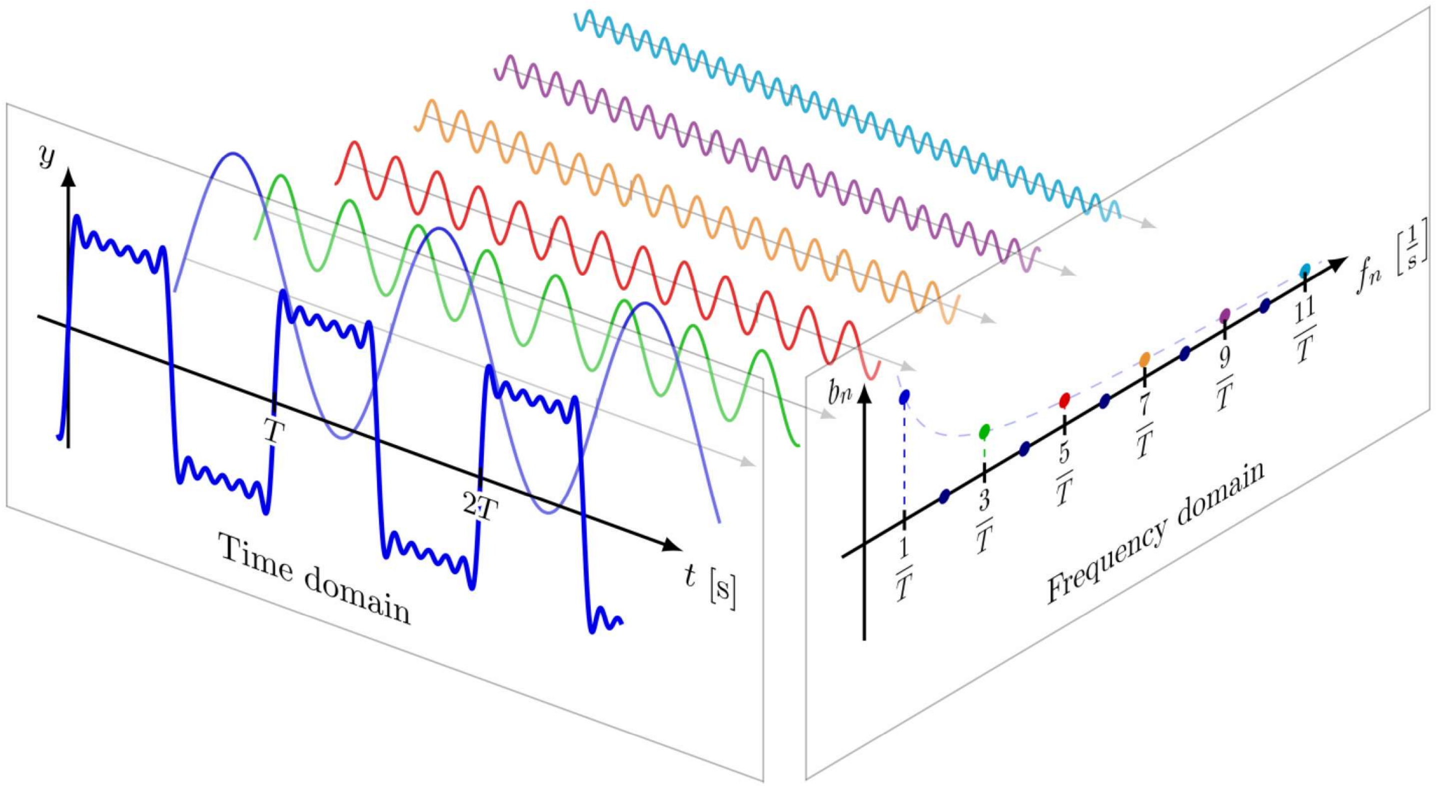
CURS 6 PD

Introducere în Procesarea Datelor



Asist. drd. ing. Iulian Vasiliev – 2024

Introducere

Clasificarea semnalelor

Eșantionarea semnalelor

Cuantificarea semnalelor

Clasificarea semnalelor

Eșantionarea semnalelor

Introducere

Cuantificarea semnalelor

- Termenii *„**date**”* și *„**informații**”* se folosesc de obicei cu aceeași semnificație, totuși ei nu înseamnă același lucru.

Definiție

Datele sunt o colecție de fapte individuale ce poate veni sub formă de număr, text, simbol, grafic, imagine, fiind o formă de cunoaștere brută ce nu posedă nici semnificație, nici scop.

Definiție

Informațiile sunt cunoaștere, fiind rezultatul analizei și interpretării datelor

Exemplu

DATE:

$x = [2.1 \quad 4.5 \quad 8.9 \quad 15 \quad \ldots \quad 4.32]$

Un set de 120 de valori numerice reale, fără nici un alt context, nu are nici o semnificație

PRELUCRARE ȘI ANALIZĂ →

CONTEXT:

Media lunară a temperaturilor citite de-a lungul a 10 ani într-o anumită locație

INFORMAȚII:

- șabloane sezoniere
- tendința de schimbare climatică

Definiție

Datele sunt o colecție de fapte individuale ce poate veni sub formă de număr, text, simbol, grafic, imagine, fiind o formă de cunoaștere brută ce nu posedă nici semnificație, nici scop.

Definiție

Informațiile sunt cunoaștere, fiind rezultatul analizei și interpretării datelor

Exemplu

DATE:

$$x = [2.1 \quad 4.5 \quad 8.9 \quad 15 \quad \ldots \quad 4.32]$$

Un set de 120 de valori numerice reale, fără nici un alt context, nu are nici o semnificație

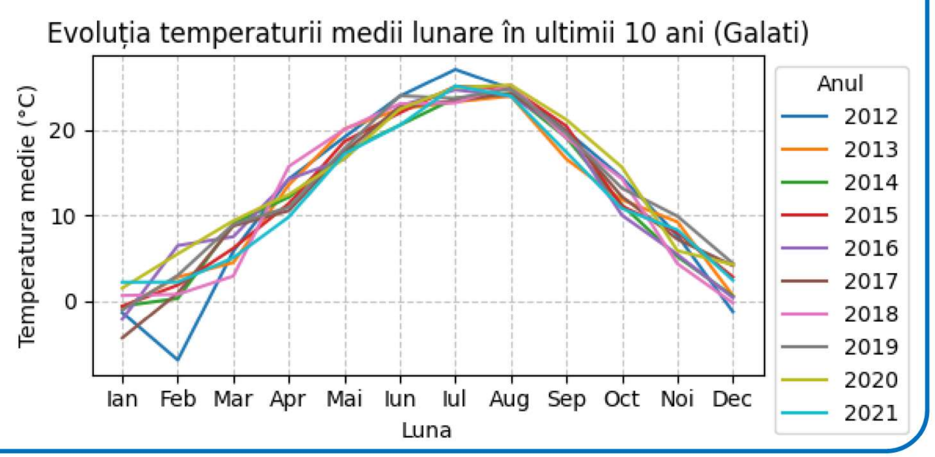
PRELUCRARE ȘI ANALIZĂ

CONTEXT:

Media lunară a temperaturilor citite de-a lungul a 10 ani într-o anumită locație

INFORMAȚII:

- șabloane sezoniere
- tendința de schimbare climatică



Definiție

Datele sunt o colecție de fapte individuale ce poate veni sub formă de număr, text, simbol, grafic, imagine, fiind o formă de cunoaştere brută ce nu posedă nici semnificație, nici scop.

Definiție

Informațiile sunt cunoaştere, fiind rezultatul analizei şi interpretării datelor

Date	Informații
O colecție de fapte	Fapte puse în context
Brute şi neorganizate	Organizate
Fiecare element este individual şi independent	Oferă o viziune de ansamblu
Lipsite de sens	Au sens
Nu depind de informații	Depind de date
Nu sunt suficiente pentru a lua decizii	Se pot lua decizii pe baza lor

Definiție

Datele sunt o colecție de fapte individuale ce poate veni sub formă de număr, text, simbol, grafic, imagine, fiind o formă de cunoaștere brută ce nu posedă nici semnificație, nici scop.

Definiție

Informațiile sunt cunoaștere, fiind rezultatul analizei și interpretării datelor

Alte exemple:

1. Pentru un supermarket, bonul fiscal al unui client reprezintă o **DATĂ**. Totuși când managerul supermarketului adună și interpretează toate bonurile fiscale dintr-o anumită perioadă de timp, acesta poate extrage **INFORMAȚII** utile precum profitul magazinului, ce produse sunt cele mai populare etc.
2. Prețurile competitorilor sunt **DATE** individuale, dar prin procesarea acestora se pot afla **INFORMAȚII** precum ce competitor are avantaj pe piață, ce oportunități se găsesc pe piață și cum poate compania să depășească concurența.
3. Temperatura citită de un termostat reprezintă o **DATĂ** individuală, totuși validarea și compararea acesteia cu un prag reprezintă o **INFORMAȚIE**: temperatura e prea mică / mare, informație ce poate sta la baza unei decizii.

Definiție

Datele sunt o colecție de fapte individuale ce poate veni sub formă de număr, text, simbol, grafic, imagine, fiind o formă de cunoaștere brută ce nu posedă nici semnificație, nici scop.

Definiție

Informațiile sunt cunoaștere, fiind rezultatul analizei și interpretării datelor

Tipuri de date:

1. **date cantitative** – exprimate de obicei numeric (exemple: greutate, volum, cost etc.)
2. **date calitative** – se referă la descriere (exemplu: gen, culoare etc.)

Definiție

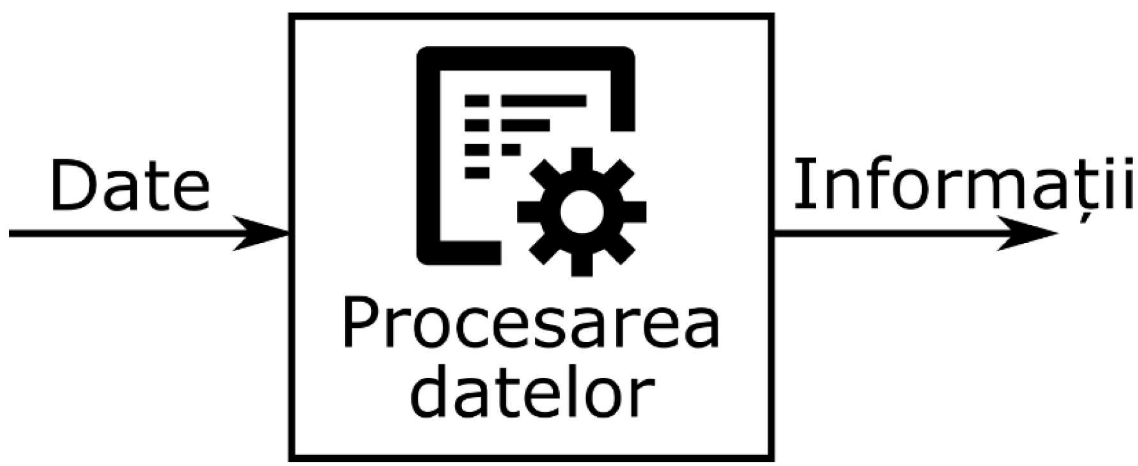
Datele sunt o colecție de fapte individuale ce poate veni sub formă de număr, text, simbol, grafic, imagine, fiind o formă de cunoaștere brută ce nu posedă nici semnificație, nici scop.

Definiție

Informațiile sunt cunoaștere, fiind rezultatul analizei și interpretării datelor

Definiție

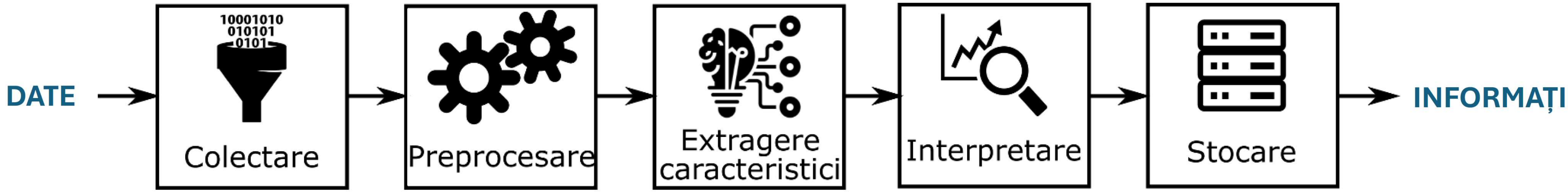
Procesarea datelor este un flux de colectare, de manipulare și de interpretare a datelor care are ca scop obținerea de informații utile ce nu au nevoie de nici un context pentru a fi înțelese.



Definiție

Procesarea datelor este un flux de colectare, de manipulare și de interpretare a datelor care are ca scop obținerea de informații utile ce nu au nevoie de nici un context pentru a fi înțelese.

Etape ale procesării datelor:



- **Colectare**: datele sunt extrase din diverse surse de încredere
- **Preprocesare**: datele brute sunt verificate de erori, curățate, datele redundante, incomplete sau inutile fiind eliminate
- **Extragere caracteristici**: procesarea propriu-zisă în care sunt extrase caracteristicile utile pentru obținerea informațiilor
- **Interpretare**: obținerea informațiilor prin interpretarea caracteristicilor extrase
- **Stocare**: informațiile sunt stocate pentru a putea fi folosite în viitor

Definiție

Procesarea datelor este un flux de colectare, de manipulare și de interpretare a datelor care are ca scop obținerea de informații utile ce nu au nevoie de nici un context pentru a fi înțelese.

Exemple de procese implicate în procesarea datelor:

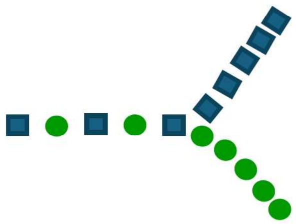
1. **Validare** – Ne asigurăm că datele sunt valide și relevante

   Exemple:
   - *validarea tipului de dată* – temperatura: număr sau text?
   - *validarea intervalului și a constrângerilor* – vârsta este un număr pozitiv? numele are o lungime mai mică de 100 de litere?

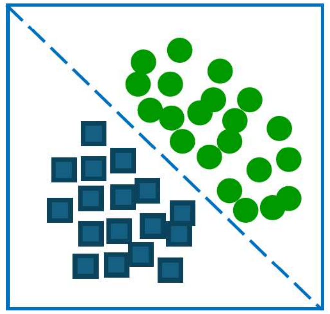
2. **Sortare** – Aranjarea datelor într-o anumită ordine

3. **Sumarizare** – Reducerea datelor detaliate

4. **Agregare** – Combinarea mai multor secvențe de date



5. **Clasificare** – obiectele dintr-o mulțime sunt distribuite pe clase într-o anumită ordine



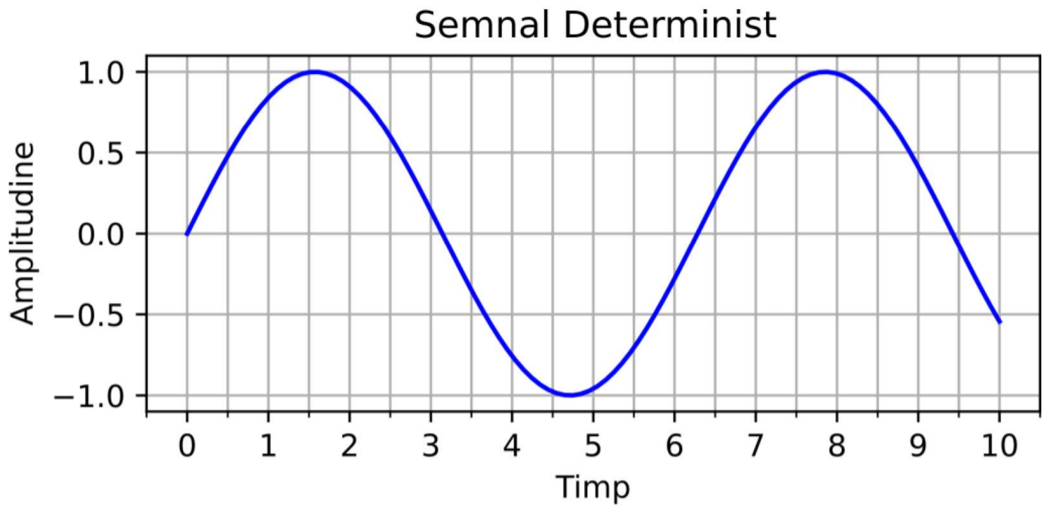
Definiție

Semnalele sunt reprezentări electrice sau electronice ale datelor care evoluează în funcție de timp. La orice moment de timp t, semnalul va fi caracterizat de o valoare numită **amplitudine.**

Definiție

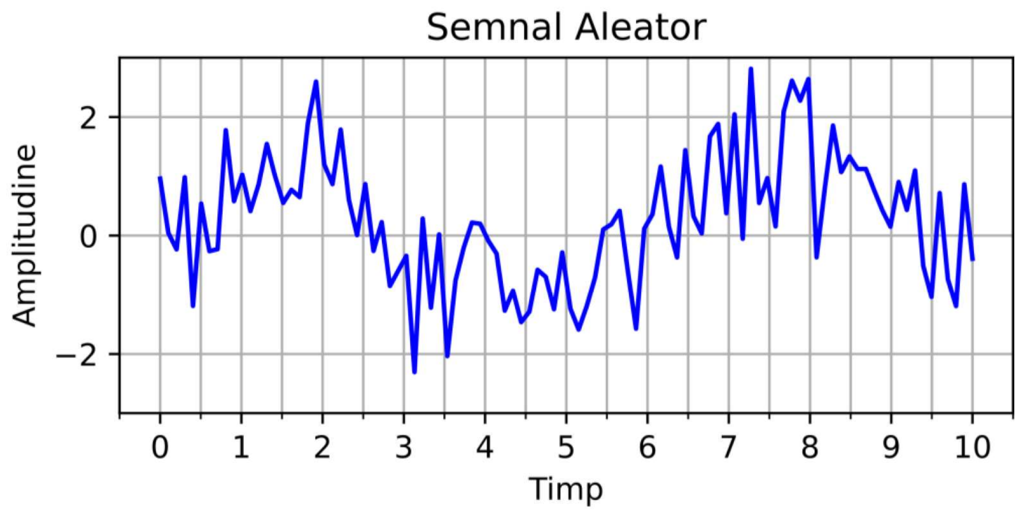
Dacă amplitudinea semnalului poate fi determinată la orice moment de timp, inclusiv în viitor, pe baza unei expresii matematice, spunem despre acest semnal că este un **semnal determinist.**

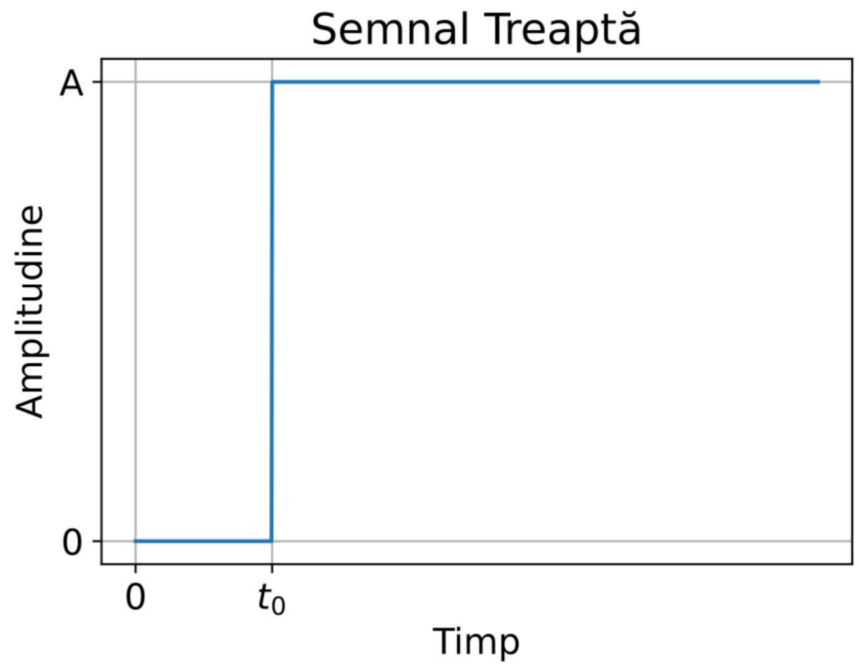
$x = \sin(t)$



Definiție

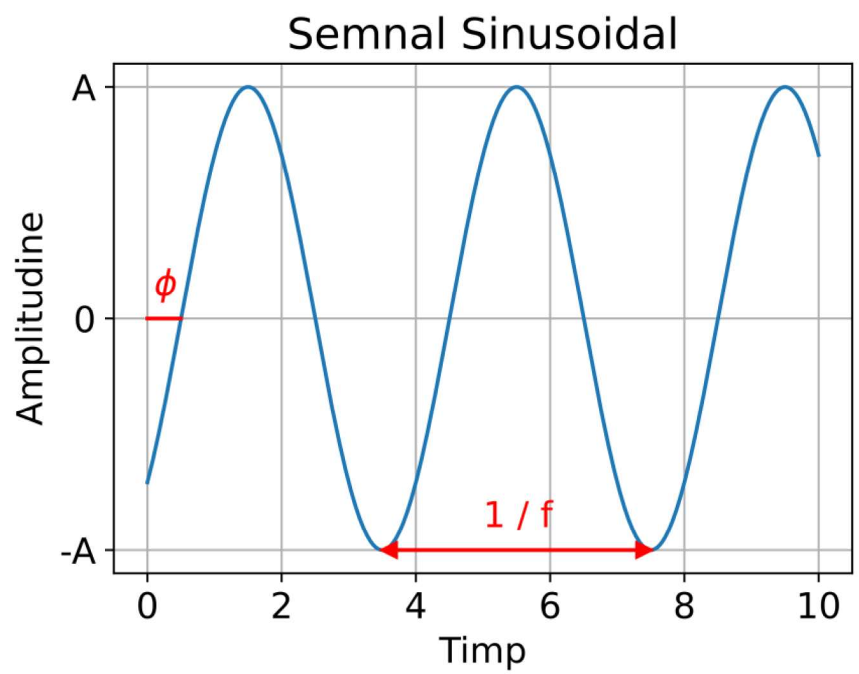
Dacă evoluția semnalului nu poate fi exprimată matematic, semnalul este un **semnal aleator.**





1. **Semnalul Treaptă** – este caracterizat de amplitudinea A și momentul de început t_0 al treptei

$$\Gamma(t) = \begin{cases} 0, t < t_0 \\ A, t \geq t_0 \end{cases}$$



2. **Semnalul sinusoidal** – este caracterizat de frecvența f, amplitudinea A și defazajul φ

$$y(t) = A \cdot \sin(2 \cdot \pi \cdot f \cdot t + \varphi)$$

Introducere

Clasificarea semnalelor

Eșantionarea semnalelor

Cuantificarea semnalelor

Introducere

Eșantionarea semnalelor

Clasificarea semnalelor

Cuantificarea semnalelor

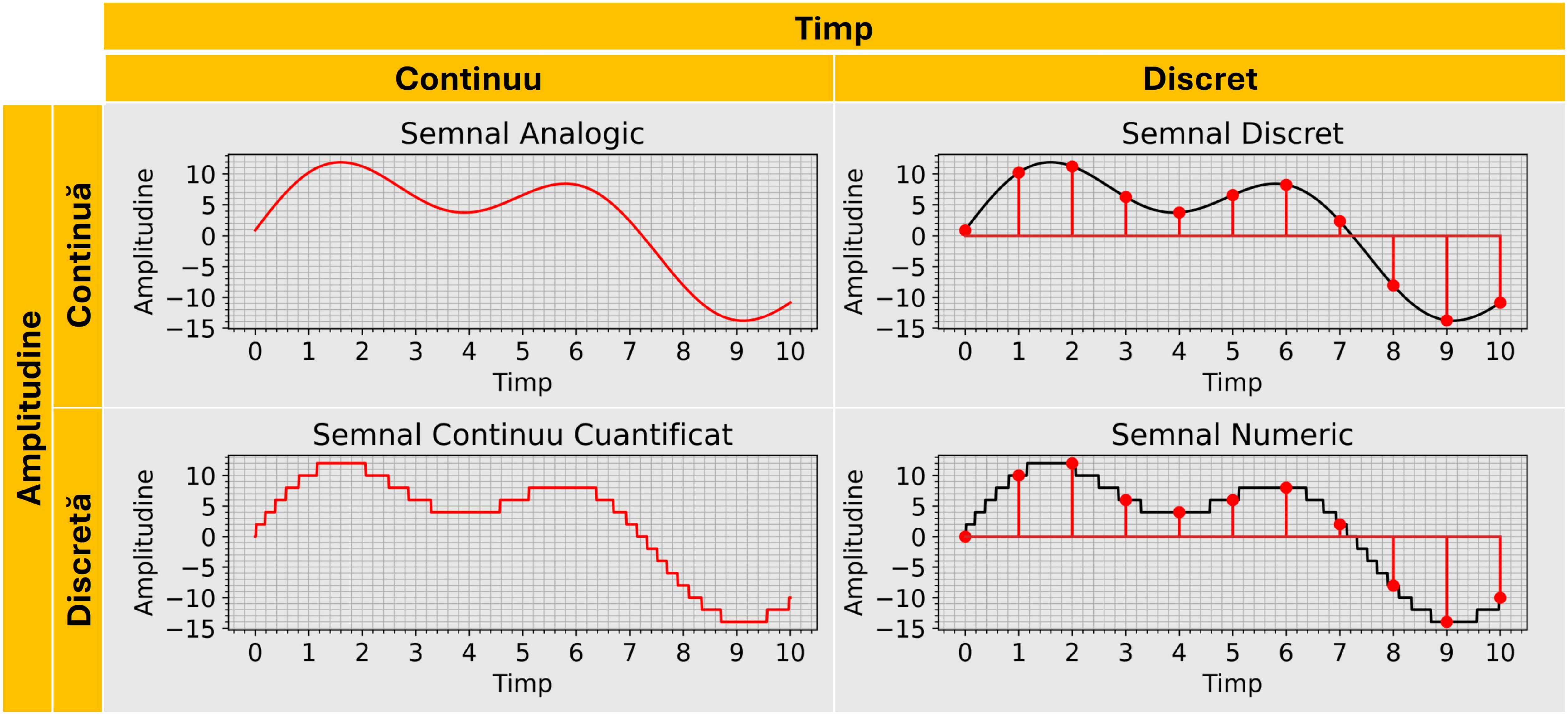
În funcție de evoluția semnalelor în timp, semnalele pot fi:

1. **semnale continue** dacă evoluția este dată de o funcție continuă
2. **semnale discrete** dacă amplitudinea semnalului este cunoscută doar pentru momente discrete de timp.

În funcție de evoluția semnalelor în amplitudine, acestea pot fi:

1. **semnale cuantificate continuu** a căror evoluție în timp a amplitudinii este continua;
2. **semnale cuantificate discret** a căror valori ale amplitudinii pot aparține unui set finit de valori.

		Timp	
		Continuu	**Discret**
Amplitudine	**Continuă**	Semnale continue (semnale analogice)	Semnale discrete
	Discretă	Semnale continue cuantificate	Semnale discrete cuantificate (semnale numerice)


Timp
Continuu
Discret
Amplitudine
Continuă
Discretă
Semnal Analogic
Semnal Discret
Semnal Continuu Cuantificat
Semnal Numeric
Amplitudine
Timp

Introducere

Eșantionarea semnalelor

Clasificarea semnalelor

Cuantificarea semnalelor

Introducere

Clasificarea semnalelor

Eșantionarea semnalelor

Cuantificarea semnalelor

Definiție

Eșantionarea reprezintă discretizarea în timp a unui semnal în timp continuu.

		Timp		
		Continuu		**Discret**
Amplitudine	**Continuă**	Semnale continue (semnale analogice)	**EȘANTIONARE** →	Semnale discrete
	Discretă	Semnale continue cuantificate		Semnale discrete cuantificate (semnale numerice)

Definiție

Eșantionarea reprezintă discretizarea în timp a unui semnal în timp continuu.

Rezultatul eșantionării unui semnal continuu $x(t)$ este un șir de eșantioane $x[n]$, un **eșantion** fiind valoarea semnalului $x(t)$ la momentul $t = n \cdot T$

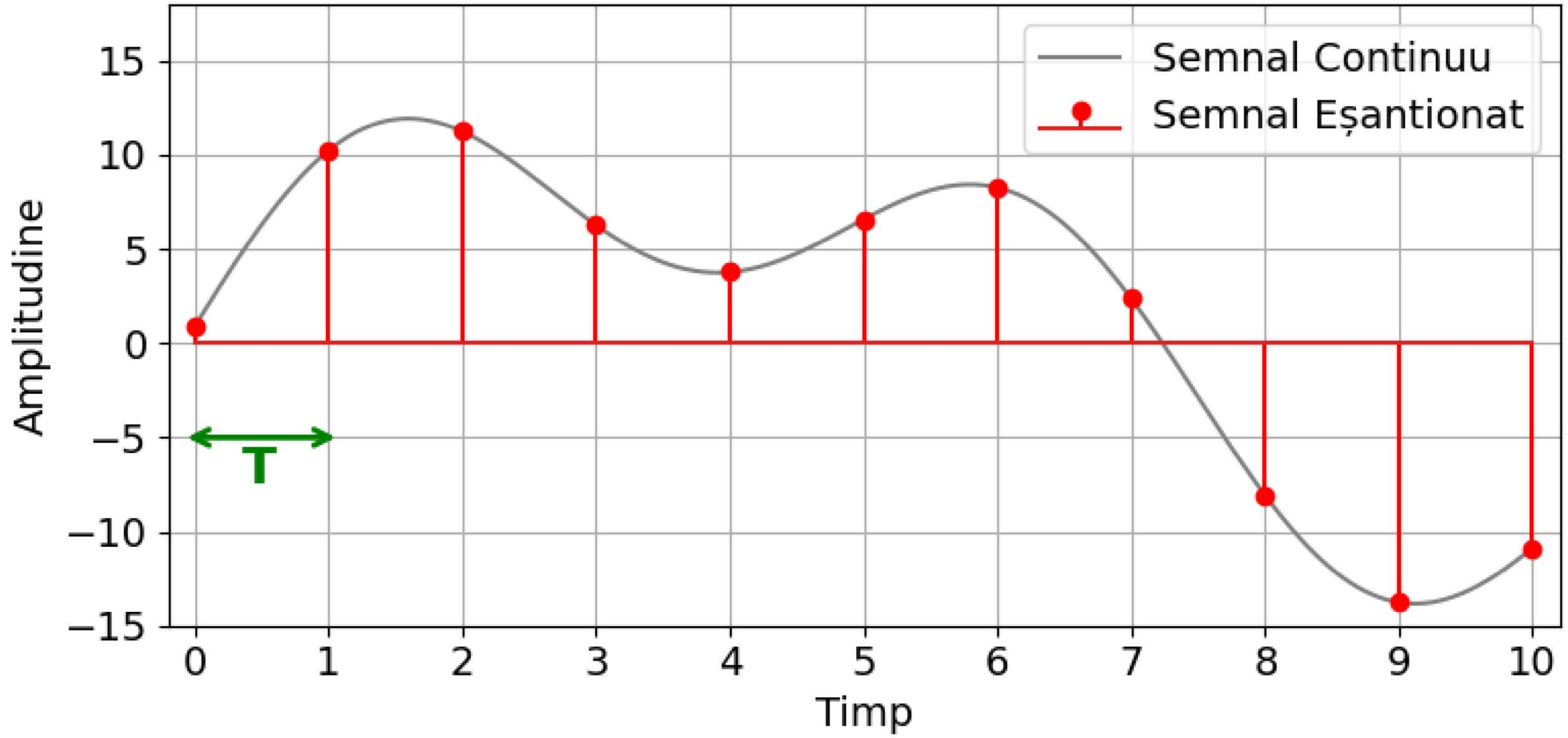
Definiție

Timpul dintre două eșantioane succesive (notat cu T) se numește **perioadă de eșantionare**

Definiție

Numărul de eșantioane obținute într-o unitate de timp (secundă) se numește **frecvență de eșantionare**, se notează cu f_s și se măsoară în Hz (Hertzi)

$$f_s = \frac{1}{T}\ [Hz]$$



Dacă dorim să avem o acuratețe cât mai mare de informație, atunci este recomandat să alegem un număr cât mai mare de eșantioane (deci o frecvență de eșantionare mai mare)

TREBUIE SĂ CUNOAȘTEM:

1. durata semnalului (t_f)
2. frecvența semnalului (f_s) sau perioada de eșantionare (T)
3. expresia matematică ce descrie semnalul

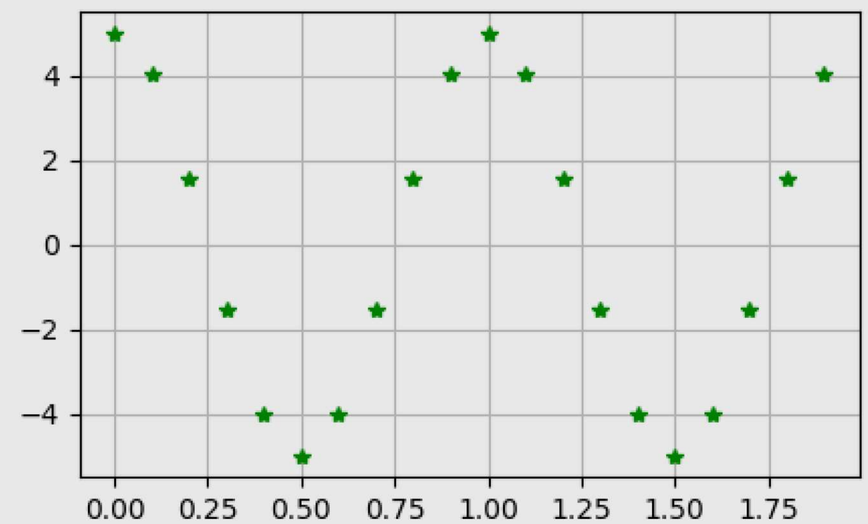
EXEMPLU

dorim generarea și afișarea grafică a eșantioanelor unui semnal sinusoidal continuu descris de $x(t) = 5 \cdot sin\left(2 \cdot \pi \cdot t + \frac{\pi}{2}\right)$, eșantionat cu frecvența de eșantionare $f_s = 10\ Hz$, pe o perioadă de 2 secunde

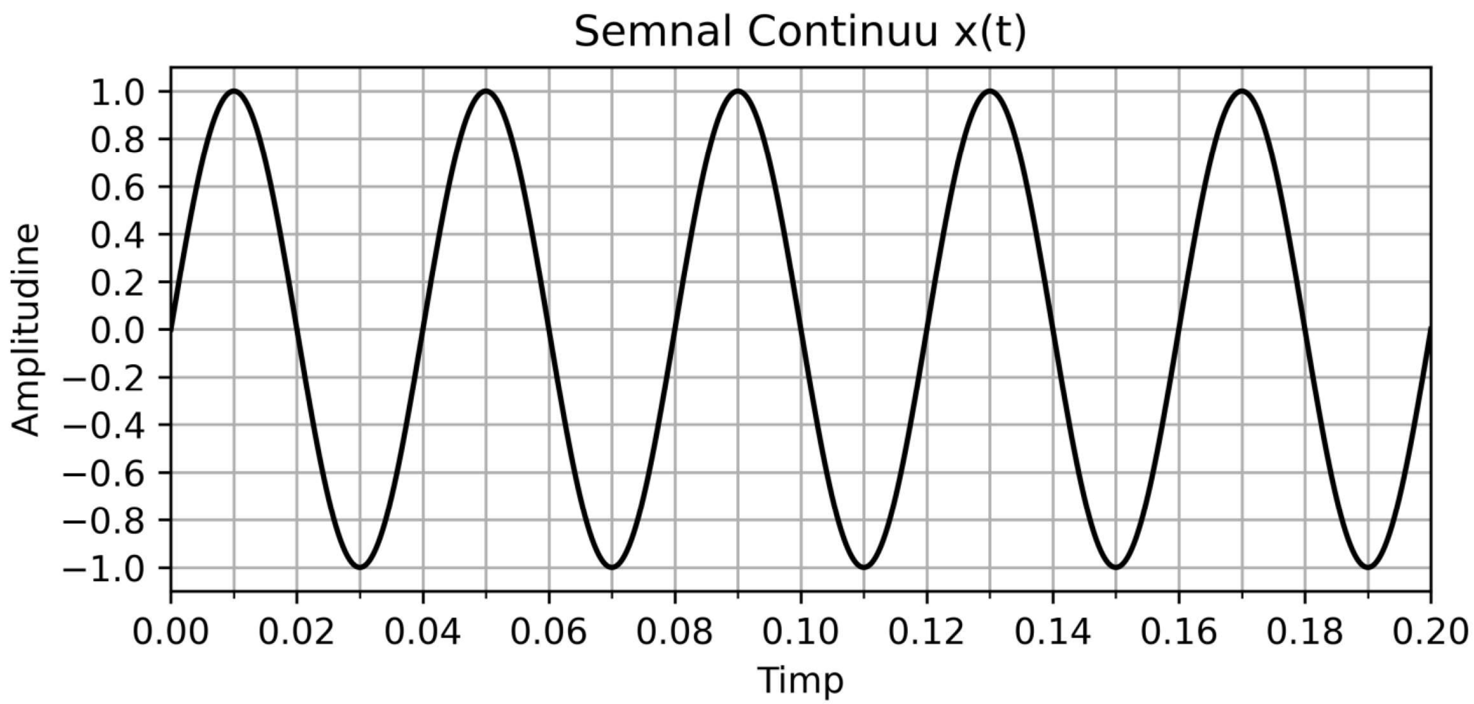
main.py

```
import numpy as np
import matplotlib.pyplot as plt

tf = 2   # durata semnalului
fs = 10  # frecventa de esantionare

t = np.arange(0, tf+1/fs, 1/fs) # generam vectorul de timp
x = 5 * np.sin(2 * np.pi * t + np.pi / 2) # generam esantioanle

plt.plot(t, x, “g*") # afisam grafic semnalul esantionat
plt.grid()
plt.show()
```

rezultat

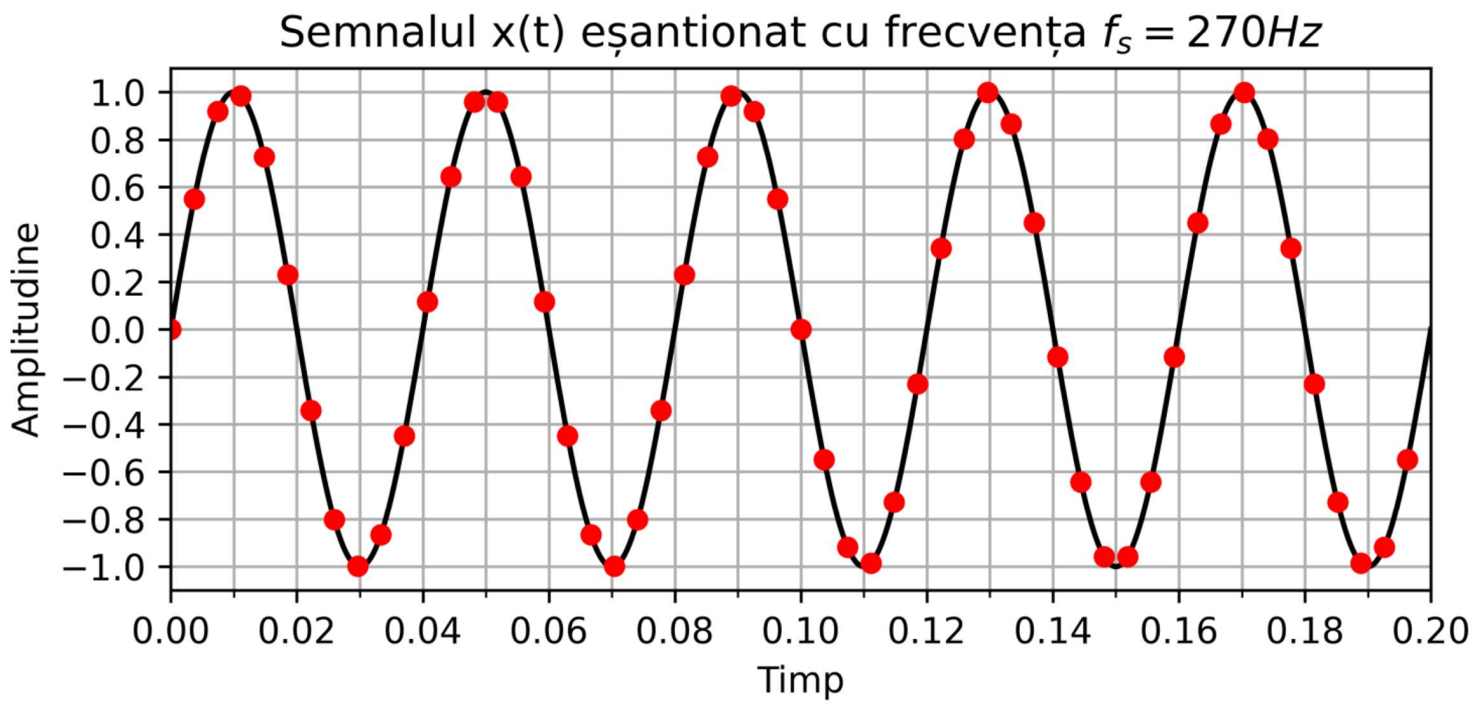


- Fie un semnal sinusoidal $x(t)$ având frecvența de $25\ Hz$.



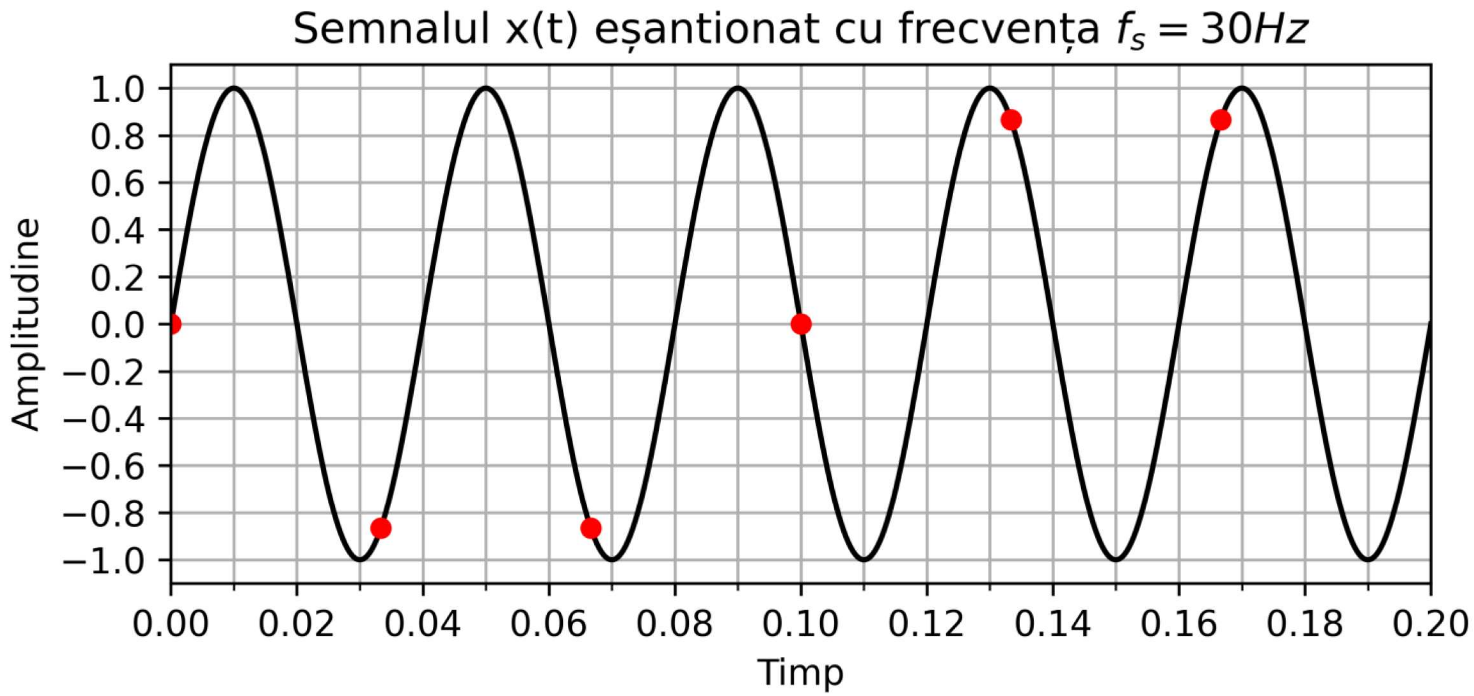
- Fie un semnal sinusoidal $x(t)$ având frecvența de $25\ Hz$.

- Dacă semnalul a fost eșantionat cu o frecvență $f_{s_1} = 270\ Hz$ observăm că putem reconstrui semnalul continuu pe baza informațiilor furnizate de eșantioane.



- Fie un semnal sinusoidal $x(t)$ având frecvența de $25\ Hz$.

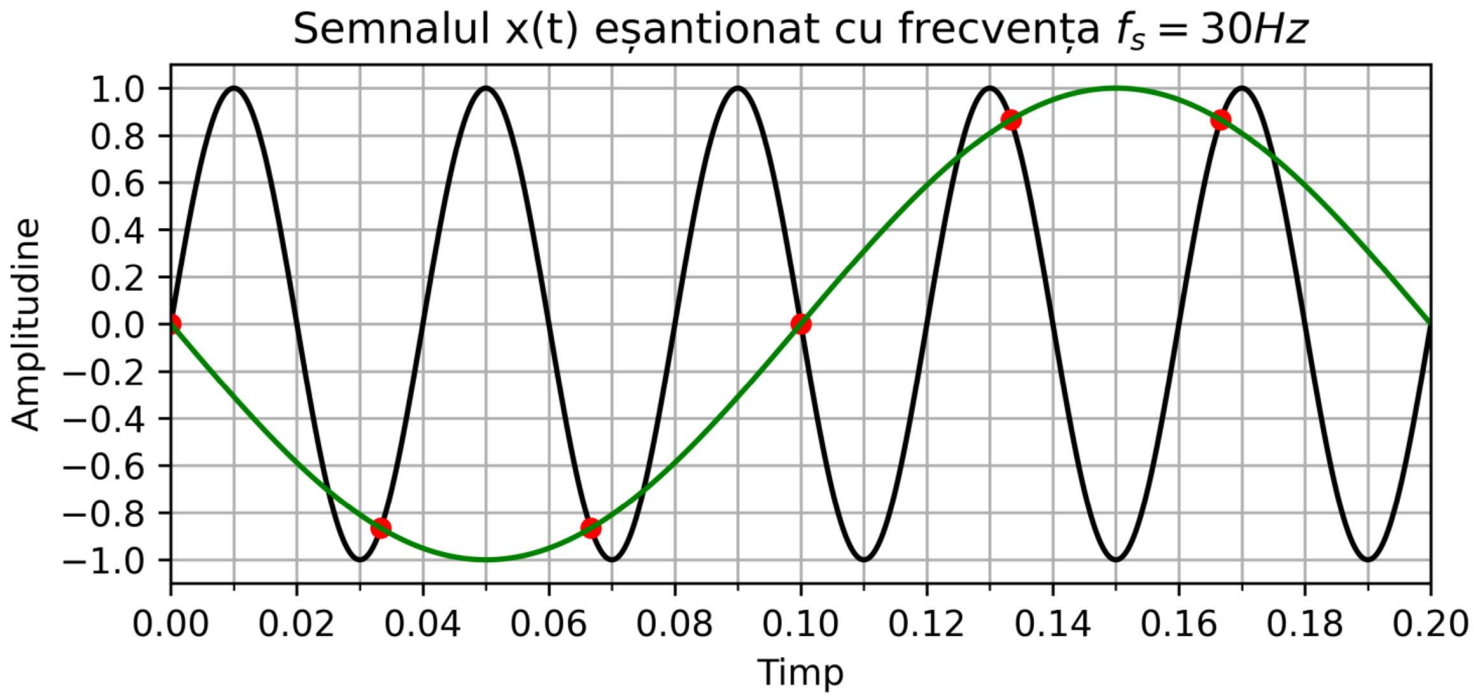
- Dacă semnalul a fost eșantionat cu o frecvență $f_{s_2} = 30\ Hz$ informațiile furnizate de eșantioane nu sunt suficiente pentru reconstruirea semnalului.



- Fie un semnal sinusoidal $x(t)$ având frecvența de $25\ Hz$.

- Dacă semnalul a fost eșantionat cu o frecvență $f_{s_2} = 30\ Hz$ informațiile furnizate de eșantioane nu sunt suficiente pentru reconstruirea semnalului.

- Mai mult, aceleași eșantioane rezultă atât din eșantionarea semnalului $x(t)$, cât și din eșantionarea unui alt semnal sinusoidal de frecvență $5\ Hz$ – apare fenomenul de **„aliere” – „aliasing”**



Teoremă

Teorema eșantionării (Shannon): Pentru ca un semnal $x(t)$ de spectru mărginit să poată fi reconstituit complet din eșantioanele sale, eșantionarea trebuie să respecte condiția ***Nyquist*** și anume, frecvența de eșantionare f_s trebuie să fie cel puțin dublul frecvenței maxime din spectrul semnalului: $f_s \geq 2 \cdot f_{\max}$.

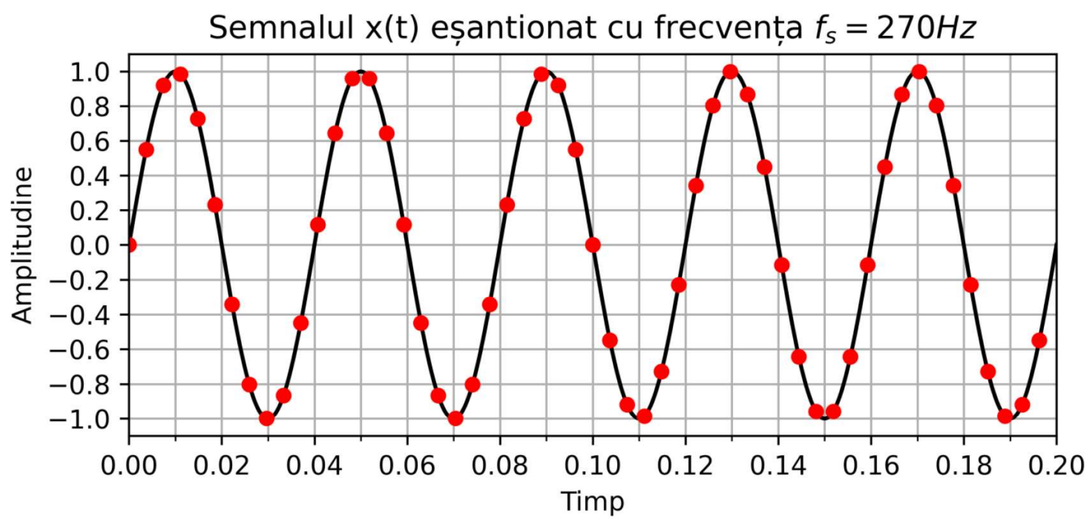
Teoremă

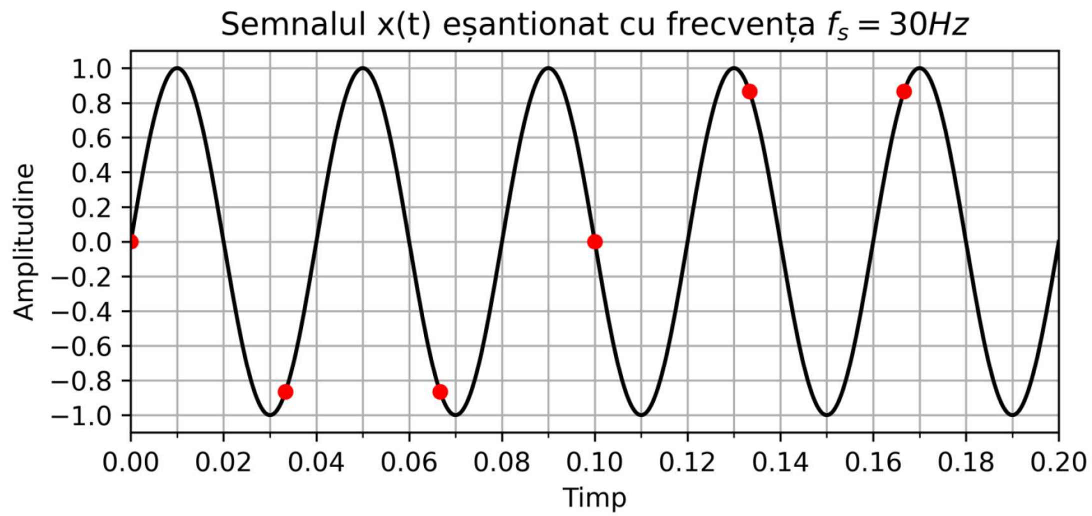
Teorema eșantionării (Shannon): Pentru ca un semnal $x(t)$ de spectru mărginit să poată fi reconstituit complet din eșantioanele sale, eșantionarea trebuie să respecte condiția ***Nyquist*** și anume, frecvența de eșantionare f_s trebuie să fie cel puțin dublul frecvenței maxime din spectrul semnalului: $f_s \geq 2 \cdot f_{\max}$.

- Fie un semnal sinusoidal $x(t)$ având frecvența de $25\ Hz$.

- Dacă semnalul a fost eșantionat cu o frecvență $f_{s_1} = 270\ Hz$, se respectă teorema eșantionării ($270\ Hz \geq 2 \cdot 25\ Hz$).

- Dacă semnalul a fost eșantionat cu o frecvență $f_{s_2} = 30\ Hz$ NU se respectă teorema eșantionării ($30\ Hz \ngeq 2 \cdot 25\ Hz$).





Teoremă

Reeșantionarea (resampling) reprezintă procesul de schimbare a frecvenței de eșantionare a unui semnal de la f_{s_1} la f_{s_2}.

- Reeșantionarea se realizează prin calcularea valorilor noilor eșantioane folosind o metodă de interpolare (liniară, pătratică, Lagrange etc.)

- Trebuie avut grijă ca și noua frecvență de eșantionare să respecte criteriul ***Nyquist*** pentru a evita fenomenul de aliere.

Teoremă

Reeșantionarea (resampling) reprezintă procesul de schimbare a frecvenței de eșantionare a unui semnal de la f_{s_1} la f_{s_2}.

INTERPOLATORUL LINIAR

- Dându-se două eșantioane succesive, interpolatorul liniar este dreapta ce trece prin cele două puncta definite de eșantioane $((t[i], x[i])$ și $(t[i+1], x[i+1]))$

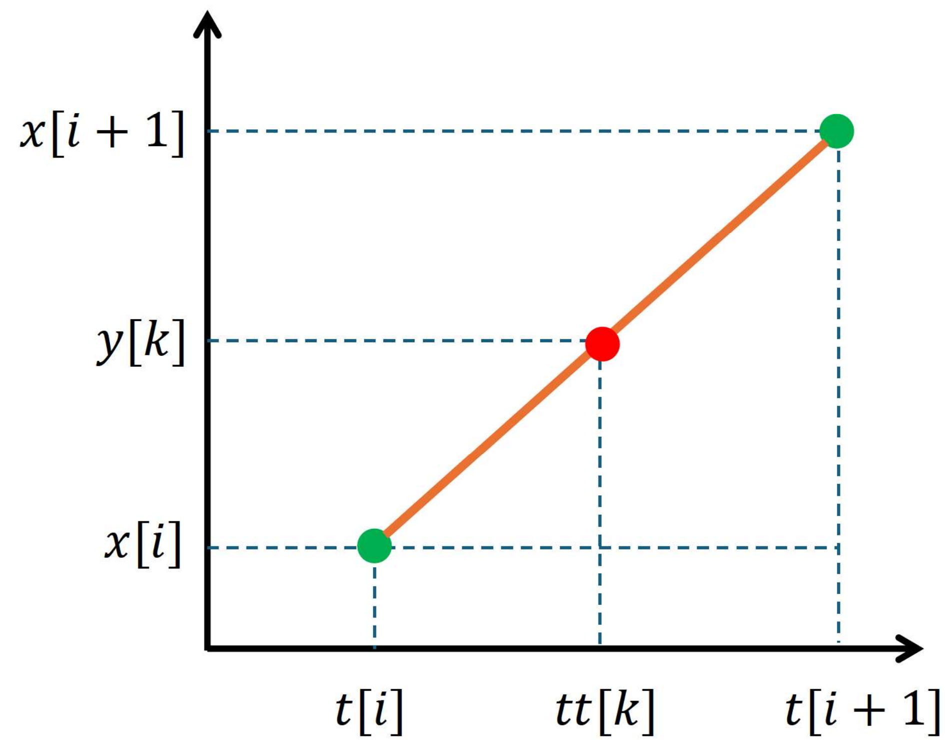
- Pentru un $tt[k] \in [t[i], t[i+1]]$ se poate determina valoarea lui $y[k]$ aflând astfel valoarea semnalului la un moment de timp aflat între două eșantioane succesive.

- Ținând cont de **Teorema lui Thales** putem scrie ecuația:

$$\frac{y[k] - x[i]}{x[i+1] - x[i]} = \frac{tt[k] - t[i]}{t[k+1] - t[i]}$$

- De unde putem scoate valoarea lui $y[k]$:

$$\boxed{y[k] = x[i] + (x[i+1] - x[i]) \cdot \frac{tt[k] - t[i]}{t[i+1] - t[i]}}$$



$$y[k] = x[i] + (x[i+1] - x[i]) \cdot \frac{tt[k] - t[i]}{t[i+1] - t[i]}$$

INTERPOLATORUL LINIAR

Vom defini o funcție Python ce implementează metoda de interpolare liniară pentru creșterea frecvenței de eșantionare de la f_{s_1} la f_{s_2}. Funcția va avea ca intrări:

- t – vectorul de timp al semnalului eșantionat cu f_{s_1}

```
t = np.arange(0, tf, 1/fs1)
```

- x – eșantioanele semnalului eșantionat cu f_{s_1}
- tt – vectorul de timp al semnalului eșantionat cu f_{s_2}

```
tt = np.arange(0, tf, 1/fs2)
```

și ca ieșire y – eșantioanele semnalului eșantionat cu f_{s_2}.

Funcție

```
import numpy as np
import matplotlib.pyplot as plt


def interpolate_linear(t, x, tt):
    N = len(t)
    NN = len(tt)
    y = np.zeros((NN,))

    for k in range(0, NN):
        for i in range(0, N - 1):
            if (t[i] <= tt[k] and tt[k] <= t[i + 1]) or i == N - 2:
                y[k] = x[i] + (x[i + 1] - x[i]) / (t[i + 1] - t[i]) * (tt[k] - t[i])
                break
    return y
```

INTERPOLATORUL LINIAR

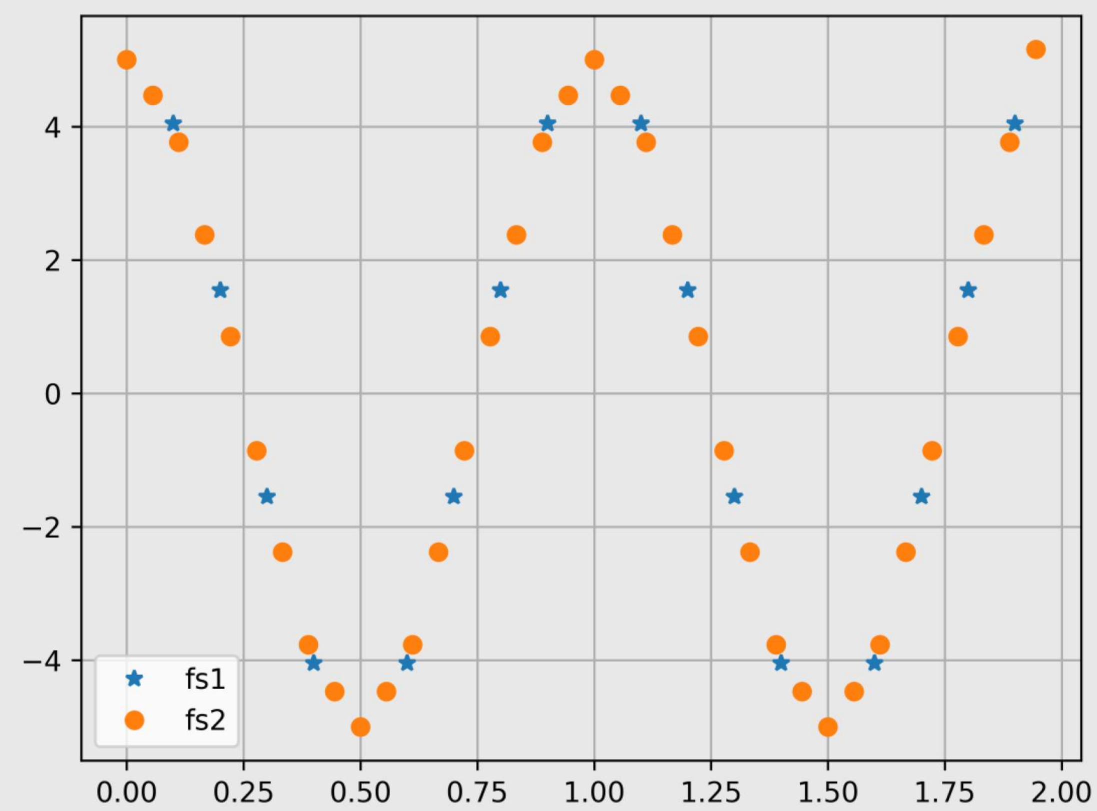
$$y[k] = x[i] + (x[i+1] - x[i]) \cdot \frac{tt[k] - t[i]}{t[i+1] - t[i]}$$

EXEMPLU DE UTILIZARE A FUNCȚIEI IMPLEMENTATE

main.py

```
import numpy as np
import matplotlib.pyplot as plt

def interpolate_linear(t, x, tt):
    ... # continutul functiei

tf = 2
fs1, fs2 = 10, 18

t = np.arange(0, tf, 1 / fs1)
x = 5 * np.sin(2 * np.pi * t + np.pi / 2)
tt = np.arange(0, tf, 1 / fs2)
y = interpolate_linear(t, x, tt)

plt.plot(t, x, "*", tt, y, 'o')
plt.grid()
plt.legend(["fs1", "fs2"])
plt.show()
```

rezultat



INTERPOLATORUL LAGRANGE

Teoremă

Teorema interpolării: Se dau N puncte distincte $(t[0], x[0]) \dots (t[N-1], x[N-1])$. Există o ecuație polinomială de gradul maxim $N-1$ care interpolează punctele date.

- Ecuația de interpolare polinomială în forma Lagrange se definește ca combinația liniară:

$$\boxed{y[k] = \sum_{i=0}^{N-1} x[i] \cdot \ell(tt[k])}$$

de ecuații polinomiale Lagrange:

$$\boxed{\ell(tt[k]) = \prod_{m=0, m\neq i}^{N-1} \frac{tt[k] - t[m]}{t[i] - t[m]}}$$

Introducere

Clasificarea semnalelor

Eșantionarea semnalelor

Cuantificarea semnalelor

Introducere

Clasificarea semnalelor

Eșantionarea semnalelor

Cuantificarea semnalelor

Definiție

Cuantizarea reprezintă discretizarea în amplitudine a unui semnal cu amplitudine continuă.

		Timp	
		Continuu	**Discret**
Amplitudine	**Continuă**	Semnale continue (semnale analogice)	Semnale discrete
		CUANTIZARE	
	Discretă	Semnale continue cuantificate	Semnale discrete cuantificate (semnale numerice)

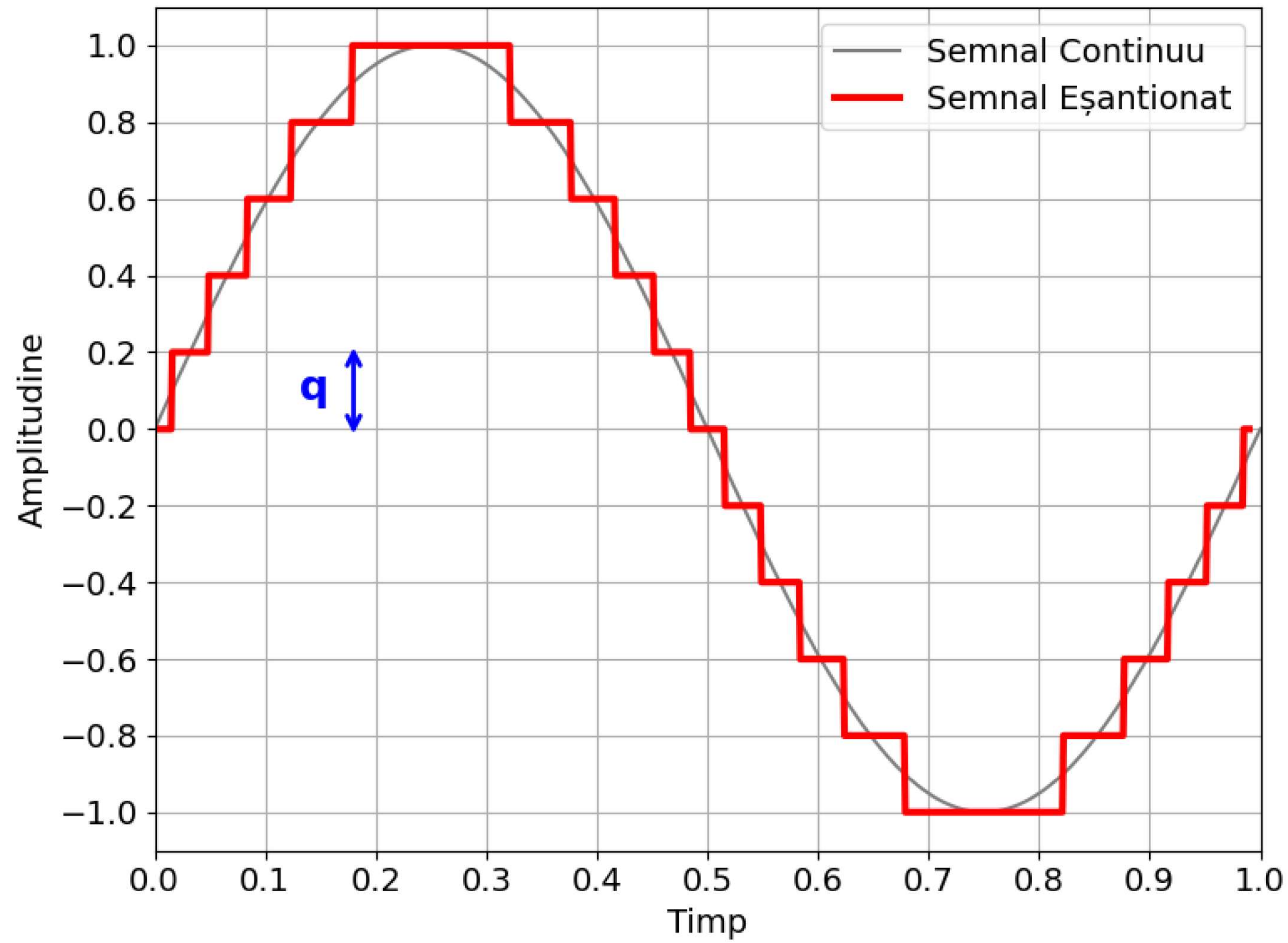
Definiție

Cuantizarea reprezintă discretizarea în amplitudine a unui semnal cu amplitudine continuă.

Prin cuantizare fiecărui moment de timp i se alocă o valoare discretă dintr-un set finit de valori

Definiție

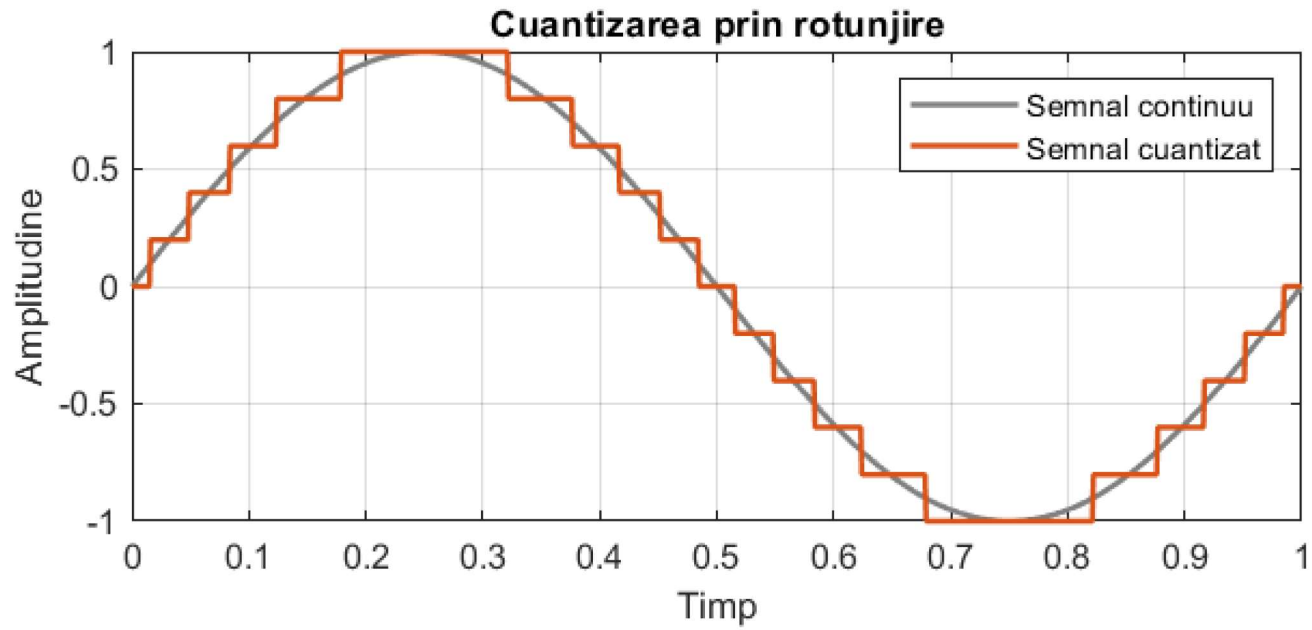
Distanța dintre două valori consecutive din setul de valori se numește **pas de cuantizare** și se notează cu q



Cele mai folosite metode de cuantizare sunt:

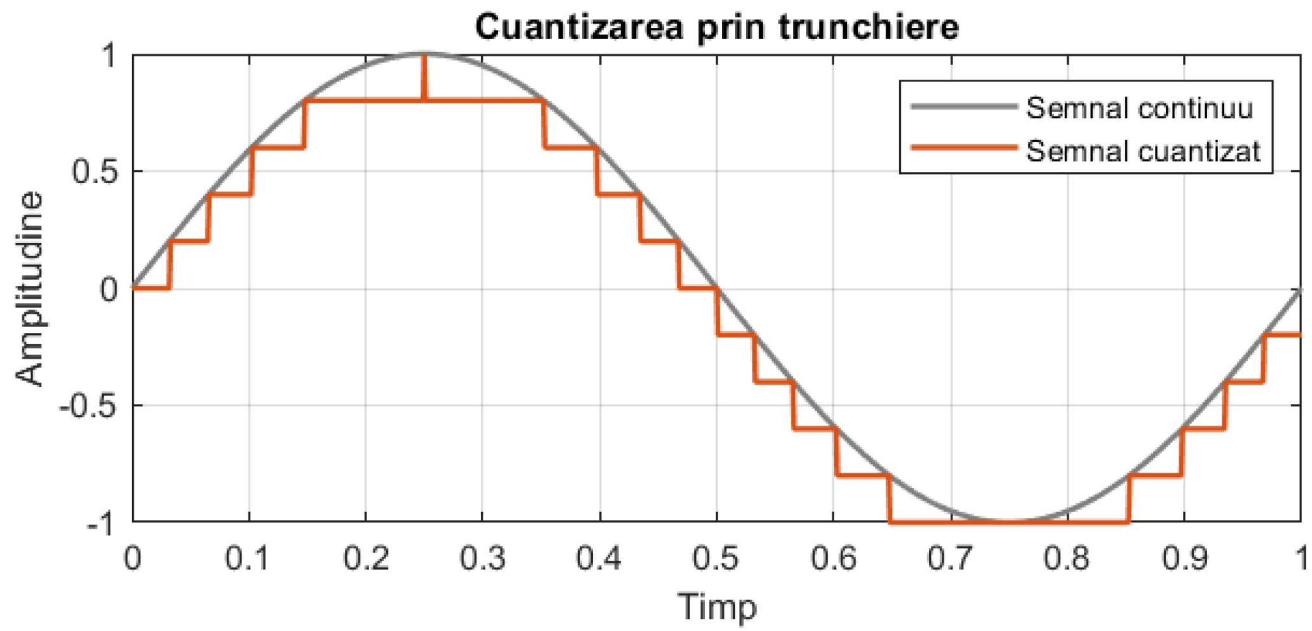
1. **Cuantizarea prin rotunjire** – valoarea discretă va fi cel mai apropiat nivel de cuantizare disponibil de valoarea reală

```
xc = np.round(x / q) * q
```



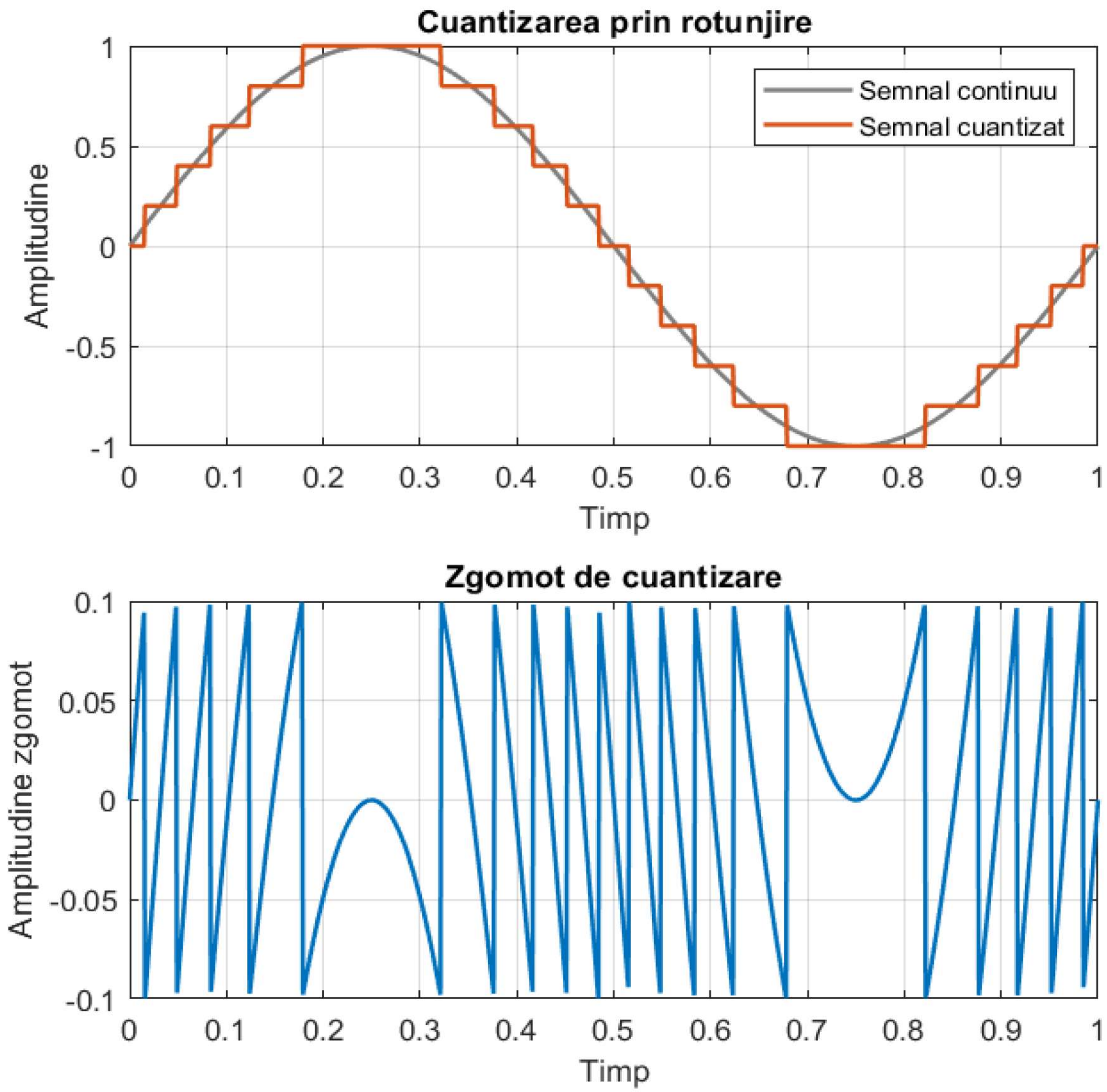
2. **Cuantizarea prin trunchiere** – valoarea discretă va fi cel mai apropiat nivel de cuantizare disponibil valoric inferior valorii reale

```
xc = np.floor(x / q) * q
```



Definiție

Zgomotul de cuantizare este semnalul obținut făcând diferența dintre semnalul cuantizat și semnalul original: $\varepsilon[t] = x_c\,[t] - x[t]$



Pentru a evalua zgomotul de cuantizare vom folosi **eroarea pătratică medie**:

$$\varepsilon_m = \frac{\sum_{t=1}^{N} \varepsilon[t]^2}{N}$$

Un eșantion cuantizat are o valoare dintr-un set finit de valori → putem descrie un eșantion precizând doar nivelul de cuantizare.

EXEMPLU: Se dau următoarele niveluri de cuantizare: $\{0, \quad 0.5, \quad 1, \quad 1.5\}$

Valoare analogică x(t)	Valoare cuantizată prin rotunjire	Nivel de cuantizare	Nivel de cuantizare (binar)
$x(t) \leq 0.25$	0	Nivel 0	00
$0.25 < x(t) \leq 0.75$	0.5	Nivel 1	01
$0.75 < x(t) \leq 1.25$	1	Nivel 2	10
$1.25 < x(t)$	1.5	Nivel 3	11

- N_q niveluri de cuantizare se pot reprezenta pe minim $b = \lceil log_2 N_q \rceil$ biți
- Având la dispoziție b biți, se pot reprezenta un număr maxim de $N_q = 2^b$ niveluri de cuantizare

Introducere

Clasificarea semnalelor

Eșantionarea semnalelor

Cuantificarea semnalelor

Introducere

Clasificarea semnalelor

Eșantionarea semnalelor

Cuantificarea semnalelor