Detection Theory
Detection theory

• A the last topic of the course, we will briefly consider detection theory.
• The methods are based on estimation theory and attempt to answer questions such as
  • Is a signal of specific model present in our time series? E.g., detection of noisy sinusoid; beep or no beep?
  • Is the transmitted pulse present at radar signal at time \( t \)?
  • Does the mean level of a signal change at time \( t \)?
  • After calculating the mean change in pixel values of subsequent frames in video, is there something moving in the scene?
Detection theory

• The area is closely related to hypothesis testing, which is widely used e.g., in medicine: Is the response in patients due to the new drug or due to random fluctuations?
• In our case, the hypotheses could be

\[ \mathcal{H}_1 : x[n] = A \cos(2\pi f_0 n + \phi) + w[n] \]
\[ \mathcal{H}_0 : x[n] = w[n] \]

• This example corresponds to detection of noisy sinusoid.
• The hypothesis \( \mathcal{H}_1 \) corresponds to the case that the sinusoid is present and is called alternative hypothesis.
• The hypothesis \( \mathcal{H}_0 \) corresponds to the case that the measurements consists of noise only and is called null hypothesis.
• Neyman-Pearson approach is the classical way of solving detection problems in an optimal manner.
• It relies on so called Neyman-Pearson theorem.
• Before stating the theorem, consider a simplistic detection problem, where we observe one sample $x[n]$ from one of two densities: $N(0, 1)$ or $N(1, 1)$.
• The task is to choose the correct density in an optimal manner.
• Our hypotheses are now

$$\mathcal{H}_1 : \mu = 1,$$

$$\mathcal{H}_0 : \mu = 0.$$  

• An obvious approach for deciding the density would choose the one, which is higher for a particular $x[0]$. 
More specifically, study the likelihoods and choose the more likely one.

The likelihoods are

\[ H_1 : p(x[0] \mid \mu = 1) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{(x[n] - 1)^2}{2} \right) \].

\[ H_0 : p(x[0] \mid \mu = 0) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{(x[n])^2}{2} \right) \].

Now, one should select \( H_1 \) if \( p(x[0] \mid \mu = 1) > p(x[0] \mid \mu = 0) \).
Introductory Example

• Let’s state this in terms of $x[0]$:

\[
p(x[0] | \mu = 1) > p(x[0] | \mu = 0)
\]

\[\iff \frac{p(x[0] | \mu = 1)}{p(x[0] | \mu = 0)} > 1\]

\[\iff \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{(x[n] - 1)^2}{2} \right) > 1\]

\[\iff \exp \left( -\frac{(x[n] - 1)^2 - x[n]^2}{2} \right) > 1\]
Introductory Example

\[ \Leftrightarrow (x[n]^2 - (x[n] - 1)^2) > 0 \]
\[ \Leftrightarrow 2x[n] - 1 > 0 \]
\[ \Leftrightarrow x[n] > \frac{1}{2}. \]

- In other words, choose \( \mathcal{H}_1 \) if \( x[0] > 0.5 \) and \( \mathcal{H}_0 \) if \( x[0] < 0.5 \).
- Studying the ratio of likelihoods on the second row of the derivation is the key.
- This ratio is called *likelihood ratio*, and comparison to a threshold \( \gamma \) (here \( \gamma = 1 \)) is called *likelihood ratio test* (LRT).
Introductory Example

• Note, that it is also possible to study posterior probability ratios \( p(H_1 | x) / p(H_0 | x) \) instead of the above likelihood ratio \( p(x | H_1) / p(x | H_0) \).

• However, using Bayes rule, this MAP test turns out to be

\[
\frac{p(x | H_1)}{p(x | H_0)} > \frac{p(H_1)}{p(H_2)},
\]

i.e., the only effect of using posterior probability is on the threshold for the LRT.
Error Types

- It might be that the detection problem is not symmetric and some errors are more costly than others.
- For example, when detecting a disease, a missed detection is more costly than a false alarm.
- The tradeoff between misses and false alarms can be adjusted using the threshold of the LRT.
Error Types

- The below figure illustrates the probabilities of the two kinds of errors. The red area on the left corresponds to the probability of choosing $\mathcal{H}_1$ while $\mathcal{H}_0$ would hold (false match). The blue area is the probability of choosing $\mathcal{H}_0$ while $\mathcal{H}_1$ would hold (missed detection).
Error Types

• It can be seen that we can decrease either probability arbitrarily small by adjusting the detection threshold.

• Left: large threshold; small probability of false match (red), but a lot of misses (blue).

• Right: small threshold; only a few missed detections (blue), but a huge number of false matches (red).
Error Types

• Probability of false alarm for the threshold $\gamma = 1.5$ is

$$P_{FA} = P(x[0] > \gamma \mid \mu = 0) = \int_{1.5}^{\infty} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{(x[n])^2}{2} \right) \, dx[n] \approx 0.0668.$$  

• Probability of missed detection is

$$P_M = P(x[0] > \gamma \mid \mu = 1) = \int_{-\infty}^{1.5} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{(x[n] - 1)^2}{2} \right) \, dx[n] \approx 0.6915.$$  

• An equivalent, but more useful term is the complement of $P_M$: probability of detection:

$$P_D = 1 - P_M = \int_{1.5}^{\infty} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{(x[n] - 1)^2}{2} \right) \, dx[n] \approx 0.3085.$$
Neyman-Pearson Theorem

- Since $P_{FA}$ and $P_D$ depend on each other, we would like to maximize $P_D$ subject to given maximum allowed $P_{FA}$. Luckily the following theorem makes this easy.
- **Neyman-Pearson Theorem:** For a fixed $P_{FA}$, the likelihood ratio test maximizes $P_D$ with the decision rule

$$L(x) = \frac{p(x; \mathcal{H}_1)}{p(x; \mathcal{H}_0)} > \gamma,$$

with threshold $\gamma$ is the value for which

$$\int_{x: L(x) > \gamma} p(x; \mathcal{H}_0) \, dx = P_{FA}.$$
Neyman-Pearson Theorem

- As an example, suppose we want to find the best detector for our introductory example, and we can tolerate 10% false alarms ($P_{FA} = 0.1$).
- According to the theorem, the detection rule is:

$$\text{Select } H_1 \text{ if } \frac{p(x | \mu = 1)}{p(x | \mu = 0)} > \gamma$$

The only thing to find out now is the threshold $\gamma$ such that

$$\int_{\gamma}^{\infty} p(x | \mu = 0) \, dx = 0.1.$$ 

This can be done with Matlab function `icdf`, which solves the inverse cumulative distribution function.
Neyman-Pearson Theorem

- Unfortunately \texttt{icdf} solves the $\gamma$ for which
  \[
  \int_{-\infty}^{\gamma} p(x \mid \mu = 0) \, dx = 0.1 \quad \text{instead of} \quad \int_{\gamma}^{\infty} p(x \mid \mu = 0) \, dx = 0.1. 
  \]
  Thus, we have to use the function like this:
  \texttt{icdf('norm', 1 - 0.1, 0, 1)}, which gives $\gamma \approx 1.2816$.
- Similarly, we can also calculate the $P_D$ with this threshold:
  \[
  P_D = \int_{1.2816}^{\infty} p(x \mid \mu = 1) \, dx \approx 0.3891. 
  \]
Detector for a known waveform

- The NP approach applies to all cases where likelihoods are available.
- An important special case is that of a known waveform $s[n]$ embedded in WGN sequence $w[n]$:
  \[ H_1 : x[n] = s[n] + w[n] \]
  \[ H_0 : x[n] = w[n]. \]
- An example of a case where the waveform is known could be detection of radar signals, where a pulse $s[n]$ transmitted by us is reflected back after some propagation time.
Detector for a known waveform

- For this case the likelihoods are

\[
p(x \mid \mathcal{H}_1) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left( -\frac{(x[n] - s[n])^2}{2\sigma^2} \right),
\]

\[
p(x \mid \mathcal{H}_0) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left( -\frac{(x[n])^2}{2\sigma^2} \right).
\]

- The likelihood ratio test is easily obtained as

\[
\frac{p(x \mid \mathcal{H}_1)}{p(x \mid \mathcal{H}_0)} = \exp \left[ -\frac{1}{2\sigma^2} \left( \sum_{n=0}^{N-1} (x[n] - s[n])^2 - \sum_{n=0}^{N-1} (x[n])^2 \right) \right] > \gamma.
\]
Detector for a known waveform

- This simplifies by taking the logarithm from both sides:

\[
-\frac{1}{2\sigma^2} \left( \sum_{n=0}^{N-1} (x[n] - s[n])^2 - \sum_{n=0}^{N-1} (x[n])^2 \right) > \ln \gamma.
\]

- This further simplifies into

\[
\frac{1}{\sigma^2} \sum_{n=0}^{N-1} x[n]s[n] - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (s[n])^2 > \ln \gamma.
\]
Detector for a known waveform

Since $s[n]$ is a known waveform ($= \text{constant}$), we can simplify the procedure by moving it to the right hand side and combining it with the threshold:

$$\sum_{n=0}^{N-1} x[n]s[n] > \sigma^2 \ln \gamma + \frac{1}{2} \sum_{n=0}^{N-1} (s[n])^2.$$  

We can equivalently call the right hand side as our threshold (say $\gamma'$) to get the final decision rule

$$\sum_{n=0}^{N-1} x[n]s[n] > \gamma'.$$
Examples

• This leads into some rather obvious results.
• The detector for a known DC level in WGN is

\[ \sum_{n=0}^{N-1} x[n] A > \gamma \Rightarrow A \sum_{n=0}^{N-1} x[n] > \gamma \]

Equally well we can set a new threshold and call it \( \gamma' = \gamma/(AN) \). This way the detection rule becomes: \( \bar{x} > \gamma' \). Note that a negative \( A \) would invert the inequality.
Examples

• The detector for a sinusoid in WGN is

\[ \sum_{n=0}^{N-1} x[n]A \cos(2\pi f_0 n + \phi) > \gamma \Rightarrow A \sum_{n=0}^{N-1} x[n] \cos(2\pi f_0 n + \phi) > \gamma. \]

• Again we can divide by \( A \) to get

\[ \sum_{n=0}^{N-1} x[n] \cos(2\pi f_0 n + \phi) > \gamma'. \]

• In other words, we check the correlation with the sinusoid. Note that the amplitude \( A \) does not affect our statistic, only the threshold which is anyway selected according to the fixed \( P_{\text{FA}} \) rate.
Examples

- As an example, the below picture shows the detection process with $\sigma = 0.5$. 

  ![Diagram showing noiseless signal, noisy signal, and detection result graphs.](image-url)
Detection of random signals

- The problem with the previous approach was that the model was too restrictive; the results depend on how well the phases match.
- The model can be relaxed by considering random signals, whose exact form is unknown, but the correlation structure is known. Since the correlation captures the frequency (but not the phase), this is exactly what we want.
- In general, the detection of a random signal can be formulated as follows.
Detection of random signals

- Suppose $\mathbf{s} \sim \mathcal{N}(0, \mathbf{C}_s)$ and $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$. Then the detection problem is a hypothesis test

  \[
  \mathcal{H}_0 : \mathbf{x} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}) \\
  \mathcal{H}_1 : \mathbf{x} \sim \mathcal{N}(0, \mathbf{C}_s + \sigma^2 \mathbf{I})
  \]

- It can be shown (see Kay-2, p. 145), that the decision rule becomes

  \[
  \text{Decide } \mathcal{H}_1, \text{ if } \mathbf{x}^T \hat{\mathbf{s}} > \gamma,
  \]

  where

  \[
  \hat{\mathbf{s}} = \mathbf{C}_s (\mathbf{C}_s + \sigma^2 \mathbf{I})^{-1} \mathbf{x}.
  \]
Detection of random signals

• The term $\hat{s}$ is in fact the estimate of the signal; more specifically, the linear Bayesian MMSE estimator, which assumes linearity for the estimator (similar to BLUE).

• A particular special case of a random signal is the Bayesian linear model.

• The Bayesian linear model assumes linearity $x = H\theta + w$ together with a prior for the parameters, such as $\theta \sim N(0, \sigma^2 I)$

• Consider the following detection problem:

$$H_0 : x = w$$
$$H_1 : x = H\theta + w$$
Detection of random signals

- Within the earlier random signal framework, this is written as

\[ H_0 : \mathbf{x} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}) \]
\[ H_1 : \mathbf{x} \sim \mathcal{N}(0, \mathbf{C}_s + \sigma^2 \mathbf{I}) \]

with \( \mathbf{C}_s = \mathbf{H} \mathbf{C}_\theta \mathbf{H}^T \).

- The assumption \( \mathbf{s} \sim \mathcal{N}(0, \mathbf{H} \mathbf{C}_\theta \mathbf{H}^T) \) states that the exact form of the signal is unknown, and we only know its covariance structure.
Detection of random signals

- This is helpful in the sinusoidal detection problem: we are not interested in the phase (included by the exact formulation $x[n] = A \cos(2\pi f_0 n + \phi)$), but only in the frequency (as described by the covariance matrix $HC_\theta H^T$).
- Thus, the decision rule becomes:

  $$\text{Decide } \mathcal{H}_1, \text{ if } x^T \hat{s} > \gamma,$$

  where

  $$\hat{s} = C_s (C_s + \sigma^2 I)^{-1} x$$

  $$= HC_\theta H^T (HC_\theta H^T + \sigma^2 I)^{-1} x$$
Detection of random signals

• Luckily the decision rule simplifies quite a lot by noticing that the last part is the MMSE estimate of $\theta$:

$$x^T \hat{s} = x^T C_\theta H^T (HC_\theta H^T + \sigma^2 I)^{-1} x$$

$$= x^T H \hat{\theta}.$$  

• An example of applying the linear model is in Kay: Statistical Signal Processing, vol. 2; Detection Theory, pages 155-158.

• In the example, a Rayleigh fading sinusoid is studied, which has an unknown amplitude $A$ and phase term $\phi$. Only the frequency $f_0$ is assumed to be known.
Detection of random signals

- This can be manipulated into a linear model form with two unknowns corresponding to $A$ and $\phi$.
- The final result is the decision rule:

\[ \left| \sum_{n=0}^{N-1} x[n] \exp(-2\pi if_0 n) \right| > \gamma. \]

- As an example, the below picture shows the detection process with $\sigma = 0.5$.
- Note the simplicity of Matlab implementation:

```matlab
h = exp(-2*pi*sqrt(-1)*f0*n);
y = abs(conv(h,x));
```
Detection of random signals

- Noiseless signal
- Noisy signal
- Detection result
A usual way of illustrating the detector performance is the Receiver Operating Characteristics curve (ROC curve).

This describes the relationship between $P_{FA}$ and $P_D$ for all possible values of the threshold $\gamma$.

The functional relationship between $P_{FA}$ and $P_D$ depends on the problem (and the selected detector, although we have proven that LRT is optimal).
Receiver Operating Characteristics

- For example, in the DC level example,

\[ P_D(\gamma) = \int_\gamma^\infty \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{(x-1)^2}{2} \right) \, dx \]

\[ P_{FA}(\gamma) = \int_\gamma^\infty \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{x^2}{2} \right) \, dx \]

- It is easy to see the relationship:

\[ P_D(\gamma) = \int_\gamma^{\gamma-1} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{x^2}{2} \right) \, dx = P_{FA}(\gamma - 1). \]
Receiver Operating Characteristics

- Plotting the ROC curve for all $\gamma$ results in the following curve.

![ROC Curve](image)
Receiver Operating Characteristics

- The higher the ROC curve, the better the performance.
- A random guess has diagonal ROC curve.
- In the DC level case, the performance increases if the noise variance $\sigma^2$ decreases. Below are the ROC plots for various values of $\sigma^2$. 
• This gives rise to a widely used measure for detector performance: the *Area Under (ROC) Curve*, or AUC criterion.
Composite hypothesis testing

• In the previous examples the parameter values specified the distribution completely; e.g., either $A = 1$ or $A = 0$.
• Such cases are called *simple hypotheses testing*.
• Often we can’t specify exactly the parameters for either case, but instead a range of values for each case.
• An example could be our DC model $x[n] = A + w[n]$ with

$$H_1 : A \neq 0$$

$$H_0 : A = 0$$
Composite hypothesis testing

- The question can be posed in a probabilistic manner as follows:

  What is the probability of observing $x[n]$ if $H_0$ would hold?

- If the probability is small (e.g., all $x[n] \in [0.5, 1.5]$, and let’s say the probability of observing $x[n]$ under $H_0$ is 1 %), then we can conclude that the null hypothesis can be rejected with 99% confidence.
An example

- As an example, consider detecting a biased coin in a coin tossing experiment.
- If we get 19 heads out of 20 tosses, it seems rather likely that the coin is biased.
- How to pose the question mathematically?
- Now the hypotheses is

\[ H_1 : \text{coin is biased}: \ p \neq 0.5 \]
\[ H_0 : \text{coin is unbiased}: \ p = 0.5, \]

where \( p \) denotes the probability of a head for our coin.
An example

- Additionally, let’s say, we want 99% confidence for the test.
- Thus, we can state the hypothesis test as: "what is the probability of observing at least 19 heads assuming \( p = 0.5 \)?"
- This is given by the binomial distribution

\[
\binom{20}{19} 0.5^{19} \cdot 0.5^1 + 0.5^{20} \approx 0.00002.
\]

- Since 0.00002 < 1%, we can reject the null hypothesis and the coin is biased.
An example

• Actually, the 99% confidence was a bit loose in this case.
• We could have set a 99.98% confidence requirement and still reject the null hypothesis.
• The upper limit for the confidence (here 99.98%) is widely used and called the *p-value*.
• More specifically,

  *The p-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true.*