Proposed Problems

Seminar 1

1. A home computer is tied to a mainframe computer via a telephone modem. The home computer will dial repeatedly until contact is made. Let c denote the fact that contact is made on a particular attempt, and n that contact is not made.
a) List the sample space of this experiment.
b) List the elementary events that constitute the following events:
A: contact is made in at most five attempts;
B: contact is made after at least 10 tries.

2. Consider a game of darts consisting of 7 concentric circles of radii $r_1 < r_2 < ... < r_7$. Let A_i be the event: the circle C_i of radius r_i is hit. Describe (in words) the following events:
a) $A = A_1 \cup ... \cup A_5$;
b) $B = A_4 \cap ... \cap A_7$;
c) $C = \overline{A_2} \cap A_3$;
d) $D = A_5 \setminus A_4$;
e) $E = A_4 \triangle A_5$ $(A \triangle B = (A \cup B) \setminus (A \cap B))$.

3. A computer system uses passwords that consist of 5 letters followed by 2 digits. There is no difference between lowercase and capital letters.
a) How many passwords are possible? (Ans: $(26)^5(10)^2$)
b) How many passwords are possible, without repetition of characters?
(Ans: $26 \cdot 25 \cdot 24 \cdot 23 \cdot 22 \cdot 10 \cdot 9$)
c) How many passwords begin with letter A and end in 5? (Ans: $(26)^4 \cdot 10$)
d) How many passwords contain the letters A, B, C (in that order) and end in an even digit?
(Ans: $3 \cdot (26)^2 \cdot 10 \cdot 5$)

4. There are 6 boys and 6 girls in a study group. For a certain project they have to work in teams of two people.
a) In how many ways can they be teamed up? (Ans: $C_{12}^2 \cdot C_{10}^2 \cdot C_8^2 \cdot C_6^2 \cdot C_4^2 \cdot C_2^2 = \frac{12!}{2^6}$)
b) In how many of those, each team consists of a boy and a girl? (Ans: $6^2 \cdot 5^2 \cdot 4^2 \cdot 3^2 \cdot 2^2 \cdot 1^2 = (6!)^2$)

Seminar 2

1. The natural numbers $1, 2, ..., n$ (where n is a fixed natural number) are placed randomly in a sequence. Find the probability of the following events:
a) A: the numbers 1 and 2 are placed consecutively, in increasing order; (Ans: $\frac{1}{n}$)
b) B: the numbers 1 and 2 are placed consecutively; (Ans: $\frac{2}{n}$)
c) C: the numbers 1 and 2 are placed in increasing order; (Ans: $\frac{1}{2}$)
d) D: the numbers i, j, k are placed consecutively, in increasing order. (Ans: $\frac{1}{n(n-1)}$)

2. A computer program is tested by 3 independent tests. When there is an error, these tests will discover it with probabilities 0.2, 0.3 and 0.5, respectively. Suppose that the program contains an error. What is the probability that it will be found by at least one test? (Ans: 0.72)

3. In a study of waters near industrial plants it was found that 30% showed signs of chemical pollution (event C), 25% showed evidence of thermal pollution (event T) and 10% showed signs of chemical and thermal pollution.
a) From the data above, is chemical pollution independent of thermal pollution in that area? (Ans: No)
b) What is the probability that a stream that shows some thermal pollution will also show signs of chemical pollution (event A)? (Ans: $\frac{2}{5}$)
c) What is the probability that a stream showing chemical pollution will not show signs of thermal pollution (event B)? (Ans: $\frac{2}{3}$)

4. Three highways Connect city A from city B and two highways connect city B to city C. During a rush hour, the traffic on each highway is blocked with probability 0.2, independently of other highways.
a) Compute the probability that there is at least one open route from A to C. (Ans: 0.9523)
b) How will a new highway, also blocked with probability 0.2 independently of other highways, change that probability if it is built
  b1) between B and C? (Ans: 0.9841)
  b2) between A and C? (Ans: 0.991)

5. A system may become infected by some spyware through the internet or e-mail. Seventy percent of the time the spyware arrives via the internet, the rest of the time via e-mail. If it enters via the internet, the system detects it with probability 0.6, whereas if via e-mail, it is detected with probability 0.8. What percentage of times is the spyware detected? (Ans: 66%)

Seminar 3

1. A person has 40 homing (messenger) pigeons. When released, the probability that a pigeon will come back is 0.7. Find the probability of the events:
a) A: 10 pigeons do not come back; (Ans: $C_{40}^{10}(0.3)^{10}(0.7)^{30} = 0.1128$)
b) B: all of them come back; (Ans: $(0.7)^{40} = 6.3668e - 07$)
c) C: at least 38 of them come back. (Ans: $\sum\limits_{k=38}^{40} C_{40}^{k}(0.7)^{k}(0.3)^{40-k} = 1.0277e - 04$)

2. Successful implementation of a new system is based on three independent modules. Module 1 works properly with probability 0.96. For modules 2 and 3, these probabilities are 0.95 and 0.9, respectively. Find the probability that
a) exactly one of these modules fails to work properly; (Ans: 0.1686)
b) the system does not work properly. (Ans: 0.1792)

3. An internet search engine looks for a keyword in 9 databases, searching them in random order. Only 5 of these databases contain the given keyword. Find the probability that it will be found in at least 2 of the first 4 searched databases. (Ans: 0.8333)

4. A flu vaccine meets specifications with probability 0.9. Would it be unusual if 7 or more vaccines have to be tested to find three that meet specifications (event A)? Explain. (Ans: Yes, the probability of that event is 0.00127, very small)

5. Students from 3 departments participate in a debate. The boys/girls ratios for the 3 groups are $(6,4)$, $(4,5)$ and $(5,5)$, respectively. If one student is chosen from each department as spokesperson, what is the probability that the spokespersons are 2 boys and 1 girl (ev. A)? (Ans: $\frac{7}{18}$)

6. An internet search engine looks for a keyword in a sequence of independent web sites. It is believed that 20% of the sites contain this keyword. Find the probability that
a) at least 5 of the first 10 sites contain the given keyword; (Ans: 0.0328)
b) the search engine had to visit at least 5 sites in order to find the first occurrence of the keyword (Ans: 0.4096)

Seminar 4

1. A computer virus attacks a folder consisting of 250 files. Files are affected by the virus independently of one another, each with probability 0.032. Let X denote the number of infected files.
a) Find the pdf of X. What type of distribution does X have? (Ans: $B(250, 0.032)$)
b) What is the probability that more than 7 files are affected by this virus? (Ans: 0.5493)

2. Messages arrive at an electronic message center at random times, with an average of 9 messages per hour. What is the probability of receiving
a) exactly 5 messages during the next hour? (Ans: 0.061)
b) at least 5 messages during the next hour? (Ans: 0.945)

3. Two out of six computers in a lab have problems with hard drives. Three computers are selected at random for inspection. Let X denote the number of computers that are found to have hard drive problems.
a) Find the pdf of X. What type of distribution does X have? (Ans: $H(6, 2, 3)$)
b) What is the probability that at most 1 of them has hard drive problems? (Ans: 0.8)

4. Eight letters are randomly distributed into 3 mailboxes. Let X be the number of letters in the 1^{st} mailbox. Find the pdf of X.

5. On any day, in a small computer lab, the number of hardware failures, X, and the number of software failures, Y, have the joint distribution $P(x, y)$, with $P(0,0) = 0.6, P(0,1) = 0.1, P(1,0) = 0.1, P(1,1) = 0.2$.
a) Write the joint pdf of the vector (X, Y);
b) Find the pdf's of X and Y, respectively;
c) Are hardware and software failures independent in this lab? (Ans: No)
d) Find the probability that no software failures occur on a given day; (Ans: 0.7)
e) Find the probability that 1 hardware failure and at most 1 software failure occur on a given day. (Ans: 0.3)

6. Network breakdowns are unexpected rare events that occur every 3 weeks, on average. Find the probability of more than 4 breakdowns occurring during a 21-week period. (Ans: 0.827)

Seminar 5

1. The time, in minutes, it takes to reboot a certain system is random variable with density

$$f(x) = \begin{cases} C(10-x)^2, & 0 < x < 10 \\ 0, & \text{otherwise} \end{cases}$$

Find
a) the constant C; (Ans: $C = 0.003$)
b) the probability that it takes between 1 and 2 minutes to reboot the system. (Ans: 0.217)

2. The random vector (X, Y) has the joint pdf given by

$$f(x, y) = k(x^2 + y), \quad \text{for } -1 \le x \le 1,\ 0 \le y \le 1.$$

Find
a) the constant k; (Ans: $k = 3/5$)
b) the marginal densities of X and Y; are X and Y independent? (Ans: No)
c) the probabilities $P(Y < 0.6)$ and $P(Y < 0.6 \mid X < 0.5)$ (Ans: 0.456 and 0.44)

3. Consider a satellite whose work is based on a certain block A. This block has an independent backup B. The satellite performs its task until both A and B fail. The lifetime of A and B are exponentially distributed random variables with average lifetime of 10 years. What is the probability that the satellite will work for more than 10 years? (Ans: 0.6004)

4. Let $X \in N(0, 1)$. Find the probability density function of $Z = X^2 - 1$.

5. Let X be a random variable with density $f_X(x) = xe^{-x}$, $x \ge 0$ and let $Y = e^X$. Find f_Y.

Seminar 6

1. (Refer to proposed problem 5, for Seminar 4) On any day, in a small computer lab, the number of hardware failures, X, and the number of software failures, Y, have the joint distribution $P(x, y)$, with $P(0, 0) = 0.6, P(0, 1) = 0.1, P(1, 0) = 0.1, P(1, 1) = 0.2$.
a) Are hardware and software failures independent in this lab? (Ans: No)
b) What is the expected total number of failures during one day? (Ans: 0.6)

2. (Optimal portfolio) Shares of company A are sold at \$10 per share and shares of company B are sold at \$50 per share. According to a market analyst, 1 share of each company can either gain \$1, with probability 0.5, or lose \$1, with the same probability, independently of the other company. Which of the following portfolios has the lowest risk:
a) 100 shares of company A; (Ans: Var = 10000)
b) 50 shares of company A + 10 shares of company B; (Ans: Var = 2600)
c) 40 shares of company A + 12 shares of company B? (Ans: Var = 1744.)

3. Lifetime (in years) of a certain hardware is a continuous random variable with pdf

$$f(x) = \begin{cases} C - x/50, & 0 < x < 10 \\ 0, & \text{otherwise} \end{cases}$$

Find
a) the constant C; (Ans: $C = 0.2$)
b) how long is this hardware expected to last. (Ans: $\dfrac{10}{3}$ years)

4. An internet search engine looks for a certain keyword in a sequence of independent web sites. It is believed that 20% of the sites contain this keyword. On the average, how many of the first 15 sites will contain that keyword?

5. (Refer to proposed problem 2, for Seminar 5) The random vector (X, Y) has the joint pdf given by

$$f(x,y) = \frac{3}{5}(x^2 + y), \quad \text{for } -1 \le x \le 1, \ 0 \le y \le 1.$$

Find
a) the means and variances of X and Y; $\left(\text{Ans} : E(X) = 0, E(Y) = \dfrac{3}{5}, V(X) = \dfrac{11}{25}, V(Y) = \dfrac{11}{150}\right)$

b) the correlation coefficient $\rho(X, Y)$. (Ans: $\rho(X, Y) = 0$, X and Y **uncorrelated**, but, as seen in proposed problem 2 (Sem 5), they are **NOT independent**)

Seminar 7

1. Let X be a r. v. with pdf $f(x) = \dfrac{x^m e^{-x}}{m!}$, $x \ge 0$. Show that $P(0 < X < 2(m+1)) \ge \dfrac{m}{m+1}$.

2. An average scanned image occupies 0.6 megabytes of memory with a standard deviation of 0.4 megabytes. If you plan to publish 80 images on your website, what is the probability that their total size is between 47 and 50 megabytes? (Ans: 0.322)

3. Every day, George takes the same street from his home to the university. There are 4 street lights along his way, and he noticed the following Markov dependence: if he sees a green light at an intersection, then 60% of the time the next light is also green, whereas if he sees a red light, then 70% of the time the next light is also red.
a) Find the transition probability matrix for the street lights;
b) If the first light is green, what is the probability that the third light is red? (Ans: 0.52)
c) George's classmate John has many street lights between his home and the university, all following the same pattern that George noticed. If the *first* street light is green, what is the probability that the *last* street light is red? (Ans: $\dfrac{4}{7} \approx 0.5714$)

4. The probability that a certain computer code runs without errors is a parameter $p \in (0, 1)$, unknown. Let X denote the number of times the computer code has to be debugged before it runs without errors. For 5 independent computer projects, a student records the number of times they each had to be corrected as: $3, 7, 5, 3, 2$. Estimate p
a) by the method of moments; (Ans: $\overline{p} = 0.2$)
b) by the method of maximum likelihood. (Ans: $\hat{p} = 0.2$)

5. Let $X_1, \ldots, X_n$ be a sample drawn from a distribution with pdf

$$f(x;\lambda) = \frac{1}{\lambda} e^{-\frac{1}{\lambda}x}, \ x > 0$$

($\lambda > 0$), which has mean $\mu = E(X) = \lambda$ and variance $\sigma^2 = V(X) = \lambda^2$. Find
a) the method of moments estimator, $\hat{\lambda}$, for λ; (Ans: $\hat{\lambda} = \overline{X}$)
b) the efficiency of $\hat{\lambda}$, $e(\hat{\lambda})$; (Ans: $e(\hat{\lambda}) = 1$)
c) an approximation for the standard error of the estimate in a), $\sigma_{\hat{\lambda}}$, if the sum of 100 observations is 150. (Ans: $\sigma_{\hat{\lambda}} \approx 0.15$)

Seminar Nr.1, Euler's Functions; Counting, Outcomes, Events

Theory Review

Euler's Gamma Function: $\Gamma : (0, \infty) \to (0, \infty), \Gamma(a) = \int\limits_0^\infty x^{a-1}e^{-x}dx.$

1. $\Gamma(1) = 1;$
2. $\Gamma(a + 1) = a\Gamma(a)$, $\forall a > 0;$
3. $\Gamma(n + 1) = n!$, $\forall n \in \mathbb{N};$
4. $\Gamma\left(\frac{1}{2}\right) = \sqrt{2}\int\limits_0^\infty e^{-\frac{t^2}{2}}dt = \int\limits_{\mathbb{R}} e^{-t^2}dt = \sqrt{\pi}.$

Euler's Beta Function: $\beta : (0, \infty) \times (0, \infty) \to (0, \infty), \beta(a, b) = \int\limits_0^1 x^{a-1}(1 - x)^{b-1}dx.$

1. $\beta(a, 1) = \frac{1}{a}$, $\forall a > 0;$
2. $\beta(a, b) = \beta(b, a)$, $\forall a, b > 0;$
3. $\beta(a, b) = \frac{a - 1}{b}\beta(a - 1, b + 1)$, $\forall a > 1, b > 0;$
4. $\beta(a, b) = \frac{b - 1}{a + b - 1}\beta(a, b - 1) = \frac{a - 1}{a + b - 1}\beta(a - 1, b)$, $\forall a > 1$, $b > 1;$
5. $\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}$, $\forall a > 0, b > 0.$

Arrangements: $A_n^k = \frac{n!}{(n - k)!};$

Permutations: $P_n = A_n^n = n!;$

Combinations: $C_n^k = \frac{A_n^k}{P_k} = \frac{n!}{k!(n - k)!}.$

De Morgan's laws:

$$\overline{\bigcup_{i\in I} A_i} = \bigcap_{i\in I} \overline{A}_i \quad \text{and} \quad \overline{\bigcap_{i\in I} A_i} = \bigcup_{i\in I} \overline{A}_i.$$

---

1. In how many ways can 10 students be seated in a classroom with
a) 15 chairs?
b) 10 chairs?

2. Find the number of possible outcomes for the following events:
a) three dice are rolled;
b) two letters and three digits are randomly selected.

3. A firm offers a choice of 10 free software packages to buyers of their new home computer. There are 25 packages available, five of which are computer games, and three of which are anti-virus programs.
a) How many selections are possible?
b) How many selections are possible, if exactly three computer games are selected?
c) How many selections are possible, if exactly three computer games and exactly two anti-virus programs are selected?

4. A person buys n lottery tickets. For $i = \overline{1, n}$, let A_i denote the event: the i^{th} ticket is a winning one. Express the following events in terms of $A_1, ..., A_n$.
a) A: all tickets are winning;

b) B: all tickets are losing;
c) C: at least one is winning;
d) D: exactly one is winning;
e) E: exactly two are winning;
f) F: at least two are winning;
g) G: at most two are winning.

5. Three shooters aim at a target. For $i = \overline{1,3}$, let A_i denote the event: the i^{th} shooter hits the target. Express the following events in terms of A_1, A_2 and A_3.
a) A: the target is hit;
b) B: the target is not hit;
c) C: the target is hit exactly three times;
d) D: the target is hit exactly once;
e) E: the target is hit exactly twice.

Seminar Nr.2, Classical Probability; Rules of Probability; Conditional Probability; Independent Events

__Theory Review__

Classical Probability: $P(A) = \dfrac{\text{nr. of favorable outcomes}}{\text{total nr. of possible outcomes}} = \dfrac{N_f}{N_t}$.

Mutually Exclusive Events: A, B **m. e. (disjoint, incompatible)** $<=> P(A \cap B) = 0$.

Rules of Probability:

$P(\overline{A}) = 1 - P(A)$;
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
$P(A \setminus B) = P(A) - P(A \cap B)$.

Conditional Probability: $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$, $P(B) \neq 0$.

Independent Events: A, B **ind.** $<=> P(A \cap B) = P(A)P(B) <=> P(A|B) = P(A)$.

Total Probability Rule: $\{A_i\}_{i \in I}$ a partition of S, then $P(E) = \sum\limits_{i \in I} P(A_i)P(E|A_i)$.

Multiplication Rule: $P\left(\bigcap\limits_{i=1}^{n} A_i\right) = P(A_1)\, P(A_2|A_1)\, P(A_3|A_1 \cap A_2)\, \ldots\, P\left(A_n|\bigcap\limits_{i=1}^{n-1} A_i\right)$.

1.The faces of a cube are painted each in a different color (the cube is transparent on the inside). Then the cube is broken into 1000 smaller, equally-sized cubes and one such cube is randomly picked. Find the probability of the following events:
a) A: the cube picked has exactly three colored faces;
b) B: the cube picked has exactly two colored faces;
c) C: the cube picked has exactly one colored face;
d) D: the cube picked has no colored faces.

2. (Pigeonhole Principle) A postman distributes n letters in N mailboxes. What is the probability of the event A: there are m letters in a given (fixed) mailbox ($0 \leq m \leq n$)?

3. (Breaking Passwords) An account uses 8-character passwords, consisting of letters (distinguishing between lower-case and capital letters) and digits. A spy program can check about 1 million passwords per second.
a) On the average, how long will it take the spy program to guess your password?
b) What is the probability that the spy program will break your password within a week?
c) Same questions, if capital letters are not used.

4. (System Reliability) Compute the reliability of the system in Figure 1 if each of the five components is operable with probability 0.92, independently of each other.

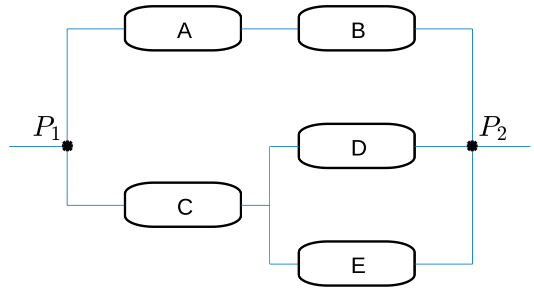

Figure 1: System Reliability

5. Among employees of a certain firm, 70% know C/C++, 60% know Fortran and 50% know both. What portion of programmers
a) does not know Fortran?
b) does not know C/C++ and does not know Fortran?
c) knows C/C++, but not Fortran?
d) Are "knowing C/C++" and "knowing Fortran" independent of each other?
e) What is the probability that someone who knows Fortran, also knows C/C++?
f) What is the probability that someone who knows C/C++, does not also know Fortran?

6. Three shooters aim at a target. The probabilities that they hit the target are 0.4, 0.5 and 0.7, respectively. Find the probability that the target is hit exactly once.

7. Under good weather conditions, 80% of flights arrive on time. During bad weather, only 50% of flights arrive on time. Tomorrow, the chance of good weather is 60%. What is the probability that your flight will arrive on time?

Bonus Problems:

8. There are 15 people waiting at a subway station. A train having 5 cars arrives and the passengers get in randomly. If each car has a capacity exceeding 15 persons, find the probability that at least one passenger will get into each car (event A).

9. There are two piles of used computer parts. All items in one pile are in working condition, while in the other pile, 1/4 of the items are defective. An item is chosen randomly from one of the two piles and it is in working condition. If it is put back in the same pile, what is the probability that a second item, chosen at random *from the same pile* as the first one, will turn out to be defective?

Seminar Nr.3, Probabilistic Models

Theory Review

Binomial Model: The probability of k successes in n Bernoulli trials, with probability of success p $(q = 1 - p)$, is

$$P(n,k) = C_n^k p^k q^{n-k},\ k = \overline{0,n}.$$

Hypergeometric Model: The probability that in n trials, we get k successes out of n_1 and $n-k$ failures out of $N - n_1$ $(0 \le k \le n_1,\ 0 \le n-k \le N - n_1)$, is

$$P(n;k) = \frac{C_{n_1}^k C_{N-n_1}^{n-k}}{C_N^n}.$$

Poisson Model: The probability of k successes $(0 \le k \le n)$ in n trials, with probability of success p_i in the i^{th} trial $(q_i = 1 - p_i)$, $i = \overline{1,n}$, is

$$P(n;k) = \sum_{1 \le i_1 < \ldots < i_k \le n} p_{i_1} \ldots p_{i_k} q_{i_{k+1}} \ldots q_{i_n}, \qquad i_{k+1}, \ldots, i_n \in \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\}$$
$$= \text{the coefficient of } x^k \text{ in the polynomial expansion } (p_1x + q_1)(p_2x + q_2)\ldots(p_nx + q_n).$$

Pascal (Negative Binomial) Model: The probability of the n^{th} success occurring after k failures in a sequence of Bernoulli trials with probability of success p $(q = 1 - p)$, is

$$P(n;k) = C_{n+k-1}^{n-1} p^n q^k = C_{n+k-1}^{k} p^n q^k.$$

Geometric Model: The probability of the 1^{st} success occurring after k failures in a sequence of Bernoulli trials with probability of success p $(q = 1 - p)$, is

$$p_k = pq^k.$$

---

1. Five percent of computer parts produced by a certain supplier are defective. What is the probability that a sample of 16 parts contains
a) exactly 3 defective parts (ev. A)?
b) more than 3 defective parts? (ev. B)?
c) at least one defective part (ev. C)?
d) less than 3 defective parts (ev. D)?

2. There are 200 seats in a theater, 10 of which are reserved for the press. 150 people come to the show one night, and are seated randomly. What is the probability of all the seats reserved for the press to be occupied (ev. A)?

3. Among 10 laptop computers, seven are good, the rest have defects. Unaware of this, a customer buys 5 laptops.
a) What is the probability of exactly 2 defective ones among them (ev. A)?
b) Knowing that *at least* 2 purchased laptops are defective, what is the probability that *exactly* 2 are defective (ev. B)?

4. A computer program is tested by 5 independent tests. If there is an error, these tests will detect it with probabilities 0.1, 0.2, 0.3, 0.4 and 0.5, respectively. Suppose that the program contains an error. What is the probability that it will be found by
a) at least one test (ev. A)?
b) more than two tests (ev. B)?
c) all five tests (ev. C)?

5. In a public library, 1 out of 10 people using the computers do not close Windows properly. What is the probability that Windows is closed properly only by the 3^{rd} user (event A)?

6. An engineer tests the quality of produced computers. Suppose that 5% of computers have defects and defects occur independently of each other. Find the probability
a) of exactly 3 defective computers in a shipment of 20 (ev. A);
b) that the engineer has to test at least 5 computers in order to find 2 defective ones (ev. B).

7. (Banach's Problem). A person buys 2 boxes of aspirin, each containing n pills. He takes one aspirin at a time, randomly from one of the two boxes. After a while, he realizes that one box is empty.
a) Find the probability of event A: when he notices that one box is empty, there are k ($k \leq n$) pills left in the other box.
b) Use part a) to find a formula for $S_n = C_{2n}^n + 2 \cdot C_{2n-1}^n + \ldots + 2^n \cdot C_n^n$.

Bonus Problems:

8. An urn contains 4 white balls and 5 black balls. Four balls are randomly removed without replacement. If the second and the third removed balls are black (while the first and forth removed balls are white), then 4 dice are rolled. Otherwise, 3 dice are rolled. Find the probability that exactly two rolled numbers are divisible by 3.

9. Three students take a test consisting of 10 questions. Their probabilities of answering a question correctly are 0.5, 0.8 and 0.6, respectively. Each question is worth one point and no fractions of a point are given. If each student answers every question, find the probability that two students will pass the exam (passing grades are $5, 6, \ldots, 10$) and one will fail (event A).

Seminar Nr. 4, Discrete Random Variables and Discrete Random Vectors

Theory Review

Bernoulli Distribution with parameter $p \in (0,1)$ pdf: $X \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$

Binomial Distribution with parameters $n \in \mathbb{N}, p \in (0,1)$ pdf: $X \begin{pmatrix} k \\ C_n^k p^k q^{n-k} \end{pmatrix}_{k=\overline{0,n}}$

Discrete Uniform Distribution with parameter $m \in \mathbb{N}$ pdf: $X \begin{pmatrix} k \\ \dfrac{1}{m} \end{pmatrix}_{k=\overline{1,m}}$

Hypergeometric Distribution with parameters $N, n_1, n \in \mathbb{N}$ $(n_1 \le N)$ pdf: $X \begin{pmatrix} k \\ \dfrac{C_{n_1}^k C_{N-n_1}^{n-k}}{C_N^n} \end{pmatrix}_{k=\overline{0,n}}$

Poisson Distribution with parameter $\lambda > 0$ pdf: $X \begin{pmatrix} k \\ \dfrac{\lambda^k}{k!} e^{-\lambda} \end{pmatrix}_{k=0,1,\ldots}$

X represents the number of "rare events" that occur in a fixed period of time; λ represents the frequency, the average number of events during that time.

(Negative Binomial) Pascal Distribution with parameters $n \in \mathbb{N}, p \in (0,1)$ pdf:

$$X \begin{pmatrix} k \\ C_{n+k-1}^k p^n q^k \end{pmatrix}_{k=0,1,\ldots}$$

Geometric Distribution with parameter $p \in (0,1)$ pdf: $X \begin{pmatrix} k \\ pq^k \end{pmatrix}_{k=0,1,\ldots}$

Cumulative Distribution Function (cdf) $F_X : \mathbb{R} \to \mathbb{R}$, $F_X(x) = P(X \le x) = \displaystyle\sum_{x_i \le x} p_i$

$(X,Y) : S \to \mathbb{R}^2$ **discrete random vector**:
− **(joint) pdf** $p_{ij} = P(X = x_i, Y = y_j)$, $(i,j) \in I \times J$,
− **(joint) cdf** $F = F_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R},\ F(x,y) = P(X \le x, Y \le y) = \displaystyle\sum_{x_i \le x} \sum_{y_j \le y} p_{ij},\ \forall (x,y) \in \mathbb{R}^2$,
− **marginal densities** $p_i = P(X = x_i) = \displaystyle\sum_{j \in J} p_{ij},\ \forall i \in I,\ q_j = P(Y = y_j) = \sum_{i \in I} p_{ij},\ \forall j \in J.$

For $X \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}, Y \begin{pmatrix} y_j \\ q_j \end{pmatrix}_{j \in J}$,
X and Y are **independent** $<=>$ $p_{ij} = P(X = x_i, Y = y_j) = P(X = x_i)\, P(Y = y_j) = p_i q_j$.
$X+Y \begin{pmatrix} x_i + y_j \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J}, \alpha X \begin{pmatrix} \alpha x_i \\ p_i \end{pmatrix}_{i \in I}, XY \begin{pmatrix} x_i y_j \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J}, X/Y \begin{pmatrix} x_i / y_j \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J} (y_j \ne 0)$

---

1. A computer virus is trying to corrupt two files. The first file will be corrupted with probability 0.4. Independently of it, the second file will be corrupted with probability 0.3. Find the probability distribution function (pdf) of X, the number of corrupted files.

2. A coin is flipped 3 times. Let X denote the number of heads that appear.
a) Find the pdf of X. What type of distribution does X have?
b) Find $P(X \le 2)$ and $P(X < 2)$.

3. (New Accounts) Customers of an internet service provider initiate new accounts at the average rate of 10 accounts per day.
a) Find the probability that more than 8 new accounts will be initiated today;
b) Find the probability that at most 16 new accounts will be initiated within 2 days.

4. It was found that the probability to log on to a computer from a remote terminal is 0.7. Let X denote the number of attempts that must be made to gain access to the computer:
a) Find the pdf of X;
b) Find the probability (express it in terms of the cdf F_X) that at most 4 attempts must be made to gain access to the computer;
c) Find the probability that at least 3 attempts must be made to gain access to the computer.

5. A number is picked randomly out of 1, 2, 3, 4 and 5. Let X denote the number picked. Let Y be 1 if the number picked was 1, 2 if the number was prime and 3, otherwise.
a) Find the pdf's of X, Y;
b) Find the pdf's of $X+Y$, XY.

6. Same problem with 2 numbers being picked randomly. Variable X refers to the 1^{st} number, variable Y to the 2^{nd}. Is there a difference in the answers, from the previous problem?

7. An internet service provider charges its customers for the time of the internet use. Let X be the used time (in hours, rounded to the nearest hour) and Y the charge per hour (in cents). The joint pdf for (X, Y) is given in the following table:

X \ Y	1	2	3
1	0	0.10	0.40
2	0.06	0.10	0.10
3	0.06	0.04	0
4	0.10	0.04	0

Find
a) the marginal pdf's of X and Y;
b) the probability that a customer will be charged only 1 cent per hour when being online for 2 hours (event B);
c) the probability that a customer will be charged at most 2 cents per hour when being online for at least 3 hours (event C);
d) the pdf of Z, the total charge for a customer.

Bonus Problems:

8. Let X and Y be two independent random variables such that X has a discrete Uniform distribution with parameter 2 and Y has a Bernoulli distribution with parameter $\frac{1}{3}$. Let $U = X+Y$ and $V = X-Y$.
a) Find the joint pdf of (U, V).
b) Find the marginal pdfs of U and V.
c) Are U and V independent? Justify the answer.

9. A point in plane has coordinates (X, Y), where the values of X and Y are determined by rolling two dice (one for X and one for Y). Find the probability that the point (X, Y) is on the circle $x^2 + y^2 = 10$.

Seminar Nr. 5, Continuous Random Variables and Continuous Random Vectors

Theory Review

$X : S \to \mathbb{R}$ continuous random variable with pdf $f : \mathbb{R} \to \mathbb{R}$ and cdf $F : \mathbb{R} \to \mathbb{R}$. Properties:

1. $F(x) = P(X \le x) = \int\limits_{-\infty}^{x} f(t)dt$

2. $f(x) \ge 0, \forall x \in \mathbb{R}, \int\limits_{\mathbb{R}} f(x) = 1$

3. $P(X = x) = 0, \forall x \in \mathbb{R}, P(a < X < b) = P(a \le X \le b) = \int\limits_{a}^{b} f(t)dt$

4. $F(-\infty) = 0, F(\infty) = 1$

$(X, Y) : S \to \mathbb{R}^2$ continuous random vector with pdf $f = f_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R}$ and

cdf $F = F_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R},\ F(x, y) = P(X \le x, Y \le y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(u, v)\, dv\, du,\ \forall (x, y) \in \mathbb{R}^2$. Properties:

1. $P(a_1 < X \le b_1, a_2 < Y \le b_2) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$
2. $F(\infty, \infty) = 1, F(-\infty, y) = F(x, -\infty) = 0,\ \forall x, y \in \mathbb{R}$
3. $F_X(x) = F(x, \infty),\ F_Y(y) = F(\infty, y),\ \forall x, y \in \mathbb{R}$ (marginal cdf's)
4. $P((X, Y) \in D) = \iint\limits_{D} f(x, y)\, dy\, dx$
5. $f_X(x) = \int\limits_{\mathbb{R}} f(x, y)dy,\ \forall x \in \mathbb{R},\ f_Y(y) = \int\limits_{\mathbb{R}} f(x, y)dx,\ \forall y \in \mathbb{R}$ (marginal densities)
6. X and Y are independent <=> $f_{(X,Y)}(x, y) = f_X(x)f_Y(y),\ \forall (x, y) \in \mathbb{R}^2$.

Function $Y = g(X)$: X r.v., $g : \mathbb{R} \to \mathbb{R}$ differentiable with $g' \ne 0$, strictly monotone

$f_Y(y) = \dfrac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|},\ y \in g(\mathbb{R})$

Uniform distribution $U(a, b),\ -\infty < a < b < \infty$: pdf $f(x) = \dfrac{1}{b-a}, x \in [a, b]$.

Normal distribution $N(\mu, \sigma), \mu \in \mathbb{R}, \sigma > 0$: pdf $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$.

Gamma distribution $Gamma(a, b),\ a, b > 0$: pdf $f(x) = \dfrac{1}{\Gamma(a)b^a} x^{a-1} e^{-\frac{x}{b}},\ x > 0$.

Exponential distribution $Exp(\lambda) = Gamma(1, 1/\lambda),\ \lambda > 0$: pdf $f(x) = \lambda e^{-\lambda x}, x > 0$.
- Exponential distribution models *time*: waiting time, interarrival time, failure time, time between rare events, etc; the parameter λ represents the frequency of rare events, measured in time^{-1}.
- Gamma distribution models the *total* time of a multistage scheme.
- For $\alpha \in \mathbb{N}$, a $Gamma(\alpha, 1/\lambda)$ variable is the sum of α independent $Exp(\lambda)$ variables.

---

1. The lifetime, in years, of some electronic component is a random variable with density

$$f(x) = \begin{cases} \dfrac{k}{x^4}, & \text{for } x \ge 1 \\ 0, & \text{for } x < 1. \end{cases}$$

Find
a) the constant k;

b) the corresponding cdf F;
c) the probability for the lifetime of the component to exceed 2 years.

2. (The Uniform property) Let $X \in U(a, b)$. For any $h > 0$ and $t, s \in [a, b - h]$,

$$P(s < X < s + h) = P(t < X < t + h).$$

The probability is only determined by the length of the interval, but not by its location.
Example: A certain flight can arrive at any time between 4:50 and 5:10 pm. Let X denote the arrival time of the flight.
a) What distribution does X have?
b) When is the flight more likely to arrive: between 4:50 and 4:55 or between 5 and 5:05; before 5 or after 5?

3. On the average, a computer experiences breakdowns every 5 months. The time until the first breakdown and the times between any two consecutive breakdowns are independent Exponential random variables. After the third breakdown, a computer requires a special maintenance.
a) Find the probability that a special maintenance is required within the next 9 months;
b) Given that a special maintenance was not required during the first 12 months, what is the probability that it will not be required within the next 4 months?

4. The joint density for (X, Y) is $f_{(X,Y)}(x, y) = \frac{1}{16}x^3y^3$, $x, y \in [0, 2]$.
a) Find the marginal densities f_X, f_Y.
b) Are X and Y independent?
c) Find $P(X \leq 1)$.

5. Let X be a random variable with density $f_X(x) = \frac{1}{4}xe^{-\frac{x}{2}}$, $x \geq 0$ and let $Y = \frac{1}{2}X + 2$. Find f_Y.

6. Let $X \in N(0, 1)$. Find the probability density function of $Y = |X|$.

Bonus Problems:

7. Let X denote the velocity of a random gas molecule. According to the Maxwell-Boltzmann law, the pdf of X is given by

$$f_X(x) = cx^2e^{-\beta x^2},\ x > 0,$$

where the constants c and β depend on the gas involved, its mass and its temperature. The kinetic energy of the molecule is given by $Y = \frac{1}{2}mX^2$, where $m > 0$. For a gas molecule with $c = 2, \beta = 1$ and $m = 1$, find the pdf of the kinetic energy of the molecule.

8. A gamer shoots at a virtual shooting board centered at the origin of the Cartesian coordinate system such that the coordinates of the hit are two independent random variables that follow the $N(0, 1)$ distribution. Find the probability that the shooter hits the upper half-plane of the shooting board at a distance between 1 and 2 from the origin.

Seminar Nr. 6, Numerical Characteristics of Random Variables

Theory Review

Expectation:

- if $X \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$ is discrete, then $E(X) = \sum\limits_{i \in I} x_i p_i$.

- if X is continuous with pdf f, then $E(X) = \int\limits_{\mathbb{R}} x f(x) dx$.

Variance: $V(X) = E\left((X - E(X))^2\right) = E\left(X^2\right) - (E(X))^2$.

Standard Deviation: $\sigma(X) = \mathrm{Std}(X) = \sqrt{V(X)}$.

Moments:
- **moment of order k**: $\nu_k = E\left(X^k\right)$.
- **absolute moment of order k**: $\underline{\nu_k} = E\left(|X|^k\right)$.
- **central moment of order k**: $\mu_k = E\left((X - E(X))^k\right)$.

Properties:
1. $E(aX + b) = aE(X) + b$, $V(aX + b) = a^2 V(X)$
2. $E(X + Y) = E(X) + E(Y)$
3. if X and Y are independent, then $E(XY) = E(X)E(Y)$ and $V(X + Y) = V(X) + V(Y)$
4. if $h : \mathbb{R} \to \mathbb{R}$ is a measurable function, X a random variable;
- if X is discrete, then $E\left(h(X)\right) = \sum\limits_{i \in I} h(x_i) p_i$
- if X is continuous, then $E\left(h(X)\right) = \int\limits_{\mathbb{R}} h(x) f(x) dx$

Covariance: $\mathrm{cov}(X, Y) = E\left((X - E(X))(Y - E(Y))\right)$

Correlation Coefficient: $\rho(X, Y) = \dfrac{\mathrm{cov}(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$

Properties:
1. $\mathrm{cov}(X, Y) = E(XY) - E(X)E(Y)$
2. $V\left(\sum\limits_{i=1}^{n} a_i X_i\right) = \sum\limits_{i=1}^{n} a_i^2 V(X_i) + 2 \sum\limits_{1 \le i < j \le n} a_i a_j \mathrm{cov}(X_i, X_j)$
3. X, Y independent $=>$ $\mathrm{cov}(X, Y) = \rho(X, Y) = 0$ (X and Y are *uncorrelated*)
4. $-1 \le \rho(X, Y) \le 1$; $\rho(X, Y) = \pm 1 \iff \exists\, a, b \in \mathbb{R},\ a \ne 0$ s.t. $Y = aX + b$

Let (X, Y) be a continuous random vector with pdf $f(x, y)$, let $h : \mathbb{R}^2 \to \mathbb{R}^2$ a measurable function, then
$$E\left(h(X, Y)\right) = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} h(x, y) f(x, y) dx dy$$

---

1. Every day, the number of network blackouts has the following pdf

$$X \begin{pmatrix} 0 & 1 & 2 \\ 0.7 & 0.2 & 0.1 \end{pmatrix}.$$

A small internet trading company estimates that each network blackout costs them $\$500$.
a) How much money can the company expect to lose each day because of network blackouts?
b) What is the standard deviation of the company's daily loss due to blackouts?

2. (Refer to Problem 5 in Sem. 3) About ten percent of computer users in a public library do not close Windows properly. On the average, how many users *do* close Windows properly before someone *does not*?

3. (Refer to Problem 1 in Sem. 5) The lifetime, in years, of some electronic component is a random variable with density

$$f(x) = \begin{cases} \dfrac{3}{x^4}, & \text{for} \quad x \geq 1 \\ 0, & \text{for} \quad x < 1. \end{cases}$$

How many years, on the average, can we expect that electronic equipment to last?

4. (Optimal portfolio) A businessman wants to invest $\$600$ and has two companies to choose from, company A, where shares cost $\$20$ each and company B, where shares cost $\$30$ per share. The market analysis shows that for company A the return per share is distributed as follows: lose $\$1$ with probability 0.2, win $\$2$ with probability 0.6, or win/lose nothing. For company B: lose $\$1$ with probability 0.3, win $\$3$ with probability 0.6, or win/lose nothing. The returns from the two companies are independent. In order to maximize the expected return and minimize the risk, which way is better to invest:
a) all money in company A;
b) all money in company B;
c) half the amount in each?

5. (Reduced Variables). Let X be a random variable with mean $E(X)$ and standard deviation $\sigma(X) = \sqrt{V(X)}$. Find the mean and variance of $Y = \dfrac{X - E(X)}{\sigma(X)}$.

6. The joint density function of the vector (X, Y) is $f(x, y) = x + y,\ (x, y) \in [0, 1] \times [0, 1]$. Find
a) the means and variances of X and Y;
b) the correlation coefficient $\rho(X, Y)$.

7. Let X be a discrete random variable with pdf $X \begin{pmatrix} -1 & 0 & 1 \\ \sin^2 a & \cos 2a & \sin^2 a \end{pmatrix}$, $a \in \left(0, \frac{\pi}{4}\right)$. For any $k \in \mathbb{N}^*$, let $Y_k = X^{2k-1}$ and $Z_k = X^{2k}$. Find $\rho(Y_k, Z_k)$. (In particular, X and X^2 are uncorrelated, but *not* independent).

Bonus Problems

8. In a random array of distinct characters, an "interchange" is exchanging the positions of exactly two elements with each other. If X_n denotes the minimum (ideal) number of interchanges necessary to sort an array of length n, then

$$X_n = X_{n-1} + I,$$

where $I = 0$, if the last element of the array is in its correct position and $I = 1$, otherwise. A random number generator is used to generate sets of 100 distinct three-decimal numbers between 0 and 1. What is the ideal average number of interchanges needed to sort such an array?

9. Two independent customers are scheduled to arrive in the afternoon. Their arrival times are uniformly distributed between 2 pm and 8 pm. What is the expected time of
a) the first (earlier) arrival;
b) the last (later) arrival?

Seminar Nr. 7, Inequalities; Central Limit Theorem; Point Estimators

Theory Review

Markov's Inequality: $P(|X| \geq a) \leq \dfrac{1}{a} E(|X|), \forall a > 0.$

Chebyshev's Inequality: $P(|X - E(X)| \geq \varepsilon) \leq \dfrac{V(X)}{\varepsilon^2}, \forall \varepsilon > 0.$

Central Limit Theorem(CLT) Let $X_1, \ldots, X_n$ be independent random variables with the same expectation $\mu = E(X_i)$ and same standard deviation $\sigma = \sigma(X_i) = \mathrm{Std}(X_i)$ and let $S_n = \sum_{i=1}^{n} X_i$. Then, as $n \to \infty$,

$Z_n = \dfrac{S_n - E(S_n)}{\sigma(S_n)} = \dfrac{S_n - n\mu}{\sigma\sqrt{n}} \longrightarrow Z \in N(0,1)$, in distribution (in cdf), i.e. $F_{Z_n} \to F_Z = \Phi$.

Point Estimators
- method of moments: solve the system $\nu_k = \overline{\nu}_k$, for as many parameters as needed ($k = 1, \ldots,$ nr. of unknown parameters);
- method of maximum likelihood: solve $\dfrac{\partial \ln L(X_1, \ldots, X_n; \theta)}{\partial \theta_j} = 0$, where $L(X_1, \ldots, X_n; \theta) = \prod_{i=1}^{n} f(X_i; \theta)$ is the likelihood function;
- **standard error** of an estimator $\overline{\theta}$: $\sigma_{\hat{\theta}} = \sigma(\overline{\theta}) = \sqrt{V(\overline{\theta})}$;
- **Fisher information** $I_n(\theta) = -E\left[\dfrac{\partial^2 \ln L(X_1, \ldots, X_n; \theta)}{\partial \theta^2}\right]$; if the range of X does not depend on θ, then $I_n(\theta) = nI_1(\theta)$;
- **efficiency** of an absolutely correct estimator $\overline{\theta}$: $e(\overline{\theta}) = \dfrac{1}{I_n(\theta)V(\overline{\theta})}$.
- an estimator $\overline{\theta}$ for the target parameter θ is
  - **unbiased**, if $E(\overline{\theta}) = \theta$;
  - **absolutely correct**, if $E(\overline{\theta}) = \theta$ and $V(\overline{\theta}) \to 0$, as $n \to \infty$;
  - **MVUE** (minimum variance unbiased estimator), if $E(\overline{\theta}) = \theta$ and $V(\overline{\theta}) \leq V(\hat{\theta}),\ \forall \hat{\theta}$ unbiased estimator;
  - **efficient**, if $e(\overline{\theta}) = 1$.

- $\overline{\theta}$ efficient => $\overline{\theta}$ MVUE.

---

1. (**The 3σ Rule**). For any random variable X, most of the values of X lie within 3 standard deviations away from the mean.

2. True or False: There is at least a 90% chance of the following happening: when flipping a coin 1000 times, the number of "heads" that appear is between 450 and 550.

3. Installation of some software package requires downloading 82 files. On the average, it takes 15 sec to download a file, with a variance of 16 sec^2. What is the probability that the software is installed in less than 20 minutes?

4. A sample of 3 observations, $X_1 = 0.4, X_2 = 0.7, X_3 = 0.9$, is collected from a continuous distribution with pdf

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1}, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases},$$

with $\theta > 0$, unknown. Estimate θ by the method of moments and by the method of maximum likelihood.

5. A sample $X_1, \ldots, X_n$ is drawn from a distribution with pdf

$$f(x;\theta) = \frac{1}{2\theta} e^{-\dfrac{x}{2\theta}},\ x > 0$$

($\theta > 0$), which has mean $\mu = E(X) = 2\theta$ and variance $\sigma^2 = V(X) = 4\theta^2$. Find
a) the method of moments estimator, $\overline{\theta}$, for θ;
b) the efficiency of $\overline{\theta}$, $e(\overline{\theta})$;
c) an approximation for the standard error of the estimate in a), $\sigma_{\overline{\theta}}$, if the sum of 100 observations is 200.