Decizie și Estimare în Prelucrarea Informației

Capitolul II. Elemente de teorie statistică a detecției

II.1 Introducere

Introducere

- Detecția semnalelor = a decide care semnal este prezent dintre două sau mai multe posibilități
  - inclusiv că nu există nici un semnal
- Avem la dispoziție observații **cu zgomot**
  - semnalele sunt afectate de zgomot

Schema bloc a detecției semnalelor

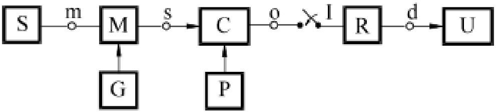

Figure 1: Signal detection model

- Conținut:
  - Sursa de informație: generează mesajele a_n cu probabilitățile $p(a_n)$
  - Modulator: transmite semnalul $s_n(t)$ la mesajul a_n
  - Canal: adaugă zgomot aleator
  - Eșantionare: prelevă eșantioane din semnalul $s_n(t)$
  - Receptor: **decide** ce mesaj a_n s-a fost recepționat

Scenarii practice

- Transmisie de date
  - nivele constante de tensiune (e.g. $s_n(t) = \text{constant}$)
  - modulație PSK (Phase Shift Keying): $s_n(t) =$ cosinus cu aceeași frecvență dar faze inițiale diferite
  - modulație FSK (Frequency Shift Keying): $s_n(t) =$ cosinus cu frecvențe diferite
  - modulație OFDM (Orthogonal Frequency Division Multiplexing): caz particular de FSK
- Radar
  - se emite un semnal; în cazul unui obstacol, semnalul se reflectă înapoi
  - receptorul așteaptă posibilele reflecții ale semnalului emis și decide
    - nu este prezentă o reflecție -> nici un obiect
    - semnalul reflectat este prezent -> obiect detectat

Generalizări

- Decizie între mai mult de două semnale
- Numărul de eșantioane (observații):
  - un singur eșantion
  - mai multe eșantioane
  - observarea întregului semnal continuu, pentru un timp T

II.2 Detecția semnalelor constante

Detecție unui semnal constant, 1 eșantion

- Cel mai simplu caz: detecția unui semnal constant afectat de zgomot, folosind un singur eșantion
  - două mesaje a_0 și a_1
  - mesajele sunt modulate cu semnale constante
    - pentru a_0: se emite $s_0(t) = 0$
    - pentru a_1: se emite $s_1(t) = A$
  - peste semnal se suprapune zgomot aditiv
  - eșantionarea preia un singur eșantion
  - decizie: se compară eșantionul cu un prag

Decizia pe bază de prag

- Valoarea eșantionului este $r = s + n$
  - s este semnalul adevărat ($s_0 = 0$ or $s_1 = A$)
  - n este un eșantion de zgomot
- n este o variabilă aleatoare continuă
- r este de asemenea o variabilă aleatoare
  - cum depinde distribuția lui r de cea a lui n
- Decizia se ia prin compararea lui r cu un prag T:
  - dacă $r < T$, se ia decizia D_0: semnalul adevărat este s_0
  - dacă $r \geq T$, se ia decizia D_1: semnalul adevărat este s_1

Ipoteze

- Receptorul decide între **două ipoteze**:
  - H_0: semnalul adevărat este s_0 (s-a transmis a_0)
  - H_1: semnalul adevărat este s_1 (s-a transmis a_1)
- Rezultate posibile
  1. Semnalul nu este prezent (s_0), si nu este detectat
     - Decizia D_0 în ipoteza H_0
     - Probabilitatea sa este $P_n = P(D_0 \cap H_0)$
  2. **Alarmă falsă**: semnalul nu este prezent (s_0), dar este detectat (eroare!)
     - Decizia D_1 în ipoteza H_0
     - Probabilitatea este $P_{fa} P(D_1 \cap H_0)$
  3. **Ratare**: semnalul este prezent (s_1), dar nu este detectat (eroare!)
     - Decizia D_0 în ipoteza H_1
     - Probabilitatea este $P_m = P(D_0 \cap H_1)$
  4. Semnal detectat corect: semnalul este prezent, și este detectat
     - Decizia D_1 în ipoteza H_1
     - Probabilitatea este $P_d = P(D_1 \cap H_1)$

Criteriul plauzibilității maxime (Maximum Likelihood)

- Se alege ipoteza care pare **cea mai plauzibilă** dat fiind eșantionul observat r
- **Plauzibilitatea** ("*likelihood*") unei observații r = densitatea de probabilitate a lui r dată fiind ipoteza H_0 sau H_1
- Plauzibilitatea în cazul ipotezei H_0: $w(r|H_0)$
  - r este doar zgomot, deci provine din distribuția zgomotului de pe canal
- Plauzibilitatea în cazul ipotezei H_1: $w(r|H_1)$
  - r este A + zgomot, deci valoarea sa provine din distribuția (A + zgomot)
- Raportul de plauzibilitate

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

Interpretare grafică

- Fie cazul în care zgomotul are distribuție normală
- Desen: cele două densități de probabilitate pentru H_0 și H_1

Decizia pe bază de prag

- Decizie ML pe baza raportului de plauzibilitate = compararea lui r cu un prag T
- Pragul = punctul de intersecție a celor două distribuții

Zgomot cu distribuție normală

- Caz particular: zgomotul are distribuția normală $\mathcal{N}(0, \sigma^2)$
- Raportul de plauzibilitate este $\frac{w(r|H_1)}{r|H_0} = \frac{e^{\frac{(r-A)^2}{2\sigma^2}}}{e^{\frac{r^2}{2\sigma^2}}} \underset{H_0}{\overset{H_1}{\gtrless}} 1$
- Pentru distribuția normală, e preferabil să aplicăm *logaritmul natural*
  - logaritmul este o funcție monoton crescătoare, deci nu schimbă rezultatul comparației
  - dacă $A < B$, atunci $\log(A) < \log(B)$
- **log-likelihood** al unui observații = logaritmul plauzibilității (likelihood)
  - de obicei este vorba de logaritmul natural, dar poate fi orice bază

Testul “log-likelihood” în cazul ML

- Pentru zgomot cu distribuție normală, decizia ML înseamnă compararea *log-likelihood*

$$\frac{(r-A)^2}{r^2} \underset{H_0}{\overset{H_1}{\gtrless}} 1$$

- Se extrage radicalul

$$\frac{|r-A|}{|r|} \underset{H_0}{\overset{H_1}{\gtrless}} 1$$

- $|r-A| =$ distanța de la r la A, $|r| =$ distanța de la r la 0
- Decizie ML în zgomot normal: se alege valoarea 0 sau A **cea mai apropiată** de r
  - principiu foarte general, întâlnit în multe alte scenarii
  - principiul **cel mai apropiat vecin** (“*nearest neighbor*”)
  - receptorul ML se mai numește **receptor de distanță minimă** (“*minimum distance receiver*”)
  - echivalent cu setarea unui prag $T = \frac{A}{2}$

Generalizări

- Dacă zgomotul are altă distribuție?
  - Pragul T rămâne punctul de intersecție, oricare ar fi acela
  - Pot fi mai multe puncte de intersecție, deci mai multe praguri
  - axa $\mathbb{R}$ este împărțită în **regiuni de decizie** R_0 și R_1
- Dacă distribuția zgomotului este diferită în cazurile H_0 și H_1?
  - Pragul T (sau pragurile) rămân punctele de intersecție, oricare ar fi acelea
- Dacă semnalul $s_0(t)$ (pentru ipoteza H_0, simbolul a_0) nu este 0, ci o altă valoare constantă B?
  - Pragul T (sau pragurile) rămân punctele de intersecție, dar distribuțiile sunt centrate pe B și A
  - Pentru zgomot gaussian, se alege B sau A, cel mai apropiat de eșantion (pragul este la mijlocul distanței dintre B și A)

Generalizări

- Mai mult de două semnale?
  - De ex. 4 nivele de semnal posibile: -6, -2, 2, 6
  - Se alege cea mai plauzibilă ipoteză, pe baza celor 4 plauzibilități
  - Nu mai există un singur prag T, sunt în mod necesar mai multe

Exerciții

- Un semnal poate avea două valori posibile, 0 sau 5. Receptorul ia un singur eșantion cu valoarea $r = 2.25$
  1. Dacă zgomotul este gaussian, ce semnal este detectat pe baza criteriului plauzibilității maxime?
  2. Dar dacă semnalul 0 este afectat de zgomot gaussian $\mathcal{N}(0, 0.5)$, iar semnalul 5 de zgomot uniform $\mathcal{U}[-4, 4]$?
  3. Repetați a. și b. dacă valoarea 0 se înlocuiește cu -1
- Un semnal poate avea patru valori posibile: -6, -2, 2, 6. Fiecare valoare durează timp de o secundă. Semnalul este afectat de zgomot alb cu distribuție normală. Receptorul ia un singur eșantion pe secundă. Folosind criteriul plauzibilității maxime, decideți ce semnal s-a transmis, dacă receptorul primește eșantioanele următoare:

$$4, 6.6, -5.2, 1.1, 0.3, -1.5, 7, -7, 4.4$$

Dezavantaje ale criteriului plauzibilității maxime

- Raportul de plauzibilitate utilizează densitățile de probabilitate **condiționate**
  - condiționate de ipotezele H_0 sau H_1
- Condiționarea de ipotezele H_0 și H_1 ignoră probabilitatea celor două ipoteze H_0 și H_1
- Reamintire (TCI): regula lui Bayes

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

- Interpretare
  - Probabilitatea $P(A)$ este extrasă din $P(B|A)$
  - $P(B|A)$ nu mai conține nici o informație despre $P(A)$, șansele ca A chiar să aibă loc
  - Exemplu: P(gol | șut la poartă)
- Practic: dacă $p(H_0) >> p(H_1)$, am vrea să împingem pragul T înspre H_1

Criteriul probabilității minime de eroare

- Se iau în calcul probabilitățile $P(H_0)$ și $P(H_1)$
- Se urmărește **minimizarea probabilității totale de eroare** P_e
  - erori $=$ alarme false și ratări
- Există de asemenea un prag T astfel încât
  - decidem D_0 dacă $r < T$
  - decidem D_1 dacă $r \geq T$
- Trebuie să găsim valoarea lui T

Probabilitatea de eroare

- Probabilitatea unei alarme false

$$\begin{aligned} P(D_1 \cap H_0) =& P(D_1|H_0) \cdot P(H_0) \\ =& \int_T^{\infty} w(r|H_0)dx \cdot P(H_0) \\ =& (1 - \int_{-\infty}^{T} w(r|H_0)dx \cdot P(H_0) \end{aligned}$$

- Probabilitatea unei ratări

$$\begin{aligned} P(D_0 \cap H_1) =& P(D_0|H_1) \cdot P(H_1) \\ =& \int_{-\infty}^{T} w(r|H_1)dx \cdot P(H_1) \end{aligned}$$

- Suma lor este

$$P_e = P(H_0) + \int_{-\infty}^{T} [w(r|H_1) \cdot P(H_1) - w(r|H_0) \cdot P(H_0)]dx$$

Probabilitatea de eroare minimă

- Urmărim minimizarea P_e, adică să minimizăm integrala
- Pentru a minimiza integrala, se alege T astfel încât pentru toți $r < T$, termenul din integrala este **negative**
  - integrarea pe întregul interval în care o funcție este negativă conduce la valoarea minimă
- Așadar, când $w(r|H_1) \cdot P(H_1) - w(r|H_0) \cdot P(H_0) < 0$ avem $r < T$, adică decizia D_0
- Invers, dacă $w(r|H_1) \cdot P(H_1) - w(r|H_0) \cdot P(H_0) > 0$ avem $r > T$, adică decizia D_1
- Astfel

$$w(r|H_1) \cdot P(H_1) - w(r|H_0) \cdot P(H_0) \underset{H_0}{\overset{H_1}{\gtrless}} 0$$

$$\frac{w(r|H_1)}{w(r|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P(H_0)}{P(H_1)}$$

Interpretare

- Similar cu criteriul plauzibilității maxime, dar depinde de probabilitățile celor două ipoteze (cazuri, simboluri)
  - Când una dintre ipoteze este mai probabilă decât cealaltă, pragul este împins în favoarea sa, înspre cealaltă ipoteză
- De asemenea bazat pe raportul de plauzibilitate, ca și primul criteriu

Criteriul probabilității minime de eroare - zgomot gaussian

- Presupunând că zgomotul este gaussian (normal), $\mathcal{N}(0, \sigma^2)$

$$w(r|H_1) = e^{\frac{(r-A)^2}{2\sigma^2}}$$

$$w(r|H_0) = e^{\frac{r^2}{2\sigma^2}}$$

- Se aplică logaritmul natural

$$\frac{(r-A)^2}{2\sigma^2} - \frac{r^2}{2\sigma^2} \underset{H_0}{\overset{H_1}{\gtrless}} \ln\left(\frac{P(H_0)}{P(H_1)}\right)$$

- Echivalent

$$(r-A)^2 \underset{H_0}{\overset{H_1}{\gtrless}} (r-0)^2 + \underbrace{2\sigma^2 \cdot \ln\left(\frac{P(H_0)}{P(H_1)}\right)}_{C}$$

Exerciții

- O sursă de informație furnizează două mesaje cu probabilitățile $p(a_0) = \frac{2}{3}$ și $p(a_1) = \frac{1}{3}$. Mesajele se transmit prin semnale constante cu valorile -5 (a_0) și 5 (a_1). Semnalele sunt afectate de zgomot alb cu distribuție gaussiană $\mathcal{N}(0, \sigma^2 = 1)$ Receptorul ia un singur eșantion cu valoarea r. Decizia se face prin compararea valorii r cu un prag T, astfel: dacă $r < T$ se decide că s-a transmis mesajul a_0, altfel se decide mesajul a_1.
    1. Să se găsească valoarea pragului T conform criteriul probabilității minime de eroare
    2. Dar dacă semnalul 5 este afectat de zgomot uniform $\mathcal{U}[-4, 4]$?

Minimum risk (cost) criterion

- What if we care more about one type of errors (e.g. false alarms) than other kind (e.g. miss)?
- Minimum risk (cost) criterion: assign costs to decisions, minimize average cost
    - C_{ij} = cost of decision D_i when true hypothesis was H_j
    - C_{00} = cost for good detection D_0 in case of H_0
    - C_{10} = cost for false alarm (detection D_1 in case of H_0)
    - C_{01} = cost for miss (detection D_0 in case of H_1)
    - C_{11} = cost for good detection D_1 in case of H_1
- The risk = the average cost

$$R = C_{00}P(D_0 \cap H_0) + C_{10}P(D_1 \cap H_0) + C_{01}P(D_0 \cap H_1) + C_{11}P(D_1 \cap H_1)$$

- Minimum risk criterion: **minimize the risk R**

Computations

- Proof on table:
  - Use Bayes rule
  - Notations: $w(r|H_j)$ (*likelihood*)
  - Probabilities: $\int_{R_i} w(r|H_j)dV$
- Conclusion, **decision rule is**

$$\frac{w(r|H_1)}{w(r|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{(C_{10} - C_{00})p(H_0)}{(C_{01} - C_{11})p(H_1)}$$

Interpretation

- Similar to ML and to minimum probability of error criteria
  - also uses a **likelihood ratio** test
- Both probabilities and the assigned costs can move threshold towards one side or the other
- If $C_{10} - C_{00} = C_{01} - C_{11}$, reduces to previous criterion (minimum probability of error)
  - e.g. if $C_{00} = C_{11} = 0$, and $C_{10} = C_{01}$

In gaussian noise

- If the noise is gaussian (normal), then similar to other criteria, apply logarithm
- Equivalently

$$(r - A)^2 \underset{H_0}{\overset{H_1}{\gtrless}} (r - 0)^2 + \underbrace{2\sigma^2 \cdot \ln\left(\frac{(C_{10} - C_{00})p(H_0)}{(C_{01} - C_{11})p(H_1)}\right)}_{C}$$

Example

- Example at blackboard: random noise with $N(0, \sigma^2)$, one sample

Generalization: two non-zero levels

- What if the s_0 signal is not 0, but another constant signal $s_0 = B$
- Noise distribution $w(r|H_0)$ is centered on B, not 0
- Otherwise, everything else stays the same
- Performance is defined by the gap between the two levels $(A - B)$
  - same performance if $s_0 = 0$, $s_1 = A$ or if $s_0 = -\frac{A}{2}$ and $s_1 = fracA2$

Differential vs single-ended signalling

- Single-ended signaling: one signal is 0, other is non-zero
  - $s_0 = 0$, $s_1 = A$
- Differential signaling: use two non-zero levels with different sign, same absolute value
  - $s_0 = 0$, $s_1 = A$
- Which is better?

Differential vs single-ended signalling

- If gap difference between levels is the same, performance is the same
- Average power of a signal = average squared value
- For differential signal: $P = \left(\pm\frac{A}{2}\right)^2 = \frac{A^2}{4}$
- For signal ended signal: $P = P(H_0) \cdot 0 + P(H_1)\,(A)^2 = \frac{A^2}{2}$
- assuming equal probabilities of 0 and 1, $P(H_0) = P(H_1) = \frac{1}{2}$
- Differential uses half the power of single-ended (i.e. better)

Summary of criteria

- We have seen decision based on 1 sample r, between 2 constant levels
- All decisions are based on a likelihood-ratio test

$$\frac{w(r|H_1)}{w(r|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} K$$

- Different criteria differ in the chosen value of K (likelihood threshold)
- Depending on the noise distributions, the real axis is partitioned into regions
  - region R_0: if r is in here, decide D_0
  - region R_1: if r is in here, decide D_1
  - e.g. $R_0 = (-infty, \frac{A+B}{/}2]$, $R_1 = (\frac{A+B}{/}2, \infty)$ (ML)

Receiver Operating Characteristic

- The receiver performance is usually represented with **"Receiver Operating Characteristic"** graph
- It is a graph of correct detection probability $P_d = P(D_1|H_1)$ as a function of false alarm probability $P_{fa} = P(D_1 \cap H_0)$
- Picture here

Receiver Operating Characteristic

- There is always a **tradeoff** between good P_d and bad P_{fa}
  - to increase P_d one must also increase P_{fa}
  - if we want to make sure we don't miss any real detections (increase P_d), we pay by increasing the chances of false alarms
- Different criteria = different likelihood thresholds K = different points on the graph = different tradeoffs
  - but the tradeoff cannot be avoided
- How to improve the receiver?
  - i.e. increase P_D while keeping P_{fa} the same

Performance of likelihood-ratio decoding in WGN

- WGN = "White Gaussian Noise"
- Assume equal probabilities $P(H_0) = P(H_1) = \frac{1}{2}$
- All decisions are based on a likelihood-ratio test

$$\frac{w(r|H_1)}{w(r|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} K$$

- Detection probability is

$$\begin{aligned}
P_D =& P(D_1|H_1)P(H_1) \\
=& P(H_1) \int_T^{\infty} w(r|H_1) \\
=& P(H_1)(F(\infty) - F(T)) \\
=& P(H_1) \left(1 - \frac{1}{2} \left(1 + erf \left(\frac{r - A}{\sqrt{2}\sigma}\right)\right)\right) \\
=& \frac{1}{4} \left(1 - erf \left(\frac{r - A}{\sqrt{2}\sigma}\right)\right)
\end{aligned}$$

Performance of likelihood-ratio decoding in WGN

- False alarm probability is

$$\begin{aligned}
P_{fa} =& P(D_1|H_0)P(H_0) \\
=& P(H_0)\int_T^{\infty} w(r|H_0) \\
=& P(H_0)(F(\infty) - F(T)) \\
=& P(H_0)\left(1 - \frac{1}{2}\left(1 + erf\left(\frac{r-0}{\sqrt{2}\sigma}\right)\right)\right) \\
=& \frac{1}{4}\left(1 - erf\left(\frac{r-0}{\sqrt{2}\sigma}\right)\right)
\end{aligned}$$

- Therefore