Task 1: Converting a linear programming problem to mathematics

The Dyson company is starting to build electric cars in three factories in Singapore.

The company is planning on building two models of electric car: a standard model and a sports utility vehicle (SUV). Factory A produces 50 standard and 37 SUV electric cars per day. Factory B produces 63 SUV electric cars, but no standard cars. Factory C produces 51 standard electric cars, but no electric SUVs. The cost of operating factories A,B and C per day are: £210,000 , £182,000 , and £170,000 respectively.

Dyson plans on producing 1500 electric SUVs and 1050 standard electric cars in September. Design a linear program to find out, how many days each factory should operate in September, for Dyson to produce the required number of cars, and to minimize the cost. You do not need to worry about weekends or public holidays. You do not need to solve the resulting problem.

Task 2: Plotting linear constraints using Matlab

Use the graphical method to solve the following linear programming problems with two decision variables.

Develop a Matlab script to plot the following constraints in the $x - y$ plane.

$$\begin{aligned} 6x + 3y &\leq 36 \\ 2x + 3y &\leq 24 \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

The linear programming problem is to find the maximum of the objective function:

$$f(x, y) = 4x + 3y$$

Task 3: solution of linear programming problem

A linear programming problem is defined by the inequalities below.

$$\begin{aligned} x_1 + x_2 &\geq 1 \\ x_2 - 5x_1 &\leq 0 \\ x_1 - x_2 &\geq -1 \\ x_1 &\geq 0 \\ x_2 &\geq 0 \end{aligned}$$

The objective function is

$$f(x_1, x_2) = 10x_1 + 2x_2$$

and the objective is to mimimise f. Use the method described in section 2.3 of the course notes on linear programming to fill in the table below. The variables x_3, x_4 and x_5 are the slack variables. Hence find the minimum solution.

x_1	x_2	x_3	x_4	x_5	feasible?	f
0	0	*	*	*	y/n	*
0	*	0	*	*	y/n	*
0	*	*	0	*	y/n	*
0	*	*	*	0	y/n	*
*	0	0	*	*	y/n	*
*	0	*	0	*	y/n	*
*	0	*	*	0	y/n	*
*	*	0	0	*	y/n	*
*	*	0	*	0	y/n	*
*	*	*	0	0	y/n	*

Task 4: using Matlab to solve a linear programming problem

Use Matlab to solve the following linear programming problem.

$$
\begin{aligned}
x_1 + x_2 + 4x_3 &\le 20 \\
2x_2 + 5x_3 &\le 40 \\
2x_1 + 6x_3 + 4x_4 &\le 50 \\
x_2 + x_3 + 3x_4 &\le 30 \\
x_1 &\ge 0 \\
x_2 &\ge 0 \\
x_3 &\ge 0 \\
x_4 &\ge 0
\end{aligned}
$$

The objective function is

$$f(x_1, x_2, x_3, x_4) = 3x_1 + 4x_2 + x_3 + 5x_4$$

Find the maximum value of f.

Task 5: algorithms for linear programming

In the module the simplex algorithm has been introduced to solve linear programming problems. Another algorithm for solving linear programming problems is called the interior-point method. Briefly describe the key difference between the interior point algorithm and the simplex algorithm, and the computational advantages of the interior point algorithm over the simplex algorithm.

Modify the code below (from the lectures) to use the interior point method.

```
A = [1 1
    1 1/4
    1 -1
    -1/4 -1
    -1 -1
    -1 1];

b = [2 1 2 1 -1 2];

f = [-1 -1/3];
x = linprog(f,A,b)
```

Task 6:

The Erlang probability density function (PDF), which measures the time between incoming calls, can be used to model the performance of a telphone call centres. The age distribution of cancer incidence has been found to follow the Erlang PDF.

The Erlang PDF is

$$p(x \mid k, \lambda) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}$$

where k is an integer and λ is a positive real number. The parameter k is called the shape parameter and λ is the rate parameter. See `http://mathworld.wolfram.com/ErlangDistribution.html` for background information. The Erlang distribution is a special case of the gamma distribution (`http://mathworld.wolfram.com/GammaDistribution.html`.)

As part of a NHS project to develop a Monte Carlo simulation of the treatment requirements for patients with cancer, you are asked to generate random numbers from the Erlang PDF. See `https://www.nature.com/articles/s41598-017-12448-7` for some background.

Use Matlab to generate 10000 random numbers from the Erlang probability density function with $\lambda = 0.5$ and $k = 3$. Plot a histogram of the random numbers and include the probability density function in the same plot. Compare the mean of the random numbers to the the theoretical expression of k/λ.

There is no function in Matlab to generate random numbers from the Erlang distribution, however there is a function to generate random numbers from the gamma PDF **gamrnd(A,B)**, with the identification of parameters of the Erlang distribution $\lambda = \frac{1}{B}$ and $k = A$.

You may find the code below, which plots the normal probability density function for the normal PDF, with mean 5 and $\sigma = 2$, as a good starting point. The example also generates 5000 random numbers generated from the normal PDF with mean 5 and $\sigma = 2$ (hint: you need to add the code for the histogram.)

```
x = 2*randn(5000,1) + 5;

hold on
y = -5:0.1:15;
mu = 5;
sigma = 2;
f = exp(-(y-mu).^2./(2*sigma^2))./(sigma*sqrt(2*pi));
plot(y,f,'LineWidth',1.5)
```

Task 7:

Consider the the 4-dimensional integral:

$$I_4(u_1, a_1, a_2, a_3, a_4) = \int_0^1 \int_0^1 \int_0^1 \int_0^1 \cos\left(2\pi u_1 + \sum_{i=1}^{4} a_i x_i\right) dx_1 dx_2 dx_3 dx_4$$

The integrand depends on the parameters u_1, a_1, a_2, a_3, a_4.

This integral can be computed (after some work) analytically to obtain:

$$I_4^{exact}(u_1, a_1, a_2, a_3, a_4) = \cos\left(2\pi u_1 + \frac{1}{2}\sum_{i=1}^{4} a_i\right) \prod_{k=1}^{4} \frac{2}{a_k} \sin\left(\frac{a_k}{2}\right)$$

Use the basic Monte Carlo integration method, with 10^6 Monte Carlo samples, to compute the 4 dimensional integral $I_4(1, 1, 1, 1, 1)$ including an error estimate. Compare the result to that computed using I_4^{exact} above. Repeat the calculation with 4×10^6 Monte Carlo samples. Does the error change as the theory predicted?

Use the basic Monte Carlo integration method, with 10^6 Monte Carlo samples, to compute the 4 dimensional integral $I_4(1, 0.1, 0.1, 0.1, 0.1)$ including an error estimate. By considering the integrand can you qualitatively explain the difference in the Monte Carlo error between $I_4(1, 0.1, 0.1, 0.1, 0.1)$ and $I_4(1, 1, 1, 1, 1)$.

Task 8: Monte Carlo integration with stratified sampling

The Monte Carlo method is a great method for computing high dimensional integrals, unfortunately because the error depends on the number of samples N_{sample} like

$$\text{error} \sim \frac{A}{\sqrt{N_{sample}}}$$

it requires a huge number of samples (and thus computing time) to get a small error.

One technique to reduce the number of samples required for a given accuracy is called stratified sampling. Although, the Monte Carlo method of integration is most useful for high dimensional integrals, here we will test stratified sampling on a 1 dimensional integral to understand the basic ideas.

Use Matlab's **integral** function to compute the integral:

$$I = \int_0^{\pi/2} f(x)dx$$

with

$$f(x) = \left(\frac{x}{\sin x}\right)^4$$

Hint: $f(0.5) = 1.18303$.

With some work the integral can be computed

$$I = -\frac{\pi^3}{12} + 2\pi\log(2) + \frac{\pi^3\log(2)}{3} - \frac{3}{2}\pi\zeta(3)$$

where the constant is Riemann zeta function with argument 3 (`https://en.wikipedia.org/wiki/Riemann_zeta_function`)

$$\zeta(3) = 1.202056903159594285399738161511449990764986292$$

which is sometimes known as Apery's constant (`https://en.wikipedia.org/wiki/Ap%C3%A9ry%27s_constant`). Compute the exact value using Matlab.

Use the standard Monte Carlo method to estimate the integral with an error using 10000 samples. By breaking the integral into two equal intervals, use stratified sampling to estimate the integral with an error. Is the error reduced? Repeat the analysis with 4 equal spaced intervals. Hint, the stratified sampling should use the same number of function evaluations, as used in the standard Monte Carlo estimate of the integral.

Task 9:

A leisure company is planning on developing a downtown area of Plymouth. The company is interested in knowing how many pubs and cafes, should be built in the location. As part of their planning, they want to know the popularity of the different choices of drinking establishments.

The data set `https://archive.ics.uci.edu/ml/datasets/Tarvel+Review+Ratings` contains user ratings for various types of leisure activity in Europe. Users rate each activity between 1 and 5. These numbers are averaged over different locations in Europe.

Download the dataset and load it into Matlab using the data import feature (or another method).

Restrict your analysis to following two rows:

- Attribute 11 : Average ratings on pubs/bars
- Attribute 22 : Average ratings on cafes

The histograms of the above two columns suggest that the mean many not be a good measure of the central value, so it is suggested to look at the median instead. Compute the median of above two columns.

To do a comparison of the two medians require the confidence interval to be computed. For this analysis, I want you to use the 95% confidence interval computed using the bootstrap method. There are some resources below for the bootstrap method.

Write a few sentences to explain how your analysis will help the company decide on the number of bars and cafes to build.

Task 10:

This task is to simulate a basic betting game, as might be done in an online casino. For example, you could be a consultant working for an online betting company. Your job is to study the betting systems used by customer to ensure that your company makes a good profit, while the customers still have fun.

The game is between the player A and the bank B. One game consists of drawing a uniform random number between 0 and 1. If the random number is less than 0.49, then A wins and B pays A £X (we start with X=50). If the random number is greater than or equal to 0.49, then B wins and A pays B £X. The game continues until either A or B loses all their money. This is a standard game. See the Matlab script **Q10_hint.m** on the DLE, which simulates this game, when A starts with £2000 and B starts with £2000. The bet for each random number is £50. Modify the code so that the game is repeated 1000 times and compute the probability of A winning the game. Compute the mean length of the game with the error.

The management of the Bank wants to win more money from their customers. One way to improve the Bank's betting strategy is use the Kelly criterion (`https://en.wikipedia.org/wiki/Kelly_criterion`) to set the amount of money bet for each random number, rather than to bet a fixed amount through out the game.

The fraction of their current money to bet is

$$f^{\star} = \frac{p(b+1)-1}{b}$$

where b are the net odds received on the wager ("b to 1"); that is, you could win £b (on top of getting back your £1 wagered) for a £1 bet, and p is the probability of winning. In a round of the game, if the bank had £100, then it would bet £$f^{\star}100$. In this task we will use $b = 1$ and p = 0.51 for the probability of the bank winning. Modify your Monte Carlo code to use the Kelly criterion for the money bet by the bank. Would you recommend that the bank uses the Kelly criterion? Is the Kelly criterion useful for the player A?

Task 11: The simplex method

Solve the following linear programming problem analytically using the simplex method described in section 2.4 of the course notes, and the podcast.

The problem has 3 decision variables x_1, x_2, and x_3, all of which are positive. The objective function is

$$f = 2x_1 + x_2 + x_3$$

and the objective is to maximize the objective function. The constraints are:

$$\begin{aligned} 3x_1 + 2x_2 - x_3 &\leq 6 \\ x_1 + x_2 + x_3 &\leq 4 \\ 2x_1 - x_2 + 2x_3 &\leq 4 \end{aligned}$$

Introduce slack variables and then solve the problem by hand using the simplex method

Note that you can use Matlab to check your workings. (The Matlab code does not need to be submitted.)

Task 12: integer programming

The head of a school has decided to use online advertising to attract more students into studying for A-levels in English and Mathematics at her school. She wants to advertise her school on YouTube (costing 10p per click) or in the online Guardian (costing 20p per click). The budget for online advertising is £1000. It is only possible to have 32 students in each of the English and Mathematics classes.

From past experience 0.2 % and 0.1 % of the people who click on an advert on YouTube or the Guardian respectively, become students of english at the school. Also, from past experience 0.1 % and 0.3 % of the people who click on an advert on YouTube or the Guardian respectively, become students of mathematics at the school.

The decision variables x_1 and x_2 are for the number people who click on an advert on YouTube or in Guardian newspaper respectively. The variables x_3 and x_4 are the number of students in English and Mathematics, who sign up from the online campaign. All the variables x_1, x_2, x_3 and x_4 are integers and positive (or zero). The linear program for above situation is below.

The objective is to maximize the number of english and mathematics students. The objective function is:

$$f = x_3 + x_4$$

The constraints are a mixture of inequlities and equalities.

$$x_3 \leq 32$$
$$x_4 \leq 32$$
$$0.1x_1 + 0.2x_2 \leq 1000$$

$$x_3 = 0.002x_1 + 0.001x_2$$

$$x_4 = 0.001x_1 + 0.003x_2$$

Use the Matlab function **intlinprog** to solve the above integer linear programming problem. As this problem involves equalities, you will need to use the variables **Aeq** and **beq**.

What are the number of students in English and Mathematics from the solution of the linear program? Use a modification of the linear programming model to suggest a way to increase the number of students.

Task 13: Binary programming

A manager works for a company that does consultancy using mathematics analysis for external companies. For a new project, she wants to assign one of the members of her group to a single one of the 4 parts of the project.

Part 1 Define the problem by talking to the clients.

Part 2 Mathematical analysis of the problem.

Part 3 Write analysis code in Matlab and produce the results

Part 4 Write the report for the client.

Different members of her group have different levels of experience, in for example writing documents or writing code. Given the group members experience, the table below shows the amount of hours which group member will take to complete each part of the project. The manager wants to use this data to estimate which member should undertake each task, in order to minimize the total hours required to complete the project.

See `https://www.techrepublic.com/article/use-this-process-to-estimate-effort-hours/` for background.

Member	Part			
	1	2	3	4
(Sarah) 1	145	122	130	95
(Gavin) 2	80	63	85	48
(Bob) 3	121	107	93	69
(Ruth) 4	118	83	116	80

The decision variables can be written as x_{ij} with both the i and j indices running from 1 to 4, are binary (they either be 0 or 1.) The i index is the member and the j index is part of the project.

$$\sum_{i=1}^{4} x_{ij} = 1 \quad \text{for j = 1..4}$$

$$\sum_{j=1}^{4} x_{ij} = 1 \quad \text{for i = 1..4}$$

Write down the objective and objective function for this problem.

Use the Matlab function **intlinprog** to solve this problem. As this problem involves equalities, you will need to use the variables **Aeq** and **beq**.

Task 14: Linear programming with Maple

The Maple computer algebra also has routines to solve linear programming problems. Use the **LPSolve** Maple function to solve the following problem. You can modify the example in the help section.

This problem has 2 decision variables x, and y. all of which are positive **integers** The objective function is

$$f = 14x + 18y$$

and the objective is to maximize the objective function.

$$\begin{aligned} -x + 3y &\leq 6 \\ 7x + y \leq 35 \end{aligned}$$

Task 15: Linear programming and databases

A manufacturing company (called the funkykeyboard company) has a factory in China which can produce 1200 keyboards per week, and a factory in Korea which can produce 1000 keyboards a week. They supply their keyboards to the three companies, which assemble computers, which are then sold to the public. The three companies use "just in time manufacturing" , so that they (`https://en.wikipedia.org/wiki/Just-in-time_manufacturing`) only order parts for computers, when they have orders.

For example, one week the funkykeyboard company has orders for keyboards. Company 1 wants 900 keyboards, company 2 wants 700 and company 3 wants 400.

Assuming that the cost of making the keyboards are the same at both factories, then a critical issue is the cost of shipping the keyboards to the three companies (1 to 3). The company wants to fulfill the orders to the three companies, but with minimum shipping cost. The unit shipping costs in £ are in the table below:

Factory	Company 1	Company 2	Company 3
China	14	13	11
Korea	12	13	13

The objective function is

$$f = 14x_1 + 13x_2 + 11x_3 + 12y_1 + 13y_2 + 13y_3$$

and the objective is to minimize f. What do the the decision variables (x_i, y_i, i=1..3) represent?

The constraints are:

$$
\begin{aligned}
x_1 + x_2 + x_3 &\leq 1200 \\
y_1 + y_2 + y_3 &\leq 1000 \\
x_1 + y_1 &= 900 \\
x_2 + y_2 &= 700 \\
x_3 + y_3 &= 400
\end{aligned}
$$

The above integer programming model is coded up in the Matlab script **Q15_hint.m** on the DLE. However, the company wants to make itself more efficient, so instead of entering the orders by hand into the Matlab script, it wants to load the numbers from a database using Structured Query Language (SQL).

Modify the script **Q15_hint.m** so that the orders on April 1st 2019 are loaded in from the database file **keyboard_orders.db** using SQL and used in the linear programming problem. The name of the table is **keyboard_record**. Report the cost of shipping.