Decision and Estimation in Data Processing

Chapter II. Elements of Signal Detection Theory

II.1 Introduction

- Signal detection = the problem of deciding which signal is present from 2 or more possibilities
  - one possibility may be that there is no signal
- Based on **noisy** observations
  - signals are affected by noise
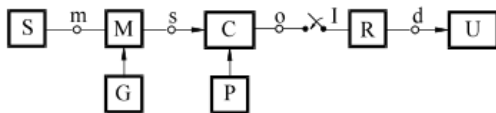
The model for signal detection

Figure 1: Signal detection model

- Contents:
  - Information source: generates messages a_n with probabilities $p(a_n)$
  - Modulator: transmits a signal $s_n(t)$ for message a_n
  - Channel: adds random noise
  - Sampler: takes samples from the signal $s_n(t)$
  - Receiver: **decides** what message a_n has been transmitted

Practical scenarios

- Data transmission
    - constant voltage levels (e.g. $s_n(t) = $ constant)
    - PSK modulation (Phase Shift Keying): $s_n(t) = $ cosine with same frequency but various initial phase
    - FSK modulation (Frequency Shift Keying): $s_n(t) = $ cosines with different frequencies
    - OFDM modulation (Orthogonal Frequency Division Multiplexing): particular case of FSK
- Radar
    - a signal is emitted; if there is an obstacle, the signal gets reflected back
    - the receiver waits for possible reflections of the signal and must decide
        - no reflection is present -> no object
        - reflected signal is present -> object detected

- Decide between more than two signals
- Number of observations:
    - use only one sample
    - use multiple samples
    - observe the whole continuous signal for some time T

II.2 Detection of constant signals

- Simplest case: detection of a constant signal contaminated with white normal noise, using 1 sample
  - two messages a_0 and a_1
  - messages are encoded as constant signals
    - for a_0: send $s_0(t) = 0$
    - for a_1: send $s_1(t) = A$
  - over the signals there is additive noise
  - receiver takes just 1 sample
  - decision: compare sample with a threshold

Threshold-based decision

- The value of the sample taken is $r = s + n$
  - s is the true underlying signal ($s_0 = 0$ or $s_1 = A$)
  - n is a sample of the noise
- n is a (continuous) random variable
- r is a random variable also
  - what distribution does r have compared to n?
- Decision is taken by comparing with a threshold T:
  - if $r < T$, take decision D_0: decide the true signal is s_0
  - if $r \geq T$, take decision D_1: decide the true signal is s_1

Hypotheses

- Receiver chooses between **two hypotheses**:
    - H_0: true signal is s_0 (a_0 has been transmitted)
    - H_1: true signal is s_1 (a_1 has been transmitted)

- Possible results
    1. Correct rejection: no signal present, no signal detected.
        - Decision D_0 when hypothesis is H_0
        - Probability is $P_n = P(D_0 \cap H_0)$
    2. **False alarm**: no signal present, signal detected (error)
        - Decision D_1 when hypothesis is H_0
        - Probability is $P_{fa}P(D_1 \cap H_0)$
    3. **Miss**: signal present, no signal detected (error)
        - Decision D_0 when hypothesis is H_1
        - Probability is $P_m = P(D_0 \cap H_1)$
    4. **Hit**: signal present, signal detected
        - Decision D_1 when hypothesis is H_1
        - Probability is $P_d = P(D_1 \cap H_1)$

Maximum likelihood criterion

- Choose the hypothesis that **seems most likely** given the observed sample r
- The **likelihood** of an observation r = the probability density of r given a hypothesis H_0 or H_1
- Likelihood in case of hypothesis H_0: $w(r|H_0)$
    - r is only noise, so value is taken from the noise distribution
- Likelihood in case of hypothesis H_1: $w(r|H_1)$
    - r is A + noise, so value is taken from the distribution of (A + noise)
- **Likelihood ratio** test:
$$\frac{w(r|H_1)}{w(r|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} 1$$

- Consider noise having a normal distribution
- Plot the two density functions for H_0, H_1

Decision via threshold

- Likelihood ratio test for ML = comparing r with a threshold T
- The threshold = the cross-over point of the two distributions

Normal noise

- ▶ Particular case: the noise has normal distribution $\mathcal{N}(0, \sigma^2)$
- ▶ Likelihood ratio is $\frac{w(r|H_1)}{r|H_0} = \frac{e^{-\frac{(r-A)^2}{2\sigma^2}}}{e^{-\frac{r^2}{2\sigma^2}}} \underset{H_0}{\overset{H_1}{\gtrless}} 1$
- ▶ For normal distribution, it is easier to apply *natural logarithm* to the terms
    - ▶ logarithm is a monotonic increasing function, so it won't change the comparison
    - ▶ if $A < B$, then $\log(A) < \log(B)$
- ▶ The **log-likelihood** of an observation = the logarithm of the likelihood value
    - ▶ usually the natural logarithm, but any one can be used

Log-likelihood test for ML

- For normal noise, the ML decision means the log-likelihood test

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

- Applying square root

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

- $|r - A|$ = distance from r to A, $|r|$ = distance from r to 0
- ML decision with normal noise: choose the value 0 or A which is **nearest** to r
    - very general principle, encountered in many other scenarios
    - also known as **nearest neighbor** principle / decision
    - ML receiver is also known as **minimum distance receiver**
    - equivalent with setting a threshold $T = \frac{A}{2}$

Generalizations

- What if the noise has another distribution?
  - Threshold T is still the cross-over point, whatever that is
  - There can be more cross-overs, so multiple thresholds
  - Can think that $\mathbb{R}$ axis is split into **decision regions** R_0 and R_1

- What if the noise distributions are different for H_0 and H_1?
  - Threshold T is the cross-over point, whatever that is

- What if the signal $s_0(t)$ (for H_0) is not 0, but another constant value B?
  - T is the crossover point, the distributions are centered on B and A
  - In case of normal noise, choose B or A whichever is nearest (threshold is at middle point)

Generalizations

- What if we have more than two signal levels?
  - e.g. 4 possible signals: -6, -2, 2, 6
  - Just choose the most likely hypothesis, out of 4 likelihood functions
  - Not a single threshold value, now there are more

Exercises

- A signal can have two possible values, 0 or 5. The receiver takes one sample with value $r = 2.25$

    1. Considering that the noise is white gaussian noise, what signal is decided based on the Maximum Likelihood criterion?
    2. What if the signal 0 is affected by gaussian noise $\mathcal{N}(0, 0.5)$, while the signal 5 is affected by uniform noise $\mathcal{U}[-4, 4]$?
    3. Repeat a. and b. assuming the value 0 is replaced by -1

- A signal can have four possible values: -6, -2, 2, 6. Each value lasts for 1 second. The signal is affected by white noise with normal distribution. The receiver takes 1 sample per second. Using ML criterion, decide what signal has been transmitted, if the received samples are:

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

Computing conditional error probabilities

- ▶ We can compute the conditional probabilities of errors
- ▶ Consider the decision regions:
  - ▶ R_0: when $r \in R_0$, decision is D_0, i.e. (∞, T) for gaussian noise
  - ▶ R_1: when $r \in R_1$, decision is D_1, i.e. $[T, \infty)$ for gaussian noise
- ▶ Probability of false alarm *if original signal is* $s_0(t)$

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

- ▶ Probability of miss *if original signal is* $s_1(t)$

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

- ▶ These probabilities do not account for the probability that the signal actually is $s_0(t)$ or $s_1(t)$
  - ▶ they are **conditional** ("if")
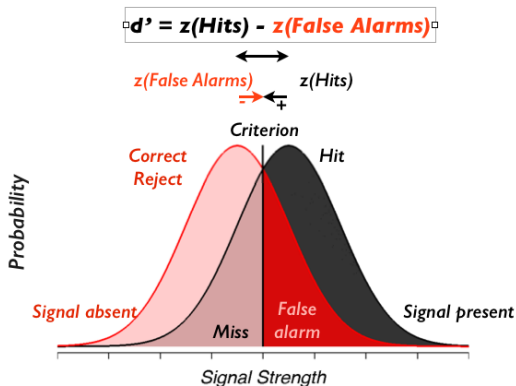
Computing conditional error probabilities

Figure 2: Conditional error probabilities

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

- Reminder: the Bayes rule

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

- Interpretation
  - The probability $P(A)$ is taken out from $P(B|A)$
  - $P(B|A)$ gives no information on $P(A)$, the chances of A actually happening
  - Example: P(score | shoot) $= \frac{1}{2}$. How many goals are scored?

▶ A signal can have two possible values, 0 or 5. The signal 0 is affected by gaussian noise $\mathcal{N}(0, 0.5)$, while the signal 5 is affected by uniform noise $\mathcal{U}[-4, 4]$. The receiver performs ML decision based on a single sample.

1. Compute the probability of a wrong decision when the original signal is $s_0(t)$
2. Compute the probability of a wrong decision when the original signal is $s_1(t)$

Pitfalls of ML decision criterion

- The ML is based on comparing **conditional** probability density functions
    - conditioned by H_0 or by H_1
- Conditioning by H_0 and H_1 ignores the probability of H_0 or H_1 actually happening
    - We don't know how $p(H_0)$ or $P(H_1)$
- If $p(H_0) > p(H_1)$, we may want to move the threshold towards H_1, and vice-versa
    - because it is more likely that the signal is $s_0(t)$
    - and thus we want to "encourage" decision D_0

- Takes into account the probabilities $P(H_0)$ and $P(H_1)$
- Goal is to **minimize the total probability of error** P_e
    - errors = false alarms and misses
- We need to find the decision regions R_0 and R_1

Probability of error

▶ Probability of false alarm

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

▶ Probability of miss

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

▶ The sum is

$$
P_e = P(H_0) + \int_{R_0} [w(r|H_1) \cdot P(H_1) - w(r|H_0) \cdot P(H_0)]dx
$$

Minimum probability of error

- We want to minimize P_e, i.e. to minimize the integral
- To minimize the integral, we choose R_0 such that for all $r \in R_0$, the term inside the integral is **negative**
  - because integrating over all the interval where the function is negative ensures minimum value of integral
- So, when $w(r|H_1) \cdot P(H_1) - w(r|H_0) \cdot P(H_0) < 0$ we have $r \in R_0$, i.e. decision D_0
- Conversely, when $w(r|H_1) \cdot P(H_1) - w(r|H_0) \cdot P(H_0) > 0$ we have $r \in R_1$, i.e. decision D_1
- Therefore

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

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

- Similar to ML, but threshold depends on probabilities of the two hypotheses
    - When one hypotheses is more likely than the other, the threshold is pushed in its favor, towards the other
- Also based on a **likelihood ratio** test, just like ML

Minimum probability of error - gaussian noise

- Assuming the noise is gaussian (normal), $\mathcal{N}(0, \sigma^2)$

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

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

- Apply natural logarithm

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

- Equivalently

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

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

Decision regions

- We still compare with a threshold T, but its value is shifted towards the less probable hypothesis

    - T depends on the ratio $\frac{P(H_0)}{P(H_1)}$

- Decision regions

    - $R_0 = (-\infty, T]$
    - $R_1 = [T, \infty)$
    - will be different for other noise distributions (non-gaussian)

- An information source provides two messages with probabilities $p(a_0) = \frac{2}{3}$ and $p(a_1) = \frac{1}{3}$. The messages are encoded as constant signals with values -5 (a_0) and 5 (a_1). The signals are affected by gaussian noise $\mathcal{N}(0, \sigma^2 = 1)$ The receiver takes one sample r. Decision is done by comparing r with a threshold value T, as follows: if $r < T$ it is decided that the transmitted message is a_0, otherwise it is a_1.

    1. Find the threshold value T according to the minimum probability of error criterion
    2. What if the signal 5 is affected by uniform noise $\mathcal{U}[-4, 4]$?
    3. What are the probabilities of false alarm and of miss?

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**

- 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**

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

- 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

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

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

- Example at blackboard: 0 / 5, random noise with $N(0, \sigma^2)$, one sample

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 $s_1 = \frac{A}{2}$
- Valid for all decision criteria

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
  - $s_0 = -\frac{A}{2}$, $s_1 = \frac{A}{2}$
- Which is better?

- 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 $P(H_0) = P(H_1) = \frac{1}{2}$
- 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" (ROC)** graph
- It is a graph of hit probability $P_d = P(D_1 \cap H_1)$ (correct detection) as a function of false alarm probability $P_{fa} = P(D_1 \cap H_0)$
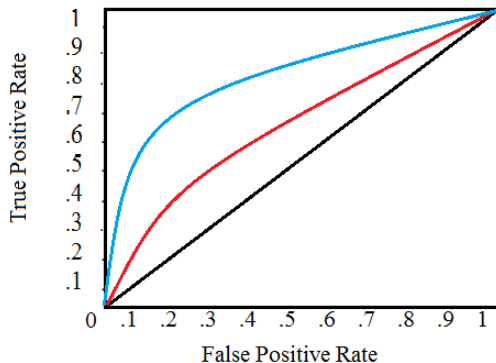

Figure 3: Sample ROC curves

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

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

- Hit probability is

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

Performance of likelihood-ratio decoding in WGN

- False alarm probability is

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

- Therefore $\frac{T}{\sqrt{2}\sigma} = Q^{-1}(P_{fa})$
- Replacing in P_{hit} yields

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

Signal-to-noise ratio

- **Signal-to-noise ratio (SNR)** $= \frac{\text{power of original signal}}{\text{power of noise}}$
- Average power of a signal = average squared value $= \overline{X^2}$
    - Original signal power is $\frac{A^2}{2}$
    - Noise power is $\overline{X^2} = \sigma^2$ (when noise mean value $\mu = 0$)
- In our case, SNR $= \frac{A^2}{2\sigma^2}$

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

- For a fixed P_{fa}, P_{hit} increases with SNR
    - Q is a monotonic decreasing function

Performance depends on SNR

- Receiver performance increases with SNR increase
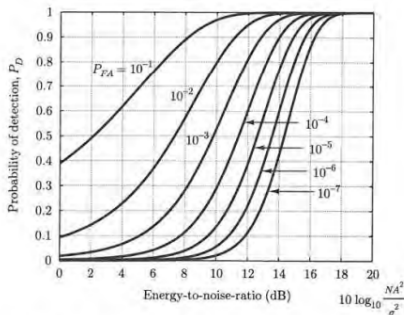 - high SNR: good performance
  - poor SNR: bad perfomance

Figure 4: Detection performance depends on SNR

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