SDSE Homework 2

Due date: Tuesday, March 4th, 2025 (by midnight)

Note: Wherever you use a computer please include the code.

Problem 1 `[ 4 + 4 + 2 points ]`

Consider the optimization problem:

$$\underset{(x_1,x_2)\in\mathbb{R}^D}{\text{minimize}} \quad \frac{x_1^3}{3} - 4x_1 + \frac{x_2^3}{3} - 16x_2$$

(a) Find the stationary points.

(b) Plot the objective function using one of Matplotlib's surface plotters (plot_surface or plot_wireframe)[1]. Indicate the local minima, local maxima, and saddle points.

(c) Does this problem have a solution? If so, what is it? If not, why not?

Problem 2 `[ 2 + 2 + 4 + 4 + 4 points ]`

A city's environmental monitoring agency tracks CO2 levels in the air. It collects the following samples (values in ppm and assumed to be iid):

$$\mathcal{D} = \{410.3, 415.6, 409.8, 412.5, 416.2, 413.9, 408.7, 414.8\}$$

(a) Calculate the sample mean.

(b) Calculate the unbiased sample variance.

For the remainder, assume that the true standard deviation of CO2 levels is known to be 2.5 ppm.

(c) Assume the measurements are Gaussian. Find the probability that the sample mean falls within 1 ppm of the true mean.

(d) Do *not* assume in this part that the measurements are Gaussian. Find the minimum number of samples needed to ensure that the probability of the sample mean falling farther than 1 ppm from the true mean does not exceed 0.05 (5% chance).

[1] See Module 5, Part 4 of "Python for Engineers" (linked on the course homepage) for a tutorial on 3D plotting in Python.

(e) Repeat part (d) under the assumption that the measurements are Gaussian. (Hint: Use `scipy.stats.ppf`, or the inverse normal cdf table).

Problem 3 `[ 5 points ]`

The following pdf is called a *beta distribution with a single parameter* θ. It is defined on the sample space $x \in [0, 1]$.

$$p(x; \theta) = (\theta + 1)x^{\theta}$$

Suppose that the following five measurements were taken from a system: [0.35, 0.85, 0.59, 0.64, 0.57]. Assuming the measurements follow a beta distribution, find the maximum likelihood estimate of θ. Hint: Use the log likelihood.

Problem 4 `[ 2 + 4 + 3 points ]`

$\hat{\Theta}_1$ and $\hat{\Theta}_2$ are two estimators for a parameter θ. Both are unbiased, and their variances are $Var[\hat{\Theta}_1] = 3$ and $Var[\hat{\Theta}_2] = 5$.

It is suggested that a better estimator might be found by taking a linear combination of the two:

$$\hat{\Theta}_3 = \alpha\hat{\Theta}_1 + (1 - \alpha)\hat{\Theta}_2$$

where $\alpha \in (0, 1)$.

1. Find the bias of $\hat{\Theta}_3$.
2. Find the value of α that minimizes the variance of $\hat{\Theta}_3$.
3. What is the MSE of $\hat{\Theta}_3$ corresponding to the optimal α? Is it better than $\hat{\Theta}_1$ and $\hat{\Theta}_2$?

Problem 5 `[ 4 points ]`

The duration of a single wash at the car wash is 9.5 minutes, with a standard deviation of 1.5 minutes. What is the probability that 20 vehicles can be washed in less than 3 hours? Assume car washes are Gaussian.