Exercise 1: *Two unbiased dice are thrown. Compute the probability that*

1. *the number of spots on the first die is odd (event A),*
2. *the number of spots on the second die is greater than 4 (event B),*
3. *the number of spots on each die is equal to 6 (event C),*
4. *the sum of spots on both dice is less than 5 (event D).*

<u>Solution:</u>
The results of the trial are given by the ordered pairs where the first number corresponds to the number of spots on the first die and the second number is the number of spots on the second die:

$$\begin{array}{l}
(1{,}1)\ (1{,}2)\ (1{,}3)\ (1{,}4)\ (1{,}5)\ (1{,}6)\\
(2{,}1)\ (2{,}2)\ (2{,}3)\ (2{,}4)\ (2{,}5)\ (2{,}6)\\
(3{,}1)\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (3{,}6)\\
(4{,}1)\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (4{,}6)\\
(5{,}1)\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (5{,}6)\\
(6{,}1)\ (6{,}2)\ (6{,}3)\ (6{,}4)\ (6{,}5)\ (6{,}6)
\end{array}$$

i.e. the number of all elementary events is equal to 36.

1. The number of elementary events corresponding to the event A is 18 (the first row, the third row and the fifth row), thus $P(A) = \frac{18}{36} = \frac{1}{2}$.
2. The number of elementary events corresponding to the event B is 12 (the fifth column and the sixth column), thus $P(B) = \frac{12}{36} = \frac{1}{3}$.
3. The number of elementary events corresponding to the event C is 1 (only the pair (6,6)), thus $P(C) = \frac{1}{36}$.
4. The number of elementary events corresponding to the event D is 12 (the pairs (1,1),(1,2),(1,3),(2,1),(2,2) and (3,1)), thus $P(D) = \frac{6}{36} = \frac{1}{6}$.

Exercise 2: *In a standard deck of 32 well-shuffled cards, there are 4 queens. A card is drawn twice at random from the deck. Find the probability that at least 2 queens are drawn if after the first turn, the card*

1. is returned (event E)

2. is not returned (event F)

back to the deck.

<u>Solution:</u>
The results of the trial are given by the ordered pairs where the first element corresponds to the card in the first draw and the second element corresponds to the card in the second draw.

1. The number of all elementary events is equal to 32^2 (32 possibilities in each draw). The elementary events corresponding to the event E are (queen - queen), (queen - not queen), (not queen - queen). The number of such events is $4 \cdot 4 + 4 \cdot 28 + 28 \cdot 4$. Thus
$$P(E) = \frac{4 \cdot 4 + 4 \cdot 28 + 28 \cdot 4}{32^2} = \frac{15}{64}.$$
An other solution:
The complementary event to the event E is the event E^c...there is no queen in the two draws, i.e. we draw a pair (not queen - not queen). The number of elementary events corresponding to the event E^c is $28 \cdot 28$. Thus
$$P(E) = 1 - P(E^c) = 1 - \frac{28 \cdot 28}{32^2} = \frac{15}{64}.$$

2. The number of all elementary events is now equal to $32 \cdot 31$, because in the second draw, one card is missing in the deck. The elementary events corresponding to the event F are (queen - queen), (queen - not queen), (not queen - queen). The number of such events is $4 \cdot 3 + 4 \cdot 28 + 28 \cdot 4$, because in the case (queen - queen), one queen is missing after the first draw. Thus
$$P(F) = \frac{4 \cdot 3 + 4 \cdot 28 + 28 \cdot 4}{32 \cdot 31} = \frac{59}{248}.$$
An other solution:
The number of elementary events corresponding to the event F^c is $28 \cdot 27$, because one "not queen" is missing after the first draw. Thus
$$P(F) = 1 - P(F^c) = 1 - \frac{28 \cdot 27}{32 \cdot 31} = \frac{59}{248}.$$

Exercise 3*: *Alan, Bob and Chris are playing tennis. Alan is better player than Bob. Chris is going to play three games against Alan and Bob while he can choose the order either Alan-Bob-Alan or Bob-Alan-Bob. Chris will obtain an award if he wins two games in a row. Which order should be chosen by Chris in order to increase the chance of wining the award?*

Solution:
Let a the probability that Chris wins the game with Alan and b the probability that Chris wins the game with Bob, then $a < b$.
In order to win the award, Chris must play either (win - win - win) or (win - win - loss) or (loss - win - win).
The probability to win the award in the order Alan - Bob - Alan is

$$aba + ab(1-a) + (1-a)ba = ab(2-a)$$

and the probability to win the award in the order Bob - Alan - Bob is

$$bab + ba(1-b) + (1-b)ab = ab(2-b).$$

Since $ab(2-a) > ab(2-b)$, it is better to choose the order Alan - Bob - Alan.

Exercise 4: *8 boys and 6 girls are making a party at dormitory. Find the probability that in a group of 4 students randomly chosen to play a game, there is*

1. *no boy (event G),*
2. *exactly 1 boy (event H),*
3. *at least 1 boy (event I).*

Solution:
The number of all groups consisting of 4 students is equal to $\binom{14}{4}$.

1. The number of groups consisting of 4 girls is $\binom{6}{4}$, thus

$$P(G) = \frac{\binom{6}{4}}{\binom{14}{4}} \doteq 0.015.$$

2. The number of groups consisting of 3 girls is $\binom{6}{3}$ and the number of groups consisting of 1 boy is $\binom{8}{1} = 8$ while we can combine all triples of girls with all boys. Thus

$$P(H) = \frac{8 \cdot \binom{6}{3}}{\binom{14}{4}} \doteq 0.16.$$

3. The event I is the complement to the event G, thus

$$P(I) = 1 - P(G) \doteq 0.985.$$

Exercise 5*: *What is the probability that in a group of n people, there is at least one pair having the birthday on the same day? (For simplicity, assume that the year has 365 days and the people are born uniformly during the whole year.)*

Solution:
Let G be the event that in the group of n people, there is at least one pair having the birthday on the same day. The probability that there is no pair having the birthday on the same day is

$$P(G^c) = \frac{365 \cdot 364 \cdot \ldots \cdot (365 - n + 1)}{365^n}.$$

Thus

$$P(G) = 1 - \frac{365 \cdot 364 \cdot \ldots \cdot (365 - n + 1)}{365^n}.$$

Exercise 6: *Consider two random numbers X and Y, both of them taking the values between 0 and 1 uniformly. Find the probability that both of them are greater than 0.3 while their sum is less than 1.*

Solution:
Denote A the event that both X and Y are greater than 0.3 while their sum is less than 1.
The set of all elementary events is the square with the side length equal to 1. The set of the elementary events corresponding to the event A is the triangle which is formed by intersection of three half-spaces:

- $y < -x + 1$
- $x > 0,3$
- $y > 0,3.$

The area of the triangle is equal to 0.08, so the probability is

$$P(A) = \frac{0.08}{1} = 0.08.$$

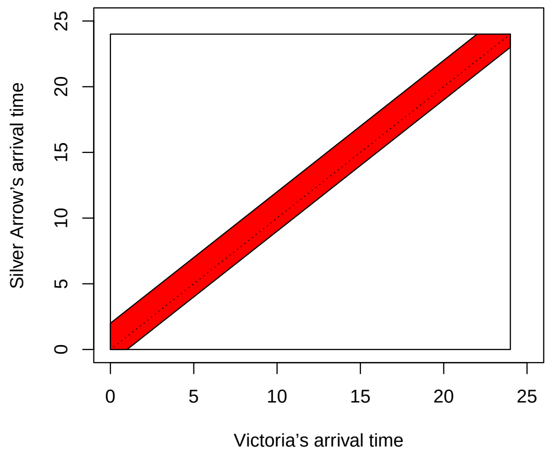
Exercise 7*: *On a square grid with the side length equal to a, a coin with the radius b is thrown, where $2b < a$. Find the probability that the coin will intersect a line of the grid.*

Solution:
Denote A the event that the coin will intersect a line of the grid. The coin does not intersect any line of the grid if its center lies inside the square (with the side length equal to a) in the distance at least b from the square side, i.e. it lies in a smaller square with the side length equal to $a - 2b$ Thus the probability that the coin does not intersect any line of the grid is $(a-2b)^2/a^2$, so the probability of intersection is

$$P(A) = 1 - \frac{(a-2b)^2}{a^2}.$$

Exercise 8:** *The boats Victoria and Silver Arrow are going to arrive in a given harbour in the following 24 hours, independently on each other. Victoria will stay for 2 hours and then, it will leave, Silver Arrow will stay for one hour and then, it will leave. Find the probability that the boats meet each other.*

Solution:
Denote A the event that the boats meet each other.
Consider the square with the side length equal to 24 playing the role of all possible pairs of times when Victoria (the x axis) and Silver Arrow (the y axis) can come. Then the part of the square bounded by the lines $y = x+2$ and $y = x - 1$ corresponds to the times when the boats meet each other, see the following picture:



Its area is equal to 69.5 (the area of the whole square minus the area of two triangles, i.e. $24^2 - (22^2/2 + 23^2/2)$), so the probability is

$$P(A) = \frac{69.5}{24^2} \doteq 0.12.$$

Exercise 9: *Two unbiased dice are thrown. Denote the events*
A... "the number of spots on the first die is odd",
B... "the number of spots on the second die is even,",
C... "the numbers of spots on both dice are equal".

1. *Are the events A, B, C independent?*
2. *Are they pairwise independent?*

<u>Solution:</u>
The set of all elementary events (sample space) is:

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

i.e. it has 36 elements.
The number of elementary events corresponding to the event A is 18 (the first row, the third row and the fifth row), the number of elementary events corresponding to the event B is 18 (the second column, the fourth column and the sixth column), and the number of elementary events corresponding to the event C is 6 (the diagonal), therefore

$$P(A) = \frac{18}{36} = \frac{1}{2}, \quad P(B) = \frac{18}{36} = \frac{1}{2}, \quad P(C) = \frac{1}{6}.$$

The number of elementary events corresponding to the event $A \cap B$ is 9 (the first row, the third row and the fifth row in the second column, the fourth column and the sixth column), The number of elementary events corresponding to the event $A \cap C$ is 3 (the elements (2,2),(4,4) and (6,6)), and the number of elementary events corresponding to the event $B \cap C$ is 3 (the elements (1,1),(3,3) and (5,5)), therefore

$$P(A \cap B) = \frac{1}{4}, \quad P(A \cap C) = \frac{1}{12}, \quad P(A \cap C) = \frac{1}{12}.$$

Further, $A \cap B \cap C$ is an impossible event, so $P(A \cap B \cap C) = 0$.
Since

$$P(A \cap B) = P(A)P(B), \; P(A \cap C) = P(A)P(C), \; P(B \cap C) = P(B)P(C),$$

the events are pairwise independent, but

$$P(A \cap B \cap C) \neq P(A)P(B)P(C),$$

so the events are not (totally/mutually) independent.

Exercise 10: *For the independent events A, B, C, it holds that $P(A) = 0.2$, $P(B) = 0.3$ and $P(A) = 0.4$. Find the probability $P(D)$, where $D = (A \cup B) \cap C$.*

Solution:
Since A, B and C are independent, it holds that

$$P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B) = 0.2 + 0.3 - 0.2 \cdot 0.3 = 0.44,$$
$$P(D) = P((A \cup B) \cap C) = P(A \cup B) \cdot P(C) = 0.44 \cdot 0.4 = 0.176\,.$$

Exercise 11: *Come back to the tennis players from Exercise 3. Chris is exhausted after the three games ;-), so now, Alan is playing against Bob. The probability than Alan wins a game is 0.7. Find the probability that in the row of 10 games, Bob wins*

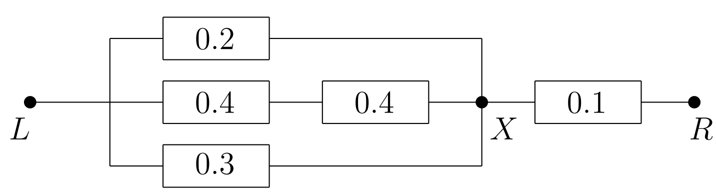
1. *once at least (the event A),*
2. *twice at most (the event B)?*

Solution:

1. $P(A) = 1 - P(A^c) = 1 - (0,7)^{10}$.
2. The event I means that Alan wins all ten games or Alan wins nine games and Bob wins one game (independently on the order) or Alan wins eight games and Bob wins two games (again independently on the order). Then

$$P(B) = (0.7)^{10} + \binom{10}{1} 0.3 \cdot (0.7)^9 + \binom{10}{2} (0.3)^2 \cdot (0.7)^8.$$

Exercise 12: *A signal is running from the left side of a machine (L) to the right side (R). Components in the machine can be broken with different probabilities as seen in the picture below. Note that the signal comes through a branch if no the component in the branch is broken, and it comes through the block from the knot L to the knot X if it comes through at least one branch. Find the probability that the signal will be delivered.*



Solution:
The events A, B, C mean that there occurs an error in the upper branch,

the middle branch and the lower branch, respectively, while $P(A) = 0.2$, $P(C) = 0.3$ and

$$P(B) = P(B_1 \cup B_2) = P(B_1) + P(B_2) - P(B_1 \cap B_2) = 0.4 + 0.4 - 0.4 \cdot 0.4 = 0.64\,,$$

where B_1 and B_2 are the events that an error occurs in the first and in the second component of the middle branch, respectively.
The probability that the signal does not go through the block from the knot L to the knot X is

$$P(D_1) = P(A \cap B \cap C) = 0.2 \cdot 0.64 \cdot 0.3 = 0.0384\,,$$

so the probability that it goes through the block is

$$P(D_1^c) = 1 - P(D_1) = 0.9616.$$

The probability that the signal does not go from the knot X to the knot R is

$$P(D_2) = 1 - 0.1 = 0.9\,,$$

so the probability that it goes from X to R is

$$P(D_2^c) = 1 - P(D_2) = 0.9.$$

Thus the probability that the signal goes from the knot L to the knot R is

$$P = P(D_1^c) \cdot P(D_2^c) = 0.86544\,.$$

Exercise 13*: *For the independent events A and B, it holds that $P(A \cup B) = 0.545$ and $P(A \cap B) = 0.105$. Find the probability $P(A)$, $P(B)$ and $P(A \cap B^c)$.*

<u>Solution:</u>
Since A and B are independent, it holds that

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

Denote $P(A) = x$ and $P(B) = y$. Then we get

$$0.545 = x + y - 0.105, \qquad x \cdot y = 0.105 \Rightarrow y = \frac{0.105}{x}\,.$$

So
$$x^2 - 0.65x + 0.105 = 0 \Rightarrow x_1 = 0.35, \quad x_2 = 0.3\,.$$
Thus,
$$P(A) = 0.35, \quad P(B) = 0.3 \qquad \text{or} \qquad P(A) = 0.3, \quad P(B) = 0.35\,.$$
Then
$$P(A \cap B^c) = P(A) \cdot P(B^c) = 0.35 \cdot 0.7 = 0.245$$
or
$$P(A \cap B^c) = 0.3 \cdot 0.65 = 0.195,$$
respectively.

Exercise 14: *At Faculty of Mathematics and Physics (Charles University), 30% of students study mathematics, 20% of them study physics and 50% of them study computer science. Among the students of mathematics, there are 30% of girls, among the students of physics, there are 20% of girls and 10% of girls are among the students of computer science. Find the probability that*

1. *a randomly chosen student is a girl,*
2. *a randomly chosen girl studies mathematics?*

<u>Solution:</u>
Denote
B... a randomly chosen student is a girl,
A_1... a randomly chosen student studies mathematics,
A_2... a randomly chosen student studies physics,
A_3... a randomly chosen student studies computer science.

1. From the law of total probability, we get
$$P(B) = \sum_{j=1}^{3} P(B|A_j) \cdot P(A_j),$$
while we have $P(A_1) = 0.3$, $P(B|A_1) = 0.3$, $P(A_2) = 0.2$, $P(B|A_2) = 0.2$, $P(A_3) = 0.5$ and $P(B|A_3) = 0.1$. Thus
$$P(B) = 0.3 \cdot 0.3 + 0.2 \cdot 0.2 + 0.1 \cdot 0.5 = 0.18.$$

2. We want to calculate $P(A_1|B)$. From Bayes theorem, we have

$$P(A_1|B) = \frac{P(B|A_1) \cdot P(A_1)}{\sum_{j=1}^{3} P(B|A_j) \cdot P(A_j)},$$

so

$$P(A_1|B) = \frac{0.3 \cdot 0.3}{0.3 \cdot 0.3 + 0.2 \cdot 0.2 + 0.1 \cdot 0.5} = \frac{1}{2}.$$

Exercise 15:** *It is known that* 1% *of drivers drink alcohol when driving. Further, the statistics show that among the drivers causing accidents, there are* 10% *of drivers who have drunk alcohol. How many times does the risk of accident increase when a driver drinks alcohol?*

Solution:
Denote
A... a randomly chosen driver has drunk alcohol,
H... a randomly chosen driver caused an accident.
Then $P(A) = 0.01$ and $P(A|H) = 0.1$. So

$$0.1 = P(A|H) = \frac{P(H|A) \cdot P(A)}{P(H|A) \cdot P(A) + P(H|A^c) \cdot P(A^c)} =$$

$$= \frac{P(H|A) \cdot 0.01}{P(H|A) \cdot 0.01 + P(H|A^c) \cdot 0.99} = \frac{1}{1 + \frac{P(H|A^c)}{P(H|A)} \cdot 99} .$$

Thus the result is the ratio

$$\frac{P(H|A)}{P(H|A^c)} = 11 .$$

Exercise 16: *Let's toss a coin, i.e.* $\Omega = \{\omega_1, \omega_2, \omega_3\}$, *where*
ω_1... *there is the head on the coin, while* $P(\omega_1) = 0.49$,
ω_2... *there is the tail on the coin, while* $P(\omega_2) = 0.49$,
ω_3... *a special situation occurred (the coin was lost, stolen... :-)), while* $P(\omega_3) = 0.02$*).*
Construct two different random variables and draw their (cumulative) distribution functions.

Solution:
The random variables could be for example

1. $X : X(\omega_1) = 1, X(\omega_2) = -1, X(\omega_3) = 0$. The distribution function is then partly constant with jumps of the size 0.49, 0.02 and 0.49 in -1,0 and 1, respectively, i.e.

$$F_X(x) = \begin{cases} 0 & \text{for } x \ < -1 \\ 0.49 & \text{for } -1 \leq x \ < 0 \\ 0.51 & \text{for } 0 \leq x \ < 1 \\ 1 & \text{for } x \ \geq 1 \end{cases}$$

2. $Y : Y(\omega_1) = Y(\omega_2) = 1, Y(\omega_3) = 2.7$. The distribution function is then partly constant with jumps of the size 0.98 and 0.02 in 1 and 2.7, respectively, i.e.

$$F_Y(x) = \begin{cases} 0 & \text{for } x \ < 1 \\ 0.98 & \text{for } 1 \leq x \ < 2.7 \\ 1 & \text{for } x \ \geq 2.7 \end{cases}$$

Exercise 17: *Consider*

$$f(x) = \begin{cases} cxe^{-2x}\,, & x \in (0,1), \\ 0\,, & otherwise\,. \end{cases}$$

Find the constant c such that $f(x)$ is probability density of a random variable.

Solution:
The basic property of probability density is

$$\int_{\mathbf{R}} f(x)dx = 1.$$

Using per partes method, we get

$$\int_{\mathbf{R}} f(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{1} cxe^{-2x}dx + \int_{1}^{\infty} 0dx = c \cdot \frac{1 - e^{-2}}{2},$$

thus $c = \frac{2}{1-e^{-2}}$.

Exercise 18: *Consider*

$$f(x) = \begin{cases} 3e^{-3x}\,, & x \in (0,\infty), \\ 0\,, & \textit{otherwise}\,. \end{cases}$$

1. *Verify that f is probability density.*
2. *Construct the corresponding distribution function.*
3. *Find $P(-1 \le X \le 1)$, where X is a random variable with the probability density f.*
4. *Calculate $\mathbb{E}X$ and* var *X.*

<u>Solution:</u>

1. The function f in non-negative and
$$\int_{\mathbf{R}} f(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{\infty} 3e^{-3x}dx = 1,$$
therefore it is probability density.

2. The corresponding distribution function is
$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} 0dx + \int_{0}^{x} 3e^{-3x}dx = 1 - e^{-3x} \text{ for } x \in (0,\infty)$$
and
$$F(x) = \int_{-\infty}^{x} 0dt = 0 \text{ for } x \le 0.$$

3. The probability is
$$P(-1 \le X \le 1) = \int_{-1}^{1} f(x)dx = \int_{-1}^{0} 0dx + \int_{0}^{1} 3e^{-3x}dx = 1 - e^{-3}.$$
We can find it also using the distribution function:
$$P(-1 \le X \le 1) = P(X \le 1) - P(X < -1) = P(X \le 1) - P(X \le -1) =$$
$$= F(1) - F(-1) = 1 - e^{-3\cdot 1} - 0 = 1 - e^{-3}.$$

4. Using per partes integration, we get

$$\mathbb{E}X = \int_{\mathbf{R}} xf(x)dx = \int_0^\infty 3xe^{-3x}dx = \frac{1}{3},$$

$$\mathbb{E}X^2 = \int_{\mathbf{R}} x^2 f(x)dx = \int_0^\infty 3x^2e^{-3x}dx = \frac{2}{9}.$$

So the expected value is $\mathbb{E}X = 1/3$ and the variance is

$$\text{var } X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{2}{9} - \frac{1}{9} = \frac{1}{9}.$$

Exercise 19: *Consider 6 runners in a sport club. The probability that a randomly chosen runner pass a given limit (e.g. 100 m in 12 s) is 0.7.*

1. *What is the distribution of the random variable X describing whether the runner pass ($X = 1$) or fail ($X = 0$) the limit? Calculate $\mathbb{E}X$ and* var X.
2. *What is the distribution of the random variable Y describing the number of runners passing the limit? Calculate $\mathbb{E}Y$ and* var Y.
3. *What is the probability that at least 4 runners pass the limit?*

Solution:

1. $X \sim Alt(0.7)$, i.e. $P(X = 1) = 0.7$ and $P(X = 0) = 0.3$.
   $\mathbb{E}X = 0.7$, $\text{var } X = 0.7 \cdot 0.3 = 0.21$.
2. $Y \sim Binom(6, 0.7)$, i.e. $P(Y = k) = \binom{6}{k}0.7^k 0.3^{6-k}$ for $k = 0, 1, ..., 6$.
   $\mathbb{E}Y = 6 \cdot 0.7 = 4.2$, $\text{var } Y = 6 \cdot 0.7 \cdot 0.3 = 1.26$.
3. $P(Y \geq 4) = \sum_{k=4}^{6} \binom{6}{k}0.7^k 0.3^{6-k} = \binom{6}{4}0.7^4 0.3^2 + \binom{6}{5}0.7^5 0.3^1 + \binom{6}{6}0.7^6 0.3^0$.

Exercise 20: *On a textile bar, there is in average one defect on 10 m of the bar. Suppose that the defects are independent on each other. Find the probability that on 50 m of the textile bar, there is*

1. *exactly 10 defects,*
2. *at most 3 defects,*
3. *exactly 5 defects while 4 of them are on the first 20 m?*

Solution:
Denote X the random variable describing the number of defects on 50m of the textile bar. Then

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, ...$$

Since $\mathbb{E}X = 5 \Rightarrow \lambda = 5$, i.e.

$$P(X = k) = \frac{5^k}{k!} e^{-5}, k = 0, 1, ...$$

Thus,

1. $P(X = 10) = \frac{5^{10}}{10!} e^{-5}$.
2. $P(X \leq 3) = e^{-5} \left(\frac{5^0}{0!} + \frac{5^1}{1!} + \frac{5^2}{2!} + \frac{5^3}{3!} \right)$.
3. Denote X_1 the random variable describing the number of defects on the first 20m and X_2 the random variable describing the number of defects on the remaining 30m. Analogously to the previous approach

$$P(X_1 = k) = \frac{\lambda_1^k}{k!} e^{-\lambda_1} \quad \text{and} \quad P(X_2 = k) = \frac{\lambda_2^k}{k!} e^{-\lambda_2},$$

where $\lambda_1 = 2$ and $\lambda_2 = 3$. Thus

$$P(X_1 = 4, X_2 = 1) = P(X_1 = 4)P(X_2 = 1) = \frac{2^4}{4!} e^{-2} \frac{3^1}{1!} e^{-3} = 2e^{-5}.$$

Exercise 21: *The probability that a born child is a boy is 0.51. A couple intend to continue having children until they have a girl. What is the probability that they will have at most 4 children?*

Solution:
Denote X the number of boys born before the first girl. Then

$$X \sim Geom(0.49), \text{ i.e. } P(X = k) = 0.51^k \cdot 0.49, \; k = 0, 1, ...,$$

thus

$$P(X < 4) = \sum_{k=0}^{3} 0.51^k \cdot 0.49 = 0.49 \sum_{k=0}^{3} 0.51^k = 0.49 \cdot \frac{1 - 0.51^4}{1 - 0.51} = 1 - 0.51^4.$$

An other approach:
Denote Y the random variable describing the number of girls born among the first four children. Then

$$Y \sim Binom(4, 0.49), \text{ i.e. } P(X = k) = \binom{4}{k} 0.49^k \cdot 0.51^{4-k}, \; k = 0, 1, 2, 3, 4.$$

Thus

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{4}{0} 0.49^0 \cdot 0.51^4 = 1 - 0.51^4.$$

Exercise 22: *During a numerical calculation, the numbers are rounded to one decimal place. What is the probability that the distance between the real number and the rounded number is greater than 0.04?*

Solution:
Denote X the random variable describing the distance between the real and the rounded number. Then

$$X \sim Ro(0, 0.05),$$

i.e. X has the density $f(x) = 20$ for $x \in (0, 0.05)$ and $f(x) = 0$ otherwise. Thus

$$P(X > 0.04) = 1 - P(X \leq 0.04) = 1 - \int_0^{0.04} 20 dx = 0.2.$$

Exercise 23: *The mean number of calls per hour received by a call centre is 12. The calls are independent on each other.*

1. *What is the probability that the call centre waits at least 10 minutes for the next call?*

2. *Find the time t such that the call centre waits at least t minutes for the next call with the probability 0.7.*

<u>Solution:</u>

1. There exist more ways of solutions again. First, denote X the random variable describing the time of waiting [min] for the next call. Then

$$X \sim Exp(1/5),$$

i.e. X has the density

$$\begin{aligned} f(x) &= \frac{1}{5}e^{-x/5} \quad x > 0 \\ &= 0 \quad x \leq 0. \end{aligned}$$

and the distribution function

$$\begin{aligned} F(x) &= 1 - e^{-x/5} \quad x > 0 \\ &= 0 \quad x \leq 0. \end{aligned}$$

The value $1/5$ follows from the fact that for $X \sim Exp(\lambda)$, it holds that $\mathrm{E}X = 1/\lambda$ while the mean waiting time is 5 minutes.
Thus

$$P(X > 10) = 1 - P(X \leq 10) = 1 - F(10) = 1 - (1 - e^{-10/5}) = e^{-2},$$

which can be also calculated as

$$P(X > 10) = \int_{10}^{\infty} \frac{1}{5}e^{-x/5}dx = e^{-2}.$$

An other approach:
Denote Y the random variable describing the number of calls per 10 minutes. Then

$$Y \sim Po(2), \text{ i.e. } P(Y = k) = \frac{2^k}{k!}e^{-2}, k = 0, 1, ...$$

The value 2 follows from the fact that for $Y \sim Po(\lambda)$, it holds that $\mathrm{E}X = \lambda$ while the mean number of calls per 10 minutes is equal to 2. Thus

$$P(Y = 0) = \frac{2^0}{0!} e^{-2} = e^{-2}.$$

2. Consider again X the random variable describing the time of waiting [min] for the next call. Then

$$\begin{aligned}
P(X > t) &= 0.7 \\
1 - P(X \leq t) &= 0.7 \\
1 - F(t) &= 0.7 \\
1 - (1 - e^{-t/5}) &= 0.7 \\
e^{-t/5} &= 0.7 \\
t &= -5 \ln 0.7 \\
t &= 1.78.
\end{aligned}$$

Exercise 24: *The height of children [in cm] in the first class of primary school is the random variable $X \sim N(130, 36)$. What is the probability that a randomly chosen child is*

1. *taller than 136 cm,*
2. *shorter than 118 cm,*
3. *between 127 cm and 133 cm tall?*

Solution:
Denote Z the random variable having standardised normal distribution, i.e. $Z \sim N(0, 1)$, and Φ its distribution function. Then

1.

$$\begin{aligned}
P(X > 136) &= P\left(\frac{X - 130}{\sqrt{36}} > \frac{136 - 130}{\sqrt{36}}\right) = P(Z > 1) = 1 - P(Z \leq 1) \\
&= 1 - \Phi(1) = 1 - 0.8413 = 0.1587.
\end{aligned}$$

2.

$$\begin{aligned} P(X < 118) &= P\left(\frac{X-130}{\sqrt{36}} < \frac{118-130}{\sqrt{36}}\right) = P(Z < -2) = \Phi(-2) \\ &= 1 - \Phi(2) = 1 - 0.9772 = 0.0228. \end{aligned}$$

3.

$$\begin{aligned} P(127 < X < 133) &= P\left(\frac{127-130}{\sqrt{36}} < \frac{X-130}{\sqrt{36}} < \frac{133-130}{\sqrt{36}}\right) \\ &= P(-0.5 < Z < 0.5) = P(Z < 0.5) - P(Z \leq -0.5) \\ &= \Phi(0.5) - \Phi(-0.5) = \Phi(0.5) - (1 - \Phi(0.5)) \\ &= 2\Phi(0.5) - 1 = 2 \cdot 0.6915 - 1 = 0.383. \end{aligned}$$

Exercise 25: *Let $X \sim U(0,2)$ and $Y = X^2 + 1$.*

1. *Find the distribution function of the random variable Y.*
2. *Find $cov(X,Y)$.*
3. *Are the random variables X and Y independent? Why?*

Solution:

1. The distribution function of X is

$$\begin{aligned} F_X(x) &= 0 & x < 0 \\ &= \frac{x}{2} & 0 \leq x \leq 2 \\ &= 1 & x > 2. \end{aligned}$$

The distribution function of Y is then

$F_Y(y) = P(Y \leq y) = P(X^2+1 \leq y) = P(X \leq \sqrt{y-1}) = F_X(\sqrt{y-1})$,

i.e.

$$\begin{aligned} F_Y(y) &= 0 & y < 1 \\ &= \frac{\sqrt{y-1}}{2} & 1 \leq y \leq 5 \\ &= 1 & y > 5. \end{aligned}$$

2. The covariance is

$$\begin{aligned}\operatorname{cov}(X,Y) &= \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = \mathbb{E}X(X^2+1) - \mathbb{E}X\mathbb{E}(X^2+1)\\ &= \mathbb{E}X^3 + \mathbb{E}X - \mathbb{E}X\mathbb{E}X^2 - \mathbb{E}X = \mathbb{E}X^3 - \mathbb{E}X\mathbb{E}X^2,\end{aligned}$$

   where

$$\begin{aligned}\mathbb{E}X &= \int_0^2 x\cdot\frac{1}{2}dx = 1,\\ \mathbb{E}X^2 &= \int_0^2 x^2\cdot\frac{1}{2}dx = \frac{4}{3},\\ \mathbb{E}X^3 &= \int_0^2 x^3\cdot\frac{1}{2}dx = 2.\end{aligned}$$

   Thus,

$$\operatorname{cov}(X,Y) = 2 - 1\cdot\frac{4}{3} = \frac{2}{3}.$$

3. The random variables X and Y are not independent $\Leftarrow \operatorname{cov}(X,Y)\neq 0$.

Exercise 26: *Join probability of the random variables X and Y is given by the following table:*

	X=0	*X=1*	*X=2*
Y=0	*1/4*	*1/8*	*0*
Y=1	*1/4*	*1/4*	*1/8*

1. *Find their marginal distributions.*
2. *Find their covariance matrix and correlation matrix.*
3. *Are the random variables X and Y independent? Why?*

Solution:

1. The marginal distribution of the random variable X is

$$P(X=0)=P(X=0,Y=0)+P(X=0,Y=1)=\frac{1}{4}+\frac{1}{4}=\frac{1}{2},$$
$$P(X=1)=P(X=1,Y=0)+P(X=1,Y=1)=\frac{1}{8}+\frac{1}{4}=\frac{3}{8},$$
$$P(X=2)=P(X=2,Y=0)+P(X=2,Y=1)=\frac{1}{8}+0=\frac{1}{8}.$$

The marginal distribution of the random variable Y is

$$P(Y=0)=\frac{1}{4}+\frac{1}{8}+0=\frac{3}{8},$$
$$P(Y=1)=\frac{1}{4}+\frac{1}{4}+\frac{1}{8}=\frac{5}{8}.$$

2. The covariance is $\operatorname{cov}(X,Y)=\mathbb{E}XY-\mathbb{E}X\mathbb{E}Y$:

$$\begin{aligned}
\mathbb{E}X =& 0\cdot\frac{1}{2}+1\cdot\frac{3}{8}+2\cdot\frac{1}{8}=\frac{5}{8},\\
\mathbb{E}Y =& 0\cdot\frac{3}{8}+1\cdot\frac{5}{8}=\frac{5}{8},\\
\mathbb{E}XY =& 0\cdot 0\cdot\frac{1}{4}+0\cdot 1\cdot\frac{1}{4}+1\cdot 0\cdot\frac{1}{8}+1\cdot 1\cdot\frac{1}{4}+2\cdot 0\cdot 0+2\cdot 1\cdot\frac{1}{8}=\frac{1}{2},\\
\operatorname{cov}(X,Y) =& \mathbb{E}XY-\mathbb{E}X\mathbb{E}Y=\frac{1}{2}-\frac{5}{8}\cdot\frac{5}{8}=\frac{7}{64}.
\end{aligned}$$

Further, since

$$\begin{aligned}
\mathbb{E}X^2 =& 0^2\cdot\frac{1}{2}+1^2\cdot\frac{3}{8}+2^2\cdot\frac{1}{8}=\frac{7}{8},\\
\mathbb{E}Y^2 =& 0^2\cdot\frac{3}{8}+1^2\cdot\frac{5}{8}=\frac{5}{8},
\end{aligned}$$

the variances are

$$\begin{aligned}
\operatorname{var}X =& \mathbb{E}X^2-(\mathbb{E}X)^2=\frac{7}{8}-\left(\frac{5}{8}\right)^2=\frac{31}{64},\\
\operatorname{var}Y =& \mathbb{E}Y^2-(\mathbb{E}Y)^2=\frac{5}{8}-\left(\frac{5}{8}\right)^2=\frac{15}{64}.
\end{aligned}$$

The covariance matrix is then

$$\mathrm{Cov}(X,Y) = \begin{pmatrix} \frac{31}{64} & \frac{7}{64} \\ \frac{7}{64} & \frac{15}{64} \end{pmatrix}.$$

The correlation of X and Y is

$$corr(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}\ X}\sqrt{\mathrm{var}\ Y}} = \frac{7/64}{\sqrt{31/64}\sqrt{15/64}} = \frac{7}{\sqrt{465}},$$

so the correlation matrix is

$$\rho(X,Y) = \begin{pmatrix} 1 & \frac{7}{\sqrt{465}} \\ \frac{7}{\sqrt{465}} & 1 \end{pmatrix}.$$

3. The random variables X and Y are not independent. One of the reasons is $\mathrm{cov}(X,Y) \neq 0$. An other reason is that e.g.

$$P(X=0, Y=0) = \frac{1}{4} \neq P(X=0)P(Y=0) = \frac{1}{2} \cdot \frac{3}{8}.$$

(Note that in case of independence, it must hold that $P(X=i, Y=j) = P(X=i)P(Y=j)$, $\forall i,j$.)

Exercise 27: *Join density of the random variables X and Y is*

$$f_{(X,Y)}(x,y) = \begin{cases} \frac{1}{2}e^{-x-\frac{y}{2}}\,, & x>0, y>0\,, \\ 0\,, & otherwise\,. \end{cases}$$

1. *Find their marginal distributions.*
2. *Are the random variables X and Y independent? Why?*
3. *Find their covariance matrix and correlation matrix.*

Solution:

1. The marginal densities are

$$\begin{aligned} f_X(x) &= \textstyle\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \\ &= \textstyle\int_0^{\infty} \frac{1}{2}e^{-x-\frac{y}{2}}dy = \frac{1}{2}e^{-x}\cdot[-2e^{-\frac{y}{2}}]_0^{\infty} = e^{-x} \text{ for } x>0, \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \textstyle\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \\ &= \textstyle\int_0^{\infty} \frac{1}{2}e^{-x-\frac{y}{2}}dx = \frac{1}{2}e^{-\frac{y}{2}}\cdot[-e^{-x}]_0^{\infty} = \frac{1}{2}e^{-\frac{y}{2}} \text{ for } y>0, \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

2. The random variables X and Y are independent if and only if $f_{X,Y}(x,y) = f_X(x)\cdot f_Y(y), \forall x,y$. From 1., it follows that this equality is satisfied.

3. Since $X \sim Exp(1)$ and $Y \sim Exp(1/2) \Rightarrow \text{var } X = 1$ and $\text{var } Y = 4$. From the independence of X and Y, it follows that $\text{cov}(X,Y) = 0 \Rightarrow \text{corr}(X,Y) = 0 \Rightarrow$
$$\text{Cov}_{(X,Y)} = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$$
and
$$\rho_{(X,Y)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Exercise 28: *A store chain has 100 shops. The probability that a shop will be in the black at the end of the year, is 0.9. Find the probability that at the end of the year, at least 85 shops will be in the black.*

Solution:
We use CLT (Central Limit Theorem):

$$P(\frac{\sum_{i=1}^{n} X_i - n\mathbb{E}X_1}{\sqrt{n\text{var } X_1}} \leq a) \doteq \Phi(a).$$

Denote $X_i = 1$ if the $i-$th shop is in the black and $X_i = 0$ if the $i-$th shop is in the lost at the end of the year, i.e. $X_i \sim Bern(0.9) \Rightarrow \mathbb{E}X_1 = 0.9$ and var $X_1 = 0.9 \cdot 0.1 = 0.09$, $i = 1, 2, \ldots, 100$. Further denote

$$Z = \frac{\sum_{i=1}^{100} X_i - 100 \cdot 0.9}{\sqrt{100 \cdot 0.09}}$$

Then

$$P(\sum_{i=1}^{100} X_i \geq 85) = P\left(\frac{\sum_{i=1}^{100} X_i - 100 \cdot 0.9}{\sqrt{100 \cdot 0.09}} \geq \frac{85 - 100 \cdot 0.9}{\sqrt{100 \cdot 0.09}}\right) = P(Z \geq -1.\bar{6})$$
$$= 1 - P(Z < -1.\bar{6}) \doteq 1 - \Phi(-1.\bar{6}) = 1 - (1 - \Phi(1.\bar{6}))$$
$$= \Phi(1.\bar{6}) \doteq 0.953.$$

Exercise 29: *I use tram number 8 for traveling to the faculty. The time intervals between tram arrivals are 10 minut. Find the probability that during 24 working days when traveling to the faculty and back home, the total time of waiting for the tram is shorter than 3 hours?*

Solution:
We use CLT again. Denote X_i the time of waiting for a tram during the $i-$th journey, $i = 1, 2, \ldots, 48$. Then $X_i \sim U(0, 10) \Rightarrow \mathbb{E}X_1 = 5$ and var $X_1 = \frac{25}{3}$. Further, denote

$$Z = \frac{\sum_{i=1}^{48} X_i - 48 \cdot 5}{\sqrt{48 \cdot \frac{25}{3}}}.$$

Then

$$P(\sum_{i=1}^{48} X_i \leq 180) = P\left(\frac{\sum_{i=1}^{48} X_i - 48 \cdot 5}{\sqrt{48 \cdot \frac{25}{3}}} \leq \frac{180 - 48 \cdot 5}{\sqrt{48 \cdot \frac{25}{3}}}\right) = P(Z \leq -3) \doteq$$
$$\doteq \Phi(3) = 1 - \Phi(3) = 1 - 0.9987 = 0.0013.$$

Exercise 30: *Consider the following data:*

1. *the number of given plants on a surface of the area 1* m^2*: 0, 2, 1, 4, 4, 5, 2, 3, 7;*

2. *time intervals between two spikes in a brain: 4.25, 0.65, 1.35, 0.20, 0.55, 6.63, 1.38, 0.22, 0.27;*
3. *the highest temperature during an annual November celebration: 8.07, 19.23, 9.27, 5.71, 12.62, 11.24, 11.92, 17.30, 14.87.*

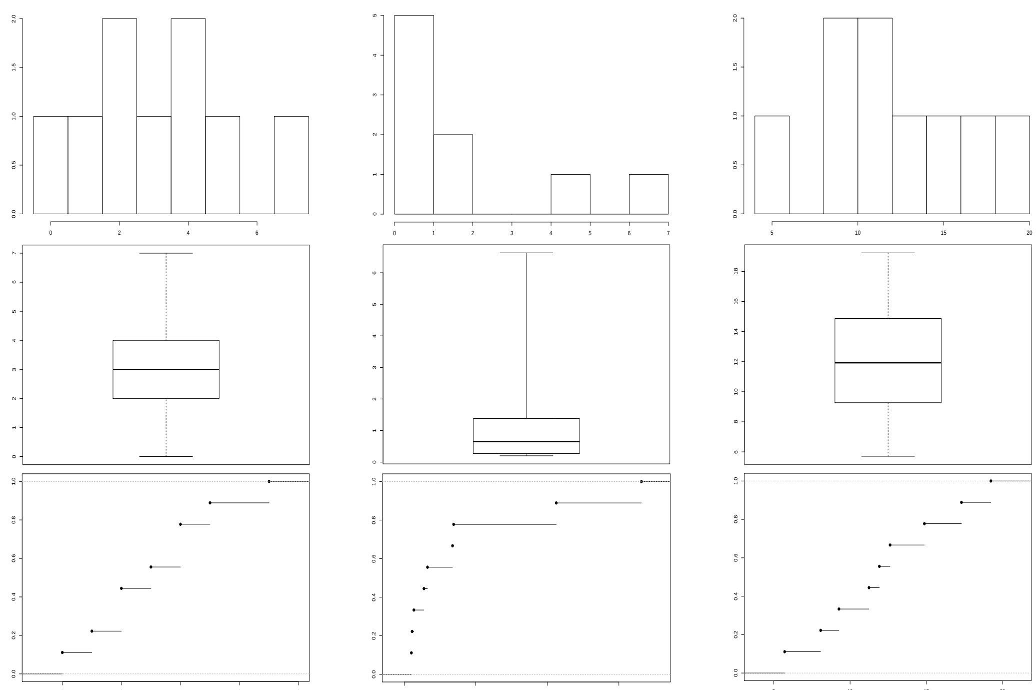
For the data, draw their

1. *histograms,*
2. *boxplots,*
3. *empirical distribution function*

and estimate their distribution.

Solution:

1. 2. 3.



1. The distribution is discrete and the number of the plants on the given area is theoretically not bounded $\Rightarrow$ data $\sim Po(\lambda)$.

2. The distribution is continuous and the shape of the histogram is similar to the density of exponential distribution (or the empirical distribution function is similar to the theoretical distribution function of exponential distribution) $\Rightarrow$ data $\sim Exp(\lambda)$.

3. The distribution is continuous and the shape of the histogram is similar to the Gaussian curve, i.e density of normal distribution $\Rightarrow$ data $\sim N(\mu, \sigma^2)$.

Exercise 31: *Time intervals between two breakdowns of a device were 4 days, 7 days, 12 days, 2.5 days a 24.5 days. The time intervals are supposed to come from exponential distribution. Use the maximum likelihood method to estimate the parameter λ.*

Solution:
For the random variable X describing the time intervals between two breakdowns, it holds $X \sim Exp(\lambda)$, so we have $f(x) = \lambda e^{-\lambda x}$ for $x > 0$.
Likelihood function is

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda e^{-\lambda \cdot 4} \lambda e^{-\lambda \cdot 7} \lambda e^{-\lambda \cdot 12} \lambda e^{-\lambda \cdot 2.5} \lambda e^{-\lambda \cdot 24.5} = \lambda^5 e^{-\lambda \cdot 50}.$$

Log-likelihood function is

$$l(\lambda) = \ln L(\lambda) = 5 \ln \lambda - 50\lambda.$$

Its derivation is

$$\frac{\partial l(\lambda)}{\partial \lambda} = \frac{5}{\lambda} - 50.$$

So the solution is

$$\frac{5}{\lambda} - 50 = 0 \Rightarrow \hat{\lambda} = \frac{1}{10}.$$

Exercise 32: *During influenza epidemic, the number of patients visiting a doctor was observed. The average number calculated from data from 64 doctors was 23 patients per day while the sample variance was equal to 36. The distribution of the number of patients is unknown. Construct 95%-CI (asympotic confidence interval) of the expected number of patients per day.*

Solution:
From CLT, we can derive that $(1-\alpha)\cdot 100\%$-CI for $\mathbb{E}X$ is

$$(\bar{X}_n - u_{1-\alpha/2}\frac{S_n}{\sqrt{n}}, \bar{X}_n + u_{1-\alpha/2}\frac{S_n}{\sqrt{n}}).$$

We have $n = 64$, $\bar{X}_{64} = 23$, $S_{64}^2 = 36$, $\alpha = 0.05$ and in statistical tables, we can find $u_{0.975} = 1.96$. Thus,

$$95\%\, CI = \left(23 - 1.96 \cdot \frac{6}{8}, 23 - 1.96 \cdot \frac{6}{8}\right) = (21.53, 24.47).$$

Exercise 33: *Observed heights of volleyball players in a team are:*

174	186	181	189	195	197	188	177	188	192	181	188	201	180	183

1. *Estimate the sample mean and the sample variance ($\sum x_i = 2800$, $\sum(x_i - \bar{x})^2 = 797.33$).*

2. *Test whether the mean height of a randomly chosen player is 190cm.*

Solution:

1. $\bar{X}_{15} = 2800/15 = 186.\bar{6}$, $S_{15}^2 = 797.33/14 \doteq 57$.

2. We want to test

   H_0 : mean height = 190cm,
   H_A : mean height $\neq$ 190cm.

   Under the null hypothesis H_0, it holds that

   $$T = \frac{\bar{X}_n - 190}{\sqrt{57}}\sqrt{15} \sim t_{14}.$$

   For our data, we have

   $$T = \frac{186.\bar{6} - 190}{\sqrt{57}} \cdot \sqrt{15} \doteq -1.7.$$

Since $|T| < t_{14;0.975} = 2.14$, we do not reject H_0 with respect to H_A at the level $\alpha = 5\%$.

Note that it makes also sense to test

H_0 : mean height $=$ 190cm,
H_A : mean height $\leq$ 190cm.

Since $|T| < t_{14;0.95} = 1.76$, we again do not reject H_0 with respect to H_A at the level $\alpha = 5\%$.

Exercise 34: *We observed the colours of hair and eyes of 100 people. The joint frequencies were:*

eyes / hair	dark	light
blue	10	20
green or grey	10	10
brown	40	10

Are the colour of hair and eyes independent?

Solution:
We want to test

H_0 : the colours are independent,
H_A : they are not independent.

Under the null hypothesis H_0, it holds that

$$\sum_{i=1}^{3}\sum_{i=1}^{2}\frac{\left(n_{ij}-\frac{n_{i.}\cdot n_{.j}}{n}\right)^2}{\frac{n_{i.}\cdot n_{.j}}{n}} \sim \chi^2_{(3-1)(2-1)} = \chi^2_2.$$

We add the marginal sums to the table:

eyes / hair	dark	light	marginal sum
blue	10	20	$n_{1.} = 30$
green or grey	10	10	$n_{2.} = 20$
brown	40	10	$n_{3.} = 50$
marginal sum	$n_{.1} = 60$	$n_{.2} = 40$	$n = 100$

and calculate

$$\chi^2 = \frac{(10 - \frac{30\cdot 60}{100})^2}{\frac{30\cdot 60}{100}} + \frac{(20 - \frac{30\cdot 40}{100})^2}{\frac{30\cdot 40}{100}} + \frac{(10 - \frac{20\cdot 60}{100})^2}{\frac{20\cdot 60}{100}} + \frac{(10 - \frac{20\cdot 40}{100})^2}{\frac{20\cdot 40}{100}} + \frac{(40 - \frac{50\cdot 60}{100})^2}{\frac{50\cdot 60}{100}} + \frac{(10 - \frac{50\cdot 40}{100})^2}{\frac{50\cdot 40}{100}} \doteq 18.$$

Since $\chi^2 > \chi^2_{0.95,2} = 5.992$, we reject H_0 with respect to H_A at the level $\alpha = 5\%$.

Exercise 35: *A company has three offices. For two years, the company was observing which office has the best economical result in the actual month. The first office had the best result ten times, the second one six times and the third one eight times. Are the offices approximately equally successful?*

Solution:
Denote p_i the probability that the ith office has the best economical result in the actual month, and X_i the number of months when the ith office has the best economical result. Then we want to test

$H_0 : p_1 = p_2 = p_3 = \frac{1}{3}$,
H_A : at least one p_i

Under the null hypothesis H_0, it holds that

$$\sum_{i=1}^{3} \frac{(X_i - 24 \cdot p_i)^2}{24 \cdot p_i} \sim \chi^2_2.$$

We calculate

$$\chi^2 = \frac{(10-8)^2}{8} + \frac{(6-8)^2}{8} + \frac{(8-8)^2}{8} = 1.$$

Since $\chi^2 < \chi^2_{2;0.95} = 5.992$, we do not reject H_0 with respect to H_A at the level $\alpha = 5\%$.

Exercise 36: *Every morning, a quest in a hotel has breakfast. He chooses the breakfast from three meals, let's denote them 1, 2 and 3. If he takes the* *meal 1 or 3, he will take the same meal the next day with probability 0.4. If he takes the meal 2, he will take the same meal in the next day with probability 0.6. In all three cases, it holds that if he change the meal in the next day, the probability is uniformly divided between the two remaining possibilities. The process* $\{X_n, n \in \mathbb{N}_0\}$ *describes the sequence of chosen breakfasts, where* X_n *is the number of the meal chosen in the* n*-th day.*

1. *Construct the matrix of transit probabilities.*
2. *Find the stationary distribution.*