Introduction

- Today we consider **parametric detection** problems for signals with one or more unknown parameters.
- We focus on the case of an $N$-sample discrete time observation $Y \in \mathcal{Y}$ with binary hypotheses

  \[
  \mathcal{H}_0 : \quad Y \sim p_x(y) \quad \text{for} \quad x \in \mathcal{X}_0
  \]

  \[
  \mathcal{H}_1 : \quad Y \sim p_x(y) \quad \text{for} \quad x \in \mathcal{X}_1
  \]

- Our approach: Absorb the unknown parameters into our notion of the state $x$ and the associated state space $\mathcal{X}$.
- In many textbooks, it is common to denote the random parameter(s) as $\Theta$ and a realization of these parameters as $\theta$. We will stay consistent with our earlier notation, however, and use $x$ here.
Example

Suppose we have a known signal $s \in \mathbb{R}^n$ and we have a communication system that transmits either $s_0 = -s$ or $s_1 = s$. The signal arrives with unknown amplitude $a > 0$ and is corrupted by zero-mean AWGN with variance $\sigma^2$. The hypotheses can be written as

\[ H_0 : Y \sim \mathcal{N}(as_0, \sigma^2 I) \]
\[ H_1 : Y \sim \mathcal{N}(as_1, \sigma^2 I) \]

for unknown $a > 0$. What is the state space $\mathcal{X}$ and what are the sets $\mathcal{X}_0$ and $\mathcal{X}_1$?

An equivalent, but more clever way to write these hypotheses is

\[ H_0 : Y \sim \mathcal{N}(as, \sigma^2 I) \text{ with } a < 0 \]
\[ H_1 : Y \sim \mathcal{N}(as, \sigma^2 I) \text{ with } a > 0 \]

for unknown $a \in \mathbb{R}\backslash 0$. What is the state space $\mathcal{X}$ and what are the sets $\mathcal{X}_0$ and $\mathcal{X}_1$?
Types of Detectors: What we Already Know

Problem: Find a decision rule to decide between

\[ \mathcal{H}_0 : Y \sim \mathcal{N}(as, \sigma^2 I) \text{ with } a < 0 \]
\[ \mathcal{H}_1 : Y \sim \mathcal{N}(as, \sigma^2 I) \text{ with } a > 0 \]

for unknown \( a \in \mathbb{R} \setminus 0 \)

- We know that these types of problems are **composite** binary hypothesis testing problems.
- Bayes decision rules involve computation of two discriminant functions (or commodity costs) \( g_0(y, \pi) \) and \( g_1(y, \pi) \) and subsequent comparison.
- Under the N-P criterion with significance level (or size) \( \alpha \):
  - Our strategy: **reduce the composite HT problem to a simple one**.
  - Check for the existence of a UMP decision rule (check the critical region and/or monotone likelihood ratio) that maximizes \( P_{D,x} \) for all \( x \in \mathcal{X}_1 \) subject to \( P_{fp,x} \leq \alpha \) for all \( x \in \mathcal{X}_0 \).
  - If a UMP rule doesn’t exist, we can usually find an LMP decision rule.
Example: Known Signal with Unknown Amplitude

We would like to decide on the presence of a known signal $s \in \mathbb{R}^n$ ($H_1$) versus the absence of the signal ($H_0$).

The known signal $s$ arrives at the detector corrupted by noise $W$ and with unknown amplitude $a$.

We observe a realization $y \in \mathbb{R}^n$ of the random variable $Y = as + W$.

- $a = 0$ when no signal is present ($H_0$)
- $a > 0$ when a signal is present ($H_1$)

What kind of hypothesis testing problem is this? Binary, composite, one-sided.

What is the state space here? $x = a \in \mathcal{X} = [0, \infty)$. 
Bayes Detector for this Example

To develop a Bayes detector for this problem, we need 3 things:

1. We need a conditional pdf or pmf statistically describing the observations \( y \in \mathcal{Y} \) for each state \( x \in \mathcal{X} \): \( p_x(y) \).
2. We need a cost assignment for each state \( x \in \mathcal{X} \) and each hypothesis \( \mathcal{H}_i \): \( C_0(x) \) and \( C_1(x) \).
3. We need a prior (pdf) on the states: \( \pi(x) \).

The Bayes risk of decision rule \( \rho \) in this case can be written as

\[
r(\rho, \pi) = \int_{\mathcal{X}} \pi(x) \int_{\mathcal{Y}} \rho^\top(y) C(x) p_x(y) \, dy \, dx
\]

\[
= \int_{\mathcal{Y}} \rho^\top(y) \left( \int_{\mathcal{X}} C(x) \pi(x) p_x(y) \, dx \right) \, dy
\]

We know how to solve this problem: For each \( y \in \mathcal{Y} \), the Bayes decision rule just compares \( g_0(y, \pi) \) to \( g_1(y, \pi) \) and sets \( \delta^{B\pi} \) equal to the index of the smaller discriminant function (or commodity cost).
Bayes Detector for this Example

Suppose the noise is distributed as $W \sim \mathcal{N}(0, \Sigma)$. Conditioned on $x = a$, we can write the density of the vector observation as

$$p_{x=a}(y) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left( -\frac{(y - as)^\top \Sigma^{-1} (y - as)}{2} \right)$$

Suppose also that we have the UCA and the prior

$$\pi(x) = \pi_0 \delta(x) + (1 - \pi_0) (\mathbb{I}_x(0) - \mathbb{I}_x(1))$$

where $\mathbb{I}_x(t) = 1$ for all $x > t$ and $\mathbb{I}_x(t) = 0$ otherwise.

Note that the state space here is $\mathcal{X} = [0, 1]$. 
Bayes Decision Rule: $\pi_0 = 0.5$ and $\bar{s} = [1.5, 0.5]^\top$
Bayes Decision Rule: $\pi_0 = 0.4$ and $\bar{s} = [1.5, 0.5]^\top$
Bayes Decision Rule: $\pi_0 = 0.75$ and $\bar{s} = [1.5, 0.5]^\top$
% Bayes examples for lecture 9
% DRB 01-March-2008
% -----------------------------------
% User variables
% -----------------------------------
pi0 = 0.75; % prior state parameter
ybartest = -1:0.1:2; % space of observations to test (transformed coordinate space)
sbar = [1.5;0.5]; % signal vector (transformed coordinate space)
% -----------------------------------
v = sbar'*sbar/2;
syms a real
discrimratio = zeros(length(ybartest),length(ybartest));

i0 = 0;
for y0 = ybartest
    i0 = i0+1;
i1 = 0;
    for y1 = ybartest
        i1 = i1+1;
        ybar = [y0;y1];
        u = sbar'*ybar;
        f=exp(a*u-a^2*v);
        discrimratio(i1,i0) = (1-pi0)/pi0 * int(f,0,1);
    end
end

contourf(ybartest,ybartest,discrimratio,[1 1]);
xlabel('ybar0'); ylabel('ybar1');
hold on; plot(sbar(1),sbar(2),'rp'); hold off; grid on; axis square
Neyman-Pearson Detector

To develop a N-P detector for this problem, we need 2 things:

1. We need a conditional pdf or pmf statistically describing the observations $y \in \mathcal{Y}$ for each state $x \in \mathcal{X}$: $p_x(y)$.

2. We need a significance-level for the test: $P_{fp} \leq \alpha$.

How should we approach the problem?

- Check to see if a UMP detector exists.
- If not, derive an LMP detector with good performance for a near zero.
- A new option: If we can specify a prior on the states, denoted as $\pi(x)$, we can reduce each composite hypothesis to a simple one by taking a weighted average of the density with respect to the prior, i.e.

$$H_j' : Y \sim p_j(y) = \int_{X_j} p_x(y) \pi_j(x) \, dx$$

where $\pi_j(x) = \frac{\pi(x)}{\text{Prob}(x \in X_j)}$ when $x \in X_j$ and is equal to zero otherwise. Then testing $H_0'$ versus $H_1'$ is a simple hypothesis testing problem for which the N-P lemma gives a unique optimal solution.
Neyman-Pearson Detector

Suppose the noise is distributed as $W \sim \mathcal{N}(0, \sigma^2 I)$. Conditioned on $x = a$, we can write the density of the vector observation as

$$p_{x=a}(y) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{(y - as)^\top (y - as)}{2\sigma^2}\right)$$

$$= \prod_{k=0}^{n-1} p_{x=a}(y_k)$$

$$= \prod_{k=0}^{n-1} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(y_k - as_k)^2}{2\sigma^2}\right)$$

Note that any results we derive for this case can also be applied to the case when $W \sim \mathcal{N}(0, \Sigma)$ by using our coordinate-transformation (decorrelation) trick.
Let’s set up a simple HT test: $H_0 : a = 0$ versus $H_1 : a = a_1 > 0$. The N-P lemma says that, for this simple HT problem, we can find an $\alpha$-level decision rule of the form

$$
\rho(y) = \begin{cases} 
1 & \ln (L(y)) > \ln v \\
\gamma & \ln (L(y)) = \ln v \\
0 & \ln (L(y)) < \ln v 
\end{cases}
$$

where $\ln (L(y)) := \ln \left( \frac{p_1(y)}{p_0(y)} \right)$ with $v \geq 0$ and $\gamma \in [0, 1]$ chosen such that $P_{fp} = \alpha$. 
Log-Likelihood Ratio

The log-likelihood ratio for this simple HT test:

\[
\ln (L(y)) = \ln \left( \frac{p_1(y)}{p_0(y)} \right)
\]

\[
= \ln \left( \frac{\prod_{k=0}^{n-1} \exp \left( -\frac{(y_k-a_1 s_k)^2}{2\sigma^2} \right)}{\prod_{k=0}^{n-1} \exp \left( -\frac{y_k^2}{2\sigma^2} \right)} \right)
\]

\[
= \ln \left( \prod_{k=0}^{n-1} \exp \left( \frac{2y_k a_1 s_k - a_1^2 s_k^2}{2\sigma^2} \right) \right)
\]

\[
= \sum_{k=0}^{n-1} \frac{a_1}{\sigma^2} \left( y_k s_k - \frac{a_1 s_k^2}{2} \right)
\]

\[
= \frac{a_1}{\sigma^2} \left( s^\top y - \frac{a_1 \|s\|^2}{2} \right)
\]
False Positive Probability

\[ P_{fp} = \text{Prob} \left( \ln (L(Y)) > \ln v \mid a = 0 \right) + \gamma \text{Prob} \left( \ln (L(Y)) = \ln v \mid a = 0 \right) \]

\[ = \text{Prob} \left( s^\top Y > \frac{\sigma^2}{a_1} \ln v + \frac{a_1 \|s\|^2}{2} \mid a = 0 \right) \]

How is \( s^\top Y \) distributed when \( a = 0 \)?

Hence, \( v \) must be selected such that

\[ P_{fp} = Q \left( \frac{\sigma^2}{a_1} \ln v + \frac{a_1 \|s\|^2}{2} \mid \|s\| \sigma \right) = \alpha \]
Neyman-Pearson Detector: UMP Decision Rule

Does a UMP decision rule exist? To answer this question, we need to determine if the critical region \( \Gamma_1 = \{ y \in \mathcal{Y} : \rho(y) \text{ decides } \mathcal{H}_1 \} \) depends on our choice of \( a_1 \).

We decide \( \mathcal{H}_1 \) for sure when

\[
s^\top y > \frac{\sigma^2}{a_1} \ln v + \frac{a_1\|s\|^2}{2}
\]

and we decide \( \mathcal{H}_1 \) with probability \( \gamma \) when

\[
s^\top y = \frac{\sigma^2}{a_1} \ln v + \frac{a_1\|s\|^2}{2}
\]

(which happens with probability zero).

Critical region depends on \( a_1 \Rightarrow \) no UMP decision rule.

What should we do now?
Held on. Does the critical region \( \Gamma_1 = \{ y \in \mathbb{R}^n : s^\top y > \frac{\sigma^2}{a_1} \ln v + \frac{a_1 \|s\|^2}{2} \} \) really depend on \( a_1 \)?

Given \( 0 \leq \alpha \leq 1 \) and \( t > 0 \), the unique solution to \( Q(z/t) = \alpha \) is \( z = tQ^{-1}(\alpha) \). Hence, the unique solution to

\[
Q \left( \frac{\sigma^2}{a_1} \ln v + \frac{a_1 \|s\|^2}{2} \right) = \alpha
\]

is

\[
\frac{\sigma^2}{a_1} \ln v + \frac{a_1 \|s\|^2}{2} = \|s\| \sigma Q^{-1}(\alpha).
\]

Hence the critical region for a significance-level \( \alpha \) N-P decision rule can be written as

\[
\Gamma_1 = \{ y \in \mathbb{R}^n : s^\top y > \|s\| \sigma Q^{-1}(\alpha) \}\]

Does this depend on \( a_1 \)? Does a UMP decision rule exist?
Neyman-Pearson Detector: UMP Decision Rule

\[ \rho_{\text{UMP}}(y) = \begin{cases} 
1 & s^\top y > \|s\|\sigma Q^{-1}(\alpha) \\
\gamma & s^\top y = \|s\|\sigma Q^{-1}(\alpha) \\
0 & s^\top y < \|s\|\sigma Q^{-1}(\alpha) \end{cases} \]

How would this change if the noise was distributed as \( W \sim \mathcal{N}(0, \Sigma) \)?
Neyman-Pearson Detector: LMP Decision Rule

If the UMP detector did not exist or was too complicated, we could find an LMP detector for this example by comparing

$$\frac{d}{da} L_a(y) \bigg|_{a=0}$$

to a threshold. When the noise samples are i.i.d., the likelihood ratio can be written as

$$L_a(y) = \frac{p_{x=a}(y)}{p_{x=0}(y)} = \prod_{k=0}^{n-1} \frac{q(y_k - a s_k)}{q(y_k)}$$

where $q(x)$ is the pmf/pdf of the $k$th noise sample. In our example, $q(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-x^2/\sigma^2}$. We’ll continue our analysis here for i.i.d. noise with general distribution $q(x)$...
Neyman-Pearson Detector: LMP Decision Rule

Taking the derivative of $L_a(y)$ with respect to $a$ yields

$$\frac{d}{da} L_a(y) = - \sum_{k=0}^{n-1} s_k \frac{q'(y_k - a s_k)}{q(y_k)} \prod_{j \neq k} \frac{q(y_j - a s_j)}{q(y_j)}$$

where $q'(x) = \frac{d}{dx} q(x)$. Setting $a = 0$ yields

$$\frac{d}{da} L_a(y) |_{a=0} = - \sum_{k=0}^{n-1} s_k \frac{q'(y_k)}{q(y_k)} \prod_{j \neq k} \frac{q(y_j)}{q(y_j)}$$

$$= - \sum_{k=0}^{n-1} s_k \frac{q'(y_k)}{q(y_k)}$$

$$= \sum_{k=0}^{n-1} s_k h^{\text{LMP}}(y_k)$$
The locally most powerful decision rule then takes the form

\[ \rho_{LMP}(y) = \begin{cases} 
1 & \sum_{k=0}^{n-1} s_k h_{LMP}(y_k) > \tau \\
\gamma & \sum_{k=0}^{n-1} s_k h_{LMP}(y_k) = \tau \\
0 & \sum_{k=0}^{n-1} s_k h_{LMP}(y_k) < \tau 
\end{cases} \]

where \( \gamma \) and \( \tau \) are selected such that \( P_{fp} = \alpha \).

In our example, \( h_{LMP}(y_k) := -\frac{q'(y_k)}{q(y_k)} = \frac{2y_k}{\sigma^2} \). Hence, if \( \tau' = \tau \sigma^2 / 2 \), the LMP decision rule then takes the form

\[ \rho_{LMP}(y) = \begin{cases} 
1 & s^\top y > \tau' \\
\gamma & s^\top y = \tau' \\
0 & s^\top y < \tau' 
\end{cases} \]

with \( \tau' \) and \( \gamma \) selected so that \( P_{fp} = \alpha \). How does this compare to the UMP decision rule? \( \rho_{LMP}(y) = \rho_{UMP}(y) \) here.
For $b > 0$, the Laplacian density is given as $q(x) = \frac{b}{2}e^{-b|x|}$.

For Laplacian noise,

$$h_{\text{LMP}}(y_k) = -\frac{q'(y_k)}{q(y_k)} = b\text{sgn}(x).$$

Hence, the locally most powerful decision rule can be written as

$$\rho_{\text{LMP}}(y) = \begin{cases} 
1 & \sum_{k=0}^{n-1} s_k \text{sgn}(y_k) > \tau \\
\gamma & \sum_{k=0}^{n-1} s_k \text{sgn}(y_k) = \tau \\
0 & \sum_{k=0}^{n-1} s_k \text{sgn}(y_k) < \tau 
\end{cases}$$

In this case, $\rho_{\text{LMP}}(y) \neq \rho_{\text{UMP}}(y)$. 
The Third Alternative: Average Density Over a Prior

The approach followed in pages 64-65 of your textbook (as well as Example III.B.5) is to reduce the composite HT problem to a simple HT problem by using the a prior $\pi(x)$ to compute a single density as a weighted average of the family of densities associated with $\mathcal{H}_j$, i.e.

$$p_j(y) = \int_{\mathcal{X}_j} p_x(y) \pi_j(x) \, dx$$

where $\pi_j(x) = \frac{\pi(x)}{\text{Prob}(x \in \mathcal{X}_j)}$ when $x \in \mathcal{X}_j$ and is equal to zero otherwise. In this case, given a significance level $\alpha$, the N-P lemma tells us that a unique optimal solution must exist based on a threshold test of the likelihood ratio

$$L(y) = \frac{p_1(y)}{p_0(y)} = \frac{\int_{\mathcal{X}_1} p_x(y) \pi_1(x) \, dx}{\int_{\mathcal{X}_0} p_x(y) \pi_0(x) \, dx}.$$
Remarks on the Third Alternative

If there is some uncertainty as to the prior, the usual approach is to choose $\pi(x)$ such that it gives as little information about $x$ as possible, i.e. such that the simple HT problem has “maximum difficulty”.

Example III.B.5: Noncoherent Detection of a Modulated Sinusoidal Carrier. In this example, the unknown parameter is the phase of the received signal. What prior on this phase would give the least information?

$$\pi(x) \sim \mathcal{U}(0, 2\pi)$$

This is exactly the prior used in the example in your textbook.

Please read this example and, in particular, note the development of the “catalyst” for deciding between two different (but known) signals with unknown phase on page 71.
Conclusions

- Detection of known signals in noise: **simple hypothesis testing**.
- Detection of signals with one or more unknown parameters in noise: **composite hypothesis testing**.
- You are encouraged to at least skim the section on “Stochastic Signals” in section III.B (up to page 81).
- I also plan to discuss section III.D “Sequential Detection” in the optional lecture after the final exam.
Midterm Exam: What You Need to Know

- Different types of hypothesis testing problems (binary, $M$-ary, simple, composite)
- Mathematical model of hypothesis testing problems.
- Intuition about “good” and “bad” decision rules.
- Poor textbook chapter II (whole chapter):
  - Bayesian hypothesis testing (binary and $M$-ary, simple and composite)
  - Minimax hypothesis testing (binary, simple)
  - Neyman-Pearson hypothesis testing (binary, simple and composite)
- Poor textbook chapter III (up to page 72):
  - Decorrelation of signals observed in correlated Gaussian noise.
  - Detection of known discrete-time signals (binary and $M$-ary)
  - Detection of discrete-time signals with random parameters (binary)