Artificial Vision

Course Summary - Master's Degree

Carlos Alberto Botina Carpio
Universidad Internacional de la Rioja
carlos.botina621@comunidadunir.net

November 18, 2025

Abstract

This document contains a summary of the Artificial Vision course syllabus for the Master's Degree. It includes a summary of the main topics covered during the sessions, as well as additional explanations and extensions of the concepts and techniques referenced in class. The purpose of this document is to serve as study material and reference for the course contents.

Contents

Lecture 003

1 Introduction

This document presents a summary of the **Artificial Vision** course syllabus for the Artificial Intelligence Master's Degree. The content includes a structured summary of the main topics covered during the course sessions, as well as additional explanations and extensions of the concepts, algorithms, and techniques referenced in class.

The main objective is to provide a comprehensive reference that complements the in-person sessions, facilitating the study and understanding of the fundamentals and applications of artificial vision. It includes detailed explanations of the most relevant topics, practical examples, and bibliographic references that allow for deeper exploration of the aspects covered during the course.

2 Sampling

Sampling is the process of converting a continuous-time signal into a discrete-time signal by measuring the signal's value at specific, uniformly spaced time instants. In the context of digital signal processing and computer vision, sampling is fundamental because real-world signals (such as images, sounds, or sensor measurements) are continuous in nature, but computers can only process discrete, finite sets of values.

The sampling process involves taking "snapshots" of a continuous signal at regular intervals, creating a sequence of discrete values that represent the original signal at those specific moments in time. The rate at which these samples are taken is called the **sampling frequency** or **sampling rate**, typically denoted as f_s and measured in samples per second (Hz). The time interval between consecutive samples is called the **sampling period** $T_s = 1/f_s$.

A critical question in sampling theory is: *How fast must we sample a signal to ensure that we can perfectly reconstruct the original continuous signal from its discrete samples?* This question is answered by the Nyquist-Shannon sampling theorem, which establishes the minimum sampling rate required for perfect reconstruction.

2.1 The Nyquist-Shannon Sampling Theorem

The Nyquist-Shannon sampling theorem, also known as the sampling theorem, is a fundamental principle in signal processing and digital image processing. It establishes the conditions under which a continuous signal can be perfectly reconstructed from its discrete samples.

2.1.1 Statement of the Theorem

Theorem 2.1 (Nyquist-Shannon Sampling Theorem)**.** If a function $x(t)$ contains no frequencies higher than B hertz, it is completely determined by giving its ordinates at a series of points spaced $\frac{1}{2B}$ seconds apart. In other words, a band-limited signal can be perfectly reconstructed from its samples if the sampling frequency f_s satisfies:

$$f_s \geq 2f_{\max} \tag{1}$$

where $f_{\max}$ is the highest frequency component in the signal. The frequency $f_N = \frac{f_s}{2}$ is called the **Nyquist frequency**, and $2f_{\max}$ is called the **Nyquist rate**.

Curious Fact: Frequency and Period Relationship

The relationship between frequency f and period T is fundamental in signal processing:

$$f = \frac{1}{T} \tag{2}$$

$$T = \frac{1}{f} \tag{3}$$

where:

- f is the frequency (measured in hertz, Hz, or cycles per second)
- T is the period (measured in seconds, s, or time per cycle)

This means that frequency and period are inversely related: higher frequency corresponds to shorter period, and vice versa. For example, if a signal has a frequency of $f = 10$ Hz, its period is $T = \frac{1}{10} = 0.1$ seconds. In the context of sampling, the sampling period T_s and sampling frequency f_s are related by $T_s = \frac{1}{f_s}$.

2.2 Understanding Sampling in the Time Domain

To understand the theorem, consider a continuous signal $x(t)$ that we wish to sample at regular intervals. The sampling process converts the continuous-time signal into a discrete-time signal by taking samples at uniformly spaced time instants. This relationship is mathematically expressed as:

$$x[n] = x(t_n), \quad t_n = nT_s, \quad n \in \mathbb{Z}, \quad T_s \in \mathbb{R} \tag{4}$$

where:

- $x[n]$ is the discrete-time signal (sequence of samples)
- $x(t_n)$ is the value of the continuous signal at time instant t_n
- n is an integer index representing the sample number
- T_s is the sampling period (time interval between consecutive samples)
- $f_s = \frac{1}{T_s}$ is the sampling frequency

Figure 1 illustrates this sampling process, showing how a continuous signal is converted into a discrete sequence of samples.

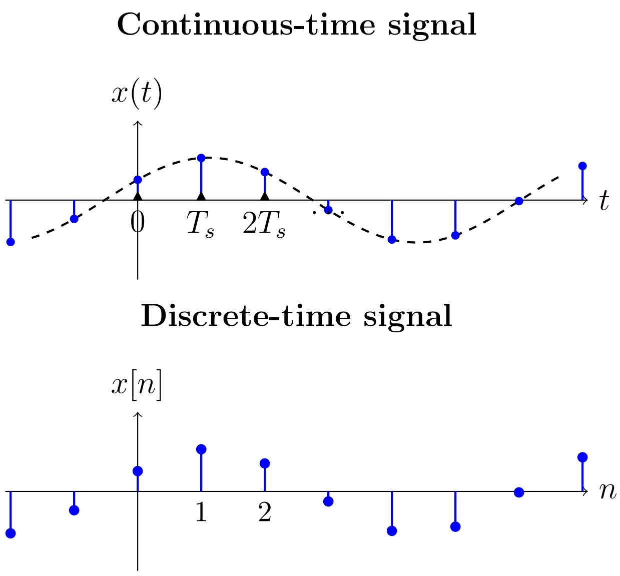

Figure 1: Sampling process: conversion from continuous-time signal $x(t)$ to discrete-time signal $x[n]$. The top plot shows the continuous signal with sampling instants marked, and the bottom plot shows the resulting discrete sequence.

2.3 Signal Reconstruction with Sinc Interpolation

Once a signal has been sampled according to the Nyquist-Shannon theorem, the original continuous signal can be perfectly reconstructed from its discrete samples. This reconstruction is achieved through **sinc interpolation**, which uses the sinc function to interpolate between sample points.

The sinc function is defined as:

$$\mathrm{sinc}(t) = \frac{\sin(\pi t)}{\pi t} \tag{5}$$

with the special case $\mathrm{sinc}(0) = 1$ (by L'Hôpital's rule).

The reconstruction formula, also known as the **Whittaker-Shannon interpolation formula**, expresses the continuous signal $x(t)$ as a weighted sum of sinc functions centered at each sample point:

$$x(t) = \sum_{n=-\infty}^{\infty} x[n] \cdot \mathrm{sinc}\left(\frac{t - nT_s}{T_s}\right) \tag{6}$$

where:

- $x[n]$ are the discrete samples
- T_s is the sampling period
- Each sinc function is centered at a sampling instant nT_s
- The sinc function has zeros at all other sampling instants, ensuring that $x(t)$ equals $x[n]$ at $t = nT_s$

This reconstruction works because:

1. At each sampling instant $t = nT_s$, only the sinc function centered at that point contributes (all others are zero), so $x(nT_s) = x[n]$.
2. Between sampling points, the sinc functions smoothly interpolate the signal values.
3. In the frequency domain, the sinc function acts as an ideal low-pass filter, removing all frequency components above the Nyquist frequency while preserving those below it.

Curious Fact: Band-Limited Signals and Perfect Reconstruction

Since sinc interpolation acts as a low-pass filter (removing all frequencies above the Nyquist frequency $f_N = f_s/2$), if we have a band-limited signal with maximum frequency $f_{\max}$ lower than the Nyquist frequency, then no frequencies are going to be removed and therefore, the result is a theoretically perfect reconstruction.

2.4 Aliasing

Aliasing is a distortion phenomenon that occurs when a signal is sampled at a rate that is too low (below the Nyquist rate). When aliasing occurs, high-frequency components of the signal are "folded back" or "aliased" into lower frequencies, making them indistinguishable from actual low-frequency components in the sampled signal.

2.4.1 Why Aliasing Occurs

In the frequency domain, sampling creates periodic replicas of the signal's spectrum at integer multiples of the sampling frequency. When the sampling frequency f_s is less than $2f_{\max}$, these replicas overlap. The overlapping high-frequency components appear as lower frequencies in the sampled signal, causing aliasing.

For example, consider a signal with frequency $f = 8$ Hz sampled at $f_s = 10$ Hz:

- The Nyquist frequency is $f_N = f_s/2 = 5$ Hz
- The signal frequency (8 Hz) is above the Nyquist frequency
- The aliased frequency is $f_{\text{alias}} = f_s - f = 10 - 8 = 2$ Hz
- The sampled signal incorrectly appears to have a 2 Hz component instead of the original 8 Hz

2.4.2 Consequences of Aliasing

Once aliasing occurs, the original signal cannot be perfectly reconstructed because the high-frequency information has been irretrievably mixed with lower frequencies. This is why the Nyquist-Shannon theorem requires sampling at or above the Nyquist rate to ensure perfect reconstruction.

Curious Fact: The Wagon Wheel Effect

A classic example of aliasing in everyday life is the **wagon wheel effect** (also known as the stroboscopic effect) seen in videos. When a wheel with spokes rotates at a certain speed and is filmed at a fixed frame rate, the wheel can appear to rotate backward, slowly, or even stand still. This occurs because the wheel's rotation frequency is being undersampled by the camera's frame rate. The high-frequency rotation is aliased into a lower apparent frequency, creating the illusion of reverse or slow motion. This is a temporal aliasing effect, where time (rather than space) is being sampled.

These are the three different sampling scenarios:

- **Adequate sampling** ($f_s > 2f_{\max}$): The signal can be perfectly reconstructed.
- **Nyquist rate sampling** ($f_s = 2f_{\max}$): The minimum sampling rate that theoretically allows perfect reconstruction.
- **Insufficient sampling** ($f_s < 2f_{\max}$): Aliasing occurs, and the original signal cannot be recovered.

2.5 Exercise: Determining Minimum Sampling Rate

Consider the following signal:

$$x(t) = \cos(100\pi t) + \sin(200\pi t) + \cos(500\pi t + \pi/4) + 7 \tag{7}$$

Problem: Determine the minimum sampling rate required to perfectly reconstruct this signal.

Solution:

To find the minimum sampling rate, we need to identify the maximum frequency component in the signal. Let's analyze each term:

- $\cos(100\pi t)$:

$$\begin{aligned}
\text{Angular frequency: } \omega_1 &= 100\pi \text{ rad/s} \\
\text{To convert to frequency: } f_1 &= \omega_1 \times \frac{1 \text{ cycle}}{2\pi \text{ rad}} \\
&= 100\pi \text{ rad/s} \times \frac{1 \text{ cycle}}{2\pi \text{ rad}} \\
&= \frac{100\cancel{\pi \text{ rad}}/\text{s} \times 1 \text{ cycle}}{2\cancel{\pi \text{ rad}}} \\
&= \frac{100}{2}\frac{\text{cycle}}{\text{s}} = 50 \text{ cycles/s} = 50 \text{ Hz}
\end{aligned}$$

- $\sin(200\pi t)$: $f_2 = \frac{200\pi}{2\pi} = 100$ Hz
- $\cos(500\pi t + \pi/4)$: $f_3 = \frac{500\pi}{2\pi} = 250$ Hz
- 7: This is a constant (DC component) with frequency $f_0 = 0$ Hz

The maximum frequency in the signal is $f_{\max} = 250$ Hz (from the $\cos(500\pi t + \pi/4)$ term).

According to the Nyquist-Shannon theorem, the minimum sampling rate (Nyquist rate) is:

$$f_s \geq 2f_{\max} = 2 \times 250 = 500 \text{ Hz} \tag{8}$$

Therefore, the minimum sampling rate required is **500 Hz**.

Figure 2 demonstrates the signal reconstruction process using **sinc interpolation** (Whittaker-Shannon interpolation formula) for three different sampling scenarios. The reconstruction is performed using the formula:

$$x(t) = \sum_{n=-\infty}^{\infty} x[n] \cdot \text{sinc}\left(\frac{t - nT_s}{T_s}\right) \tag{9}$$

where $\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$ and $T_s = 1/f_s$ is the sampling period.

Each row of the figure shows three plots: (1) the original continuous signal, (2) the sampled signal with sample points marked, and (3) the reconstructed signal using sinc interpolation overlaid with the original for comparison. The three rows correspond to:

- **Adequate sampling** ($f_s = 600$ Hz $> 2f_{\max}$): The reconstructed signal perfectly matches the original, demonstrating perfect reconstruction when sampling above the Nyquist rate.

- **Nyquist rate sampling** ($f_s = 500$ Hz $= 2f_{\max}$): The reconstructed signal matches the original, showing that the Nyquist rate is the theoretical minimum for perfect reconstruction.

- **Insufficient sampling** ($f_s = 300$ Hz $< 2f_{\max}$): The reconstructed signal does not match the original due to aliasing, demonstrating that perfect reconstruction is impossible when sampling below the Nyquist rate.

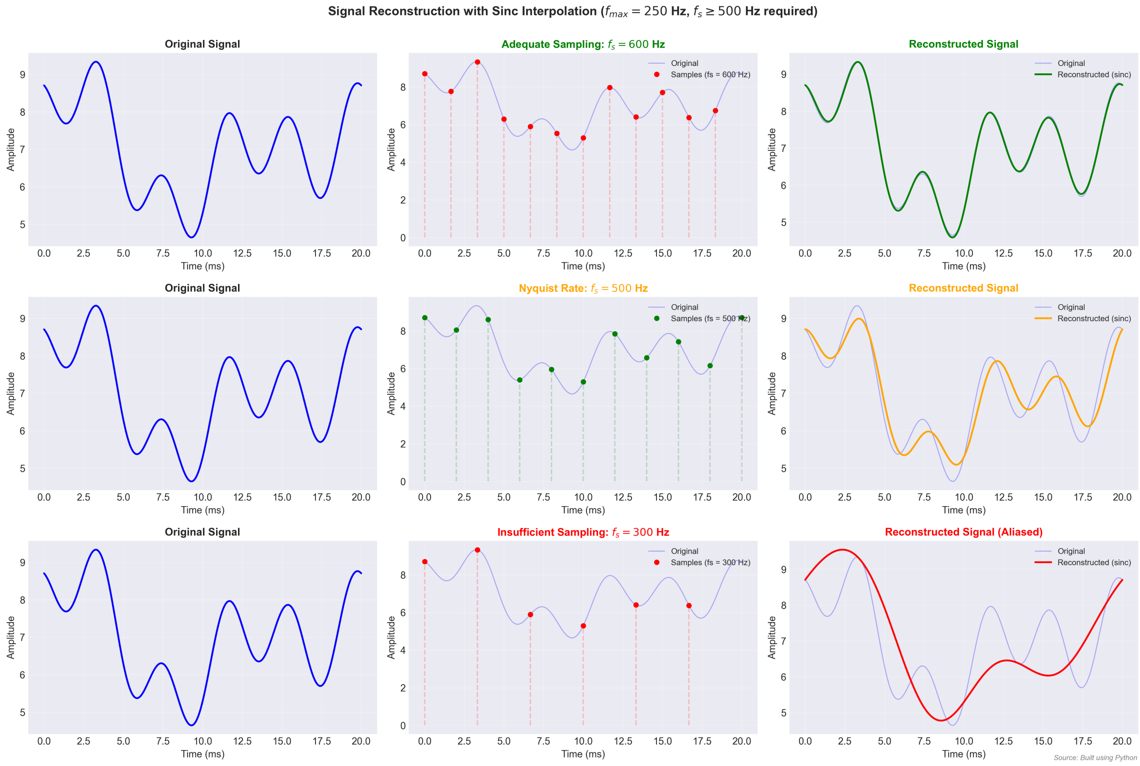

Figure 2: Signal reconstruction using sinc interpolation for $x(t) = \cos(100\pi t) + \sin(200\pi t) + \cos(500\pi t + \pi/4) + 7$ with $f_{\max} = 250$ Hz. Each row shows: original signal (left), sampled signal (center), and reconstructed signal using sinc interpolation (right). The three rows demonstrate adequate sampling (600 Hz), Nyquist rate (500 Hz), and insufficient sampling (300 Hz) where aliasing prevents perfect reconstruction.

Lecture 004

3 Entropy: Concept and Estimation

3.1 Noise

Noise is any unwanted signal of random nature that modifies the intensity of the original signal to be perceived.

In the real world, signals are affected by uncontrollable elements that generate noise. This noise is typically superimposed as **additive noise**:

$$S(t) = f(t) + r(t) \tag{10}$$

where $S(t)$ is the received signal, $f(t)$ is the original signal, and $r(t)$ is the noise component.

The first stage in signal processing focuses on identifying and eliminating noisy artifacts, though complete elimination is usually not feasible. The random nature of noise means that signals with noise are not deterministic but rather **stochastic processes**, where repeated measurements of the same signal produce different results.

3.1.1 Types of Noise

Atmospheric Noise Atmospheric noise comes from electrical signals derived from natural discharges that occur under the ionosphere. Storms or electrical charges in clouds are sources of this type of noise, which generally affects communication systems using the radio spectrum more significantly. Approximately, the power of atmospheric noise is inversely proportional to frequency. Thus, atmospheric noise has greater impact on low and medium frequency bands, while lower power noise affects VHF and UHF bands. As a result, atmospheric noise affects AM communication bands and decreases significantly at TV and FM frequencies. Beyond 30 MHz, atmospheric noise has less negative impact than the receiver's own noise.

Man-Made Noise This refers to electrical artifacts generated by sources such as automobiles, electric motors, switches, high-voltage lines, etc. It is also known as **industrial noise**. The intensity of these noisy signals is greater in large urban centers and industrial areas. In these areas, noise of this nature prevails over other noise sources in the frequency range between 1 MHz and 600 MHz.

Impulsive or Shot Noise This type of noise causes the appearance of anomalous values (outliers) in the signal. It is characterized by a sudden increase in intensity during a short period of time. Generally, its origin is an external agent to the information system: a lightning strike or interference from a motor spark. However, it should not be confused with atmospheric or man-made noise, as the duration of these is more prolonged in time.

Galactic Noise It originates from disturbances produced beyond the Earth's atmosphere. The main sources of galactic noise are the sun and other stars.

- **Solar**: The sun is a major source of energy emission in the form of electromagnetic radiation. These signals affect telecommunications systems. The frequency range of these emissions is very wide, including bands commonly used for radio communication systems. The intensity of the emission produced by the sun varies cyclically, with a period of approximately eleven years. At the highest levels, this radiation can make some frequency bands unusable.

- **Cosmic**: Like the sun, other stars near our planet emit energy in the form of electromagnetic radiation that can affect our signals and communication systems.

Thermal Noise This noise source is due to the random agitation of electrons in the elements of an electronic circuit. This movement could only be canceled under absolute zero temperature conditions. Therefore, it is an unavoidable noise source that will always be present in a signal acquisition and processing system. The movement of electrons increases as the temperature of the conductor increases, giving rise to small electrical currents. This noisy signal is distributed over a wide range of frequencies, so it will always affect the system to some degree, despite carrying out different filtering stages.

Flicker Noise or 1/f Noise It is called 1/f because its power decays below 1 kHz when frequency increases. Therefore, it has greater impact on low frequencies. The physical causes of this type of noise are not entirely clear. It originates in elements such as transistors or resistors, and it is hypothesized that it is due to intermodulation processes in these elements.

3.1.2 Signal-to-Noise Ratio (SNR)

When an information source is affected by noisy artifacts, the **Signal-to-Noise Ratio (SNR)** quantitatively indicates the quality of the signal of interest. This ratio is defined as the quotient between the power of the received signal and the estimated noise power. A value greater than unity (1) indicates a greater presence of the signal compared to the noise. The relationship between these power terms is generally expressed in decibels (dB).

$$\text{SNR} = 10 \log_{10}\left(\frac{P_S}{P_N}\right) \tag{11}$$

where:

- P_S corresponds to the signal power.
- P_N corresponds to the noise power.

Example: Consider a communication system where the signal power is $P_S = 100$ and the noise power is $P_N = 10$. The SNR is calculated as:

$$\begin{aligned} \text{SNR} &= 10 \log_{10}\left(\frac{P_S}{P_N}\right) \\ &= 10 \log_{10}\left(\frac{100}{10}\right) \\ &= 10 \log_{10}(10) \\ &= 10 \times 1 = 10 \text{ dB} \end{aligned}$$

This means the signal power is 10 times greater than the noise power (a ratio of 10:1), resulting in an SNR of 10 dB.

Interpretation of SNR values:

- **SNR greater than 0 dB**: Signal power exceeds noise power (good quality)
- **SNR = 0 dB**: Signal and noise powers are equal
- **SNR lower than 0 dB**: Noise power exceeds signal power (poor quality)
- **SNR = 20 dB**: Signal is 100 times stronger than noise (excellent quality)
- **SNR = 3 dB**: Signal is approximately 2 times stronger than noise (minimum acceptable for many applications)

3.2 Entropy

Signals contain information and are affected by various noise sources. In this context, the concept of **entropy** arises. Similar to physics, the term refers to the complexity of the signal. The addition of noise increases the degree of complexity of a signal, resulting in higher entropy.

In information theory, **entropy** is defined as the amount of information from a random source (on average). Therefore, entropy serves to **characterize a random variable**. Signals can be modeled as a sequence of realizations of a random variable over time (stochastic process), so we will see how to extend the definition of entropy to random elements of this nature.

3.2.1 Shannon's Entropy Definition

Given a discrete random variable X that takes values from the set $\{x_1, x_2, \ldots, x_M\}$ with probability distribution $P(X = x_i) = p_i$, Shannon defined entropy as:

$$H(X) = E\{-\log_2[P(X)]\} = \sum_{i=1}^{M} -\log_2[P(x_i)] \cdot P(x_i) = \sum_{i=1}^{M} -p_i \log_2(p_i) \tag{12}$$

where $-\log_2[P(x_i)]$ is interpreted as the **quantity of information** (or **self-information**) associated with outcome x_i.

Curious Fact: What Does E Mean?

The capital E denotes the **expected value** (also called expectation or mean). For a discrete random variable, the expected value of a function $g(X)$ is calculated as:

$$E[g(X)] = \sum_{i=1}^{M} g(x_i) \cdot P(x_i) \tag{13}$$

In the entropy formula, $E\{-\log_2[P(X)]\}$ means we take the expected value of the information content $-\log_2[P(X)]$, which gives us the average information across all possible outcomes.

3.2.2 Understanding the Formula

The key insight is that **less probable values carry more information** (surprise effect) compared to more probable values. For example:

- If an event is very likely ($p_i \approx 1$), then $-\log_2(p_i) \approx 0$: we learn little new information.
- If an event is very unlikely ($p_i \approx 0$), then $-\log_2(p_i)$ is large: we learn a lot of new information.

Entropy $H(X)$ is the **expected value** (average) of this information content across all possible outcomes.

3.2.3 Example: Bernoulli Distribution

Consider a random variable X with only two possible outcomes, $\{x_1, x_2\}$ (a **Bernoulli distribution**). Let $P(X = x_1) = p$ and $P(X = x_2) = 1 - p$. The entropy is:

$$H(X) = -p \log_2(p) - (1-p)\log_2(1-p) \tag{14}$$

The entropy reaches its **maximum value of 1** when $p = 0.5$. In this case, both events have equal probability, and on average we obtain the same amount of information from X. When p approaches 0 or 1, the entropy approaches 0, meaning we can almost predict the outcome with certainty, so we learn little new information.

Concrete Example: Consider a fair coin flip where $p = 0.5$:

$$\begin{aligned} H(X) &= -0.5\log_2(0.5) - 0.5\log_2(0.5) \\ &= -0.5 \cdot (-1) - 0.5 \cdot (-1) \\ &= 0.5 + 0.5 = 1 \end{aligned}$$

This means each coin flip provides an entropy of 1 on average. If the coin is biased (e.g., $p = 0.9$ for heads), then:

$$\begin{aligned} H(X) &= -0.9\log_2(0.9) - 0.1\log_2(0.1) \\ &\approx 0.469 \end{aligned}$$

The entropy is lower because we can predict the outcome more easily (heads is very likely), so we learn less information.

Figure 3 shows the variation of entropy $H(X)$ as a function of the probability $P(X = x_1) = p$ for a Bernoulli distribution. The curve is symmetric and reaches its maximum value of 1 when $p = 0.5$, demonstrating that uncertainty (entropy) is highest when both outcomes are equally probable.

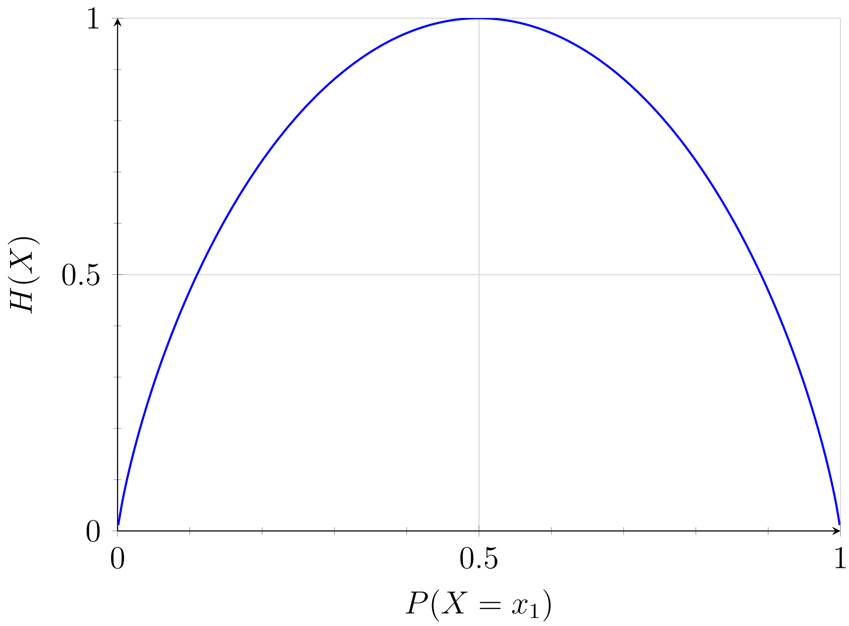

Figure 3: Entropy $H(X)$ as a function of probability $P(X = x_1) = p$ for a Bernoulli distribution. The entropy reaches its maximum value of 1 when $p = 0.5$ (equal probability for both outcomes) and approaches 0 when p approaches 0 or 1 (certain outcomes).

3.3 Signals as Stochastic Processes

Signals can be modeled mathematically as a set of random variables (a **stochastic process**). For example, a voice signal of a certain duration can be viewed as a finite time series, where each sample represents a realization of a random variable.

Including new samples in the series increases the information content, showing that process entropy depends on its length. Therefore, it makes sense to measure the variation of signal entropy due to the inclusion of a new sample. This is called the **entropy rate** or **differential entropy**.

3.3.1 Entropy of a Stochastic Process

Consider a signal of length N as a sequence of N random variable realizations: $\mathbf{x} = x_1, x_2, \ldots, x_N$. Figure 4 illustrates this concept, showing a signal where each sample x_i represents a realization of a random variable at time index i.

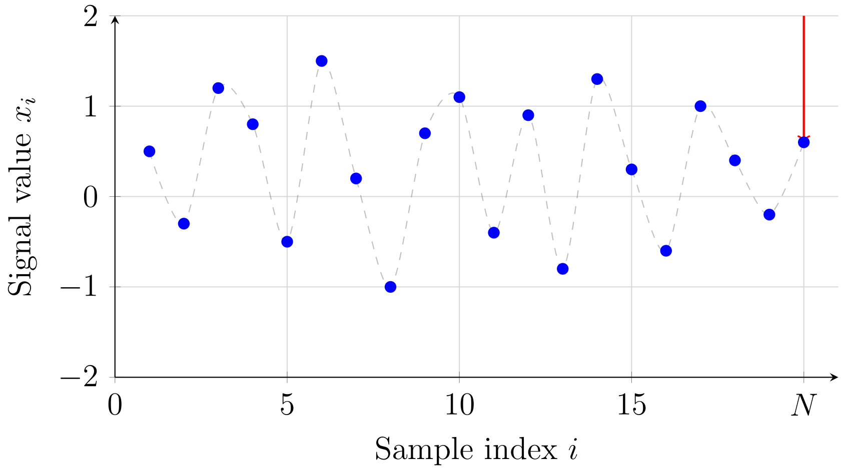

Figure 4: A signal of length $N = 20$ as a sequence of random variable realizations. Each point (i, x_i) represents a sample where i is the time index and x_i is the realization of the random variable at that time.

The entropy H_N of this stochastic process is:

$$\begin{aligned} H_N &= E\{-\log_2[p(x_1, x_2, \ldots, x_N)]\} \\ &= -\int_{-\infty}^{\infty} \log_2[p(x_1, x_2, \ldots, x_N)] \cdot p(x_1, x_2, \ldots, x_N)\, dx_1 \ldots dx_N \end{aligned} \tag{15}$$

where $p(x_1, x_2, \ldots, x_N)$ is the joint probability density function (PDF) of the variables composing the stochastic process.

Note: Computational Complexity

Computing entropy using the stochastic process formula is **much more computationally expensive** than ApEn:

- **Stochastic process entropy**: Requires estimating an N-dimensional joint PDF and computing an N-dimensional integral. The complexity grows exponentially with N (curse of dimensionality), making it impractical for long signals.
- **ApEn**: Works with fixed-length subseries (m is typically 2-3), requiring only pattern matching. Complexity is approximately $O(N^2)$, which is manageable even for long signals.

This computational advantage is one of the main reasons why ApEn is widely used in practice.

3.3.2 Entropy Rate

The **entropy rate** E_N of the signal is defined as:

$$E_N = \lim_{N\to\infty} (H_{N+1} - H_N) \tag{16}$$

This represents the change in entropy when a new sample is added to an infinitely long sequence.

Note

The entropy rate measures how much entropy increases (on average) when you add one more sample to a long signal.

3.3.3 Approximate Entropy (ApEn)

There are various methods for estimating signal entropy. **Approximate Entropy (ApEn)** is one such estimation procedure. The algorithm involves estimating the entropy of subseries of length m and $m+1$. The final entropy value is obtained by taking the difference between these two estimations.

Methodology: Consider an original time series $\mathbf{x} = [x_1, x_2, \ldots, x_N]$.

Step 1: Extract subseries. Extract all subseries of length m, denoted as $x_i^{(m)} = [x_i, x_{i+1}, \ldots, x_{i+m-1}]$ for $i = 1, 2, \ldots, N-m+1$.

Step 2: Find similar subseries. Given a tolerance r, count the number of subseries $N^{(m)}(i)$ that are similar to $x_i^{(m)}$, where similarity is determined by a distance metric $d[x_i^{(m)}, x_j^{(m)}] \leq r$.

Step 3: Calculate probability. The probability of finding a subseries similar to $x_i^{(m)}$ in the original series is:

$$C^{(m)}(i) = \frac{N^{(m)}(i)}{N-m+1} \tag{17}$$

where $N-m+1$ is the total number of subseries of length m that can be extracted from the original series.

Step 4: Estimate entropy. The term $C^{(m)}(i)$ provides a discrete estimation of the probability density function. Using Shannon's entropy definition, the entropy of the process represented by $x^{(m)}$ is:

$$H_N^{(m)} = -\frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \log_2[C^{(m)}(i)] \tag{18}$$

The Approximate Entropy is then calculated as:

$$\text{ApEn}(m, r, N) = H_N^{(m)} - H_N^{(m+1)} \tag{19}$$

Note: ApEn vs. Entropy Rate

Approximate Entropy (ApEn) is **not exactly the same** as the entropy rate, but they are related concepts:

- **Entropy rate** $E_N = \lim_{N\to\infty}(H_{N+1} - H_N)$ measures the change in entropy when adding one more **sample** to the sequence.
- **ApEn** $= H_N^{(m)} - H_N^{(m+1)}$ measures the change in entropy when increasing the **subseries length** from m to $m+1$.

ApEn is an **approximation** of the entropy rate. Instead of computing the theoretical entropy rate (which requires the full joint PDF), ApEn estimates it by analyzing patterns in subseries of different lengths. Both measure how entropy changes with sequence length, but ApEn uses a practical, pattern-based approach rather than the theoretical limit.

Example: Consider a time series $\mathbf{x} = [1.0, 1.2, 0.9, 1.1, 1.3, 0.8]$ with $N = 6$. Let $m = 2$ and $r = 0.2$.

Step 1: Extract subseries of length $m = 2$:

- $x_1^{(2)} = [1.0, 1.2]$
- $x_2^{(2)} = [1.2, 0.9]$
- $x_3^{(2)} = [0.9, 1.1]$
- $x_4^{(2)} = [1.1, 1.3]$
- $x_5^{(2)} = [1.3, 0.8]$

Step 2: Find similar subseries. Using the Chebyshev distance (maximum absolute difference between corresponding elements), we count how many subseries are within tolerance $r = 0.2$ of each $x_i^{(2)}$. Two subseries are similar if $d[x_i^{(m)}, x_j^{(m)}] = \max_k |x_{i+k} - x_{j+k}| \leq r$.

For example, comparing $x_1^{(2)} = [1.0, 1.2]$ with $x_4^{(2)} = [1.1, 1.3]$:

$$\begin{aligned} d[x_1^{(2)}, x_4^{(2)}] &= \max(|1.0 - 1.1|, |1.2 - 1.3|) \\ &= \max(0.1, 0.1) = 0.1 \leq 0.2 \end{aligned}$$

Since $0.1 \leq 0.2$, these subseries are similar.

Results for all subseries:

- For $x_1^{(2)} = [1.0, 1.2]$: $N^{(2)}(1) = 2$ (matches itself and $x_4^{(2)} = [1.1, 1.3]$)
- For $x_2^{(2)} = [1.2, 0.9]$: $N^{(2)}(2) = 1$ (only itself)
- For $x_3^{(2)} = [0.9, 1.1]$: $N^{(2)}(3) = 1$ (only itself)
- For $x_4^{(2)} = [1.1, 1.3]$: $N^{(2)}(4) = 2$ (matches itself and $x_1^{(2)}$)
- For $x_5^{(2)} = [1.3, 0.8]$: $N^{(2)}(5) = 1$ (only itself)

Step 3: Calculate probabilities. With $N - m + 1 = 6 - 2 + 1 = 5$ total subseries:

$$\begin{aligned}
C^{(2)}(1) &= \frac{2}{5} = 0.4 \\
C^{(2)}(2) &= \frac{1}{5} = 0.2 \\
C^{(2)}(3) &= \frac{1}{5} = 0.2 \\
C^{(2)}(4) &= \frac{2}{5} = 0.4 \\
C^{(2)}(5) &= \frac{1}{5} = 0.2
\end{aligned}$$

Step 4: Estimate entropy.

$$\begin{aligned}
H_6^{(2)} &= -\frac{1}{5} \sum_{i=1}^{5} \log_2[C^{(2)}(i)] \\
&= -\frac{1}{5}[\log_2(0.4) + \log_2(0.2) + \log_2(0.2) + \log_2(0.4) + \log_2(0.2)] \\
&\approx -\frac{1}{5}[-1.32 - 2.32 - 2.32 - 1.32 - 2.32] \\
&\approx 1.92
\end{aligned}$$

Similarly, we would compute $H_6^{(3)}$ for subseries of length $m + 1 = 3$, and then:

$$\text{ApEn}(2, 0.2, 6) = H_6^{(2)} - H_6^{(3)}$$

Takeaway: Why is ApEn Useful?

Approximate Entropy is a practical and powerful tool for signal analysis because:

- **Practical estimation**: It provides a computationally feasible way to estimate entropy without requiring knowledge of the full probability distribution, making it applicable to real-world signals.
- **Pattern detection**: By analyzing subseries patterns, ApEn can detect regularity and predictability in signals, which is useful for characterizing signal complexity.
- **Noise robustness**: The tolerance parameter r allows ApEn to be robust to noise, focusing on overall patterns rather than exact matches.
- **Wide applications**: ApEn is widely used in biomedical signal processing (EEG, ECG), time series analysis, and any domain where quantifying signal complexity or regularity is important.
- **Comparative analysis**: It enables comparison of entropy between different signals or different segments of the same signal, helping identify changes in signal characteristics.

3.4 Entropy in Images

The entropy value of a signal can be interpreted as its degree of uncertainty. Equivalently, it reflects the capacity to predict a future state or value from the knowledge or observation of previous signal values. A higher entropy value reflects greater complexity and chaos in the signal under study.

As a result, entropy gives us an idea of the level of noise impact on a signal. If we take a sample of the same signal under the same conditions but at different time instants, the signal with higher entropy will be the one with a higher noise level.

3.4.1 Images vs. One-Dimensional Signals

The nature and mathematical modeling of images are different from one-dimensional time-dependent signals. An image does not have an implicit time variable, as occurs in a voice signal or an electrocardiogram, but rather represents light captured at each spatial position. Additionally, image information is represented in two dimensions.

Therefore, in images, just as in time series we characterized the rate of entropy increase with respect to new samples, we could think of an entropy rate with respect to the unit area represented. For entropy estimation in an image, the histogram of intensity levels is used. The final estimation is obtained as the entropy of the random variable characterized by this histogram.

As with one-dimensional signals, entropy will tend to increase, or at least remain the same, if the image area considered for estimation is enlarged. Thus, lower entropy values will be associated with repetitive patterns in the image that lead to a histogram with marked peaks (texture). In contrast, entropy increases if there is greater variability in the intensity values observed in the image, with no marked patterns producing a flatter histogram. In this sense, it follows that noise contributes to increasing the entropy of the image, as it causes the variability of the observed intensity levels to increase.

3.5 Mathematical Characterization of Noise: Stochastic Processes

The term **stochastic process** has been previously used in this topic to refer to a random signal. In our case, any signal will be the result of the combination of the signal of interest and an unwanted signal, of random and chaotic nature, which contributes to increasing entropy. This unwanted signal is noise.

Therefore, the resulting signal is, in itself, a random signal. Just as happens with a random variable, from which we take a sample and obtain values according to a probability density function, the signals we handle are realizations of a stochastic process. Each time we extract a sample from the information source, we obtain a different signal.

In this section, a formal definition of stochastic process is provided that allows understanding the modeling and characterization of noise in signal processing.

3.5.1 Random Variables

A random variable is characterized by the following three elements:

- **Sample space**: The set of all possible outcomes that can be observed in the realization of an experiment.
- **Set of events**: Subset of the sample space.
- **Probability law**: Assignment of probability to each of the observable events.

Example: Rolling a Fair Die Consider the experiment of rolling a fair six-sided die:

- **Sample space**: $\Omega = \{1, 2, 3, 4, 5, 6\}$ (all possible outcomes)
- **Set of events**: Examples include:
  - Event A: "Rolling an even number" $= \{2, 4, 6\}$
  - Event B: "Rolling a number greater than 4" $= \{5, 6\}$
  - Event C: "Rolling a 3" $= \{3\}$
- **Probability law**: For a fair die, each outcome has equal probability:
  - $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$
  - $P(A) = P(\{2, 4, 6\}) = \frac{3}{6} = \frac{1}{2}$
  - $P(B) = P(\{5, 6\}) = \frac{2}{6} = \frac{1}{3}$

Example: Signal Intensity Measurement In signal processing, consider measuring the intensity of a signal at a specific time:

- **Sample space**: $\Omega = [0, 255]$ (all possible intensity values, e.g., for an 8-bit image)
- **Set of events**: Examples include:
  - Event D: "Intensity between 100 and 150" $= [100, 150]$
  - Event E: "Intensity greater than 200" $= (200, 255]$
- **Probability law**: Defined by a probability density function (PDF) $f(x)$ that assigns probabilities to intervals, such as:
  - $P(D) = \int_{100}^{150} f(x)\, dx$
  - $P(E) = \int_{200}^{255} f(x)\, dx$

3.5.2 Stochastic Processes

A stochastic process can be viewed as a random variable for which the result of an experiment is given in the form of a signal. In the same way as a random variable, it is characterized by the three elements mentioned: sample space, set of events, and probability assignment law.

In practice, as previously mentioned in this topic, we will have noisy signals that, from a mathematical point of view, will be modeled as a stochastic process. The noise component will be assumed to be additive, so the captured signal will have the following form:

$$y(t) = x(t) + \varepsilon(t) \tag{20}$$

where:

- $x(t)$ reflects the signal of interest.
- $\varepsilon(t)$ corresponds to the noise.

Example: Sinusoidal Signal with Gaussian Noise Consider, for example, that the signal of interest corresponds to a tone of frequency f and that the noise component follows a Gaussian distribution with zero mean and variance σ^2.

In Figure 5, we can see this target signal (top) and a realization of the stochastic process that corresponds to the observed signal (bottom).

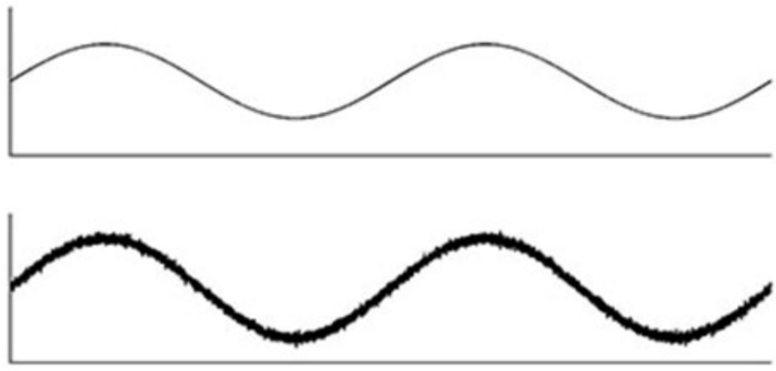

Figure 5: Comparison of a clean sinusoidal signal (top) and a noisy realization of the stochastic process (bottom). The noise component gives the signal a random nature that prevents us from knowing its exact value at any instant t.

As can be seen, the noise component gives the signal a random nature that prevents us from knowing its exact value at any instant t. In order to characterize the stochastic process, the objective will be to know its statistical properties. Probability distribution and density functions allow us to model the process statistically.

These functions are given as follows:

- **Distribution function**: $F_X(x,t) = P(X(t) \leq x)$
- **Probability density function**: $f_X(x,t) = \frac{dF_X(x,t)}{dx}$

From these functions, the stationarity of the process can be defined:

- A process is stationary in **strict sense** if the probability density function that characterizes the process does not vary with time. That is, for a constant c such that $c > 0$, the following will hold: $f_X(x,t) = f_X(x,t+c)$
- A process is stationary in **wide sense** if the statistical moments that characterize it (mean, variance, etc.) do not vary with respect to time.

Example: Strict-Sense Stationarity Consider a process $X(t) = A\sin(2\pi ft + \phi)$, where A and ϕ are random variables, and f is a constant frequency. If A and ϕ are independent, with A having a fixed distribution and ϕ uniformly distributed on $[0, 2\pi]$, then the PDF of $X(t)$ at any time t is the same (it depends only on the distribution of A and ϕ, not on t). This process is stationary in strict sense because $f_X(x,t) = f_X(x,t+c)$ for any c.

Example: Wide-Sense Stationarity Consider white noise $\varepsilon(t)$ with zero mean and constant variance σ^2:

- Mean: $E[\varepsilon(t)] = 0$ (constant, independent of t)
- Variance: $\mathrm{Var}[\varepsilon(t)] = \sigma^2$ (constant, independent of t)
- Autocorrelation: $E[\varepsilon(t)\varepsilon(t+\tau)] = \sigma^2\delta(\tau)$ (depends only on τ, not on t)

This process is stationary in wide sense because its statistical moments (mean and variance) do not vary with time, even though we may not know the full PDF.

Example: Non-Stationary Process Consider a process $Y(t) = t + \varepsilon(t)$, where $\varepsilon(t)$ is white noise. The mean of this process is $E[Y(t)] = t$, which clearly varies with time. Therefore, this process is **not stationary** (neither in strict sense nor in wide sense) because its statistical properties change over time.

Example: Non-Stationary Signal with Trend Let us return to the previous example. In this case, the captured signal shows, in addition to Gaussian noise, another component that causes a clear trend over time. Figure 6 shows this new example.

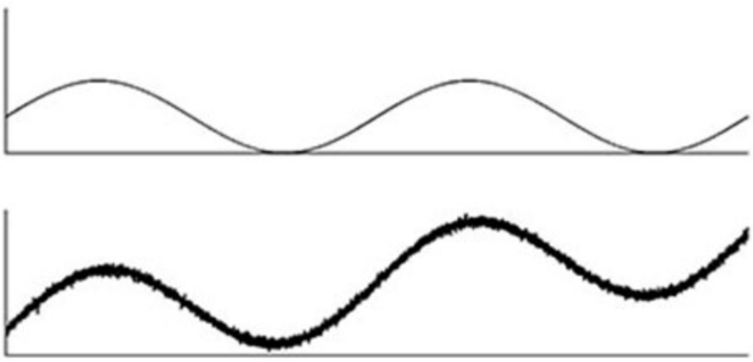

Figure 6: Non-stationary signal with a trend component. The signal exhibits both Gaussian noise and a clear trend over time, making its statistical properties vary along the temporal axis.

As a result of this trend, the statistical properties of the signal do not remain constant along the temporal axis, so it cannot be considered a stationary signal. It will be necessary to eliminate the noise component that causes this trend to remove the non-stationarity present in our information.