Session 11

Recall...

11.1

GLRT-continued

When we carry out this procedure, we get a GLRT of the form

$$L_g(\underline{X}) = \frac{\max_{\underline{\theta} \in \Theta_1} f_{\underline{\theta}}(\underline{X})}{\max_{\underline{\theta} \in \Theta_0} f_{\underline{\theta}}(\underline{X})} \underset{H_0}{\overset{H_1}{\gtrless}} L_0,$$

which yields a statistical test of the form

$$\phi(\underline{X}) = \begin{cases} 1, & \text{for } L_g(\underline{X}) > L_0, \\ \gamma, & \text{for } L_g(\underline{X}) = L_0, \\ 0, & \text{for } L_g(\underline{X}) < L_0, \end{cases}$$

where, in principle, L_0 and γ are selected to yield a size α test.

Recall...

L_0 and γ

$$\phi(\underline{X}) = \begin{cases} 1, & \text{for } L_g(\underline{X}) > L_0, \\ \gamma, & \text{for } L_g(\underline{X}) = L_0, \\ 0, & \text{for } L_g(\underline{X}) < L_0. \end{cases}$$

Selecting L_0 and γ to yield a size α test is difficult.

The reason is that the size of the test is still defined as

$$\alpha = \sup_{\underline{\theta} \in \Theta_0} \mathrm{E}_{\underline{\theta}} \left[\phi(\underline{X}) \right].$$

We do not use $\mathrm{E}_{\hat{\underline{\theta}}_0} [\phi(\underline{X})]$ as the size of the test.

Example: *Suppose we wish to test the hypotheses H_0 versus H_1 that the random sample $\underline{X} = (X_1, \ldots, X_N)$ comes from a density*

$$f_\theta(x_n) = \frac{1}{\sqrt{2\pi}} \exp\left\{ \frac{-(x_n - \theta)^2}{2} \right\},$$

where under $H_0 : \Theta_0 = [-1, 1]$, and under $H_1 : \Theta_0 = \{\theta \in \mathbf{R} : |\theta| > 1\}$.

We note that in general,

$$f_\theta(\underline{X}) = \frac{1}{(2\pi)^{N/2}} \exp\left\{ -\frac{1}{2} \sum_{n=1}^{N} (x_n - \theta)^2 \right\}, \tag{1}$$

from which it follows that the (unconstrained) maximum likelihood estimate $\hat{\theta}_{ML}$ of θ can be found by solving

$$\frac{\partial}{\partial \theta} f_\theta(\underline{X}) = 0$$

for θ, yielding

$$\hat{\theta}_{ML} = \frac{1}{N} \sum_{n=1}^{N} X_n.$$

For this case, we can easily see that

$$\hat{\theta}_0 = \arg\max_{\underline{\theta}\in\Theta_0} f_{\underline{\theta}}(\underline{X}) = \begin{cases} \hat{\theta}_{ML}, & \text{for } \hat{\theta}_{ML} \in [-1,1], \\ -1, & \text{for } \hat{\theta}_{ML} < -1, \\ 1, & \text{for } \hat{\theta}_{ML} > 1, \end{cases}$$

and

$$\hat{\theta}_1 = \arg\max_{\underline{\theta}\in\Theta_1} f_{\underline{\theta}}(\underline{X}) = \begin{cases} \hat{\theta}_{ML}, & \text{for } \hat{\theta}_{ML} \notin [-1,1], \\ -1, & \text{for } \hat{\theta}_{ML} \in [-1,0], \\ 1, & \text{for } \hat{\theta}_{ML} \in (0,1]. \end{cases}$$

Using this $\hat{\theta}_0$ *and* $\hat{\theta}_1$ *we can now construct the GLRT*

$$L_g(\underline{X}) = \frac{f_{\hat{\theta}_1}(\underline{X})}{f_{\hat{\theta}_0}(\underline{X})} \underset{H_0}{\overset{H_1}{\gtrless}} L_0$$

with an appropriately chosen threshold L_0.

Recall: $\alpha = \sup_{\underline{\theta}\in\Theta_0} \mathrm{E}_{\underline{\theta}}\left[\phi(\underline{X})\right]; \quad \alpha \neq \mathrm{E}_{\hat{\underline{\theta}}_0}\left[\phi(\underline{X})\right].$

Bayesian Detection Theory

In *classical* detection, we decide between $H_0 : \underline{\theta} \in \Theta_0$ versus $H_1 : \underline{\theta} \in \Theta_1$ based on an observation $\underline{X}$ governed by the parameterized cdf $F_{\underline{\theta}}(\underline{x})$

Here $\underline{\theta}$ was assumed to be an *unknown* but fixed parameter. The parameter $\underline{\theta}$ was *not* assumed to be random, just unknown.

In the Bayesian detection framework, we again assume that our observation $\underline{X}$ is governed by a distribution $F_{\underline{\theta}}(\underline{X})$,

but we now assume that $\underline{\theta}$ is a random vector taking on one of two possible values: $\underline{\theta}_0$ or $\underline{\theta}_1$. These values are taken on with probabilities

$$\begin{aligned} p_0 &= P(\{\underline{\theta} = \underline{\theta}_0\}), \\ p_1 &= P(\{\underline{\theta} = \underline{\theta}_1\}) = 1 - p_0, \end{aligned}$$

where $p_0 \in [0,1]$.

We have a random experiment with probability space $(\mathcal{S}, \mathcal{F}, P)$ having two random variables $\underline{\theta}$ and $\underline{X}$ defined on it.

When the random experiment is performed, $\underline{\theta}$ takes on a value from the set $\{\underline{\theta}_0, \underline{\theta}_1\}$ with probabilities p_0 and p_1, respectively.

The observed value of $\underline{X}$ takes on a value consistent with the distribution $F_{\underline{\theta}}(\underline{x})$ for the value $\underline{\theta}$ takes on.

Thus we have a conditional distribution function for $\underline{X}$:

$$F(\underline{x}|\underline{\theta}_j) = P(\{\underline{X} \leq \underline{x}\}|\{\underline{\theta} = \underline{\theta}_j\}) = F_{\underline{\theta}_j}(\underline{x}), \quad j = 0, 1.$$

The joint distribution of $\underline{\theta}$ and $\underline{X}$ is given by

$$F(\underline{\theta}, \underline{x}) = F(\underline{x}|\underline{\theta})P(\underline{\theta}) = \begin{cases} p_0 F(\underline{x}|\underline{\theta}_0), & \text{for } \underline{\theta} = \underline{\theta}_0, \\ p_1 F(\underline{x}|\underline{\theta}_1), & \text{for } \underline{\theta} = \underline{\theta}_1. \end{cases}$$

It follows that

$$P(\underline{\theta}) = \int_{\mathbf{R}^n} dF(\underline{\theta}, \underline{x}) = \int_{\mathbf{R}^n} f(\underline{\theta}, \underline{x})\, d\underline{x} = \begin{cases} p_0, & \text{for } \underline{\theta} = \underline{\theta}_0, \\ p_1, & \text{for } \underline{\theta} = \underline{\theta}_1, \end{cases}$$

and

$$\begin{aligned} F(\underline{x}) &= F(\underline{x}|\underline{\theta}_0)P(\underline{\theta}_0) + F(\underline{x}|\underline{\theta}_1)P(\underline{\theta}_1) \\ &= p_0 F(\underline{x}|\underline{\theta}_0) + p_1 F(\underline{x}|\underline{\theta}_1), \end{aligned}$$

where $\mathbf{R}^n$ is the observation space of the n-dimensional observation vector.

Bayes Risk for Hypothesis Testing

Assume we have a decision rule of the form

$$\phi(\underline{x}) = \begin{cases} 1, & \text{for } \underline{x} \in S_1, \\ 0, & \text{for } \underline{x} \in S_0, \end{cases}$$

for deciding between $H_0 : \underline{\theta} = \underline{\theta}_0$ and $H_1 : \underline{\theta} = \underline{\theta}_1$, where $\{S_0, S_1\}$ are a partition of the observation space $\mathbf{R}^n$.

For any such decision rule, we expect that we will sometimes make errors in deciding between H_0 and H_1.

In Bayesian detection, we assign a *cost* or *loss* to our decisions.

Bayes Risk (Cont.)

Definition: For each $\underline{\theta}_i$ $(i = 0, 1)$, to the decision $\phi(\underline{x}) = j$, $j = 0, 1$, we assign a non-negative *loss* $L_{ij} = L[\underline{\theta} = \underline{\theta}_i, \phi(\underline{x}) = j]$ for deciding that H_j is in effect when in fact $H_i : \underline{\theta} = \underline{\theta}_i$ is true.

In the binary hypothesis testing problem, we have four particular losses of interest:

$$\begin{aligned} L_{01} &= \text{loss incurred when } H_0 \text{ is true, but } H_1 \text{ is decided;} \\ L_{00} &= \text{loss incurred when } H_0 \text{ is true, and } H_0 \text{ is decided;} \\ L_{10} &= \text{loss incurred when } H_1 \text{ is true, but } H_0 \text{ is decided;} \\ L_{11} &= \text{loss incurred when } H_1 \text{ is true, and } H_1 \text{ is decided.} \end{aligned}$$

Generally, we will have $L_{00} < L_{01}$ and $L_{11} < L_{10}$, because we would expect that it is more costly to make an error than decide correctly. Most often, we will take $L_{00} = L_{11} = 0$.

Bayes Risk (Cont.)

Definition: The *Risk* associated with test ϕ of H_0 versus H_1 for a given $\underline{\theta}$ is defined as

$$\begin{aligned}
R(\underline{\theta}, \phi) &\triangleq \mathrm{E}_{\underline{\theta}}\left[L[\underline{\theta}, \phi(\underline{X})]\right] \\
&= \int_{\mathbf{R}^n} L[\underline{\theta}, \phi(\underline{x})]\, dF_{\underline{\theta}}(\underline{x}) \\
&= \int_{\mathbf{R}^n} L[\underline{\theta}, \phi(\underline{x})] f_{\underline{\theta}}(\underline{x})\, d\underline{x} \\
&= \begin{cases} L_{00} P_{\underline{\theta}_0}(\{\phi(\underline{X}) = 0\}) + L_{01} P_{\underline{\theta}_0}(\{\phi(\underline{X}) = 1\}) & \text{for } \underline{\theta} = \underline{\theta}_0, \\ L_{10} P_{\underline{\theta}_1}(\{\phi(\underline{X}) = 0\}) + L_{11} P_{\underline{\theta}_1}(\{\phi(\underline{X}) = 1\}) & \text{for } \underline{\theta} = \underline{\theta}_1, \end{cases} \\
&= \begin{cases} L_{00} P_{00} + L_{01} P_{01} & \text{for } \underline{\theta} = \underline{\theta}_0, \\ L_{10} P_{10} + L_{11} P_{11} & \text{for } \underline{\theta} = \underline{\theta}_1, \end{cases}
\end{aligned}$$

where

$$P_{ij} = P(H_i \cap \{\phi(\underline{X}) = j\}) = P(\{H_i \text{ is true and } H_j \text{ is decided}\}), \quad i, j = 0, 1.$$

Bayes Risk (Cont.)

We can adopt the notation of the *size* α and power β of the test and write

$$\begin{aligned}
P_{00} &= 1 - \alpha, \\
P_{01} &= \alpha = P_{FA}, \\
P_{10} &= 1 - \beta, \\
P_{11} &= \beta = P_D.
\end{aligned}$$

We then have

$$\begin{aligned}
R(\underline{\theta}, \phi) &= \begin{cases} L_{00}(1-\alpha) + L_{01}\alpha, & \text{for } \underline{\theta} = \underline{\theta}_0, \\ L_{10}(1-\beta) + L_{11}\beta, & \text{for } \underline{\theta} = \underline{\theta}_1, \end{cases} \\
&= \begin{cases} (L_{01} - L_{00})\alpha + L_{00}, & \text{for } \underline{\theta} = \underline{\theta}_0, \\ (L_{10} - L_{11})(1-\beta) + L_{11}, & \text{for } \underline{\theta} = \underline{\theta}_1. \end{cases}
\end{aligned}$$

Bayes Risk (Cont.)

Noting that α and β are the Type I and Type II errors of the test ϕ, examination of the form of $R(\underline{\theta}, \phi)$ reveals that, without loss of generality, we can take $L_{00} = L_{11} = 0$. Thus we will consider a risk of the form

$$R(\underline{\theta}, \phi) = \begin{cases} L_{01}\alpha, & \text{for } \underline{\theta} = \underline{\theta}_0, \\ L_{10}(1-\beta), & \text{for } \underline{\theta} = \underline{\theta}_1. \end{cases}$$

Note that this form of the risk is just a weighted sum of the type I error probability α and the type II error probability $1 - \beta$.

Definition: The *Bayes Risk* of a test ϕ is the average risk over the prior distribution $P(\underline{\theta}) = \mathbf{p} = (p_0, p_1)$, and is denoted

$$\begin{aligned} R(\mathbf{p}, \phi) &= \mathrm{E}[R(\underline{\theta}, \phi)] \\ &= p_0 R(\underline{\theta}_0, \phi) + p_1 R(\underline{\theta}_1, \phi) \\ &= p_0 L_{01}\alpha + p_1 L_{10}(1-\beta). \end{aligned}$$

The Bayes Test

In Bayesian detection, our goal is to design a decision rule that minimizes the Bayes risk.

Definition: The *Bayes Decision Rule* $\phi_B(\cdot)$ for priors $\mathbf{p} = (p_0, p_1)$ and losses L_{01} and L_{10} is a decision rule that minimizes $R(\mathbf{p}, \phi)$ among all decision rules ϕ.

11.14

The Bayes test for testing $H_0 : \underline{\theta} = \underline{\theta}_0$ versus $H_1 : \underline{\theta} = \underline{\theta}_1$ is the test that minimizes

$$R(\mathbf{p}, \phi) = p_0 L_{01} \alpha + p_1 L_{10}(1 - \beta),$$

where

$$\alpha = P_{\underline{\theta}_0}(\{\phi(\underline{X}) = 1\}) = \mathrm{E}_{\underline{\theta}_0}[\phi(\underline{X})] = \int_{S_1} dF_{\underline{\theta}_0}(\underline{x}),$$

and

$$1 - \beta = 1 - P_{\underline{\theta}_1}(\{\phi(\underline{X}) = 1\}) = 1 - \mathrm{E}_{\underline{\theta}_1}[\phi(\underline{X})] = 1 - \int_{S_1} dF_{\underline{\theta}_1}(\underline{x}),$$

where $S_1 = \{\underline{x} \in \mathbf{R}^n : \phi(\underline{x}) = 1\}$.

Thus we have

$$\begin{aligned} R(\mathbf{p}, \phi) &= p_0 L_{01} \int_{S_1} dF_{\underline{\theta}_0}(\underline{x}) + p_1 L_{10} \left(1 - \int_{S_1} dF_{\underline{\theta}_1}(\underline{x})\right) \\ &= p_1 L_{10} + \int_{S_1} \left[p_0 L_{01}\, dF_{\underline{\theta}_0}(\underline{x}) - p_1 L_{10}\, dF_{\underline{\theta}_1}(\underline{x})\right]. \end{aligned}$$

11.15

To design the Bayes test, we must select S_1 to minimize $R(\mathbf{p}, \phi)$ since S_1 determines the test ϕ.

$$\begin{aligned} R(\mathbf{p}, \phi) &= p_0 L_{01} \int_{S_1} dF_{\underline{\theta}_0}(\underline{x}) + p_1 L_{10} \left(1 - \int_{S_1} dF_{\underline{\theta}_1}(\underline{x})\right) \\ &= p_1 L_{10} + \int_{S_1} \left[p_0 L_{01}\, dF_{\underline{\theta}_0}(\underline{x}) - p_1 L_{10}\, dF_{\underline{\theta}_1}(\underline{x})\right]. \end{aligned}$$

This can be done by including any $\underline{x} \in \mathbf{R}^n$ that makes the integrand

$$\psi(\underline{x}) = \left[p_0 L_{01}\, dF_{\underline{\theta}_0}(\underline{x}) - p_1 L_{10}\, dF_{\underline{\theta}_1}(\underline{x})\right]$$

negative.

This yields

$$S_1 = \left\{ \underline{x} \in \mathbf{R}^n : \frac{dF_{\theta_1}(\underline{x})}{dF_{\theta_0}(\underline{x})} > \frac{p_0 L_{01}}{p_1 L_{10}} \right\}$$

If the cdfs $F_{\underline{\theta}_0}(\underline{x})$ and $F_{\underline{\theta}_1}(\underline{x})$ have corresponding pdfs (or p.m.f.s) $f_{\underline{\theta}_0}(\underline{x})$ and $f_{\underline{\theta}_1}(\underline{x})$, then we have

$$\frac{dF_{\theta_1}(\underline{x})}{dF_{\theta_0}(\underline{x})} = \frac{f_{\theta_1}(\underline{x})}{f_{\theta_0}(\underline{x})} = L(\underline{x}).$$

The resulting Bayes test is a likelihood ratio test of the form

$$\phi_B(\underline{x}) = \begin{cases} 1, & \text{for } L(\underline{x}) > L_0; \\ 0, & \text{for } L(\underline{x}) \leq L_0, \end{cases}$$

where the threshold

$$L_0 = \frac{p_0 L_{01}}{p_1 L_{10}}$$

is a function of both the priors and the losses.

Two Difficulties with the Bayesian Approach

There are two basic problems with the Bayesian approach when applied to radar detection:

1. How do you determine reasonable losses L_{01} and L_{10}, or at least determine the ratio L_{01}/L_{10}?

2. How do you determine, or estimate, the priors $\mathbf{p} = (p_0, p_1)$?

Assuming you can determine reasonable losses L_{01} and L_{10}, the determination of the priors $\mathbf{p} = (p_0, p_1)$ is still a significant problem. One way to deal with this problem is to construct a *minimax test*.

11.18

The Minimax Test

In the minimax test, we design a Bayes test for the worst possible prior (also known as the *Least Favorable Prior*.) So the resulting Bayes risk would be

$$\min_{\phi} \max_{\mathbf{p}} R(\mathbf{p}, \phi).$$

Effectively, we assume there is an adversarial relationship in which the particular detector you would use for every possible prior $\mathbf{p}$ is known by your adversary (i.e., the adversary knows your cost function,) and the adversary has the freedom to pick the prior $\mathbf{p}$.

Among all sets of tests and priors the adversary picks the prior $\mathbf{p}$ yielding the largest Bayes risk.

The logical way to confront such an adversary is to use the *minimax strategy:* Find the worst prior $\mathbf{p}^*$, yielding the largest Bayes risk, and use the Bayes test ϕ_B^* for this prior. This is the *minimax test:*

$$\phi_{MM} = \phi_B^* = \arg\min_{\phi} \max_{\mathbf{p}} R(\mathbf{p}, \phi).$$

11.19

Key Property of the Minimax Test

By selecting the minimax test (i.e., designing the Bayes test for the least favorable priors,) it can be shown that we will never incur a cost greater than

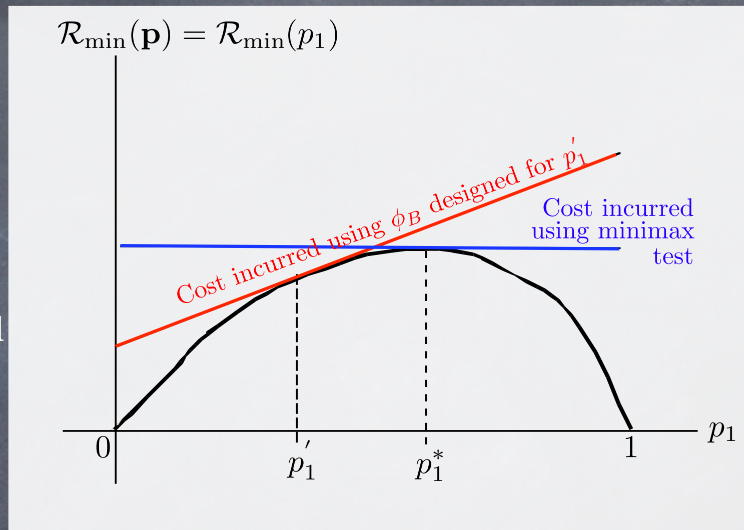
$$R(\mathbf{p}^*, \phi_B^*) = \min_{\phi} \max_{\mathbf{p}} R(\mathbf{p}, \phi),$$

even if $\mathbf{p} \neq \mathbf{p}^*$.

This is not true of other tests!

There is a geometric argument that easily illustrates this thaqt is based on the fact that Bayesian tests are likelihood ratio tests.

Suppose we plot the minimum Bayes risk for a given set of costs as a function of the prior $\mathbf{p} = (p_0, p_1)$



The resulting curve is convex $\cap$ in p_1.

For any Bayes test designed for a $p_1^{'} \neq p_1^*$, there exist values of p_1 such that $\mathcal{R}_{\min}(p_1) > \mathcal{R}_{\min}(p_1^*)$.

So with the minimax test, you know the average cost you will incur regardless of the prior.

If this average cost is acceptable, the minimax test is a reasonable test.