Decizie și Estimare în Prelucrarea Informației

Capitolul III. Elemente de Teoria Estimării

III.1 Introducere

Ce înseamnă "estimare"?

necunoscut

$s_\Theta(t) = \Theta$

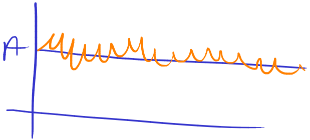
$s_\Theta(t) = \cos(2\pi \Theta n)$

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

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

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

Ce înseamnă “estimare”?



- Exemple:
  - Amplitudinea unui semnal constant: $r(t) = \underbrace{A}_{s_\theta(t)} + zgomot$, trebuie estimat A
  - Faza unui semnal sinusoidal: $r(t) = \underbrace{\cos(2\pi f t + \phi)}_{s_\theta(t)} + zgomot$, de estimat ϕ
  - Exemple mai complicate:
    - De estimat/decis ce cuvânt este pronunțat într-un semnal vocal

Estimare și Detecție/Decizie

Decizie $s(t) = \begin{cases} s_0(t) = 0 \\ s_1(t) = 5 \end{cases}$ A

$r(t) = s(t) + zgomot$ → $D_0: 0$, $D_1: 5$

Estimare $s_\theta(t) = A$ $\quad r(t) = s_\theta(t) + zgomot$ $\quad A = ?$

- Fie următoarea problemă de estimare:

  $s_\theta(t)$

  Se recepționează un semnal $r(t) = A + zgomot$, estimați-l pe A
- La detecție: 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 => se alege între **o infinitate de opțiuni** ale A
  - A poate fi orice valoare din $\mathbb{R}$, în general

Estimare și Detecție/Decizie

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

Semnalul recepționat

- 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

$$\mathbf{r} = [r_1, r_2, ... r_N]$$

- Eșantioanele depind de valoarea lui Θ

Semnalul recepționat

- Fiecare eșantion r_i este o variabilă aleatoare ce depinde de Θ (și de zgomot)
  - Fiecare eșantion are o distribuție care depinde de Θ

$$\longrightarrow \quad w_i(r_i|\Theta)$$

$w\left(r_i \,\middle|\, \begin{smallmatrix} H_0 \\ H_1 \end{smallmatrix}\right)$

- Întregul vector de eșantioane $\mathbf{r}$ este o variabilă aleatoare N-dimensională ce depinde de Θ (și de zgomot)
  - Are o distribuție N-dimensională ce depinde de Θ
  - Egală cu produsul tuturor $w_i(r_i|\Theta)$

$$w_N(\mathbf{r}|\Theta) = w_1(r_1|\Theta) \cdot w_2(r_2|\Theta) \cdot \ldots \cdot w_N(r_N|\Theta)$$

Tipuri de estimare

estimare ML

- 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

Bayes

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** unui valori Θ, dat fiind vectorul de observații $\mathbf{r}$:

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

Likelihood = L

- $L(\Theta|\mathbf{r})$ 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

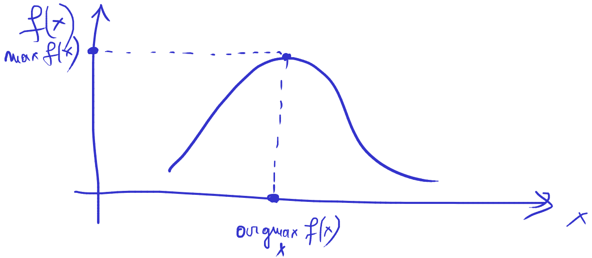
Estimarea de plauzibilitate maximă (Maximum Likelihood, ML):

- Estimatul $\hat{\Theta}_{ML}$ este **valoarea care maximizează plauzibilitatea, dat fiind valorile observate r**
  - i.e. valoarea care maximizează $L(\Theta|\mathbf{r})$, adică maximizează $w(\mathbf{r}|\Theta)$

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

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

Notații matematice



- Notații matematice generale
  - $\arg\max_x f(x)$ = "valoarea x are maximizează funcția f(x)"
  - $\max_x f(x)$ = "valoarea maximă a funcției 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":

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

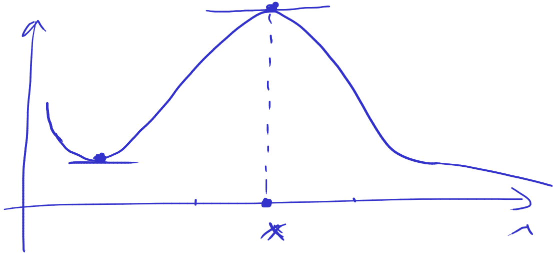
- Estimare ML:
    - "se alege valoarea care maximizează plauzibilitatea" = "alege Θ cu plauzibilitatea maximă"

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

Găsirea maximului

- Cum se rezolvă problema de maximizare?
  - adică cum se găsește estimatul $\hat{\Theta}_{ML}$ care maximizează $L(\Theta|\mathrm{v}\cancel{\mathrm{ec}}\mathrm{r})$
- 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”)

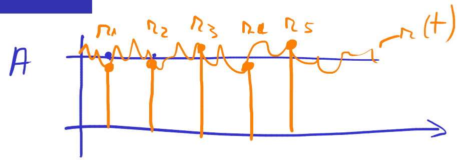
$$\frac{d\ln\left(L(\Theta|\mathbf{r})\right)}{d\Theta} = 0$$

Procedura de găsire a estimatului ML

Procedura de găsire a estimatului ML:

1. Se găsește expresia funcției
$$L(\Theta|\mathbf{r}) = w(\mathbf{r}|\Theta) = w(r_1|\theta) \cdot w(r_2|\theta) \cdot \ldots \cdot w(r_N|\theta)$$
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



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

  r_1, r_2, r_3, r_4, r_5

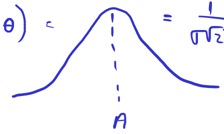
  $\hat{A}_{ML} = ?$

- Soluție: la tablă

- Estimatul $\hat{A}_{ML}$ este chiar valoarea medie a eșantioanelor
  - (deloc surprinzător)

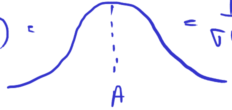
1) $w(r|\Theta) = w(r_1|\Theta) \cdot \ldots \cdot w(r_5|\Theta)$

$w(r_1|\theta) =$



$= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(r_1-A)^2}{2\sigma^2}}$

$w(r_2|\theta) =$



$= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(r_2-A)^2}{2\sigma^2}}$

$w(r_3|\theta)$

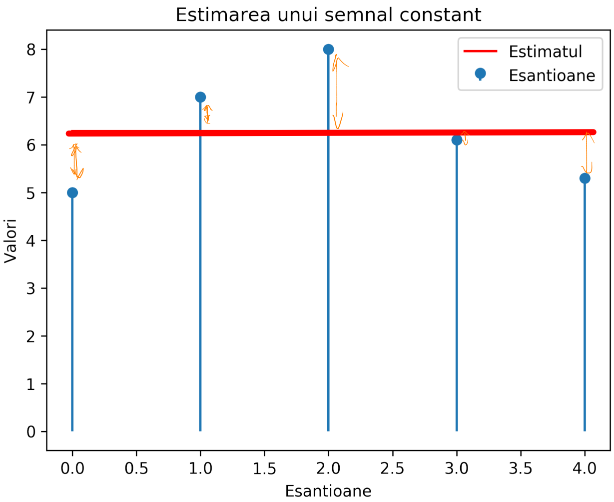
$w(r_4|\theta)$

$w(r_5|\theta)$

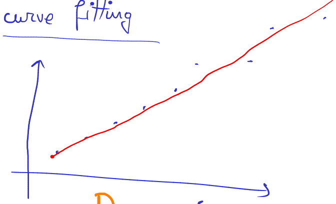
$$w(r|\Theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^5 \cdot e^{-\frac{(5-A)^2+(7-A)^2+(8-A)^2+(6.1-A)^2+(5.3-A)^2}{2\sigma^2}}$$

Vreau maxim!

Simulare numerică



curve fitting



Vreau maxim!:

$$\ln w(r|\theta) = \ln\left(\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^5\right) - \frac{\overbrace{\sum (r_i - A)^2}^{D}}{2\sigma^2}$$

Vreau minim!

2) $\frac{dD}{dA} = 0 \iff$

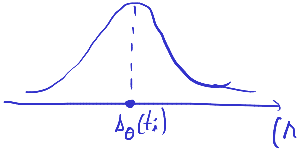
$$\cancel{2}(5-A)(\cancel{-1}) + \cancel{2}(7-A)(\cancel{-1}) + \cancel{2}(8-A)(\cancel{-1}) + \cancel{2}(6.1-A)(\cancel{-1}) + \cancel{2}(5.3-A)(\cancel{-1}) = 0$$

$$31.4 - 5A = 0 \Rightarrow \hat{A}_{ML} = \frac{31.4}{5} = 6.28$$

Aproximare a unei curbe

- Estimare = aproximare a unei curbe
  - se găsește cea mai bună potrivire a lui $s_\Theta(t)$ prin datele $\mathbf{r}$
- Din exemplul grafic anterior:
  - avem un set de date $\mathbf{r}$
  - se cunoaște forma semnalului = o dreaptă orizontală (A constant)
  - se aproximează în mod optim dreapta prin setul de date

Semnal oarecare în AWGN



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

- 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

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

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \cdot e^{-\frac{\sum_i (r_i - s_\Theta(t_i))^2}{2\sigma^2}}$$

Vrem Θ a.î. să fie maxim!

Semnal oarecare în AWGN

- Logaritmul plauzibilității ("log-likelihood") este

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

Vrem Θ a.î. să fie minim!

Vrem Θ a.î. să fie maxim!

Semnal oarecare în AWGN

- Maximul funcției = minimul exponentului

$$\hat{\Theta}_{ML} = \arg\max_{\Theta} L(\Theta|\mathbf{r}) = \arg\min_{\Theta} \sum_{i} (r_i - s_{\Theta}(t_i))^2$$

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

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

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

$$r = [r_1 \; r_2 \; \dots \; r_N]$$

$$s_\theta = [s_\theta(t_1) \; s_\theta(t_2) \; \dots \; s_\theta(t_N)]$$

$$d(r, s_\theta)^2 = (r_1 - s_\theta(t_1))^2 + (r_2 - s_\theta(t_2))^2 + \dots$$

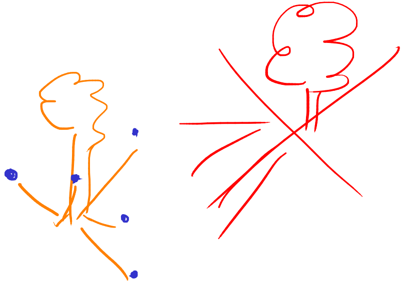
Semnal oarecare în AWGN

- Estimarea ML se poate rescrie 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

Semnal oarecare în AWGN

Procedura pentru estimarea tip ML în zgomot AWGN:

$r = [\dots]$

$s_\theta = [\dots]$

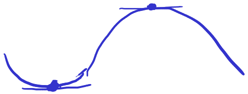
1. Se scrie expresia pentru pătratul distanței:

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

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 $\boxed{\hat{\Theta}_{ML}}$
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



Simulare numerică

Estimarea frecvenței f a unui semnal sinusoidal

- Găsiți estimatul Maximum Likelihood pentru frecvența f a unui semnal $s_\Theta(t) = cos(2\pi f t_i)$, din 10 măsurători afectate de zgomot $r_i = cos(2\pi f t_i) + zgomot$ 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

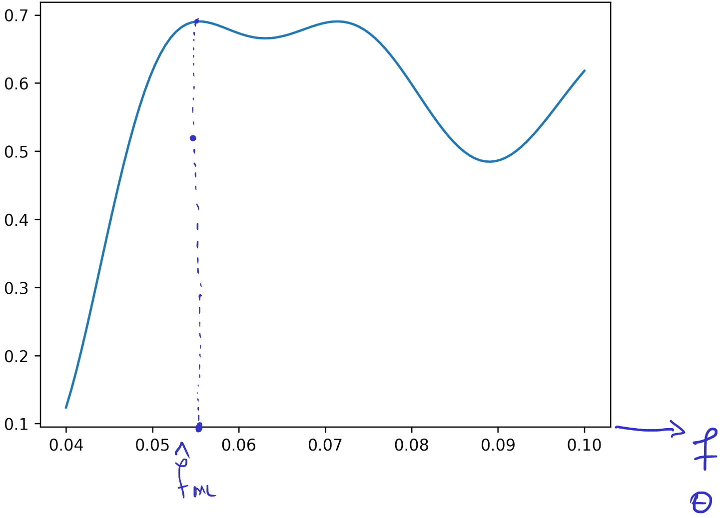
$$R = [\, \cdot \; \cdots \; \cdot \; \cdots \,]$$

$$s_\Theta = [\cos 2\pi f \cdot 0 \quad \cos(2\pi f 1) \; \ldots \; \cos(2\pi f 9)]$$

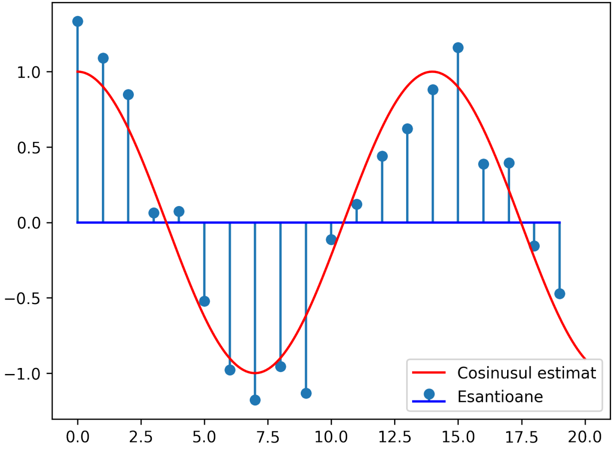
$$D = (d(r - s_\Theta))^2$$

Simulare numerică

Funcția de plauzibilitate este $L(\theta|r) = w(r|\theta)$



Simulare numerică



Parametri multipli

- Dacă semnalul depinde de mai mulți parametri?
  - de ex. amplitudinea, frecvența și faza inițială a unui cosinus:

$$s_{\theta}(t) = A\cos(2\pi f t + \phi)$$

- Se va considera Θ ca fiind un vector:

$$\boldsymbol{\Theta} = [\Theta_1, \Theta_2, ...\Theta_M]$$

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

Parametri multipli

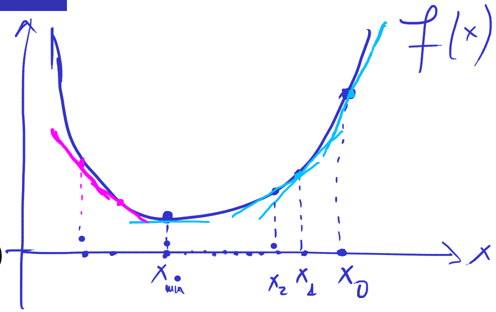
- Se rezolvă cu aceeași procedură, dar în loc de o singură derivată vom avea M derivate
- Se rezolvă sistemul:

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

$$\begin{cases} \frac{\partial L}{\partial A} = 0 \\ \frac{\partial L}{\partial f} = 0 \\ \frac{\partial L}{\partial \phi} = 0 \end{cases}$$

  - uneori este dificil/imposibil

Coborâre după gradient (Gradient Descent)



- Cum se estimează parametrii $\mathbf{\Theta}$ în cazuri complicate?
  - în aplicații reale, unde pot fi foarte mulți parametri ($\mathbf{\Theta}$ 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)

$$\boxed{x_{k+1} = x_k - \mu \cdot f'(x)\Big|_{x=x_k}}$$

Coborâre după gradient (Gradient Descent)

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

$\Theta = [\Theta_1, \Theta_2, \ldots \Theta_N]$

2. Repetă la fiecare iterație k:

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

   $d(r, s_\Theta)^2$

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

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

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

   - sau, sub formă vectorială:

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

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

Rețele Neurale

- 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

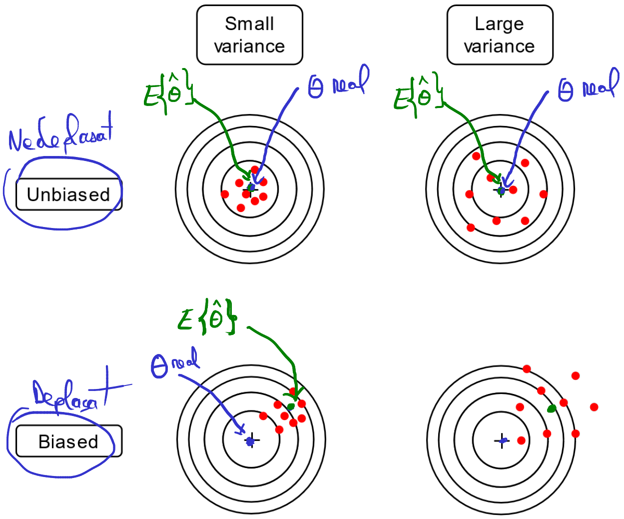
Deplasarea și varianta estimatorilor

$\Theta \simeq \hat{\Theta}_{ML}$

- Cum caracterizăm calitatea unui estimator?
- Un estimator $\hat{\Theta}$ este **o variabilă aleatoare**
  - poate avea diverse valori, pentru că se calculează pe baza eșantioanelor recepționate, care depind de zgomot
  - exemplu: se repetă aceeași estimare pe calculatoare diferite => valori estimate ușor diferite
- Fiind o variabilă aleatoare, se pot defini:
  - valoarea medie a estimatorului: $E\left\{\hat{\Theta}\right\}$
  - varianța estimatorului: $E\left\{(\hat{\Theta} - \Theta)^2\right\}$

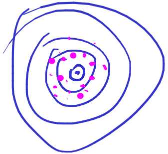
Deplasarea și varianța estimatorilor

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



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

Figure 1: Deplasarea și varianța estimatorilor



Deplasarea unui estimator

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

$$Deplasare = E\left\{\hat{\Theta}\right\} - \Theta$$

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

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

- Estimator **deplasat** = valoarea medie a estimatorului diferă de valoarea adevărată a parametrului Θ
  - diferența $E\left\{\hat{\Theta}\right\} - \Theta$ este **deplasarea** estimatorului

Deplasarea unui estimator

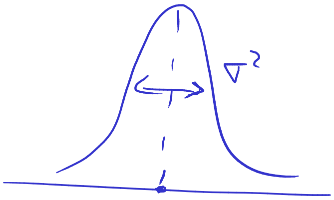
- Exemplu: semnal constant A, zgomot Gaussian (cu media), estimatorul de plauzibilitate maximă este $\hat{A}_{ML} = \frac{1}{N}\sum_i r_i$
- Atunci:

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

- Acest estimator este nedeplasat

Varianța unui estimator

$$\sigma^2 = E\{(\hat{\theta} - E\{\hat{\theta}\})^2\}$$



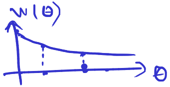
- **Varianța** unui estimator măsoară "abaterile" estimatorului în jurul valorii medii
  - aceasta e definiția varianței σ^2 în general
- Dacă un estimator are **varianța** mare, valoarea estimată poate fi departe de cea reală, chiar daca estimatorul este nedeplasat
- De obicei se preferă estimatori cu **varianță mică**, tolerându-se o eventuală mică deplasare

II.3 Estimare Bayesiană

Estimare Bayesiană

Bayes

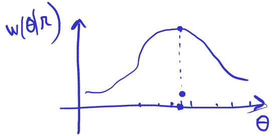
- **Estimarea Bayesiană** ia în calcul termeni suplimentari pe lângă $w(\mathbf{r}|\Theta)$:
  - o distribuție *a priori* $w(\Theta)$
  - opțional, o funcție de cost



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

Estimare Bayesiană

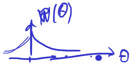
- Se definește **distribuția *a posteriori*** a lui Θ, dat fiind observațiile $\mathbf{r}$, folosind **regula lui Bayes**:



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

constant

- Termenii:
  - Θ este parametrul necunoscut
  - $\mathbf{r}$ este vectorul de observații cunoscut
  - $w(\Theta|\mathbf{r})$ este probabilitatea ca parametrul să aibă valoarea Θ, dat fiind vectorul de observații $\mathbf{r}$;
  - $w(\mathbf{r}|\Theta)$ este funcția de plauzibilitate
  - $w(\Theta)$ este distribuția **a priori** a lui Θ
  - $w(\mathbf{r})$ este distribuția **a priori** a lui $\mathbf{r}$; se presupune a fi constantă

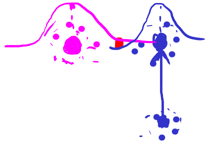


Regula lui Bayes:

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

$$\Downarrow$$

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



Regula lui Bayes

- Relația precedentă arată că, în general, estimarea lui Θ depinde de două lucruri:
  1. De vectorul observațiilor $\mathbf{r}$, prin termenul $w(\mathbf{r}|\Theta)$
  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
  - Noțiunile bazate pe regula lui Bayes poartă deseori numele de "Bayesiane"

Distribuția *a priori*

„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*
  - estimatul va fi "tras" înspre valori mai probabile

Estimatorul MAP

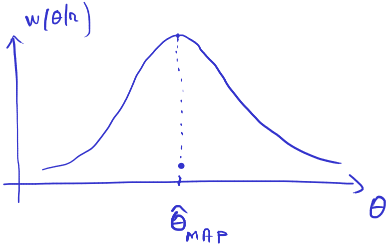
$w(r|\theta) \cdot w(\theta) \,/\, w(r)$

$=$

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

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

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



Estimatorul MAP

Exemplu: Imagine

Relația dintre estimarea MAP și ML

- Estimatorul ML:
$$\hat{\Theta}_{ML} = \arg\max_{\Theta} w(\mathbf{r}|\Theta)$$
- Estimatorul MAP:
$$\hat{\Theta}_{MAP} = \arg\max_{\Theta} w(\mathbf{r}|\Theta) \cdot w(\Theta)$$
- Estimatorul ML este un caz particular de MAP pentru $w(\Theta)$ constant
  - $w(\Theta) =$ constant înseamnă că toate valorile lui Θ sunt *a priori* echiprobabile
  - i.e. nu avem extra informații despre valoarea lui Θ

Relația cu detecția semnalelor

- Criteriul probabilității minime de eroare: $\boxed{\frac{w(r|H_1)}{w(r|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P(H_0)}{P(H_1)}}$ $\cdot \frac{P(H_1)}{P(H_0)}$
- Se poate rescrie ca $w(r|H_1) \cdot P(H_1) \underset{H_0}{\overset{H_1}{\gtrless}} w(r|H_0)P(H_0)$ $w(r|H_i) \cdot P(H_i)$, max pt. i = 0 sau 1
  - 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!

Estim MAP $\longleftrightarrow$ Crit. M.P.E.

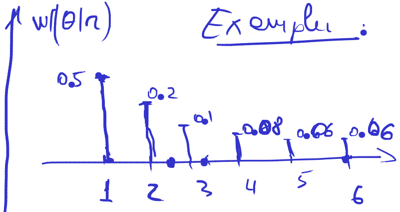
Funcția de cost

- Vrem s 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

Exemplu:



1).

2). $(\hat{\theta} - \theta)^2$

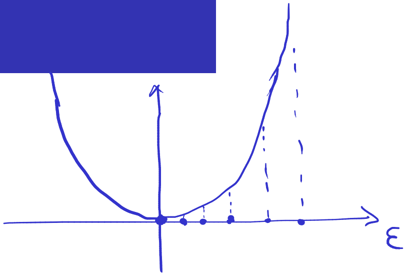
est. = 1 $\left[\begin{array}{ll} 50\% & \to 10 \\ 6\% & \to 25 \end{array}\right.$

est = 2 $\left[\begin{array}{ll} 6\% & \to 16 \\ 50\% & \to 1 \end{array}\right.$

Funcția de cost

- Funcții de cost uzuale:
  - Pătratică: "Eroare pătratică"

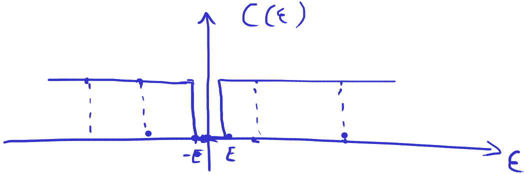
$$C(\epsilon) = \epsilon^2 = \left(\hat{\Theta} - \Theta\right)^2$$



  - Uniformă:

total sau nimic

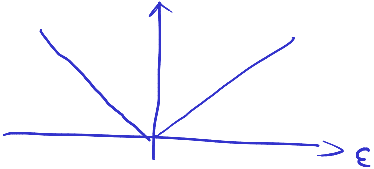
$$C(\epsilon) = \begin{cases} 0, & \text{if } |\epsilon| = |\hat{\Theta} - \Theta| \leq E \\ 1, & \text{if } |\epsilon| = |\hat{\Theta} - \Theta| > E \end{cases}$$



  - 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 detecție
  - la detecție 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 $w(\Theta|\mathbf{r})$

IUmportanța funcției de cost

- Fie distribuția *a posteriori* următoare:

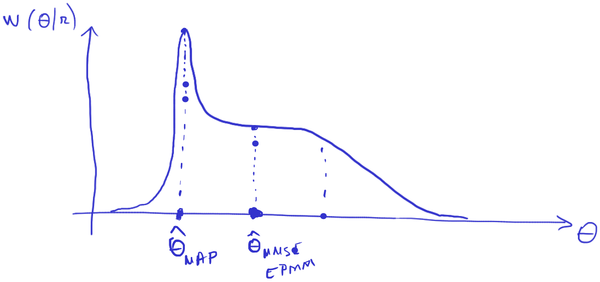


Figure 2: Unbalanced posterior distribution

- Care este estimatorul 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 ? :)

Riscul Bayesian

- Distribuția *a posteriori* $w(\Theta|\mathbf{r})$ 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:

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

The Bayesian risk

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

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

- 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)

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

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

- 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

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

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

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

Interpretare

$$\overline{x} = \int_{-\infty}^{\infty} x \cdot w(x)\, dx$$

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

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

  - EPMM = "Eroare Pătratică Medie Minimă"
  - valoarea medie = sumă (integrală) din fiecare Θ ori probabilitatea sa $w(\Theta|\mathbf{r})$

MMSE = "Minimum Mean Squared Error"

- 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

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

$E \longrightarrow 0$

- Știm că $\Theta = \hat{\Theta} - \epsilon$
- Se obține

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

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

Estimatorul MAP

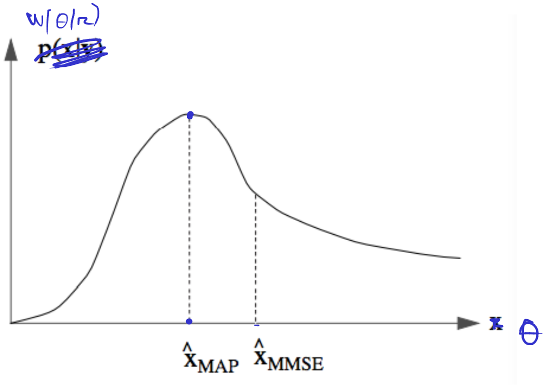
- Pentru minimizarea R, trebuie să maximizăm $\int_{\hat{\Theta}-E}^{\hat{\Theta}+E} w(\Theta|\mathbf{r}) d\Theta$, integrala din jurul punctului $\hat{\Theta}$
- Pentru E foarte mic, funcția $w(\Theta|\mathbf{r})$ este aproximativ constantă, deci se va alege punctul unde funcția este maximă
- **Estimatorul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

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

Interpretare

- Estimatorul MAP: $\hat{\Theta}$ = valoarea care maximizează distribuția *a posteriori*

- Estimatorul EPMM: $\hat{\Theta}$ = valoarea medie a distribuției *a posteriori*



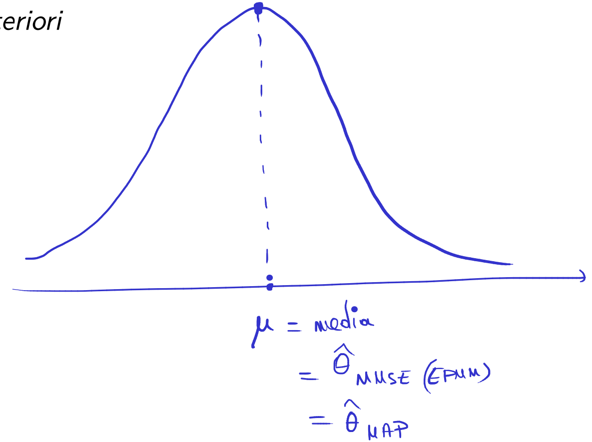


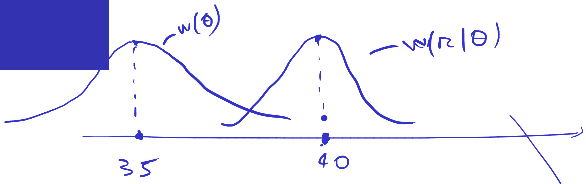
Figure 3: Estimatorul MAP vs EPMM(MMSE)

[image from https://allenlu2007.wordpress.com]

Relația între estim. MAP and EPMM

- Estimatorul MAP = minimizează costul mediu, folosind funcția de cost uniformă
  - ca le detecție: 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



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

$r = 40 = T + \text{zgomot}$

- 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) → $w(r|\theta)$
- Se știe că de obicei în această perioadă a anului temperatura este în jur de 35 grade, cu o distribuție Gaussiană $\mathcal{N}(35, \sigma^2 = 2)$. → $w(T)$
- Estimați valoarea reală a temperaturii folosind estimarea ML, MAP și EPMM(MMSE)