Digital Image Processing

Daniel T. McGuiness, PhD

Version: γ.2024.1

MCI

Table of Contents

Lecture Structure

Table of Contents

Lecture Structure

- The goal of this lecture is to give you the fundamentals of digital image processing and understanding of mathematical principles.
- This lecture is a total of 4 SWS with a total of sixty (60) hours.
- There are two (2) assignments for this course

  1st will be a pre-defined work which is individual based.
  2nd will be group based.

> You are to come up with a project that uses DIP using Python.

- You will work with a group of up to three (3) or two (2).
- You are to come up with a group and decide on your topic.

- The individual assignment focuses on understanding DIP principles.
- The assignment is uploaded to SAKAI for you to work on along with what is required of you for submission.
  - The assignment contains questions where applications of DIP will be needed.
- The deadline is the end day of **last lecture before presentations**.

A Help in Colour

Due to the nature of the topic, some aspects are to be presented in a colour spectrum some student may not be able to perceive. In situations like this, please let me know if there are some diagrams or some colour choices making the lecture illegible via mail and I will send you a colour correct version based on the condition.

Lecture Structure

- For your project use Python.
- Some possible project ideas:
  - License plate detection,
  - Handwriting detection,
  - Signature verification,
  - Face detection,
  - Image to text conversion,
  - Barcode detection,
  - Convert sudoku drawings to computer code.
  - Book detection.

> The use of AI/ML is allowed as long as clear explanation is given and its process is understood.

Lecture Structure

- The last three (3) appointments are reserved for group presentations.
- You will do a presentation in front of the class for 20 mins.
- The next 20 mins following your presentation will be the Q&A.
- The Q&A will involve two (2) questions from your relevant work.
- You are also to submit a report with your project detailing the work.

> Each student needs to declare the part the student worked on.

- i.e., Student A has done the writing, edge detection
- i.e., Student B has done the data analysis, figure generation.
- You are to submit your reports and all relevant resources to SAKAI no later than 2 weeks before your assigned presentation.

Lecture Structure

Assessment Type	Overall Points	Breakdown	%
Homework	40		
		Report	20
		Solution(s)	60
		Code Analysis	20
Group Project	60		
		Report	40
		Presentation	40
		Q & A	20

Table 1: Assessment Grade breakdown for the lecture.

Lecture Structure

Covered Topic	Appointment
Mathematical Fundamentals	1
Perception	2
Camera	2-3
Display	4
Noise	4-5
Histogram Operations	6
Morphological Operations	7
Blurring Filters	8
Feature Analysis	9
Edge Detection	10
Neural Networks for Image Processing	11-12
Group Assignment Presentations	13-15

Table 2: Distribution of materials across the semester.

Mathematical Fundamentals

- 2D Convolution,
- Discrete Fourier Transform,
- Sampling Theorem

Lecture Structure

Perception

- Colour Blindness,
- Colour Standards,
- Colour Models

Cameras

- Used sensors,
- Lenses,
- Sensitivity

Displays

- Dithering,
- Interlacing,
- Display Technologies

Lecture Structure

Noise

- Types of noises,
- Modelling Noises,
- Random Noise generation

Lecture Structure

Histogram Operations

- Colour Channels,
- Masking,
- Dynamic Range

Morphological Operations

- Opening,
- Closing,
- Erosion,
- Dilation.

Lecture Structure

Blurring Filters

- Gaussian Blurring,
- Multivariate Distribution,
- Bilinear Filtering

Feature Analysis

- ORB Feature Extractor,
- Adaptive Threshold,
- Scale Invariant Feature Transform.

Egde Detection

- Defining an Edge to the computer,
- Types of Kernels,
- Canny Edge Detection.

Neural Networks for Image Processing

- Defining ANNs,
- OCR,
- ResNet.

Lecture Structure

Books

- Forsyth, Ponce "*Computer Vision: A Modern Approach*" Prentice-Hall, 2003.
- Young I. "*Fundamentals of Image Processing*" Delft 1998.
- Szeliski R. "*Computer Vision: Algorithms and Applications*" Springer 2022.
- Nixon M. et. al "*Feature Extraction and Image Processing for Computer Vision*" Academic press 2019.
- Gonzalez R. "*Digital Image Processing*" Pearson 2009

White Papers

- Luminera "*Getting it Right: Selecting a Lens for a Vision System*",
- Luminera "*The Complete Guide to Industrial Camera Lenses*",
- Fowler B, et. al, *Read Noise Distribution Modeling for CMOS Image Sensors*.
- Oxford Instruments *Understanding Read Noise in sCMOS Cameras*.

Lecture Structure

Lecture Notes

- Applied Multi-variable Statistical Analysis "*Lesson 4: Multivariate Normal Distribution*",
- Statistical Theory and Methods I "*Chapter 3: Multivariate Distributions*", Stephen M. Stigler
- The Discrete Fourier Transform "*Signal Processing & Filter Design*", Stephen Roberts.
- Procedural Generation: 2D Perlin Noise *Game Programming*, Mount .E, Eastman R.
- Foundations of computer vision: Lecture notes, Carreira-Perpinan M.
- Computer Vision, CMU School of Computer Science
- Computer Vision, University of Cambridge
- Computer Vision, NYU Computer Science

Web Resources

- Scikit-image documentation
- OpenCV documentation
- Pillow (fork of PIL) documentation

Mathematical Fundamentals

Table of Contents

Learning Outcomes

(LO1) An Overview of Mathematical Methods,

(LO2) Description of Analogue and Digital,

(LO3) Fourier Analysis Overview,

(LO4) Convolution Introduction.

Learning Outcomes

(LO1) An Overview of Mathematical Methods,
(LO2) Description of Analogue and Digital,
(LO3) Fourier Analysis Overview,
(LO4) Convolution Introduction.

Learning Outcomes

(LO1) An Overview of Mathematical Methods,

(LO2) Description of Analogue and Digital,

(LO3) Fourier Analysis Overview,

(LO4) Convolution Introduction.

Learning Outcomes

(LO1) An Overview of Mathematical Methods,

(LO2) Description of Analogue and Digital,

(LO3) Fourier Analysis Overview,

(LO4) Convolution Introduction.

Mathematical Fundamentals

- Computer Vision encompasses multiple disciplines, including digital image processing, cameras and displays.
- To better prepare, it is important to refresh/learn some fundamental mathematical principles & concepts.

Concepts and Principles

- Convolution
- Fourier Analysis
  - Properties
  - Discrete Fourier Transform
- Shannon-Nyquist Sampling Theorem
- A brief introduction to Information Theory
  - Entropy in Information Theory

Mathematical Fundamentals

- Computer Vision encompasses multiple disciplines, including digital image processing, cameras and displays.
- To better prepare, it is important to refresh/learn some fundamental mathematical principles & concepts.

Concepts and Principles

- Convolution
- Fourier Analysis
  - Properties
  - Discrete Fourier Transform
- Shannon-Nyquist Sampling Theorem
- A brief introduction to Information Theory
  - Entropy in Information Theory

Mathematical Fundamentals

- Computer Vision encompasses multiple disciplines, including digital image processing, cameras and displays.
- To better prepare, it is important to refresh/learn some fundamental mathematical principles & concepts.

Concepts and Principles

- Convolution
- Fourier Analysis
  - Properties
  - Discrete Fourier Transform
- Shannon-Nyquist Sampling Theorem
- A brief introduction to Information Theory
  - Entropy in Information Theory

Mathematical Fundamentals

- Computer Vision encompasses multiple disciplines, including digital image processing, cameras and displays.
- To better prepare, it is important to refresh/learn some fundamental mathematical principles & concepts.

Concepts and Principles

- Convolution
- Fourier Analysis
  - Properties
  - Discrete Fourier Transform
- Shannon-Nyquist Sampling Theorem
- A brief introduction to Information Theory
  - Entropy in Information Theory

Mathematical Fundamentals

- Computer Vision encompasses multiple disciplines, including digital image processing, cameras and displays.
- To better prepare, it is important to refresh/learn some fundamental mathematical principles & concepts.

Concepts and Principles

- Convolution
- Fourier Analysis
  - Properties
  - Discrete Fourier Transform
- Shannon-Nyquist Sampling Theorem
- A brief introduction to Information Theory
  - Entropy in Information Theory

- Computer Vision encompasses multiple disciplines, including digital image processing, cameras and displays.
- To better prepare, it is important to refresh/learn some fundamental mathematical principles & concepts.

Concepts and Principles

- Convolution
- Fourier Analysis
  - Properties
  - Discrete Fourier Transform
- Shannon-Nyquist Sampling Theorem
- A brief introduction to Information Theory
  - Entropy in Information Theory

Mathematical Fundamentals

- Computer Vision encompasses multiple disciplines, including digital image processing, cameras and displays.
- To better prepare, it is important to refresh/learn some fundamental mathematical principles & concepts.

Concepts and Principles

- Convolution
- Fourier Analysis
  - Properties
  - Discrete Fourier Transform
- Shannon-Nyquist Sampling Theorem
- A brief introduction to Information Theory
  - Entropy in Information Theory

Mathematical Fundamentals

- Computer Vision encompasses multiple disciplines, including digital image processing, cameras and displays.
- To better prepare, it is important to refresh/learn some fundamental mathematical principles & concepts.

Concepts and Principles

- Convolution
- Fourier Analysis
  - Properties
  - Discrete Fourier Transform
- Shannon-Nyquist Sampling Theorem
- A brief introduction to Information Theory
  - Entropy in Information Theory

Mathematical Fundamentals

- Computer Vision encompasses multiple disciplines, including digital image processing, cameras and displays.
- To better prepare, it is important to refresh/learn some fundamental mathematical principles & concepts.

Concepts and Principles

- Convolution
- Fourier Analysis
  - Properties
  - Discrete Fourier Transform
- Shannon-Nyquist Sampling Theorem
- A brief introduction to Information Theory
  - Entropy in Information Theory

Mathematical Fundamentals

- Convolution, mathematically is defined as:

$$(f * g)(t) = \int_{-\infty}^{+\infty} f(\tau)\, g(t - \tau)\, \mathrm{d}\tau.$$

> In layman's terms convolution is just fancy multiplication.

Example

Imagine you manage a hospital treating patients with a single disease.

You have:

Treatment Plan `3` Every patient gets 3 units of the cure on their first day.

Patient List `[1, 2, 3, 4, 5]` Your patient count for the week (1 person Monday, 2 people on Tuesday, etc.).

How much medicine do you use each day?

Example

Imagine you manage a hospital treating patients with a single disease.

You have:

Treatment Plan `3` Every patient gets 3 units of the cure on their first day.

Patient List `[1, 2, 3, 4, 5]` Your patient count for the week (1 person Monday, 2 people on Tuesday, etc.).

How much medicine do you use each day?

Solution

The answer is a quick multiplication:

$$\text{Plan} \times \text{Patients} = \text{Daily Usage}$$

$$\texttt{3} \times \texttt{[1, 2, 3, 4, 5]} = \texttt{[3, 6, 9, 12, 15]}$$

Multiplying the plan by the patient list gives usage for upcoming days:

`[3, 6, 9, 12, 15]`

Everyday multiplication of (3×4) means using the plan with a single day of patients:

$$\texttt{[3]} \times \texttt{[4]} = \texttt{[12]}$$

Example

Now the disease mutates and needs multi-day treatment. A new plan:

Plan: `[3, 2, 1]`

Meaning:

- 3 units of the cure on day one,
- 2 units on day two,
- 1 unit on day three.

Given the same patient schedule of:

Patient: `[1, 2, 3, 4, 5]`

what's our medicine usage each day?

Example

Now the disease mutates and needs multi-day treatment. A new plan:

Plan: `[3, 2, 1]`

Meaning:

- 3 units of the cure on day one,
- 2 units on day two,
- 1 unit on day three.

Given the same patient schedule of:

Patient: `[1, 2, 3, 4, 5]`

what's our medicine usage each day?

Example

Now the disease mutates and needs multi-day treatment. A new plan:

Plan: `[3, 2, 1]`

Meaning:

- 3 units of the cure on day one,
- 2 units on day two,
- 1 unit on day three.

Given the same patient schedule of:

Patient: `[1, 2, 3, 4, 5]`

what's our medicine usage each day?

Solution

Let's see

- On day 1, 1 patient `A` comes in. It's their first day, so 3 units.
- On day 2, `A` gets 2 units (second day), but two new patients (`B1` & `B2`) arrive, who get 3 each ($2 \times 3 = 6$).
  - The total is $2 + (2 \times 3) = 8$ units.
- On Wednesday, it's trickier: The patient `A` finishes (1 unit, her last day), the `B1` and `B2` get 2 units (2 * 2), and there are 3 new Wednesday people....

The patients are overlapping and it's hard to track. How can we organise this calculation?

Solution

Let's see

- On day 1, 1 patient `A` comes in. It's their first day, so 3 units.
- On day 2, `A` gets 2 units (second day), but two new patients (`B1` & `B2`) arrive, who get 3 each ($2 \times 3 = 6$).
  - The total is $2 + (2 \times 3) = 8$ units.
- On Wednesday, it's trickier: The patient `A` finishes (1 unit, her last day), the `B1` and `B2` get 2 units (2 * 2), and there are 3 new Wednesday people....

The patients are overlapping and it's hard to track. How can we organise this calculation?

Solution

Let's see

- On day 1, 1 patient `A` comes in. It's their first day, so 3 units.
- On day 2, `A` gets 2 units (second day), but two new patients (`B1` & `B2`) arrive, who get 3 each ($2 \times 3 = 6$).
  - The total is $2 + (2 \times 3) = 8$ units.
- On Wednesday, it's trickier: The patient `A` finishes (1 unit, her last day), the `B1` and `B2` get 2 units (2 * 2), and there are 3 new Wednesday people....

The patients are overlapping and it's hard to track. How can we organise this calculation?

Solution

Let's see

- On day 1, 1 patient `A` comes in. It's their first day, so 3 units.
- On day 2, `A` gets 2 units (second day), but two new patients (`B1` & `B2`) arrive, who get 3 each ($2 \times 3 = 6$).
  - The total is $2 + (2 \times 3) = 8$ units.
- On Wednesday, it's trickier: The patient `A` finishes (1 unit, her last day), the `B1` and `B2` get 2 units (2 * 2), and there are 3 new Wednesday people....

The patients are overlapping and it's hard to track. How can we organise this calculation?

Solution

An idea worth considering is to reverse the order of the patient list:

New Patient List: `[5, 4, 3, 2, 1]`

Next, imagine we have 3 separate rooms where we apply the proper dose:

Rooms: `[3, 2, 1]`

On your first day, you walk into the first room and get 3 units of medicine. The next day, you walk into room #2 and get 2 units. On the last day, you walk into room #3 and get 1 unit. There's no rooms afterwards, and your treatment is done.

Solution

To calculate the total medicine usage, line up the patients and walk them through the rooms:

```
Monday
----------------------------
Rooms                  3 2 1
Patients       5 4 3 2 1

Usage                  3
```

On Monday (our first day), we have a single patient in the first room. `A` gets 3 units, for a total usage of 3.

Makes sense, right?

Solution

On Tuesday, everyone takes a step forward:

```
Tuesday
----------------------------
Rooms                  3 2 1
Patients ->      5 4 3 2 1

Usage                  6 2      = 8
```

The first patient is now in the second room, and there's 2 new patients in the first room. We multiply each room's dose by the patient count, then combine.

Solution

```
Wednesday
----------------------------
Rooms                      3 2 1
Patients ->            5 4 3 2 1
Usage                      9 4 1     = 14

Thursday
-----------------------------
Rooms                      3 2 1
Patients ->              5 4 3 2 1
Usage                     12 6 2     = 20

Friday
-----------------------------
Rooms                      3 2 1
Patients ->                5 4 3 2 1
Usage                     15 8 3     = 26
```

Solution

It's intricate, but we figured it out, right? We can find the usage for any day by reversing the list, sliding it to the desired day, and combining the doses.

The total day-by-day usage looks like this (don't forget Sat and Sun, since some patients began on Friday):

```
Plan      *  Patient List   = Total Daily Usage

[3 2 1]   *  [1 2 3 4 5]    = [3 8 14 20 26 14 5]
              M T W T F         M T W  T  F  S  S
```

This calculation is the convolution of the plan and patient list. It's a fancy multiplication between a list of input numbers and a "program".

Example

Write a script which does convolution of the following two (2) arrays:

$$A = [1,\ 1,\ 2,\ 2,\ 1] \qquad B = [1,\ 1,\ 1,\ 3]$$

Solution

```python
import numpy as np
def convolve_1d(signal, kernel):
    kernel = kernel[::-1]
    k = len(kernel)
    s = len(signal)
    signal = [0]*(k-1)+signal+[0]*(k-1)
    n = s+(k-1)
    res = []
    for i in range(s+k-1):
        res.append(np.dot(signal[i:(i+k)], kernel))
    return res
```

```python
A = [1,1,2,2,1]
B = [1,1,1,3]

print(convolve_1d(A, B))
```

- An operation on two functions (f and g) that produces $f * g$.

> It expresses how the shape of one is modified by the other.

- There are several notations to indicate convolution with the most common is:

$$c = f\left(t\right) * g\left(t\right) = \left(f * g\right)\left(t\right),$$

Mathematical Fundamentals

- An operation on two functions (f and g) that produces $f * g$.

It expresses how the shape of one is modified by the other.

- There are several notations to indicate convolution with the most common is:

$$c = f(t) * g(t) = (f * g)(t),$$

- In 2D continuous space (i.e., analogue):

$$
\begin{aligned}
c(x,\ y) &= f(x,\ y) * g(x,\ y)\,, \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(\chi,\ \xi)\ g(x-\chi,\ y-\xi)\ \mathrm{d}\chi\,\mathrm{d}\xi.
\end{aligned}
$$

- In 2D discrete space (i.e., digital):

$$
\begin{aligned}
c[m,\ n] &= f[m,\ n] * g[m,\ n]\,, \\
&= \sum_{j=-\infty}^{+\infty}\sum_{k=-\infty}^{+\infty} f[j,\ k]\ g[m-j,\ n-k]\,.
\end{aligned}
$$

- In 2D continuous space (i.e., analogue):

$$
\begin{aligned}
c\left(x,\, y\right) &= f\left(x,\, y\right) * g\left(x,\, y\right), \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f\left(\chi,\, \xi\right)\, g\left(x - \chi,\, y - \xi\right)\, \mathrm{d}\chi\, \mathrm{d}\xi.
\end{aligned}
$$

- In 2D discrete space (i.e., digital):

$$
\begin{aligned}
c\left[m,\, n\right] &= f\left[m,\, n\right] * g\left[m,\, n\right], \\
&= \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} f\left[j,\, k\right]\, g\left[m - j,\, n - k\right].
\end{aligned}
$$

- It is the single most important technique in Digital Signal Processing.
- Using the strategy of impulse decomposition, systems are described by a signal called the impulse response.
- Convolution is important because it relates the three (3) signals of interest:
  1. Input signal,
  2. Output signal,
  3. Impulse response.

- It is the single most important technique in Digital Signal Processing.
- Using the strategy of impulse decomposition, systems are described by a signal called the impulse response.
- Convolution is important because it relates the three (3) signals of interest:
  1. Input signal,
  2. Output signal,
  3. Impulse response.

- It is the single most important technique in Digital Signal Processing.
- Using the strategy of impulse decomposition, systems are described by a signal called the impulse response.
- Convolution is important because it relates the three (3) signals of interest:
  1. Input signal,
  2. Output signal,
  3. Impulse response.

Mathematical Fundamentals

- It is the single most important technique in Digital Signal Processing.
- Using the strategy of impulse decomposition, systems are described by a signal called the impulse response.
- Convolution is important because it relates the three (3) signals of interest:
    1. Input signal,
    2. Output signal,
    3. Impulse response.

- It is the single most important technique in Digital Signal Processing.
- Using the strategy of impulse decomposition, systems are described by a signal called the impulse response.
- Convolution is important because it relates the three (3) signals of interest:
    1. Input signal,
    2. Output signal,
    3. Impulse response.

Mathematical Fundamentals

- It is the single most important technique in Digital Signal Processing.
- Using the strategy of impulse decomposition, systems are described by a signal called the impulse response.
- Convolution is important because it relates the three (3) signals of interest:
  1. Input signal,
  2. Output signal,
  3. Impulse response.

Commutative

- The order in which we convolve two signals does not change the result:

$$f(t) * g(t) = g(t) * f(t)$$

Distributive

- if there are three signals $f(t), g(t), h(t)$, then the convolution of $f(t)$ is distributive:

$$f(t) * [g(t) + h(t)] = [f(t) * g(t)] + [f(t) * h(t)]$$

Commutative

- The order in which we convolve two signals does not change the result:

$$f\,(t) * g\,(t) = g\,(t) * f\,(t)$$

Distributive

- if there are three signals $f\,(t)\,, g\,(t)\,, h\,(t)$, then the convolution of $f\,(t)$ is distributive:

$$f\,(t) * [g\,(t) + h\,(t)] = [f\,(t) * g\,(t)] + [f\,(t) * h\,(t)]$$

Associative

- The way in which the signals are grouped in a convolution does not change the result:

$$f(t) * [g(t) * h(t)] = [f(t) * g(t)] * h(t)$$

Shift Property

- The convolution of a signal with a time shifted signal results a shifted version of that signal.
- i.e.,

$$f(t) * g(t) = y(t)$$

- Then according to the shift property of convolution:

$$f(t) * f(t - T_0) = y(t - T_0)$$

- Similarly:

$$f(t - T_0) * f(t) = y(t - T_0)$$

- Therefore:

$$f(t - T_1) * f(t - T_2) = y(t - T_1 - T_2)$$

Shift Property

- The convolution of a signal with a time shifted signal results a shifted version of that signal.
- i.e.,

$$f(t) * g(t) = y(t)$$

- Then according to the shift property of convolution:

$$f(t) * f(t - T_0) = y(t - T_0)$$

- Similarly:

$$f(t - T_0) * f(t) = y(t - T_0)$$

- Therefore:

$$f(t - T_1) * f(t - T_2) = y(t - T_1 - T_2)$$

Mathematical Fundamentals

Shift Property

- The convolution of a signal with a time shifted signal results a shifted version of that signal.
- i.e.,

$$f(t) * g(t) = y(t)$$

- Then according to the shift property of convolution:

$$f(t) * f(t - T_0) = y(t - T_0)$$

- Similarly:

$$f(t - T_0) * f(t) = y(t - T_0)$$

- Therefore:

$$f(t - T_1) * f(t - T_2) = y(t - T_1 - T_2)$$

Shift Property

- The convolution of a signal with a time shifted signal results a shifted version of that signal.
- i.e.,

$$f(t) * g(t) = y(t)$$

- Then according to the shift property of convolution:

$$f(t) * f(t - T_0) = y(t - T_0)$$

- Similarly:

$$f(t - T_0) * f(t) = y(t - T_0)$$

- Therefore:

$$f(t - T_1) * f(t - T_2) = y(t - T_1 - T_2)$$

Shift Property

- The convolution of a signal with a time shifted signal results a shifted version of that signal.
- i.e.,

$$f(t) * g(t) = y(t)$$

- Then according to the shift property of convolution:

$$f(t) * f(t - T_0) = y(t - T_0)$$

- Similarly:

$$f(t - T_0) * f(t) = y(t - T_0)$$

- Therefore:

$$f(t - T_1) * f(t - T_2) = y(t - T_1 - T_2)$$

Mathematical Fundamentals

$$\underset{I}{\begin{pmatrix} 0_{\times 1} & 1_{\times 0} & 1_{\times 1} & 1 & 0 & 0 & 0 \\ 0_{\times 0} & 0_{\times 1} & 1_{\times 0} & 1 & 1 & 0 & 0 \\ 0_{\times 1} & 0_{\times 0} & 0_{\times 1} & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}} * \underset{K}{\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}} = \underset{I * K}{\begin{pmatrix} \boxed{1} & 4 & 3 & 4 & 1 \\ 1 & 2 & 4 & 3 & 3 \\ 1 & 2 & 3 & 4 & 1 \\ 1 & 3 & 3 & 1 & 1 \\ 3 & 3 & 1 & 1 & 0 \end{pmatrix}}$$

Figure 1: A visual representation of how convolution works in 2D.

- Converting from a continuous 2D data $a(x, y)$ to its digital representation $a[x, y]$ requires the process of sampling.
- An ideal sampling system is defined as the image $a(x, y)$ multiplied by an ideal 2D impulse train $\delta(x, y)$:

$$
\begin{aligned}
b[m, n] &= a(x, y) \sum_{m=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \delta(x - mX_0, y - nY_0) \\
&= \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} a(mX_0, nY_0)\, \delta(x - mX_0, y - nY_0).
\end{aligned}
$$

where X_0 and Y_0 are the sampling distance or intervals and δ is the Dirac delta function.

- If you were to sample in square shapes $X_0 = Y_0$ where you could think of each individual block a pixel

- Converting from a continuous 2D data $a(x, y)$ to its digital representation $a[x, y]$ requires the process of sampling.
- An ideal sampling system is defined as the image $a(x, y)$ multiplied by an ideal 2D impulse train $\delta(x, y)$:

$$b[m, n] = a(x, y) \sum_{m=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \delta(x - mX_0, y - nY_0)$$
$$= \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} a(mX_0, nY_0)\, \delta(x - mX_0, y - nY_0).$$

where X_0 and Y_0 are the sampling distance or intervals and δ is the Dirac delta function.

- If you were to sample in square shapes $X_0 = Y_0$ where you could think of each individual block a pixel

- Converting from a continuous 2D data $a(x, y)$ to its digital representation $a[x, y]$ requires the process of sampling.
- An ideal sampling system is defined as the image $a(x, y)$ multiplied by an ideal 2D impulse train $\delta(x, y)$:

$$
\begin{aligned}
b[m, n] &= a(x, y) \sum_{m=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \delta(x - mX_0, y - nY_0) \\
&= \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} a(mX_0, nY_0)\, \delta(x - mX_0, y - nY_0) \, .
\end{aligned}
$$

where X_0 and Y_0 are the sampling distance or intervals and δ is the Dirac delta function.

- If you were to sample in square shapes $X_0 = Y_0$ where you could think of each individual block a pixel

- Converting from a continuous 2D data $a(x, y)$ to its digital representation $a[x, y]$ requires the process of sampling.
- An ideal sampling system is defined as the image $a(x, y)$ multiplied by an ideal 2D impulse train $\delta(x, y)$:

$$b[m, n] = a(x, y) \sum_{m=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \delta(x - mX_0, y - nY_0)$$
$$= \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} a(mX_0, nY_0)\, \delta(x - mX_0, y - nY_0).$$

  where X_0 and Y_0 are the sampling distance or intervals and δ is the Dirac delta function.
- If you were to sample in square shapes $X_0 = Y_0$ where you could think of each individual block a pixel

Mathematical Fundamentals

- To reconstruct a continuous analog signal from its sampled version accurately, the sampling rate must be at least twice the highest frequency present in the signal.
- This ensures that there are enough samples taken per unit of time to capture all the details of the original waveform without introducing aliasing, which can cause distortion or artifacts in the reconstructed signal.

$$f_s \geq 2f_m$$

where f_s is the signal frequency, f_m is the maximum sample frequency.

> This is only a theoretical limit, not a practical one.

Mathematical Fundamentals

- To reconstruct a continuous analog signal from its sampled version accurately, the sampling rate must be at least twice the highest frequency present in the signal.
- This ensures that there are enough samples taken per unit of time to capture all the details of the original waveform without introducing aliasing, which can cause distortion or artifacts in the reconstructed signal.

$$f_s \geq 2f_m$$

where f_s is the signal frequency, f_m is the maximum sample frequency.

> This is only a theoretical limit, not a practical one.

Mathematical Fundamentals

- To reconstruct a continuous analog signal from its sampled version accurately, the sampling rate must be at least twice the highest frequency present in the signal.
- This ensures that there are enough samples taken per unit of time to capture all the details of the original waveform without introducing aliasing, which can cause distortion or artifacts in the reconstructed signal.

$$f_\mathrm{s} \geq 2f_\mathrm{m}$$

where f_s is the signal frequency, f_m is the maximum sample frequency.

> This is only a theoretical limit, not a practical one.

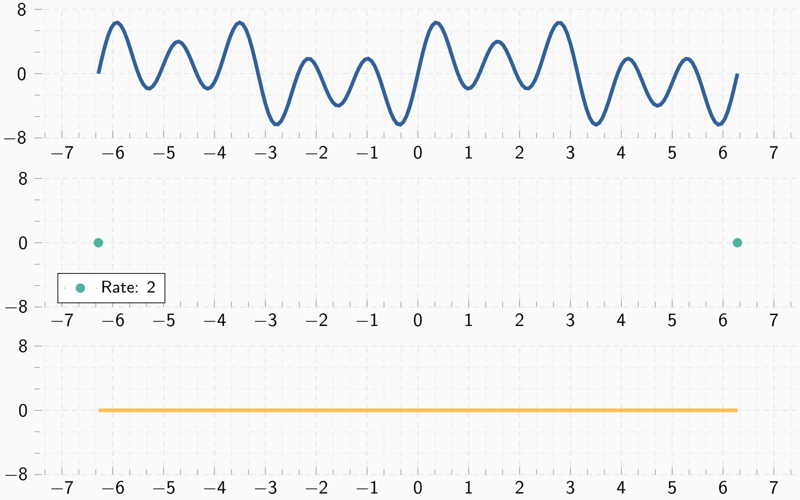

Figure 2: The effects of signal reconstruction on the sampling rate.

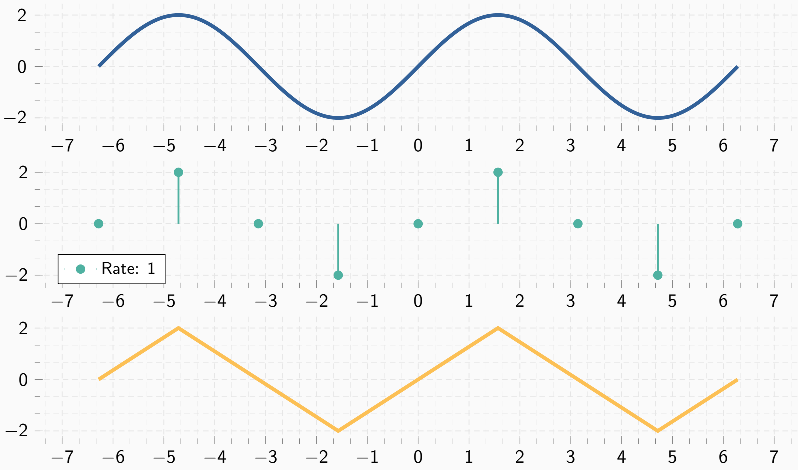

Figure 3: Reconstruction of the signal with 1 times the signal frequency.

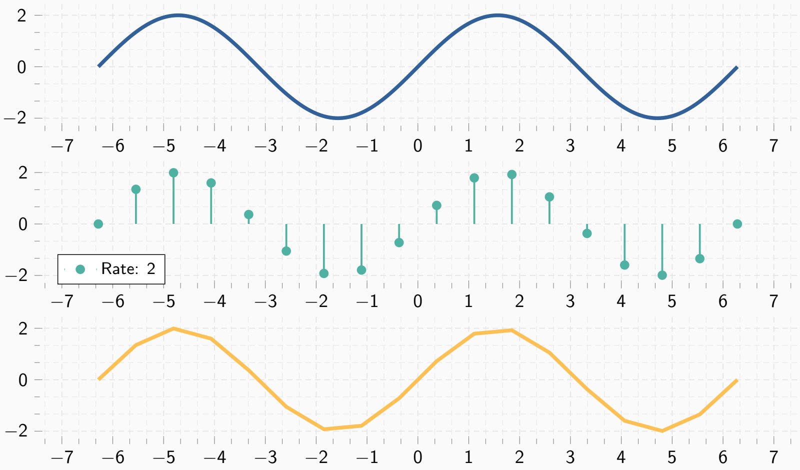

Figure 4: Reconstruction of the signal with 2 times the signal frequency.

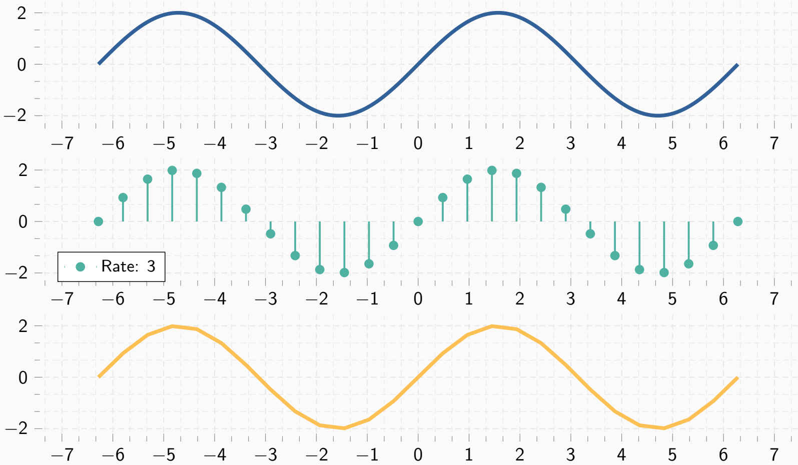

Figure 5: Reconstruction of the signal with 3 times the signal frequency.

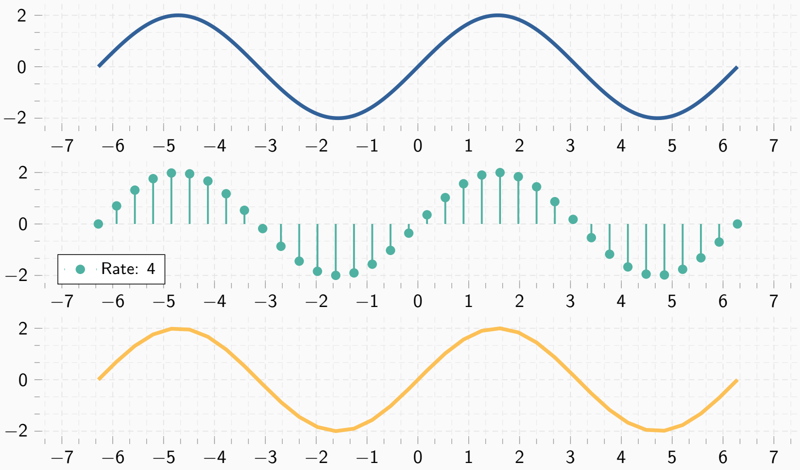

Figure 6: Reconstruction of the signal with 4 times the signal frequency.

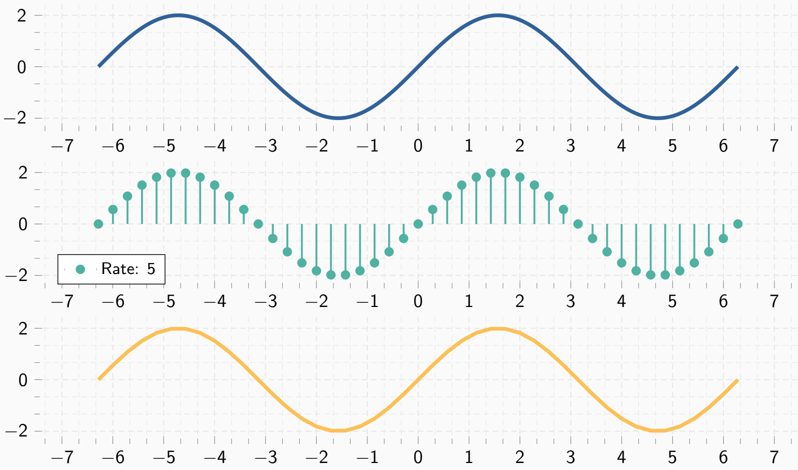

Figure 7: Reconstruction of the signal with 5 times the signal frequency.

Reconstruction of an Audio Signal

- In practice, doubling frequency is not enough recreate the signal.
- Approaching Nyquist frequency will create a siren like sound, and reaching exact frequency will record a pulse-wave approximation of a sine wave at an amplitude that will vary based on phase.
- Even at 4 times the sampling will only reconstruct a triangle wave and shifting the phase will create tonal distortion.
- For practical cases at least 6 times sampling rate is needed to accurately reconstruct the sine wave

Reconstruction of an Audio Signal

- In practice, doubling frequency is not enough recreate the signal.
- Approaching Nyquist frequency will create a siren like sound, and reaching exact frequency will record a pulse-wave approximation of a sine wave at an amplitude that will vary based on phase.
- Even at 4 times the sampling will only reconstruct a triangle wave and shifting the phase will create tonal distortion.
- For practical cases at least 6 times sampling rate is needed to accurately reconstruct the sine wave

Reconstruction of an Audio Signal

- In practice, doubling frequency is not enough recreate the signal.
- Approaching Nyquist frequency will create a siren like sound, and reaching exact frequency will record a pulse-wave approximation of a sine wave at an amplitude that will vary based on phase.
- Even at 4 times the sampling will only reconstruct a triangle wave and shifting the phase will create tonal distortion.
- For practical cases at least 6 times sampling rate is needed to accurately reconstruct the sine wave

Reconstruction of an Audio Signal

- In practice, doubling frequency is not enough recreate the signal.
- Approaching Nyquist frequency will create a siren like sound, and reaching exact frequency will record a pulse-wave approximation of a sine wave at an amplitude that will vary based on phase.
- Even at 4 times the sampling will only reconstruct a triangle wave and shifting the phase will create tonal distortion.
- For practical cases at least 6 times sampling rate is needed to accurately reconstruct the sine wave

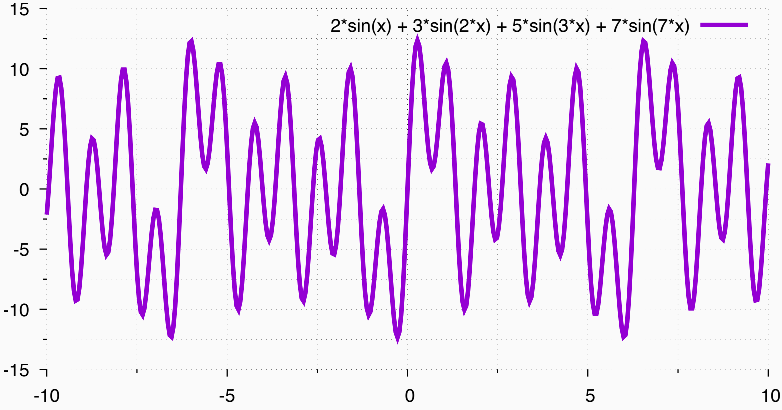

Figure 8: A sample signal with containing sample sine waves.

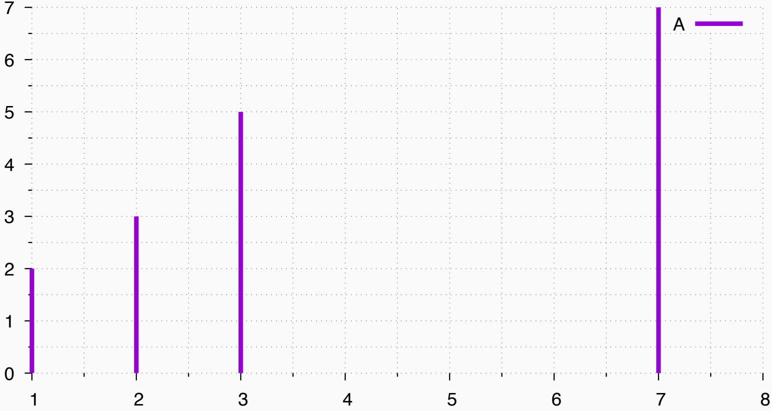

Figure 9: The FFT of the previous complex signal.

- Two (2) key problems arise when conducting spectral analysis of finite, discrete time series (not an infinite time series):

Aliasing we only resolve frequencies lower than the Nyquist frequency and frequencies higher than this get aliased to lower frequencies.

Spectral Leakage we assume that all waveforms stop and start at = 0 and =, but in the real world, many of these wave numbers may not complete a full integer number of cycles throughout the domain, causing spectral leakage to other wave numbers.

Mathematical Fundamentals

- Two (2) key problems arise when conducting spectral analysis of finite, discrete time series (not an infinite time series):

Aliasing we only resolve frequencies lower than the Nyquist frequency and frequencies higher than this get aliased to lower frequencies.

Spectral Leakage we assume that all waveforms stop and start at $= 0$ and $=$, but in the real world, many of these wave numbers may not complete a full integer number of cycles throughout the domain, causing spectral leakage to other wave numbers.

Mathematical Fundamentals

- Two (2) key problems arise when conducting spectral analysis of finite, discrete time series (not an infinite time series):

Aliasing we only resolve frequencies lower than the Nyquist frequency and frequencies higher than this get aliased to lower frequencies.

Spectral Leakage we assume that all waveforms stop and start at $= 0$ and $=$, but in the real world, many of these wave numbers may not complete a full integer number of cycles throughout the domain, causing spectral leakage to other wave numbers.

Aliasing

- If the initial samples are not sufficiently closely spaced to represent high-frequency components present in the underlying function, then the DFT values will be corrupted by aliasing.
- The solution is either to increase the sampling rate (if possible) or to pre-filter the signal in order to minimise its high-frequency spectral content.

Aliasing

- If the initial samples are not sufficiently closely spaced to represent high-frequency components present in the underlying function, then the DFT values will be corrupted by aliasing.
- The solution is either to increase the sampling rate (if possible) or to pre-filter the signal in order to minimise its high-frequency spectral content.

Leakage

- The continuous Fourier transform of a periodic waveform requires the integration to be performed over the interval $-\infty$ to $+\infty$ or over an integer number of cycles of the waveform.
- If we attempt to complete the DFT over a non-integer number of cycles of the input signal, then we might expect the transform to be corrupted in some way.

Leakage

- The continuous Fourier transform of a periodic waveform requires the integration to be performed over the interval $-\infty$ to $+\infty$ or over an integer number of cycles of the waveform.
- If we attempt to complete the DFT over a non-integer number of cycles of the input signal, then we might expect the transform to be corrupted in some way.

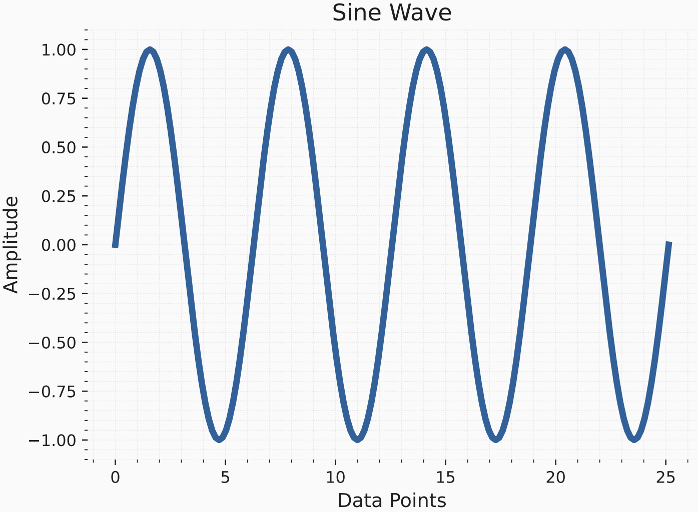

Figure 10: An example of a sine wave with four (4) complete cycles.

- Computing the discrete power spectrum gives:

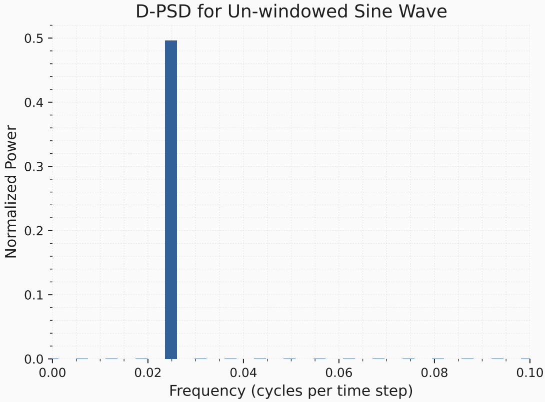

Figure 11: The PSD of an un-windowed sine wave.

- As expected, a single spectral peak corresponding to the frequency of our sine wave.
- Let's see what happens if we apply a window to our sine wave that cuts off the sine wave such that the sine function does not complete an integer number of cycles within the time domain.

Mathematical Fundamentals

- As expected, a single spectral peak corresponding to the frequency of our sine wave.
- Let's see what happens if we apply a window to our sine wave that cuts off the sine wave such that the sine function does not complete an integer number of cycles within the time domain.

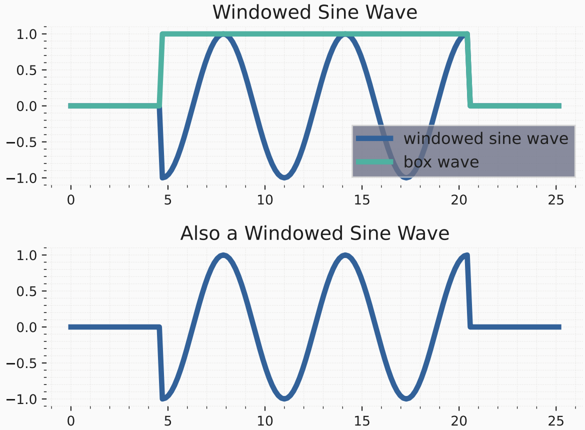

Figure 12: Windowed sine wave.

- To demonstrate what spectral leakage is, we will now compute the discrete power spectrum of the windowed sine wave to see what happens.

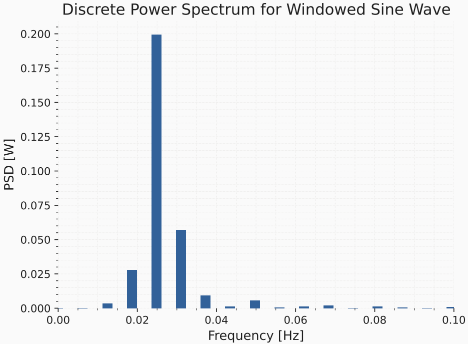

Figure 13: PSD of a windowed sine wave.

Example

Below is a signal with 1 Hz, Amplitude of 1 and 8 Sampling points.

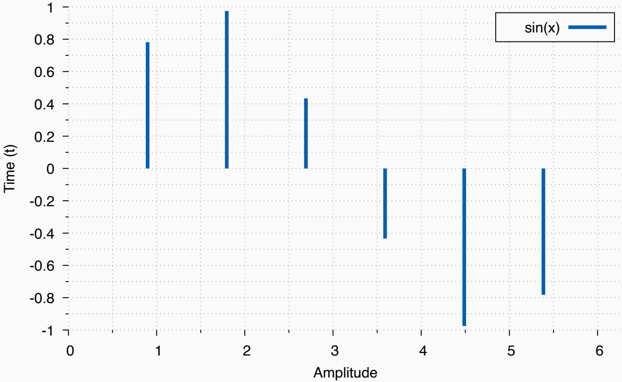

Figure 14: A Sampled Sine wave.

Solution

As it is a single sine function with 1 Hz, we expect a single value of 1 in the frequency domain (at 1 Hz).

The sampling points will sample the signal and retrieve the following data points as shown in the array below:

$$x_k = \left[0, 0.707, 1, 0.707, 0, -0.707, -1, -0.707\right]$$

Once we have these sampling points (x_n), we can turn our attention to the DFT formula:

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-(j\,2\pi k n)/\ N}$$

where X_k is the k^{th} frequency bin.

For the case of $x_0 = 0$ the exponential part is removed and we are left with $X_0 = 0$.
For the cases of X_1:

$$X_1 = \sum_{n=0}^{7} x_n \cdot e^{-(\mathbf{j}\,2\pi\,(1)\,n)/\ N}$$

$$= x \cdot \begin{bmatrix} 0 \\ e^{-(\mathbf{j}\,2\pi)/\ N} \\ e^{-(\mathbf{j}\,4\pi)/\ N} \\ e^{-(\mathbf{j}\,6\pi)/\ N} \\ e^{-(\mathbf{j}\,8\pi)/\ N} \\ e^{-(\mathbf{j}\,10\pi)/\ N} \\ e^{-(\mathbf{j}\,12\pi)/\ N} \\ e^{-(\mathbf{j}\,14\pi)/\ N} \end{bmatrix} = 0 - \mathbf{j}\,4 \quad \blacksquare$$

For the case of $x_0 = 0$ the exponential part is removed and we are left with $X_0 = 0$.
For the cases of X_2:

$$X_2 = \sum_{n=0}^{7} x_n \cdot e^{-(\mathbf{j}\, 2\pi\, (2)\, n)/\ N}$$

$$= x \cdot \begin{bmatrix} 0 \\ e^{-(\mathbf{j}\, 4\pi)/\ N} \\ e^{-(\mathbf{j}\, 8\pi)/\ N} \\ e^{-(\mathbf{j}\, 12\pi)/\ N} \\ e^{-(\mathbf{j}\, 16\pi)/\ N} \\ e^{-(\mathbf{j}\, 20\pi)/\ N} \\ e^{-(\mathbf{j}\, 24\pi)/\ N} \\ e^{-(\mathbf{j}\, 28\pi)/\ N} \end{bmatrix} = 0 \quad \blacksquare$$

Mathematical Fundamentals

For the case of $x_0 = 0$ the exponential part is removed and we are left with $X_0 = 0$.
For the cases of X_3:

$$X_3 = \sum_{n=0}^{7} x_n \cdot e^{-(\mathbf{j}\, 2\pi\,(3)\, n)/\ N}$$

$$= x \cdot \begin{bmatrix} 0 \\ e^{-(\mathbf{j}\, 6\pi)/\ N} \\ e^{-(\mathbf{j}\, 12\pi)/\ N} \\ e^{-(\mathbf{j}\, 18\pi)/\ N} \\ e^{-(\mathbf{j}\, 24\pi)/\ N} \\ e^{-(\mathbf{j}\, 30\pi)/\ N} \\ e^{-(\mathbf{j}\, 36\pi)/\ N} \\ e^{-(\mathbf{j}\, 42\pi)/\ N} \end{bmatrix} = 0 \quad \blacksquare$$

Mathematical Fundamentals

For the case of $x_0 = 0$ the exponential part is removed and we are left with $X_0 = 0$.
For the cases of X_4:

$$X_4 = \sum_{n=0}^{7} x_n \cdot e^{-(\mathbf{j}\,2\pi\,(4)\,n)/\ N}$$

$$= x \cdot \begin{bmatrix} 0 \\ e^{-(\mathbf{j}\,8\pi)/\ N} \\ e^{-(\mathbf{j}\,16\pi)/\ N} \\ e^{-(\mathbf{j}\,24\pi)/\ N} \\ e^{-(\mathbf{j}\,32\pi)/\ N} \\ e^{-(\mathbf{j}\,40\pi)/\ N} \\ e^{-(\mathbf{j}\,48\pi)/\ N} \\ e^{-(\mathbf{j}\,56\pi)/\ N} \end{bmatrix} = 0 \quad \blacksquare$$

Mathematical Fundamentals

For the case of $x_0 = 0$ the exponential part is removed and we are left with $X_0 = 0$.
For the cases of X_5:

$$X_5 = \sum_{n=0}^{7} x_n \cdot e^{-(\mathbf{j}\, 2\pi\,(5)\, n)/\ N}$$

$$= x \cdot \begin{bmatrix} 0 \\ e^{-(\mathbf{j}\, 10\pi)/\ N} \\ e^{-(\mathbf{j}\, 20\pi)/\ N} \\ e^{-(\mathbf{j}\, 30\pi)/\ N} \\ e^{-(\mathbf{j}\, 40\pi)/\ N} \\ e^{-(\mathbf{j}\, 50\pi)/\ N} \\ e^{-(\mathbf{j}\, 60\pi)/\ N} \\ e^{-(\mathbf{j}\, 70\pi)/\ N} \end{bmatrix} = 0 \quad \blacksquare$$

For the case of $x_0 = 0$ the exponential part is removed and we are left with $X_0 = 0$.
For the cases of X_6:

$$X_6 = \sum_{n=0}^{7} x_n \cdot e^{-(\mathbf{j}\, 2\pi\,(6)\,n)/\ N}$$

$$= x \cdot \begin{bmatrix} 0 \\ e^{-(\mathbf{j}\, 12\pi)/\ N} \\ e^{-(\mathbf{j}\, 24\pi)/\ N} \\ e^{-(\mathbf{j}\, 36\pi)/\ N} \\ e^{-(\mathbf{j}\, 48\pi)/\ N} \\ e^{-(\mathbf{j}\, 60\pi)/\ N} \\ e^{-(\mathbf{j}\, 72\pi)/\ N} \\ e^{-(\mathbf{j}\, 84\pi)/\ N} \end{bmatrix} = 0 \quad \blacksquare$$

For the case of $x_0 = 0$ the exponential part is removed and we are left with $X_0 = 0$.
For the cases of X_7:

$$X_7 = \sum_{n=0}^{7} x_n \cdot e^{-(\mathbf{j}\,2\pi\,(7)\,n)/\ N}$$

$$= x \cdot \begin{bmatrix} 0 \\ e^{-(\mathbf{j}\,14\pi)/\ N} \\ e^{-(\mathbf{j}\,28\pi)/\ N} \\ e^{-(\mathbf{j}\,42\pi)/\ N} \\ e^{-(\mathbf{j}\,56\pi)/\ N} \\ e^{-(\mathbf{j}\,70\pi)/\ N} \\ e^{-(\mathbf{j}\,84\pi)/\ N} \\ e^{-(\mathbf{j}\,98\pi)/\ N} \end{bmatrix} = 0 + \mathbf{j}\,4 \quad \blacksquare$$

- Therefore the values are of the transform are:

$$X_k = \left[0, 0 - \mathbf{j}\,4, 0, 0, 0, 0, 0, 0 + \mathbf{j}\,4\right]$$

- We can see only the first and the seventh bins have values other than zero.
- Calculating the magnitudes of the bins, we arrive at 4.

$$|X_k| = \left[0, 4, 0, 0, 0, 0, 0, 4\right]$$

- The frequency resolution of the plot is the sampling frequency divided by the number of samples (i.e., f S/N).
- This means we can get values for every integer frequency values.

- Therefore the values are of the transform are:

$$X_{\mathrm{k}} = \left[0, 0 - \mathbf{j}\,4, 0, 0, 0, 0, 0, 0 + \mathbf{j}\,4\right]$$

- We can see only the first and the seventh bins have values other than zero.
- Calculating the magnitudes of the bins, we arrive at 4.

$$|X_{\mathrm{k}}| = \left[0, 4, 0, 0, 0, 0, 0, 4\right]$$

- The frequency resolution of the plot is the sampling frequency divided by the number of samples (i.e., f S/N).
- This means we can get values for every integer frequency values.

- Therefore the values are of the transform are:

$$X_k = \left[0, 0 - \mathbf{j}\,4, 0, 0, 0, 0, 0, 0 + \mathbf{j}\,4\right]$$

- We can see only the first and the seventh bins have values other than zero.
- Calculating the magnitudes of the bins, we arrive at 4.

$$|X_k| = \left[0, 4, 0, 0, 0, 0, 0, 4\right]$$

- The frequency resolution of the plot is the sampling frequency divided by the number of samples (i.e., f S/N).
- This means we can get values for every integer frequency values.

- Therefore the values are of the transform are:

$$X_k = \left[0, 0 - \mathbf{j}\,4, 0, 0, 0, 0, 0, 0 + \mathbf{j}\,4\right]$$

- We can see only the first and the seventh bins have values other than zero.
- Calculating the magnitudes of the bins, we arrive at 4.

$$|X_k| = \left[0, 4, 0, 0, 0, 0, 0, 4\right]$$

- The frequency resolution of the plot is the sampling frequency divided by the number of samples (i.e., f S/N).
- This means we can get values for every integer frequency values.

Mathematical Fundamentals

- Therefore the values are of the transform are:

$$X_{\mathsf{k}} = \left[0, 0 - \mathbf{j}\,4, 0, 0, 0, 0, 0, 0 + \mathbf{j}\,4\right]$$

- We can see only the first and the seventh bins have values other than zero.
- Calculating the magnitudes of the bins, we arrive at 4.

$$|X_{\mathsf{k}}| = \left[0, 4, 0, 0, 0, 0, 0, 4\right]$$

- The frequency resolution of the plot is the sampling frequency divided by the number of samples (i.e., f S/N).
- This means we can get values for every integer frequency values.

Solution

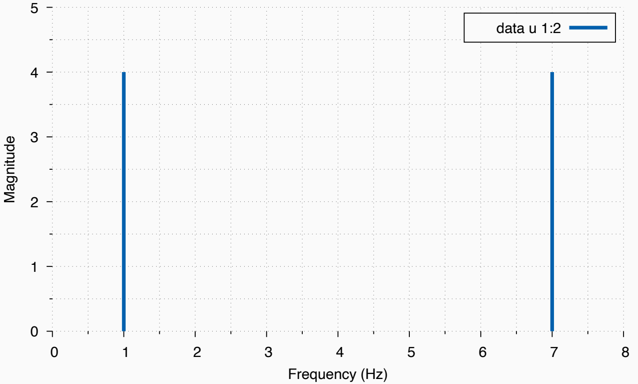

Figure 15: Sampled dataset of the original signal. There is still another step.

Solution

- We can see we get a value for the first frequency bin (1 Hz) and it makes sense.
- The reason we get a frequency bin is due to the plot being a two-sided frequency plot where it shows the energy in both the positive and negative domains.
- The negative frequencies are always complex conjugate to the positive frequencies, so there is no additional information in the negative frequencies.

Solution

- We can see we get a value for the first frequency bin (1 Hz) and it makes sense.
- The reason we get a frequency bin is due to the plot being a two-sided frequency plot where it shows the energy in both the positive and negative domains.
- The negative frequencies are always complex conjugate to the positive frequencies, so there is no additional information in the negative frequencies.

Solution

- We can see we get a value for the first frequency bin (1 Hz) and it makes sense.
- The reason we get a frequency bin is due to the plot being a two-sided frequency plot where it shows the energy in both the positive and negative domains.
- The negative frequencies are always complex conjugate to the positive frequencies, so there is no additional information in the negative frequencies.

Solution

- Therefore, to convert from a two-sided spectrum to a single-sided spectrum, discard the second half of the array and multiply every point except for DC by two.
- The last operation is to divide the magnitudes of the lower frequencies by the number of samples used in deriving these bins

$$|X_k| = \left[0, 8, 0, 0\right] \rightarrow |X_k| / N = \left[0, 1, 0, 0\right] \quad \blacksquare$$

Solution

- Therefore, to convert from a two-sided spectrum to a single-sided spectrum, discard the second half of the array and multiply every point except for DC by two.
- The last operation is to divide the magnitudes of the lower frequencies by the number of samples used in deriving these bins

$$|X_{\mathrm{k}}| = \left[0, 8, 0, 0\right] \to |X_{\mathrm{k}}| / N = \left[0, 1, 0, 0\right] \quad \blacksquare$$

Parseval's Theorem

- The sum (or integral) of the square of a function is equal to the sum (or integral) of the square of its transform.
- For continuous signals:

$$\int_{-\infty}^{\infty} |f(t)|^2 \, \mathrm{d}t = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2 \, \mathrm{d}\omega = \int_{-\infty}^{\infty} |F(2\pi f)|^2$$

- This signal energy is not to be confused with physical energy.

Parseval's Theorem

- The sum (or integral) of the square of a function is equal to the sum (or integral) of the square of its transform.
- For continuous signals:

$$\int_{-\infty}^{\infty} |f(t)|^2 \, \mathrm{d}t = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2 \, \mathrm{d}\omega = \int_{-\infty}^{\infty} |F(2\pi f)|^2$$

- This signal energy is not to be confused with physical energy.

Parseval's Theorem

- The sum (or integral) of the square of a function is equal to the sum (or integral) of the square of its transform.
- For continuous signals:

$$\int_{-\infty}^{\infty} |f(t)|^2 \, \mathrm{d}t = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2 \, \mathrm{d}\omega = \int_{-\infty}^{\infty} |F(2\pi f)|^2$$

- This signal energy is not to be confused with physical energy.

Average Value (μ)

- Defined as the sample mean of a given region.
- The equation is defined as below (it is also known as expected value):

$$\mu = \frac{1}{N} \sum_{i=0}^{N} x_{\mathrm{N}}$$

Standard Deviation (σ)

- The standard deviation is a measure of the amount of variation of the values of a variable about its mean (μ):

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2}$$

Mathematical Fundamentals

Average Value (μ)

- Defined as the sample mean of a given region.
- The equation is defined as below (it is also known as expected value):

$$\mu = \frac{1}{N}\sum_{i=0}^{N} x_N$$

Standard Deviation (σ)

- The standard deviation is a measure of the amount of variation of the values of a variable about its mean (μ):

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

Mathematical Fundamentals

Average Value (μ)

- Defined as the sample mean of a given region.
- The equation is defined as below (it is also known as expected value):

$$\mu = \frac{1}{N} \sum_{i=0}^{N} x_N$$

Standard Deviation (σ)

- The standard deviation is a measure of the amount of variation of the values of a variable about its mean (μ):

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2}$$

Mode

- The mode is the value appears most often in a set of data values.

  i.e., in an data pool of:

$$X = [1, 2, 2, 3, 4, 7, 9]$$

- The mode of is 2 as it is the most frequent value of the data set.
- whereas in:

$$X = [2, 4, 9, 6, 4, 6, 6, 2, 8, 2]$$

  the mode is 2, 6 as there are two values with same frequency.

Mode

- The mode is the value appears most often in a set of data values. i.e., in an data pool of:

$$X = [1, 2, 2, 3, 4, 7, 9]$$

- The mode of is 2 as it is the most frequent value of the data set.
- whereas in:

$$X = [2, 4, 9, 6, 4, 6, 6, 2, 8, 2]$$

the mode is 2, 6 as there are two values with same frequency.

Mode

- The mode is the value appears most often in a set of data values.
  i.e., in an data pool of:

$$X = [1, 2, 2, 3, 4, 7, 9]$$

- The mode of is 2 as it is the most frequent value of the data set.
- whereas in:

$$X = [2, 4, 9, 6, 4, 6, 6, 2, 8, 2]$$

  the mode is 2, 6 as there are two values with same frequency.

Mode

- The mode is the value appears most often in a set of data values.

  i.e., in an data pool of:

$$X = [1, 2, 2, 3, 4, 7, 9]$$

- The mode of is 2 as it is the most frequent value of the data set.
- whereas in:

$$X = [2, 4, 9, 6, 4, 6, 6, 2, 8, 2]$$

the mode is 2, 6 as there are two values with same frequency.

Mode

- The mode is the value appears most often in a set of data values. i.e., in an data pool of:

$$X = [1, 2, 2, 3, 4, 7, 9]$$

- The mode of is 2 as it is the most frequent value of the data set.
- whereas in:

$$X = [2, 4, 9, 6, 4, 6, 6, 2, 8, 2]$$

the mode is 2, 6 as there are two values with same frequency.

Median

- The median is the middle value separating the greater and lesser halves of the data set.
- For a ordered data set X with n elements,
  - if n is odd, med (x) = x (n+1)/2,
  - if n is even, med (x) = x (n/2) + x ((n/2)+1)/2
- i.e., in a ordered data set of:

$$X = [1, 2, 2, 3, 4, 7, 9] ,$$

- the median is 3 and in:

$$Y = [1, 2, 3, 4, 5, 6, 8, 9] ,$$

the median is 4.5.

Median

- The median is the middle value separating the greater and lesser halves of the data set.
- For a ordered data set X with n elements,
  - if n is odd, med (x) = x (n+1)/2,
  - if n is even, med (x) = x (n/2) + x ((n/2)+1)/2
- i.e., in a ordered data set of:

$$X = [1, 2, 2, 3, 4, 7, 9],$$

- the median is 3 and in:

$$Y = [1, 2, 3, 4, 5, 6, 8, 9],$$

the median is 4.5.

Median

- The median is the middle value separating the greater and lesser halves of the data set.
- For a ordered data set X with n elements,
  - if n is odd, med (x) = x (n+1)/2,
  - if n is even, med (x) = x (n/2) + x ((n/2)+1)/2
- i.e., in a ordered data set of:

$$X = [1, 2, 2, 3, 4, 7, 9]\,,$$

- the median is 3 and in:

$$Y = [1, 2, 3, 4, 5, 6, 8, 9]\,,$$

the median is 4.5.

Median

- The median is the middle value separating the greater and lesser halves of the data set.
- For a ordered data set X with n elements,
  - if n is odd, med (x) = x (n+1)/2,
  - if n is even, med (x) = x (n/2) + x ((n/2)+1)/2
- i.e., in a ordered data set of:

$$X = [1, 2, 2, 3, 4, 7, 9],$$

- the median is 3 and in:

$$Y = [1, 2, 3, 4, 5, 6, 8, 9],$$

the median is 4.5.

Median

- The median is the middle value separating the greater and lesser halves of the data set.
- For a ordered data set X with n elements,
  - if n is odd, med (x) = x (n+1)/2,
  - if n is even, med (x) = x (n/2) + x ((n/2)+1)/2
- i.e., in a ordered data set of:

$$X = [1, 2, 2, 3, 4, 7, 9],$$

- the median is 3 and in:

$$Y = [1, 2, 3, 4, 5, 6, 8, 9],$$

the median is 4.5.

Median

- The median is the middle value separating the greater and lesser halves of the data set.
- For a ordered data set X with n elements,
  - if n is odd, med (x) = x (n+1)/2,
  - if n is even, med (x) = x (n/2) + x ((n/2)+1)/2
- i.e., in a ordered data set of:

$$X = [1, 2, 2, 3, 4, 7, 9]\,,$$

- the median is 3 and in:

$$Y = [1, 2, 3, 4, 5, 6, 8, 9]\,,$$

the median is 4.5.

Median

- The median is the middle value separating the greater and lesser halves of the data set.
- For a ordered data set X with n elements,
  - if n is odd, med (x) = x (n+1)/2,
  - if n is even, med (x) = x (n/2) + x ((n/2)+1)/2
- i.e., in a ordered data set of:

$$X = [1, 2, 2, 3, 4, 7, 9]\,,$$

- the median is 3 and in:

$$Y = [1, 2, 3, 4, 5, 6, 8, 9]\,,$$

  the median is 4.5.

Mathematical Fundamentals

- The signal-to-noise ratio (SNR) can have several definitions depending on the study field.
- Noise is characterised by its standard deviation, σ.
- The characterisation of the signal can differ.
- If the signal is known to lie between two boundaries, $a_{\text{min}} \leq a \leq a_{\text{max}}$, then the SNR is defined as:

$$\text{SNR} = 20 \log_{10} \left(\frac{a_{\text{max}} - a_{\text{min}}}{s_{\text{n}}} \right) \text{ dB}.$$

- If the signal is not bounded but has a statistical distribution then two other definitions are known:

$$\text{SNR} = 20 \log_{10} \left(\frac{\mu}{\sigma} \right) \text{ dB}.$$

Mathematical Fundamentals

- The signal-to-noise ratio (SNR) can have several definitions depending on the study field.
- Noise is characterised by its standard deviation, σ.
- The characterisation of the signal can differ.
- If the signal is known to lie between two boundaries, $a_{\text{min}} \leq a \leq a_{\text{max}}$, then the SNR is defined as:

$$\text{SNR} = 20\log_{10}\left(\frac{a_{\text{max}} - a_{\text{min}}}{s_{\text{n}}}\right)\ \text{dB}.$$

- If the signal is not bounded but has a statistical distribution then two other definitions are known:

$$\text{SNR} = 20\log_{10}\left(\frac{\mu}{\sigma}\right)\ \text{dB}.$$

Mathematical Fundamentals

- The signal-to-noise ratio (SNR) can have several definitions depending on the study field.
- Noise is characterised by its standard deviation, σ.
- The characterisation of the signal can differ.
- If the signal is known to lie between two boundaries, $a_{\mathrm{min}} \leq a \leq a_{\mathrm{max}}$, then the SNR is defined as:

$$\mathrm{SNR} = 20 \log_{10} \left(\frac{a_{\mathrm{max}} - a_{\mathrm{min}}}{s_{\mathrm{n}}} \right) \ \mathrm{dB}.$$

- If the signal is not bounded but has a statistical distribution then two other definitions are known:

$$\mathrm{SNR} = 20 \log_{10} \left(\frac{\mu}{\sigma} \right) \ \mathrm{dB}.$$

- The signal-to-noise ratio (SNR) can have several definitions depending on the study field.
- Noise is characterised by its standard deviation, σ.
- The characterisation of the signal can differ.
- If the signal is known to lie between two boundaries, $a_{\text{min}} \leq a \leq a_{\text{max}}$, then the SNR is defined as:

$$\text{SNR} = 20 \log_{10} \left(\frac{a_{\text{max}} - a_{\text{min}}}{s_{\text{n}}} \right) \text{ dB.}$$

- If the signal is not bounded but has a statistical distribution then two other definitions are known:

$$\text{SNR} = 20 \log_{10} \left(\frac{\mu}{\sigma} \right) \text{ dB.}$$

- The signal-to-noise ratio (SNR) can have several definitions depending on the study field.
- Noise is characterised by its standard deviation, σ.
- The characterisation of the signal can differ.
- If the signal is known to lie between two boundaries, $a_{\text{min}} \leq a \leq a_{\text{max}}$, then the SNR is defined as:

$$\text{SNR} = 20 \log_{10} \left(\frac{a_{\text{max}} - a_{\text{min}}}{s_{\text{n}}} \right) \text{ dB}.$$

- If the signal is not bounded but has a statistical distribution then two other definitions are known:

$$\text{SNR} = 20 \log_{10} \left(\frac{\mu}{\sigma} \right) \text{ dB}.$$

Mathematical Fundamentals

- In 1948, Claude Shannon published a paper called A Mathematical Theory of Communication.
- This paper heralded a transformation in our understanding of information.
- Before Shannon's paper, information had been viewed as a kind of poorly defined ethereal concept.
- But after Shannon's paper, it became apparent that information is a well-defined and, above all, measurable quantity.

Mathematical Fundamentals

- In 1948, Claude Shannon published a paper called A Mathematical Theory of Communication.
- This paper heralded a transformation in our understanding of information.
- Before Shannon's paper, information had been viewed as a kind of poorly defined ethereal concept.
- But after Shannon's paper, it became apparent that information is a well-defined and, above all, measurable quantity.

Mathematical Fundamentals

- In 1948, Claude Shannon published a paper called A Mathematical Theory of Communication.
- This paper heralded a transformation in our understanding of information.
- Before Shannon's paper, information had been viewed as a kind of poorly defined ethereal concept.
- But after Shannon's paper, it became apparent that information is a well-defined and, above all, measurable quantity.

Mathematical Fundamentals

- In 1948, Claude Shannon published a paper called A Mathematical Theory of Communication.
- This paper heralded a transformation in our understanding of information.
- Before Shannon's paper, information had been viewed as a kind of poorly defined ethereal concept.
- But after Shannon's paper, it became apparent that information is a well-defined and, above all, measurable quantity.

Mathematical Fundamentals

- Information theory defines definite, unbreachable limits on precisely how much information can be communicated between any two components of any system, whether this system is man-made or natural.
- The basic laws of information can be summarised as follows.

1. there is a definite upper limit, the channel capacity, to the amount of information that can be communicated through that channel,
2. this limit shrinks the amount of noise in the channel increases,
3. this limit can very nearly be reached by judicious packaging, or encoding, of data.

Mathematical Fundamentals

- Information theory defines definite, unbreachable limits on precisely how much information can be communicated between any two components of any system, whether this system is man-made or natural.
- The basic laws of information can be summarised as follows.

1. there is a definite upper limit, the channel capacity, to the amount of information that can be communicated through that channel,
2. this limit shrinks the amount of noise in the channel increases,
3. this limit can very nearly be reached by judicious packaging, or encoding, of data.

Mathematical Fundamentals

- Information theory defines definite, unbreachable limits on precisely how much information can be communicated between any two components of any system, whether this system is man-made or natural.
- The basic laws of information can be summarised as follows.

1. there is a definite upper limit, the channel capacity, to the amount of information that can be communicated through that channel,
2. this limit shrinks the amount of noise in the channel increases,
3. this limit can very nearly be reached by judicious packaging, or encoding, of data.

Mathematical Fundamentals

- Information theory defines definite, unbreachable limits on precisely how much information can be communicated between any two components of any system, whether this system is man-made or natural.
- The basic laws of information can be summarised as follows.

1. there is a definite upper limit, the channel capacity, to the amount of information that can be communicated through that channel,
2. this limit shrinks the amount of noise in the channel increases,
3. this limit can very nearly be reached by judicious packaging, or encoding, of data.

Mathematical Fundamentals

- Information theory defines definite, unbreachable limits on precisely how much information can be communicated between any two components of any system, whether this system is man-made or natural.
- The basic laws of information can be summarised as follows.

1. there is a definite upper limit, the channel capacity, to the amount of information that can be communicated through that channel,
2. this limit shrinks the amount of noise in the channel increases,
3. this limit can very nearly be reached by judicious packaging, or encoding, of data.

Bits are Not Binary Digits

- The word bit is derived from binary digit,
  - but a bit and a binary digit are fundamentally different types of quantities.
- A binary digit is the value of a binary variable, whereas a bit is an amount of information.

Bits are Not Binary Digits

- The word bit is derived from binary digit,
  - but a bit and a binary digit are fundamentally different types of quantities.
- A binary digit is the value of a binary variable, whereas a bit is an amount of information.

Bits are Not Binary Digits

- The word bit is derived from binary digit,
  - but a bit and a binary digit are fundamentally different types of quantities.
- A binary digit is the value of a binary variable, whereas a bit is an amount of information.

- Consider a coin which lands heads up 90% of the time:

$$p(x_{\mathrm{h}}) = 0.9.$$

- When this coin is flipped, we expect it to land heads up ($x = x_{\mathrm{h}}$),
- When it does, we are less surprised than when it lands tails ($x = x_{\mathrm{t}}$).

> The more improbable a particular outcome is, the more surprised we are to observe it.

- If we use $\log_2$ then the Shannon information or surprisal of each outcome is measured in bits.

$$\text{Shannon Information} = \log_2 \frac{1}{p(x_{\mathrm{h}})}$$

Mathematical Fundamentals

- Consider a coin which lands heads up 90% of the time:

$$p\left(x_{\mathrm{h}}\right) = 0.9.$$

- When this coin is flipped, we expect it to land heads up ($x = x_{\mathrm{h}}$),
- When it does, we are less surprised than when it lands tails ($x = x_{\mathrm{t}}$).

> The more improbable a particular outcome is, the more surprised we are to observe it.

- If we use $\log_2$ then the Shannon information or surprisal of each outcome is measured in bits.

$$\text{Shannon Information} = \log_2 \frac{1}{p\left(x_{\mathrm{h}}\right)}$$

Mathematical Fundamentals

- Consider a coin which lands heads up 90% of the time:

$$p(x_{\mathrm{h}}) = 0.9.$$

- When this coin is flipped, we expect it to land heads up ($x = x_{\mathrm{h}}$),
- When it does, we are less surprised than when it lands tails ($x = x_{\mathrm{t}}$).

> The more improbable a particular outcome is, the more surprised we are to observe it.

- If we use $\log_2$ then the Shannon information or surprisal of each outcome is measured in bits.

$$\text{Shannon Information} = \log_2 \frac{1}{p(x_{\mathrm{h}})}$$

Mathematical Fundamentals

- Consider a coin which lands heads up 90% of the time:

$$p\left(x_{\mathrm{h}}\right) = 0.9.$$

- When this coin is flipped, we expect it to land heads up ($x = x_{\mathrm{h}}$),
- When it does, we are less surprised than when it lands tails ($x = x_{\mathrm{t}}$).

> The more improbable a particular outcome is, the more surprised we are to observe it.

- If we use $\log_2$ then the Shannon information or surprisal of each outcome is measured in bits.

$$\text{Shannon Information} = \log_2 \frac{1}{p\left(x_{\mathrm{h}}\right)}$$

Entropy is Average Shannon Information

- We can represent the outcome of a coin flip as the random variable x, such that a head is $x = x_{\mathrm{h}}$ and a tail is $x = x_{\mathrm{t}}$.
- In practice, we are not usually interested in the surprise of a particular value of a random variable, but we are interested in how much surprise, on average, is associated with the entire set of possible values.
- The average surprise of a variable x is defined by its probability distribution $p(x)$, and is called the entropy of $p(x)$, represented as $H(x)$.

Entropy is Average Shannon Information

- We can represent the outcome of a coin flip as the random variable x, such that a head is $x = x_{\mathrm{h}}$ and a tail is $x = x_{\mathrm{t}}$.
- In practice, we are not usually interested in the surprise of a particular value of a random variable, but we are interested in how much surprise, on average, is associated with the entire set of possible values.
- The average surprise of a variable x is defined by its probability distribution $p(x)$, and is called the entropy of $p(x)$, represented as $H(x)$.

Entropy is Average Shannon Information

- We can represent the outcome of a coin flip as the random variable x, such that a head is $x = x_{\mathrm{h}}$ and a tail is $x = x_{\mathrm{t}}$.
- In practice, we are not usually interested in the surprise of a particular value of a random variable, but we are interested in how much surprise, on average, is associated with the entire set of possible values.
- The average surprise of a variable x is defined by its probability distribution $p(x)$, and is called the entropy of $p(x)$, represented as $H(x)$.

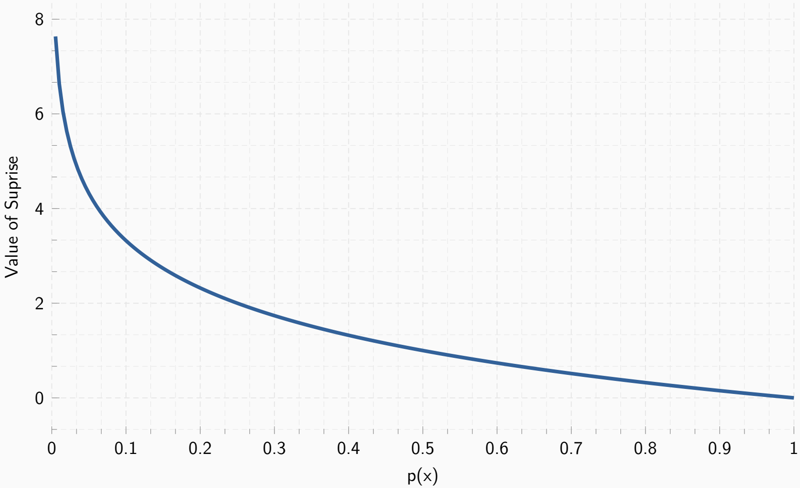

Figure 16: The quantifiable surprise with respect to increasing probability.

Entropy of a Fair Coin

- If a coin is fair or unbiased then:

$$p_{x_h} = p_{x_t} = 0.5$$

- the Shannon information gained when a head or a tail is observed is:

$$\log 1/0.5 = 1 \text{ bit}$$

- The average Shannon information gained after each coin flip is also 1 bit.
- Because entropy is defined as average Shannon information, the entropy of a fair coin is $H(x) = 1$ bit.

Entropy of a Fair Coin

- If a coin is fair or unbiased then:

$$p_{x_h} = p_{x_t} = 0.5$$

- the Shannon information gained when a head or a tail is observed is:

$$\log 1/0.5 = 1 \ \text{bit}$$

- The average Shannon information gained after each coin flip is also 1 bit.
- Because entropy is defined as average Shannon information, the entropy of a fair coin is $H(x) = 1$ bit.

Entropy of a Fair Coin

- If a coin is fair or unbiased then:

$$p_{x_h} = p_{x_t} = 0.5$$

- the Shannon information gained when a head or a tail is observed is:

$$\log 1/0.5 = 1 \text{ bit}$$

- The average Shannon information gained after each coin flip is also 1 bit.
- Because entropy is defined as average Shannon information, the entropy of a fair coin is $H(x) = 1$ bit.

Entropy of a Fair Coin

- If a coin is fair or unbiased then:

$$p_{x_h} = p_{x_t} = 0.5$$

- the Shannon information gained when a head or a tail is observed is:

$$\log 1/0.5 = 1 \text{ bit}$$

- The average Shannon information gained after each coin flip is also 1 bit.
- Because entropy is defined as average Shannon information, the entropy of a fair coin is $H(x) = 1$ bit.

Entropy of an Unfair Coin

- If a coin is biased such that the probability of a head is $p(x_{\mathrm{h}}) = 0.9$.
  - it is easy to predict the result of each coin flip (i.e. with 90% accuracy if we predict a head for each flip)
- If the outcome is a head then the amount of Shannon information gained is $\log(1/0.9) = 0.15$ bits.
- But if the outcome is a tail then the amount of Shannon information gained is $\log(1/0.1) = 3.32$ bits.
- Notice that more information is associated with the more surprising outcome.

Entropy of an Unfair Coin

- If a coin is biased such that the probability of a head is $p(x_{\mathrm{h}}) = 0.9$.
  - it is easy to predict the result of each coin flip (i.e. with 90% accuracy if we predict a head for each flip)
- If the outcome is a head then the amount of Shannon information gained is $\log(1/0.9) = 0.15$ bits.
- But if the outcome is a tail then the amount of Shannon information gained is $\log(1/0.1) = 3.32$ bits.
- Notice that more information is associated with the more surprising outcome.

Entropy of an Unfair Coin

- If a coin is biased such that the probability of a head is $p(x_{\mathrm{h}}) = 0.9$.
  - it is easy to predict the result of each coin flip (i.e. with 90% accuracy if we predict a head for each flip)
- If the outcome is a head then the amount of Shannon information gained is $\log(1/0.9) = 0.15$ bits.
- But if the outcome is a tail then the amount of Shannon information gained is $\log(1/0.1) = 3.32$ bits.
- Notice that more information is associated with the more surprising outcome.

Entropy of an Unfair Coin

- If a coin is biased such that the probability of a head is $p(x_{\mathrm{h}}) = 0.9$.
  - it is easy to predict the result of each coin flip (i.e. with 90% accuracy if we predict a head for each flip)
- If the outcome is a head then the amount of Shannon information gained is $\log(1/0.9) = 0.15$ bits.
- But if the outcome is a tail then the amount of Shannon information gained is $\log(1/0.1) = 3.32$ bits.
- Notice that more information is associated with the more surprising outcome.

Entropy of an Unfair Coin

- If a coin is biased such that the probability of a head is $p(x_{\mathrm{h}}) = 0.9$.
  - it is easy to predict the result of each coin flip (i.e. with 90% accuracy if we predict a head for each flip)
- If the outcome is a head then the amount of Shannon information gained is $\log(1/0.9) = 0.15$ bits.
- But if the outcome is a tail then the amount of Shannon information gained is $\log(1/0.1) = 3.32$ bits.
- Notice that more information is associated with the more surprising outcome.

Entropy of an Unfair Coin

- Given that the proportion of flips that yield a head is p(xh), and that the proportion of flips that yield a tail is p(xt) (where p(xh) + p(xt) = 1), the average surprise is

$$H(x) = p(x_h) \log \frac{1}{p(x_h)} + p(x_t) \log \frac{1}{p(x_t)},$$

- Which comes to 0.469bits.
- If we define a tail as $x_1 = x_t$ and a head as $x_2 = x_h$ then the above equation is written as:

$$H(x) = \sum_{i=1}^{2} p(x_i) \log \frac{1}{p(x_i)} \text{ bits.}$$

Entropy of an Unfair Coin

- Given that the proportion of flips that yield a head is p(xh), and that the proportion of flips that yield a tail is p(xt) (where p(xh) + p(xt) = 1), the average surprise is

$$H(x) = p(x_{\mathrm{h}}) \log \frac{1}{p(x_{\mathrm{h}})} + p(x_{\mathrm{t}}) \log \frac{1}{p(x_{\mathrm{t}})},$$

- Which comes to 0.469bits.
- If we define a tail as $x_1 = x_{\mathrm{t}}$ and a head as $x_2 = x_{\mathrm{h}}$ then the above equation is written as:

$$H(x) = \sum_{i=1}^{2} p(x_{\mathrm{i}}) \log \frac{1}{p(x_{\mathrm{i}})} \text{ bits.}$$

Entropy of an Unfair Coin

- Given that the proportion of flips that yield a head is p(xh), and that the proportion of flips that yield a tail is p(xt) (where p(xh) + p(xt) = 1), the average surprise is

$$H(x) = p(x_{\mathrm{h}}) \log \frac{1}{p(x_{\mathrm{h}})} + p(x_{\mathrm{t}}) \log \frac{1}{p(x_{\mathrm{t}})},$$

- Which comes to 0.469bits.
- If we define a tail as $x_1 = x_{\mathrm{t}}$ and a head as $x_2 = x_{\mathrm{h}}$ then the above equation is written as:

$$H(x) = \sum_{i=1}^{2} p(x_{\mathrm{i}}) \log \frac{1}{p(x_{\mathrm{i}})} \text{ bits.}$$

- More generally, a random variable x with a probability distribution p(x) = p(x1), . . . , p(xm) has an entropy of

$$H(x) = \sum_{i=1}^{m} p(x_i) \log \frac{1}{p(x_i)} \text{ bits.}$$

Mathematical Fundamentals

- Entropy is a measure of uncertainty.
- When our uncertainty is reduced, we gain information,
  - so information and entropy are two sides of the same coin.
- However, information has a rather subtle interpretation, which can easily lead to confusion.
- Average information shares the same definition as entropy,
  - but whether we call a given quantity information or entropy depends on whether it is being given to us or taken away.
- For example, if a variable has high entropy the initial uncertainty of the variable is large and is, by definition, exactly equal to its entropy.
- If we are told the variable value, on average, we have been given information equal to the uncertainty (entropy) we had about its value.
- Thus, receiving an amount of information is equivalent to having exactly the same amount of entropy (uncertainty) taken away.

Mathematical Fundamentals

- Entropy is a measure of uncertainty.
- When our uncertainty is reduced, we gain information,
  - so information and entropy are two sides of the same coin.
- However, information has a rather subtle interpretation, which can easily lead to confusion.
- Average information shares the same definition as entropy,
  - but whether we call a given quantity information or entropy depends on whether it is being given to us or taken away.
- For example, if a variable has high entropy the initial uncertainty of the variable is large and is, by definition, exactly equal to its entropy.
- If we are told the variable value, on average, we have been given information equal to the uncertainty (entropy) we had about its value.
- Thus, receiving an amount of information is equivalent to having exactly the same amount of entropy (uncertainty) taken away.

Mathematical Fundamentals

- Entropy is a measure of uncertainty.
- When our uncertainty is reduced, we gain information,
  - so information and entropy are two sides of the same coin.
- However, information has a rather subtle interpretation, which can easily lead to confusion.
- Average information shares the same definition as entropy,
  - but whether we call a given quantity information or entropy depends on whether it is being given to us or taken away.
- For example, if a variable has high entropy the initial uncertainty of the variable is large and is, by definition, exactly equal to its entropy.
- If we are told the variable value, on average, we have been given information equal to the uncertainty (entropy) we had about its value.
- Thus, receiving an amount of information is equivalent to having exactly the same amount of entropy (uncertainty) taken away.

Mathematical Fundamentals

- Entropy is a measure of uncertainty.
- When our uncertainty is reduced, we gain information,
  - so information and entropy are two sides of the same coin.
- However, information has a rather subtle interpretation, which can easily lead to confusion.
- Average information shares the same definition as entropy,
  - but whether we call a given quantity information or entropy depends on whether it is being given to us or taken away.
- For example, if a variable has high entropy the initial uncertainty of the variable is large and is, by definition, exactly equal to its entropy.
- If we are told the variable value, on average, we have been given information equal to the uncertainty (entropy) we had about its value.
- Thus, receiving an amount of information is equivalent to having exactly the same amount of entropy (uncertainty) taken away.

Mathematical Fundamentals

- Entropy is a measure of uncertainty.
- When our uncertainty is reduced, we gain information,
  - so information and entropy are two sides of the same coin.
- However, information has a rather subtle interpretation, which can easily lead to confusion.
- Average information shares the same definition as entropy,
  - but whether we call a given quantity information or entropy depends on whether it is being given to us or taken away.
- For example, if a variable has high entropy the initial uncertainty of the variable is large and is, by definition, exactly equal to its entropy.
- If we are told the variable value, on average, we have been given information equal to the uncertainty (entropy) we had about its value.
- Thus, receiving an amount of information is equivalent to having exactly the same amount of entropy (uncertainty) taken away.

Mathematical Fundamentals

- Entropy is a measure of uncertainty.
- When our uncertainty is reduced, we gain information,
  - so information and entropy are two sides of the same coin.
- However, information has a rather subtle interpretation, which can easily lead to confusion.
- Average information shares the same definition as entropy,
  - but whether we call a given quantity information or entropy depends on whether it is being given to us or taken away.
- For example, if a variable has high entropy the initial uncertainty of the variable is large and is, by definition, exactly equal to its entropy.
- If we are told the variable value, on average, we have been given information equal to the uncertainty (entropy) we had about its value.
- Thus, receiving an amount of information is equivalent to having exactly the same amount of entropy (uncertainty) taken away.

Mathematical Fundamentals

- Entropy is a measure of uncertainty.
- When our uncertainty is reduced, we gain information,
  - so information and entropy are two sides of the same coin.
- However, information has a rather subtle interpretation, which can easily lead to confusion.
- Average information shares the same definition as entropy,
  - but whether we call a given quantity information or entropy depends on whether it is being given to us or taken away.
- For example, if a variable has high entropy the initial uncertainty of the variable is large and is, by definition, exactly equal to its entropy.
- If we are told the variable value, on average, we have been given information equal to the uncertainty (entropy) we had about its value.
- Thus, receiving an amount of information is equivalent to having exactly the same amount of entropy (uncertainty) taken away.

Mathematical Fundamentals

- Entropy is a measure of uncertainty.
- When our uncertainty is reduced, we gain information,
  - so information and entropy are two sides of the same coin.
- However, information has a rather subtle interpretation, which can easily lead to confusion.
- Average information shares the same definition as entropy,
  - but whether we call a given quantity information or entropy depends on whether it is being given to us or taken away.
- For example, if a variable has high entropy the initial uncertainty of the variable is large and is, by definition, exactly equal to its entropy.
- If we are told the variable value, on average, we have been given information equal to the uncertainty (entropy) we had about its value.
- Thus, receiving an amount of information is equivalent to having exactly the same amount of entropy (uncertainty) taken away.

- Entropy is a measure of uncertainty.
- When our uncertainty is reduced, we gain information,
  - so information and entropy are two sides of the same coin.
- However, information has a rather subtle interpretation, which can easily lead to confusion.
- Average information shares the same definition as entropy,
  - but whether we call a given quantity information or entropy depends on whether it is being given to us or taken away.
- For example, if a variable has high entropy the initial uncertainty of the variable is large and is, by definition, exactly equal to its entropy.
- If we are told the variable value, on average, we have been given information equal to the uncertainty (entropy) we had about its value.
- Thus, receiving an amount of information is equivalent to having exactly the same amount of entropy (uncertainty) taken away.

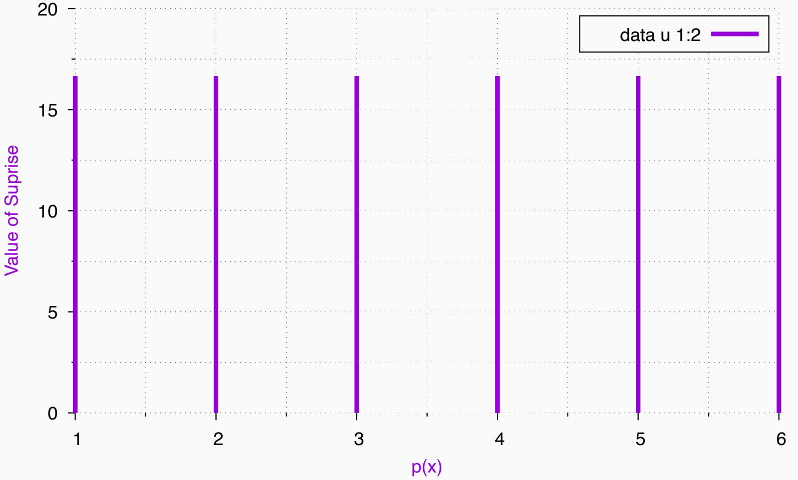

Figure 17: The probability distribution of 1 dice(s).

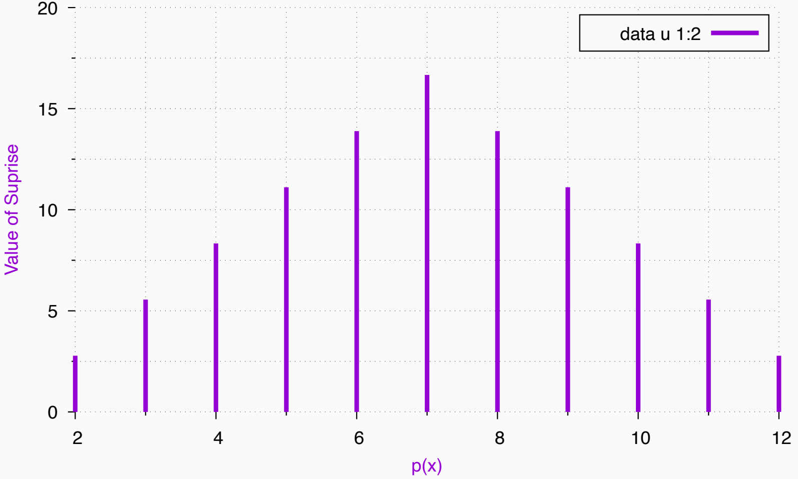

Figure 18: The probability distribution of 2 dice(s).

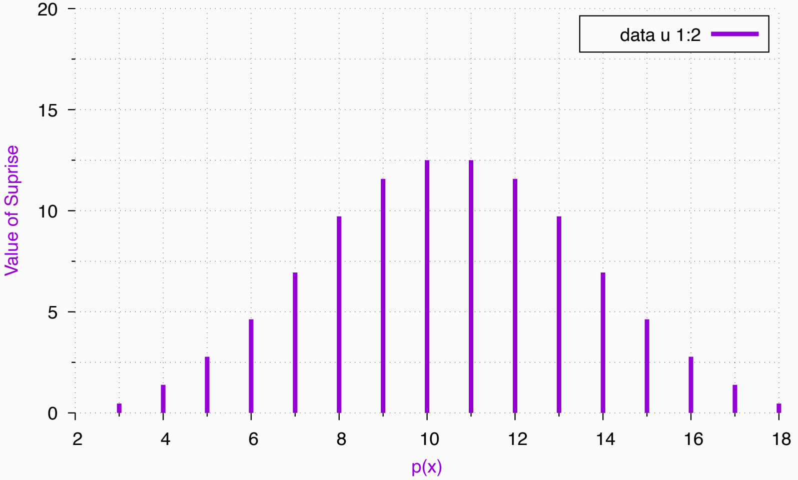

Figure 19: The probability distribution of 3 dice(s).

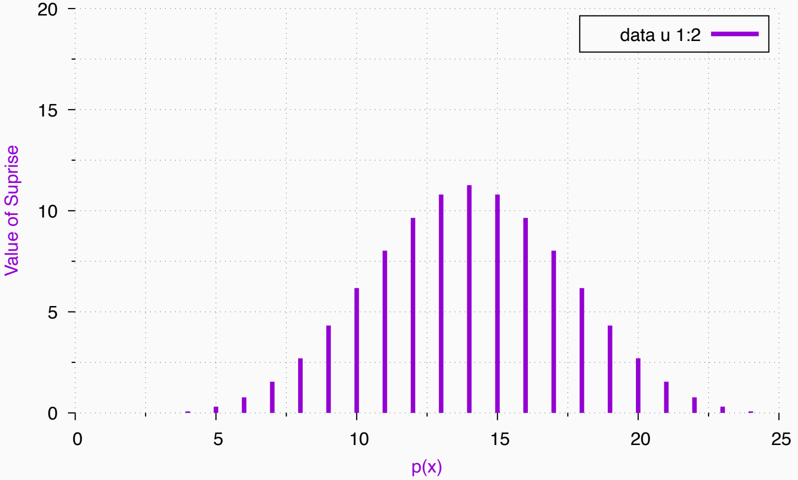

Figure 20: The probability distribution of 4 dice(s).

- Throwing a pair of 6-sided dice produces an outcome in the form of an ordered pair of numbers.
  - There are a total of 36 equiprobable outcomes,
- If we define an outcome value as the sum of this pair of numbers then there are $m = 11$ possible outcome values:

$$A_x = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\} .$$

Dividing the frequency of each outcome value by 36 yields the probability p of each outcome value.

Mathematical Fundamentals

- Throwing a pair of 6-sided dice produces an outcome in the form of an ordered pair of numbers.
  - There are a total of 36 equiprobable outcomes,
- If we define an outcome value as the sum of this pair of numbers then there are $m = 11$ possible outcome values:

$$A_x = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}\ .$$

Dividing the frequency of each outcome value by 36 yields the probability p of each outcome value.

- Throwing a pair of 6-sided dice produces an outcome in the form of an ordered pair of numbers.
  - There are a total of 36 equiprobable outcomes,
- If we define an outcome value as the sum of this pair of numbers then there are $m = 11$ possible outcome values:

$$A_x = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}.$$

Dividing the frequency of each outcome value by 36 yields the probability p of each outcome value.

- Throwing a pair of 6-sided dice produces an outcome in the form of an ordered pair of numbers.
  - There are a total of 36 equiprobable outcomes,
- If we define an outcome value as the sum of this pair of numbers then there are $m = 11$ possible outcome values:

$$A_x = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}.$$

Dividing the frequency of each outcome value by 36 yields the probability p of each outcome value.

Mathematical Fundamentals

- We can use these 11 probabilities to find the entropy.

$$
\begin{aligned}
H(x) &= p(x_1)\log\frac{1}{p(x_1)} + p(x_1)\log\frac{1}{p(x_1)} + \cdots + p(x_{11})\log\frac{1}{p(x_{11})} \\
&= 3.27 \text{ bits.}
\end{aligned}
$$