ESE 524
Detection and Estimation Theory

Joseph A. O’Sullivan
Samuel C. Sachs Professor
Electronic Systems and Signals
Research Laboratory
Electrical and Systems Engineering
Washington University
211 Urbauer Hall
314-935-4173 (Lynda answers)
jao@wustl.edu
Detection Theory

- Which model best fits the measured data?
- Decision or detection or hypothesis testing?
- Principled approach
  - Define the concept of a model
  - Define the concept of a decision
  - Define an objective function to be minimized
  - Given these, derive an optimal decision
  - Quantify the performance (as a function of parameters)
Example: Throwing Darts

- Suppose that when people throw darts at a blackboard, there is a distribution of the locations of the darts.
- The first individual is very good and his darts land in the vicinity of the target, with a circularly symmetric distribution.
- The second individual has a systematic error that he does not seem to be able to correct. His darts land in a circularly symmetric distribution offset from the truth.
- The third individual has no systematic error, but his darts, while still landing according to a circularly symmetric distribution around the truth, are more widespread than the first individuals.
- Problem 1: Individual 1 or 2
- Problem 2: Individual 1 or 3
- Other problems
Example: Throwing Darts

- These problems are still not well posed. Need additional assumptions such as:
  - A1: The distributions are two-dimensional Gaussian distributions around the truth with known covariance matrix.
  - A2: All throws are independent of all other throws.
  - A3: The covariance matrix is diagonal with equal variance in each direction.
function errf=dartslecture2
errf=0;
x=randn(2,100);
y=2*ones(2,100)+randn(2,100);
z=3*randn(2,100);
figure
plot(x(1,:),x(2,:),'bx')
hold on
plot(y(1,:),y(2,:),'ro')
plot(z(1,:),z(2,:),'gv')
axis equal;
figure
plot(x(1,:),x(2,:),'bx')
hold on
plot(2+y(1,:),2+y(2,:),'ro')
axis equal;
errf=1;

Can decide among hypotheses if more “throws” (data) are observed.
Probability Density Functions: Matlab Mesh Plots
function errf=dartslecture2pdf
errf=0;
x=-5:.1:5;
[x,y]=meshgrid(x);
p1=exp(-(x.^2+y.^2)/2)/(2*pi);
p2=exp(-(x-2).^2+(y-2).^2)/2)/(2*pi);
p3=exp(-(x.^2+y.^2)/18)/(18*pi);
figure
mesh(x,y,p1)
figure
mesh(x,y,p2)
figure
mesh(x,y,p3)
figure
mesh(x,y,max(max(p1,p2),p3))
errf=1;
Example: Throwing (One Dim) Darts

- Suppose that one of two possible known signals is transmitted during the interval $[0,T]$:
  - $s(t)$ or $-s(t)$
- The signal is measured in white Gaussian noise:
  \[ r(t) = a\sqrt{E}s(t) + w(t), \quad 0 \leq t \leq T \]
  \[ a \in \{+1, -1\} \]
- The receiver first computes the integral of $r(t)$ times $s(t)$ (unit energy) over the interval to get
  \[ r = a\sqrt{E} + n \]
  \[ n \sim \mathcal{N}(0, N_0/2) \]
- Problem: $a=+1$ or $a=-1$
Example: Throwing 2D Darts

- In QAM (quadrature amplitude modulation), one of four signals is sent over \([0,T]\):

\[
s(t) \in \left\{ \sqrt{\frac{2}{T}} \cos(\omega_c t), -\sqrt{\frac{2}{T}} \cos(\omega_c t), \sqrt{\frac{2}{T}} \sin(\omega_c t), -\sqrt{\frac{2}{T}} \cos(\omega_c t) \right\}
\]

- Assume measurements are in white Gaussian noise

- Assuming the cosine and sine are orthogonal over the interval, the measurements are integrated against each to get two measurements.

- Problem: which of four signals was sent?

\[
r(t) = \sqrt{E}s(t) + w(t), \ 0 \leq t \leq T
\]

\[
r_I = \int_0^T r(t) \frac{2}{T} \cos(\omega_c t) dt = \sqrt{E} \int_0^T s(t) \frac{2}{T} \cos(\omega_c t) dt + n_I
\]

\[
r_Q = \int_0^T r(t) \frac{2}{T} \sin(\omega_c t) dt = \sqrt{E} \int_0^T s(t) \frac{2}{T} \sin(\omega_c t) dt + n_Q
\]

\[
E \begin{bmatrix} r_I \\ r_Q \end{bmatrix} \in \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\}
\]

\[
n_I \sim \mathcal{N}(0, N_0 / 2)
\]

\[
n_Q \sim \mathcal{N}(0, N_0 / 2)
\]

\[
n_I \perp n_Q
\]
Finite Dimensional
Binary Detection Theory

- **Problem:** Given a measurement of a random vector $r$, decide which of two models it is drawn from.

- Source produces $H_0$ or $H_1$ probabilities $P_0$ and $P_1 = 1 - P_0$. Probabilities may not be known.
- Random measurement vector results. Transition probabilities (system) are often assumed known.
- Decision rule is a partition of measurement space into two sets $Z_0$ and $Z_1$ corresponding to decisions $H_0$ and $H_1$. 

---

J. A. O'S. ESE 524, Lecture 2, 01/15/09
Outcomes, Bayes Criterion

- There are four possible outcomes:
  - Decide $H_0$ and $H_1$ is true: $P_1(1-P_D)$
  - Decide $H_1$ and $H_1$ is true: $P_1P_D$
  - Decide $H_0$ and $H_0$ is true: $P_0(1-P_F)$
  - Decide $H_1$ and $H_0$ is true: $P_0P_F$

- Probability of False Alarm: $P_F$
  - Decide $H_1$ given $H_0$ is true

- Probability of Detection: $P_D$
  - Decide $H_1$ given $H_1$ is true

- Probability of Miss: $P_M=1-P_D$
  - Decide $H_0$ given $H_1$ is true

- Error probabilities equal integrals of conditional (or transition) probability density functions over decision regions
Bayes Criterion

- **Assumptions:**
  - Source prior probabilities $P_0$ and $P_1$ are known
  - Transition densities are known
  - There are costs of decision outcomes and these are known: $C_{ij}$ cost of deciding $H_i$ and $H_j$ is true
    - $C_{01} - C_{11} > 0$ is the relative cost of a miss
    - $C_{10} - C_{00} > 0$ is the relative cost of a false alarm

- **Bayes decision criterion:** Pick the decision rule to minimize the average risk (cost)

$$R = C_{00}P_0(1 - P_F) + C_{01}P_1(1 - P_D) + C_{10}P_0P_F + C_{11}P_1P_D$$

$$= (C_{01} - C_{11})P_1(1 - P_D) + (C_{10} - C_{00})P_0P_F + C_{00}P_0 + C_{11}P_1$$

$$= C_M P_1(1 - P_D) + C_F P_0 P_F \quad \leftarrow \text{Simplified notation}$$

$$+ C_{00}P_0 + C_{11}P_1$$
Results

- Theorem: Likelihood ratio test is optimal
- Corrolary: The log-likelihood ratio test is optimal
- For proofs, see notes and text
- Example: deterministic signal in Gaussian noise
## Decision Rules

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Transition Densities</th>
<th>Priors $P_0 &amp; P_1$</th>
<th>Costs $C_{ij}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayes: minimize expected risk or cost</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Minimum Probability of Error</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Minimax</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Neyman-Pearson</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

The entire space is divided into two regions: Decide $H_0$ or $H_1$

The integrals of the probability density functions determine the probabilities of error

Integrate over the “other” or “wrong” region

$$\mathcal{R} = C_M P_M P_1 + C_F P_F P_0$$

$$P_M = \int_{Z_0}^{} p_{r|H_1}(R \mid H_1) dR$$

$$P_F = \int_{Z_1}^{} p_{r|H_0}(R \mid H_0) dR$$
Bayes Criterion

- Define the two decision regions. Minimize Bayes risk over the decision regions.
- Every point in the space must be included in one integral or the other.
- Put a point in a region if it contributes less to that integral than to the other.
- Introduce new notation for the result of the comparison

\[ \mathcal{R} = C_M P_M P_1 + C_F P_F P_0 \]

\[ = C_M P_1 \int_{Z_0} p_{r|H_1}(\mathbf{R} \mid H_1) d\mathbf{R} + C_F P_0 \int_{Z_1} p_{r|H_0}(\mathbf{R} \mid H_0) d\mathbf{R} \]

\[ C_M P_1 p_{r|H_1}(\mathbf{R} \mid H_1) \quad C_F P_0 p_{r|H_0}(\mathbf{R} \mid H_0) \]
Bayes Criterion

- Compare $C_M P_1 p_{r|H_1}(R \mid H_1)$ and $C_F P_0 p_{r|H_0}(R \mid H_0)$
- Introduce new notation for the result of the comparison
  - If the left side is bigger, decide $H_1$
  - If the right side is bigger, decide $H_0$
  - For continuous pdfs, do not worry about equality

\[
R = C_M P_M P_1 + C_F P_F P_0 \\
= C_M P_1 \int_{Z_0} p_{r|H_1}(R \mid H_1) dR + C_F P_0 \int_{Z_1} p_{r|H_0}(R \mid H_0) dR
\]

\[
\begin{array}{c|c|c}
  & H_1 & \triangleleft \\
  \triangleright & H_0 & \end{array}
\]

\[
C_M P_1 p_{r|H_1}(R \mid H_1) \quad C_F P_0 p_{r|H_0}(R \mid H_0)
\]
Bayes Criterion: Derivation of Likelihood Ratio Test

- Compare \( C_M p_{r|H_1}(R \mid H_1) \) and \( C_F p_{r|H_0}(R \mid H_0) \)
- There are several equivalent ways to compare these values including the likelihood ratio and the log-likelihood ratio

\[
\begin{align*}
\Lambda(R) &\triangleq \frac{p_{r|H_1}(R \mid H_1)}{p_{r|H_0}(R \mid H_0)} \\
&\overset{H_1}{>} \frac{C_F P_0}{C_M P_1} \overset{\eta}{\triangleq} \eta \\
l(R) &\triangleq \ln \left( \frac{p_{r|H_1}(R \mid H_1)}{p_{r|H_0}(R \mid H_0)} \right) \\
&\overset{H_1}{>} \ln \left( \frac{C_F P_0}{C_M P_1} \right) \overset{\gamma}{\triangleq} \gamma
\end{align*}
\]
The likelihood ratio is a nonnegative-valued function of the observed data.

Evaluated at a random realization of the data, the likelihood ratio is a random variable.

Comparing the likelihood ratio to a threshold is the optimal test.

The probabilities of error can be computed in terms of the probability density functions for either the likelihood ratio or the log-likelihood ratio.
Optimal Decision Rules

- Likelihood Ratio Test
- Threshold for Bayes Rule:
  \[ \eta = \frac{C_F P_0}{C_M P_1} = \frac{(C_{10} - C_{00}) P_0}{(C_{01} - C_{11}) P_1} \]
- Computations of the error probabilities

\[
\Lambda(R) = \frac{p_{r|H_1}(R | H_1)}{p_{r|H_0}(R | H_0)} \begin{cases} H_1 & \text{if } \Lambda > \eta \\ H_0 & \text{if } \Lambda < \eta \end{cases}
\]

\[
P_M = \int_0^\eta p(\Lambda | H_1) d\Lambda
\]

\[
P_F = \int_\eta^\infty p(\Lambda | H_0) d\Lambda
\]

\[
l(R) = \ln \left[ \frac{p_{r|H_1}(R | H_1)}{p_{r|H_0}(R | H_0)} \right] \begin{cases} H_1 & \text{if } l > \ln \eta \\ H_0 & \text{if } l < \ln \eta \end{cases}
\]

\[
P_M = \int_{-\infty}^\gamma p_{l|H_1}(L | H_1) dL
\]

\[
P_F = \int_{\gamma}^{\infty} p_{l|H_0}(L | H_0) dL
\]
Minimum Probability of Error

- Minimum Probability of Error
  \[ P_e = P_M P_1 + P_F P_0 \]
  \[ C_M = C_F = 1 \]
  \[ \eta = \frac{P_0}{P_1} \]

- Loglikelihood Ratio Test
Minimum Probability of Error

Alternative View of Decision Rule:

- Compute Posterior Probabilities on Hypotheses
- Select Most Likely Hypothesis

\[
p_{r|H_1}(R \mid H_1) > P_0 \\
p_{r|H_0}(R \mid H_0) < P_1
\]

\[
\frac{p_{r|H_1}(R \mid H_1)P_1}{g(R)} > \frac{p_{r|H_0}(R \mid H_0)P_0}{g(R)}
\]

\[
p(R) < \frac{p_{r|H_0}(R \mid H_0)P_0}{p(R)}
\]

\[
p(R) = p_{r|H_1}(R \mid H_1)P_1 + p_{r|H_0}(R \mid H_0)P_0
\]

\[
\frac{P(H_1 \mid R)}{P(H_0 \mid R)} > \frac{P(H_0 \mid R)}{P(H_1 \mid R)}
\]

for any \( g(R) > 0 \)
Gaussian Example
Minimax Decision Rule

- Find the decision rule that minimizes the maximum Bayes Risk over all possible priors.

\[ \min_{Z_1} \max_{P_1} R \]

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Transition Densities</th>
<th>Priors ( P_0 ) &amp; ( P_1 )</th>
<th>Costs ( C_{ij} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayes: minimize expected risk or cost</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Minimum Probability of Error</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Minimax</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
</tr>
<tr>
<td>Neyman-Pearson</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
Minimax Decision Rule: Analysis

- For any fixed decision rule the risk is linear in $P_1$
- The maximum over $P_1$ is achieved at an end point
- To make that end point as low as possible, the risk should be constant with respect to $P_1$
- To minimize that constant value, the risk should achieve the minimum risk at some $P_1^*$.
- At that value of the prior, the best decision rule is a likelihood ratio test
  \[ R = C_M P_M P_1 + C_F P_F P_0 \]
  \[ P_M = \int_{Z_0} p_{r|H_1}(\mathbf{R} \mid H_1) d\mathbf{R} \]
  \[ P_F = \int_{Z_1} p_{r|H_0}(\mathbf{R} \mid H_0) d\mathbf{R} \]
Minimax Decision Rule: Analysis

- Bayes risk is concave (it is always below its tangent)
- Minimax is achieved at an end point or at an interior point on the Bayes risk curve where the tangent is zero
Minimax Decision Rule:

Example

\[ H_0 : x \sim 5e^{-5x}, X \geq 0 \]
\[ H_1 : x \sim e^{-x}, X \geq 0 \]
\[ l = 4X - \ln 5 \]
\[ P_F = \int_{y'}^{\infty} 5e^{-5x} \, dx = e^{-5y'} \]
\[ P_M = \int_{0}^{\gamma'} e^{-x} \, dx = 1 - e^{-\gamma'} \]

**Matlab Code**

```matlab
pf=0:0.01:1;
pd=pf.^0.2;
etap=0.2*(pf.^(-0.8)); % eta=dP_D/dP_F
figure
plot(pf,pd);xlabel('P_F');ylabel('P_D')

cm=10;cf=100;
p1star=1./(1+cm*eta/cf);
riskoptimal=cm*(1-pd).*p1star+cf*pf.*(1-p1star);
figure
plot(p1star,riskoptimal,'b'), hold on
p1=0:0.01:1;
r1=cm*(1-pd(10))*p1+cf*pf(10)*(1-p1);
plot(p1,r1,'r'), hold on
r2=cm*(1-pd(20))*p1+cf*pf(20)*(1-p1);
plot(p1,r2,'g'), hold on
r3=cm*(1-pd(30))*p1+cf*pf(30)*(1-p1);
plot(p1,r3,'c')
xlabel('P_1');ylabel('Risk')
```
Neyman-Pearson Decision Rule

\[ F = P_M + \eta(P_F - \alpha) \]

- Minimize \( P_M \) subject to \( P_F \leq \alpha \)
- Variational approach
- Upper bound usually achieved
- Likelihood ratio test; threshold?

\[
= \int_{Z_0}^{Z_1} p_{r|H_1}(\mathbf{R} \mid H_1) d\mathbf{R} + \eta \left( \int_{Z_0}^{Z_1} p_{r|H_0}(\mathbf{R} \mid H_0) d\mathbf{R} - \alpha \right)
\]

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Transition Densities</th>
<th>Priors ( P_0 ) &amp; ( P_1 )</th>
<th>Costs ( C_{ij} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayes: minimize expected risk or cost</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Minimum Probability of Error</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Minimax</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Neyman-Pearson</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
Neyman-Pearson Decision Rule

Minimize $P_M$ subject to $P_F \leq \alpha$

Variational approach

Upper bound usually achieved

Likelihood ratio test; threshold?

Plot the ROC: $P_D$ versus $P_F$ for the family of likelihood ratio tests

Draw a vertical line where $P_F = \alpha$

Find the corresponding $P_D$

At that point, the threshold equals the derivative of the ROC

$$F = P_M + \eta(P_F - \alpha)$$

$$= \int_{Z_0} p_{r|H_1}(\mathbf{R} | H_1) d\mathbf{R} + \eta \left( \int_{Z_1} p_{r|H_0}(\mathbf{R} | H_0) d\mathbf{R} - \alpha \right)$$

$$\eta = \frac{dP_D}{dP_F}$$

$$= \frac{dP_D}{d\eta} / \frac{dP_F}{d\eta}$$
Summary

- Several decision rules
- Likelihood (and loglikelihood) ratio test is optimal
- Receiver operating characteristic (ROC) plots probability of detection versus probability of false alarm with the threshold as a parameter → all possible optimal performance
- Neyman-Pearson is a point on the ROC \( (P_F = \alpha) \)
- Minimax is a point on the ROC \( (P_F C_F = P_M C_M) \)
- Probability of error is a point on the ROC (slope \( \eta = (1-P_1)/P_1 \) )