Middle East Technical University — Department of Computer Engineering

# CENG 222

## Statistical Methods for Computer Engineering

Spring 2023-2024

## Homework 1

Due date: March 21 2024, Thursday, 23:55

While answering the questions, please show your work and the steps of your calculations. If you think the calculations are not feasible to do with hand, *e.g.* Q1d, you can use a computer. When applicable, please give your answers using **3 digits after the decimal point**. Except for Q2c, you can use `binocdf`, `poisscdf` of Octave/MATLAB when you feel in need of Binomial or Poisson CDFs.

## Question 1 (25 pts)

Let $X$ be a random discrete random variables with outcomes $x \in \{1, 2, 3, 4, 5\}$, and $P$ be the associated probability mass function such that $P(x) = \mathbf{P}\{X = x\} = N/x$.

**a) ( 5 pts)** Determine the constant $N$ so that $\mathbf{P}$ qualifies as a probability mass function. Use the $N$ you found in answering the following questions.

**b) ( 5 pts)** Calculate the expected value of $X$.

**c) ( 5 pts)** Calculate the variance of $X$.

**d) (10 pts)** Let $Y$ be a random variable with $P(y) = y/15$ where $y \in \{1, 2, 3, 4, 5\}$ and let the joint distribution be $P(x, y) = P(x)P(y)$. Calculate the covariance $\mathrm{Cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X)\,\mathbf{E}(Y)$ by first calculating $\mathbf{E}(Y)$ and $\mathbf{E}(XY)$. Interpret the number you found, *i.e.*, what does $\mathrm{Cov}(X, Y)$ say about the relation between $X$ and $Y$?

## Question 2 (50 pts)

**a) (15 pts)** In an anectodal note, it is said that Thomas Edison was successful on his 1000th attempt in inventing the light bulb. Assuming each attempt can only result in failure or success with equal probability, what should be the probability of success for an *individual attempt* so that the probability that at least one attempt is successful in 1000 trials is 95%?

**b) (20 pts)** The Elo rating in chess is used to quantify a player's skill level. The average chess player's Elo rating is 1400, an average international master's (IM) Elo rating is 2400, while for grandmasters (GM), the average Elo is 2650. However, these numbers may not accurately portrait the difference between the skill levels since the scale does not progress linearly. For a better understanding,

i) assume that the probability for the average player to win against an IM is $p = 3 \times 10^{-3}$. What is the likelihood of having to play more than 500 games in order to win twice against an IM for an average player?

ii) assume that the probability for the average player to win against a GM is $p = 10^{-4}$. What is the likelihood of having to play more than 10,000 games in order to win twice against a GM for an average player?

**c) (15 pts)** Assume that as a healthy person, the probability of you not feeling sick is 98% for any given day of the year. Throughout the current year, which has 366 days, what is the probability of you not feeling sick for at least 360 days? Use the most fitting values from **the Appendix of the textbook**.

## Question 3 (25 pts)

In answering this question, use Octave or MATLAB. Please include **both** your code and plots in your reports.

**a) ( 5 pts)** Calculate the more accurate value for **Q2c** via Octave or MATLAB. Is it lower or higher? Why?

**b) (10 pts)** In **Q2c**, let the probability of feeling well be $p = 98\%$, the number of total days ($n$) be a variable $n \in \{50, 51, \cdots, 400\}$, number of maximum sick days ($x$) be $x = 6$. For each $n$, calculate the corresponding probability using Binomial CDF and Poisson CDF. In a **single graph** provide the plots of your answers. The $y$ axis is for the probabilities you calculated, while the $x$ axis will be the number of total days.

**c) (10 pts)** Do the same as in **Q3b** with the only change that $p = 78\%$. Provide a brief comparison of how the Poisson distribution's approximation to the Binomial distribution changed as $p$ varied.

While plotting the distributions in Q3b and Q3c, use the command '`hold on`' after drawing your first plot. For example,

```
>> p = 0.98;
>> ns = 50:400;
>> binomial_probabilities = (...);
>> poisson_probabilities = (...);
>> close all;
>> plot(ns, binomial_probabilities, 'linewidth', 2);
>> hold on;
>> plot(ns, poisson_probabilities, '-.', 'linewidth', 2);
>> saveas(1, "p=0.98.png");
```

## Conceptual Questions (Ungraded)

Make sure that you can confidently answer the following True/False questions.

**a) (T/F)** For all events $A_1$ and $A_2$ that are dependent on the event $B$, their union, $A_3 = A_1 \cup A_2$, is also dependent on $B$.

**b) (T/F)** For random variables $X$ and $Y$, $\mathbf{E}(X)\,\mathbf{E}(Y) = \mathbf{E}(X\,Y)$ implies the independence of $X$ and $Y$.

**c) (T/F)** In general, given the marginal distributions of random variables $X$ and $Y$, their joint distribution cannot be determined.

**d) (T/F)** In general, given the joint distribution of random variables $X$ and $Y$, the marginal distributions for $X$ and $Y$ cannot be determined.

**e) (T/F)** For any discrete random variable $X$, $\mathbf{E}(X)$ yields the most likely outcome.

## Suggested Examples and Exercises (Ungraded)

Here is a suggested (non-exhaustive) list of examples and exercises from the textbook.

- Exercises 1.1, 1.2, and 1.3. These exercises may appear trivial, yet, they offer an opportunity for reflection.
- Example 2.20. A nice systematic setting for exercise. Also see Exercises 2.23 and 2.24.
- Example 2.24. A paradox that might catch you off-guard. Also see Exercise 2.30.
- Exercises 2.31, 2.35, and 2.36. For those who like to prove.

## Regulations

1. Your submission should be a single vector-based PDF document with the name `the1.pdf`. You can use the given template.
2. Please do not skip the calculation steps. Show every step of your work. When applicable, give the name of the distributions you used.
3. **Late Submission:** You have a total of 2 late days for this homework. For each day you have submitted late, you will lose 25 points. If you submit your homework at least 2 days later than the deadline, you will get zero.
4. **Cheating:** Cheating is forbidden. The violators will be punished according to the department regulations.

5. **Updates & Announces:** Follow the course page on ODTUClass for any updates and clarifications. Please ask your questions on ODTUClass instead of e-mailing if they do not contain some part of the solution. If they contain, you can send an email to "ferhata@metu.edu.tr".

6. **Evaluation:** Your `.pdf` file will be checked for plagiarism automatically using "black-box" technique and manually by assistants, so make sure to obey the specifications.

## Submission

Submission will be done via odtuclass. You will submit a single PDF file, `the1.pdf`.