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

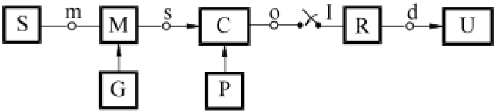


Figure 1: Signal detection model

- 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:

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

Probabilități de eroare condiționate

- Putem calcula probabilitățile de eroare condiționate
- Fie regiunile de decizie:
  - R_0: dacă $r \in R_0$, decizia este D_0, de ex. (∞, T) pentru zgomot gaussian
  - R_1: daca $r \in R_1$, decizia este D_1, de ex. $[T, \infty)$ pentru zgomot gaussian
- Probabilitatea unei alarme false ***dacă*** *semnalul original este* $s_0(t)$

$$P(D_1|H_0) = \int_{R_1} w(r|H_0)dx$$

- Probabilitatea unei ratări ***dacă*** *semnalul original este* $s_1(t)$

$$P(D_0|H_1) = \int_{R_0} w(r|H_1)dx$$

- Aceste valori nu țin cont de probabilitatea ca semnalul să fie $s_0(t)$ sau $s_1(t)$
  - sunt **condiționate** ("dacă")

Probabilități de eroare condiționate

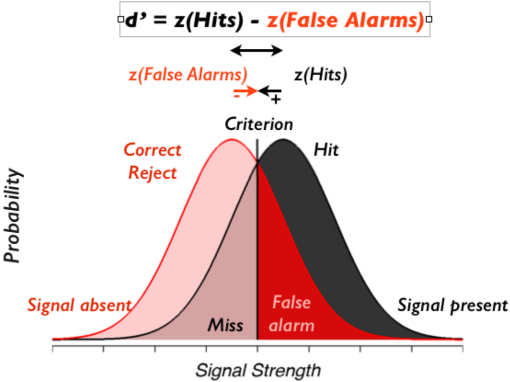


Figure 2: Probabilitățile deciziilor

[sursa: hhttp://gru.stanford.edu/doku.php/tutorials/sdt]

Reamintire (TCI): regula lui Bayes

- 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). Câte goluri se înscriu?

Exercițiu

- Un semnal poate avea două valori posibile, 0 sau 5. Semnalul 0 este afectat de zgomot gaussian $\mathcal{N}(0, 0.5)$, iar semnalul 5 de zgomot uniform $\mathcal{U}[-4, 4]$. Receptorul decide pe baza criteriului plauzibilității maxime, folosind un singur eșantion din semnal.
    1. Calculați probabilitatea unei decizii greșite când semnalul original este $s_0(t)$
    2. Calculați probabilitatea unei decizii greșite când semnalul original este $s_1(t)$

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
- Dacă $p(H_0) > p(H_1)$, am vrea să împingem pragul T înspre H_1, și vice-versa
  - pentru că este mai probabil ca semnalul să fie $s_0(t)$
  - și de aceea vrem să "favorizăm" decizia D_0

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
- Trebuie să găsim regiunile de decizie R_0 și R_1

Probabilitatea de eroare

- Probabilitatea unei alarme false

$$\begin{aligned} P(D_1 \cap H_0) =& P(D_1|H_0) \cdot P(H_0) \\ =& \int_{R_1} w(r|H_0)dx \cdot P(H_0) \\ =& (1 - \int_{R_0} w(r|H_0)dx \cdot P(H_0) \end{aligned}$$

- Probabilitatea unei ratări

$$\begin{aligned} P(D_0 \cap H_1) =& P(D_0|H_1) \cdot P(H_1) \\ =& \int_{R_0} w(r|H_1)dx \cdot P(H_1) \end{aligned}$$

- 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 R_0 astfel încât pentru toți $r \in R_0$, termenul din integrala este **negativ**
  - 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

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

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

Regiuni de decizie

- Se compară eșantionul tot cu un prag T, dar valoarea acestuia este împinsă înspre ipoteza mai puțin probabilă
  - T depinde de raportul $\frac{P(H_0)}{P(H_1)}$
- Regiuni de decizie
  - $R_0 = (-\infty, T]$
  - $R_1 = [T, \infty)$
  - pot fi diferite pentru alte tipuri de zgomot

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 $\mathcal{U}[-4, 4]$?
  3. Calculați probabilitatea unei alarme false și a unei ratări

Criteriul riscului (costului) minim

- Dacă ne afectează mai mult un anume tip de erori (de ex. alarme false) decât celelalte?
- Criteriul riscului (sau costului) minim: deciziile au un cost, se minimizează costul mediu
  - C_{ij} = costul deciziei D_i când ipoteza adevărată este H_j
  - C_{00} = costul unei rejecții corecte
  - C_{10} = costul unei alarme false
  - C_{01} = costul unei ratări
  - C_{11} = costul unei detecții corecte
- Definim **riscul** = costul mediu

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

- Criteriul riscului minim: **se minimizează riscul R**

Calcule

- Demonstrație la tablă
  - se folosește regula lui Bayes
- Concluzie: regula de decizie este

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

Interpretare

- Similar cu primele două criterii, bazat tot pe **raportul de plauzibilitate**
- Atât probabilitățile cât și costurile pot împinge pragul T într-o parte sau alta
- Caz particular: dacă $C_{10} - C_{00} = C_{01} - C_{11}$, se reduce la criteriul probabilității de eroare minime
  - de ex.: dacă $C_{00} = C_{11} = 0$ și $C_{10} = C_{01}$

În zgomot gaussian

- Dacă zgomotul este gaussian (normal), se aplică logaritmul natural, ca la celelalte criterii
- Se obține valoarea pragului T:

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

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

Exemplu

- Exemplu la tablă: 0 / 5, zgomot alb $N(0, \sigma^2)$, un eșantion

Două nivele de semnal nenule

- Dacă semnalul $s_0(t)$ nu este 0, ci are o altă valoare constantă $s_0(t) = B$?
- Distribuția zgomotului $w(r|H_0)$ va fi centrată pe B în loc de 0
- În rest, totul rămâne la fel
- Performanțele sunt determinate de diferența dintre cele două valori $(A - B)$
    - cazul $s_0 = 0$, $s_1 = A$ este identic cu cazul $s_0 = -\frac{A}{2}$, $s_1 = \frac{A}{2}$
- Valabil pentru toate criteriile de decizie

Semnale diferențiale sau unipolare

- Semnal unipolar: o valoare este 0, cealaltă este nenulă
  - $s_0 = 0$, $s_1 = A$
- Semnal diferențial: două valori nenule cu semne contrare, aceeași valoare absolută
  - $s_0 = -\frac{A}{2}$, $s_1 = \frac{A}{2}$
- Care metodă este mai bună?

Semnale diferențiale sau unipolare

- Cu aceeași diferență între nivele, performanțele deciziei sunt identice
- Dar puterea medie a semnalelor diferă
- Pentru semnale diferențiale: $P = \left(\pm \frac{A}{2}\right)^2 = \frac{A^2}{4}$
- Pentru semnale unipolare: $P = P(H_0) \cdot 0 + P(H_1)\,(A)^2 = \frac{A^2}{2}$
  - presupunând probabilități egale $P(H_0) = P(H_1) = \frac{1}{2}$
- Semnalul diferențial necesită putere la jumătate față de cel unipolar (mai bine)

Sumar: criterii de decizie

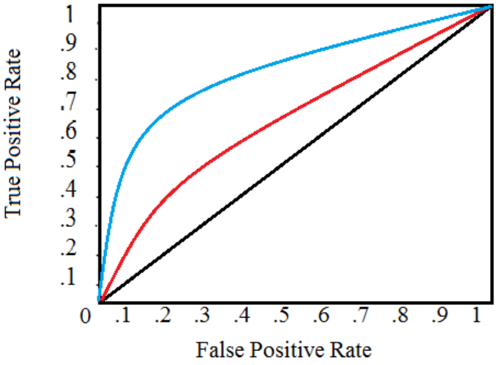
- Am văzut: decizie între două nivele constante, bazată pe 1 eșantion r
- Toate criteriile au la bază un test al raportului de plauzibilitate

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

- Criterii diferite conduc la valori diferite pentru K (pragul de plauzibilitate)
- În funcție de distribuția zgomotului, axa reală este împărțită în regiuni
  - regiunea R_0: dacă r este aici, se decide D_0
  - regiunea R_1: dacă r este aici, se decide D_1
  - de ex. $R_0 = (-\infty, \frac{A+B}{2}]$, $R_1 = (\frac{A+B}{2}, \infty)$ (pentru crit. plauz. max)

Caracteristica de operare a receptorului (ROC)

- Performanța unui receptor este ilustrată cu un grafic numit **"Caracteristica de operare a receptorului"** (**"Receiver Operating Characteristic", ROC)**
- Reprezintă probabilitatea detecției corecte $P_d = P(D_1 \cap H_1)$ în funcție de probabilitatea alarmei false $P_{fa} = P(D_1 \cap H_0)$



Caracteristica de operare a receptorului (ROC)

- Există întotdeauna un **compromis** între P_d și P_{fa}
  - creșterea P_d implică și creșterea P_{fa}
  - pentru a fi siguri că nu ratăm nici un semnal (creșterea P_d), plătim prin creșterea probabilității de alarme false
- Criterii diferite = diferite praguri K = diferite puncte pe grafic = compromisuri diferite
- Cum să creștem performanțele unui receptor?
  - adică să creștem P_D menținând P_{fa} la aceeași valoare

Performanțele detecției în zgomot alb gaussian

- Considerăm probabilități egale $P(H_0) = P(H_1) = \frac{1}{2}$
- Deciziile se iau pe baza raportului de plauzibilitate

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

- Probabilitatea detecției corecte este

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

Performanțele detecției în zgomot alb gaussian

- Probabilitatea alarmei false este

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

- Rezultă că $\frac{T}{\sqrt{2}\sigma} = Q^{-1}\left(P_{fa}\right)$
- Înlocuind în P_d se obține

$$P_d = Q\left(\underbrace{Q^{-1}\left(P_{fa}\right)}_{constant} - \frac{A}{\sqrt{2}\sigma}\right)$$

Raportul semnal zgomot

- **Raportul semnal zgomot (SNR)** $= \frac{\text{puterea semnalului original}}{\text{puterea zgomotului}}$
- Puterea medie a unui semnal = valoarea pătratică medie $= \overline{X^2}$
  - Puterea semnalului original este $\frac{A^2}{2}$
  - Puterea zgomotului este $\overline{X^2} = \sigma^2$ (pentru valoare medie nulă $\mu = 0$)
- În cazul nostru, SNR $= \frac{A^2}{2\sigma^2}$

$$P_d = Q\left(\underbrace{Q^{-1}\left(P_{fa}\right)}_{constant} - \sqrt{SNR}\right)$$

- Pentru P_{fa} de valoare fixă, P_d crește odată cu SNR
  - Q este o funcție monoton descrescătoare

Performanța depinde de SNR

- Performanța receptorului crește odată cu creșterea SNR
  - SNR mare: performanță bună
  - SNR mic: performanță slabă

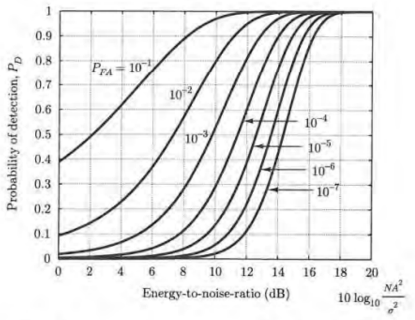


Figure 4: Performanțele detecției depind de SNR

[sursa: Fundamentals of Statistical Signal Processing, Steven Kay]