EE226a - Summary of Lecture 8
Binary Detection under AWGN
Antonis Dimakis

I. SUMMARY

Here are the key ideas and results:

- Section [III] binary detection with vector observations under AWGN. The ideas from L7 carry over to the case of vector observations.
- Definition: Sufficient statistic, in Section [III-A].
- BER calculations and system design, in Section [IV].

II. EXAMPLE: SYSTEM DESIGN UNDER POWER CONSTRAINTS

Let us consider the basic (scalar observation) model of binary detection in L7, under Gaussian noise:

\[ Y = \mu X + W, \]

where \( X \in \{0, 1\} \) is the r.v. we want to guess, \( W \) is noise \( N(0, \sigma^2) \), and \( Y \) is the observation. Assume \( X \) has the distribution \( P(X = 0) = p_0, P(X = 1) = p_1. \)

In L7 we saw that \( \hat{X} := \text{MAP}[X|Y = y] = 1 \) or 0 depending on whether \( Y > \theta \) or \( < \theta \), respectively, where

\[ \theta := \frac{\mu}{2} + \frac{\sigma^2}{\mu} \log\frac{p_0}{p_1}. \]

Then,

\[
P(X \neq \hat{X}) = P[X \neq \hat{X}|X = 0]p_0 + P[X \neq \hat{X}|X = 1]p_1
= P(N(0, \sigma^2) \geq \theta)p_0 + P(N(\mu, \sigma^2) \leq \theta)p_1
= P(\frac{N(0, 1) \geq \theta}{\sigma})p_0 + P(\frac{N(0, 1) \leq \theta - \mu}{\sigma})p_1.
\]

When \( p_0 = p_1 \) we have the further simplification

\[
P(X \neq \hat{X}) = P\left(N(0, 1) \geq \frac{\mu}{2\sigma}\right).
\]

Now, if we wanted to design our communication system such that the bit-error-rate (BER) is less than, say \( 10^{-10} \), then we could use (1), to calculate the average energy per bit required to achieve this. (Note that \( P(N(0, 1) \geq 6) \leq 10^{-10}. \) For example, if the amplitude of the signal at the sender is \( A \), and \( G \) the channel gain, then \( \mu = AG \). Since no signal is sent when \( X = 0 \), which happens half of the time \((p_0 = p_1 = 1/2)\), the average energy per bit is \( p = \frac{1}{2}A^2 \). Thus,

\[
\frac{p}{\sigma^2} \geq \frac{72}{G^2},
\]

gives a lower bound in terms of the power of the noise.

III. BINARY DETECTION WITH VECTOR OBSERVATIONS

The key ideas in L7 carry over to the vector case: Let \( X \in \{0, 1\} \) as before, \( \mu_0, \mu_1 \in \mathbb{R}^n \) and

\[ Y = \mu_1 + Z, \quad \text{when } X = 1, \]

where \( Z = N(0, K) \). This can model a variety of situations. One can think of \( Y \) as the received signals at an antenna array, each coordinate corresponding to the output of a different antenna. Another way to look at this model, is by considering \( \mu_0, \mu_1 \) as waveforms, corrupted by noise with an autocorrelation structure specified by matrix \( K \).

Let’s compute \( \text{MLE}[X|Y = y] \) for the case that \( |K| \neq 0. \)

\[
\frac{f_{Y|X}(y|1)}{f_{Y|X}(y|0)} = \exp\left(-\frac{1}{2}(y - \mu_1)^T K^{-1} (y - \mu_1) + \frac{1}{2}(y - \mu_0)^T K^{-1} (y - \mu_0)\right)
= \exp\left((\mu_1 - \mu_0)^T K^{-1} y - \frac{1}{2} \mu_1^T K^{-1} \mu_1 + \frac{1}{2} \mu_0^T K^{-1} \mu_0\right).
\]
The last expression is the likelihood-ratio. Usually, we work with the exponent, the so called log-likelihood-ratio (LLR)

\[
(\mu_1 - \mu_0)^T K^{-1} y - \frac{1}{2} \mu_1^T K^{-1} \mu_1 + \frac{1}{2} \mu_0^T K^{-1} \mu_0,
\]

which we denote by \(\text{LLR}[X|Y = y]\).

Thus,

\[
\text{MLE}[X|Y = y] = \begin{cases} 
1, & \text{if } \text{LLR}[X|Y = y] > 0 \\
0, & \text{otherwise}
\end{cases}
\]

For \(\text{MAP}[X|Y = y]\) one needs to compare \(f_{Y|X}(y|x)\) with \(f_{Y|X}(y|x_0)\). Hence,

\[
\text{MAP}[X|Y = y] = \begin{cases} 
1, & \text{if } \text{LLR}[X|Y = y] > \log(p_1/p_0) \\
0, & \text{otherwise}.
\end{cases}
\]

Whereas we started with a nonlinear problem, that of minimizing the probability of incorrect detection, we see that we get simple linear rules. To see why this happens, notice that when \(K = I\), from (2), one chooses \(i = 0\) or \(i = 1\) depending on which of \(\mu_0, \mu_1\) is closest to the observation vector \(y\), in the sense of Euclidean distance \(|y - \mu_1|\). Hence, the decision rule is specified by the two halfspaces divided by the line of equidistant points from \(\mu_0, \mu_1\).

A. Sufficient statistic

Notice that even though the dimensionality \(n\) of the observations might be high, what is actually sufficient in determining the LLR -and hence MAP and MLE-, is a linear function of \(Y\). Such a function is called a sufficient statistic. More precisely, we will call \(g(Y)\) a sufficient statistic\(^1\), if

\[
f_{Y|X}(y|x) = f(g(y), x)G(y),
\]

for some real functions \(F, G\). There can be many sufficient statistics; a trivial one is \(Y\) itself. In the definition, \(X, x\) may be also vectors.

B. Matched filter

One common way to compute \(\mu_1^T y\) needed in the sufficient statistic, is by a matched filter. Note that

\[
\mu_1^T y = \sum_{t=1}^{n} \mu_i(t) y_t = \sum_{t=1}^{n} h_i(n-t) y_t,
\]

where \(\mu_i = (\mu_i(1), \ldots, \mu_i(n))^T\), and \(h_i(t) = \mu_i(n-t)\). Thus, \(\mu_1^T y\) can be computed by passing \(y\) through a filter with an impulse response \(h_i(\cdot)\) “matched” to the signal \(\mu_i\).

IV. BIT ERROR RATE CALCULATIONS

Assume \(K = \sigma^2 I\), for simplicity. The probability of incorrect detection is

\[
P(\text{error}) = P \left( \text{LLR}[X|Y] \geq \log \frac{p_0}{p_1} | X = 0 \right) p_0 + P \left( \text{LLR}[X|Y] < \log \frac{p_0}{p_1} | X = 1 \right) p_1.
\]

Hence, we need to determine the distribution of \(\text{LLR}(X|Y)\). \(Y\) is jointly Gaussian, and

\[
E[\text{LLR}(Y)|X = 0] = \frac{(\mu_1 - \mu_0)^T \mu_0}{\sigma^2} + \frac{\mu_0^T \mu_0}{2\sigma^2} - \frac{\mu_1^T \mu_1}{2\sigma^2}
\]

\[
= \frac{1}{\sigma^2} (\mu_1 - \mu_0)^T \mu_0 - \frac{1}{2\sigma^2} (\mu_1 + \mu_0)^T (\mu_0 - \mu_1)
\]

\[
= -\frac{1}{2\sigma^2} \|\mu_1 - \mu_0\|^2
\]

Similarly, we find

\[
\text{Var}(\text{LLR}(Y)|X = 0) = \frac{\|\mu_1 - \mu_0\|^2}{\sigma^2}.
\]

Both the variance and conditional mean depend only on the ratio \(\|\mu_1 - \mu_0\|^2/\sigma =: \gamma\).

Thus, (4) becomes

\[
P(\text{error}) = P \left( N(0, 1) \geq \frac{1}{\gamma} \log \frac{p_0}{p_1} + \frac{\gamma}{2} \right) p_0 + P \left( N(0, 1) \leq -\frac{1}{\gamma} \log \frac{p_0}{p_1} + \frac{\gamma}{2} \right) p_1.
\]

\(^1\)This is one of many equivalent definitions, that suffices for our purposes.
Observe that BER depends on the energy in the difference of the two signals, $\mu_1$ and $\mu_0$, and how these compare to the power of the noise $\sigma^2$.

When $p_0 = p_1$, the above simplifies to

$$P(\text{error}) = P\left(N(0, 1) \geq \frac{\gamma}{2}\right).$$

Usually one wants to minimize this, under some power constraint. Part of the problem is to determine the signals $\mu_i$. One way to do this is to set $\mu_0 = 0$, i.e. allocate no power to this signal, and transmit any $\mu_1$ at full power.

V. EXAMPLE: RELATION WITH THE SCALAR CASE

Assume that a symbol is a bit-string of fixed length $n$, where each of the $n$ bits is picked independently. Independent noise $W = N(0, I_{n \times n})$ corrupts each bit in a symbol. What is the MAP detector?

In this case, we have $2^n$ possible symbols in the set $\{0, 1\}^{1,...,n} =: S$, but we don’t need to have $2^n$ matched filters for implementing MAP. One has,

$$\text{MAP}[X|Y = y] = \arg \max_{x \in S} f_{Y|X}(y|x)P(X = z)$$

$$= \arg \max_{x \in S} \left( \prod_{i=1}^{n} f_{Y_i|X_i}(y_i|x_i) \right) \left( \prod_{j=1}^{n} P(X_j = z_j) \right)$$

$$= \arg \max_{x \in S} \prod_{i=1}^{n} f_{Y_i|X_i}(y_i|x_i)P(X_i = z_i),$$

where the second line follows from independence of the $X_i$’s. Thus, $\text{MAP}[X|Y = y]$ will estimate the $i$-th bit of $X$ as $\hat{X}_i = \text{MAP}[X_i|Y_i = y_i]$. 
