About linear dependence.**
$|\rho_{x,y}| = 1 \Rightarrow \exists a, b \in \mathbb{R} : Y = aX + b$.
Proof. Consider the function of t:

$$f(t) := \mathbb{D}(X + tY)$$

We use the linearity of covariance in each of its arguments:

$$\mathbb{D}(X + tY) = \text{Cov}(X + tY, X + tY) = \text{Cov}(X, X) + 2t \cdot \text{Cov}(X, Y) + t^2 \cdot \text{Cov}(Y, Y) \geq 0 \quad \forall t.$$

With respect to t, we obtain a quadratic polynomial, and since this is a variance, it must be non-negative. This means that the discriminant is less than or equal to zero:

$$4(\text{Cov}(X,Y))^2 - 4\mathbb{D}X \cdot \mathbb{D}Y \le 0$$

When the correlation coefficient equals one, the inequality above becomes an equality:

$$|r(X,Y)| = 1 \Rightarrow (\text{Cov}(X,Y))^2 = \mathbb{D}X \cdot \mathbb{D}Y \Rightarrow \exists t: \quad \mathbb{D}(X+tY) = 0.$$

Then we know that

$$X + tY = \text{const}.$$

$$\mathbb{D}X \pm 2t\sqrt{\mathbb{D}X \cdot \mathbb{D}Y} + t^2\mathbb{D}Y = 0$$

$$t = \pm\sqrt{\frac{\mathbb{D}X}{\mathbb{D}Y}} \Rightarrow X = -tY + \text{const}$$

problems.

Calculate the correlation of the following random variables:
1)if (X,Y) is given as
$P(X=-1, Y=1) = 0.1; \quad P(X=-1, Y=2) = 0.3; \quad P(X=1, Y=1) = 0.4; \quad P(X=1, Y=2) = 0.2$
2) if (X,Y) is given by the following density:

$$f_{(X,Y)}(x,y) = \begin{cases} 2, & x,y \in [0,1], x > y; \\ 0, & else. \end{cases}$$

3)Let $\xi, \xi_1, ..., \xi_n$ be independent identically distributed random variable with $E(\xi) = a$, $D(\xi) = \sigma^2$; find the correlation for the following r.v's.
a)$X = \xi_1 + \xi_2; \quad Y = \xi_2;$
b) $X = \xi_1 + \xi_2; \quad Y = \xi_2 + \xi_3;$
c) $X = \xi_1 + \xi_2 + \xi_3; \quad Y = \xi_1;$
d) $X = \xi_1 + \xi_2 + \xi_3; \quad Y = \xi_1 + \xi_2;$
e) $X = \xi_1 + \xi_2 + \xi_3; \quad Y = \xi_2 + \xi_3 + \xi_4;$

Moments of a Random Vector

Expectation Vector

$$\mathbb{E}[\mathbf{X}] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_n])^T$$

Covariance Matrix

$$\Sigma = \text{Cov}(\mathbf{X}) = \mathbb{E}\left[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T\right]$$

It is symmetric and positive semi-definite:

- Diagonal: variances of the components
- Off-diagonal: covariances $\text{Cov}(X_i, X_j)$

Proof of positive semi-definition.

$$\forall y \in \mathbb{R}^n \quad \langle y, By \rangle \geq 0.$$

$$\langle y, By \rangle = \sum_{t,j=1}^{n} y_t \cdot \mathrm{Cov}(X_t, X_j) \cdot y_j$$

Let us rearrange and simplify:

$$= \mathrm{Cov}\left(\sum_{t=1}^{n} y_t X_t, \sum_{j=1}^{n} y_j X_j\right) = \mathbb{D}\left(\sum_{t=1}^{n} y_t X_t\right) \geq 0$$

$$\exists\, y \neq 0 \quad : \quad \langle y, By \rangle = 0 \Rightarrow \mathbb{D}(\langle y, X \rangle) = 0$$

Therefore,

$$\exists\, \alpha \quad : \quad \langle y, X \rangle = \alpha \quad \text{almost surely.}$$

Linear Transformations

Let A be a matrix of size $m \times n$ and $\mathbf{X} \in \mathbb{R}^n$, then:

$$\mathbf{Y} = A\mathbf{X} + \mathbf{b}$$

is also a random vector, with:

$$\mathbb{E}[\mathbf{Y}] = A\mathbb{E}[\mathbf{X}] + \mathbf{b} \quad \text{and} \quad \mathrm{Cov}(\mathbf{Y}) = A\mathrm{Cov}(\mathbf{X})A^T$$

Multivariate Normal Distribution

A very important special case:

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

where:

- $\boldsymbol{\mu} \in \mathbb{R}^n$ is the mean vector
- $\Sigma \in \mathbb{R}^{n \times n}$ is the covariance matrix

The probability density function is:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$

property.

1. X,Y-are uncorrelated $\Leftrightarrow$ X,Y are independent.

proof

Firstly find the joint density for 2 random variables.

Let us consider a two-dimensional normal random vector $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$, whose probability density function we want to derive.

Assume $X \sim \mathcal{N}(\mu, \Sigma)$, where:

- $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ is the mean vector - $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$ is the covariance matrix - $\rho \in [-1, 1]$ is the correlation coefficient between X_1 and X_2

The probability density function of the bivariate normal distribution is given by:

$$f(x) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

where:

- $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ - $|\Sigma|$ is the determinant of the covariance matrix - Σ^{-1} is the inverse of the covariance matrix

The determinant is:

$$|\Sigma| = \sigma_1^2\sigma_2^2(1-\rho^2)$$

The inverse of the covariance matrix is:

$$\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}$$

Substituting these into the general formula yields:

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right)$$

• This is the probability density function of the bivariate normal distribution.
We want to demonstrate that when the correlation coefficient $\rho = 0$, the components of a bivariate normal random vector are independent.

Start with the bivariate normal density function. When $\rho = 0$, the formula simplifies to:

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right)$$

Note that the exponent is a sum of two terms. Therefore, we can factor the expression:

$$f(x_1, x_2) = \left(\frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right)\right) \cdot \left(\frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)\right)$$

This means:

$$f(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$$

where:

- $f_{X_1}(x_1)$ is the density of $\mathcal{N}(\mu_1, \sigma_1^2)$ - $f_{X_2}(x_2)$ is the density of $\mathcal{N}(\mu_2, \sigma_2^2)$

Since the joint density factors into the product of the marginal densities, it follows that X_1 and X_2 are independent.

2. if $X \sim N(\mu, \Sigma) \Rightarrow AX + b \sim N(A\mu + b, A\Sigma\, A^T)$, where A- a matrix $m \times n$.

Problems.

Let $(X_1, X_2, X_3)^T \sim N(\mu, \Sigma)$, where $\mu = (-1, 1, 2), \Sigma = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 4 & 3 \\ 2 & 3 & 5 \end{bmatrix}$

Find distributions for the following expressions:
a) Find margin distributions;
b) Distribution of pair sum;
c) $X_1 + X_2 + X_3$;
d) $(3X_1 - 2X_2, X_3 + 2X_1)$.

§8. Transformation of random vectors.

Unidimensional Function of random vectors.

Let random vector $\mathbb{X}$ is given by density function $f_{\mathbb{X}}(\cdot)$.
If $Y = g(\mathbf{X})$, then:

$$F_Y(y) = P(g(\mathbf{X}) \leq y)$$

Sometimes it is convenient to express this probability as a multivariate integral:

$$F_Y(y) = \int_{g(\mathbf{x}) \leq y} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}$$

Convolution of random variables.

Suppose we want to find the distribution function of the sum

$$\eta = \xi_1 + \xi_2 + \ldots + \xi_n,$$

where $p(x_1, x_2, \ldots, x_n)$ is the probability density function of the random vector $(\xi_1, \xi_2, \ldots, \xi_n)$.

The desired distribution function equals the probability that the point $(\xi_1, \xi_2, \ldots, \xi_n)$ lies in the half-space

$$\xi_1 + \xi_2 + \ldots + \xi_n < x,$$

and therefore,

$$F(x) = \int \!\ldots\! \int_{\sum x_k < x} p(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n.$$

• If X and Y are **not independent**, the convolution formula is no longer valid in this form, since the joint distribution $f_{X,Y}(x, y)$ cannot be factorized. In this case, the density of $Z = X + Y$ is:

$$F_Z(z) = \int \int\limits_{x+y<z} f(x,y) dx dy = \int \int\limits_{-\infty}^{z-x} f(x,y) dx dy = \int \int\limits_{-\infty}^{z} f(x, t-x) dt\, dx = \int \int\limits_{-\infty}^{z} f(t-y, y) dt\, dy;$$

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(z-y, y)\, dy = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)\, dx.$$

This is a **generalized convolution formula** that works even when X and Y are dependent.

• Let X and Y be independent random variables with density functions $f_X(x)$ and $f_Y(y)$, respectively. Then the **density function** of the sum $Z = X + Y$ is given by the **convolution** of f_X and f_Y:

$$f_Z(z) = (f_X * f_Y)(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx$$

For discrete random variables, the convolution becomes:

$$P(Z=z) = \sum_{k=-\infty}^{\infty} P(X = z-k) \cdot P(Y=k)$$

Problems.

Find the distribution of convolutions of the following random variables:
1) $X_1, X_2 \sim Norm(\mu_i, \sigma_i^2)$ are independent;
2) $X_1, X_2 \sim Pois(\lambda)$ and for $Pois(\lambda_i)$ are independent;
3) $X_1, X_2 \sim Unif[0,1]$.

The distribution of the quotient of random variables.

If $\zeta := \frac{\xi}{\eta}$, when the joint distribution of a random vector (ξ, η) is given by $f_{\xi,\eta}$.

$$F_\zeta(x) = \mathbb{P}\left(\frac{X}{Y} < x\right).$$

According to the general formula, the desired probability is

$$F_\zeta(x) = \int_0^\infty \int_{-\infty}^{zx} p(y,z)\, dy\, dz + \int_{-\infty}^0 \int_{zx}^\infty p(y,z)\, dy\, dz.$$

It follows that if ξ and η are independent, and $p_1(x)$, $p_2(x)$ are their respective densities, then

$$F_\zeta(x) = \int_0^\infty F_1(xz) p_2(z)\, dz + \int_{-\infty}^0 (1 - F_1(xz)) p_2(z)\, dz.$$

Differentiating, we obtain

$$p_\zeta(x) = \int_0^\infty z p(xz, z)\, dz - \int_{-\infty}^0 z p(xz, z)\, dz.$$

• If random variables are independent then

$$p_\zeta(x) = \int_0^\infty z p_1(xz) p_2(z)\, dz - \int_{-\infty}^0 z p_1(xz) p_2(z)\, dz.$$

Problems.

1. For $X, Y \sim Exp(a)$,
2. For $X, Y \sim Norm(0,1)$.
3. Find X+Y, where $X, Y \sim Exp(a)$.

Multidimensional Function of random vectors.

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a random vector in $\mathbb{R}^n$, and let

$$g : \mathbb{R}^n \to \mathbb{R}^m$$

be a measurable function. Then the image $\mathbf{Y} = g(\mathbf{X}) \in \mathbb{R}^m$ is also a random vector.

If $\mathbf{X}$ has a joint probability density function $f_{\mathbf{X}}(x_1, \ldots, x_n)$, and if g is bijective and differentiable with inverse g^{-1}, then the density of $\mathbf{Y}$ is given by the change of variables formula:

$$f_{\mathbf{Y}}(y_1, \ldots, y_m) = f_{\mathbf{X}}(g^{-1}(y)) \cdot \left|\det\left(J_{g^{-1}}(y)\right)\right|,$$

where $J_{g^{-1}}(y)$ is the Jacobian matrix of the inverse transformation.

Reminder. Jacobian Determinant

Let $\mathbf{Y} = g(\mathbf{X}) = (Y_1, Y_2, \ldots, Y_n)$ be a differentiable transformation of the vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$.

The **Jacobian matrix** of the transformation g is:

$$J_g(\mathbf{x}) = \begin{bmatrix} \frac{\partial Y_1}{\partial X_1} & \cdots & \frac{\partial Y_1}{\partial X_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial Y_n}{\partial X_1} & \cdots & \frac{\partial Y_n}{\partial X_n} \end{bmatrix}$$

The **Jacobian determinant** is $\det(J_g(\mathbf{x}))$.

In general, for a non-invertible or non-linear mapping, the distribution of $\mathbf{Y}$ can be found using transformation methods, moment-generating functions, or numerical integration.

Problems.

1. Let X,Y $\sim Exp(a)$, consider the new random variable $U = X+Y$, $V = \frac{X}{Y}$, find $f_{u,v}$. Check the independence.
2. Let (X,Y) are given by the following density:

$$f_{X,Y}(x,y) = 4xy, \quad x, y \in [0,1]; \qquad 0; \quad \text{else.}$$

Consider $U = \frac{X}{Y}, \quad V = X \cdot Y$. Sketch the support of (U,V) and joint distribution.

§9. Conditional mathematic expectation.

Reminder. Conditional probability $P(B) > 0$:

$$P(A|B) = \frac{P(AB)}{P(B)}, \quad P(\xi \in \Delta) = E\left(\mathbb{1}_{\Delta}(\xi)\right), \quad P(AB) = E(\mathbb{1}_A \cdot \mathbb{1}_B).$$

Moreover for discrete random variable:

$$P(A|Y=y) = \frac{P(A \cap \{Y=y\})}{P(\{Y=y\})}, \quad P(\{Y=y\}) > 0.$$

In this case we can **define** conditional expectation with respect to a set B

$$E(\mathbb{1}_A|B) = \frac{E(\mathbb{1}_A \cdot \mathbb{1}_B)}{P(B)}, \qquad E(\xi|B) = \frac{E(\xi \cdot \mathbb{1}_B)}{P(B)}.$$

Remark. If $C_1, ..., C_k$ are some sets which form σ-algebry $\mathfrak{A}$ and $\forall i \neq j \quad C_i \cap C_j = \emptyset$, then we can write the following:

$$E(\xi|\mathfrak{A}) = \sum_{i=1}^{k} E(\xi|C_i)\mathbf{1}_{C_i} = \sum_{i=1}^{k} \frac{E(\xi\, \mathbf{1}_{C_i})}{P(C_i)} \cdot \mathbf{1}_{C_i},$$

ans for $\omega \in C_j \quad E(\xi|\mathfrak{A}) := E(\xi|C_j) = \frac{E(\xi 1_{C_j})}{P(C_j)}$.
Moreover for an arbitrary $B \in \mathfrak{A}$ such as $B = B_{k1} \cup B_{k2} \cup ... \cup B_{km}$,

$$\int_B E(X|\mathfrak{A})dP = \sum_{i=1}^{m} E(X|B_{ki})P(B_{ki}) = \sum_{i=1}^{m} E(X 1_{B_{ki}}) = \int_B X dP.$$

Definition. Let X be a random variable, $E(X) < \infty$ and $\mathfrak{F}$ - sigma-algebra, then $Z := E(X|\mathfrak{F})$ is such random variable, that
1. Z – $\mathfrak{F}$-measurable.
2. $\forall B \in \mathfrak{F} \quad E(X\mathbf{1}_B) = E(Z\mathbf{1}_B)$.

Lemma. $E(X|\mathfrak{F})$ exists and is unique.
• Existence by The Radon-Nikodym theorem:
The measure $\widehat{P}(B) := \int_B X dP, \quad B \in \mathfrak{F}$ is absolute continuous with respect to measure P (i.e. $U \in \mathfrak{F} \quad P(U) = 0 \Rightarrow \widehat{P}(U) = 0$), conditional probability $A \in \mathfrak{F} \quad P(A|\mathfrak{F})$ with respect to σ-algebra $\mathfrak{B} \subset \mathfrak{F}$ is defined by formula $P(A|\mathfrak{B}) := E(1_A|\mathfrak{B})$and $\int_D P(A|\mathfrak{B})dP = P(AD), \quad D \in \mathfrak{B}, A \in \mathfrak{F}$.
• Uniqueness: Consider Z_1, Z_2 are two option of $E(X|\mathfrak{F})$, then

$$\forall B \in \mathfrak{A} \quad E(Z_i \cdot 1_B) = E(X \cdot 1_B) \Rightarrow E((Z_1 - Z_2)1_B) = 0,$$

then consider $B = 1(Z_1 > Z_2)$ and $B = 1(Z_2 > Z_1)$. Hence $P(Z_1 > Z_2) = P(Z_1 < Z_2) = 0$ then $P(Z_1 = Z_2) = 1$.

Lemma.
$E(X|Y) =: g(Y)$, where
in discrete case $g(y) = \frac{E(X \cdot \mathbf{1}(Y=y))}{P(Y=y)} = \sum_i x_i \cdot P(X = x_i|Y = y)$;
in abs. continuous case $g(y) = \int\limits_R x \frac{f_{x,y}(x,y)}{f_Y(y)} dx$.

Proof

$$\mathbb{E}\{X \mid Y = y\} := \lim_{\delta \downarrow 0} \mathbb{E}\{X \mid Y \in [y, y + \delta)\} = \lim_{\delta \downarrow 0} \frac{\mathbb{E}\{X\mathbb{I}_{[y,y+\delta)}(Y)\}}{\mathbb{P}(Y \in [y, y + \delta))};$$

$$\mathbb{E}\{X\mathbb{I}_{[y,y+\delta)}(Y)\} = \int_{Y\in[y,y+\delta)} X\, d\mathbb{P} = \int_{Y\in[y,y+\delta)} \mathbb{E}\{X \mid \sigma(Y)\}\, d\mathbb{P} = \int_{Y\in[y,y+\delta)} g(Y)\, d\mathbb{P}$$
$$= \int_{[y,y+\delta)} g(z)\, d\mathbb{P}(Y < z) = (g(y) + o(1))\mathbb{P}(Y \in [y, y + \delta)).$$

Then

$$\mathbb{E}\{X \mid Y = y\} = \lim_{\delta \downarrow 0} \frac{\mathbb{E}\{X\mathbb{I}_{[y,y+\delta)}(Y)\}}{\mathbb{P}(Y \in [y, y + \delta))} = \lim_{\delta \downarrow 0} \frac{\int_y^{y+\delta} \int_{-\infty}^{\infty} x f(x, z)\, dx\, dz}{\int_y^{y+\delta} f_Y(z)\, dz}$$
$$= \int_{-\infty}^{\infty} x \frac{f(x, y)}{f_Y(y)}\, dx.$$

Problems.

1. The distribution of a random vector (X,Y) is given by the following table:

X/Y	Y=0	Y=1
X=1	0.1	0.2
X=2	0.3	0.4

Find $E(X|Y)$ and $E(Y|X)$.
2. Let (X,Y) be given by the following joint density function:

$$f(x,y) = \begin{cases} 2, & x, y \in [0,1], \quad x > y. \\ 0, & \text{else.} \end{cases}$$

Find $E(X|Y)$ and $E(Y|X)$
3. Find $E(X \in \Delta | Y = y)$;
4. Find $E(g(X,Y))|Y = y$.
5. Prove $E(g(X,Y)|Y = y) = E(g(X,y)|Y = y)$.
6. Consider the probability space $([-1/2, 1/2], \mathfrak{B}([-1/2, 1/2]), P(d\omega) = d\omega$. Consider $Y(\omega) = \omega^3$. Find $(E(Y|\sigma(X)))$ if a) $X(\omega) = \omega^2$ $\quad X(\omega) = \max(0, |\omega| - 1/4)$
Find $E(Y|X = x)$.

Properties.

1. $E(const|\mathfrak{F}) = const$;
2. $E(aX|\mathfrak{F}) = aE(X|\mathfrak{F})$;
3. $E(X + Y|F) = E(X|\mathfrak{F}) + E(Y|\mathfrak{F})$;
4. $E(E(X|\mathfrak{F})) = E(X)$;
5. if $X \le Y \Rightarrow E(X|\mathfrak{F}) \le E(Y|\mathfrak{F})$;
6. If X is $\mathfrak{F}$-measurable, $E(X|\mathfrak{F}) = X$;
7. If X is indepemdent of $\mathfrak{F}$, than $E(X|\mathfrak{F}) = E(X)$.
8. If $E|X| < \infty$ then $|E(X|\mathfrak{F}) \le E(|X||\mathfrak{F})$;
9. If Y is $\mathfrak{F}$-measurable, then $E(XY|\mathfrak{F}) = YE(X|\mathfrak{F})$.
10. If $\mathfrak{A} \subset \mathfrak{B} \subset \mathfrak{F}$, then $E(\xi|\mathfrak{A}) = E\left(E(\xi|\mathfrak{B})|\mathfrak{A}\right)$.

§10. Probability inequalities for random variables, their sums and their characteristics.

1. Jensen's Inequality

Let X be a random variable such that $\mathbb{E}[X]$ exists, and let $\varphi(x)$ be a **convex function**. Then:

$$\varphi(\mathbb{E}[X]) \le \mathbb{E}[\varphi(X)]$$

If φ is **concave**, the inequality sign is reversed.

Examples:

- $\varphi(x) = x^2 \Rightarrow (\mathbb{E}[X])^2 \le \mathbb{E}[X^2]$
- $\varphi(x) = \log x \Rightarrow \log \mathbb{E}[X] \ge \mathbb{E}[\log X] \quad$ (for $X > 0$)

Remark. For conditional mathematical expectation it is correct too.

2. Markov's Inequality

Let $X \geq 0$ be a non-negative random variable with finite expectation. Then for any $a > 0$:

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

corollary Generalized Markov's Inequality

Let X be a random variable, and let $\varphi(x)$ be a non-negative, monotonically increasing function such that $\varphi(X) \geq 0$ almost surely and $\mathbb{E}[\varphi(X)] < \infty$. Then for any $a > 0$:

$$P(X \geq a) = P(\varphi(X) \geq \varphi(a)) \leq \frac{\mathbb{E}[\varphi(X)]}{\varphi(a)}$$

Example: If $\varphi(x) = x^p$ for $x \geq 0$ and $p > 0$, then:

$$P(X \geq a) \leq \frac{\mathbb{E}[X^p]}{a^p}$$

3. Chebyshev's Inequality

Let X be a random variable with finite expectation $\mu = \mathbb{E}[X]$ and finite variance $\sigma^2 = \mathbb{D}[X]$. Then for any $\varepsilon > 0$:

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

Markov's and Chebyshev's Inequalities for Sums of Random Variables

Markov's Inequality for the Sum of Random Variables

Let $S_n = X_1 + X_2 + \cdots + X_n$ be the sum of random variables (not necessarily independent), and assume $S_n \geq 0$ almost surely. Then, for any $a > 0$:

$$\mathbb{P}(S_n \geq a) \leq \frac{\mathbb{E}[S_n]}{a}$$

(Linearity of expectation does not require independence.)

Chebyshev's Inequality for the Sum of Random Variables

Suppose S_n has a finite expectation $\mathbb{E}[S_n]$ and variance $\mathrm{Var}(S_n)$. Then, for any $\varepsilon > 0$:

$$\mathbb{P}\left(|S_n - \mathbb{E}[S_n]| \geq \varepsilon\right) \leq \frac{\mathrm{Var}(S_n)}{\varepsilon^2}$$

When Random Variables are Independent

- In **Markov's inequality**, independence is **not needed**.
- In **Chebyshev's inequality**, independence is also **not required** for the inequality itself.
- However, if $X_1, \ldots, X_n$ are **independent**, then the variance simplifies:

$$\mathrm{Var}(S_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

Without independence, variance expands with covariances:

$$\mathrm{Var}(S_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j)$$

§11. Types of Convergence of Random Variables

Let (X_n) be a sequence of random variables, and X a random variable. There are several important types of convergence:

1. Almost Sure Convergence

$$X_n \xrightarrow{\text{a.s.}} X \quad \text{meaning} \quad \mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$$

2. Convergence in Probability

$$X_n \xrightarrow{\mathbb{P}} X \quad \text{meaning that for any } \varepsilon > 0, \quad \lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0$$

3. Convergence in L^p (Mean Convergence)

$$X_n \xrightarrow{L^p} X \quad \text{meaning} \quad \lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0$$

where $p \geq 1$.

4. Convergence in Distribution (Weak Convergence)

$$X_n \xrightarrow{d} X \quad \text{meaning} \quad \lim_{n\to\infty} F_{X_n}(x) = F_X(x) \quad \text{at all continuity points of } F_X$$

Relationships Between Types of Convergence

- Almost sure convergence implies convergence in probability.
- L^p convergence implies convergence in probability.
- Convergence in probability implies convergence in distribution.
- The converse statements generally do **not** hold.

Type of Convergence	Notation	Implies From	Implies To
Almost Sure (a.s.)	$X_n \xrightarrow{\text{a.s.}} X$	–	$X_n \xrightarrow{\mathbb{P}} X$
In L^p Mean	$X_n \xrightarrow{L^p} X$	–	$X_n \xrightarrow{\mathbb{P}} X$
In Probability	$X_n \xrightarrow{\mathbb{P}} X$	From a.s. or L^p	$X_n \xrightarrow{d} X$
In Distribution	$X_n \xrightarrow{d} X$	From in probability	–

How to check the almost surely convergence?

Borel–Cantelli Lemma (General Version)

Let (A_n) be a sequence of events in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

First Borel–Cantelli Lemma (for events)

If

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < +\infty,$$

then

$$\mathbb{P}\left(A_n \text{ occurs infinitely often}\right) = 0.$$

In other words,

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0,$$

where where

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Second Borel–Cantelli Lemma (for events)

If

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = +\infty,$$

and the events A_n are independent, then

$$\mathbb{P}\left(A_n \text{ occurs infinitely often}\right) = 1.$$

Borel–Cantelli Lemma for Almost Sure Convergence

Remark. For $\epsilon > 0$ let $A_n(\epsilon) = \{\omega : \left|X_n(\omega) - X(\omega)\right| > \epsilon\}$. Then the property of almost surely convergence as equivalent to saying that for every $\epsilon > 0$:

$$P\left(\{A_n(\epsilon) \quad \textbf{finitely often}\right) = 1.$$

This is why the Borel-Cantelli lemma is so useful in studying almost surely limits. Let X_n be a sequence of random variables, and X its limit.

Define for each $\varepsilon > 0$ the events

$$A_n(\varepsilon) = \{\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}.$$

Then:

- **If for each** $\varepsilon > 0$

  $$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) < +\infty,$$

  then by the First Borel–Cantelli Lemma,

  $$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1,$$

  which means $X_n \xrightarrow{\text{a.s.}} X$.

- **If for each** $\varepsilon > 0$

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = +\infty,$$

and the events $A_n(\varepsilon)$ are independent, then by the Second Borel–Cantelli Lemma,

$$\mathbb{P}\left(|X_n - X| \geq \varepsilon \text{ infinitely often}\right) = 1,$$

and thus $X_n \overset{\text{a.s.}}{\not\longrightarrow} X$ (Not almost surely convergences).

Condition on Tail Probabilities

Lemma. If for every $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) < +\infty,$$

then

$$X_n \xrightarrow{\text{almost surely}} X.$$

Remark. In general case, the inverse statemint of the Lemma is not correct.
For example, consider

$$\begin{cases} P(X_k = 0) = 1 - \frac{1}{k}, \\ P(X_k = k) = \frac{1}{k}. \end{cases}$$

$X_n \overset{a.s}{\to} 0$, but the corresponding series don't convergence.

Some important examples.

1. $(X_n \overset{P}{\to} X,\ X_n \overset{a.s.}{\nrightarrow} X,\ X_n \overset{L^1}{\to} X,\ X_n \overset{L^p}{\nrightarrow} X)$
For $n \geq 1$ put $m = \lfloor \log_2 n \rfloor$, i.e., let $m \geq 0$ be such that $2^m \leq n < 2^{m+1}$. Consider the events

$$A_n = \left[\frac{n - 2^m}{2^m}, \frac{n + 1 - 2^m}{2^m}\right) \subset [0, 1]$$

in the canonical probability space, and let $X_n(\omega) = \mathbf{1}_{A_n}(\omega)$. Because

$$P(|A_n| > 0) = P(A_n) = \mathbb{E}(\mathbf{1}_{A_n}) = 2^{-\lfloor \log_2 n \rfloor} < \frac{2}{n} \to 0$$

as $n \to \infty$, the sequence X_n converges to $X \equiv 0$ in probability and in L^p, for each fixed $p > 0$. However,

$$\{\omega \in \Omega : X_n(\omega) \to X(\omega) = 0 \text{ as } n \to \infty\} = \emptyset,$$

i.e., there is no point $\omega \in \Omega$ for which the sequence $X_n(\omega) \in \{0, 1\}$ converges to $X(\omega) = 0$; in fact, for each $\omega \in \Omega$ the real sequence $X_n(\omega)$ never stops jumping between 0 and 1.
2. Consider the following X_n

$$\begin{cases} P(X_k = 0) = 1 - \frac{1}{k}, \\ P(X_k = 1) = \frac{1}{k}. \end{cases}$$

In this case $X_k \overset{Lp}{\to} 0, \quad \forall p > 0$, but there is not the almost surely condition, because in case of such sequence

$$\begin{cases} P(X_k = 0) = 1 - p_k, \\ P(X_k = 1) = p_k, \end{cases}$$

the almost surely convergence to 0 is equivalent to the condition $\sum_{n=1}^{\infty} p_n < \infty$.
3. Consider the following X_n:

$$\begin{cases} P(X_k = 0) = 1 - p_k, \\ P(X_k = a_k \neq 0) = p_k, \end{cases}$$

Easy to find the restrictions on a_k and p_k for convergences almost surely, Lp, in probability. And then choose such a_k and p_k for which one type of convergences is perform and some other not.

In general, to verify convergence with probability one is not immediate. The following lemma gives a sufficient condition of almost sure convergence.

§12. Limit theorems for sums of random variables.

Law of Large Numbers (Hinčin, Lévy, Markov and Chebyshev).

Hinčin's Law of Large Numbers

Let $X_1, X_2, X_3, \dots$ be a sequence of independent random variables with identical distributions, $\mathbb{E}[X_1] = \mu$ and $\mathrm{Var}(X_1) = \sigma^2$ (finite).

Then for any $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| \geq \varepsilon\right) \to 0 \quad \text{as} \quad n \to \infty,$$

or equivalently,

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \mu.$$

Lévy's Law of Large Numbers

Let $X_1, X_2, X_3, \dots$ be a sequence of random variables with identical distributions, $\mathbb{E}[X_1] = \mu$, and there exists a constant C such that for all n,

$$\mathbb{E}[X_1^2] \leq C.$$

Then,

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \mu.$$

Remark. The necessary condition on moments of random variables can be weaken. So in this case $E(|X_i|) < \infty$ is enough.

Corollary.

1. Let $\{X_i\}_i$ be i.i.d.r.v's with distribution function $F(\cdot)$. Consider $F_n(t) = \frac{1}{n}\sum_{i=1}^{1} 1\{X_i < t\}$. By the Law of Large Number $F_n(t) \xrightarrow{P} F(t), n \to \infty$.
2. If $\xi_n \xrightarrow{P} C = const$ and $f(\cdot)$ is a continuous function at the point C, then $g(\xi_n) \xrightarrow{P} g(C)$.

Central Limit Theorem (CLT)

Let $X_1, X_2, X_3, \dots$ be a sequence of independent random variables with identical distributions, $\mathbb{E}[X_1] = \mu$, $\mathrm{Var}(X_1) = \sigma^2$, and $\sigma^2 > 0$.

Then for the sum of these variables:

$$S_n = \sum_{k=1}^{n} X_k,$$

the normalized sum

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1),$$

where $N(0,1)$ is the standard normal distribution.

Remark

$$P(S_n \in (a,b)) \approx \Phi\left(\frac{b - E(S_n)}{\sqrt{D(S_n)}}\right) - \Phi\left(\frac{a - E(S_n)}{\sqrt{D(S_n)}}\right),$$

where $\Phi(\cdot)$- distribution function of Norm(0,1).

Strong law of large numbers.

I. Kolmogorov's Strong Law of Large Numbers

Let $b_n > 0$ be a numerical sequence with $b_n \nearrow$ (increasing), and let $X_1, X_2, \dots$ be independent random variables such that

$$\mathbb{E}X_i^2 < \infty \quad (\mathbb{D}X_i < \infty).$$

Then

$$\sum_n \frac{\mathbb{D}X_n}{b_n^2} < \infty \quad \Rightarrow \quad \frac{S_n - \mathbb{E}S_n}{b_n} \xrightarrow[n\to\infty]{\text{a.s.}} 0,$$

where $S_n = X_1 + \dots + X_n$.

II. Kolmogorov's Strong Law of Large Numbers.

Let $\{X_n\}$ be a sequence of independent and identically distributed random variables such that

$$\mathbb{E}|X_1| < \infty.$$

Then, with probability 1,

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to\infty]{\text{a.s.}} \mathbb{E}X_1.$$

Problems.

0. Let $X_1, X_2, \dots, X_k$ be independent non-negative random variables, identically distributed such that

$$E(X_i) = 2, \quad D(X_i) = 1 \quad \text{for} \quad i = 1, 2, \dots, 100.$$

Let

$$S = \sum_{i=1}^{100} X_i.$$

a) Estimate the upper bound for the probability that $S \geq 300$.

b) Estimate the upper bound for the probability that the deviation of the sum from its expected value exceeds 50, i.e., $P(|S - E(S)| \geq 50)$.

1. Every 5 minutes, a random number of passengers arrives at a subway platform, distributed according to a Poisson distribution with parameter $\lambda = 250$. At the same time, trains depart from the platform, each capable of taking on a number of passengers uniformly distributed in the interval $[195, 205]$. Can we assume that a platform with capacity N will not be overloaded, if this regime continues indefinitely?

2. Let $\{X_n\}$ be a sequence of independent random variables such that

$$X_k \in \{-2^k, -1, 1, 2^k\}$$

with probabilities 2^{-2k-1}, $\frac{1-2^{-2k}}{2}$, $\frac{1-2^{-2k}}{2}$, 2^{-2k-1}, respectively. Does the law of large numbers (LLN) hold for the sequence $\{X_n\}$?

3. A fair die is rolled repeatedly until the total sum of outcomes exceeds 700. Estimate the probability that this will take: (a) more than 210 rolls; (b) fewer than 180 rolls.

4 When shooting at a target, the shooter hits: a 10 with probability 0.35, a 9 with probability 0.3, an 8 with probability 0.2, a 7 with probability 0.1, a 6 with probability 0.05. The shooter fires 100 shots. What is the probability that they score at least 900 points?