

Decizie și Estimare în Prelucrarea Informației

Capitolul III. Elemente de Teoria Estimării

III.1 Introducere

Ce înseamnă "estimare"?

- ▶ Un emițător transmite un semnal $s_\Theta(t)$ care depinde de parametru necunoscut Θ
- ▶ Semnalul este afectat de zgomot, se recepționează

$$\boxed{r(t)} = s_\Theta(t) + \underline{\text{zgomot}}$$

- ▶ Vrem să găsim valoarea parametrului Θ
 - ▶ pe baza eșantioanelor din semnalul recepționat, sau a întregului semnal
 - ▶ datele recepționate au zgomot \Rightarrow parametrul este "estimat"
- ▶ Valoarea găsită este $\hat{\Theta}$, estimatul lui Θ
 - ▶ există întotdeauna eroare de estimare $\epsilon = \hat{\Theta} - \Theta$

$$A = \begin{cases} \infty \\ 5 \end{cases}$$

$$\begin{aligned}s_\Theta(t) &= \textcircled{A} \\ &= \cos(\underline{\omega t}) \\ &= \textcircled{A} \cdot \cos(2\pi f t)\end{aligned}$$

$$\begin{cases} s_0(t) = \cos(2\pi f t) \\ s_1(t) = 0 \end{cases}$$

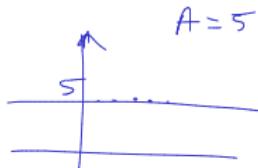
$$s_\Theta(t) = A \cdot \cos(2\pi f t), A = \begin{cases} \infty \\ 1 \end{cases}$$

Ce înseamnă "estimare"?

7.1 7.15 6.93 6.74 7.02

$$\hat{A} =$$

$$D(t) = A$$



► Exemple:

► Amplitudinea unui semnal constant: $r(t) = A + \text{zgomot}$, trebuie estimat A

$$\varepsilon = \hat{A} - A = -0.1$$

► Faza unui semnal sinusoidal: $r(t) = \cos(2\pi ft + \phi) + \text{zgomot}$, de estimat ϕ

► Exemple mai complicate:

► De estimat/decis ce cuvânt este pronunțat într-un semnal vocal

Estimare și Detectie/Decizie

$$\delta(A) = A \quad , \quad \hat{A} = ?$$

$$\delta(A) = A, \quad A = \begin{cases} 0 \\ 5 \end{cases} \quad , \quad \hat{A} = \begin{cases} 0 \\ 5 \end{cases} ?$$

- ▶ Fie următoarea problemă de estimare:

Se recepționează un semnal $r(t) = A + \text{zgomot}$, estimați-l pe A

- ▶ La detectie: se alege între două valori cunoscute ale A :

- ▶ de ex. A poate fi 0 sau 5 (ipotezele H_0 și H_1)
- ▶ La estimare: A poate fi oricât \Rightarrow se alege între o infinitate de opțiuni ale A
 - ▶ A poate fi orice valoare din \mathbb{R} , în general

Decizie

- ▶ Detectie = Estimare **restrânsă** doar la un set discret de opțiuni
- ▶ Estimare = Detectie cu un număr infinitt de opțiuni posibile
- ▶ Metodele statistice sunt similare
 - ▶ În practică, distincția între estimare și detectie nu este strictă
 - ▶ (de ex. când trebuie să alegem între 1000 de ipoteze, este “detectie” sau “estimare”?)

$$\Delta_0(t) \subset -1000000$$

$$\Delta_1(t) = -599\dots59$$

$$\Delta_2(t) = -999\dots58$$

$w(x)$

- ▶ Semnalul recepționat este $r(t) = s_{\Theta}(t) + zgomot$
 - ▶ este afectat de zgomot
 - ▶ depinde de parametrul necunoscut Θ
- ▶ Considerăm N eșantioane din $r(t)$, luate la momentele de timp t_i

$$\underline{r} = [r_1, r_2, \dots, r_N]$$

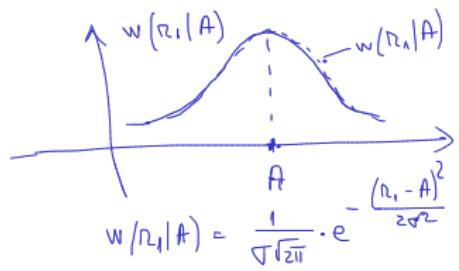
- ▶ Eșantioanele depind de valoarea lui Θ \Rightarrow $\hat{\Theta}$

Semnalul recepționat

Exemplu:

$$\Delta_{\Theta}(t) = A$$

$$r_i(t) = A + z_{\text{gomer}} \sim N(\mu=0, \sigma^2)$$



- ▶ Fiecare eşantion r_i este o variabilă aleatoare ce depinde de Θ (și de zgomer)

- ▶ Fiecare esantion are o distribuție care depinde de Θ

$$w_i(r_i | \Theta)$$

$$w_i(r_i | \Theta)$$

~

$$w(r_i | \frac{\Theta_0}{\Theta_1})$$

- ▶ Întregul vector de esantioane r este o variabilă aleatoare N-dimensională ce depinde de Θ (și de zgomer)

- ▶ Are o distribuție N-dimensională ce depinde de Θ
 - ▶ Egală cu produsul tuturor $w_i(r_i | \Theta)$

$$w(r | \Theta) = \underbrace{w_1(r_1 | \Theta)} \cdot \underbrace{w_2(r_2 | \Theta)} \cdot \dots \cdot \underbrace{w_N(r_N | \Theta)}$$

Tipuri de estimare

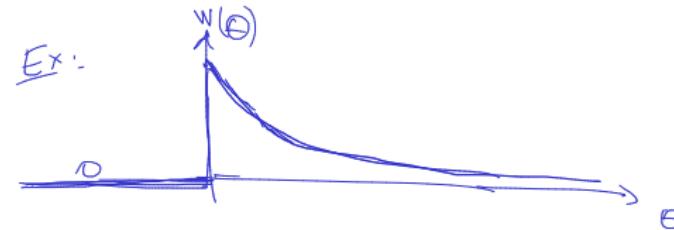
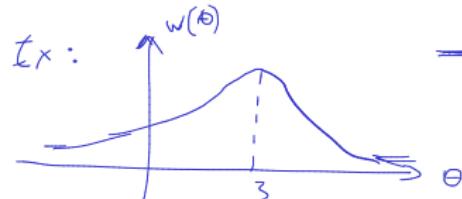
- Considerăm două tipuri de estimare:

1. Estimare de plauzibilitate maximă (Maximum Likelihood)

Estimation, MLE): În afară de r nu se cunoaște nimic despre Θ , decât cel mult vreun domeniu de existență (de ex. $\Theta > 0$)

2. Estimare Bayesiană: În afară de r se mai cunoaște o distribuție a priori $w(\Theta)$ a lui Θ , care indică ce valori ale lui Θ sunt mai probabile / mai puțin probabile

- caz mai general decât primul



Estimare ML (Maximum Likelihood)

~ Adjectiv M.L.

~ Detection M.P.E.
M.R.

$P(H_0)$
 $P(H_1)$

C₀₀

C₁₀

C₀₁

C₁₁

II.2 Estimarea de plauzibilitate maximă (Maximum Likelihood)

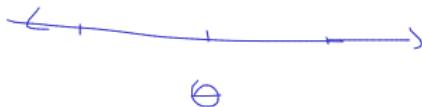
Estimarea tip Maximum Likelihood

- Dacă nu se cunoaște vreo distribuție *a priori* se folosește metoda estimării de plauzibilitate maximă ("Maximum Likelihood", ML)
- Se definește **plauzibilitatea** unei valori Θ , dat fiind vectorul de observații r :

$$L(\Theta|r) = \cancel{w(\Theta|r)} \cdot w(r|\Theta) = \text{probabilitatea de a se fi generat } r \text{ dacă param. erau fi avut valoarea } \Theta$$

- $L(\Theta|r)$ reprezintă funcția de plauzibilitate
- A se compara cu formula din Cap. 2, slide 20

- e aceeași
- aici "ghicim" pe Θ , acolo "ghiceam" pe H_i



Estimarea tip Maximum Likelihood

$$\hat{\theta}_{ML}$$

$L(\theta|r)$ = o funcție de θ

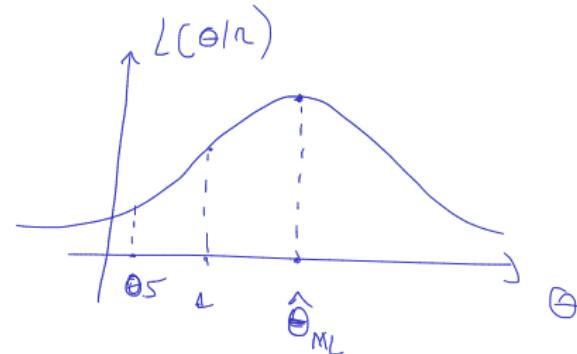
Estimarea de plauzibilitate maximă (Maximum Likelihood, ML):

- ▶ Estimatul $\hat{\theta}_{ML}$ este valoarea care maximizează plauzibilitatea, dat fiind valorile observate r

- ▶ i.e. valoarea care maximizează $L(\theta|r)$, adică maximizează $w(r|\theta)$

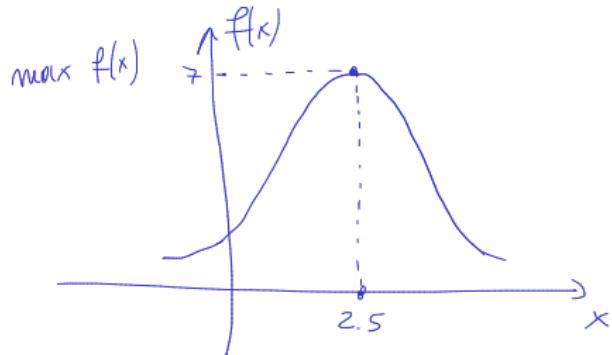
$$\hat{\theta}_{ML} = \arg \max_{\theta} L(\theta|r) = \arg \max_{\theta} w(r|\theta)$$

- ▶ Dacă θ aparține doar unui anumit interval, se face maximizarea doar pe acel interval

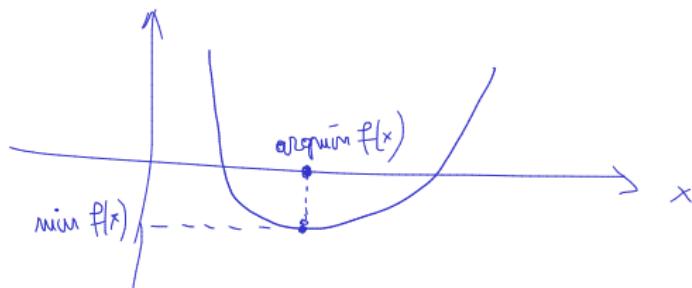


▶ Notări matematice generale

- ▶ $\arg \max_x f(x) =$ "valoarea x care maximizează funcția $f(x)$ "
- ▶ $\max_x f(x) =$ "valoarea maximă a funcției $f(x)$ "



$$2.5 = \arg\max f(x)$$
$$7 = \max f(x)$$



Estimare vs decizie Maximum Likelihood

- ▶ Estimarea ML este foarte similară cu decizia ML!
- ▶ Criteriul de decizie ML:
 - ▶ “se alege ipoteza cu plauzibilitate mai mare”:

$$\frac{L(H_1|r)}{L(H_0|r)} = \frac{w(r|H_1)}{w(r|H_0)} \stackrel{H_1}{\gtrless} 1$$

- ▶ Estimare ML:
 - ▶ “se alege valoarea care maximizează plauzibilitatea”

$$\hat{\Theta}_{ML} = \arg \max_{\Theta} L(\Theta|r) = \arg \max_{\Theta} w(r|\Theta)$$

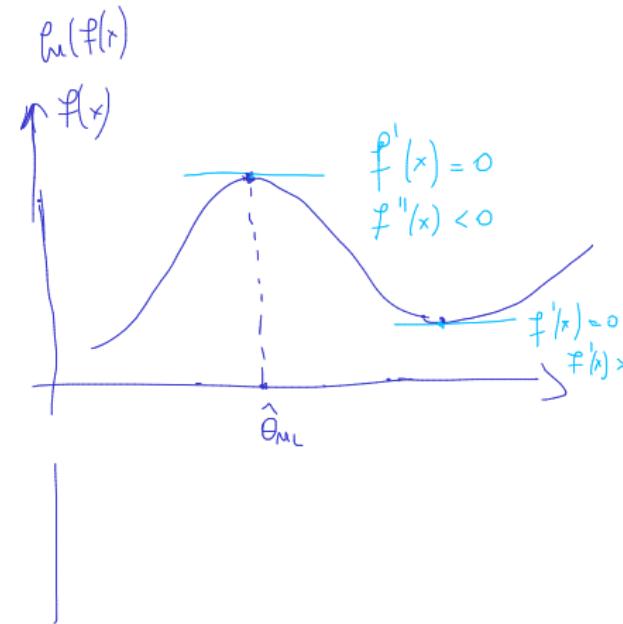
Găsirea maximului

- ▶ Cum se rezolvă problema de maximizare?
 - ▶ adică cum se găsește estimatul $\hat{\Theta}_{ML}$ care maximizează $L(\Theta|\text{vec}r)$
- ▶ Maximul se găsește prin derivare și egalare cu 0

$$\frac{dL(\Theta|\mathbf{r})}{d\Theta} = 0$$

- ▶ Se poate aplica **logaritmul natural** asupra funcției $L(\Theta|\mathbf{r})$ înainte de derivare (funcția "log-likelihood")

$$\frac{d \ln(L(\Theta|\mathbf{r}))}{d\Theta} = 0$$



arcpnolx $f(x) = \operatorname{argmax} \ln(f(x))$

Procedura de găsire a estimatului

Procedura de găsire a estimatului ML:

1. Se găsește expresia funcției

$$\underline{L(\Theta|\mathbf{r})} = w(\mathbf{r}|\Theta)$$

2. Se pune condiția ca derivata lui $L(\Theta|\mathbf{r})$ sau a lui $\ln(L(\Theta|\mathbf{r}))$ să fie 0

$$\frac{dL(\Theta)}{d\Theta} = 0, \text{ sau } \frac{d \ln(L(\Theta))}{d\Theta} = 0$$

3. Se rezolvă ecuația, se găsește valoarea $\hat{\Theta}_{ML}$
4. Se verifică că derivata a doua în punctul $\hat{\Theta}_{ML}$ este negativă, pentru a verifica că este un punct de maxim

- ▶ Întrucât derivata = 0 și pentru maxime și pentru minime
- ▶ uneori sărim peste această etapă

Exemple

$$D_\theta(t) = A$$

$$r_i(t) = A + \text{zgomot}, \quad N(\mu=0, \sigma^2 = \sigma^2),$$

$$r_i = [5, 7, 8, 6.1, 5.3]$$

$$L(\theta|r_i) = w(r_i|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(r_i - \theta)^2}{2\sigma^2}}$$

- ▶ Estimarea unui semnal constant în zgomot gaussian:

Găsiți estimatul Maximum Likelihood pentru un semnal de valoare constantă $s_\theta(t) = A$ din 5 măsurători afectate de zgomot

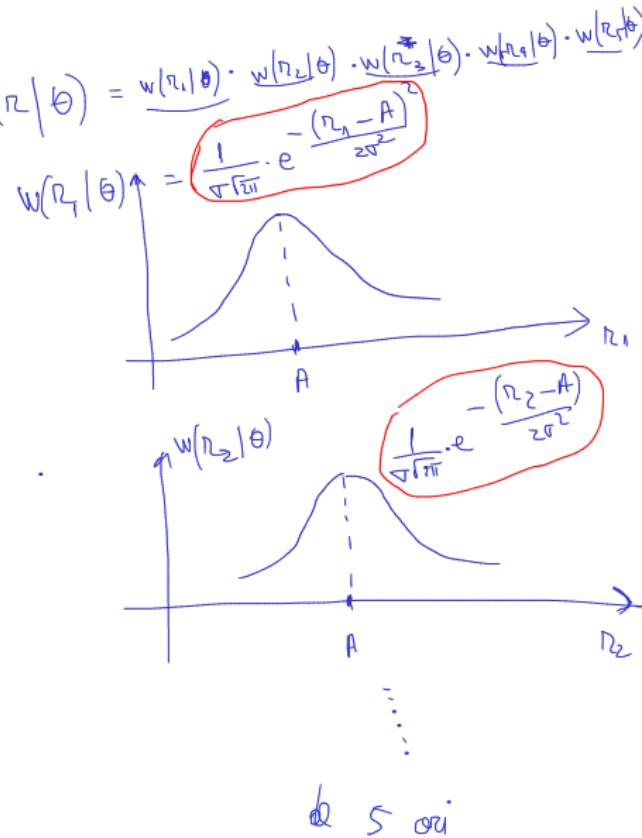
$r_i = A + \text{zgomot}$, cu valori egale cu $[5, 7, 8, 6.1, 5.3]$. Zgomotul este AWGN $N(\mu = 0, \sigma^2)$.

- ▶ Soluție: la tablă
- ▶ Estimatul \hat{A}_{ML} este chiar valoarea medie a eșantioanelor

▶ (deloc surprinzător)

$$L(A|r_i) = w(r_i|\theta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^5 \cdot e^{-\frac{(r_1-A)^2 + (r_2-A)^2 + (r_3-A)^2 + (r_4-A)^2 + (r_5-A)^2}{2\sigma^2}}$$

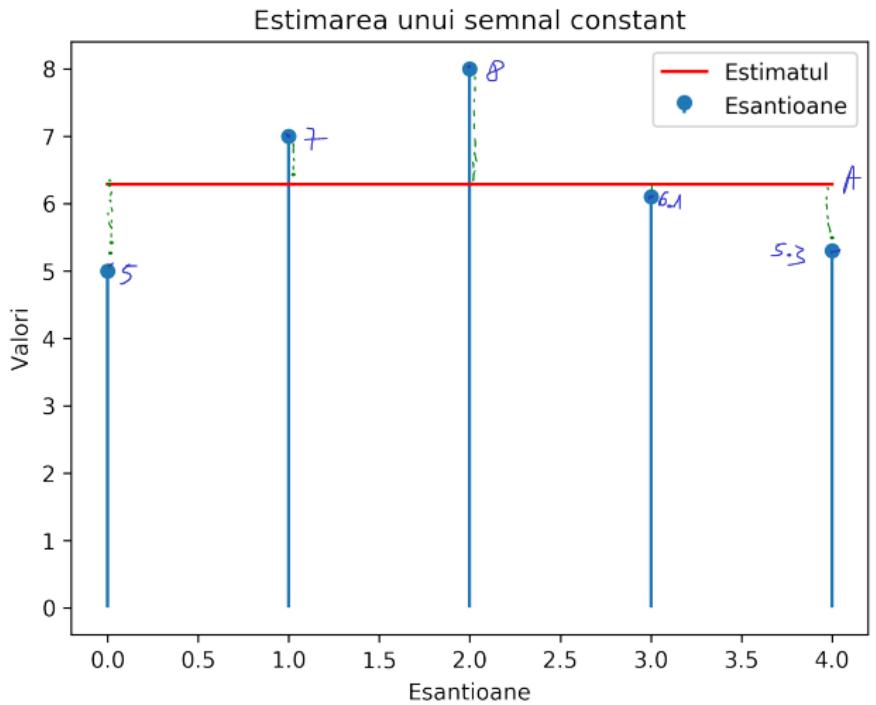
$$\ln(L(A|r_i)) = \ln\left(\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^5\right) + \frac{(5-A)^2 + (7-A)^2 + (8-A)^2 + (6.1-A)^2 + (5.3-A)^2}{2\sigma^2}$$



d 5 ori

Simulare numerică

$$\left| \frac{\partial}{\partial k} (\xi - A)^2 \right| = 2(5-A) \cdot (-1)$$



2) $\frac{d}{dA} \left(\ln L(A|z) \right) = 0 \quad (\Rightarrow)$

$$\frac{1}{2} \left(2(5-A) + 2(7-A) + 2(8-A) + 2(6.1-A) + 2(5.3-A) \right) = 0$$

$$(\Rightarrow) 5-A + 7-A + 8-A + 6.1-A + 5.3-A = 0$$

$$(\Rightarrow) 31.4 = 5A$$

$$\Rightarrow \hat{A}_{ML} = \frac{31.4}{5} = \frac{5+7+8+6.1+5.3}{5} = 6.28$$

Aproximare a unei curbe

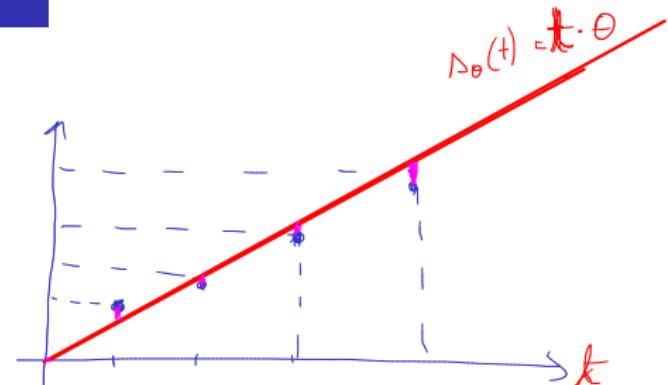
curve fitări

► Estimare = aproximare a unei curbe

- se găsește cea mai bună potrivire a lui $s_\theta(t)$ ^{cu} prin datele r

► Din exemplul grafic anterior:

- avem un set de date r
- se cunoaște forma semnalului = o dreaptă orizontală (A constant)
- se aproximează în mod optim dreapta prin setul de date



$$s_\theta(t) = ?$$

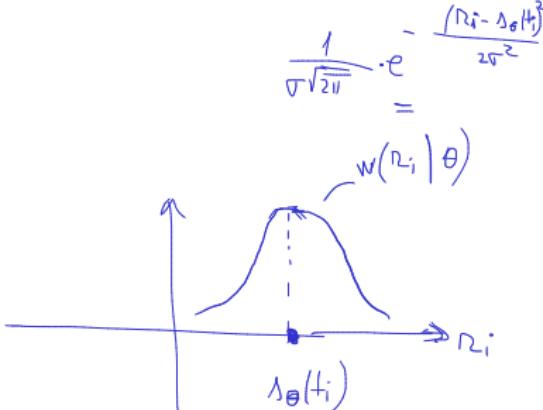
$$s_\theta(t) = \theta \cdot t$$

$$r_i = \underbrace{\theta \cdot t_i}_{s_\theta(t_i)} + z_{\text{găsat}}$$

Semnal oarecare în AWGN (Additive White Gaussian Noise)

- ▶ Fie semnalul original $s_\Theta(t)$
- ▶ Zgomotul este AWGN $\mathcal{N}(\mu = 0, \sigma^2)$
- ▶ Eșantioanele r_i sunt luate la momentele t_i
- ▶ Eșantioanele r_i au distribuție normală, cu media $\mu = s_\Theta(t_i)$ și varianța σ^2
- ▶ Funcția de plauzibilitate globală = produsul plauzibilităților fiecărui eșantion r_i

$$\begin{aligned} & s_\Theta(t) + z_{\text{gumot}} \\ t_i: r_i &= s_\Theta(t_i) + z_{\text{gumot}} \end{aligned}$$



$$\begin{aligned} L(\Theta | \mathbf{r}) = w(\mathbf{r} | \Theta) &= \prod_{i=1}^N \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(r_i - s_\Theta(t_i))^2}{2\sigma^2}} \\ &= \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^N \cdot e^{-\frac{\sum_{i=1}^N (r_i - s_\Theta(t_i))^2}{2\sigma^2}} \end{aligned}$$

Semnal oarecare în AWGN

- Logaritmul plauzibilității ("log-likelihood") este

$$\ln(L(\Theta|\mathbf{r})) = \underbrace{\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N}_{\text{constant}} - \frac{\sum(r_i - s_\Theta(t_i))^2}{2\sigma^2}$$

✓ din Θ care face funcția asta să fie maximă

Vrem termenul acelor
căt mai mici !

$$d(r, s_\Theta)$$

$$r = [r_1 \ r_2 \ \dots \ r_n]$$
$$s_\Theta = [s_\Theta(t_1) \ s_\Theta(t_2) \ \dots \ s_\Theta(t_n)]$$

Semnal oarecare în AWGN

- Maximul funcției = minimul exponentului

$$\hat{\Theta}_{ML} = \arg \max_{\Theta} L(\Theta | \mathbf{r}) = \arg \min_{\Theta} \sum_{i=1}^n (r_i - s_{\Theta}(t_i))^2 = \underset{\Theta}{\operatorname{arg\,min}} d(\mathbf{r}, s_{\Theta})^2$$

AWGN
număratorul

- Termenul $\sum(r_i - s_{\Theta}(t_i))^2$ este **distanța $d(\mathbf{r}, s_{\Theta})$ la pătrat**

$$d(\mathbf{r}, s_{\Theta}) = \sqrt{\sum (r_i - s_{\Theta}(t_i))^2}$$

$$(d(\mathbf{r}, s_{\Theta}))^2 = \sum (r_i - s_{\Theta}(t_i))^2$$

- ▶ Estimarea ML se poate scrie sub forma:

$$\hat{\Theta}_{ML} = \arg \max_{\Theta} L(\Theta | \mathbf{r}) = \arg \min_{\Theta} d(\mathbf{r}, \mathbf{s}_{\Theta})^2$$

- ▶ Estimatul de plauzibilitate maximă (ML) $\hat{\Theta}_{ML}$ = valoarea care face $s_{\Theta}(t_i)$ **cel mai apropiat de vectorul recepționat r**
 - ▶ mai aproape = potrivire mai bună = mai probabil
 - ▶ cel mai aproape = cea mai bună potrivire = cel mai probabil = plauzibilitate maximă

Semnal oarecare în AWGN

- ▶ Estimare ML în zgomot gaussian = minimizarea distanței
- ▶ Aveam aceeași interpretare și la decizia ML!
 - ▶ dar la decizie alegeam minimul din 2 opțiuni
 - ▶ aici alegem minimul dintre toate opțiunile posibile
- ▶ Relația e valabilă pentru orice fel de spații vectoriale
 - ▶ vectori cu N elemente, semnale continue, etc
 - ▶ doar se înlocuiește definiția distanței Euclidiene

Exemplu: problema de pe slide-ul 18 :

$$\mathbf{r} = [5 \quad 7 \quad 8 \quad 6.1 \quad 5.3]$$

$$\mathbf{s}_0 = [A \quad A \quad A \quad A \quad A]$$

$$D = \sigma^2(\mathbf{r}, \mathbf{s}_0) = (A-5)^2 + (A-7)^2 + (A-8)^2 + (A-6.1)^2 + (A-5.3)^2$$

$$\frac{\partial D}{\partial A} = \cancel{2(A-5)} + \cancel{2(A-7)} + \cancel{2(A-8)} + \cancel{2(A-6.1)} + \cancel{2(A-5.3)} \\ = 0$$

$$\Rightarrow \hat{A}_{ML} = \frac{5+7+8+6.1+5.3}{5} = 6.28$$

$$\hat{A}_{ML} = \frac{r_1 + r_2 + \dots + r_N}{N} = \frac{1}{N} \sum_{i=1}^N r_i \quad \begin{matrix} \text{medie} \\ \text{aritmetică!} \end{matrix}$$

Semnal oarecare în AWGN

Procedura pentru estimarea tip ML în zgomot AWGN:

1. Se scrie expresia pentru pătratul distanței:

$$D = (d(\mathbf{r}, s_\Theta))^2 = \sum (r_i - s_\Theta(t_i))^2$$

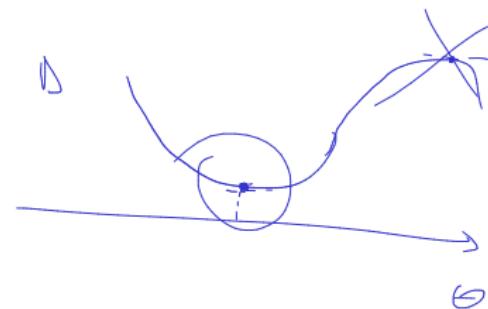
2. Vrem minimul, deci egalăm derivata cu 0:

$$\frac{dD}{d\Theta} = \sum 2(r_i - s_\Theta(t_i))(-\frac{ds_\Theta(t_i)}{d\Theta}) = 0$$

3. Se rezolvă și obținem valoarea $\hat{\Theta}_{ML}$

4. Se verifică că derivata a doua în punctul $\hat{\Theta}_{ML}$ este pozitivă, pentru a se verifica că punctul este un minim

- ▶ uneori sărim peste această etapă



Estimarea frecvenței f a unui semnal sinusoidal

- ▶ Găsiți estimatul Maximum Likelihood pentru frecvența f a unui semnal $s_\Theta(t) = \cos(2\pi ft_i)$, din 10 măsurători afectate de zgomot $r_i = \cos(2\pi ft_i) + \text{zgomot}$ de valori [...]. Zgomotul este AWGN $\mathcal{N}(\mu = 0, \sigma^2)$. Momentele de eșantionare sunt

$$t_i = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]$$

- ▶ Soluție: la tablă

$$\underline{r} = [r_1, r_2, \dots, r_{10}]$$

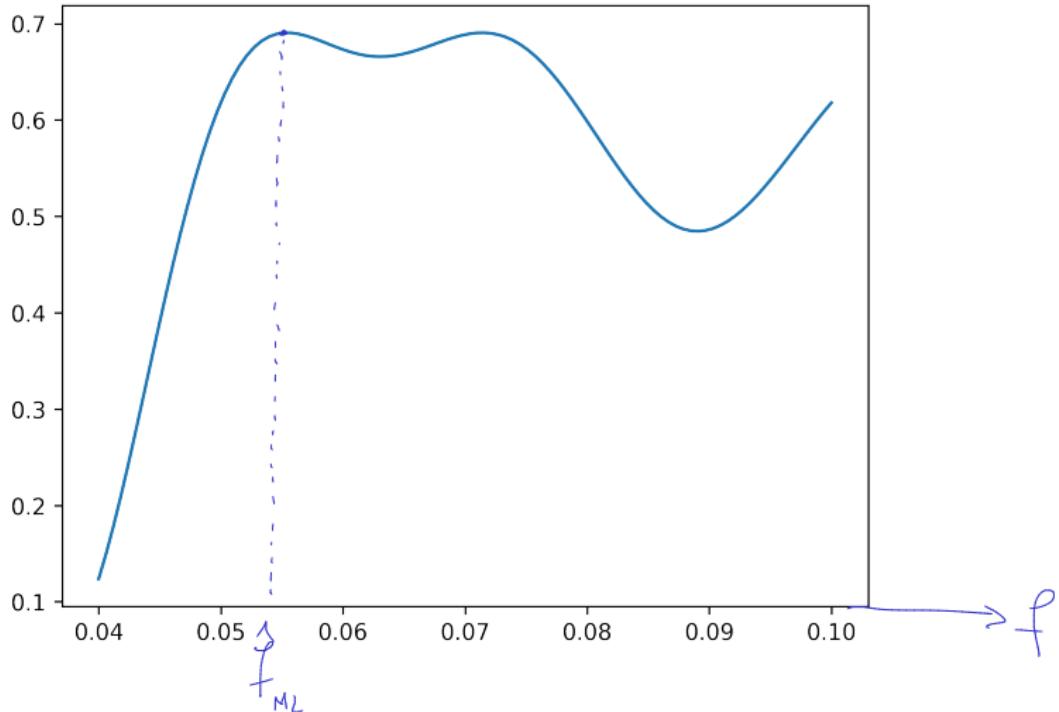
$$\underline{s}_0 = [\cos(2\pi f t_0), \cos(2\pi f t_0), \dots, \cos(2\pi f t_0)]$$

$$D = d(r, s_0)^2$$

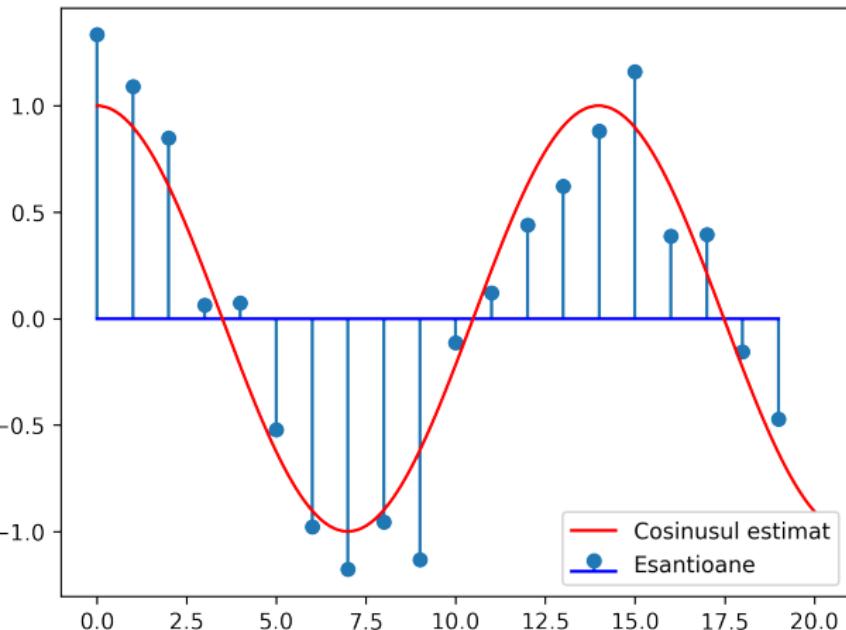
Simulare numerică

Funcția de plauzibilitate este

$$L(\theta | \mathbf{r}) = L(f | \mathbf{r}) \quad (\theta = f)$$



Simulare numerică



Parametri multipli

- Dacă semnalul depinde de mai mulți parametri?

- de ex. amplitudinea, frecvența și faza inițială a unui cosinus:

$$s(t) = A \cos(2\pi ft + \phi)$$

- Se va considera Θ ca fiind un vector:

$$\Theta = [A, f, \phi]$$

$$\Theta = [\Theta_1, \Theta_2, \dots, \Theta_M]$$

- e.g. $\Theta = [\Theta_1, \Theta_2, \Theta_3] = [A, f, \phi]$

Parametri multipli

rezonat gaussian:

$$D = d(r, \underbrace{s_0}_{A, f, \varphi})^2 = D(A, f, \varphi) \text{ depinde de mai multe necunoscute}$$

- Se rezolvă cu aceeași procedură, dar în loc de o singură derivată vom avea M deriveate
- Se rezolvă sistemul:

$$\begin{cases} \frac{\partial L}{\partial \Theta_1} = 0 \\ \frac{\partial L}{\partial \Theta_2} = 0 \\ \dots \\ \frac{\partial L}{\partial \Theta_M} = 0 \end{cases}$$

$$r = [r_1, r_2, r_3, r_4, \dots]$$

$$s_0 = [A \cos(2\pi f_1 t_1 + \varphi), A \cos(2\pi f_2 t_2 + \varphi), A \cos(2\pi f_3 t_3 + \varphi), A \cos(2\pi f_4 t_4 + \varphi), \dots]$$

$$D = d(r, s_0)^2 = (A \cos(2\pi f_1 t_1 + \varphi) - r_1)^2 + \dots + (A \cos(2\pi f_4 t_4 + \varphi) - r_4)^2$$

- uneori este dificil/imposibil

$$\left\{ \begin{array}{l} \frac{\partial D}{\partial A} = 0 \\ \frac{\partial D}{\partial f} = 0 \\ \frac{\partial D}{\partial \varphi} = 0 \end{array} \right. \Rightarrow A, f, \varphi = \dots$$

Coborâre după gradient (Gradient Descent)

- ▶ Cum se estimează parametrii Θ în cazuri complicate?
 - ▶ În aplicații reale, unde pot fi foarte mulți parametri (Θ este vector)
- ▶ De obicei nu se pot găsi valorile optime prin formule directe
- ▶ Se îmbunătățesc valorile în mod iterativ cu algoritmi tip coborâre după gradient (Gradient Descent)

Coborâre după gradient (Gradient Descent)

1. Se initializează parametrii cu valori aleatoare $\Theta^{(0)}$

2. Repetă la fiecare iteratie k :

2.1 Se calculează funcția $J(\Theta^{(k)} | \mathbf{r})$

2.2 Se calculează derivatele $\frac{\partial J}{\partial \Theta_i^{(k)}}$ pentru toți Θ_i ("Gradient")

2.3 Se actualizează toate valorile Θ_i , prin scăderea derivatei ("Descent"):

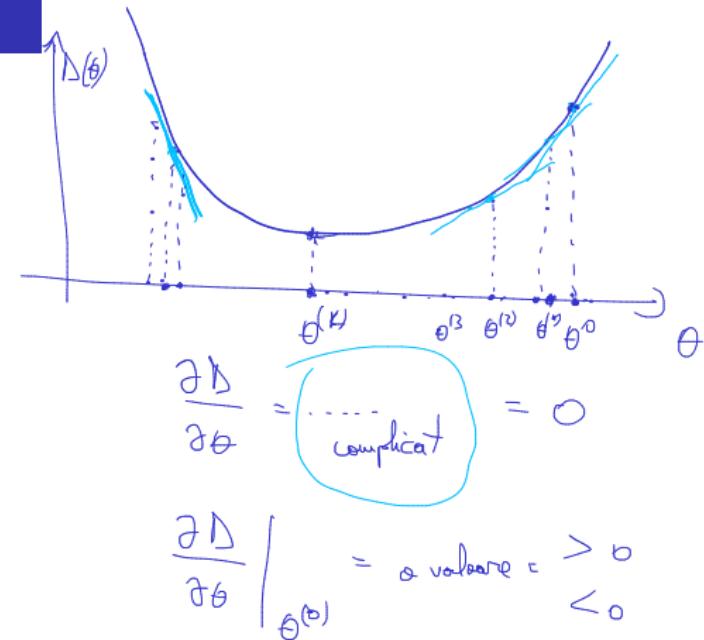
$$\Theta_i^{(k+1)} = \Theta_i^{(k)} - \mu \frac{\partial J}{\partial \Theta_i^{(k)}}$$

new value old value step size derivative

► sau, sub formă vectorială:

$$\Theta^{(k+1)} = \Theta^k - \mu \frac{\partial J}{\partial \Theta^{(k)}}$$

3. Până la îndeplinirea unui criteriu de terminare (de ex. parametrii nu se mai modifică mult)



Coborâre după gradient (Gradient Descent)

- ▶ Explicații la tablă
- ▶ Exemplu: regresia logistică cu valori 2D
 - ▶ exemplu la tablă

- ▶ Cel mai proeminent exemplu: **Rețele Neurale Artificiale** (a.k.a. “Rețele Neurale”, “Deep Learning”, etc.)
 - ▶ Pot fi văzute ca un exemplu de estimare ML
 - ▶ Se utilizează algoritmul Gradient Descent, pentru găsirea parametrilor
 - ▶ Aplicații de vârf: recunoașterea de imagini, automated driving etc.
- ▶ Mai multe informații despre rețele neurale / machine learning:
 - ▶ căutați cursuri sau cărți online
 - ▶ IASI AI Meetup

Deplasarea și varianta estimatorilor

Exemplu:

$$n = \begin{bmatrix} 5 & 7 & 8 & 6.1 & 5.3 \end{bmatrix}$$

$$\hat{\theta} = \begin{bmatrix} A & A & A & A & A \end{bmatrix}$$

$$\hat{\theta}_{ML} = \frac{1}{N} \sum_{i=1}^N n_i$$

- ▶ Cum caracterizăm calitatea unui estimator?
- ▶ Un estimator $\hat{\theta}$ este o variabilă aleatoare
 - ▶ poate avea diverse valori, pentru că se calculează pe baza eșantioanelor receptionate, care depend de zgomot
 - ▶ exemplu: se repetă aceeași estimare pe calculatoare diferite \Rightarrow valori estimate ușor diferite
- ▶ Fiind o variabilă aleatoare, se pot defini:
 - ▶ valoarea medie a estimatorului: $E\{\hat{\theta}\}$ = media valorilor estimate
 - ▶ varianta estimatorului: $E\{(\hat{\theta} - \bar{\theta})^2\}$

estimatorul
nostru

$E\{\hat{\theta}\}$

= media
estimatorilor

Deplasarea și varianța estimatorilor

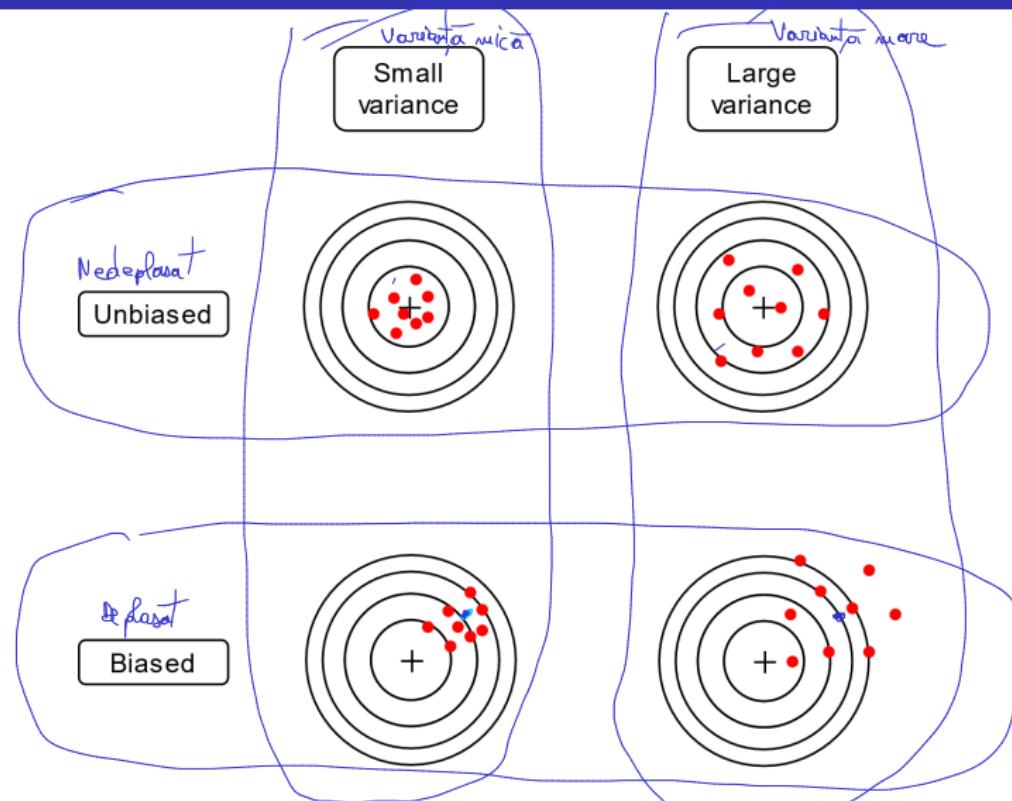


Figure 1: Deplasarea și varianța estimatorilor

Deplasarea unui estimator

- **Deplasarea** ("bias") unui estimator = diferența dintre valoarea medie a estimatorului și valoarea adevărată Θ

$$\text{Deplasare} = E \{ \hat{\Theta} \} - \Theta$$

- Estimator **nedeplasat** = valoarea medie a estimatorului este egală cu valoarea adevărată a parametrului Θ

$$E \{ \hat{\Theta} \} = \Theta$$

- Estimator **deplasat** = valoarea medie a estimatorului diferă de valoarea adevărată a parametrului Θ

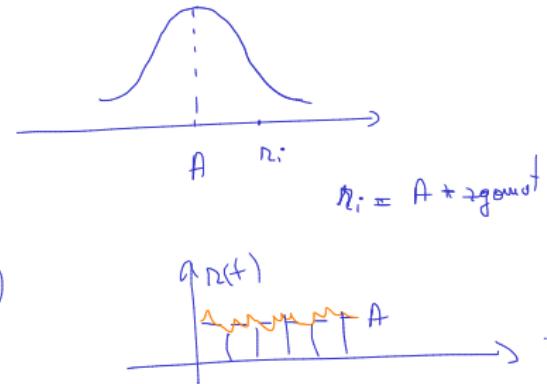
- diferența $E \{ \hat{\Theta} \} - \Theta$ este **deplasarea** estimatorului

Deplasarea unui estimator

- Exemplu: semnal constant A , zgomot Gaussian (cu media $\mu = 0$ și variansă $\sigma^2 = 1$), estimatorul de plauzibilitate maximă este $\hat{A}_{ML} = \frac{1}{N} \sum_i r_i$

► Atunci:

$$\begin{aligned} E\{\hat{A}_{ML}\} &= \frac{1}{N} E\left\{\sum_i r_i\right\} \\ &= \frac{1}{N} \sum_{i=1}^N E\{r_i\} \\ &= \frac{1}{N} \sum_{i=1}^N E\{A + \text{zgomot}\} \\ &= \frac{1}{N} \sum_{i=1}^N A + E\{\text{zgomot}\} \\ &= \frac{1}{N} \sum_{i=1}^N A \\ &= A \end{aligned}$$



- Acum acest estimator este nedeplasat

$$\begin{cases} E\{x+y\} = E\{x\} + E\{y\} \\ E\{a \cdot x\} = a \cdot E\{x\} \end{cases}$$

Varianța unui estimator

$$\Delta_\theta(\hat{\theta}) = A$$

Exemplu:

$$\hat{A}_{ML} = \frac{1}{N} \sum_{i=1}^N r_i$$

$$\nabla^2 \hat{A} = ?$$

Vrem să afiliem

$$Z = X + Y$$

$$\nabla^2 Z = \nabla^2 X + \nabla^2 Y$$

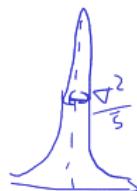
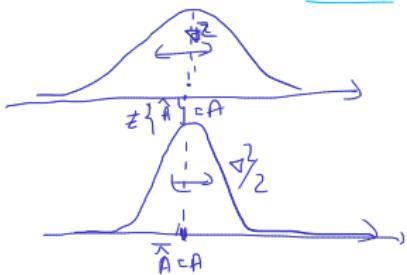
dacă X, Y necorelate

$$Z = a \cdot X$$

$$\nabla^2 Z = a^2 \cdot \nabla^2 X$$

- ▶ **Varianța** unui estimator măsoară "abaterile" estimatorului în jurul valorii medii

- ▶ aceasta e definiția varianței σ^2 în general
- ▶ Dacă un estimator are **varianță mare**, valoarea estimată poate fi departe de cea reală, chiar dacă estimatorul este nedeplasat
- ▶ De obicei se preferă estimatori cu **varianță mică**, tolerându-se o eventuală mică deplasare



$$\nabla^2 \{ \hat{A}_{ML} \} = \frac{1}{N} \cdot \nabla^2 z_{\text{genat}}$$

∴

$$\begin{aligned} \nabla^2 \{ \hat{A}_{ML} \} &= \nabla^2 \left\{ \frac{1}{N} \cdot \sum_{i=1}^N r_i \right\} \\ &= \left(\frac{1}{N} \right)^2 \cdot \nabla^2 \left\{ \sum_{i=1}^N r_i \right\} \\ &= \left(\frac{1}{N} \right)^2 \cdot \sum \nabla^2 \{ r_i \} \\ &= \left(\frac{1}{N} \right)^2 \cdot \sum \nabla^2 \{ A + z_{\text{genat}} \} \\ &\quad \text{Nabla}^2 \{ A \} = 0 \quad \text{Nabla}^2 \{ z_{\text{genat}} \} = 0 \\ &= \left(\frac{1}{N} \right)^2 \cdot \sum_{i=1}^N \nabla^2 z_{\text{genat}} \\ &= \frac{1}{N} \cdot \nabla^2 z_{\text{genat}} \end{aligned}$$

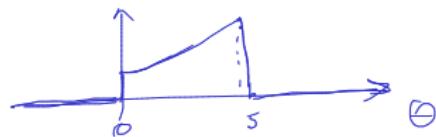
II.3 Estimare Bayesiană

- ▶ **Estimarea Bayesiană** ia în calcul termeni suplimentari pe lângă

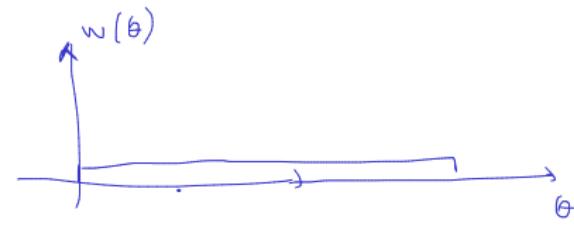
$$L(\theta|n) = w(r|\Theta)$$

- ▶ o distribuție a priori $w(\Theta)$
- ▶ optional, o funcție de cost

$$w(\theta)$$



- ▶ Se obține echivalentul din estimare pentru criteriile de decizie MPE și MR



Estimare Bayesiană

- ▶ Se definește **distribuția a posteriori** a lui Θ , dat fiind observațiile r , folosind **regula lui Bayes**:

$$w(\theta|r) = \frac{w(r|\theta) \cdot w(\theta)}{w(r)}$$

w(r) = ct = nu ne influențează

distribuția a posteriori

distribuția a priori

- ▶ Termenii:

- ▶ Θ este parametrul necunoscut
- ▶ r este vectorul de observații
- ▶ $w(\theta|r)$ este probabilitatea ca parametrul să aibă valoarea θ , dat fiind vectorul de observații r ;
- ▶ $w(r|\theta)$ este funcția de plauzibilitate $= L(\theta|r)$
- ▶ $w(\theta)$ este distribuția **a priori** a lui θ
- ▶ $w(r)$ este distribuția **a priori** a lui r ; se presupune a fi constantă

r = cunoscut $\Rightarrow w(r) = \text{un număr}$

M.L.: $L(\theta|r) = w(r|\theta)$

Contur: nu e o lățime fixă de θ

Bayes: $w(\theta|r)$

Regula lui Bayes:

$$P(A \cap B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

(=)

$$\begin{aligned} P(A \cap B) &= P(B|A) \cdot P(A) \\ &= P(A|B) \cdot P(B) \end{aligned}$$

Exemplu

$$\begin{aligned} P(NV|D) &= 90\% \\ P(V|D) &= 10\% \\ P(D|NV) &= \frac{P(NV|D) \cdot P(D)}{P(NV)} \\ P(D|V) &= \frac{P(V|D) \cdot P(D)}{P(V)} \end{aligned}$$



Regula lui Bayes

$$w(\theta | r)$$

- Relația precedentă arată că, în general, estimarea lui Θ depinde de două lucruri:

1. De vectorul observațiilor r , prin termenul $w(r|\Theta)$
2. De informația "a priori" avută despre Θ , prin termenul $w(\Theta)$

- Numele este "estimare Bayesiană"

- Thomas Bayes = matematician englez, a descoperit regula cu acest nume
- Notiunile bazate pe regula lui Bayes poartă deseori numele de "Bayesiane"

$$\left\{ \begin{array}{l} \frac{P(D|NV)}{P(D|V)} = \frac{P(NV|D) \cdot P(V)}{P(NV)} \cdot \frac{P(V)}{P(V|D) \cdot P(N)} \\ \frac{P(D|NV)}{P(D|V)} = \frac{P(NV|D)}{P(V|D)} \cdot \frac{P(V)}{P(NV)} \end{array} \right.$$

$$\frac{P(D|NV)}{P(D|V)} = \frac{0.9}{0.1} \cdot \frac{0.3}{0.7}$$

$$\frac{P(D|NV)}{P(D|V)} = \frac{9}{1} \cdot \frac{3}{7} = \frac{27}{7} \approx 4$$

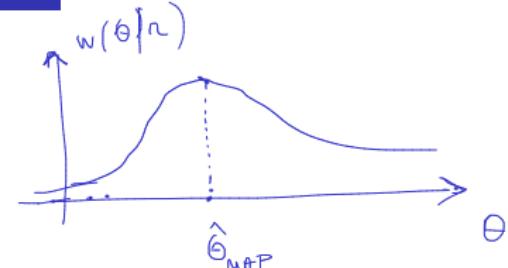
Distribuția a priori

- ▶ Presupunem că se știe de dinainte o distribuție a lui Θ , $w(\Theta)$
 - ▶ știm de dinainte care e probabilitatea de a fi a anume valoare sau alta
 - ▶ se numește distribuția a priori
- ▶ Estimarea trebuie să ia în calcul și distribuția a priori $w(\Theta)$
 - ▶ estimatul va fi “tras” înspre valori mai probabile

Estimatorul MAP

- ▶ Cunoaștem distribuția a posteriori $w(\Theta|r)$. Care este valoarea estimată?
- ▶ Se poate alege valoarea care are probabilitate maximă
- ▶ Estimatorul Maximum A Posteriori (MAP) este :

$$\hat{\Theta}_{MAP} = \arg \max_{\Theta} w(\Theta|r) = \arg \max_{\Theta} w(r|\Theta) \cdot w(\Theta)$$



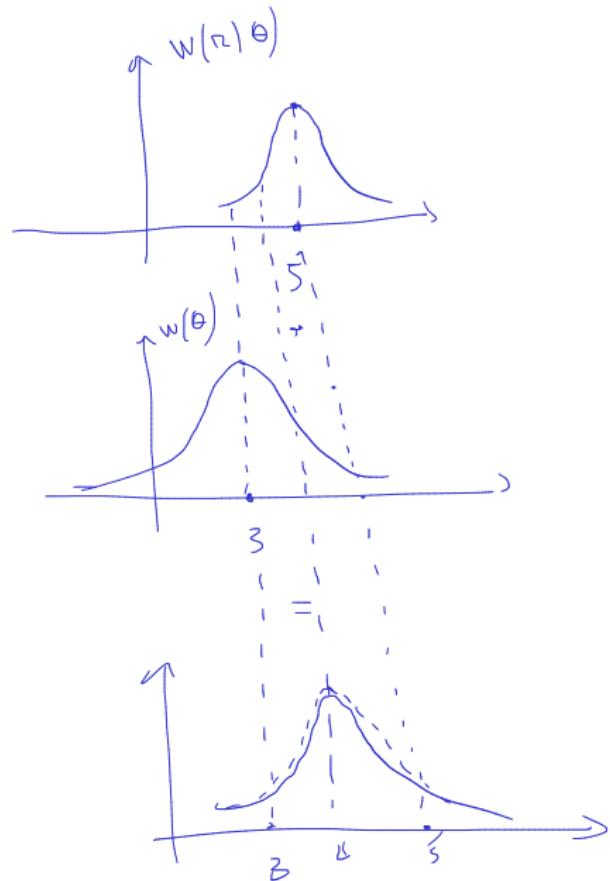
- ▶ Estimatorul MAP alege acea valoare Θ unde distribuția a posteriori $w(\Theta|r)$ este maximă
- ▶ Estimatorul MAP maximizează **produsul** dintre plauzibilitate și **distribuția a priori** $w(\Theta)$

Estimatorul MAP

$$\text{r}_n(t) = f + \text{nois} + N(\mu=0, \sigma^2=3)$$

$$\sigma_n = 5$$

Exemplu: Imagine



Relația dintre estimarea MAP și ML

- ▶ Estimatorul ML:

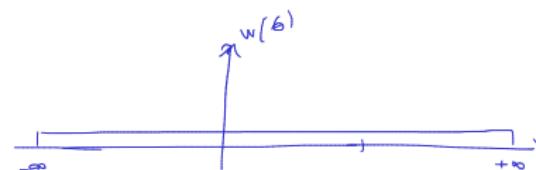
$$\hat{\Theta}_{ML} = \arg \max w(\mathbf{r}|\Theta)$$

- ▶ Estimatorul MAP:

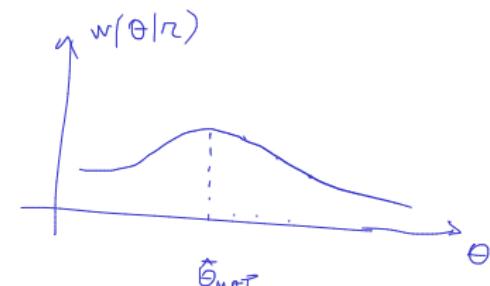
$$\hat{\Theta}_{MAP} = \arg \max w(\mathbf{r}|\Theta) \cdot w(\Theta)$$

- ▶ Estimatorul ML este un caz particular de MAP pentru $w(\Theta)$ constant

- ▶ $w(\Theta) = \text{constant}$ înseamnă că toate valorile lui Θ sunt *a priori* echiprobabile
- ▶ i.e. nu avem extra informații despre valoarea lui Θ



- ▶ Criteriul probabilității minime de eroare:
$$\frac{w(r|H_1)}{w(r|H_0)} \stackrel{H_1}{\gtrless} \stackrel{H_0}{\lless} \frac{P(H_0)}{P(H_1)}$$
- ▶ Se poate rescrie ca $w(r|H_1) \cdot P(H_1) \stackrel{H_1}{\gtrless} \stackrel{H_0}{\lless} w(r|H_0)P(H_0)$
 - ▶ adică se alege ipoteza pentru care $w(r|H_i) \cdot P(H_i)$ este mai mare
- ▶ **Criteriul de decizie MPE:** se alege ipoteza care maximizează $w(r|H_i) \cdot P(H_i)$
 - ▶ dintre cele două ipoteze H_0, H_1
- ▶ **Estimarea MAP:** se alege valoarea care maximizează $w(\mathbf{r}|\Theta) \cdot w(\Theta)$
 - ▶ dintre toate valorile posibile pentru Θ
- ▶ Același principiu!



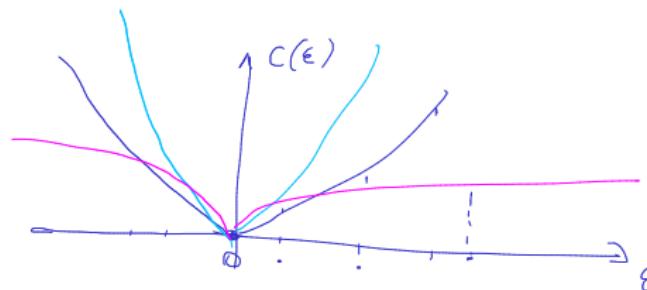
Funcția de cost

- ▶ Vrem să găsim un echivalent și pentru criteriul MR
- ▶ Avem nevoie de un echivalent pentru costurile C_{ij}
- ▶ Eroarea de estimare = diferența între estimatul $\hat{\Theta}$ și valoarea reală Θ

$$\epsilon = \hat{\Theta} - \Theta$$

- ▶ Funcția de cost $C(\epsilon)$ = atribuie un cost pentru fiecare eroare de estimare posibilă

- ▶ când $\epsilon = 0$, costul $C(0) = 0$
- ▶ erori ϵ mici au costuri mici
- ▶ erori ϵ mari au costuri mari



Functia de cost

- ▶ Funcții de cost uzuale:

- ▶ Pătratică:

$$C(\epsilon) = \epsilon^2 = (\hat{\Theta} - \Theta)^2$$

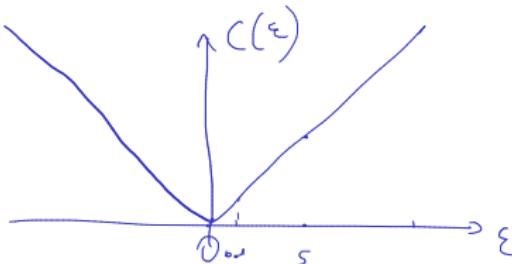
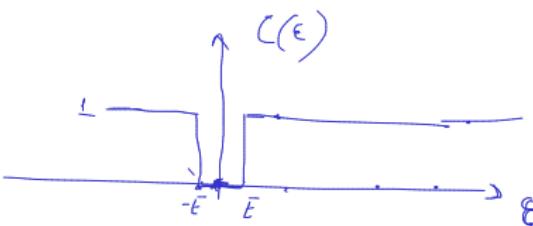
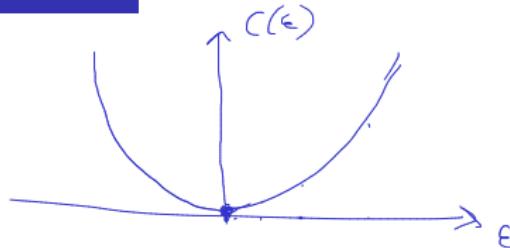
- ▶ Uniformă:

$$C(\epsilon) = \begin{cases} 0, & \text{if } |\epsilon| = |\hat{\Theta} - \Theta| \leq E \\ 1, & \text{if } |\epsilon| = |\hat{\Theta} - \Theta| > E \end{cases}$$

- ▶ Liniară:

$$C(\epsilon) = |\epsilon| = |\hat{\Theta} - \Theta|$$

- ▶ De desenat la tablă



Funcția de cost

- ▶ Funcția de cost $C(\epsilon)$ reprezintă echivalentul costurilor C_{ij} de la detectie
 - ▶ la detectie aveam doar 4 valori: C_{00} , C_{01} , C_{10} , C_{11}
 - ▶ aici avem un cost pentru fiecare eroare posibilă ϵ
- ▶ Funcția de cost dictează ce valoarea alegem din distribuția $w(\Theta|r)$

I Importanța funcției de cost

- Fie distribuția a posteriori următoare:

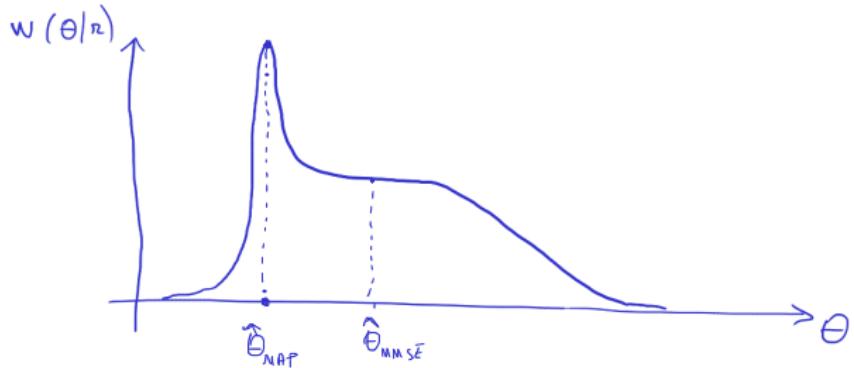


Figure 2: Unbalanced posterior distribution

- Care este estimatorul MAP? $\hat{\theta}_{MAP} =$

- Dacă avem funcția de cost următoare:

- dacă estimarea $\hat{\theta}$ este $<$ valoarea reală θ , te costă 1000 \$
- dacă estimarea $\hat{\theta}$ este $>$ valoarea reală θ , platești 1 \$
- schimbăm valoarea estimată ? :)

Joc :

1	2	3	4	5	6
50%	5%	5%	5%	5%	30%

→ 1). $C(\epsilon) = \text{uniformă} \Rightarrow \text{Ghicesc 1}$

→ 2). $C(\epsilon) = \epsilon^2 = (\hat{\theta} - \theta)^2$

Ghicesc 1 :

$$50\% \cdot 0 + 5\% \cdot 1^2 + 5\% \cdot 2^2 + 5\% \cdot 3^2 +$$

$$+ 5\% \cdot 4^2 + 30\% \cdot 25 = \\ = 0.05 \cdot 30 + 0.3 \cdot 25 =$$

Ghicesc 2 :

$$50\% \cdot 1^2 + 5\% \cdot 0 + 5\% \cdot 1^2 + 5\% \cdot 2^2 +$$

$$5\% \cdot 3^2 + 30\% \cdot 4^2 = \\ = 0.5 + 0.05 \cdot 19 + 0.3 \cdot 16 =$$

- ▶ Distribuția a posteriori $w(\Theta|r)$ dă probabilitatea fiecărei valori $\hat{\Theta}$ de a fi cea corectă
- ▶ Alegerea unui estimat $\hat{\Theta}$ implică o anume eroare ϵ
- ▶ Eroarea de estimare are un anumit cost $C(\epsilon)$
- ▶ **Riscul** = valoarea medie a costului = $C(\epsilon) \times$ probabilitatea:

$$R = \int_{-\infty}^{\infty} C(\epsilon) w(\Theta|r) d\Theta$$

$$= \int_{-\infty}^{\infty} C(\epsilon) \cdot w(\Theta|r) d\Theta$$

The Bayesian risk

- ▶ Alegem valoarea $\hat{\Theta}$ care minimizează costul mediu R

$$\hat{\Theta} = \arg \min_{\Theta} \int_{-\infty}^{\infty} C(\epsilon) w(\Theta | \mathbf{r}) d\Theta$$

$\hat{\Theta}$ (blue bracket)

- ▶ O obținem înlocuind $C(\epsilon)$ cu definiția sa, și derivând după $\hat{\Theta}$
 - ▶ Atenție: se derivează după $\hat{\Theta}$, nu Θ !

Estimatorul EPMM (eroare pătratică medie minimă) (MMSE)

- Când funcția de cost este pătratică $C(\epsilon) = \epsilon^2 = (\hat{\Theta} - \Theta)^2$

$$R = \int_{-\infty}^{\infty} (\hat{\Theta} - \Theta)^2 w(\Theta | \mathbf{r}) d\Theta$$

? C(ε) w(Θ|r) a posteriori

$$\frac{d}{d\hat{\Theta}} \left[\dots \right] d\Theta = \int_{-\infty}^{\infty} \frac{d}{d\hat{\Theta}} (\dots) d\Theta$$

- Vrem $\hat{\Theta}$ care minimizează R , deci derivăm

$$\frac{dR}{d\hat{\Theta}} = 2 \int_{-\infty}^{\infty} (\hat{\Theta} - \Theta) w(\Theta | \mathbf{r}) d\Theta = 0$$

- Echivalent cu

$$\hat{\Theta} \underbrace{\int_{-\infty}^{\infty} w(\Theta | \mathbf{r}) d\Theta}_{1} = \int_{-\infty}^{\infty} \Theta w(\Theta | \mathbf{r}) d\Theta$$

1 media distri. $\int x \cdot w(x) dx$

- Estimatorul de eroare pătratică medie minimă (EPMM) ("Minimum Mean Squared Error, MMSE"):

$$\hat{\Theta}_{EPMM} = \int_{-\infty}^{\infty} \Theta \cdot w(\Theta | \mathbf{r}) d\Theta$$

$$\begin{aligned}
 \frac{dR}{d\hat{\Theta}} &= \left(\frac{d}{d\hat{\Theta}} \int_{-\infty}^{\infty} (\hat{\Theta} - \Theta)^2 w(\Theta | \mathbf{r}) d\Theta \right) = \\
 &= \int_{-\infty}^{\infty} 2(\hat{\Theta} - \Theta) \cdot w(\Theta | \mathbf{r}) d\Theta = 0 \\
 &= \int_{-\infty}^{\infty} \hat{\Theta} \cdot w(\Theta | \mathbf{r}) d\Theta - \int_{-\infty}^{\infty} \Theta \cdot w(\Theta | \mathbf{r}) d\Theta = 0
 \end{aligned}$$

- ▶ **Estimatorul EPMM:** estimatorul $\hat{\Theta}$ este **valoarea medie a distribuției a posteriori** $w(\Theta|\mathbf{r})$

$$\hat{\Theta}_{EPMM} = \int_{-\infty}^{\infty} \Theta \cdot w(\Theta|\mathbf{r}) d\Theta$$

$$\int x \cdot w(x) dx$$

- ▶ EPMM = “Eroare Pătratică Medie Minimă”
- ▶ valoarea medie = sumă (integrală) din fiecare Θ ori probabilitatea sa $w(\Theta|\mathbf{r})$
- ▶ Estimatprul EPMM se obține din distribuția a posteriori $w(\Theta|\mathbf{r})$, considerând funcția de cost pătratică

Estimatorul MAP

- Dacă funcția de cost este uniformă:

$$C(\epsilon) = \begin{cases} 0, & \text{if } |\epsilon| = |\hat{\theta} - \theta| \leq E \\ 1, & \text{if } |\epsilon| = |\hat{\theta} - \theta| > E \end{cases}$$

- Stim că $\theta = \hat{\theta} - \epsilon$
- Se obține

$$R = \int_{-\infty}^{\infty} C(\epsilon) \cdot w(\theta|r) d\theta$$

$$R = \int_{-\infty}^{\hat{\theta}-E} w(\theta|r) d\theta + \int_{\hat{\theta}+E}^{\infty} w(\theta|r) d\theta$$

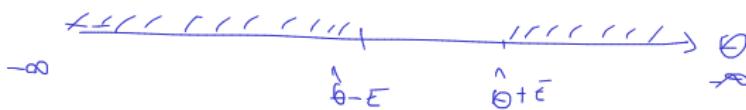
$$R = 1 - \int_{\hat{\theta}-E}^{\hat{\theta}+E} w(\theta|r) d\theta$$

$$\int_{-\infty}^{\infty} w(\theta|r) d\theta = 1$$

$$\begin{aligned} \epsilon &= \hat{\theta} - \theta \Rightarrow \theta = \hat{\theta} - \epsilon \\ C(\epsilon) &\stackrel{\text{cond}}{=} 0 \quad |\epsilon| \leq E \\ &-E \leq \epsilon \leq E \\ &-\epsilon \leq \hat{\theta} - \theta \leq E \quad |-\hat{\theta}| \\ &-E - \hat{\theta} \leq -\theta \leq E - \hat{\theta} \\ \boxed{\hat{\theta} - E \geq \theta \geq \hat{\theta} + E} \end{aligned}$$

$\text{Cond} \quad \theta \in [\hat{\theta} - E, \hat{\theta} + E] \Rightarrow C(\epsilon) = 0$

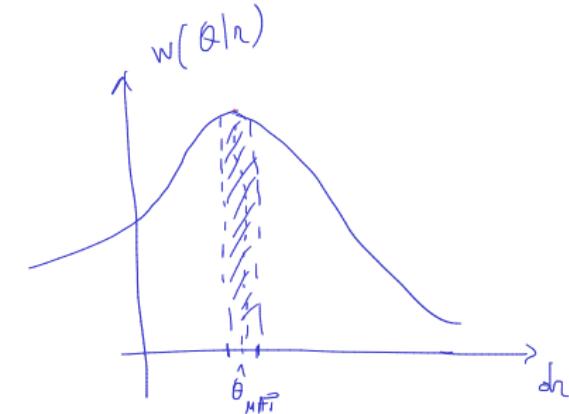
$\text{Cond} \quad \theta \notin [\hat{\theta} - E, \hat{\theta} + E] \Rightarrow C(\epsilon) = 1$



Estimatorul MAP

- ▶ Pentru minimizarea R , trebuie să maximizăm $\int_{\hat{\Theta}-E}^{\hat{\Theta}+E} w(\Theta|r)d\Theta$, integrala din jurul punctului $\hat{\Theta}$
- ▶ Pentru E foarte mic, funcția $w(\Theta|r)$ este aproximativ constantă, deci se va alege punctul unde funcția este maximă
- ▶ **Estimatorul Maximum A Posteriori (MAP)** = valoarea $\hat{\Theta}$ care maximizează $w(\Theta|r)$

$$\hat{\Theta}_{MAP} = \arg \max_{\Theta} w(\Theta|r) = \arg \max_{\Theta} w(r|\Theta) \cdot w(\Theta)$$



Interpretare

- ▶ Estimatorul MAP: $\hat{\Theta} = \text{valoarea care maximizează distribuția a posteriori}$
- ▶ Estimatorul EPMM: $\hat{\Theta} = \text{valoarea medie a distribuției a posteriori}$

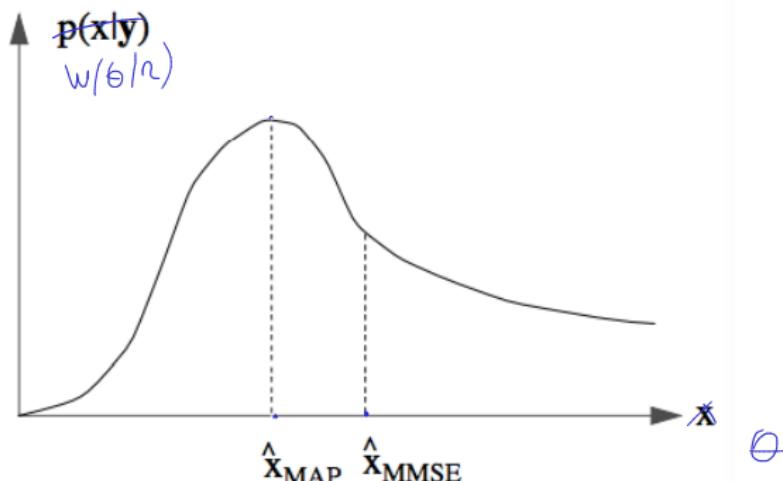


Figure 3: Estimatorul MAP vs EPMM(MMSE)

- ▶ Estimatorul MAP = minimizează costul mediu, folosind funcția de cost uniformă
 - ▶ ca le detectie: criteriul MPE = criteriul MR când costurile sunt la fel
- ▶ Estimatorul MMSE = minimizează costul mediu, folosind funcția de cost pătratică
 - ▶ similar cu criteriul MR, dar la estimare



Exercițiu: valoare constantă, 1 măsurătoare, zgomot Gaussian același σ

- ▶ Vrem să estimam temperatura de astăzi din Sahara
- ▶ Termometrul indică 40 grade, dar valoarea este afectată de zgomot Gaussian $\mathcal{N}(0, \sigma^2 = 2)$ (termometru ieftin)
- ▶ Se știe că de obicei în această perioadă a anului temperatura este în jur de 35 grade, cu o distribuție Gaussiană $\mathcal{N}(35, \sigma^2 = 2)$.
- ▶ Estimați valoarea reală a temperaturii folosind estimarea ML, MAP și EPMM(MMSE)

Exercițiu: valoare constantă, 1 măsurătoare, zgomot Gaussian același σ

- ▶ Dacă avem trei termometre, care indică 40, 38, 41 grade?

Exercițiu: valoare constantă, 1 măsurătoare, zgomot Gaussian σ diferit

- ▶ Dacă temperatura în această perioadă a anului are distribuție Gaussiană $\mathcal{N}(35, \sigma_2^2 = 3)$
 - ▶ cu varianță diferită, $\sigma_2 \neq \sigma$

Semnal oarecare în zgomot Gaussian (AWGN)

- ▶ Fie semnalul original “curat” $s_\Theta(t)$
- ▶ Zgomotul este Gaussian (AWGN) $\mathcal{N}(\mu = 0, \sigma^2)$
- ▶ Ca în cazul estimării de plauzibilitate maximă, funcția de plauzibilitate este:

$$w(\mathbf{r}|\Theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\sum(r_i - s_\Theta(t_i))^2}{2\sigma^2}}$$

- ▶ Dar acum aceasta **se înmulțește cu** $w(\Theta)$

$$w(\mathbf{r}|\Theta) \cdot w(\Theta)$$

Semnal oarecare în zgomot Gaussian (AWGN)

- ▶ Estimatorul MAP estimator este cel care maximizează produsul

$$\hat{\Theta}_{MAP} = \arg \max w(\mathbf{r}|\Theta)w(\Theta)$$

- ▶ Logaritmând:

$$\begin{aligned}\hat{\Theta}_{MAP} &= \arg \max \ln(w(\mathbf{r}|\Theta)) + \ln(w(\Theta)) \\ &= \arg \max -\frac{\sum(r_i - s_{\Theta}(t_i))^2}{2\sigma^2} + \ln(w(\Theta))\end{aligned}$$

Distribuție “a priori” Gaussiană

- Dacă distribuția “a priori” este de asemenea Gaussiană $\mathcal{N}(\mu_\Theta, \sigma_\Theta^2)$

$$\ln(w(\Theta)) = -\frac{\sum(\Theta - \mu_\Theta)^2}{2\sigma_\Theta^2}$$

- Estimatorul MAP devine

$$\hat{\Theta}_{MAP} = \arg \min \frac{\sum(r_i - s_\Theta(t_i))^2}{2\sigma^2} + \frac{\sum(\Theta - \mu_\Theta)^2}{2\sigma_\Theta^2}$$

- Poate fi rescris

$$\hat{\Theta}_{MAP} = \arg \min d(\mathbf{r}, s_\Theta)^2 + \underbrace{\frac{\sigma^2}{\sigma_\Theta^2}}_{\lambda} \cdot d(\Theta, \mu_\Theta)^2$$

- ▶ Estimatorul MAP în zgomot Gaussian și cu distribuție “a priori” Gaussiană

$$\hat{\Theta}_{MAP} = \arg \min d(\mathbf{r}, s_{\Theta})^2 + \underbrace{\frac{\sigma^2}{\sigma_{\Theta}^2} \cdot d(\Theta, \mu_{\Theta})^2}_{\lambda}$$

- ▶ $\hat{\Theta}_{MAP}$ este apropiat de valoarea medie μ_{Θ} și de asemenea face ca semnalul adevărat să fie apropiat de eșantioanele recepționate \mathbf{r}

- ▶ Exemplu: “caut locuință aproape de serviciu dar și aproape de Mall”
 - ▶ λ controlează importanța relativă a celor doi termeni

- ▶ Cazuri particulare

- ▶ σ_{Θ} foarte mic = distribuția “a priori” este foarte specifică (îngustă) = λ mare = termenul al doilea este dominant = $\hat{\Theta}_{MAP}$ foarte apropiat de μ_{Θ}
 - ▶ σ_{Θ} foarte mare = distribuția “a priori” este foarte nespecifică = λ mic = primul termen este dominant = $\hat{\Theta}_{MAP}$ apropiat de estimatorul de plauzibilitate maximă

- ▶ În general, aplicațiile practice:
 - ▶ utilizează diverse tipuri de distribuții “a priori”
 - ▶ estimează **mai mulți parametri** (un vector de parametri)
- ▶ Aplicații
 - ▶ reducerea zgromotului din semnale
 - ▶ restaurarea semnalelor (parti lipsă din imagini, imagini *blurate* etc)
 - ▶ compresia semnalelor