Identificarea sistemelor

Ingineria sistemelor, anul 3
Universitatea Tehnică din Cluj-Napoca

Lucian Buşoniu

Partea III

Baze matematice: Regresie liniară şi statistică

Motivare

Până acum, am discutat analiza în domeniul timp a răspunsurilor la treaptă, folosind concepte cunoscute din teoria sistemelor liniare.

Majoritatea metodelor de identificare care urmează necesită elemente noi: regresia liniară şi concepte de teoria probabilităţilor şi statistică. Le vom discuta aici.

În această parte anumite notaţii (de ex. x, A) au o semnificaţie diferită de cea din restul materialului de curs.

Conţinut

Conţinut

Problema de regresie

Elementele problemei:

- Un şir de eşantioane cunoscute $y(k) \in \mathbb{R}$, indexate de $k = 1, \dots, N$: y este *variabila dependentă* (măsurătoarea).
- Pentru fiecare k, un vector cunoscut $\varphi(k) \in \mathbb{R}^n$: conţine *regresorii* $\varphi_i(k)$, $i = 1, \dots, n$, $\varphi(k) = [\varphi_1(k), \varphi_2(k) \dots, \varphi_n(k)]^\top$.
- Un *vector de parametri* $\theta \in \mathbb{R}^n$ necunoscut.

Obiectiv: identificarea comportamentului variabilei dependente din date, folosind modelul liniar:

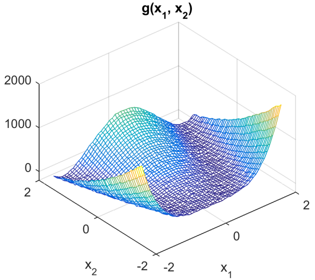
$$\boxed{y(k) = \varphi^\top(k)\theta}$$

Regresia liniară este o metodă clasică şi des folosită, de ex. Gauss a folosit-o pentru a calcula orbitele planetelor în 1809.

Problema de regresie: Două utilizări importante

1. k este o variabilă de timp, şi modelăm seria temporală $y(k)$.
2. k este doar un index de eşantion, şi $\varphi(k) = \phi(x(k))$ unde x este intrarea unei funcţii necunoscute g. În acest caz $y(k)$ este ieşirea corespunzătoare, posibil afectată de zgomot, iar obiectivul este identificarea unui model al funcţiei g din date. Această problemă se numeşte şi *aproximarea unei funcţii*, sau *învăţarea supervizată*.

$g(x)$ necunoscut



$\hat{g}(x)$

?

Aproximare: Funcţii de bază

Pentru aproximarea unei funcţii, regresorii $\phi_i(k)$ din:

$$\phi(x(k)) = [\phi_1(x(k)), \phi_2(x(k)), \ldots, \phi_n(x(k))]^\top$$

se numesc *funcţii de bază*.

Aproximare: Exemplul 1

Studiem venitul anual y (în EUR) al unei persoane bazat pe nivelul de educaţie x_1 şi experienţa profesională x_2 (ambele măsurate în ani).

Se dă un set de date $(x_1(k), x_2(k), y(k))$ de la un grup reprezentativ de persoane. Obiectivul este predicţia venitului oricărei alte persoane din nivelul său de educaţie (x_1) şi experienţă (x_2).

- Alegem funcţiile de bază $\phi(x) = [x_1, x_2, 1]^\top$. Ne aşteptăm ca venitul să evolueze conform cu $\theta_1 x_1 + \theta_2 x_2 + \theta_3 = \phi^\top(x)\theta$, crescând liniar cu educaţia and experienţa (de la un nivel minimal). Regresia implică găsirea parametrilor θ pentru care expresia este cel mai aproape de valorile reale ale venitului.
- În realitate, lucrurile sunt mai complicate... vom avea nevoie de variabile de intrare suplimentare, funcţii de bază mai bune, etc.

Aproximare: Exemplul 2

Studiem timpul de reacţie y (în ms) al unui şofer în funcţie de vârsta sa x_1 (în ani) şi oboseala x_2 (de ex. pe o scară de la 0 la 1).

Se dă un set de date $(x_1(k), x_2(k), y(k))$ de la un grup reprezentativ de şoferi de diferite vârste şi nivele de oboseală. Obiectivul este predicţia timpului de reacţie al oricărui alt şofer folosind vârsta (x_1) şi oboseala (x_2) sa.

Exemplu regresori 1: Polinom în k

Util pentru modelarea seriilor temporale.

$$
\begin{aligned}
y(k) &= \theta_1 + \theta_2 k + \theta_3 k^2 + \ldots + \theta_n k^{n-1} \\
&= \begin{bmatrix} 1 & k & k^2 & \ldots & k^{n-1} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \ldots \\ \theta_n \end{bmatrix} \\
&= \varphi^\top(k)\theta
\end{aligned}
$$

Exemplu regresori 2: Polinom în x

Util pentru aproximarea funcţiilor. De exemplu, polinomul de gradul 2 cu două variabile de intrare $x = [x_1, x_2]^\top$ este:

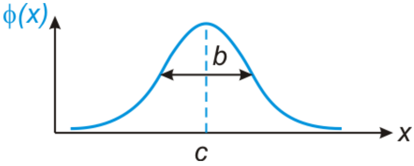
$$
\begin{aligned}
y(k) &= \theta_1 + \theta_2 x_1(k) + \theta_3 x_2(k) + \theta_4 x_1^2(k) + \theta_5 x_2^2(k) + \theta_6 x_1(k) x_2(k) \\
&= \begin{bmatrix} 1 & x_1(k) & x_2(k) & x_1^2(k) & x_2^2(k) & x_1(k)x_2(k) \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \end{bmatrix} \\
&= \phi^\top(x(k))\theta = \varphi^\top(k)\theta
\end{aligned}
$$

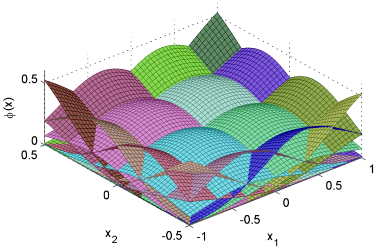
Conexiune: Proiect partea 1

Exemplu regresori 3: Funcţii de bază Gaussiene

Utile pentru aproximarea funcţiilor:

$$\phi_i(x) = \exp\left[-\frac{(x-c_i)^2}{b_i^2}\right] \qquad \text{(1-dim)};$$
$$= \exp\left[-\sum_{j=1}^{d}\frac{(x_j-c_{ij})^2}{b_{ij}^2}\right] \qquad \text{(d-dim)}$$

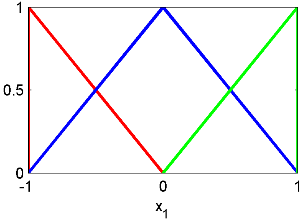


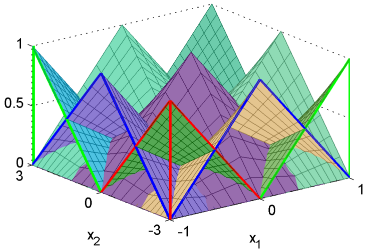


Exemplu regresori 4: Interpolare

Utilă pentru aproximarea funcţiilor.

- Grilă *d*-dimensională de puncte în spaţiul intrărilor.
- Interpolare (multi)-liniară între aceste puncte.
- Echivalent cu funcţii de bază *piramidale* (triunghiulare într-o singură dimensiune)





Sistem liniar

Scriind modelul pentru fiecare din cele N date, obţinem un sistem de ecuaţii liniare:

$$\begin{aligned} y(1) &= \varphi_1(1)\theta_1 + \varphi_2(1)\theta_2 + \ldots \varphi_n(1)\theta_n \\ y(2) &= \varphi_1(2)\theta_1 + \varphi_2(2)\theta_2 + \ldots \varphi_n(2)\theta_n \\ &\ldots \\ y(N) &= \varphi_1(N)\theta_1 + \varphi_2(N)\theta_2 + \ldots \varphi_n(N)\theta_n \end{aligned}$$

Reamintim că în aproximarea funcţiilor, $\varphi_i(k) = \phi_i(x(k))$

Sistemul se poate scrie în *formă matriceală*:

$$\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix} = \begin{bmatrix} \varphi_1(1) & \varphi_2(1) & \ldots & \varphi_n(1) \\ \varphi_1(2) & \varphi_2(2) & \ldots & \varphi_n(2) \\ \ldots & \ldots & \ldots & \ldots \\ \varphi_1(N) & \varphi_2(N) & \ldots & \varphi_n(N) \end{bmatrix} \cdot \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \ldots \\ \theta_n \end{bmatrix}$$

$$\boxed{Y = \Phi\theta}$$

cu noile variabile $Y \in \mathbb{R}^N$ şi $\Phi \in \mathbb{R}^{N\times n}$.

Problema celor mai mici pătrate (CMMP)

Dacă $N = n$, sistemul se poate rezolva cu egalitate.

În practică, este mai bine să folosim $N > n$, de ex. datorită zgomotului. În acest caz, sistemul nu mai poate fi rezolvat cu egalitate, ci doar cu aproximare.

- *Eroarea* la k: $\varepsilon(k) = y(k) - \varphi^\top(k)\theta$, vectorul de eroare $\varepsilon = [\varepsilon(1), \varepsilon(2), \dots, \varepsilon(N)]^\top$.
- Funcţia obiectiv ce trebuie minimizată:

$$V(\theta) = \boxed{\frac{1}{2}\sum_{k=1}^{N} \varepsilon(k)^2} = \frac{1}{2}\varepsilon^\top \varepsilon$$

Problema CMMP

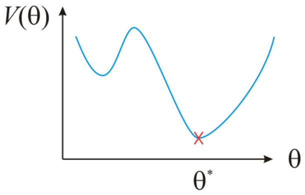
Găseşte vectorul de parametri $\widehat{\theta}$ care minimizează funcţia obiectiv:

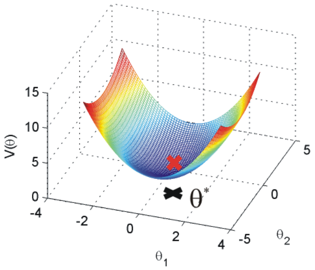
$$\widehat{\theta} = \arg\min_{\theta} V(\theta)$$

Paranteză: Problema de optimizare

Dată fiind o funcţie V de variabilele θ, care poate fi de ex. obiectivul nostru CMMP, sau în general oricare altă funcţie:

găseşte *valoarea optimă a funcţiei* $\min_\theta V(\theta)$ şi valorile $\theta^* = \arg\min_\theta V(\theta)$ ale variabilelor pentru care minimul este atins





De notat că în cazul regresiei liniare, folosim notaţia $\widehat{\theta}$; vectorul $\widehat{\theta}$ este soluţia reală a problemei de optimizare dat fiind setul de date, dar rămâne totuşi o estimare datorită zgomotului din date

Soluţia formală a problemei de regresie

După câţiva paşi de algebră liniară:

$$\boxed{\widehat{\theta} = (\Phi^\top \Phi)^{-1} \Phi^\top Y}$$

Observaţii:

- Valoarea optimă a funcţiei obiectiv este $V(\widehat{\theta}) = \frac{1}{2}[Y^\top Y - Y^\top \Phi(\Phi^\top \Phi)^{-1} \Phi^\top Y]$.
- Matricea $\Phi^\top \Phi$ trebuie să fie inversabilă, ceea ce necesită o alegere bună a modelului (ordin n, regresori φ), şi folosirea unui set informativ de date.

Expresie alternativă

$$\Phi^\top \Phi = \sum_{k=1}^{N} \varphi(k)\varphi^\top(k), \Phi^\top Y = \sum_{k=1}^{N} \varphi(k)y(k)$$

Soluţia poate fi scrisă:

$$\widehat{\theta} = \left[\sum_{k=1}^{N} \varphi(k)\varphi^\top(k)\right]^{-1} \left[\sum_{k=1}^{N} \varphi(k)y(k)\right]$$

Avantaj: matricea Φ cu dimensiunile $N \times n$ nu mai trebuie calculată; este nevoie doar de matrici şi vectori mai mici, de dimensiuni $n \times n$ respectiv n.

Rezvolarea sistemului liniar

În practică, ambele metode bazate pe inversarea de matrici se comportă prost din punct de vedere numeric. Există algoritmi mai buni, cum ar fi triangularizarea ortogonală.

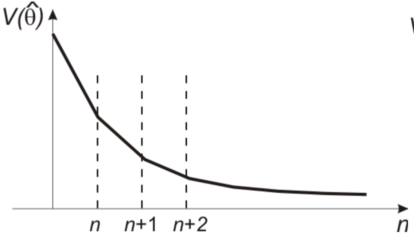
În majoritatea cazurilor, MATLAB alege automat un algoritm potrivit. Dacă Φ este stocată în variabila `PHI` şi Y în `Y`, comanda care rezolvă sistemul de ecuaţii în sensul CMMP este împărţirea matriceală la stânga (backslash):

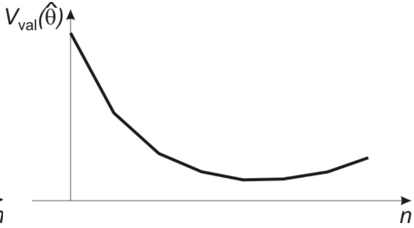
```
theta = PHI \ Y;
```

Dacă se doreşte un control mai detaliat al algoritmului, se poate folosi funcţia `linsolve` în loc de \.

Alegerea modelului

Considerăm că dată fiind o complexitate a modelului (număr de parametri) n, putem genera regresori $\varphi(k)$ care fac modelul mai expresiv (de ex., funcţii de bază pe o grilă mai fină). Ne aşteptăm ca funcţia obiectiv (CMMP) să se comporte în următorul fel:





Putem aşadar creşte treptat valoarea lui n până când eroarea V nu mai scade, sau eroarea V_{val} pe datele de validare începe să crească.

Observaţie: Dacă datele sunt afectate de zgomot, creşterea exagerată a lui n va duce la supraantrenare: performanţe bune pe datele de identificare, dar proaste pe date diferite. Validarea pe un set separat de date este esenţială în practică!

Conţinut

Exemplu analitic: Estimarea unui scalar

Model:

$$y(k) = b = 1 \cdot b = \varphi(k)\theta$$

unde $\varphi(k) = 1 \forall k$, $\theta = b$.

Pentru N date:

$$y(1) = \varphi(1)\theta = 1 \cdot b$$
$$\cdots$$
$$y(N) = \varphi(N)\theta = 1 \cdot b$$

În formă matriceală:

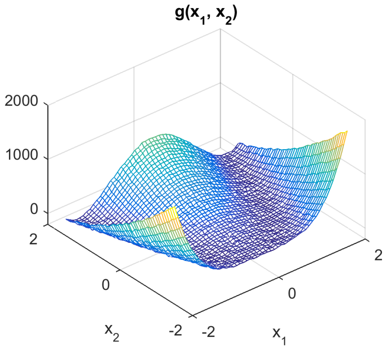
$$\begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \theta$$
$$Y = \Phi\theta$$

Exemplu analitic: Estimarea unui scalar (continuare)

$$
\begin{aligned}
\widehat{\theta} &= (\Phi^\top \Phi)^{-1} \Phi^\top Y \\
&= \left(\begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} \\
&= N^{-1} \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} \\
&= \frac{1}{N} (y(1) + \ldots + y(N))
\end{aligned}
$$

Intuiţie: Estimarea este media tuturor măsurătorilor, filtrând zgomotul.

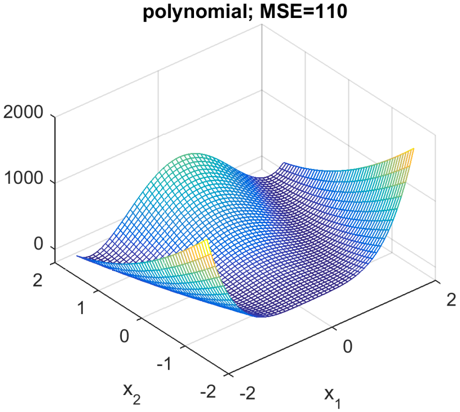
Exemplu: Aproximarea funcţiei lui Rosenbrock



- Funcţia Rosenbrock: $g(x_1, x_2) = (1 - x_1)^2 + 100[(x_2 + 1.5) - x_1^2]^2$ (necunoscută de algoritm).
- Date de identificare: 200 puncte de intrare (x_1, x_2), distribuite aleator în spaţiul $[-2, 2] \times [-2, 2]$; şi ieşirile corespunzătoare $y = g(x_1, x_2)$, afectate de zgomot.
- Date de validare: grilă uniformă cu 31×31 puncte în $[-2, 2] \times [-2, 2]$ cu ieşirile corespunzătoare (afectate de zgomot).

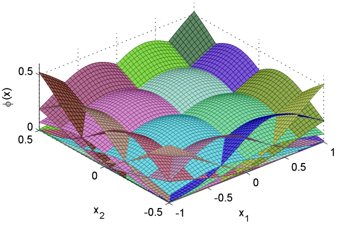
Funcţia Rosenbrock: Aproximare polinomială

Polinom de gradul 4 în cele două intrări (15 parametri):

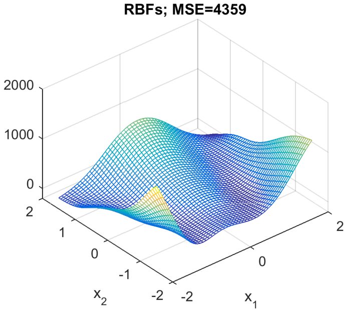


Funcţia Rosenbrock: Funcţii de bază radiale

Reamintim funcţiile de bază radiale:

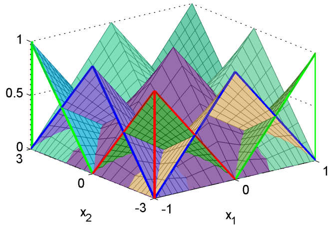


Rezultate cu 6×6 RBF-uri, cu centrele pe o grilă echidistantă şi lăţimea egală cu distanţa între centre:

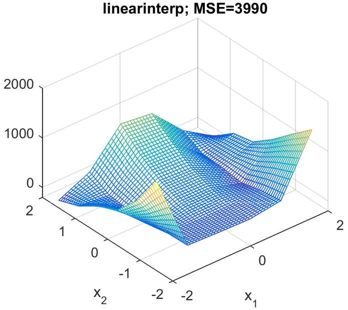


Funcţia Rosenbrock: Interpolare

Reamintim funcţiile de bază piramidale, pentru interpolare:



Rezultate cu grilă de interpolare 6×6 (corespunzând la 6×6 funcţii de bază):



Conţinut

Variabilă aleatoare discretă

Fie un set $\mathcal{X}$ conţinând valori posibile x. O variabilă aleatoare X corespunzând acestui set este descrisă de funcţia de frecvenţă:

Definiţie

Funcţia de frecvenţă a variabilei X este lista probabilităţilor tuturor valorilor individuale $p(x_0), p(x_1), \ldots$. Suma probabilităţilor trebuie să fie 1: $P(X \in \mathcal{X}) = \sum_{x \in \mathcal{X}} p(x) = 1$.

Exemplu: Numărul obţinut ca urmare a aruncării unui zar este o variabilă aleatoare discretă, cu şase valori posibile: $\mathcal{X} = \{1, 2, \ldots, 6\}$. Pentru un zar corect, funcţia de frecvenţă este $p(x) = 1/6$ pentru toate valorile x, de la 1 la 6.

Observaţie: Setul $\mathcal{X}$ poate conţine un număr finit sau infinit de elemente, dar în al doilea caz setul trebuie să fie numărabil (Intuitiv: valorile se pot enumera, sau pot fi asociate numerelor naturale $0, 1, 2, \ldots$).

Variabilă aleatoare continuă

Fie acum un interval $\mathcal{X} \subseteq \mathbb{R}$ de numere reale, cu x o valoare individuală. O variabilă aleatoare X corespunzând acestui set este descrisă de densitatea de probabilitate:

Definiţie

Densitatea de probabilitate a unei variabile aleatoare continue X este o funcţie $f : \mathcal{X} \to [0, \infty)$, astfel încât probabilitatea de a obţine o valoare într-un interval $[a, b] \subseteq \mathcal{X}$ este:

$$P(X \in [a, b]) = \int_a^b f(x)dx$$

Densitatea trebuie să satisfacă $P(X \in \mathcal{X}) = \int_{X \in \mathcal{X}} f(x)dx = 1$.

Observaţie: Intervalul $\mathcal{X}$ poate fi şi setul complet al numerelor reale.

Exemplu 1: Densitate uniformă

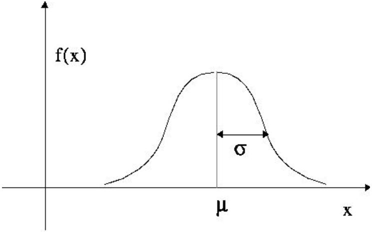
Dorim să caracterizăm unghiul la care se va opri roata unei rulete, ca o fracţiune dintr-o rotaţie completă: $x \in [0, 1]$. În acest caz, $\mathcal{X} = [0, 1]$. Pentru o ruletă corectă, fiecare valoare x are aceeaşi probabilitate.

Vom încerca întâi să definim o funcţie de frecvenţă; aceasta trebuie să aibă aceeaşi valoare peste tot. Luăm orice valoare nenulă c; cum există o infinitate nenumărabilă de astfel de valori în intervalul $[0, 1]$, probabilitatea întregului set $P(X \in [0, 1])$, care însumează constanta c de o infinitate de ori, este infinită. Deci $p(x)$ trebuie să fie 0 pentru toate valorile x, şi nu are nici o utilitate practică; nu putem aşadar folosi funcţia de frecvenţă pentru a descrie acest caz.

Putem defini probabilităţi doar pentru intervale, şi de aceea avem nevoie de densitatea f. Mai exact, pentru a obţine probabilităţi uniforme, dorim ca $P(X \in [a, b]) = b - a$, ceea ce înseamnă că $f(x) = 1$.

Exemplu 2: Densitate Gaussiană

Are formă similară cu funcţiile de bază Gaussiene, dar semnificaţie diferită.



$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Parametri: media μ şi varianţa σ^2 (vor fi explicaţi mai târziu)

Distribuţia Gaussiană intervine adeseori în natură: de ex., distribuţia IQ-urilor într-o populaţie umană. Este numită de aceea şi distribuţia normală, şi se notează $\mathcal{N}(\mu, \sigma^2)$.

Valoarea medie

Definiţie

$$\boxed{\mathrm{E}\left\{X\right\} = \begin{cases} \sum_{x \in \mathcal{X}} p(x)x & \text{pentru variabile aleatoare discrete} \\ \int_{x \in \mathcal{X}} f(x)x & \text{pentru variabile aleatoare continue} \end{cases}}$$

Intuiţie: media tuturor valorilor, ponderate de probabilitatea lor; valoarea "aşteptată" în avans, dată fiind distribuţia de probabilitate.

Valoarea medie se mai numeşte şi *valoare aşteptată* sau *speranţă*.

Exemple:

- Pentru un zar unde X este numărul feţei obţinute, $\mathrm{E}\left\{X\right\} = \frac{1}{6}1 + \frac{1}{6}2 + \ldots + \frac{1}{6}6 = 7/2$.
- Dacă X are densitate Gaussiană $\mathcal{N}(\mu, \sigma^2)$, atunci $\mathrm{E}\left\{X\right\} = \mu$.

Valoarea medie a unei funcţii

Considerăm o funcţie $g : \mathcal{X} \rightarrow \mathbb{R}$ care depinde de variabila aleatoare X. Atunci, $g(X)$ este o şi ea o variabilă aleatoare, cu valoarea medie:

$$\mathrm{E}\left\{g(X)\right\} = \begin{cases} \sum_{x \in \mathcal{X}} p(x)g(x) & \text{discret} \\ \int_{x \in \mathcal{X}} f(x)g(x) & \text{continuu} \end{cases}$$

Varianţa

Definiţie

$$\mathrm{Var}\left\{X\right\} = \boxed{\mathrm{E}\left\{(X - \mathrm{E}\left\{X\right\})^2\right\}} = \mathrm{E}\left\{X^2\right\} - (\mathrm{E}\left\{X\right\})^2$$

Intuiţie: cât de “răspândite” sunt valorile aleatoare în jurul valorii medii.

$$\begin{aligned}\mathrm{Var}\left\{X\right\} &= \begin{cases}\sum_{x\in\mathcal{X}} p(x)(x - \mathrm{E}\left\{X\right\})^2 & \text{discret}\\ \int_{x\in\mathcal{X}} f(x)(x - \mathrm{E}\left\{X\right\})^2 & \text{continuu}\end{cases}\\ &= \begin{cases}\sum_{x\in\mathcal{X}} p(x)x^2 - (\mathrm{E}\left\{X\right\})^2 & \text{discret}\\ \int_{x\in\mathcal{X}} f(x)x^2 - (\mathrm{E}\left\{X\right\})^2 & \text{continuu}\end{cases}\end{aligned}$$

Exemple:

- Pentru un zar, $\mathrm{Var}\left\{X\right\} = \frac{1}{6}1^2 + \frac{1}{6}2^2 + \ldots + \frac{1}{6}6^2 - (7/2)^2 = 35/12$.
- Dacă X are densitate Gaussiană $\mathcal{N}(\mu, \sigma^2)$, atunci $\mathrm{Var}\left\{X\right\} = \sigma^2$.

Notaţie

Vom nota generic $\mathrm{E}\left\{X\right\} = \mu$ şi $\mathrm{Var}\left\{X\right\} = \sigma^2$.

Cantitatea $\sigma = \sqrt{\mathrm{Var}\left\{X\right\}}$ se numeşte *abaterea standard*.

Probabilitate: Independenţă

Definiţie

Două variabile aleatoare X şi Y sunt independente dacă:

- în cazul continuu, $f_{X,Y}(x,y) = f_Y(x) \cdot f_Y(y)$.
- în cazul discret, $p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$.

unde $f_{X,Y}$ este densitatea comună a vectorului (X, Y), f_X şi f_Y sunt densităţile pentru X şi Y, şi la fel pentru funcţiile de frecvenţă p.

Exemple:

- Evenimentul de a arunca 6 cu un zar este independent de evenimentul 6 la aruncarea anterioară (de fapt, de oricare altă valoare la orice aruncare anterioară).
- Evenimentul de a arunca două valori 6 *consecutive* nu este independent de aruncarea anterioară!

(Primul fapt este contra-intuitiv şi multă lume nu îl înţelege, ducând la aşa-numita *gambler's fallacy*. O secvenţă mai lungă de jocuri norocoase sau proaste nu are nici absolut nici o influenţă asupra jocului următor!)

Covarianţă

Definiţie

$$\mathrm{Cov}\{X, Y\} = \boxed{\mathrm{E}\{(X - \mathrm{E}\{X\})(Y - \mathrm{E}\{Y\})\}} = \mathrm{E}\{(X - \mu_X)(Y - \mu_Y)\}$$

unde μ_X, μ_Y sunt valorile medii ale celor două variabile.

Intuiţie: cât de “aliniate” sunt schimbările celor două variabile (covarianţă pozitivă dacă variabilele se schimbă în direcţii similare, negativă dacă se schimbă în direcţii opuse).

Observaţie: $\mathrm{Var}\{X\} = \mathrm{Cov}\{X, X\}$.

Variabile necorelate

Definiţie

Variabilele aleatoare X şi Y sunt necorelate dacă $\mathrm{Cov}\{X, Y\} = 0$.
Altfel, ele se numesc corelate.

Exemple:

- Nivelul de educaţie al unei persoane este corelat cu venitul său.
- Culoarea părului este necorelată cu venitul (sau ar trebui să fie, în cazul ideal).

Observaţii:

- Dacă X şi Y sunt independente, atunci sunt şi necorelate.
- Dar nu şi invers! Putem avea variabile necorelate care sunt totuşi dependente.

Conţinut

Vectori de variabile aleatoare

Considerăm vectorul $\boldsymbol{X} = [X_1, \ldots, X_N]^\top$ unde fiecare X_i este o variabilă aleatoare cu valori reale continue. Acest vector are o *funcţie de densitate comună* $f(\boldsymbol{x})$, cu $\boldsymbol{x} \in \mathbb{R}^N$.

Definiţii

Valoarea medie şi *matricea de covarianţă* a lui $\boldsymbol{X}$:

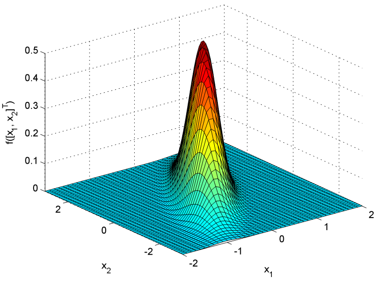
$$\mathrm{E}\left\{\boldsymbol{X}\right\} := \left[\mathrm{E}\left\{X_1\right\}, \ldots, \mathrm{E}\left\{X_N\right\}\right]^\top = [\mu_1, \ldots, \mu_N]^\top, \text{ notată } \boldsymbol{\mu} \in \mathbb{R}^N$$

$$\mathrm{Cov}\left\{\boldsymbol{X}\right\} := \begin{bmatrix} \mathrm{Cov}\left\{X_1, X_1\right\} & \mathrm{Cov}\left\{X_1, X_2\right\} & \cdots & \mathrm{Cov}\left\{X_1, X_N\right\} \\ \mathrm{Cov}\left\{X_2, X_1\right\} & \mathrm{Cov}\left\{X_2, X_2\right\} & \cdots & \mathrm{Cov}\left\{X_2, X_N\right\} \\ \cdots & \cdots & \cdots & \cdots \\ \mathrm{Cov}\left\{X_N, X_1\right\} & \mathrm{Cov}\left\{X_N, X_2\right\} & \cdots & \mathrm{Cov}\left\{X_N, X_N\right\} \end{bmatrix}$$

$$= \mathrm{E}\left\{(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^\top\right\}, \text{ notată } \Sigma \in \mathbb{R}^{N,N}$$

Observaţii: $\mathrm{Cov}\left\{X_i, X_i\right\} = \mathrm{Var}\left\{X_i\right\}$. De asemenea, $\mathrm{Cov}\left\{X_i, X_j\right\} = \mathrm{Cov}\left\{X_j, X_i\right\}$, deci matricea Σ este simetrică.

Exemplu: vector Gaussian



Densitatea comună Gaussiană a unui vector $\boldsymbol{X}$ se poate scrie:

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^N \sqrt{\det(\Sigma)}} \exp\left(-(\boldsymbol{x} - \boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})^\top\right)$$

unde $\boldsymbol{\mu}$ este vectorul de valori medii şi Σ matricea de covarianţă (care trebuie să fie pozitiv definită, pentru ca $\det(\Sigma) > 0$ şi Σ^{-1} să existe).

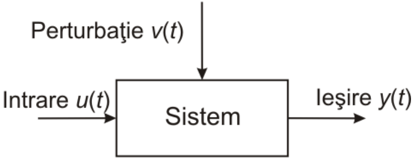
Proces stohastic

Definiţie

Un proces stohastic $\boldsymbol{X}$ este o secvenţă de variabile aleatoare $\boldsymbol{X} = (X_1, \ldots, X_k, \ldots, X_N)$.

Avem aşadar de-a face tot cu un vector de variabile aleatoare, cu o structură specifică: fiecare index din vector este asociat unui pas discret de timp k.

În identificarea sistemelor, semnalele (intrări, ieşiri, perturbaţii etc.) vor fi adesea procese stohastice.



Zgomot alb de medie zero

Definiţie

Un proces stohastic $\boldsymbol{X}$ este zgomot alb de medie zero dacă: $\forall k$, $\mathrm{E}\left\{X_k\right\} = 0$ (medie zero), şi $\forall k, k' \neq k$, $\mathrm{Cov}\left\{X_k, X_{k'}\right\} = 0$ (valorile la paşi diferiţi de timp sunt necorelate). În plus, varianţa $\mathrm{Var}\left\{X_k\right\}$ trebuie să fie finită $\forall k$.

Cu notaţie vectorială, aceste proprietăţi se pot scrie compact: media $\boldsymbol{\mu} = \mathrm{E}\left\{\boldsymbol{X}\right\} = \mathbf{0} \in \mathbb{R}^N$ şi matricea de covarianţă $\Sigma = \mathrm{Cov}\left\{\boldsymbol{X}\right\}$ este diagonală (cu diagonala formată din numere finite şi pozitive).

În identificarea sistemelor, măsurătorile sunt adesea afectate de zgomote, şi vom presupune câteodată că aceste zgomote sunt albe şi de medie zero.

Proces staţionar

Valorile unui semnal la diferite moment de timp pot fi corelate (de ex. când semnalul depinde de ieşirea unui sistem dinamic). Vom presupune însă câteodată că semnalele sunt staţionare:

Definiţie

Procesul stohastic $\boldsymbol{X}$ este staţionar dacă $\forall k$, $\mathrm{E}\left\{X_k\right\} = \mu$, şi $\forall k, k', \tau$, $\mathrm{Cov}\left\{X_k, X_{k+\tau}\right\} = \mathrm{Cov}\left\{X_{k'}, X_{k'+\tau}\right\}$.

Media este aceeaşi la fiecare pas, iar covarianţa depinde doar de poziţiile relative ale paşilor de timp (şi nu de poziţiile lor absolute).