## MM20B007 - MLE Tutorial Homework

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy.optimize import minimize
```

## Getting the data

```
df =  pd.read_csv('/content/drive/MyDrive/SEM 7/ID5055/MLE_Tutorial/student-por_2.csv', sep = ';')
df
```

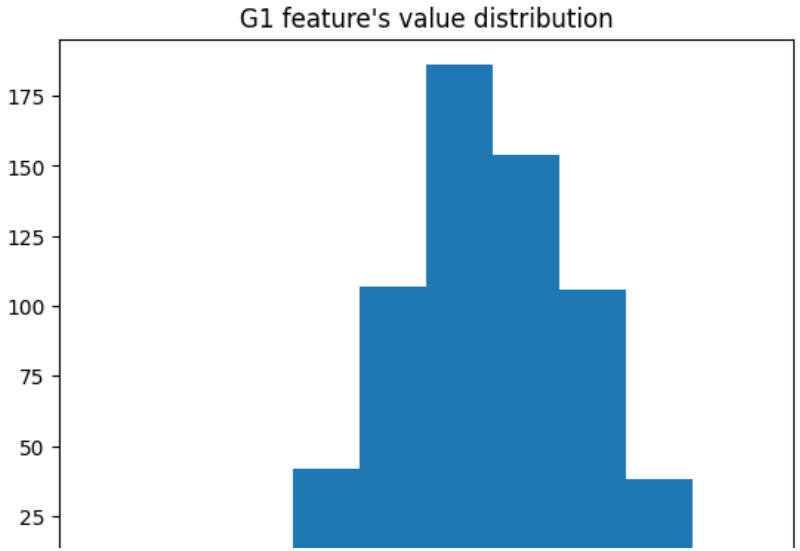
| | school | sex | age | address | famsize | Pstatus | Medu | Fedu | Mjob | Fjob | ... | famrel | freetime | goout | Dalc | Walc | health |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | GP | F | 18 | U | GT3 | A | 4 | 4 | at_home | teacher | ... | 4 | 3 | 4 | 1 | 1 | 3 |
| **1** | GP | F | 17 | U | GT3 | T | 1 | 1 | at_home | other | ... | 5 | 3 | 3 | 1 | 1 | 3 |
| **2** | GP | F | 15 | U | LE3 | T | 1 | 1 | at_home | other | ... | 4 | 3 | 2 | 2 | 3 | 3 |
| **3** | GP | F | 15 | U | GT3 | T | 4 | 2 | health | services | ... | 3 | 2 | 2 | 1 | 1 | 5 |
| **4** | GP | F | 16 | U | GT3 | T | 3 | 3 | other | other | ... | 4 | 3 | 2 | 1 | 2 | 5 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **644** | MS | F | 19 | R | GT3 | T | 2 | 3 | services | other | ... | 5 | 4 | 2 | 1 | 2 | 5 |
| **645** | MS | F | 18 | U | LE3 | T | 3 | 1 | teacher | services | ... | 4 | 3 | 4 | 1 | 1 | 1 |
| **646** | MS | F | 18 | U | GT3 | T | 1 | 1 | other | other | ... | 1 | 1 | 1 | 1 | 1 | 5 |
| **647** | MS | M | 17 | U | LE3 | T | 3 | 1 | services | services | ... | 2 | 4 | 5 | 3 | 4 | 2 |
| **648** | MS | M | 18 | R | LE3 | T | 3 | 2 | services | other | ... | 4 | 4 | 1 | 3 | 4 | 5 |

649 rows × 33 columns

## (a) Plot the histogram and guess the distribution the given feature follows

```
plt.title('G1 feature\'s value distribution')
plt.hist(df['G1'])
```

```
(array([  1.,   0.,   7.,  42., 107., 186., 154., 106.,  38.,   8.]),
 array([ 0. ,  1.9,  3.8,  5.7,  7.6,  9.5, 11.4, 13.3, 15.2, 17.1, 19. ]),
 <BarContainer object of 10 artists>)
```




The above histogram has the representation of a Gaussian Distribution. The parameter $\mu$ is the mean or expectation of the distribution (and also its median and mode), while the parameter $\sigma$ is its standard deviation. The variance of the distribution is $\sigma^2$. The probability density function of the Gaussian distribution is given by:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

## (b) Estimate the parameters of the distribution by maximizing the log-likelihood

```
# Log-likelihood function for Gaussian distribution
def log_likelihood(params, data):
    mu, sigma = params
    log_likelihood = -0.5 * np.sum(np.log(2 * np.pi * sigma**2) + ((data - mu)**2) / (2 * sigma**2))
    return -log_likelihood  # We'll minimize the negative log-likelihood

# Initial parameter guess
initial_params = [0.0, 1.0]

# Maximize log-likelihood to estimate parameters
result = minimize(log_likelihood, initial_params, args= (df['G1'],) , method='Powell')

# Estimated parameters
estimated_mu, estimated_sigma = result.x
```

```
print("Estimated Parameters:")
print("Estimated mean:", estimated_mu)
print("Estimated sigma:", estimated_sigma)
```

```
Estimated Parameters:
Estimated mean: 11.399112386778784
Estimated sigma: 1.9395871813611976
```

The estimated Mean ($\mu$) is 11.399112386778784 and Standard Deviation ($\sigma$) is 1.9395871813611976.

The Variance ($\sigma^2$) is 3.761998434100675.