

Session 19

Recall... Bayesian Detection

19.1

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

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

Recall...

Bayesian Detection

19.2



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

$$\begin{aligned} F(\underline{x}) &= F(\underline{x}|\underline{\theta}_0)P(\underline{\theta}_0) + F(\underline{x}|\underline{\theta}_1)P(\underline{\theta}_1) \\ &= p_0F(\underline{x}|\underline{\theta}_0) + p_1F(\underline{x}|\underline{\theta}_1), \end{aligned}$$

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

19.3

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

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

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

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

The Bayes Test

19.8

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 .

19.9

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

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

and

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

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

Thus we have

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

$$\int_{S_1} [p_0 L_{01} f_{\underline{\theta}_0}(\underline{x}) - p_1 L_{10} f_{\underline{\theta}_1}(\underline{x})] d\underline{x}$$

19.10

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

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

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

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

negative.

19.11

This yields

$$S_1 = \left\{ \underline{x} \in \mathbf{R}^n : \frac{dF_{\underline{\theta}_1}(\underline{x})}{dF_{\underline{\theta}_0}(\underline{x})} > \frac{p_0 L_{01}}{p_1 L_{10}} \right\}$$

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

$$\frac{dF_{\underline{\theta}_1}(\underline{x})}{dF_{\underline{\theta}_0}(\underline{x})} = \frac{f_{\underline{\theta}_1}(\underline{x})}{f_{\underline{\theta}_0}(\underline{x})} = L(\underline{x}).$$

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*.

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*:

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

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

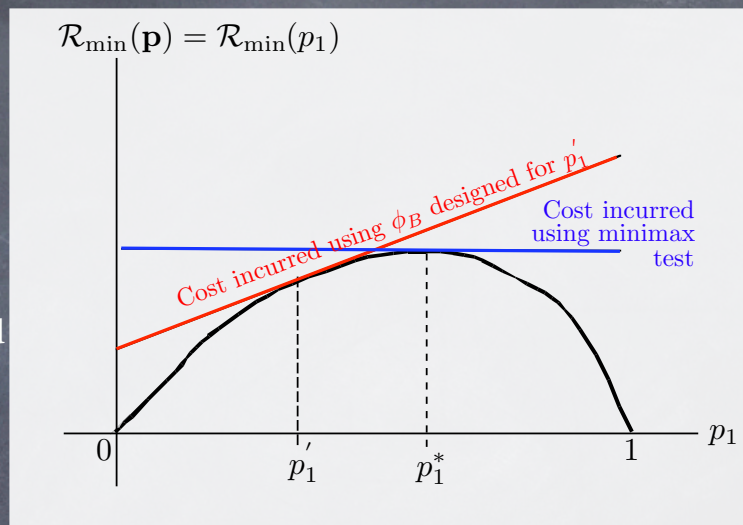
This is not true of other tests!

There is a geometric argument that easily illustrates this that 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.