Experimental Error

Debasis Kundu
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

Outline of the Talk

Outline of the Talk

- Experiment

Outline of the Talk

- Experiment
- Data

Outline of the Talk

- Experiment
- Data
- Information

Outline of the Talk

- Experiment
- Data
- Information
- Different Types of Models

Outline of the Talk

- Experiment
- Data
- Information
- Different Types of Models
- Different Forms of Errors

Outline of the Talk

- Experiment
- Data
- Information
- Different Types of Models
- Different Forms of Errors
- Estimating Models/ Model Parameters

Outline of the Talk

- Experiment
- Data
- Information
- Different Types of Models
- Different Forms of Errors
- Estimating Models/ Model Parameters
- Estimating Errors

OUTLINE OF THE TALK

- Experiment
- Data
- Information
- Different Types of Models
- Different Forms of Errors
- Estimating Models/ Model Parameters
- Estimating Errors
- Conclusions

Experiment:

Experiment:

What is an experiment?

Experiment:

What is an experiment?

Experiment is a process which produces outcomes/ data.

Experiment:

We divide the experiments into two broad classes

EXPERIMENT:

We divide the experiments into two broad classes

- Random Experiment

Experiment:

We divide the experiments into two broad classes

- Random Experiment
- Non-Random Experiment

EXPERIMENT:

What is a Random Experiment?

EXPERIMENT:

What is a Random Experiment?

Random experiment is an experiment which produces random/ non-predictable outcomes. That means even if we repeat the experiment under the same environmental conditions, the outcomes may not be the same, whatever be the reason.

Experiment:

Examples:

Experiment:

Examples:

- Toss a coin 10 times.

EXPERIMENT:

Examples:

- Toss a coin 10 times.
- Rumford Cooling experiment (1798).

Experiment:

Examples:

- Toss a coin 10 times.
- Rumford Cooling experiment (1798).
- ECG data of a particular person.

Experiment:

Examples:

- Toss a coin 10 times.
- Rumford Cooling experiment (1798).
- ECG data of a particular person.
- Stock price of a particular brand.

Experiment:

What is a Non-Random Experiment?

Experiment:

What is a Non-Random Experiment?

If we know the exact outcome each time an experiment is performed, we call it a non-random experiment.

Experiment:

What is a Non-Random Experiment?

If we know the exact outcome each time an experiment is performed, we call it a non-random experiment.

Most of the experiments in nature are random experiments.

Experiment:

Experiment:

- We are mainly interested in non-random experiments.

EXPERIMENT:

- We are mainly interested in non-random experiments.
- It is assumed that the outcomes are quantitative in nature.

Experiment:

- We are mainly interested in non-random experiments.
- It is assumed that the outcomes are quantitative in nature.
- We call the outcomes as DATA.

EXPERIMENT:

- We are mainly interested in non-random experiments.
- It is assumed that the outcomes are quantitative in nature.
- We call the outcomes as DATA.
- Data can be univariate or multivariate.

EXPERIMENT:

- We are mainly interested in non-random experiments.
- It is assumed that the outcomes are quantitative in nature.
- We call the outcomes as DATA.
- Data can be univariate or multivariate.

AIM: Trying to extract maximum information from the observed DATA

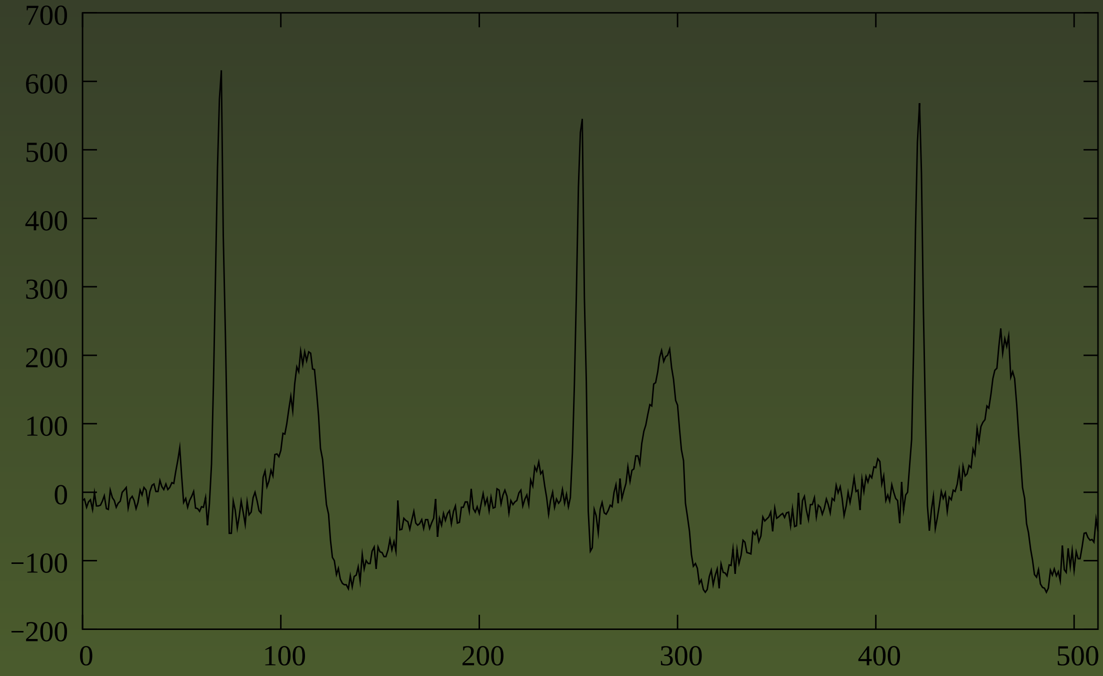
DATA

Example: Rumford Cooling Experiment
(Amount of heat generated by friction)

Time (min)	Temperature (^0F)	Time (min)	Temperature (^0F)
4	126	24	115
5	125	28	124
7	123	31	113
12	120	34	112
14	119	37.5	111
16	118	41	110
20	116		

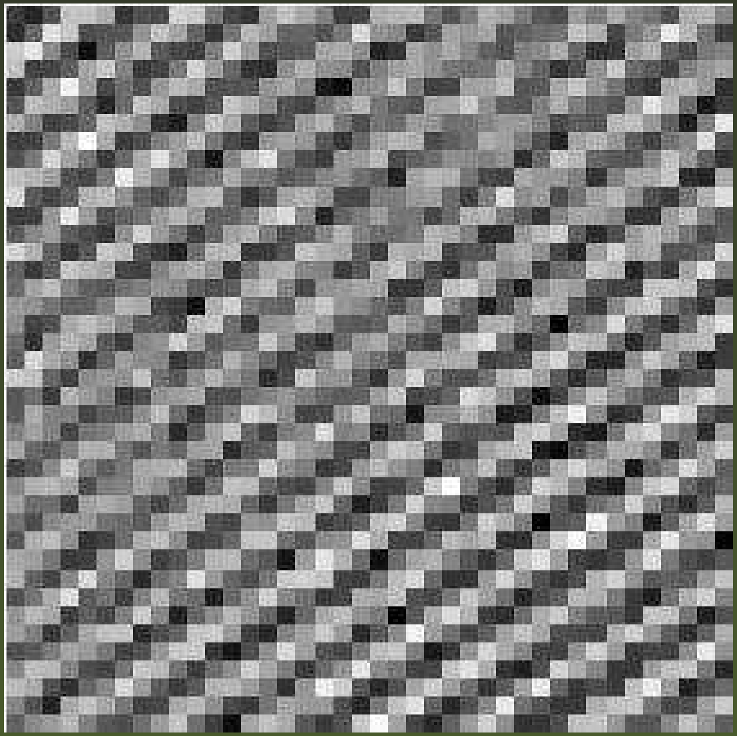
DATA

Example (ECG Data)



DATA

Example (Texture Data)



DATA:

Most of the Data can be expressed in the following form

DATA:

Most of the Data can be expressed in the following form

Data = Information/Signal + Noise/Error

DATA:

Most of the Data can be expressed in the following form

Data = Information/Signal + Noise/Error

Here, Information/ Signal is a very generic term. It may be completely deterministic or completely random or mixed. Our aim is to extract the Information/ Signal from the (noisy) Data.

INFORMATION

Why is it necessary?

INFORMATION

Why is it necessary?

- To understand the process itself.

Information

Why is it necessary?

- To understand the process itself.
- Prediction purposes.

INFORMATION

Why is it necessary?

- To understand the process itself.
- Prediction purposes.
- Storing purposes.

INFORMATION

Why is it necessary?

- To understand the process itself.
- Prediction purposes.
- Storing purposes.
- Designing the experiments.

INFORMATION

When is it possible?

Information

When is it possible?

Intuitively we feel that if the Information/ Signal is much more than the noise/ error, we should be able to extract the Information/ Signal from the noisy data. Otherwise it will be almost impossible.

INFORMATION

When is it possible?

Intuitively we feel that if the Information/ Signal is much more than the noise/ error, we should be able to extract the Information/ Signal from the noisy data. Otherwise it will be almost impossible.

Therefore, it is clear that understanding the noise/ error is very important in any $scientific$ experiment.

Information

We need to quantify Information. To quantify Information we approximate it by a *Mathematical Model*, therefore

INFORMATION

We need to quantify Information. To quantify Information we approximate it by a *Mathematical Model*, therefore

$$\text{Information} \approx \text{Model}$$

MODELS

Models

Data = Model + Error

MODELS

$$\text{Data} = \text{Model} + \text{Error}$$

Error has several components

MODELS

$$\text{Data} = \text{Model} + \text{Error}$$

Error has several components

- Modeling Error

MODELS

$$\text{Data} = \text{Model} + \text{Error}$$

Error has several components

- Modeling Error
- Experimentation Error

MODELS

Data = Model + Error

Error has several components

- Modeling Error
- Experimentation Error
- Data Collection Error

MODELS

Data = Model + Error

Error has several components

- Modeling Error
- Experimentation Error
- Data Collection Error
- Calibration Error

MODELS

Different Types of Model

MODELS

Different Types of Model

- Mechanistic Model

Models

Different Types of Model

- Mechanistic Model
- Functional Model

Models

Mechanistic Model:

MODELS

Mechanistic Model:

A mechanistic model is a model based on the exact mechanism of an experiment. It is usually quite complex but most of the time it is quite effective.

Models

Functional Model:

MODELS

Functional Model:

It is purely a mathematical model. One does not need any experimental knowledge to construct a functional model.

MODELS

Using the mechanistic/ functional model often the data can be written in the following form:

MODELS

Using the mechanistic/ functional model often the data can be written in the following form:

$$\text{Data} = Y = f(X) + \text{Error}$$

MODELS

Using the mechanistic/ functional model often the data can be written in the following form:

$$\text{Data} = Y = f(X) + \text{Error}$$

- Y and X are known

MODELS

Using the mechanistic/ functional model often the data can be written in the following form:

$$\text{Data} = Y = f(X) + \text{Error}$$

- Y and X are known
- Error is unknown

Models

Using the mechanistic/ functional model often the data can be written in the following form:

$$\text{Data} = Y = f(X) + \text{Error}$$

- Y and X are known
- Error is unknown

Aim: From $\{X_1, Y_1\}, \ldots, \{X_n, Y_n\}$ we want to estimate $f(\cdot)$.

EXAMPLES

Rumford Cooling Data

- Mechanistic Model:
- Based on Newton's law of cooling the following model has been used:

$$Y(t) = 60 + 70e^{-\theta t} + \epsilon(t)$$

Examples

ECG Data

- Functional Model:

$$Y(t) = \sum_{k=1}^{p} \{A_k \cos(\omega_k t) + B_k \sin(\omega_k t)\} + \epsilon(t)$$

Examples

Texture Data

- Functional Model:

$$Y(m,n) = \sum_{k=1}^{p} \{A_k \cos(\omega_k m + \lambda_k n) + B_k \sin(\omega_k m + \lambda_k n)\} + \epsilon(m,n)$$

Models: Estimation

Possible Approaches:

Models: Estimation

Possible Approaches:

- From Y and X estimate $f(\cdot)$. (Non-Parametric Approach)

Models: Estimation

Possible Approaches:

- From Y and X estimate $f(\cdot)$. (Non-Parametric Approach)
- Assume $f(X) = f(X, \theta)$, here the form of $f(\cdot)$ is known but the parameter θ is unknown. Therefore, estimate θ assuming $f(\cdot)$ is known. Once θ has been estimated and since $f(\cdot)$ is known, therefore $f(X, \theta)$ becomes known. (Parametric Approach)

Errors

Different forms of Errors:

ERRORS

Different forms of Errors:

- Random Errors

Errors

Different forms of Errors:

- Random Errors
- Systematic Errors

Errors

Different Assumptions on Errors:

Errors

Different Assumptions on Errors:

- Errors are purely random. They have mean zeros and same variances. They are independent.

Errors

Different Assumptions on Errors:

- Errors are purely random. They have mean zeros and same variances. They are independent.
- They have mean zeros but unequal variances. They are independent.

Errors

Different Assumptions on Errors:

- Errors are purely random. They have mean zeros and same variances. They are independent.
- They have mean zeros but unequal variances. They are independent.
- They have mean zeros, equal variances but are correlated.

Errors

Different Assumptions on Errors:

- Errors are purely random. They have mean zeros and same variances. They are independent.
- They have mean zeros but unequal variances. They are independent.
- They have mean zeros, equal variances but are correlated.
- Spurious errors.

Errors

Different Assumptions on Errors:

- Errors are purely random. They have mean zeros and same variances. They are independent.
- They have mean zeros but unequal variances. They are independent.
- They have mean zeros, equal variances but are correlated.
- Spurious errors.
- Errors have some particular probability distribution.

Errors

Why are the assumptions necessary?

ERRORS

Why are the assumptions necessary?

- Choosing the correct model.

Errors

Why are the assumptions necessary?

- Choosing the correct model.
- Effective estimation of the unknown parameters.

Errors

Why are the assumptions necessary?

- Choosing the correct model.
- Effective estimation of the unknown parameters.
- Effective prediction of the future observations if necessary.

Errors

Why are the assumptions necessary?

- Choosing the correct model.
- Effective estimation of the unknown parameters.
- Effective prediction of the future observations if necessary.
- Constructing error bars of the estimates.

ERRORS

Why are the assumptions necessary?

- Choosing the correct model.
- Effective estimation of the unknown parameters.
- Effective prediction of the future observations if necessary.
- Constructing error bars of the estimates.
- Deriving properties of the estimates.

ERRORS

How to verify the assumptions?

Errors

How to verify the assumptions?

- Choose a particular $f(\cdot, \theta)$ from a class of possible models.

Errors

How to verify the assumptions?

- Choose a particular $f(\cdot, \theta)$ from a class of possible models.
- Make some simple assumptions on the error.

ERRORS

How to verify the assumptions?

- Choose a particular $f(\cdot, \theta)$ from a class of possible models.
- Make some simple assumptions on the error.
- Estimate the parameters.

Errors

How to verify the assumptions?

- Choose a particular $f(\cdot, \theta)$ from a class of possible models.
- Make some simple assumptions on the error.
- Estimate the parameters.
- Estimate the error also.

Errors

How to verify the assumptions?

- Choose a particular $f(\cdot, \theta)$ from a class of possible models.
- Make some simple assumptions on the error.
- Estimate the parameters.
- Estimate the error also.
- Then verify the error assumptions.

ERRORS

How to verify the assumptions?

- Choose a particular $f(\cdot, \theta)$ from a class of possible models.
- Make some simple assumptions on the error.
- Estimate the parameters.
- Estimate the error also.
- Then verify the error assumptions.
- If the error assumptions are true, then be HAPPY otherwise

ERRORS

How to verify the assumptions?

- Choose a particular $f(\cdot, \theta)$ from a class of possible models.
- Make some simple assumptions on the error.
- Estimate the parameters.
- Estimate the error also.
- Then verify the error assumptions.
- If the error assumptions are true, then be HAPPY otherwise
- Change the model, change the error assumptions and continue the process.

Estimating Model Parameters

Least Squares Estimators (Most Popular One)

Estimating Model Parameters

Least Squares Estimators (Most Popular One)

$$\min_{\theta} \sum_{i=1}^{n} \left(Y_i - f(X_i, \theta)\right)^2$$

Estimating Model Parameters

Least Squares Estimators (Most Popular One)

$$\min_{\theta} \sum_{i=1}^{n} \left(Y_i - f(X_i, \theta)\right)^2$$

- It works very well if the errors are independent with mean zero and same variances.

Estimating Model Parameters

Least Squares Estimators (Most Popular One)

$$\min_{\theta} \sum_{i=1}^{n} (Y_i - f(X_i, \theta))^2$$

- It works very well if the errors are independent with mean zero and same variances.
- If the error assumptions are not met then the estimates can be quite bad.

Estimating Model Parameters

Generalized Least Squares Estimators

Estimating Model Parameters

Generalized Least Squares Estimators

$$\min_{\theta} \sum_{i=1}^{n} \frac{\left(Y_i - f(X_i, \theta)\right)^2}{\text{Variance}(\epsilon_i)}$$

Estimating Model Parameters

Generalized Least Squares Estimators

$$\min_{\theta} \sum_{i=1}^{n} \frac{\left(Y_i - f(X_i, \theta)\right)^2}{\text{Variance}(\epsilon_i)}$$

- It works very well if the errors are independent with mean zero and different variances.

Estimating Model Parameters

Generalized Least Squares Estimators

$$\min_{\theta} \sum_{i=1}^{n} \frac{(Y_i - f(X_i, \theta))^2}{\text{Variance}(\epsilon_i)}$$

- It works very well if the errors are independent with mean zero and different variances.
- It works much better than the least squares estimators.

Estimating Model Parameters

Least Absolute Deviations

Estimating Model Parameters

Least Absolute Deviations

$$\min_{\theta} \sum_{i=1}^{n} |Y_i - f(X_i, \theta)|$$

Estimating Model Parameters

Least Absolute Deviations

$$\min_{\theta} \sum_{i=1}^{n} |Y_i - f(X_i, \theta)|$$

- It works very well if the errors are spurious in nature.

Estimating Model Parameters

Least Absolute Deviations

$$\min_{\theta} \sum_{i=1}^{n} |Y_i - f(X_i, \theta)|$$

- It works very well if the errors are spurious in nature.
- The estimates are quite robust as compared to the least squares estimators.

Estimating Model Parameters

Least Absolute Deviations

$$\min_{\theta} \sum_{i=1}^{n} |Y_i - f(X_i, \theta)|$$

- It works very well if the errors are spurious in nature.
- The estimates are quite robust as compared to the least squares estimators.
- Computationally it is very demanding.

Conclusions

CONCLUSIONS

All experimental data will have errors

Conclusions

All experimental data will have errors

To extract maximum information from the data

CONCLUSIONS

All experimental data will have errors

To extract maximum information from the data

- Proper Error Assumptions/ Error Analysis Are Required

CONCLUSIONS

All experimental data will have errors

To extract maximum information from the data

- Proper Error Assumptions/ Error Analysis Are Required
- Proper Model Assumptitons/ Model Analysis Are Required

CONCLUSIONS

All experimental data will have errors

To extract maximum information from the data

- Proper Error Assumptions/ Error Analysis Are Required
- Proper Model Assumptitons/ Model Analysis Are Required

Professional Statistician’s Help is Needed

Thank You