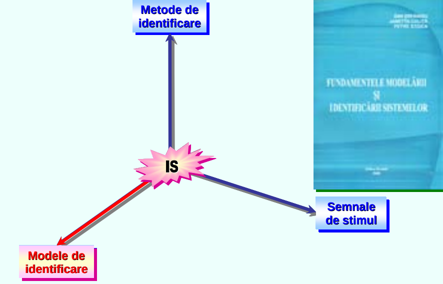
❷ Modele de identificare



❷.❶ Scurtă clasificare

M

☞ în acest curs

Analitice | **Experimentale**

- Au un caracter mai mult teoretic şi se obţin, de regulă, prin aplicarea legilor fizico-chimice care caracterizează funcţionarea unui proces.
- Dovedesc o mare aplicabilitate practică, sunt destul de generale, şi precise, dar, de multe ori, ridică pragul de complexitate dincolo de o limită admisibilă.

- Se obţin din seturi de date furnizate de către proces şi sunt, în general, uşor de implementat.
- Suferă de lipsă de generalitate, fiind specifice anumitor tipuri de procese sau chiar puncte de funcţionare ale aceluiaşi proces.
- Precizia lor este modestă.

M

în acest curs ☞

Liniare | **Neliniare**

- Procesele local liniarizabile în jurul anumitor puncte de funcţionare sunt cele mai răspîndite.
- Modelele matematice liniare, eventual cu parametri variabili în timp reprezintă adesea procesele liniare sau local liniarizabile.

- Majoritatea proceselor din natură sunt neliniare, dar cu diferite grade de neliniaritate.
- Procesele profund neliniare sunt dificil de identificat, ele necesitînd modele matematice relativ complexe, avînd tot caracter neliniar.

❷ Modele de identificare

❷.❶ Scurtă clasificare

$\mathcal{M}$

dar mai mult ☟

Parametrice

☜ în acest curs ☞

Neparametrice

- Mai precise.

$\mathcal{M}(\theta)$

- Mai puţin precise.
- Oferă mai mult descrieri calitative ale procesului de interes, dar pot fi utilizate pentru a stabili anumite caracteristici ale experimentului de identificare parametrică desfăşurat ulterior (întîrziere intrinsecă, frecvenţă de eşantionare, numărul maxim de parametri, etc.).

$\mathcal{M}$

Invariante în timp

☜ în acest curs ☞

Variabile în timp

$\mathcal{M}$

Single **I**nput **S**ingle **O**utput

Multi(ple) **I**nput(s) **M**ulti(ple) **O**utput(s)

în acest curs ☞

SISO

MIMO

☜ dar şi

SIMO

MISO

Single **I**nput **M**ulti(ple) **O**utput(s)

Multi(ple) **I**nput(s) **S**ingle **O**utput

❷ Modele de identificare

❷.❶ Scurtă clasificare

M

În timp continuu

☝ în acest curs

În timp discret

- Procesele care trebuie identificate au de regulă o evoluţie în timp continuu.
- Sunt mai degrabă modele analitice decît experimentale.
- Pentru procesele de complexitate redusă, este posibilă determinarea parametrilor funcţiilor de transfer (numiţi şi parametri fizici) sau chiar a parametrilor de stare.

- De regulă, sunt modele experimentale.
- Uşor de implementat şi de determinat..

M

în acest curs ☝

Cu parametri concentraţi

Cu parametri distribuiţi

- Cu un număr finit de parametri.

- Cu un număr infinit de parametri.

M

În domeniul timpului

☜ în acest curs ☞

În domeniul frecvenţei

❷.❷ Noţiuni de Statistică şi Prelucrare de Semnal

☝ Colecţiile de date măsurate sunt semnale avînd un caracter nedeterminist (stocastic).

Modelele de identificare construite plecînd de la astfel de date au natură statistică.

- Pentru a putea caracteriza proprietăţile modelelor de identificare sunt necesare cîteva concepte elementare de Statistică şi de Prelucrare de Semnal (PS).

Densitate de probabilitate

(distribuţie de probabilitate)

- Distribuţii de probabilitate uzuale:

Gaussiană | Student | Laplaciană | Pearson

$$\mathcal{p} : \mathcal{D}(x) \rightarrow [0,1]$$

domeniul de variaţie al variabilei aleatoare x

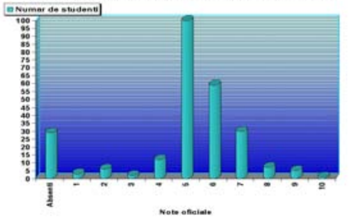
$$\int_{\mathcal{D}(x)} \mathcal{p}(x)\,dx = 1$$

$\mathcal{D}(x)$ cel mult numărabil

$\mathcal{p}$ frecvenţă de apariţie

$\mathcal{D}(x)$ finit

h histogramă



$$\mathcal{p} \equiv \frac{h}{\#\mathcal{D}(x)}$$

frecvenţă de apariţie

cardinalul domeniului de variaţie (numărul de valori)

Probabilitate

$$\mathcal{P}(x \in \mathcal{S}) \stackrel{def}{=} \int_{\mathcal{S}} \mathcal{p}(x)\,dx$$

probabilitatea ca x să aparţină submulţimii $\mathcal{S}$ din domeniul de variaţie

Apartenenţa unei variabile aleatoare la domeniul său de variaţie este un eveniment sigur.

☝ Frecvenţa de apariţie este adesea considerată o aproximare a densităţii de probabilitate.

❷ Modele de identificare

❷.❷ Noţiuni de Statistică şi Prelucrare de Semnal

Tipuri de densităţi de probabilitate utilizate în IS

Densitatea simplă de probabilitate

$$\wp(u[n],n) \overset{def}{=} \wp(u[n]) \in [0,1]$$

$$\wp(y[n],n) \overset{def}{=} \wp(y[n]) \in [0,1]$$

Probabilitatea ca valoarea intrării, respectiv ieşirii să aparţină setului de date măsurate la momentul $n \in \mathbb{N}^*$.

Densitatea de probabilitate încrucişată (intrare-ieşire)

$$\wp(u[m],y[n],m,n) \overset{def}{=} \wp(u[m],y[n]) \in [0,1]$$

Probabilitatea ca intrarea $u[m]$ să producă ieşirea $y[n]$ în datele măsurate, la momentele $m \in \mathbb{N}^*$ şi $n \in \mathbb{N}^*$.

Densitatea de probabilitate a aceleiaşi realizări

$$\wp(y[m],y[n],m,n) \overset{def}{=} \wp(y[m],y[n]) \in [0,1]$$

Probabilitatea ca valorile $y[m]$ şi $y[n]$ ale ieşirii procesului să aparţină aceleiaşi realizări, la momentele $m \in \mathbb{N}^*$ şi $n \in \mathbb{N}^*$.

☝ Cunoaşterea preliminară a acestor densităţi de probabilitate este dificilă (dacă nu imposibilă).

TLC

✍ Convenţie: distribuţiile de probabilitate ale proceselor vor fi implicit considerate de tip Gaussian, dacă nu se specifică altfel.

❷.❷ Noţiuni de Statistică şi Prelucrare de Semnal

Medie statistică
(valoarea cea mai aşteptată, speranţa matematică)

$$E\{y[n]\} \overset{def}{=} \int_{\mathcal{D}(y[n])} \boxed{p(y[n]) \cdot y[n]}\, dy[n] \quad \forall n \in \mathbb{N}^*$$

produsul dintre valoarea densităţii de probabilitate şi valoarea corespunzătoare a variabilei aleatoare

Operator de mediere statistică

E

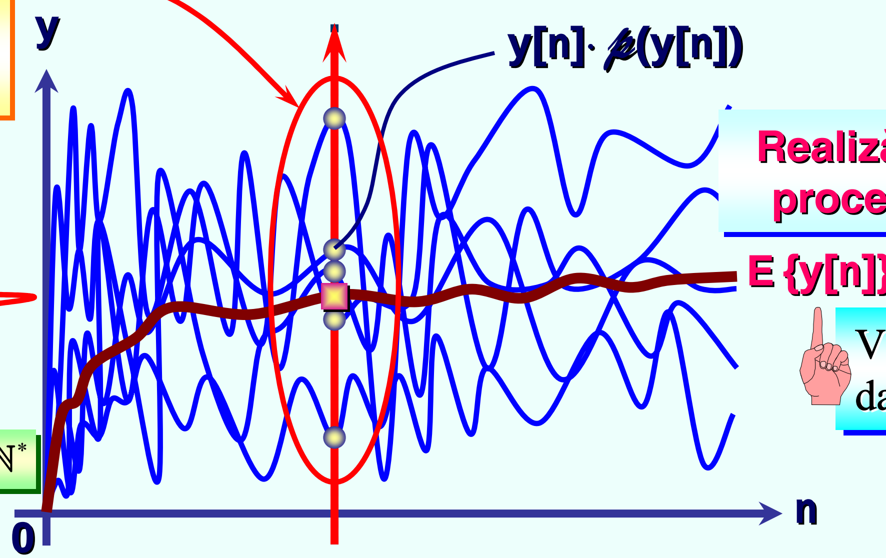
Expected

Speranţă matematică?

Folosind media statistică a unui set de date aleatoare (cu o anumită distribuţie de probabilitate), se poate obţine o predicţie (de regulă grosieră) a valorii care s-ar adăuga datelor disponibile la momentul următor de achiziţie.

☝ Timpul joacă un rol secundar în definiţia operatorului E, deoarece integrarea se efectuează peste mulţimea valorilor variabilei aleatoare şi nu de-a lungul axei timpului.

Direcţia de integrare în definiţia operatorului de mediere statistică



Realizări ale procesului

Variabilă în timp, dar deterministă.

Proprietăţi elementare ale operatorului de mediere statistică

Exerciţii

➔ Invarianţa variabilelor deterministe

$$E\{a[n]\} = a[n] \quad \forall n \in \mathbb{N}^*$$

➔ Liniaritate

$$E\{a_1[n]y_1[n] + a_2[n]y_2[n]\} = a_1[n]E\{y_1[n]\} + a_2[n]E\{y_2[n]\} \quad \forall n \in \mathbb{N}^*$$

❷ Modele de identificare

❷.❷ Noţiuni de Statistică şi Prelucrare de Semnal

Exemplu

Revenim la răspunsul indicial al unui sistem liniar de ordin I

Parametrii procesului pot fi determinaţi cu o anumită precizie folosind curba medie statistică a tuturor realizărilor acestuia.

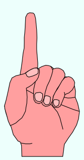
• Inconvenientele utilizării directe a definiţiei operatorului de mediere statistică în aplicaţii practice:

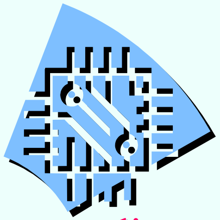
☹ De cele mai multe ori, nu se cunoaşte densitatea de probabilitate a datelor măsurate.

Ar putea fi estimată prin iniţierea unui mare număr de experimente econometrice şi trasarea histogramelor valorilor măsurate la fiecare moment de timp.

☹ Pentru evaluarea cu precizie a mediei statistice, este necesară o colecţie infinită de realizări ale procesului.

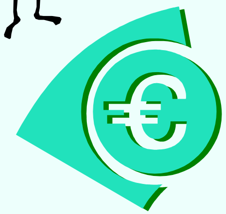
Un mare număr de experimente econometrice ar putea conduce la valori suficient de precise ale mediei statistice.





complicat

&



costisitor

❷ Modele de identificare

❷.❷ Noțiuni de Statistică şi Prelucrare de Semnal

Cum se poate utiliza operatorul de mediere statistică în cazul în care doar o singură realizare a procesului este disponibilă?

Folosind o ipoteză celebră din Fizică. → **Ipoteza Ergodică (IE) (de medie)**

Media statistică pe ansamblul realizărilor unui proces se poate evalua calculînd mediile temporale succesive ale oricărei realizări, în jurul fiecărui moment de timp.

Precizia evaluării creşte odată cu numărul datelor care reprezintă acea realizare.

$$E\{y[n]\} \cong \frac{1}{n}\sum_{i=1}^{n} y[i] \quad \forall n \in \mathbb{N}^*$$

← medie temporală

numărul de termeni ai sumei

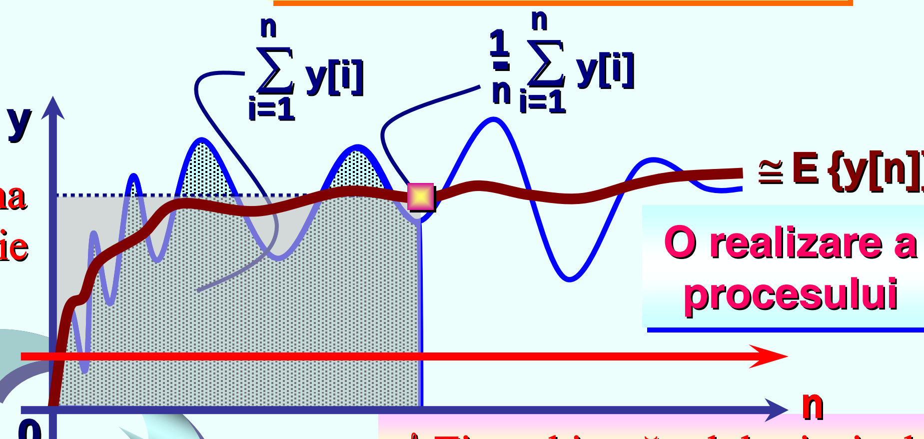
Direcţia de sumare/integrare în definiţia mediei temporale

Caz particular

Proces staționar (în medie)

$$E\{y[n]\} \overset{def}{=} \bar{y} = \text{constant} \quad \forall n \in \mathbb{N}^*$$

Teorema de medie



IE

$$E\{y[n]\} = \bar{y} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} y[i]$$

← cauzală

$$E\{y[n]\} = \bar{y} = \lim_{N\to\infty} \frac{1}{2N}\sum_{i=n-N+1}^{n+N} y[i]$$

← necauzală

☝ Timpul joacă rolul principal.

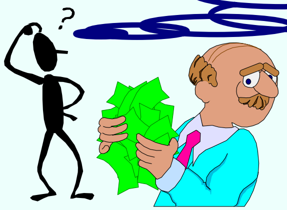
$$E\{y[n]\} \cong \frac{1}{n}\sum_{i=1}^{n} y[i] \quad \forall n \in \mathbb{N}^*$$

❷ Modele de identificare

❷.❷ Noţiuni de Statistică şi Prelucrare de Semnal

Aşadar

Folosirea **IE** conduce la posibilitatea de a aproxima operatorul de mediere statistică în aplicaţii practice **fără a cunoaşte densitatea de probabilitate** a datelor şi **fără a efectua mai mult de un experiment econometric**.



Cît de precisă este această aproximaţie?

Precizia aproximaţiei creşte odată cu numărul de date achiziţionate.

deterministă

$$E\{y\} = \boxed{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} y[i]} \quad \forall n \in \mathbb{N}^*$$

nedeterministă

$$E\{y[n]\} \cong \boxed{\frac{1}{n}\sum_{i=1}^{n} y[i]} \quad \forall n \in \mathbb{N}^*$$

Totuşi, pentru un număr constant de date achiziţionate, precizia aproximaţiei poate fi testată cu ajutorul **dispersiei datelor în jurul mediei**.

Dispersie

$$\sigma_y^2[n] \overset{def}{=} E\{y^2[n]\} = \int_{\mathcal{D}(y^2[n])} p\big(y^2[n]\big)\, y^2[n]\, dy^2[n] \quad \forall n \in \mathbb{N}^*$$

(media pătratelor valorilor)

Date staţionarizate

$$\tilde{y}[n] \overset{def}{=} y[n] - E\{y[n]\} \quad \forall n \in \mathbb{N}^*$$

(centrate pe medie)

$$E\{\tilde{y}[n]\} = 0 \quad \forall n \in \mathbb{N}^*$$

(vezi proprietăţile operatorului E)

Varianţă

$$\sigma_{\tilde{y}}^2[n] \overset{def}{=} E\{\tilde{y}^2[n]\} \quad \forall n \in \mathbb{N}^*$$

(dispersia datelor în jurul mediei)

IE

$$\sigma_y^2[n] \cong \frac{1}{n}\sum_{i=1}^{n} y^2[i] \quad \forall n \in \mathbb{N}^*$$

Deviaţie standard

$$\sigma_{\tilde{y}}[n] \overset{def}{=} \sqrt{E\{\tilde{y}^2[n]\}} \quad \forall n \in \mathbb{N}^*$$

- Altă variantă:

$$\sigma_y^2[n] \cong \frac{1}{n-1}\sum_{i=1}^{n} y^2[i] \quad \forall n \in \mathbb{N}^* \setminus \{1\}$$

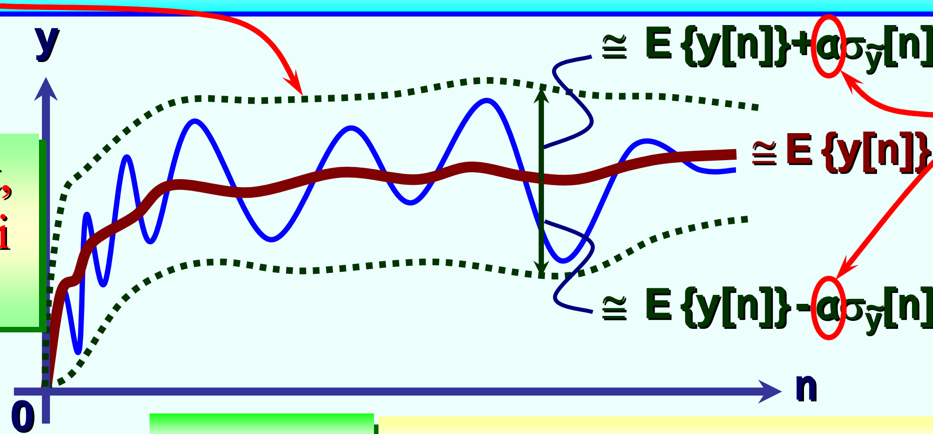
proprietăţi statistice superioare

➋ Modele de identificare

➋.➋ Noțiuni de Statistică și Prelucrare de Semnal

Gradul de grupare a datelor în jurul mediei este cuantificat de deviația standard, cu ajutorul tubului de deviație standard.

Cu cît tubul este mai îngust, cu atît precizia curbei medii este mai mare.



factor de proporționalitate care depinde de interpretarea pe care o are tubul

Exemplu Cazul procesului normal distribuit ⇒ $\alpha = 3$

Tubul de deviație standard este definit în orice moment de intervalul 3σ.

☝ Mai mult de 95% din realizările procesului trec prin acest tub.

Altă interpretare ⇒ **Precizia de predicție a procesului**

Valoarea măsurată la momentul următor celui curent (***n***) se va situa în intervalul de încredere:

$$\left[E\{y[n]\} - \alpha\sigma_{\tilde{y}}[n]\, ,\, E\{y[n]\} + \alpha\sigma_{\tilde{y}}[n] \right]$$

cu o anumită probabilitate, numită nivel de încredere.

☝ **De regulă, nivelul de încredere scade odată cu îngustarea intervalului de încredere.**

Exemplu Nivelul de încredere pentru predicția procesului normal distribuit este 0.95.