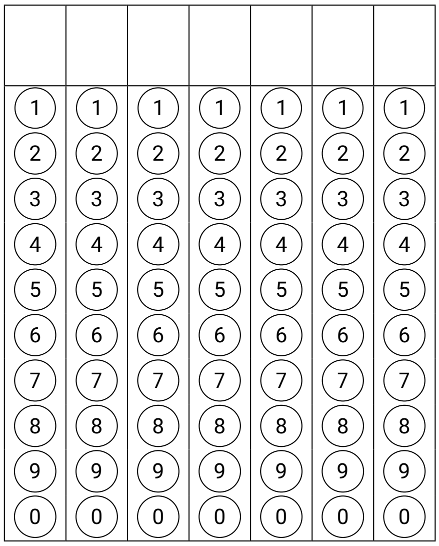
Surname, First name



Modelling, Uncertainty and Data for Engineers EXAM (CEGM1000)
Exam Q1
4 Nov, 2024, 13:30-16:30

Answer multiple-choice questions as shown in the example. Circular checkboxes have only one correct answer. Square checkboxes may have multiple correct answers.

Dear student,

Before you start the exam, a few remarks:

- Write down your first and last name in the field on the top left corner of this page
- Fill in your student number on the top right corner of this pages. Fill in the number in the boxes on top, and mark the corresponding number. Fill the corresponding circle with the number completely.
- The duration of the exam is 3 hours. If you're entitled to time extension, you have 3,5 hours.
- It is not allowed to release the staple.
- (Graphical) calculators are allowed, as long as they do not include document viewing or network connection capabilities.
- Any other tools and sources of information are not allowed.
- This answer sheet contains both the questions as an answer field. It is unique per student, as a result of which you are not able to obtain a new answer sheet. In case you want to erase and rewrite your answer, ask an invigilator for a white sticker to cover your incorrect answer. Furthermore, this answer sheet contains some additional space for your answers. You can use this if you need run out of space at the original answer field. If you use this additonal space, please indicate so at the original answer field.
- For your convenience, a separate formula sheet is provided. These appendices will not be graded.
- Scrap paper is not graded.

Good luck!

Kind regards, the MUDE team

Exercise 1: Programming

Programming part I: Object-oriented programming

Consider the following piece of code and output, and reflect on the object oriented programming (OOP) that is used:

```
[ ]: import numpy as np
     data = np.genfromtxt(…)
     print(data.shape)
     print(data.mean())
```

```
Output:
(2734, )
65.4
```

2p **1a** Which of the following are OOP **attributes**? (You can select more than 1)

- ☐ data
- ☐ np
- ☐ numpy
- ☐ genfromtext
- ☐ print
- ☐ shape
- ☐ mean
- ☐ none of the above

2p **1b** which of the following are OOP **methods**? (You can select more than 1)

- ☐ data
- ☐ np
- ☐ numpy
- ☐ genfromtext
- ☐ print
- ☐ shape
- ☐ mean
- ☐ none of the above

Programming part II: Errors

A friend has sent you a piece of code (not provided) that is supposed to compute the numerical solution to an ordinary differential equation using Taylor series, an iteration scheme and a numpy array. However, instead of getting the expected answer of 5, the code returns 9.

3p **1c** What type of error is this? Add an explanation for your reasoning.

3p **1d** What would be a useful approach to help figure out and solve the problem? Consider each of the following and select **one** that you would find useful. Add an explanation for your selection.

- Read the traceback
- Use list comprehension
- Try pandas instead of numpy
- Use assert statements

Exercise 2: Uncertainty Propagation

Let $X_1, X_2, X_3,$ be independent random variables, each following a standard normal distribution. These random variables are combined into a 3-dimensional random vector $Y = [Y_1, Y_2, Y_3]^T$, defined as:

$$Y = \begin{bmatrix} X_1 \\ X_2 + X_3 \\ X_1 + X_2 + X_3 \end{bmatrix}$$

4p **2a** What is the expectation of Y?

8p **2b** **Apply** the propagation law to show that $cov\,(Y_1, Y_2) = 0$ and **explain** why this is the case?

6p **2c** If the second element of Y is changed to $X_2 - X_3$, which elements in covariance matrix of the changed vector Y, Σ_Y, will be changed? Write out the new Σ_Y.

Exercise 3: Observation theory

An object is moving along a straight line with an unknown speed v. It starts at $t_0 = 0\,[s]$ with the known location $x_0 = 0\,[m]$. We measured its location every Δt seconds and obtained m measurements y_i, where $i = [1, \ldots, m]$, measured at $t_i = i \cdot \Delta t$.

Model 1: $y_i = x_0 + vt_i$

We assume the measurements y_i are independent and follow a normal distribution with standard deviation $\sigma_i = i \cdot \sigma$

6p **3a** Specify the functional model and the covariance matrix of the observable, Σ_Y, according to Model 1.

6p **3b** How many measurements m are required if we want the 99% confidence interval to be no larger than $\hat{v} \pm 0.2[m/s]$? Assume $\sigma = 0.4[m]$ and $\Delta t = 2[s]$

Another engineer assumes the initial position x_0 to be unknown, and that the movement may have an unknown acceleration a:

$$\textbf{Model 2: } y_i = x_0 + vt_i + \frac{1}{2}at_i^2$$

6p **3c** Which test should the engineer apply **and** what is the corresponding critical value used for significance level $\alpha = 0.05$? Answer the questions **without** specifying the functional model of Model 2.

Exercise 4: Numerical Modeling

8p **4a** Use the Newton-Raphson method to solve for x in the following equation:

$$x^5 + 3x = 100$$

Recall that:

$$x_{j+1} = x_j - \frac{g(x_j)}{g'(x_j)}$$

Where $g(x) = 0$ and j is the iteration number.
Write down the equation to be iterated in the Newton-Raphson method.
Assume a first guess of $x = 2$, then compute the second guess (i.e., one iteration).

2p **4b** What is the error order of the Taylor Series Expansion (TSE) of: $x^2 - a$ using the first 3 terms $\left[f(x_i) + \Delta x f'(x_i) + \frac{(\Delta x^2)}{2!} f''(x_i)\right]$, where a is a constant.

- (a) There is no error
- (b) $\mathcal{O}(\Delta x)$.
- (c) $\mathcal{O}(\Delta x^2)$.
- (d) $\mathcal{O}(\Delta x^3)$.
- (e) $\mathcal{O}(\Delta x^4)$.

6p **4c** Use Forward Euler to approximate the solution of the following Initial Value Problem from $t = 0$ until $t = 0.2$:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = y + \sin(t)$$

$$y(t_0 = 0) = 1$$

$$\Delta t = 0.1$$

↳

2p **4d** Is Backward Euler more accurate than Forward Euler?

(a) Yes (b) No

5p **4e** Consider the following convection equation:

$$\frac{dc}{\mathrm{d}t} + v\frac{dc}{dx}$$

Specify how many initial conditions and boundary conditions are required to solve the convection equation. **Then,** write the algebraic expression to approximate the first derivative of this PDE using: Central difference in space **and** central difference in time

6p **4f** Consider a grid with 5 nodes: $x_0, x_{1,}x_{2,}x_3, x_4$The temperature profile on this grid is described by the discretized ODE:

$$\frac{1}{\Delta x^2}\left(T_{i-1} - 2T_i + T_{i+1}\right) - \alpha\left(T_i - T_0\right) = 0$$

where α, Δx and T_s are known constants. The boundary conditions are Dirichlet:

$$T\left(x_0\right) = T_0 \text{ and } T\left(x_4\right) = T_4$$

A fellow student proposes the following system to solve the problem:

$$\begin{bmatrix} -(2+\alpha\Delta x^2) & 1 & 0 \\ 1 & -(2+\alpha\Delta x^2) & 1 \\ 0 & 1 & -(2+\alpha\Delta x^2) \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix} = \begin{bmatrix} \alpha T_s\Delta x^2 \\ \alpha T_s\Delta x^2 \\ \alpha T_s\Delta x^2 \end{bmatrix}$$

Explain what is wrong with this proposed system. How would you correct it?

Exercise 5: Probability and Reliability

The empirical or sample mean based on m outcomes (or: realizations) x_i can be computed as:

$$\hat{\mu}_X = \frac{1}{m}\sum_{i=1}^{m} x_i$$

The sample variance can be computed as:

$$\hat{\sigma}_X^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \hat{\mu}_X)^2$$

Finally, the covariance is given by:

$$Cov(X_1, X_2) = \mathbb{E}\left([X_1 - \mathbb{E}(X_1)]\,[X_2 - \mathbb{E}(X_2)]\right)$$

A group of scientists are analyzing the influence of ice water content in the clouds $\left[IC, kg/m^2\right]$ on rainfall $[RF, mm]$ as their theory is that it will improve the development of weather and climate models.
Here you have a brief definition of the variables:

- Ice water content in the clouds $\left[IC, kg/m^2\right]$. It is the amount of ice that you have in the clouds on the column of air above a given area. Thus, only positive values have a physical meaning.
- Rainfall $[RF, mm]$. Here, it is defined as the amount of rain that was accumulated during the hour when the measurement was conducted in the area used to compute IC. Again, only positive values have a physical meaning.

They have been able to conduct 10 field measurements that are shown in the table below:

$IC\left[\mathrm{kg}/m^2\right]$	$RF\,[\mathrm{mm}]$
999	10
242	8
76	7
666	4
1292	12
927	9
405	9.5
126	7
867	6
491	10

Table 1: Pair observations of ice water content and rainfall.

Disclaimer: this example and data are generated for academic purposes.

2p **5a** What do PDF and CDF stand for? Choose the right combination of words.

- (a) P-value Distribution Function, and Cumulative Distribution Function
- (b) Probability Density Function, and Cumulative Distribution Function
- (c) Probability Distribution Function, and Cumulative Density Function
- (d) Probability Density Function, and Cumulative Density Function

6p **5b** Compute and draw the empirical CDF of the variable IC.

Remember to label the axis and indicate their values. You may want to arrange your calculations in table format.

3p **5c** Is the Uniform distribution a reasonable distribution for the random variable IC? Justify your answer with at least one reason.

2p **5d** Assuming that the scientists want to quantify the joint distribution function of IC and RF using a multivariate Gaussian distribution, what are the minimum parameters that they need to compute from the observations to quantify the distribution?

- (a) The mean, the variance and standard deviation of the each variable.
- (b) The mean of each variable and the correlation coefficient between them.
- (c) The covariance and correlation coefficient between the variables and the standard deviation of each variable.
- (d) The mean and variance of each variable and the covariance between them.

8p **5e** Assuming that the scientists want to quantify the joint distribution function of IC and RF using a multivariate Gaussian distribution.

Compute the covariance matrix of IC and RF. Do not use more than two decimal figures. You may want to make use of a table format to arrange your calculations.

4p **5f** Assume that finally the scientists have changed their mind and do NOT want to quantify the joint distribution function of IC and RF using a multivariate Gaussian distribution. They have derived marginal distributions for IC, RF and the joint distribution of IC and RF as follows:

- The marginal distribution of IC is given by the PDF $f_{IC}(ic)$ and the CDF $F_{IC}(ic)$.
- The marginal distribution of RF is given by the PDF $f_{RF}(rf)$ and the CDF $F_{RF}(rf)$.
- The joint probability distribution of IC and RF is given by the PDF $f_{IC,RF}(ic, rf)$ and the CDF $F_{IC,RF}(ic, rf)$

Using the information described ONLY in this subquestion, derive the expression for the joint exceedance probability $P[IC > ic, RF > rf]$. You may want to draw a diagram to assist you.

Use the grid below if you run out of space for any exercise.
In that case, please indicate so at the original answer field.