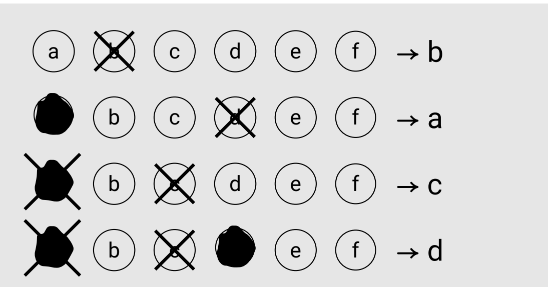
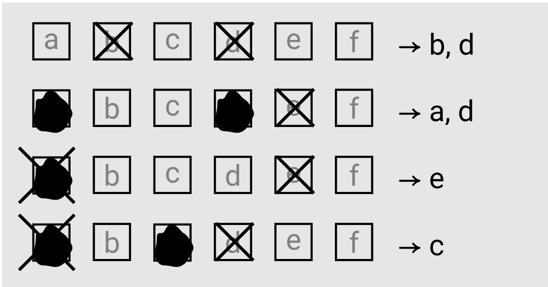
Surname, First name

1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3
4	4	4	4	4	4	4
5	5	5	5	5	5	5
6	6	6	6	6	6	6
7	7	7	7	7	7	7
8	8	8	8	8	8	8
9	9	9	9	9	9	9
0	0	0	0	0	0	0

Modelling, Uncertainty and Data for Engineers EXAM (CEGM1000)
Exam Q1 R





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

===== DO NOT OPEN THIS EXAM OR TURN IT OVER UNTIL INSTRUCTED TO DO SO =====

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.
- Write **student number** on the **top right** corner as the **number and filled in circle**.
- Place your student ID card face up on your desk.
- The duration of the exam is 3 hours (if entitled to an extension place paper on desk, face up).
- You are only allowed a pen/pencil and approved calculator; nothing else is allowed.
- Note in the multiple choice answer examples above that filled in circles/squares take precedence over an "X." **It is your responsibility to make sure that there is no ambiguity in your final answer**s (for example, add an extra note "final answer = x, y" etc). If in doubt, ask an instructor prior to submitting your exam. Examples are provided on the back page of this exam.
- Write all answers in the space provided (scrap paper is not graded), and do not remove the staple (nietje). In case you must erase, ask an invigilator for a white sticker to cover your incorrect answer. If you use any extra space to write an answer, indicate so in the original answer field.
- A summary of all questions is provided on the back page.

Good luck!

Propagation of Uncertainty

A speed detection system is used to monitor the speed of vehicles in a traffic flow. The speed of the vehicle can then be determined by analyzing the Doppler shift in the returned signal, which depends on the propagation speed c of the signal in the air. This propagation speed varies with air temperature T and humidity H as:

$c = 331.3 + 0.6T + 0.0124H^2$

where c is in m/s, T is in °C, and H is in g/m^3.

Both temperature and humidity are random variables due to environmental fluctuations.
We are interested in the precision of the propagation speed, where it is known that:

$\mu T = 25$°C, $\sigma_T = 3$ °C
$\mu_H = 10$ g/m^3, $\sigma_H = 2$ g/m^3.
$Cov\,(T, H) = 0$

8p **1a** Approximate the standard deviation of the propagation speed (give your answer to 2 decimal places). Show how you arrived at your answer.

4p **1b** Would it have more impact to reduce σ_H or σ_T , in order to reduce σ_C,? Explain your answer.

5p **1c** You find that for the humidity sensor, only 5 measurements were taken during the day. In contrast, the temperature sensor measured the temperature every hour.

Please answer the following questions, you only need to write 1-2 sentences for each:
- Which source of uncertainty affects both of them?
- Which source of uncertainty affects the humidity sensor to a larger degree than the temperature sensor?
- Describe how and if these sources of uncertainty can be reduced.

3p **1d** Which of the following is **not true** about propagation of uncertainty?

- (a) Independent variables always lead to additive variances.
- (b) Covariance terms are zero when variables are uncorrelated.
- (c) The propagation of uncertainty is only applicable to linear functions.

Observation Theory

A scientist is studying the decay of a radioactive isotope in groundwater to estimate its half-life. The concentration of the isotope $C(t)$ at time t follows the exponential decay law:

$$C(t) = C_0 e^{-\lambda t}$$

where:

- C_0 is the initial concentration of the isotope at $t = 0$,
- λ is the decay constant,
- t is the time since the sample was collected.
- The scientist measures the concentrations C_i at 6 different times, creating 6 observations with standard deviation σ_c. To simplify the analysis, the geoscientist linearizes the equation by taking the natural logarithm:

$$\ln(C_i) = \ln(C_0) - \lambda t_i.$$

Where the following is defined:

$$Y_i = \ln(C_i), \quad a = \ln(C_0), \quad b = -\lambda.$$

6p **2a** Give the functional model in the form $\mathbb{E}(\mathbf{Y}) = A\mathbf{x}$ and specifying the following:

- Provide the sizes and shapes of the A matrix and vectors $\mathbf{Y}$ and $\mathbf{x}$.
- The redundancy of the system.

4p **2b** What happens to a model and errors when the redundancy $q = 0$?
Multiple answers possible.

- [ ] The model has extra flexibility, which allows for better error estimation
- [ ] The system cannot be solved, because there is no solution
- [ ] The model becomes statistically reliable for prediction, eliminating the impact of noise
- [ ] The model perfectly fits the data, and the residuals become 0
- [ ] The model uses all available data, but cannot validate the parameters due to a lack of redundancy.

4p **2c** Assume that measurement 3 contains a bias of size ∇, making it an outlier. Formulate the alternative hypothesis by modifying the functional model to account for this bias.

4p **2d** For the propagation of uncertainty, we define the non-linear variance propagation law for $X = q(Y)$ as:

$$\sigma_X^2 = \left(\frac{\partial q}{\partial Y}\right)_0^2 \sigma_Y^2$$

Considering the transformed observations:

$$Y_i = \ln(C_i)$$

Derive the stochastic model.

6p **2e** To decide between the null hypothesis H_0 and the alternative hypothesis H_a, the scientist conducts a hypothesis test. The test statistic T_q for the hypothesis test has already been computed as:

$T_q = 6.81$

Answer the following questions:

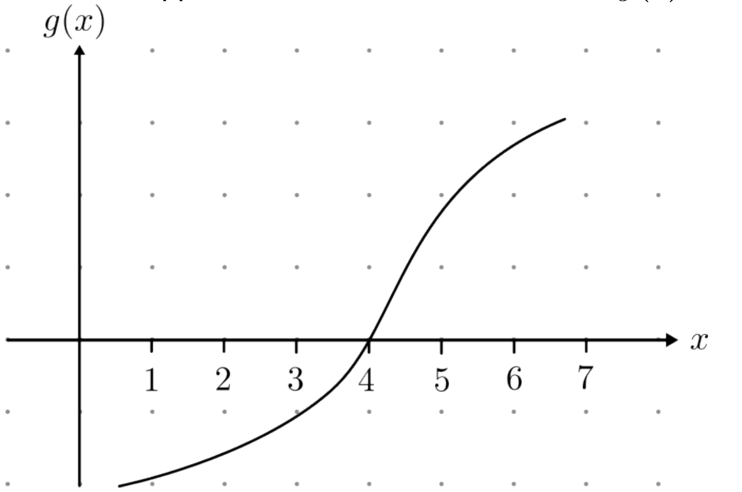
- What type of hypothesis test should the scientist use in this scenario?
- Using a false alarm probability of 5%, compute and demonstrate if the Null hypothesis is accepted or rejected

3p **2f** What does **Best Linear Unbiased Estimator (BLUE)** aim to minimize in a linear regression model?

- (a) The sum of squared residuals.
- (b) The variance of the estimator.
- (c) The observation noise.
- (d) The sum of absolute residuals.

Numerical Modelling

Below you can see a graphical representation for a given function $g(x)$. We use the Newton-Raphson method to approximate the root of the function $g(x) = 0$



6p **3a** Assuming an initial estimation of $x = 5$, indicate the second estimation of the value of x. Provide a sketch in the answer box below (and the next page) to demonstrate your estimated value of x.

Note: you don't need to make explicit calculations, as you can solve this graphically. Your sketch should clearly indicate how you arrived at your estimated value of x. Markings on the figure above will not be graded.

8p **3b** Derive the Backward Difference expression **second order accurate** of the **first derivative** around the point x_i. Assume that the distance between points is equidistant.

3p **3c** Select the correct classification of the following equation:

$$\frac{d^3y}{dx^3} - x\frac{d^2y}{dx^2} + y = 0$$

- (a) First-order and linear ODE
- (b) Third-order and linear ODE
- (c) First-order and non-linear ODE
- (d) Third-order and non-linear ODE
- (e) First-order linear PDE
- (f) Third- order linear PDE
- (g) First- order and non-linear PDE
- (h) Third- order and non-linear PDE

3p **3d** Multi stage methods ... (select all that correctly complete the sentence):

- [ ] require initialisation from another method
- [ ] do not require initialisation from another method
- [ ] allow choosing large steps
- [ ] can only be used with small steps

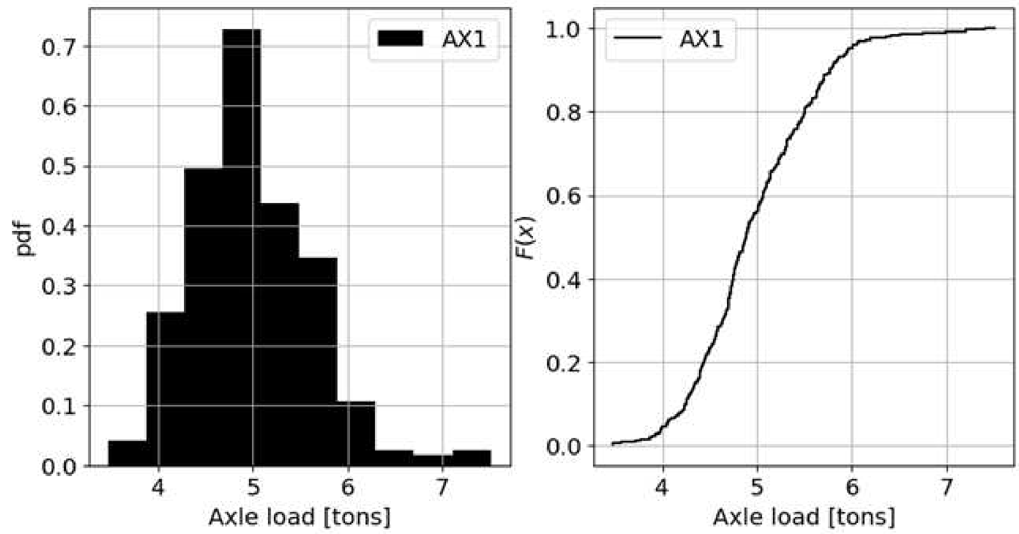
3p **3e** This type of condition defines the value of the function (to be solved) at the boundary:

- (a) Neumann
- (b) Dirichlet
- (c) Robin
- (d) Mixed

Probability
You are the designer of a bridge and want to study the traffic loads that it will face. You have measurements of axle loads in a nearby road that will connect to the bridge in the future.

Axle loads are defined as the force transmitted by the wheels from the first axle to the ground in tons. You are focusing on trucks with 4 axles as they are the heavier vehicles in that road and, thus, will transmit higher axle loads to the bridge. Then, you have four axle loads per vehicle, denoted here as AX1, AX2, AX3 and AX4.

You start studying the first axle load, AX1. You plot the empirical CDF and PDF of the data, as shown in the image below.



3p **4a** What would be the most appropriate parametric distribution for AX1?

- (a) Gaussian
- (b) Exponential
- (c) Uniform

3p **4b** Justify briefly your answer with **at least** one reason.

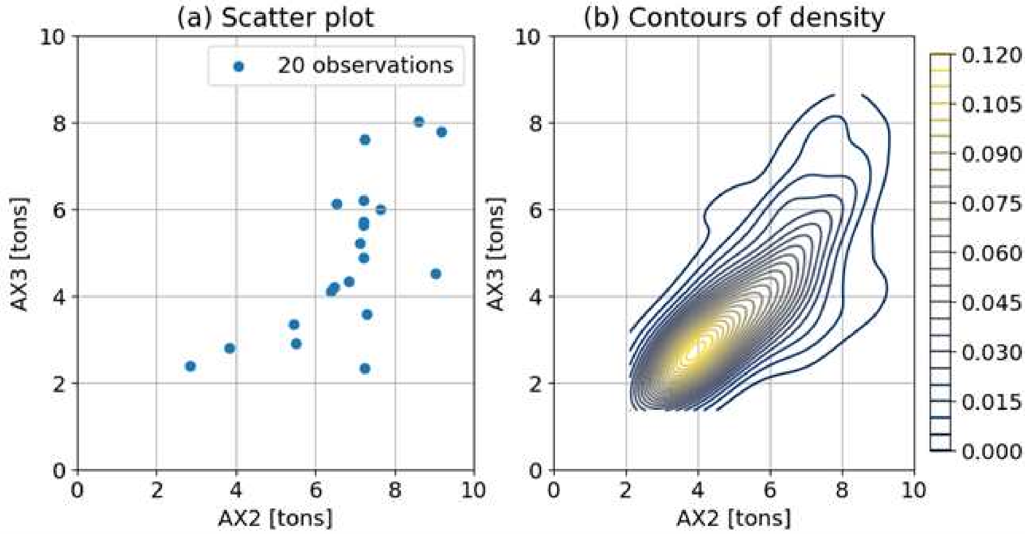
3p **4c** Using the plot above, compute approximately $P(AX1 > 5)$.

6p **4d** Another colleague is studying AX4. They have decided to fit a Gumbel distribution to the data of AX4. To assess the goodness of fit of the Gumbel distribution, they are using the Kolmogorov-Smirnov test.

Remember that the null hypothesis of the test is that the observations are coming from the fitted distribution and, thus, the tested distribution is a reasonable model for them. They have obtained a **statistic = 0.06** and a **p-value = 0.23** and asked for your help to interpret it considering a **significance level of 0.05**.

Is the Gumbel distribution a reasonable distribution for AX4 based on Kolmogorov-Smirnov test?
State yes or no, then briefly explain your reasoning using the quantitative results provided above.

The multivariate probability team is working on the relationship between AX2 and AX3 and start by plotting some of the observations and the joint PDF as shown in the figure below.



4p **4e** Using the plots above, compute $P(AX2 \leq 6 | AX3 \leq 4)$

4p **4f** Using the plots above, compute $P(AX2 > 8 \cap X3 > 6)$

The multivariate probability team has decided to use a multivariate Gaussian distribution to model the relationship between AX1, AX2, AX3 and AX4. They have obtained the following covariance matrix.

$$\Sigma = \begin{pmatrix} 0.40 & 0.76 & 0.37 & 0.35 \\ 0.76 & 3.60 & 2.30 & 2.01 \\ 0.37 & 2.30 & 2.4 & 2.21 \\ 0.35 & 2.01 & 2.21 & 2.30 \end{pmatrix}$$

3p **4g** What is the variable with the highest variability?

(a) AX1 (b) AX2 (c) AX3 (d) AX4

4p **4h** What is the pair of variables with the highest correlation?

(a) AX1-AX2

(b) AX2-AX3

(c) AX1-AX4

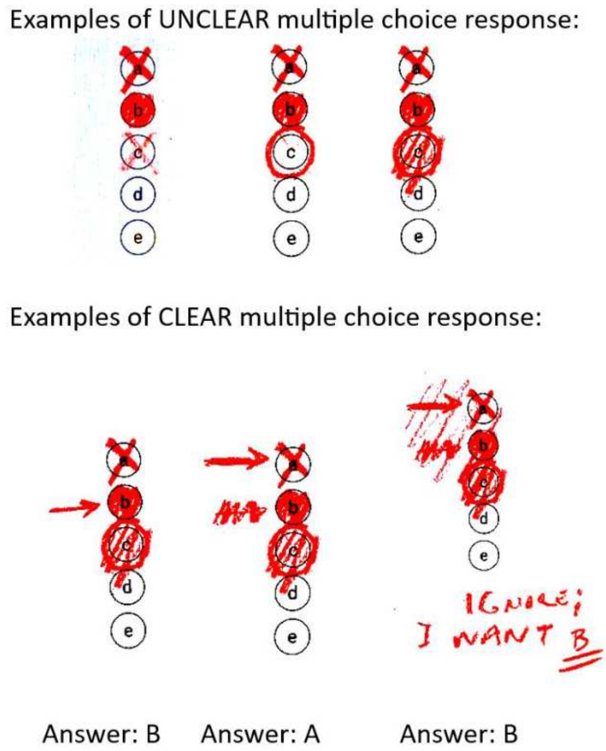
(d) AX3-AX4

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

The table below gives an overview of the questions to help you plan your time during the exam.

No.	Topic	Number of sub-parts	Points
1	Propagation of Uncertainty	4	20
2	Observation Theory	6	27
3	Numerical Modelling	5	23
4	Probability	8	30

In case you want to correct your answer for a multiple choice question put an ARROW in front of your final answer. If you also make a mistake with your arrow, write a clear message on the page. Here are a few examples:



Multiple choice examples.