Today: (1) Intro to Detection Theory

- Neal: Potential jury duty on Tue 28th. I will swap Lecture 10 (now Tue 28th) and Exam 1 Review (now on Thu 23rd). I will email on Monday evening when I find out; if needed I will make up a lecture in the future.

- Lecture today not on Exam 1.

0.1 Exam 1

Coverage:

- HW Sets 1-4
- Lectures 1-8, and associated readings.

**Format**: Short answer, derivation, proof. Don’t need a calculator.

**Portfolio**: May have your completed homework, including any written notes you want to put on your homework solutions. Must be your own writing, not photocopied from any source. Note that Tables 3.1 and 4.1 (pages 115-116, 164-165) will be copied onto your exam booklet, so you don’t need to memorize these pmfs/pdfs.

**Advice**: Do lots and lots and lots of problems from the book. Work together in groups to work out difficulties.

1 Intro to Detection Theory

This section provides a brief framework into “Detection Theory” or “Hypothesis Testing” – these are two names for the same thing.

The framework of detection theory is philosophical: how do we know what we know? We as engineers divide the answer into two parts: (1) our prior experience and (2) our current observations. The prior experience may tell us what the probability of certain events are. But, we may not have any prior knowledge, or we may choose to ignore prior knowledge which might bias our current decision.

Two approaches to hypothesis testing
Bayes approach: Use and quantify prior knowledge. This is the standard approach in digital communication systems, which we design the transmitter, therefore we know the prior probabilities of each transmitted symbol. The Bayesian approach is appropriate and is covered here at the U in ECE 5520.

Frequentist, or Neyman-Pearson, approach: Do not use prior knowledge. Instead, constrain the probability of false alarm and maximize the probability of detection. This is more appropriate for deciding on statistical models for our data, because we approach the modeling question with data, and decide upon a model. We will cover this approach in this section.

In natural systems, we may not know a priori information about the system. Here “natural” is used loosely. For example, research questions in fields of science don’t generally have prior probabilities on their answer. Some questions that we ask are in fact either true or false, even if we cannot observe the answer directly or clearly.

For example, questions you might ask in your project:

- Is there correlation in the day-to-day changes in the time series of a particular stock or commodity price?
- Has the mean traffic level on the LAN today changed mean or variance?
- Do multipath arrivals in radio channels have a Markov property?
- Does temperature over the U.S. have a given spatial correlation function?

These are questions which can affect the design of algorithms and systems. This is why it is important to answer them, and to do so in a reliable and quantifiable way.

Today we will talk about “simple” hypotheses, that is, we know that the exact expression of the two hypotheses

- $H_0$: null hypothesis, $x \sim f_{X|H_0}(x|H_0)$
- $H_1$: alternate hypothesis, $x \sim f_{X|H_1}(x|H_1)$

Again, in the N-P approach, we will make a decision solely on the measurement $X = x$. There will be one region of $x$, that if we measure $x \in \mathcal{X}_0$, we will decide $H_0$ is true. In contrast, if we measure $x \in \mathcal{X}_1$, we will decide $H_1$ is true. The sets $\mathcal{X}_0$ and $\mathcal{X}_1$ form a partition of the entire range of $X$. We will see that we will come up with a function of $x$ and compare it to a threshold, as we will show below.

We can define two things:
• Power: The probability of detection, that is, given that \( H_1 \) is true, the probability that we detect \( H_1 \) to be true.

• False alarm: The probability of false alarm, that is, given that \( H_0 \) is true, the probability that we detect \( H_1 \) to be true. This is also called \( \alpha \) or \( P_{FA} \).

This nomenclature goes back to radar problem, in which \( H_0 \) is typically that a target is not present, and \( H_1 \) was that a target is present. Deciding \( H_1 \) is to raise an alarm, and not noticing a target is a ‘miss’.

I will present the continuous r.v. case here. A brief description of the detection procedure will help with the notation:

1. Set desired false alarm rate \( P_{FA} \).
2. Know the pdfs \( f_{X|\theta_1}(x|\theta_1) \) and \( f_{X|\theta_0}(x|\theta_0) \).
3. Determine the threshold \( \eta \).
4. Take a measurement \( X = x \).
5. Compute \( \phi^* \), your decision number (1 for \( H_1 \) and 0 for \( H_0 \)).

**Theorem:** Neyman-Pearson Lemma. The most powerful (MP) test of with false alarm rate \( P_{FA} \in [0,1] \) is a decision \( \phi^* \) of the form,

\[
\phi^* = \begin{cases} 
1, & f_{X|\theta_1}(x|\theta_1) > \eta f_{X|\theta_0}(x|\theta_0) \\
0, & f_{X|\theta_1}(x|\theta_1) < \eta f_{X|\theta_0}(x|\theta_0)
\end{cases}
\]

where \( \eta \) is selected to satisfy,

\[
E_{\phi^*|\theta_0} [\phi^*|\theta_0] = \alpha
\]

**Proof:** Not covered.

Intuition: Given \( \theta_1 \) and \( \theta_0 \), we would have had two different distributions of \( X \). When \( f_{X|\theta_1}(x|\theta_1) \) is very high, that means that \( X \) was very likely under the case of the parameter being \( \theta_1 \). On the contrary, when \( f_{X|\theta_0}(x|\theta_0) \) is very high, that means that \( X \) was very likely under the case of the parameter being \( \theta_0 \). Where should we draw the line between the two? How much more likely does \( x|\theta_1 \) need to be compared to \( x|\theta_0 \) in order to call it for \( H_1 \)? This is the threshold \( \eta \).

The test given by the Neyman-Pearson lemma is also called the most-powerful likelihood ratio test (MP-LRT).
1.1 Simplification of MP-LRT Test

This test can be rewritten as:

\[ \phi^* = \begin{cases} 
1, & \frac{f_{X|\theta_1}(x|\theta_1)}{f_{X|\theta_0}(x|\theta_0)} > \eta \\
0, & \frac{f_{X|\theta_1}(x|\theta_1)}{f_{X|\theta_0}(x|\theta_0)} < \eta 
\end{cases} \]

Or even to this shorthand:

\[ \frac{f_{X|\theta_1}(x|\theta_1)}{f_{X|\theta_0}(x|\theta_0)} \stackrel{H_1}{>}_{H_0} \eta \]

The conditional probability \( f_{X|\theta_i}(x|\theta_i) \) is also called the “likelihood”. Is it clear why it is called a likelihood ratio test (LRT)?

We’re not interested in the likelihood ratio value - only whether that ratio is above or below a threshold. So we can algebraically manipulate the expression. Common example: Take the log of both sides:

\[ \log \frac{f_{X|\theta_1}(x|\theta_1)}{f_{X|\theta_0}(x|\theta_0)} \stackrel{H_1}{>}_{H_0} \eta' \]

where \( \eta' = \log \eta \). This helps when pdfs are exponential in form. (Why can we take the log of both sides?)

The threshold is a parameter. Tuning it changes the probabilities of errors in each direction.

- If you raise \( \eta \), you make it more likely that you make the error of deciding \( H_0 \) when \( H_1 \) was actually true. (Probability of Miss)
- If you lower \( \eta \), you make it more likely that you make the error of deciding \( H_1 \) when \( H_0 \) was actually true. (Probability of False Alarm)

Example: Test for Gaussian change of mean

As a simple example, consider the detection of whether the a Gaussian r.v. \( X \) has one of two different means,

- \( H_0: X \sim \mathcal{N}(\mu_0, \sigma^2) \)
- \( H_1: X \sim \mathcal{N}(\mu_1, \sigma^2) \)

where \( \mu_1 > \mu_0 \).

From the N-P Lemma, the optimal test is the MP-LRT

\[
\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(x - \mu_1)^2}{2\sigma^2} \right] \stackrel{H_1}{\geq}\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(x - \mu_0)^2}{2\sigma^2} \right] \stackrel{H_0}{\leq} \eta
\]

\[
\exp \left( \frac{(x - \mu_0)^2 - (x - \mu_1)^2}{2\sigma^2} \right) \stackrel{H_1}{\geq}\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(x - \mu_0)^2}{2\sigma^2} \right] \stackrel{H_0}{\leq} \eta
\]
Now letting $\eta' = \log \eta$ and taking the log of both sides,

$$
(x - \mu_0)^2 - (x - \mu_1)^2 \begin{cases} 
\begin{align*}
\eta'2\sigma^2 & x \begin{array}{c} \gtrless \mu_1 \gtrless \mu_0 \\
\gtrless \mu_1 \gtrless \mu_0 
\end{array}
\end{align*}
\end{cases}
$$

where $\gamma = \frac{\eta'2\sigma^2 - \mu_0^2 + \mu_1^2}{2(\mu_1 - \mu_0)}$.

But the choice of $\eta$ was as arbitrary as the choice of $\gamma$. We don’t worry about constant multiplying or additive factors (unless they would flip the signs in the decision). We’re going to choose $\gamma$ based on the false alarm rate it causes, anyway.

What is the false alarm rate?

**Solution:**

\[
\begin{align*}
P_{FA} & = P[X > \gamma|H_0] \\
& = \int_{x=\gamma}^{\infty} \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left[ -\frac{(x - \mu_0)^2}{2\sigma^2} \right] dx \\
& = Q\left( \frac{\gamma - \mu_0}{\sigma} \right)
\end{align*}
\]

Given that we are willing to tolerate a false alarm rate of $\alpha$, how should we set the threshold $\gamma$?

\[
\begin{align*}
\alpha & = Q\left( \frac{\gamma - \mu_0}{\sigma} \right) \\
\gamma & = \sigma Q^{-1}(\alpha) + \mu_0
\end{align*}
\]

Thus for a constant false alarm rate of $\alpha$, you set $\gamma = \sigma Q^{-1}(\alpha) + \mu_0$.

Note that $\mu_1$ has no part in the decision. The other mean plays a part in the “power” or probability of detection. This is

\[
\begin{align*}
P_D & = P[X > \gamma|H_1] \\
& = Q\left( \frac{\gamma - \mu_1}{\sigma} \right)
\end{align*}
\]

But, even if you don’t know $\mu_1$, and you only know that $\mu_1 > \mu_0$, you would have still designed the same test. Only the power would have changed. In this case, the test $x \begin{cases} 
\begin{align*}
\gtrsim \sigma Q^{-1}(\alpha) + \mu_0 & x \begin{array}{c} \gtrsim \mu_1 \gtrsim \mu_0 \\
\gtrsim \mu_1 \gtrsim \mu_0 
\end{array}
\end{align*}
\end{cases}$ is called the uniformly most powerful (UMP) detection test, because it is the most powerful test for a given $P_{FA}$ for any $\mu_1 \in (\mu_0, \infty)$. The fact that there is a UMP test is a good thing, it means we only need to know the ‘typical’ ($H_0$) distribution to design the best test.
Example: Compare Exponential and Pareto

Detect whether $X$ belongs to an exponential or a Pareto distribution:

- $H_0$: $f_{X|H_0}(x|H_0) = \lambda e^{-\lambda(x-a)}$ for $x > a$ and zero otherwise.
- $H_1$: $f_{X|H_0}(x|H_0) = \frac{ka^k}{x^{k+1}}$ for $x > a$ and zero otherwise.

These two distributions are shown in Figure 1. Intuitively, when (for what values of $X$) would you decide that $X$ belonged to the Pareto distribution?

![Figure 1: Distribution of exponential ($a = 1$ and $\lambda = 1$) and Pareto ($k = 1$).](image)

What is the optimal NP test for these hypotheses?

**Solution:** The LRT is given as

$$
\frac{ka^k}{x^{k+1}\lambda e^{-\lambda(x-a)}} \frac{n_1}{n_0} > n_0 \gamma \\
\log \frac{ka^k}{\lambda} - (k + 1) \log x + \lambda(x - a) \frac{n_1}{n_0} \log \gamma \\
x - \frac{k + 1}{\lambda} \log x \frac{n_1}{n_0} > n_0 \gamma'$$

This LRT is shown in Figure 2. For $\gamma$ high enough, the test will simply be

$$x \frac{n_1}{n_0} \geq \gamma'$$

But for lower $\gamma'$ (for higher $P_{FA}$) we might also decide $H_1$ when we measure very low $X$.

**Example: Test vs. Uniform**
(From A. Hero, U. of Michigan, EECS 564 notes.) Let $Z$ be a r.v.
with values in $[-1,1]$, with density function,
\[ f_{Z|\theta}(z|\theta) = \frac{3}{2(\theta z^2 + 1)} \]
where $\theta > 0$. Note $\theta$ controls the deviation of $Z$ from the uniform density. Derive the MP-LRT of level $P_{FA} = \alpha$ for the hypothesis $H_0 : \theta = 0$, vs. $H_1 : \theta = \theta_1$, where $\theta_1$ is a fixed and known positive value.

**Solution:**
\[
\frac{f_{X|\theta_1}(x|\theta_1)}{f_{X|\theta_0}(x|\theta_0)} \stackrel{H_1}{\gtrless} \frac{\eta_{H_1}}{\eta_{H_0}} \eta
\]
\[
\frac{3}{3 + \theta_1} (\theta_1 z^2 + 1) \stackrel{H_1}{\gtrless} \frac{\eta_{H_1}}{\eta_{H_0}} \eta
\]
\[
z^2 \stackrel{H_1}{\gtrless} \frac{3\eta + \theta_1\eta - 1}{3\theta_1} = \gamma
\]
Thus the test is to compare $Z^2$ to a threshold. What is the $P_{FA}$ as a function of $\gamma$?

\[
\alpha = P_{FA} = P [\gamma < Z^2 \leq 1 | H_0] = 2 \int_{\sqrt{\gamma}}^1 \frac{1}{2} dz = 1 - \sqrt{\gamma}
\]
\[
\gamma = (1 - \alpha)^2
\]

This test and threshold is not a function of $\theta_1 > 0$, so the test is UMP.

### 1.2 Multiple Measurements

There is no limitation in detection theory to single measurements $X$. In fact, we typically want to use many measurements $X_1, \ldots, X_n$ to give ourselves more confidence in our results.
Example: Independent Gaussian Measurements

Continuing example 1,

- \( H_0: X_i \sim N(\mu_0, \sigma^2) \)
- \( H_1: X_i \sim N(\mu_1, \sigma^2) \)

for \( i = 1 \ldots n \), where measurements \( \{X_i\} \) are independent (under both models). Let \( X = [X_1, \ldots, X_n]^T \). Show that the LRT becomes a comparison of the average to a threshold.

Solution:

\[
\frac{f_{X|\theta_1}(x|\theta_1)}{f_{X|\theta_0}(x|\theta_0)} \begin{cases} H_1 \geq \eta \\
 H_0 \end{cases}
\]

\[
\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(x_i - \mu_1)^2}{2\sigma^2} \right] \begin{cases} H_1 \geq \eta \\
 H_0 \end{cases}
\]

\[
\exp \left( \sum_{i=1}^{n} (x_i - \mu_0)^2 - \sum_{i=1}^{n} (x_i - \mu_1)^2 \right) \begin{cases} H_1 \geq \gamma \\
 H_0 \end{cases}
\]

You can see that the solution becomes a test of the sum or average of the measurements \( \{x_i\} \).

\[
\sum_{i=1}^{n} x_i \begin{cases} H_1 \geq \gamma \\
 H_0 \end{cases}
\]

\[
\frac{1}{n} \sum_{i=1}^{n} x_i \begin{cases} H_1 \geq \gamma' \\
 H_0 \end{cases}
\]