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

Solutions

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?

Solution:
Order *does* matter, so these are arrangements:
a) $A_{15}^{10} = 6\cdot 7\cdot\ldots\cdot 15;$
b) $A_{10}^{10} = P_{10} = 10!.$

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.

Solution:
All actions are *independent* of each other, so we simply multiply the corresponding numbers of outcomes for each action:
a) There are 6 possibilities for each die, so for 3 dice there are 6^3 possible outcomes;

b) There are 26 letters, 10 digits, so all together $26^2 \cdot 10^3$ possible outcomes.

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?

Solution:
Here, order *does not* matter, so for each individual action we use combinations.
a) Choosing 10 objects out of 25 (order does not matter): C_{25}^{10} possibilities;

b) We break the problem into subproblems:
– First, selecting exactly three computer games, i.e. choosing 3 objects out of 5 (again order does not matter). This can be done in C_5^3 ways;
– Then, choosing the rest, i.e. 7 objects out of the remaining 20 (only 20, because no more computer games can be chosen), so C_{20}^7 possibilities.
The two actions are again, independent, so we multiply the numbers.
Final answer: $C_5^3 \cdot C_{20}^7$;

c) The same type of argument as before, only now there are three independent actions, so $C_5^3 \cdot C_3^2 \cdot C_{17}^5$ possible selections.

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.

Solution:
Let's recall:
– union, $\cup$, is described with the word "or";
– intersection, $\cap$, is described with the word "and";
– opposite (contrary, complementary) event, $\overline{E}$, is described with the word "not";
– difference, $\setminus$, is described with the words "but not".

a) "All" tickets are winning, that means the first one is (event A_1) AND the second one is (event A_2) AND ... all the way to A_n. Thus, this is an intersection,

$$A = \bigcap_{i=1}^{n} A_i.$$

b) Same here, only all are now *losing* tickets, i.e. *not* winning, so

$$B = \bigcap_{i=1}^{n} \overline{A_i} = \overline{\bigcup_{i=1}^{n} A_i},$$

the last part coming from de Morgan's laws.

c) Whenever working with events (sets), always keep in mind the contrary event, as well, and choose to work with the one that is easiest. The opposite event of C would be " 0 winning tickets", which is event B, so

$$C = \overline{B} = \bigcup_{i=1}^{n} A_i.$$

d) The sole winning ticket could be either one. So, another way of describing event D would be "only the first ticket is a winning one OR only the second ticket is OR ... OR only the last one is. That is a union. Only it's not the union of events A_i, because A_i states that ticket i is winning, but it's not necessarily *the only* winning ticket.

Thus, let us denote by D_i the event: only the i^{th} ticket is a winning one, for $i = 1, \ldots, n$. Then $D = \bigcup_{i=1}^{n} D_i$.

Now, all that is left to do is express D_i in terms of A_i. Let us start with D_1 and A_1. Obviously, A_1 says more than D_1. In fact, to get D_1 (only the first ticket is winning) from A_1 (the first ticket is winning), we have to "take something out", that is make sure that none of the other tickets are winning ones. This sounds like "something ... BUT NOT something else ..." Indeed, we want A_1, but not "any of the other tickets are winning". So $D_1 = A_1 \setminus \bigcup_{j=2}^{n} A_j$ and, in general,

$$D_i = A_i \setminus \left(\bigcup_{j \neq i} A_j \right) = A_i \cap \overline{\bigcup_{j \neq i} A_j} = \overline{A}_1 \cap \ldots \cap \overline{A}_{i-1} \cap A_i \cap \overline{A}_{i+1} \cap \ldots \cap \overline{A}_n.$$

Then

$$D = \bigcup_{i=1}^{n} \left(A_i \setminus \left(\bigcup_{j \neq i} A_j \right) \right)$$

e) The same type of argument leads to

$$E = \bigcup_{1 \leq i < j \leq n} \left((A_i \cap A_j) \setminus \left(\bigcup_{k \neq i,j} A_k \right) \right)$$

f) The contrary event of F: at most one is winning, i.e. 0 or 1 winning tickets, so $\overline{F} = B \cup D$. Thus

$$F = \overline{B \cup D} = \overline{B} \cap \overline{D} = C \cap \overline{D} = C \setminus D.$$

g) At most 2 winning tickets means 0 or 1 or 2. So

$$G = B \cup D \cup E.$$

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.

<u>Solution</u>:
This is the same type of problem as the previous one. We have:

a) $A = A_1 \cup A_2 \cup A_3$;

b) $B = \overline{A} = \overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3$;

c) $C = A_1 \cap A_2 \cap A_3$;

d) $D = \left(A_1 \cap \overline{A}_2 \cap \overline{A}_3\right) \cup \left(\overline{A}_1 \cap A_2 \cap \overline{A}_3\right) \cup \left(\overline{A}_1 \cap \overline{A}_2 \cap A_3\right)$;

e) $E = \left(A_1 \cap A_2 \cap \overline{A}_3\right) \cup \left(A_1 \cap \overline{A}_2 \cap A_3\right) \cup \left(\overline{A}_1 \cap A_2 \cap A_3\right)$.

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.

Solution:

This is classical probability. First, we compute the denominator, N_t, since it is the same for all events A, B, C and D. Obviously,

$$N_t = 1000.$$

Then, for each event, first we locate the cubes and then we count them.

a) These are the cubes on corners. There are 8 corners, so $N_f = 8$ and

$$P(A) = \frac{8}{1000}.$$

b) These are the cubes on edges, but not on corners. There are 12 edges and on each edge there are 8 such cubes (without the corner ones!). Thus $N_f = 12 \cdot 8 = 96$ and

$$P(B) = \frac{96}{1000}.$$

c) These are the cubes on faces, but not on edges. There are 6 faces and $8 \cdot 8 = 64$ cubes on each face, making $N_f = 6 \cdot 64 = 384$ and

$$P(C) = \frac{384}{1000}.$$

d) We could compute this probability the same way as the previous ones (these would be the interior cubes), but let us use something else instead. Notice that in this experiment, the only possible numbers of colored faces are $0, 1, 2$ or 3. In other words, events A, B, C and D cover all possibilities, $S = A \cup B \cup C \cup D$. They are also mutually exclusive, obviously. Then they form a partition of the sample space. Recall that the sum of probabilities of all events in a partition is equal to 1 (the probability of the sure event). Since we have already computed the other ones, we have

$$P(D) = 1 - P(A) - P(B) - P(C) = \frac{512}{1000}.$$

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 ($0 \le m \le n$)?

Solution:

For the number of possible outcomes N_t:

The 1^{st} letter can be distributed in any of the N mailboxes, so there are N choices. The same thing is true for the 2^{nd} letter, so another N choices. So, if we had 2 letters, we would have N^2 cases. Since the same thing is true for each of the n letters, we have

$$N_t = N^n.$$

For the number of favorable outcomes N_f:

First, the m letters can be chosen in C_n^m ways. Then, once those m letters are determined, the other $n - m$ letters should be distributed in the remaining $N - 1$ mailboxes. From above, this can be done in $(N-1)^{n-m}$ ways. Since these two problems (choosing the m letters and distributing the other $n - m$ letters) are independent, the number of favorable outcomes is

$$N_f = C_n^m (N-1)^{n-m}$$

and

$$P(A) = \frac{C_n^m (N-1)^{n-m}}{N^n}.$$

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 (event A)?

c) Same questions, if capital letters are not used.

__**Solution**__:

a) First, let us see how long it will take the spy program to check **all** passwords.

There are 62 characters to choose from (26 lower-case letters, 26 capital letters and 10 digits) and we need to choose 8, with repetitions, so there are

$$(62)^8 = 2.183 \cdot 10^{14}$$

8-character possible passwords.

Now, how fast does the spy program check them? A speed of 1 million per second, means

$10^6 \cdot 60 \cdot 60 \cdot 24 = 8.64 \cdot 10^{10}$ per day,

$8.64 \cdot 10^{10} \cdot 7 = 6.048 \cdot 10^{11}$ per week,

$6.048 \cdot 10^{11} \cdot 52 = 3.145 \cdot 10^{13}$ per year.

So it will take it

$$\frac{21.83}{3.145} = 6.9412,$$

about 7 years, to check *all* the passwords and, hence, to guess yours. Since it's just as probable to try (and guess) a combination at *any time* in the interval $[0, 6.9412]$, on average, it will take about half of that, i.e. 3 and a half years for the spy program to crack your password.

b) This is just classical probability with $N_t = 2.183 \cdot 10^{14}$ and $N_f = 6.048 \cdot 10^{11}$, so

$$p_1 = P(A) = \frac{6.048 \cdot 10^{11}}{2.183 \cdot 10^{14}} \simeq 0.00277,$$

very small!

c) Without capital letters, there are fewer characters to choose from (only $26 + 10 = 36$), so only $(36)^8 = 2.8211 \cdot 10^{12}$ passwords to test. The spy program works at the same speed, so nothing changes there. Now it will take only

$$\frac{2.8211 \cdot 10^{12}}{8.64 \cdot 10^{10}} = 32.6516$$

DAYS to test them all. On the average, in this case, the spyware will break your password in about 16 days.

For the second part, now $N_t = 2.821 \cdot 10^{12}$ and N_f is the same. So the probability that the password will be cracked within a week is

$$p_2 = P(A) = \frac{6.048 \cdot 10^{11}}{2.821 \cdot 10^{12}} \simeq 0.2144,$$

much bigger!

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.

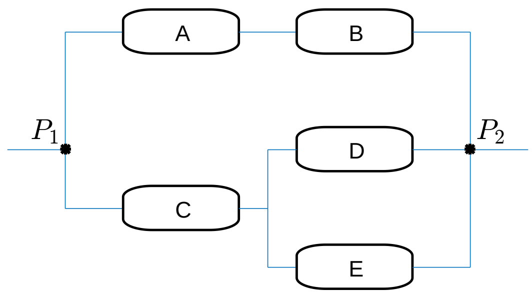


Figure 1: System Reliability

<u>**Solution**</u>**:**

The *reliability* is the probability of the system being operable. Now, how do we interpret two components being linked **sequentially**, versus being connected **in parallel**? For sequential connection, *both* components have to be operable (that means the word "and" would be used), whereas for parallel connection one OR the other would have to work. So now we know which set operations to use.

Let us denote by B1 the upper branch connecting points P_1 and P_2 and by B2 the lower branch. Then the reliability is

$$R = P(\text{the system is operable}) = P(B1 \cup B2),$$

since $B1$ and $B2$ are linked in parallel. We can compute this two ways: the direct way, or using the contrary event. Let us do it directly now and we will use the other way later. We have

$$P(B1 \cup B2) = P(B1) + P(B2) - P(B1 \cap B2) = P(B1) + P(B2) - P(B1)P(B2),$$

because the independence of all events implies the independence of any combinations/operations of/with them (so $B1$ and $B2$ are also independent).

Now, let us compute the probabilities $P(B1)$ and $P(B2)$. We have

$$P(B1) = P(A \cap B) = P(A) \cdot P(B) = (0.92)^2 = 0.8464,$$

since A and B are linked sequentially and operate independently of each other.

Similarly, we go further:

$$P(B2) = P(C \cap (D \cup E)) = P(C) \cdot P(D \cup E),$$

since C and $D \cup E$ are also independent.
For $P(D \cup E)$, we will use the complementary event. Thus,

$$\begin{aligned} P(D \cup E) &= 1 - P(\overline{D \cup E}) = 1 - P(\overline{D} \cap \overline{E}) \\ &= 1 - P(\overline{D}) \cdot P(\overline{E}) = 1 - (1 - 0.92)^2 = 0.9936. \end{aligned}$$

Then

$$P(B2) = P(C) \cdot P(D \cup E) = 0.92 \cdot 0.9936 = 0.9141.$$

Finally,

$$R = P(B1 \cup B2) = 0.8464 + 0.9141 - 0.8464 \cdot 0.9141 = 0.9868.$$

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 (event A_1)?
b) does not know C/C++ and does not know Fortran (event A_2)?
c) knows C/C++, but not Fortran (event A_3)?
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++ (event A_4)?
f) What is the probability that someone who knows C/C++, does not also know Fortran (event A_5)?

Solution:
Denote the events:
C: employee knows C/C++,
F: employess knows Fortran.
We know

$$P(C) = 0.7, \ P(F) = 0.6 \text{ and } P(C \cap F) = 0.5.$$

Then we express (carefully!) all the events with the right operations.

a) $P(A_1) = P(\overline{F}) = 1 - P(F) = 0.4$;

b) $P(A_2) = P(\overline{C} \cap \overline{F}) = P(\overline{C \cup F}) = 1 - P(C \cup F) = 1 - \Big(P(C) + P(F) - P(C \cap F)\Big)$
$= 1 - (0.7 + 0.6 - 0.5) = 0.2$;

c) $P(A_3) = P(C \cap \overline{F}) = P(C \setminus F) = P(C) - P(C \cap F) = 0.7 - 0.5 = 0.2$;

d) Since

$$P(C \cap F) = 0.5 \neq 0.42 = P(C) \cdot P(F),$$

the answer is "NO", they are not;

e) $P(A_4) = P(C|F) = \dfrac{P(C \cap F)}{P(F)} = \dfrac{0.5}{0.6} = \dfrac{5}{6} = 0.833;$

f) $P(A_5) = P(\overline{F}|C) = \dfrac{P(\overline{F} \cap C)}{P(C)} = \dfrac{P(A_3)}{P(C)} = \dfrac{0.2}{0.7} = \dfrac{2}{7} = 0.286.$

Alternatively, we can use the probability of the contrary event. The same formula holds for conditional probabilities:

$P(A_5) = P(\overline{F}|C) = 1 - P(F|C) = 1 - \dfrac{P(F \cap C)}{P(C)} = 1 - \dfrac{0.5}{0.7} = \dfrac{2}{7} = 0.286.$

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.

Solution:

Refer to Problem 5 in Seminar 1. In fact, the hard work was done last time! Denote the events

A: the target is hit exactly once,

A_i: the i^{th} shooter hits the target, $i = 1, 2, 3$.

Then $P(A_1) = 0.4, P(A_2) = 0.5, P(A_3) = 0.7$ and, from last time,

$$A = \left(A_1 \cap \overline{A_2} \cap \overline{A_3}\right) \cup \left(\overline{A_1} \cap A_2 \cap \overline{A_3}\right) \cup \left(\overline{A_1} \cap \overline{A_2} \cap A_3\right)$$

Since the union is disjoint and the events A_1, A_2 and A_3 are independent, we have

$$P(A) = 0.4 \cdot 0.5 \cdot 0.3 + 0.6 \cdot 0.5 \cdot 0.7 + 0.6 \cdot 0.5 \cdot 0.7 = 0.36.$$

Note. This is a classical example of a Poisson probabilistic model.

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?

Solution: Denote the events

A: the flight arrives on time,

G: there's good weather,

B: there's bad weather.

Then, in fact, $B = \overline{G}$, we have $P(G) = 0.6, P(\overline{G}) = 0.4$ and $\{G, \overline{G}\}$ form a partition of the sample space.

We are given $P(A|G) = 0.8$ and $P(A|\overline{G}) = 0.5$ and we want to compute $P(A)$, *without* any condition. By the Total Probability Rule,

$$P(A) = P(A|G)P(G) + P(A|\overline{G})P(\overline{G}) = 0.8 \cdot 0.6 + 0.5 \cdot 0.4 = 0.68.$$

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

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

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

$$P(n; k) = \sum_{1 \le i_1 < \ldots < i_k \le n} p_{i_1} \ldots p_{i_k} q_{i_{k+1}} \ldots q_{i_n}, \qquad i_{k+1}, \ldots, i_n \in \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\}$$
$$= \text{the coefficient of } x^k \text{ in the polynomial expansion } (p_1 x + q_1)(p_2 x + q_2) \ldots (p_n x + q_n).$$

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)?

Solution:
A *trial* here is "selecting a computer part". There are 16 independent trials.
At each trial, the outcomes are "the computer part is defective" or "the computer part is good".
We are asked about a certain number of parts being *defective*, so we define "success": "a part is defective". Probability of success in every trial is then 0.05.
So, this is a **Binomial model** with parameters $n = 16$, $p = 0.05$.

a)

$$P(A) = P(k = 3) = C_{16}^3 (0.05)^3 (0.95)^{13} \overset{\text{Matlab}}{=} \mathit{binopdf}(3, 16, 0.05) = 0.0359.$$

b)

$$\begin{aligned} P(B) &= P(k > 3) = 1 - P(k \le 3) = 1 - P\Big(\underbrace{(k = 0) \cup (k = 1) \cup (k = 2) \cup (k = 3)}_{\text{m.e.}}\Big) \\ &= 1 - \sum_{k=0}^{3} C_{16}^k (0.05)^k (0.95)^{16-k} \\ &= 1 - \mathit{binocdf}(3, 16, 0.05) = 0.007. \end{aligned}$$

c)

$$P(C) = P(k \geq 1) = 1 - P(k = 0) = 1 - C_{16}^{0}(0.05)^{0}(0.95)^{16} = 1 - binopdf(0, 16, 0.05) = 0.5599.$$

d)

$$P(D) = P(k < 3) = 1 - P(k \geq 3) = 1 - P(\underbrace{A \cup B}_{\text{m.e.}}) = 1 - P(A) - P(B) = 0.9571.$$

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)?

Solution:
A trial here is "taking a seat". There are 150 trials.
At each trial, the outcomes are "the seat is reserved for the press" or "the seat is *not* reserved for the press". There are 200 seats to choose from, 10 of which have the property of being reserved for the press and sampling is done **without** replacement.
Thus, this is a **Hypergeometric model** with parameters $N = 200, n_1 = 10, n = 150$.
Then the probability that among the 150 objects, 10 have the characteristic and 140 do not, is

$$P(A) = P(k = 10) = \frac{C_{10}^{10} \cdot C_{190}^{140}}{C_{200}^{150}} = hygepdf(10, 200, 10, 150) = 0.0521.$$

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)?

Solution:
a) There are 10 objects to choose from, 3 of which are defective and 5 are being selected, obviously *without* replacement. This is a **Hypergeometric model** with $N = 10$, $n_1 = 3$ (we are interested in the defective ones) and $n = 5$.

$$P(A) = P(k = 2) = \frac{C_3^2 \cdot C_7^3}{C_{10}^5} = hygepdf(2, 10, 3, 5) = 0.4167.$$

b) First off, this is *conditional* probability

$$P(B) = P(k = 2 | k \geq 2) = \frac{P(k = 2)}{P(k \geq 2)} = \frac{P(k = 2)}{P(k = 2) + P(k = 3)} = \frac{P(A)}{P(A) + P(k = 3)}.$$

For each probability, we use the same model with the same parameters.

$$P(B) = \frac{P(A)}{P(A) + \dfrac{C_3^3 \cdot C_7^2}{C_{10}^5}} = \frac{P(A)}{P(A) + hygepdf(3, 10, 3, 5)} = 0.8333.$$

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)?

Solution:
A trial here is "a test is checking the computer program". There are 5 independent trials. At each trial, the outcomes are "the test finds the error" or "it does not". Denote "success": "the test finds the error". What is different here from Problem 1. is that the probability of success *differs* in every trial. Thus, this is a **Poisson model** with parameters $n = 5, p_1 = 0.1$, $p_2 = 0.2$, $p_3 = 0.3$, $p_4 = 0.4$, $p_5 = 0.5$. So we write the polynomial

$$\begin{aligned} & (0.1x + 0.9)(0.2x + 0.8)(0.3x + 0.7)(0.4x + 0.6)(0.5x + 0.5) \\ = \; & 0.0012x^5 + 0.0214x^4 + 0.1274x^3 + 0.3274x^2 + 0.3714x + 0.1512. \end{aligned}$$

Then we have
a)
$$P(A) = P(k \geq 1) = 1 - P(k = 0) = 1 - 0.1512 = 0.8488.$$
b)
$$P(B) = P(k > 2) = P(k = 3) + P(k = 4) + P(k = 5) = 0.1274 + 0.0214 + 0.0012 = 0.15.$$
c)
$$P(C) = P(k = 5) = 0.0012.$$
Note We **cannot** use a Poisson distribution here!!

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)?

Solution:
A trial: a user closes his Windows session. How many trials are there? We *don't* know, theoretically, infinitely many! At each trial, the user either closes Windows correctly or not. Define "success" as "the user closes Windows properly", so that this matches a Geometric model. Then the probability of success is $p = 9/10 = 0.9$ and event A may be rephrased as: the first success in the 3^{rd} trial, i.e. after 2 failures. Now this is a **Geometric model** with parameter $p = 0.9$ and we compute

$$P(A) = P(k = 2) = 0.9(0.1)^2 = \mathit{geopdf}(2, 0.9) = 0.009.$$

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).

Solution:
a) This is a **Binomial model**, with "trial": engineer tests a computer, "success": the computer has defects, parameters $n = 20$, $p = 0.05$.

$$P(A) = P(k = 3) = C_{20}^{3}(0.05)^3(0.95)^{17} = \mathit{binopdf}(3, 20, 0.05) = 0.0596.$$

b) First off, note that the "in a shipment of 20" part refers to part a) *only*! In part b), a trial means the same thing, but there's no longer a given, finite number of trials. The experiment is repeated until something happens, namely, until the engineer finds 2 defective computers. This is about the rank of a success, so, again we define "success" as: the computer is defective. We rephrase event B as
B: the 2^{nd} success in 5 or more trials, i.e.
B: the 2^{nd} success after 3 or more failures.

This looks like a Pascal model, but there is a problem: the "3 *or more* failures" part, how many exactly are we talking about? A number greater than or equal to 3, i.e. $k = 3, 4, \ldots$. So

that probability would be an infinite sum, a *series*. How do we deal with that? By looking at the *complementary* event,
$\overline{B}$: the 2nd success after 2 or less failures;
So, for $\overline{B}$, we use the **Negative Binomial model** with parameters $n = 2, p = 0.05$ and we consider the values $k \in \{0, 1, 2\}$. Thus,

$$\begin{aligned} P(B) &= P(k \geq 3) = 1 - P(\overline{B}) = 1 - P(k \leq 2) = 1 - \sum_{k=0}^{2} C_{k+1}^{1}(0.05)^2(0.95)^k \\ &= 1 - \sum_{k=0}^{2} nbinpdf(k, 2, 0.05) = 1 - nbincdf(2, 2, 0.05) = 0.986. \end{aligned}$$

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 $S_n = C_{2n}^{n} + 2\cdot C_{2n-1}^{n} + \ldots + 2^n \cdot C_n^n$.

Solution:
a) Denote the events:
A_1: at the time he realizes that the 1st box is empty, there are k pills left in the 2nd box, and
A_2: the other way around.
Then A_1 and A_2 are disjoint and by symmetry, $P(A_1) = P(A_2)$. So,

$$P(A) = P(A_1 \cup A_2) = 2P(A_1).$$

Now, let us see, *when exactly* does he notice that the 1^{st} box is empty? It's not when he *takes* the last pill, but when he *attempts* to take another pill and there's none, i.e. when he attempts to take the $(n+1)^{st}$ pill out of it. This sounds like the rank of a success... Indeed, a trial is "he attempts to take a pill (and if he finds one, he will take it)" so, let us define "success" to be: the person opens the 1^{st} box (attempting to take a pill out of it), "failure": the 2^{nd} box. Event A_1 is then: obtain the $(n+1)^{\text{st}}$ success after $n-k$ failures.

This is then a **Negative Binomial model** with parameters $N = n+1$ and $p = P(\text{success}) = P(\text{failure}) = q = 1/2$ and we want to compute the probability for $K = n-k$. Thus,

$$P(A_1) = C_{(n+1)+(n-k)-1}^{n+1-1}\left(\frac{1}{2}\right)^{n+1}\left(\frac{1}{2}\right)^{n-k} = \frac{1}{2}C_{2n-k}^{n}\cdot\left(\frac{1}{2}\right)^{2n-k}$$

and

$$P(A) = C_{2n-k}^{n}\left(\frac{1}{2}\right)^{2n-k}.$$

b) If for $k = \overline{0, n}$, we denote by E_k the event A from a), then $\{E_k\}_{k=\overline{0,n}}$ form a partition, so $\sum\limits_{k=0}^{n} P(E_k) = 1$, i.e. by a),

$$\begin{aligned} 1 &= \sum_{k=0}^{n} C_{2n-k}^{n}\cdot\frac{2^k}{2^{2n}} \\ &= \frac{1}{2^{2n}}\sum_{k=0}^{n} 2^k C_{2n-k}^{n} \\ &= \frac{1}{2^{2n}}\cdot S_n, \end{aligned}$$

so

$$S_n = 2^{2n}.$$

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 $N, n_1, n \in \mathbb{N}$ $(n_1 \leq N)$ 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) $F_X : \mathbb{R} \to \mathbb{R}$, $F_X(x) = P(X \leq x) = \sum\limits_{x_i \leq x} p_i$

$(X,Y) : S \to \mathbb{R}^2$ **discrete random vector**:

$-$ **(joint) pdf** $p_{ij} = P\left(X = x_i, Y = y_j\right)$, $(i,j) \in I \times J$,

$-$ **(joint) cdf** $F = F_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R},\ F(x,y) = P(X \leq x, Y \leq y) = \sum\limits_{x_i \leq x} \sum\limits_{y_j \leq y} p_{ij},\ \forall (x,y) \in \mathbb{R}^2$,

$-$ **marginal densities** $p_i = P(X = x_i) = \sum\limits_{j \in J} p_{ij},\ \forall i \in I,\ q_j = P(Y = y_j) = \sum\limits_{i \in I} p_{ij},\ \forall j \in J.$

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** $<=>$ $p_{ij} = P\left(X = x_i, Y = y_j\right) = P\left(X = x_i\right) P\left(Y = y_j\right) = p_i q_j$.

$X+Y \begin{pmatrix} x_i + y_j \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J}, \alpha X \begin{pmatrix} \alpha x_i \\ p_i \end{pmatrix}_{i \in I}, XY \begin{pmatrix} x_i y_j \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J}, X/Y \begin{pmatrix} x_i / y_j \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J} (y_j \neq 0)$

---

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.

Solution:

How many files can be corrupted by the virus? $0, 1$ or 2. So X can take the values $0, 1, 2$. With what probability each?

$$\begin{aligned}
P(X=0) &= P(1^{\text{st}} \text{ no and } 2^{\text{nd}} \text{ no}) = P(1^{\text{st}} \text{ no})P(2^{\text{nd}} \text{ no}) = (1-0.4)(1-0.3) = 0.42;\\
P(X=2) &= P(1^{\text{st}} \text{ yes})P(2^{\text{nd}} \text{ yes}) = 0.4 \cdot 0.3 = 0.12;\\
P(X=1) &= 1 - P(X=0) - P(X=2)\\
&\overset{\text{OR}}{=} P(\{1^{\text{st}} \text{ yes},\ 2^{\text{nd}} \text{ no}\} \cup \{1^{\text{st}} \text{ no},\ 2^{\text{nd}} \text{ yes}\}) = 0.4 \cdot 0.7 + 0.6 \cdot 0.3 = 0.46.
\end{aligned}$$

Actually, the *easiest* way to compute *all* at once, is to use the Poisson model, with "success" meaning a file is corrupted, so parameters are $n = 2$, $p_1 = 0.4$, $p_2 = 0.3$, X is the number of successes, i.e. takes values $X = 0, 1, 2$, with probabilities from

$$\begin{aligned}
(p_1 x + q_1)(p_2 x + q_2) &= (0.4x + 0.6)(0.3x + 0.7)\\
&= 0.12x^2 + 0.46x + 0.42\\
&= P(X=2)x^2 + P(X=1)x + P(X=0).
\end{aligned}$$

So, the pdf of X is

$$X \begin{pmatrix} 0 & 1 & 2 \\ 0.42 & 0.46 & 0.12 \end{pmatrix}$$

Careful!! This is NOT a Poisson variable! A Poisson random variable has NOTHING to do with the Poisson model!!

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 $P(X < 2)$.

<u>**Solution:**</u>
a) X denotes the number of successes in $n = 3$ trials, where "success" means "heads", so probability of success is $p = 1/2$. Thus, X has a Binomial $B(3, 1/2)$ distribution. Its pdf is

$$X \begin{pmatrix} k \\ C_3^k \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{3-k} \end{pmatrix}_{k=\overline{0,3}} = \begin{pmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \end{pmatrix}.$$

b)

$$\begin{aligned}
P(X \le 2) &= F_X(2) = \frac{7}{8} = \mathit{binocdf}(2, 3, 1/2) = 0.875,\\
P(X < 2) &= P(X \le 1) = \frac{1}{2} = F_X(1) = \mathit{binocdf}(1, 3, 1/2) = 0.5.
\end{aligned}$$

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.

<u>**Solution:**</u>
a) New account initiations qualify as rare events, i.e. discrete events observed over a period of time. Then X, the number of today's new accounts has a Poisson distribution. What is the parameter? The parameter λ is the average number of new accounts initiated per day, thus, $\lambda = 10$. So

$$\begin{aligned}
P(A) &= P(X > 8) = 1 - P(X \le 8)\\
&= 1 - \sum_{k=0}^{8} \frac{10^k}{k!} e^{-10} = 1 - F_X(8)\\
&= 1 - \mathit{poisscdf}(8, 10) = 0.6672.
\end{aligned}$$

b) We argue similarly for Y, the number of new accounts opened within 2 days. But it **IS NOT** $2X$!! Instead, it's like this: if 10 new accounts are opened daily, then within 2 days, on average, 20 new accounts will be opened. Thus Y has a Poisson $\mathcal{P}(20)$ distribution. So

$$\begin{aligned} P(B) &= P(Y \le 16) = \sum_{k=0}^{16} \frac{20^k}{k!} e^{-20} \\ &= F_Y(16) = poisscdf(16, 20) \\ &= 0.2211. \end{aligned}$$

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.

Solution:
a) What values can X take? The values $1, 2, \ldots$ We compute each probability:

$$\begin{aligned} P(X = 1) &= 0.7; \\ P(X = 2) &= P(1^{\text{st}} \text{ no and } 2^{\text{nd}} \text{ yes}) = P(1^{\text{st}} \text{ no})P(2^{\text{nd}} \text{ yes}) = (1 - 0.7) \cdot 0.7 = 0.7 \cdot 0.3; \end{aligned}$$

Note: One might argue here that the two events ("1^{st} no" and "2^{nd} yes") are *not* independent, since the very fact that we got to the second attempt to log on was the effect of failing in the first attempt. While that is true, still, the probability to log on is *the same*, no matter how many times we failed before. So, even if we considered them *not* to be independent, we would have

$$P(1^{\text{st}} \text{ no} \cap 2^{\text{nd}} \text{ yes}) = P(1^{\text{st}} \text{ no})P(2^{\text{nd}} \text{ yes} \mid 1^{\text{st}} \text{ no}) = 0.3 \cdot 0.7.$$

Now, proceed further to notice a pattern:

$$\begin{aligned} P(X = 3) &= P(1^{\text{st}} \text{ no and } 2^{\text{nd}} \text{ no and } 3^{\text{rd}} \text{ yes}) = (1 - 0.7)(1 - 0.7)0.7 = 0.7 \cdot (0.3)^2; \\ \ldots & \ldots \\ P(X = k) &= 0.7 \cdot (0.3)^{k-1} \\ \ldots & \ldots \end{aligned}$$

So, the pdf of X is

$$X \begin{pmatrix} k \\ (0.7)(0.3)^{k-1} \end{pmatrix}_{k=1,2,\ldots} \quad \text{and} \quad X - 1 \begin{pmatrix} l \\ (0.7)(0.3)^{l} \end{pmatrix}_{l=0,1,\ldots}$$

The variable $X - 1$ (the number of *failures* that occurred before the first success) has a Geometric distribution with parameter $p = 0.7$. Variable X (the number of *trials* needed to get the first success) has an "almost" $Geo(0.7)$ distribution. This is known as the *Shifted Geometric* distribution.
Note: In some books, X is called a Geometric random variable.

b)

$$P(X \le 4) = P(X - 1 \le 3) = F_X(4) = F_{X-1}(3) = geocdf(3, 0.7) = 0.9919.$$

c)

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) \;=\; 1 - P(X \leq 2) \;=\; 1 - F_X(2) \\ &= 1 - F_{X-1}(1) \;=\; 1 - geocdf(1, 0.7) \;=\; 0.09. \end{aligned}$$

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.

Solution:
a) X is the number picked, so it can take the values $1, 2, \ldots, 5$, all equally probable. So the pdf is

$$X \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{pmatrix},$$

which is a discrete Uniform $U(5)$ distribution.

Y takes values 1 (if $X = 1$), 2 (if $X = 2, 3$ or 5) and 3 (if $X = 4$). So its pdf is

$$Y \begin{pmatrix} 1 & 2 & 3 \\ \frac{1}{5} & \frac{3}{5} & \frac{1}{5} \end{pmatrix}.$$

b) Recall, when doing operations with *discrete* random variables, we operate on their *values*.
So, we get all the possible values of $X + Y$ by adding each possible value of X to each possible value of Y. Thus, $X + Y$ can take values: $2, 3, 4, 5, 6, 7$ and 8. Now, for the probabilities:

$$P(X + Y = 2) \;=\; P(X = 1, Y = 1),$$

where the comma means $\cap$. Now, are the events $(X = 1)$ and $(Y = 1)$ independent? **Not at all!** The value of Y is obviously very much depending on the value of X, in fact, it is *completely* determined by the value of X. Once the value of X is known, we automatically have the value of Y. As events,

$$(X = 1) \text{ implies (induces) } (Y = 1), \text{i.e.}$$

$$\begin{aligned} (X = 1) &\subseteq (Y = 1), \text{i.e.} \\ (X = 1) \cap (Y = 1) &= (X = 1). \end{aligned}$$

So, $P(X = 1, Y = 1) = P(X = 1) = 1/5$. Similarly we compute the other probabilities. Notice that, because of the dependence of X and Y, some combinations of values are *impossible*.

$$\begin{aligned} P(X + Y = 2) &= P(X = 1) \;=\; \frac{1}{5}, \\ P(X + Y = 3) &= P\Big((X = 1, Y = 2) \cup (X = 2, Y = 1)\Big) \;=\; 0 + 0 \;=\; 0 \text{ (both are impossible)}, \\ P(X + Y = 4) &= P\Big((X = 1, Y = 3) \cup (X = 2, Y = 2) \cup (X = 3, Y = 1)\Big) \\ &= P(X = 2, Y = 2) \;=\; P(X = 2) \;=\; \frac{1}{5}, \\ \ldots \quad & \ldots \quad \ldots \end{aligned}$$

In the end (recall that we *do not* list in the pdf values with probability 0), the pdf of $X + Y$ is

$$X + Y \begin{pmatrix} 2 & 4 & 5 & 7 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{2}{5} \end{pmatrix}.$$

In a similar fashion, we find the pdf of $X \cdot Y$. The variable $X \cdot Y$ can take values $1, 2, 3, 4, 5, 6, 8, 9, 10, 12$ and 15. Again, some combinations are impossible. For example,

$$\begin{aligned} P(X \cdot Y = 4) &= P\Big((X = 2, Y = 2) \cup (X = 4, Y = 1)\Big) \\ &= P(X = 2, Y = 2) \;=\; P(X = 2) \;=\; \frac{1}{5}. \end{aligned}$$

In the end, the pdf of $X \cdot Y$ is

$$X \cdot Y \begin{pmatrix} 1 & 4 & 6 & 10 & 12 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{pmatrix}$$

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?

Solution:

Part a) is the same.

b) The difference is that now X and Y are **independent**, so *all* combinations of values are possible. For example,

$$\begin{aligned} P(X + Y = 7) &= P(X = 4, Y = 3) + P(X = 5, Y = 2) \\ &\overset{\text{ind}}{=} P(X = 4)P(Y = 3) + P(X = 5)P(Y = 2) \\ &= \frac{1}{25} + \frac{3}{25} \;=\; \frac{4}{25}. \end{aligned}$$

Thus, now the pdf's are

$$X + Y \begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \frac{1}{25} & \frac{4}{25} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{4}{25} & \frac{1}{25} \end{pmatrix}$$

$$X \cdot Y \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 8 & 9 & 10 & 12 & 15 \\ \frac{1}{25} & \frac{4}{25} & \frac{2}{25} & \frac{4}{25} & \frac{1}{25} & \frac{4}{25} & \frac{3}{25} & \frac{1}{25} & \frac{3}{25} & \frac{1}{25} & \frac{1}{25} \end{pmatrix}$$

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:

$X \setminus Y$	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.

Solution:

a) To get the *marginal* pdf's, i.e. the pdf's of the components X and Y, for each fixed value of one of them, we add all the values on that row or column, hence, getting them on the *margins*.

$X \backslash Y$	1	2	3	
1	0	0.10	0.40	0.50
2	0.06	0.10	0.10	0.26
3	0.06	0.04	0	0.10
4	0.10	0.04	0	0.14
	0.22	0.28	0.50	

So, the marginal pdf's are

$$X\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0.5 & 0.26 & 0.1 & 0.14 \end{pmatrix} \text{ and } Y\begin{pmatrix} 1 & 2 & 3 \\ 0.22 & 0.28 & 0.5 \end{pmatrix}.$$

b) We get the value from the table.

$$P(B) = P\Big((X,Y) = (2,1)\Big) = 0.06$$

c) Same here, but we have more combinations.

$$\begin{aligned} P(C) &= P(X \geq 3, Y \leq 2) \\ &= P\Big((X = 3 \cup X = 4) \cap (Y = 1 \cup Y = 2)\Big) \\ &= P(X = 3, Y = 1) + P(X = 3, Y = 2) + P(X = 4, Y = 1) + P(X = 4, Y = 2) \\ &= 0.06 + 0.04 + 0.1 + 0.04 = 0.24. \end{aligned}$$

d) The total charge for a customer is given by the number of total hours spent online multiplied by the price per each hour, i.e. $Z = X \cdot Y$. Z can take the values $\{1, 2, 3, 4, 6, 8, 9, 12\}$. We have:

$$\begin{array}{rclclcl} P(Z=1) &=& P(X=1,Y=1) &=& 0, && \\ P(Z=2) &=& P(X=1,Y=2) &+& P(X=2,Y=1) &=& 0.16, \\ P(Z=3) &=& P(X=1,Y=3) &+& P(X=3,Y=1) &=& 0.46, \\ & \ldots & \ldots &&&& \end{array}$$

So, the pdf is

$$Z\begin{pmatrix} 2 & 3 & 4 & 6 & 8 \\ 0.16 & 0.46 & 0.2 & 0.14 & 0.04 \end{pmatrix}.$$

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 $g' \neq 0$, strictly monotone

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

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

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

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

Solution:
a) We find the constant k from the condition $\int_{\mathbb{R}} f(x)\,dx = 1$. We have

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)dx &= k\int_{1}^{\infty} \frac{1}{x^4}dx = k\int_{1}^{\infty} x^{-4}\,dx\\ &= -\frac{k}{3}\,\frac{1}{x^3}\Big|_1^{\infty} = \frac{k}{3} = 1,\end{aligned}$$

so $k = 3$ and

$$f(x) = \frac{3}{x^4},\ x \geq 1.$$

b) The cdf is found by

$$F(x) = \int_{-\infty}^{x} f(t)\,dt.$$

If $x < 1$, we integrate the constant 0, so $F(x) = 0$. For $x \geq 1$, we have

$$\begin{aligned}F(x) &= \int_{-\infty}^{x} f(t)dt = \int_{1}^{x} \frac{3}{t^4}dt\\ &= 3\int_{1}^{x} t^{-4}\,dt = -\frac{1}{t^3}\Big|_1^{x} = 1 - \frac{1}{x^3}.\end{aligned}$$

Thus,

$$F(x) = \begin{cases} 0, & x < 1\\ 1 - \dfrac{1}{x^3}, & x \geq 1.\end{cases}$$

c) Recall that X is the *lifetime*, in years, of the component. So, we want

$$\begin{aligned}P(X > 2) &= 1 - P(X \leq 2) = 1 - F(2)\\ &= 1 - \left(1 - \frac{1}{8}\right) = \frac{1}{8} = 0.125.\end{aligned}$$

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 4:55 or after 5:05?

Solution:

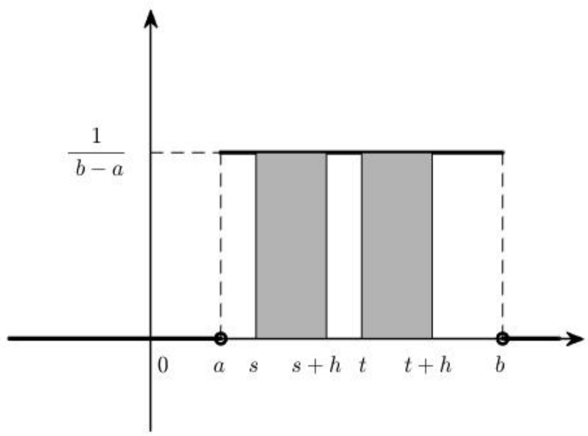
Recall that

$$P(a < X \leq b) = F(b) - F(a) = \int_a^b f(t)dt$$

and the interval can be open or closed at either endpoint. Then

$$\begin{aligned} P(s < X < s+h) &= \int_s^{s+h} \frac{1}{b-a} dx = \frac{1}{b-a} x \Big|_s^{s+h} = \frac{h}{b-a} \\ P(t < X < t+h) &= \int_t^{t+h} \frac{1}{b-a} dx = \frac{1}{b-a} x \Big|_t^{t+h} = \frac{h}{b-a}. \end{aligned}$$

Also, we can see it graphically. Each probability is the area of a rectangle whose height is $\frac{1}{b-a}$. So, as long as the two rectangles have also the same *base*, the areas will be equal, no matter where the starting point s (or t) is in the interval $[a, b]$.



Example:

a) Since the flight can arrive at *any* time in that interval, it means the time is Uniformly distributed in the interval $[4:50, 5:10]$, so $X \in U(4:50, 5:10)$.

b) The flight is *just as likely* to arrive between 4:50 and 4:55 as it is to arrive between 5 and 5:05, because the two intervals have the *same length*.

Same for before 4:55 or after 5:05, just as likely, for the same reason.

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?

Solution:

Since computer breakdowns qualify as rare events (they cannot occur simultaneously), the time between two consecutive breakdowns has Exponential distribution. Since breakdowns occur every 5 months, their frequency is $\lambda = 1/5$ per month. So the distribution is $Exp(1/5)$. Now, since breakdowns are independent of each other, the time T until the third breakdown (when a special maintenance is required) is the sum of 3 $Exp(1/5)$ variables and,

hence, has a $Gamma(3,5)$ distribution.
a) Then we want

$$P(T \leq 9) \;=\; F_T(9) \;\overset{\text{Matlab}}{=}\; \mathit{gamcdf}(9,3,5) \;=\; 0.2694.$$

b) This is *conditional* probability.

$$\begin{aligned} P(T > 12+4 \,|\, T > 12) &= \frac{P(T > 16)}{P(T > 12)} = \frac{1 - P(T \leq 16)}{1 - P(T \leq 12)} \\ &= \frac{1 - F_T(16)}{1 - F_T(12)} \overset{\text{Matlab}}{=} \frac{1 - \mathit{gamcdf}(16,3,5)}{1 - \mathit{gamcdf}(12,3,5)} = 0.6668. \end{aligned}$$

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 $P(X \leq 1)$.

Solution:
We find the marginal pdf of one of the component by integrating over $\mathbb{R}$ with respect to the *other* variable. Let us consider f_X first. In order for the vector to have a nonzero pdf, namely $\frac{1}{16}x^3y^3$, *both* x and y have to be in $[0,2]$. So, if $x \notin [0,2]$, we integrate the function 0 and, thus, $f_X(x) = 0$. For $x \in [0,2]$, we have

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f(x,y)\,dy = \frac{1}{16}\int_0^2 x^3y^3 dy \\ &= \frac{x^3}{16}\int_0^2 y^3\,dy = \frac{x^3}{16}\cdot\frac{1}{4}y^4\Big|_{y=0}^{y=2} = \frac{1}{4}x^3. \end{aligned}$$

Thus,

$$f_X(x) \;=\; \frac{1}{4}x^3,\; x \in [0,2].$$

By symmetry (both the function $f(x,y)$ and the domain $[0,2]\times[0,2]$ are symmetric), we get the same result for f_Y. So

$$f_Y(y) \;=\; \frac{1}{4}y^3,\; y \in [0,2].$$

b) We check that

$$f_{(X,Y)}(x,y) = f_X(x)\cdot f_Y(y), \forall (x,y) \in \mathbb{R}^2.$$

First, the trivial case. If $x \notin [0,2]$ or $y \notin [0,2]$ (or both), the left-hand side is 0 and at least one of the factors on the right is 0, so we have equality $0 = 0$.
If both $x, y \in [0,2]$, then we have

$$\frac{1}{16}x^3y^3 \;=\; \frac{1}{4}x^3\cdot\frac{1}{4}y^3,$$

equality again, so, yes, they are independent.

c)

$$\begin{aligned} P(X \leq 1) &= \int_{-\infty}^{1}\int_{-\infty}^{\infty} f(x,y)\,dydx = \int_{-\infty}^{1} f_X(x)\,dx \\ &= \frac{1}{4}\int_0^1 x^3\,dx = \frac{1}{4}\cdot\frac{1}{4}x^4\Big|_0^1 = \frac{1}{16}. \end{aligned}$$

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.

Solution:
Here, $Y = g(X)$, with $g : \mathbb{R} \to \mathbb{R}$

$$g(x) = \frac{1}{2}x + 2.$$

So, first off, since g is a linear function, its range is $g(\mathbb{R}) = \mathbb{R}$.
Next, we check the conditions of the theorem on the function g. We have $g'(x) = \frac{1}{2}$, which exists everywhere and is never equal to 0. Moreover, since $g'(x) = \frac{1}{2} > 0$, $\forall x \in \mathbb{R}$, g is *strictly increasing*. Thus, it satisfies all the hypotheses and

$$f_Y(y) = \frac{f_X\left(g^{-1}(y)\right)}{|g'\left(g^{-1}(y)\right)|},$$

for every $y \in \mathbb{R}$.
Since the derivative of g is a constant, the denominator is simply equal to $\frac{1}{2}$.
For the numerator, we need the *inverse* of g. Recall we find the inverse by solving for x

$$y = g(x) \;<=>\; x = g^{-1}(y).$$

We have

$$\begin{aligned} \frac{1}{2}x + 2 &= y \\ \frac{1}{2}x &= y - 2 \\ x &= 2(y-2). \end{aligned}$$

So

$$g^{-1}(y) = 2(y-2).$$

Then the numerator will be (simple composition of functions)

$$\begin{aligned} f_X\left(g^{-1}(y)\right) &= f_X\left(2(y-2)\right), \\ &= \frac{1}{4}\big(2(y-2)\big)e^{-\frac{(2(y-2))}{2}},\; 2(y-2) \geq 0, \\ &= \frac{1}{2}(y-2)e^{-(y-2)}, y \geq 2. \end{aligned}$$

Thus,

$$f_Y(y) = (y-2)e^{2-y}, y \geq 2 \;\text{(and } 0, \text{ otherwise).}$$

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

Solution:
For $X \in N(0,1)$, the pdf is

$$f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2},\; \forall x \in \mathbb{R}.$$

Main idea: Start with the cdf.

$$F_Y(x) = P(Y \leq x) = P\left(|X| \leq x\right).$$

Now stop, to analyze the situation. Since $|X| \geq 0$, it is *impossible* for $|X|$ to be less than or equal to a negative number, so for $x < 0$, $F_Y(x) = 0$ and, hence,

$$f_Y(x) = F'_Y(x) = 0,\; x < 0.$$

For $x \geq 0$, we rewrite the inequality until *it is an inequality for X*, whose pdf is *known*, so probabilities about X can be computed. We have

$$F_Y(x) \;=\; P\left(-x \leq X \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-x}^{x} e^{-\frac{1}{2}t^2}\, dt.$$

Now, the function $e^{-\frac{t^2}{2}}$ has no primitive, so what do we do, how do we compute that integral? **We don't!** Remember what we want is the *pdf*, i.e. the *derivative of F*, so we want to go with the computations just far enough to be able to find its derivative. Even if we could compute it, there's no need to! All we want is to write it in the form

$$G(x) \;=\; \int_{a}^{x} g(t)\, dt,\; a \in \mathbb{R},$$

for which the derivative is

$$G'(x) \;=\; g(x).$$

Or, more generally,

$$G(x) \;=\; \int_{a}^{h(x)} g(t)\, dt,\; a \in \mathbb{R} \;<=>\; G'(x) \;=\; g\big(h(x)\big) \cdot h'(x),$$

which also involves the derivative of a composite function.
So, to get back to our integral, let us recall another thing about integration and symmetry:

$$\int_{-a}^{a} f(x)\, dx \;=\; \begin{cases} 2\displaystyle\int_{0}^{a} f(x)\, dx, & \text{if the function is even, i.e. } f(-x) = f(x) \\ \quad 0, & \text{if the function is odd, i.e. } f(-x) = -f(x) \end{cases}.$$

Since we are integrating over a symmetric interval and the function $e^{-\frac{1}{2}t^2}$ is even, we have

$$F_Y(x) \;=\; 2\frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{1}{2}t^2}\, dt \;=\; \sqrt{\frac{2}{\pi}} \int_{0}^{x} e^{-\frac{1}{2}t^2}\, dt.$$

Now we can take the derivative

$$F_Y'(x) \;=\; \sqrt{\frac{2}{\pi}}\, e^{-\frac{1}{2}x^2}.$$

Thus, the pdf of Y is

$$f_Y(x) \;=\; \sqrt{\frac{2}{\pi}}\, e^{-\frac{1}{2}x^2},\; x \geq 0 \;\text{ (and } 0, \text{ otherwise).}$$

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. $E(aX+b)=aE(X)+b, V(aX+b)=a^2V(X)$
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. $V\left(\sum\limits_{i=1}^{n} a_iX_i\right)=\sum\limits_{i=1}^{n} a_i^2V(X_i)+2\sum\limits_{1\le i<j\le n} a_ia_j\mathrm{cov}(X_i,X_j)$
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

$$E\left(h(X,Y)\right)=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} h(x,y)f(x,y)dxdy$$

---

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?

Solution:

Variable X represents the number of daily network blackouts and each network blackout costs \$500. Then the company's daily loss due to blackouts is the variable

$$Y = 500X.$$

a) We want the *expected* daily loss due to blackouts, so $E(Y)$. First, let us compute $E(X)$. Since X is a *discrete* random variable, we compute its expectation as a sum.

$$E(X) \;=\; \sum_{i\in I} x_i p_i \;=\; 0\cdot 0.7 + 1\cdot 0.2 + 2\cdot 0.1 \;=\; 0.4.$$

Then, by the properties of expectation,

$$E(Y) \;=\; E(500X) \;=\; 500E(X) \;=\; 500\cdot 0.4 \;=\; 200 \text{ dollars}.$$

b) Now we want

$$\sigma(Y) \;=\; \sqrt{V(Y)}.$$

Again, let us first compute $V(X)$. For that, we need $E(X^2)$. The pdf of X^2 is

$$X^2 \begin{pmatrix} 0 & 1 & 4 \\ 0.7 & 0.2 & 0.1 \end{pmatrix},$$

so

$$\begin{aligned} E(X^2) &= 0.2 + 0.4 \;=\; 0.6, \\ V(X) &= E(X^2) - (E(X))^2 \;=\; 0.6 - 0.16 \;=\; 0.44. \end{aligned}$$

Then, by the properties of variance,

$$\begin{aligned} V(Y) &= V(500X) \;=\; (500)^2 V(X) \;=\; 250000\cdot 0.44 \;=\; 110000 \text{ dollars}^2 \text{ and} \\ \sigma(Y) &= \sqrt{V(Y)} \;=\; 331.6625 \text{ dollars}. \end{aligned}$$

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*?

Solution:

Let X denote the number of users that close Windows properly before someone does not. If "success" means "a person *does not* close Windows properly", then X has a Geometric distribution with parameter $p = 0.1$. We want the "average" of that number, i.e. the mean value $E(X)$. For the *Geo*(p) distribution, the mean value is $\frac{q}{p}$. Thus,

$$E(X) \;=\; \frac{q}{p} \;=\; \frac{0.9}{0.1} \;=\; 9.$$

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

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

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

Solution:
The variable X is the lifetime of the component, we want to know how long can we *expect* it to last, so, again, $E(X)$. This is a *continuous* random variable, so we compute its expectation as an integral. We have

$$\begin{aligned} E(X) &= \int_{\mathbb{R}} x f(x)\, dx = 3\int_1^\infty x \cdot \frac{1}{x^4}\, dx \\ &= 3\int_1^\infty x^{-3}\, dx = -\frac{3}{2}\frac{1}{x^2}\Big|_1^\infty = \frac{3}{2}, \end{aligned}$$

so about a year and a half.

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?

Solution:
So, we want to maximize the *expected* return, that means the expected value of the return and minimize the risk. How do we quantify the "risk"? Once we know the average (expected) value of the return, "risk" would be the amount of *variability* from that expected return, i.e. its variance (or standard deviation). For an "optimal" portfolio, we will want *high* expected return and *low* variance.
Let A denote the actual (random) return from each share of company A and B the same for company B. Then their pdf's are

$$A\begin{pmatrix} -1 & 0 & 2 \\ 0.2 & 0.2 & 0.6 \end{pmatrix} \text{ and } B\begin{pmatrix} -1 & 0 & 3 \\ 0.3 & 0.1 & 0.6 \end{pmatrix}.$$

The pdf's of A^2 and B^2 (which will be needed for the computation of the variances) are

$$A^2\begin{pmatrix} 0 & 1 & 4 \\ 0.2 & 0.2 & 0.6 \end{pmatrix} \text{ and } B^2\begin{pmatrix} 0 & 1 & 9 \\ 0.1 & 0.3 & 0.6 \end{pmatrix}.$$

We have

$$\begin{aligned} E(A) &= -0.2 + 1.2 = 1, & E(B) &= -0.3 + 1.8 = 1.5, \\ E(A^2) &= 0.2 + 2.4 = 2.6, & E(B^2) &= 0.3 + 5.4 = 5.7, \\ V(A) &= 2.6 - 1^2 = 1.6, & V(B) &= = 5.7 - (1.5)^2 = 3.45. \end{aligned}$$

Now we can make a comparison between the three investments, by looking at their expected values and their variances.
a) Investing **all money in company A**, i.e. \$600 at \$20 per share, means buying 30 shares from company A. So the return (profit) from this investment is the random variable $30A$, for which
$E(30A) = 30E(A) = 30$ (the expected return)
$V(30A) = 900V(A) = 1440$ (the variance of the return, i.e. the risk of the investment)

b) Investing **all money in company B**, i.e. \$600 at \$30 per share, means buying 20 shares from company B. The return from this investment is the random variable $20B$, for which
$E(20B) = 20E(B) = 30$ (the expected return)
$V(20B) = 400V(B) = 1380$ (the variance of the return, i.e. the risk of the investment)

c) Finally, investing **half the amount in each**, i.e. $\$300$ at $\$20$ per share, means buying 15 shares of company A stock and $\$300$ at $\$30$ per share, means buying 10 shares of company B stock. In this case, the return from this investment is the random variable $15A + 10B$, for which
$E(15A + 10B) = 15E(A) + 10E(B) = 30$ (the expected return)
and, because A and B are independent,
$V(15A + 10B) = 225V(A) + 100V(B) = 705$ (the variance of the return, i.e. the risk of the investment)

So **all** three proposed portfolios have the same expected return (which should be no surprise, since each share of each company is expected to return $1/20$ or $1.5/30$, which is 5%), but the **third** one, with the lowest variance is the **least risky**.
This is why financiers and brokers always say "in order to minimize the risk, D I V E R S I F Y the portfolio"..

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 $Y = \dfrac{X - E(X)}{\sigma(X)}$.

Solution:
This is a simple linear transformation, that can be applied to every random variable (discrete or continuous) that has a mean value and a variance (i.e. they are finite):

$$Y = \frac{1}{\sigma(X)}X - \frac{E(X)}{\sigma(X)} = aX + b.$$

The "reduction" refers to the mean value and variance of the reduced variable. Recall the properties

$$\begin{aligned} E(aX + b) &= aE(X) + b, \\ V(aX + b) &= a^2V(X). \end{aligned}$$

So

$$\begin{aligned} E(Y) &= \frac{1}{\sigma(X)}E(X) - \frac{E(X)}{\sigma(X)} = 0, \\ V(Y) &= \frac{1}{\sigma^2(X)}V(X) = 1. \end{aligned}$$

Y is called the **reduced variable of** X.

Recall that for a Normal $N(\mu, \sigma)$ variable, $E(X) = \mu, V(X) = \sigma^2$.
So, returning to reduced variables, in particular, if $X \in N(\mu, \sigma)$, then $\dfrac{X - \mu}{\sigma} \in N(0, 1)$ (Standard or Reduced Normal). We see many such reduced variables in Statistics:

$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}, \text{ or, more generally, } \frac{\overline{\theta} - \theta}{\sigma_{\overline{\theta}}}.$$

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 $\rho(X, Y)$.

Solution:
To compute *all* of these numerical characteristics, we use the formula

$$E\left(h(X, Y)\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x, y)f(x, y)dxdy.$$

a)

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)\,dx\,dy \\
&= \int_0^1\int_0^1 (x^2+xy)\,dy\,dx \;=\; \int_0^1 \left(x^2\cdot y + x\cdot\frac{1}{2}y^2\right)\Bigg|_{y=0}^{y=1} dx \\
&= \int_0^1 \left(x^2+\frac{1}{2}x\right) dx \;=\; \left(\frac{1}{3}x^3+\frac{1}{4}x^2\right)\Bigg|_0^1 \;=\; \frac{7}{12},
\end{aligned}$$

and by symmetry, $E(Y)=\dfrac{7}{12}$, also.

$$E\left(X^2\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 f(x,y)\,dx\,dy = \int_0^1\int_0^1 (x^3+x^2y)\,dx\,dy = \frac{5}{12} = E\left(Y^2\right),$$

again, by symmetry.
That means the variances are also equal.

$$V(X) = V(Y) = \frac{5}{12}-\left(\frac{7}{12}\right)^2 = \frac{11}{12^2}.$$

b) This formula is *especially* useful for computing the covariance, since it avoids the (many times) complicated calculation of the pdf of the variable XY. We have

$$\begin{aligned}
E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f(x,y)\,dx\,dy \\
&= \int_0^1\int_0^1 (x^2y+xy^2)\,dx\,dy \;=\; \frac{1}{3},
\end{aligned}$$

so

$$\begin{aligned}
\operatorname{cov}(X,Y) &= E(XY)-E(X)E(Y) \\
&= \frac{1}{3}-\left(\frac{7}{12}\right)^2 \;=\; -\frac{1}{12^2} \\
\rho(X,Y) &= \frac{\operatorname{cov}(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}} \\
&= -\frac{\frac{1}{12^2}}{\frac{11}{12^2}} \;=\; -\frac{1}{11}.
\end{aligned}$$

The value of $|\rho(X,Y)|$ is so close to 0, which means there is very weak (almost nonexistent) linear relationship between X and 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).

Solution:

Recall that

$$X, Y \text{ are independent} \Rightarrow X, Y \text{ are uncorrelated, i.e. } \rho(X, Y) = 0.$$

That makes sense, since independence means there is no relationship of *any kind* between the variables, including linear. Obviously, independence is a much stronger condition than uncorrelation. A linear relationship does not exist, but some *other* form of relationship may exist. This exercise is an example in that sense, variables that are uncorrelated, but *not* independent.

Notice that for any $k \in \mathbb{N}^*$, $Y_k = X^{2k-1}$ has the same pdf as

$$X\begin{pmatrix} -1 & 0 & 1 \\ \sin^2 a & \cos 2a & \sin^2 a \end{pmatrix}$$

and $Z_k = X^{2k}$ has the same pdf as

$$X^2\begin{pmatrix} 0 & 1 \\ \cos 2a & 2\sin^2 a \end{pmatrix}.$$

Then

$$\begin{aligned} E\left(X^{2k-1}\right) &= -\sin^2 a + \sin^2 a = 0, \\ E\left(X^{2k}\right) &= 2\sin^2 a, \\ E\left(X^{2k-1}X^{2k}\right) &= E\left(X^{4k-1}\right) = 0, \end{aligned}$$

so

$$\text{cov}\,(Y, Z) = \rho\,(Y, Z) = 0.$$

In particular, for $k = 1$, we have that $\rho\left(X, X^2\right) = 0$, so they are uncorrelated, but obviously, not independent.

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 $X_1, \dots, X_n$ 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 $n \to \infty$,

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

Point Estimators

- method of moments: solve the system $\nu_k = \overline{\nu}_k$, for as many parameters as needed ($k = 1, \dots,$ nr. of unknown parameters);

- method of maximum likelihood: solve $\dfrac{\partial \ln L(X_1, \dots, X_n; \theta)}{\partial \theta_j} = 0$, where $L(X_1, \dots, X_n; \theta) = \prod_{i=1}^{n} f(X_i; \theta)$ is the likelihood function;

- **standard error** of an estimator $\overline{\theta}$: $\sigma_{\hat{\theta}} = \sigma(\overline{\theta}) = \sqrt{V(\overline{\theta})}$;

- **Fisher information** $I_n(\theta) = -E\left[\dfrac{\partial^2 \ln L(X_1, \dots, X_n; \theta)}{\partial \theta^2}\right]$; 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.

Solution:

Let X be a r.v. with mean $E(X) = \mu$ and standard deviation $\sigma(X) = \sqrt{V(X)} = \sigma$. In Chebyshev's inequality let $\varepsilon = k\sigma$, for $k = \overline{1,3}$. Then

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

For $k = 1$ that means

$$P(|X - \mu| < \sigma) \quad \geq \quad 0 \text{ (not much).}$$

For $k = 2$, we have

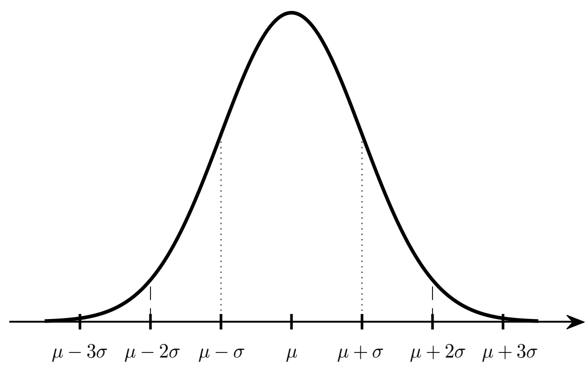
$$P(|X - \mu| < 2\sigma) \quad \geq \quad 1 - \frac{1}{4} \quad = \quad \frac{3}{4} \quad = \quad 0.75.$$

Finally, for $k = 3$, we get

$$P(|X - \mu| < 3\sigma) \;\geq\; 1 - \frac{1}{9} \;=\; \frac{8}{9} \;\approx\; 0.89.$$

The **3$\boldsymbol{\sigma}$ Rule** states that: most of the values (at least 89%) that a random variable takes, lie within 3 standard deviations (3σ) away from the mean.
Now, for *symmetric* distributions, these probabilities are *much* higher. For $X \in N(\mu, \sigma)$ (recall, for $X \in N(\mu, \sigma), E(X) = \mu, V(X) = \sigma^2$), we have:

$$\begin{aligned} P(|X - \mu| < \sigma) &\approx 0.68, \\ P(|X - \mu| < 2\sigma) &\approx 0.95, \\ P(|X - \mu| < 3\sigma) &\approx 0.99. \end{aligned}$$



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.

Solution:
Let X denote the number of "heads" that appear when tossing a coin 1000 times. Then $X \in Bino\left(1000, \frac{1}{2}\right)$, so

$$\begin{aligned} E(X) &= np = 500, \\ V(X) &= npq = 250. \end{aligned}$$

Now, the problem asks about the "chance" (i.e. probability) that this number is between 450 and 550, so about

$$P(450 < X < 550).$$

Chebyshev's inequality gives information about

$$P(|X - 500| < \varepsilon),$$

which we rewrite as

$$\begin{aligned} P(|X - 500| < \varepsilon) &= P(-\varepsilon < X - 500 < \varepsilon) \\ &= P(500 - \varepsilon < X < 500 + \varepsilon). \end{aligned}$$

So, we take $\varepsilon = 50$ in Chebyshev's inequality. Then we get

$$\begin{aligned} P(450 < X < 550) &= P(500 - 50 < X < 500 + 50) \\ &= P(-50 < X - 500 < 50) \\ &= P(|X - 500| < 50) \\ &\geq 1 - \frac{250}{\varepsilon^2} \\ &= 1 - \frac{250}{2500} = 0.9, \end{aligned}$$

so the statement is true.

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

<u>**Solution**</u>:

Let X_i denote the time it takes to download file i. Then, for every $i = 1, \ldots, 82$,

$$\begin{aligned} \mu &= E(X_i) = 15 \text{ sec}, \\ \sigma &= \sqrt{V(X_i)} = \sqrt{16} = 4 \text{ sec}. \end{aligned}$$

The *entire* software is installed in

$$S_{82} = X_1 + X_2 + \ldots X_{82} \text{ sec}.$$

We have a sample of size $n = 82$, $X_1, X_2, \ldots, X_{82}$. Convert the time into seconds, $20\ \mathrm{min} = 1200\ \mathrm{sec}$. So, we want to compute

$$P(S_{82} < 1200).$$

By the CLT, we have

$$\begin{aligned} P(S_{82} < 1200) &= P\left(\frac{S_{82} - n\mu}{\sigma\sqrt{n}} < \frac{1200 - n\mu}{\sigma\sqrt{n}}\right) \\ &= P\left(Z_n < \frac{1200 - 82 \cdot 15}{4 \cdot \sqrt{82}}\right) \\ &= P(Z_n < -0.8282) \overset{\text{CLT}}{\approx} P(Z < -0.8282) \\ &= F_Z(-0.8282) = \mathit{normcdf}(-0.8282) = 0.2038. \end{aligned}$$

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.

<u>**Solution**</u>:

<u>Method of moments</u>:

There is only one unknown, θ, so we solve the system

$$\nu_1 = \overline{\nu}_1,$$

where

$$\begin{aligned} \nu_1 &= E(X) = \int_{\mathbb{R}} x f(x)\, dx \\ &= \theta \int_0^1 x^\theta\, dx \\ &= \frac{\theta}{\theta+1} x^{\theta+1} \Big|_{x=0}^{x=1} = \frac{\theta}{\theta+1} \end{aligned}$$

and

$$\begin{aligned} \overline{\nu}_1 &= \overline{x} = \frac{x_1 + x_2 + x_3}{3} \\ &= \frac{0.4 + 0.7 + 0.9}{3} = \frac{2}{3}. \end{aligned}$$

Solve for θ

$$\begin{aligned}\frac{\theta}{\theta+1} &= \overline{x}, \\ \theta &= \overline{x}(\theta+1), \\ \theta(1-\overline{x}) &= \overline{x},\end{aligned}$$

to get

$$\overline{\theta} = \frac{\overline{x}}{1-\overline{x}} = 2.$$

<u>Method of maximum likelihood</u>:
Again, having only one unknown, θ, we have only one equation

$$\frac{\partial \ln L(x_1,\ldots,x_n;\theta)}{\partial\theta} = 0.$$

The likelihood function is the joint density of the vector (X_1, X_2, X_3) :

$$\begin{aligned}L(x_1,x_2,x_3;\theta) &= \prod_{i=1}^{3} f(x_i;\theta) \\ &= \prod_{i=1}^{3}\left(\theta x_i^{\theta-1}\right) = \theta^3\left(\prod_{i=1}^{3} x_i\right)^{\theta-1} \\ \ln L &= 3\ln\theta + (\theta-1)\sum_{i=1}^{3}\ln x_i \\ \frac{\partial \ln L}{\partial\theta} &= \frac{3}{\theta} + \sum_{i=1}^{3}\ln x_i.\end{aligned}$$

Solve $\dfrac{\partial \ln L}{\partial\theta} = 0$ for θ, to find

$$\hat{\theta} = -\frac{3}{\displaystyle\sum_{i=1}^{3}\ln x_i} = 2.1766.$$

<u>Note</u>: In the case where the two estimators *do not* coincide, the MLE is more trustworthy.

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.

<u>Solution</u>:
a) Again, there is only one unknown, θ, so we solve the system $\nu_1 = \overline{\nu}_1$, where

$$\nu_1 = E(X) = 2\theta$$

and

$$\overline{\nu}_1 = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Solve for θ

$$2\theta = \overline{X},$$

to get

$$\overline{\theta} = \frac{1}{2}\overline{X}.$$

Note In this case, this is also the MLE.
Now, *efficiency* is only computed for *absolutely correct* estimators. Let us check the absolute correctness. We have

$$\begin{aligned} E(\overline{\theta}) &= E\left(\frac{1}{2}\overline{X}\right) = \frac{1}{2}E(\overline{X}) \\ &= \frac{1}{2}\cdot\mu = \frac{1}{2}\cdot 2\theta = \theta \end{aligned}$$

and

$$\begin{aligned} V(\overline{\theta}) &= V\left(\frac{1}{2}\overline{X}\right) = \frac{1}{4}V(\overline{X}) \\ &= \frac{1}{4}\cdot\frac{\sigma^2}{n} = \frac{1}{4n}4\theta^2 = \frac{\theta^2}{n} \overset{n\to\infty}{\longrightarrow} 0, \end{aligned}$$

so $\overline{\theta}$ is an absolutely correct estimator for θ.

b) The efficiency is given by

$$e(\overline{\theta}) = \frac{1}{I_n(\theta)V(\overline{\theta})},$$

where

$$I_n(\theta) = -E\left[\frac{\partial^2 \ln L(X_1,\ldots,X_n;\theta)}{\partial\theta^2}\right]$$

is the Fisher information.
Since the range of X *does not* depend on θ, we have

$$I_n(\theta) = nI_1(\theta),$$

where

$$I_1(\theta) = -E\left[\frac{\partial^2 \ln L(X_1;\theta)}{\partial\theta^2}\right].$$

We proceed with the computations:

$$\begin{aligned} L(X_1;\theta) &= \frac{1}{2\theta}e^{-\frac{1}{2\theta}X_1}, \\ \ln L(X_1;\theta) &= -\ln(2\theta) - \frac{1}{2\theta}X_1 = -\ln 2 - \ln\theta - \frac{1}{2\theta}X_1, \\ \frac{\partial \ln L}{\partial\theta} &= -\frac{1}{\theta} + \frac{1}{2}\cdot\frac{1}{\theta^2}X_1, \\ \frac{\partial^2 \ln L}{\partial\theta^2} &= \frac{1}{\theta^2} - \frac{1}{\theta^3}X_1, \\ -E\left[\frac{\partial^2 \ln L}{\partial\theta^2}\right] &= -\frac{1}{\theta^2} + \frac{1}{\theta^3}E(X_1) = -\frac{1}{\theta^2} + \frac{2}{\theta^2} = \frac{1}{\theta^2}. \end{aligned}$$

So

$$I_n(\theta) = \frac{n}{\theta^2} \text{ and } e(\overline{\theta}) = 1,$$

which means the estimator is *efficient* and, thus, also a MVUE.

c) The standard error is

$$\begin{aligned}\sigma_{\overline{\theta}} &= \sigma(\overline{\theta}) = \sqrt{V(\overline{\theta})} \\ &= \frac{\theta}{\sqrt{n}} \approx \frac{\overline{\theta}}{\sqrt{n}}.\end{aligned}$$

If the sum of 100 observations is 200, then $n = 100$ and $\overline{X} = \dfrac{200}{100} = 2$.
The estimator is

$$\overline{\theta} = 1$$

and the standard error is

$$\sigma_{\overline{\theta}} \approx \frac{1}{10} = 0.1.$$