INTRODUCTION :

The Minimum Mean Square Error (MMSE) is a widely used technique in estimation theory, aiming to minimize the average squared errors between actual and predicted values. This report focuses on estimating the growth of bacterial populations using the MMSE estimator and fitting a linear model to a set of simulated data points. The dataset represents bacterial growth over time, modeled with random noise to reflect real-world fluctuations. The objective is to use the MMSE approach to predict the bacterial population over time and evaluate the accuracy of the model through the MMSE value.

DATA SIMULATION AND OVERVIEW :

A small dataset containing 10 points was simulated to represent bacterial growth over a span of 10 days. The bacterial population was modeled using a linear growth pattern, with added Gaussian noise to simulate natural variability.

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

# Simulating 10 data points related to bacteria growth over time

np.random.seed(42)

time = np.arange(1, 11) # Independent variable: time in days

true\_growth\_rate = 2.5 # True growth rate of the bacteria

initial\_population = 10 # Initial population of bacteria

# Simulating bacterial population using a linear growth model with some noise

population = initial\_population + true\_growth\_rate \* time + np.random.normal(0, 2, size=time.shape)

# Creating a pandas DataFrame for easier analysis

data = pd.DataFrame({

'Time (Days)': time,

'Bacterial Population': population

})

data.head()

# Plot the data

plt.figure(figsize=(8,5))

plt.scatter(time, population, color='blue', label='Observed Population')

plt.title('Bacterial Population Growth Over Time')

plt.xlabel('Time (Days)')

plt.ylabel('Bacterial Population')

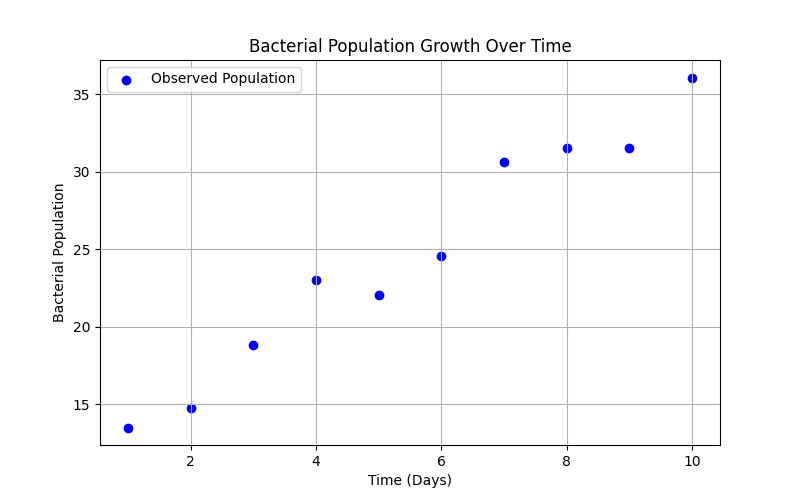
plt.legend()

plt.grid(True)

plt.show()

data

Result:



The dataset is as follows:

|  |  |
| --- | --- |
| **Time (Days)** | **Bacterial Population** |
| 1 | 13.49 |
| 2 | 14.72 |
| 3 | 18.80 |
| 4 | 23.05 |
| 5 | 22.03 |
| 6 | 24.53 |
| 7 | 30.66 |
| 8 | 31.53 |
| 9 | 31.56 |
| 10 | 36.09 |

The bacterial population follows a general upward trend over time, reflecting steady growth. Some fluctuations in the data are apparent, which may result from biological or environmental variations.

METHODOLOGY : Linear Regression and MMSE Calculation

To model the relationship between the time and bacterial population, a simple linear regression model was used. The model assumes that the bacterial population grows linearly over time, following the equation:

where:

* is the bacterial population
* is the time (in days),
* is the growth rate (slope),
* is the initial population (intercept).

After fitting the model to the observed data, we calculate the Mean Squared Error (MSE), which, in this case, is equivalent to the MMSE. The formula for MSE is:

where is the actual bacterial population, is the predicted bacterial population, and is the total number of data points.

# Performing linear regression to fit a model to the data

from sklearn.linear\_model import LinearRegression

# Reshaping time data for the model

X = time.reshape(-1, 1)

y = population

# Creating and fitting the linear regression model

model = LinearRegression()

model.fit(X, y)

# Making predictions based on the fitted model

y\_pred = model.predict(X)

# Calculating the Mean Squared Error (MSE)

mse = np.mean((y\_pred - y) \*\* 2)

mmse = mse # MMSE is equivalent to MSE

# Plotting the observed vs predicted data

plt.figure(figsize=(8,5))

plt.scatter(time, population, color='blue', label='Observed Population')

plt.plot(time, y\_pred, color='red', label='Predicted Population', linestyle='--')

plt.title('Bacterial Population Growth: Observed vs Predicted')

plt.xlabel('Time (Days)')

plt.ylabel('Bacterial Population')

plt.legend()

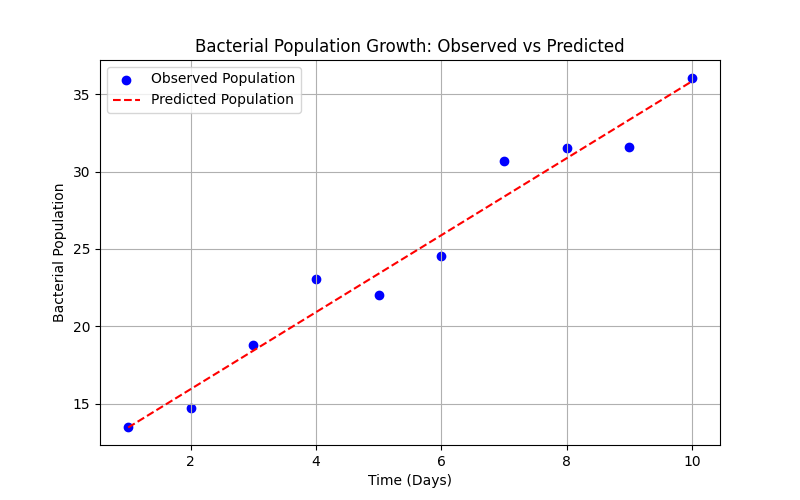
plt.grid(True)

plt.show()

# Displaying the coefficients of the fitted model and MMSE value

model.coef\_[0], model.intercept\_, mmse

(2.4861992566322373, 10.972026311920207, 1.8802964815832257)



RESULTS :

After fitting the model, we found the following key values:

* **Growth rate (slope):**
* **Initial population (intercept):**
* **Minimum Mean Square Error (MMSE):** 1.88

The growth rate of approximately 2.49 implies that the bacterial population increases by 2.49 units per day. The intercept (10.97) represents the estimated initial bacterial population at day 0, which aligns well with the data given the starting point at day 1.

The MMSE of 1.88 indicates a small average squared error between the predicted and actual values, suggesting a good fit of the model. In other words, on average, the model’s predictions deviate from the observed population by a small margin, making it a reliable predictor of bacterial growth over time.

DISCUSSION :

The linear model offers a straightforward yet effective means of understanding the relationship between time and bacterial population growth. While biological systems, such as bacterial growth, may follow more complex patterns (e.g., exponential growth phases), a linear model can still provide valuable insights in the early stages of population growth, particularly when only a small dataset is available.

The small MMSE value demonstrates that the linear model captures the essential trend in bacterial population dynamics over time. However, it is important to note that the model assumes a constant growth rate, which may not hold true in the long term, where environmental factors or resource limitations might slow down growth.

Moreover, since the data points were generated with some random noise, the model’s fit could be further improved by refining the dataset or adopting a more complex model that accounts for nonlinear growth patterns.

CONCLUSION :

The MMSE approach, coupled with linear regression, proves to be an effective method for estimating bacterial population growth over time. This analysis, based on a simulated dataset, suggests that the bacterial population grows at a steady rate of approximately 2.49 units per day. The model’s accuracy, as indicated by the MMSE value of 1.88, underscores the reliability of this approach in predicting bacterial growth.

Future studies could expand this model by including more complex growth patterns or larger datasets to capture the full spectrum of bacterial dynamics. Additionally, further experiments with real-world data could validate and improve upon the insights obtained in this study.

IREM ECE BAYRAM

2021205000