Since $\Gamma_1 = \{ y \in \mathbb{R} \mid y \geq \tau' \}$, we have

$$P_j(\Gamma_1) = \int_{\tau'}^{\infty} p_j(y)\,dy = 1 - \Phi\left(\frac{\tau' - \mu_j}{\sigma}\right)$$

$$= \begin{cases} 1 - \Phi\left(\frac{\log \tau}{d} + \frac{d}{2}\right), & j = 0 \\ 1 - \Phi\left(\frac{\log \tau}{d} - \frac{d}{2}\right), & j = 1 \end{cases}$$

where Φ is the cumulative probability distribution function (cdf) of $\mathcal{N}(0,1)$

$d = \frac{\mu_1 - \mu_0}{\sigma}$ ~ $\boxed{\text{a simple version of signal-to-noise ratio}}$

$\Rightarrow$ The minimum Bayes risk $r(\delta_B)$ can be calculated: For uniform cost and equal priors,

$$r(\delta_B) = 1 - \Phi\left(\frac{d}{2}\right)$$

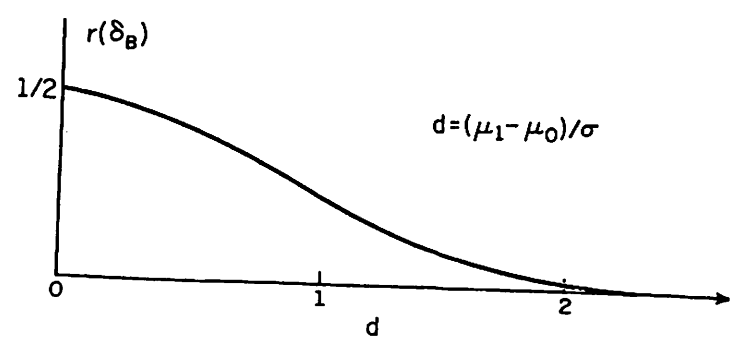


FIGURE II.B.3. Bayes risk in location testing with Gaussian err

§2. Minimax Hypothesis Testing

* In many practical applications, the prior probabilities, π_0 and π_1, may not be known. In this case, the Bayes risk may not be a good objective function to minimize ($r(\delta) = \pi_0 R_0(\delta) + \pi_1 R_1(\delta)$).

We need to look for an alternative design criterion.

As one can see, if both $R_0(\delta)$ and $R_1(\delta)$ are small, the risk function $r(\delta)$ is small too. This motivates to use the following objective function

$$\max\{R_0(\delta), R_1(\delta)\}.$$

* The minimax criterion :

$$\min_{\delta} \max\{R_0(\delta), R_1(\delta)\}$$

This is max of two.

* Mathematically, it is sometime more conventinent to consider max/min over a continuom region (not a finite set, such as, two above) so that oncan take derivative to study the extremal points.

* For a given prior $\pi_0 \in [0, 1]$, let

$$r(\pi_0, \delta) = \pi_0 R_0(\delta) + (1-\pi_0) R_1(\delta)$$

which is function of π_0 defined on $[0, 1]$.

* For a fixed δ, $r(\pi, \delta)$ is straight line in terms of π_0. It is not hard to see

$$\max\{R_0(\delta), R_1(\delta)\} = \max_{0 \le \pi_0 \le 1} r(\pi_0, \delta)$$

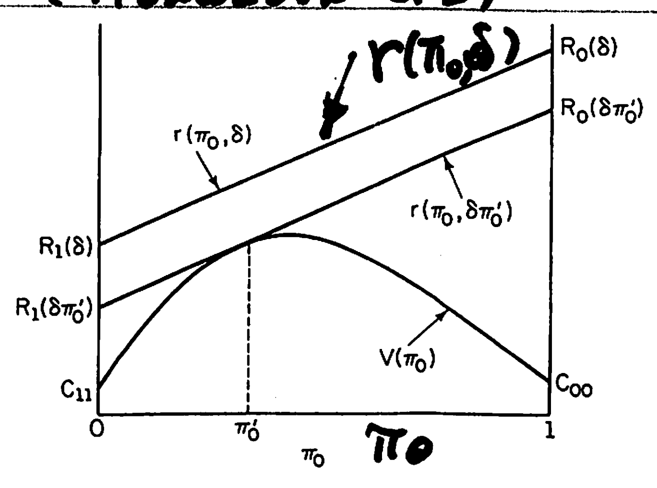
$\Rightarrow$ max over two $\xrightarrow{\text{becomes}}$ max over a continuous region

* For each prior $\pi_0 \in [0, 1]$, let δ_{π_0} denote a Bayes rule corresponding to π_0.

Let $$V(\pi_0) = r(\pi_0, \delta_{\pi_0})$$

the minimum possible Bayes risk for the prior π_0.

Then $V(\pi_0)$ is a continuous concave function of π_0 (Homework #1)



* Let
$$\pi_L = \arg \max_{0 \le \pi_0 \le 1} V(\pi_0)$$
called the least-favorable prior.

* Proposition II.C.1: The Minimax Test

Suppose that π_L is a solution to
$$V(\pi_L) = \max_{0 \le \pi_0 \le 1} V(\pi_0).$$
Suppose further that either $\pi_L = 0$, $\pi_L = 1$, or $R_1(\delta_{\pi_L}) = R_0(\delta_{\pi_L})$. Then, δ_{π_L} is a minimax rule.

Proof:

1) Consider the case $R_1(\delta_{\pi_L}) = R_0(\delta_{\pi_L})$

For any π_0, we have

$$\max_{0 \le \pi_0 \le 1} \min_{\delta} r(\pi_0, \delta) \overset{\text{definition}}{=\!=} r(\pi_L, \delta_{\pi_L})$$
$$= \pi_L R_0(\delta_{\pi_L}) + (1-\pi_L) R_1(\delta_{\pi_L})$$
$$\overset{*}{=} \pi_L R_0(\delta_{\pi_L}) + (1-\pi_L) R_0(\delta_{\pi_L})$$
$$= R_0(\delta_{\pi_L})$$
$$= \pi_0 R_0(\delta_{\pi_L}) + (1-\pi_0) R_0(\delta_{\pi_L})$$
$$\overset{*}{=} \pi_0 R_0(\delta_{\pi_L}) + (1-\pi_0) R_1(\delta_{\pi_L})$$
$$= r(\pi_0, \delta_{\pi_L}), \text{ independent of } \pi_0$$

Given condition

Thus,

$$\max_{0\le\pi_0\le1} \min_{\delta} r(\pi_0,\delta) = \max_{0\le\pi_0\le1} r(\pi_0, \delta_{\pi_L})$$

$$\ge \min_{\delta} \max_{0\le\pi_0\le1} r(\pi_0,\delta) \qquad (*)$$

On the other hand, for each δ, we have

$$\max_{0\le\pi_0\le1} r(\pi_0,\delta) \ge \max_{0\le\pi_0\le1} \min_{\delta} r(\pi_0,\delta)$$

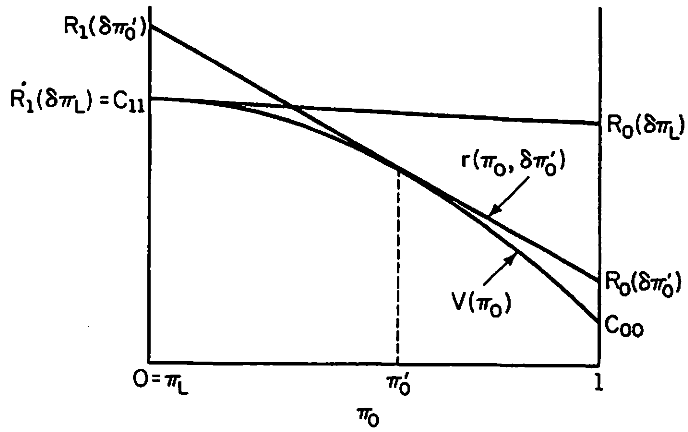
$$\Rightarrow \min_{\delta} \max_{0\le\pi_0\le1} r(\pi_0,\delta) \ge \max_{0\le\pi_0\le1} \min_{\delta} r(\pi_0,\delta) \qquad (**)$$

$$\Rightarrow \min_{\delta} \max_{0\le\pi_0\le1} r(\pi_0,\delta) = \max_{0\le\pi_0\le1} \min_{\delta} r(\pi_0,\delta)$$

$$= r(\pi_L, \delta_{\pi_L})$$

$\Rightarrow$ δ_{π_L} is a minimax rule.

2) Consider the case $\pi_L = 0$.



Since $V(\pi_0)$ is a concave function of π_0,

$$V(0) = \max_{0 \le \pi_0 \le 1} V(\pi_0) = R_1(\delta\pi_L),$$

$$R_1(\delta\pi_L) \ge R_0(\delta\pi_L)$$

Thus,

$$\max_{0 \le \pi_0 \le 1} r(\pi_0, \delta\pi_L)$$

$$= \max_{0 \le \pi_0 \le 1} \pi_0 (R_0(\delta\pi_L) - R_1(\delta\pi_L)) + R_1(\delta\pi_L)$$

$$= R_1(\delta\pi_L) = r(\pi_L, \delta\pi_L).$$

And

$$\min_{\delta} \max_{0 \le \pi_0 \le 1} r(\pi_0, \delta) \le \max_{0 \le \pi_0 \le 1} r(\pi_0, \delta\pi_L)$$

$$= r(\pi_L, \delta\pi_L).$$

On the other hand,

$$\max_{0 \le \pi_0 \le 1} r(\pi_0, \delta) \ge \max_{0 \le \pi_0 \le 1} \min_{\delta} r(\pi_0, \delta)$$

$$= \max_{0 \le \pi_0 \le 1} r(\pi_0, \delta_{\pi_0})$$

$$= r(\pi_L, \delta_{\pi_L})$$

$$\Rightarrow \min_{\delta} \max_{0 \le \pi_0 \le 1} r(\pi_0, \delta) \ge r(\pi_L, \delta_{\pi_L}).$$

$$\Rightarrow \min_{\delta} \max_{0 \le \pi_0 \le 1} r(\pi_0, \delta) = r(\pi_L, \delta_{\pi_L})$$

$\Rightarrow \delta_{\pi_L}$ is a minimax rule. (Homework 2)

3) The case $\pi_L = 1$ is similar.

$\square$

$\Rightarrow$ Under the assumptions of Prob., δ_{π_L} a minimax rule is the Bayes rule for the least-favorable prior.

* Decision rule δ satisfying $R_0(\delta) = R_1(\delta)$ is called an equalizer rule.

* When $\pi_L \in (0,1)$ and $V(\pi_0)$ is differentiable in $(0,1)$, δ_{π_L} is also a minimax rule.

a minimax rule $\Leftrightarrow$ a Bayes rule for the least-favorable prior

$\Leftrightarrow$ an equalizer rule

When $V(\pi_0)$ is differentiable in $(0,1)$,

$$V'(\pi_L) = 0 \Rightarrow R_1(\pi_L) = R_0(\pi_L), \quad 0 < \pi_L < 1$$

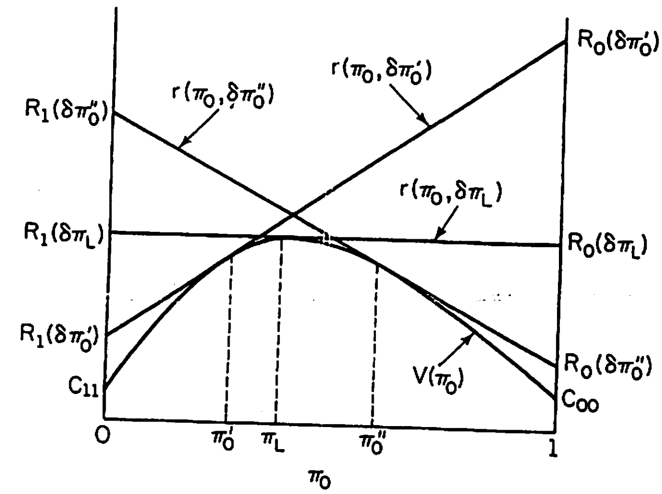


FIGURE II.C.2. Illustration of the minimax rule when V has an interior maximum.

* When $\pi_L \in (0,1)$ and $V(\pi_0)$ is not differentiable at π_L

Let $\delta_{\pi_L}^- = \lim_{\pi_0 \nearrow \pi_L} \delta_{\pi_0}$

$$\xrightarrow{\pi_0} \pi_L$$

$$\delta_{\pi_L}^+ = \lim_{\pi_0 \searrow \pi_L} \delta_{\pi_0}$$

$$\pi_L \xleftarrow{} \pi_0$$

It is clear that the critical region for δ_{π}^- is

$$\Gamma_1^- = \{ y \in \Gamma \mid (1-\pi_L)(C_{11}-C_{01})p_1(y) \leq \pi_L (C_{00}-C_{10}) p_0(y) \}$$

Also for $\delta^{+}_{\pi_L}$ (assume $C_{11} < C_{01}$)

$$\Gamma_1^{+} = \{ y \in \Gamma \mid (1-\pi_L)(C_{11}-C_{01})p_1(y) \leq \pi_L (C_{01}-C_{10}) p_0(y) \}$$

For a number $q \in [0,1]$, we consider the decision rule $\tilde{\delta}_{\pi_L}$ that uses Γ_1^{-} with probability q

Γ_1^{+} - - - $1-q$

i.e., $\tilde{\delta}_{\pi_L}$ chooses H_1 if $y \in \Gamma_1^{+}$

H_0 if $y \in (\Gamma_1^{-})^c$

chooses H_1 with probability q if y is on the boundary of Γ_1^{-}

$\tilde{\delta}_{\pi_L}$ is still a Bayes rule for π_L since the risk is not affected by the boundary.

Also $r(\pi_L, \tilde{\delta}_{\pi_L}) = V(\pi_L)$

The conditional risks do depend on the boundary:

$$R_j(\tilde{\delta}_{\pi_L}) = q R_j(\delta^{-}_{\pi_L}) + (1-q) R_j(\delta^{+}_{\pi_L})$$

and

$R_0(\tilde{\delta}_{\pi_L}) = R_1(\tilde{\delta}_{\pi_L})$ is achieved by choosing

$$q = \frac{R_0(\delta_{\pi_L}^+) - R_1(\delta_{\pi_L}^+)}{R_0(\delta_{\pi_L}^+) - R_1(\delta_{\pi_L}^+) + R_1(\delta_{\pi_L}^-) - R_0(\delta_{\pi_L}^-)}$$

This, thus, satisfies the proposition and therefore $\tilde{\delta}_{\pi_L}$ with q chosen above is a minimax rule.

<u>Remark:</u> If the region

$$\{y \in \Gamma \mid (1-\pi_L)(C_{11}-C_{01})p_1(y) = \pi_L(C_{00}-C_{10})p_0(y)\}$$

occurs with zero probability under H_0 and H_1, then $\tilde{\delta}_{\pi_L} = \delta_{\pi_L}$.

The the decision rule $\tilde{\delta}_{\pi_L}$ is an example of a <u>randomized decision rule</u> (it will be discussed later in more details)

$$q = \frac{V'(\pi_L^+)}{V'(\pi_L^+) - V'(\pi_L^-)}$$

$V'(\pi_L^+)$
$= R_0(\delta_{\pi_L}^+)$
$- R_1(\delta_{\pi_L}^+)$

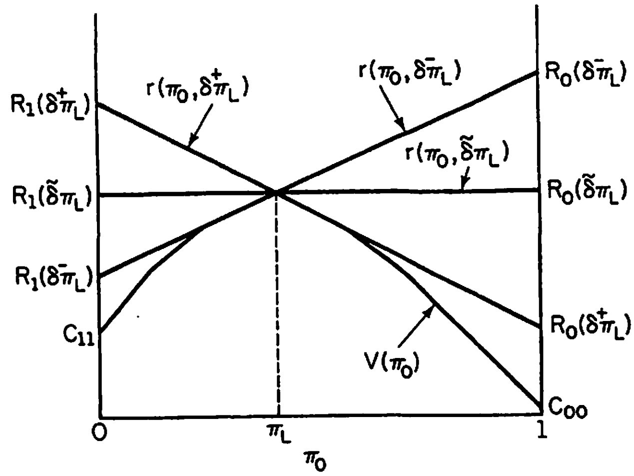


FIGURE II.C.4. Depiction of a randomized decision rule.

$V'(\pi_L^-)$
$= R_0(\delta_{\pi_L}^-)$
$- R_1(\delta_{\pi_L}^-)$

The lines $r(\pi_0, \delta_{\pi_L}^-)$ and $r(\pi_0, \delta_{\pi_L}^+)$ cross at $\pi_0 = \pi_L$ and have slopes equal to $V'(\pi_0^-)$ and $V'(\pi_0^+)$, respectively.

left derivative right derivative

of $V(\pi_0)$

& concave, it always has left and right derivatives

$$\lim_{\pi_0 \nearrow \pi_L} \frac{V(\pi_0) - V(\pi_L)}{\pi_0 - \pi_L}, \quad \lim_{\pi_0 \searrow \pi_L} \frac{V(\pi_0) - V(\pi_L)}{\pi_0 - \pi_L}$$

By varying the probability q from 0 to 1, any line between these two lines can be obtained. The particular choice of q above yields the horizontal line that lies between these two.

We next consider two examples, respectively.

<u>Example I:</u> Location Testing with Gaussian Error

Consider the location testing problem we considered previously with Gaussian noise and uniform costs.

For every π_0, it is found

$$V(\pi_0) = \pi_0 \left(1 - \Phi\left(\frac{\tau' - \mu_0}{\sigma}\right)\right) + (1-\pi_0)\,\Phi\left(\frac{\tau' - \mu_1}{\sigma}\right)$$

where $\tau' = \frac{\sigma^2}{\mu_1 - \mu_0} \log\left(\frac{\pi_0}{1-\pi_0}\right) + \frac{\mu_1 + \mu_0}{2}$

Since $V(0) = C_{11} = 0 = C_{00} = V(1)$, the least favorable prior π_L has to be inside $(0,1)$, i.e., $\pi_L \in (0, 1)$, in the interior in $(0,1)$.

Also, $V(\pi_0)$ is differentiable in $(0,1)$, randomization is not necessary.

In this case, π_L can be found by setting

$$R_0(\delta_{\pi_L}) = R_1(\delta_{\pi_L})$$

$$\Rightarrow 1 - \Phi\left(\frac{\tau' - \mu_0}{\sigma}\right) = \Phi\left(\frac{\tau' - \mu_1}{\sigma}\right)$$

$$\Rightarrow \tau'_L = \frac{\mu_0 + \mu_1}{2}$$

$\Rightarrow$ The minimax decision rule is $\delta_{\pi_L}(y) = \begin{cases} 1 & \text{if } y \geq \frac{\mu_0+\mu_1}{2} \\ 0 & \text{if } y < \frac{\mu_0+\mu_1}{2} \end{cases}$

The least-favorable prior $\pi_L = \frac{1}{2}$ and the minimax risk is

$$V\left(\tfrac{1}{2}\right) = 1 - \Phi\left(\frac{\mu_1 - \mu_0}{2\sigma}\right).$$

Remark: It is not surprising that the minimax decision rule is when no prior information is known. In this case, one treats $\pi_0 = \pi_1$

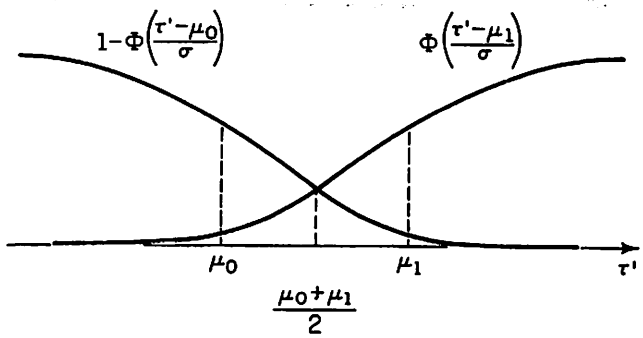


FIGURE II.C.5. Conditional risks for location testing with Gaussian error and uniform costs.

Example II: The Binary Channel

Consider the previous example of binary channel with uniform costs.

In this case, the threshold for the Bayes rule:

$$\tau = \frac{\pi_0}{\pi_1} = \frac{\pi_0}{1-\pi_0}$$

The minimum Bayes risk function is

$$V(\pi_0) = \min\{(1-\pi_0)\lambda_1, \pi_0(1-\lambda_0)\} + \min\{(1-\pi_0)(1-\lambda_1), \pi_0\lambda_0\}$$

(Homework 3)

$$(1-\pi_0)\lambda_1 \lesseqgtr \pi_0(1-\lambda_0) \Rightarrow \pi_0 \gtreqless \frac{\lambda_1}{1-\lambda_0+\lambda_1},$$

$$(1-\pi_0)(1-\lambda_1) \lesseqgtr \pi_0\lambda_0 \Rightarrow \pi_0 \gtreqless \frac{1-\lambda_1}{1-\lambda_1+\lambda_0}.$$

$$V(\pi_0) = \begin{cases} \pi_0 & \text{if } 0 \le \pi_0 \le \underline{\pi} \\ \underline{\pi} + C\pi_0 & \text{if } \underline{\pi} \le \pi_0 < \overline{\pi} \\ 1-\pi_0 & \text{if } \overline{\pi}_0 \le \pi_0 \le 1 \end{cases}$$

where

$$\underline{\pi} = \min\left\{\frac{\lambda_1}{1-\lambda_0+\lambda_1}, \frac{1-\lambda_1}{1-\lambda_1+\lambda_0}\right\}$$

$$\overline{\pi} = \max\left\{\frac{\lambda_1}{1-\lambda_0+\lambda_1}, \frac{1-\lambda_1}{1-\lambda_1+\lambda_0}\right\}$$

$$C = (1-\overline{\pi}-\underline{\pi})/(\overline{\pi}-\underline{\pi})$$

$\Rightarrow V(\pi_0)$ is piecewise linear with changes in slope at $\underline{\pi}$ and $\overline{\pi}$.

$V(0) = V(1) = 0$

If $C < 0$, $\max V(\pi_0) = \underline{\pi}$

and $\boxed{\pi_L = \underline{\pi}}$

If $C > 0$, $\max V(\pi_0) = \underline{\pi} + C\pi_0 \Big|_{\text{when } \pi_0 = \overline{\pi}}$

and $\boxed{\pi_L = \overline{\pi}}$

If $C = 0$, π_L is any prior in $[\underline{\pi}, \overline{\pi}]$.

In either case, $V(\pi_L) = \max\{\underline{\pi}, 1 - \overline{\pi}\}$

It is clear that $V(\pi_0)$ is not differentiable at $\pi_0 = \pi_L = \underline{\pi}$ or $\overline{\pi}$,

we need to consider a randomized test.

Let us consider the case $C > 0$. The case $C < 0$ is similar.

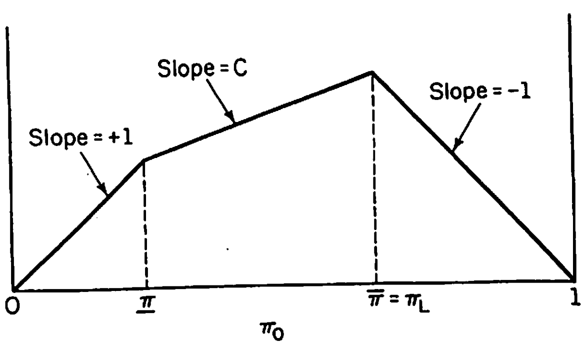


FIGURE II.C.6. $V(\pi_0)$ for the binary channel.

In this case, $\pi_L = \bar{\pi}$.

$\Rightarrow V'(\pi_L^+) = -1$

$V'(\pi_L^-) = C$

$\Rightarrow q = \dfrac{-1}{-1-C} = \dfrac{\bar{\pi} - \underline{\pi}}{1-2\underline{\pi}}$

It is not hard to see that

If $\pi_0 > \bar{\pi}$, $\delta_{\pi_0}(0) = \delta_{\pi_0}(1) = 0$

$\Rightarrow \delta^+_{\pi_L}(0) = \delta^+_{\pi_L}(1) = 0$, Γ_1^+ is empty.

For $\underline{\pi} < \pi_0 < \bar{\pi}$,

$$\delta_{\pi_0}(y) = \begin{cases} y, & \text{if } \lambda_1 < 1-\lambda_0 \\ 1-y, & \text{if } \lambda_1 > 1-\lambda_0. \end{cases}$$

Thus, for example, if $\lambda_1 < 1-\lambda_0$, we have $\delta^-_{\pi_L}(y) = y$

& the minimax rule $\bar{\delta}_{\pi_L}$ chooses 0 if $y=0$

and chooses $\begin{cases} 1 \text{ with probability } q \text{ if } y=1 \\ 0 \quad \text{"} \quad \text{"} \quad \text{"} \quad 1-q \end{cases}$

boundary

$V(\bar{\pi}_L) = 1-\bar{\pi} = \dfrac{\lambda_0}{1-\lambda_1+\lambda_0}$

If $\lambda_0 = \lambda_1 = \lambda$ (in which case $\lambda_1 < 1 - \lambda_0$ if $\lambda < \frac{1}{2}$),

then $q = 1$ and

$$V(\pi_L) = \lambda$$

$\Rightarrow$ the same as the Bayes risk for any $\pi_0 \in [\lambda, 1-\lambda]$.

$\Rightarrow$ Uncertainty in the prior does not cost any performance as long as $\lambda \le \pi_0 \le 1-\lambda$

In summary: Optimal tests can be designed even without the assumption of known priors by using a minimax ~~rule~~ design criterion. The solution is a Bayes test for the least-favorable prior, with randomization on the boundary of the decision region being necessary in some problems.

§3. Neyman–Pearson Hypothesis Testing

* In both Bayes and minimax design criteria, cost functions are needed by minimizing a kind of cost/risk. In many applications, a specific cost function may not be possible or desirable. In these cases, the Neyman-Pearson criterion is often imposed.

* For the binary hypothesis test:

$$H_0 \sim P_0$$
$$\text{vs.}$$
$$H_1 \sim P_1$$

there are two types of errors to be made:

H_0 can be falsely rejected:
Type I error / false alarm

H_1 can be falsely rejected:
Type II error / a miss

* False alarm or a miss comes from radar in which H_0 and H_1 usually represent the absense and presence of a target, respectively.

Reject the absense of a target
→ target detected falsely.

Reject the presence of a target
→ a target is missed.

* For a decision rule δ, the probability of Type I error is known as the size or <u>false-alarm probability</u> (or false-alarm rate) of δ. We denote it by $P_F(\delta)$.

The probability of Type II error is called the <u>miss probability</u>, $P_M(\delta)$. However, we often talk above the <u>detection probability</u>:

$$P_D(\delta) = 1 - P_M(\delta)$$

also called the power of δ.

* The design of a test for H_0 vs. H_1 involves a trade-off between the probabilities of the two types of errors. It is because one can always be made arbitrarily small at the expense of the other.

The Bayes and minimax criteria are two ways of trading these off (such as average of two or weighted two into a single one).

* The <u>Neyman-Pearson criterion</u> for making this trade-off is to place a bound on the false-alarm probability and then to minimize the miss probability within this constraint, i.e.,

$$\max_{\delta} P_D(\delta) \text{ subject to } P_F(\delta) \le \alpha$$

where α is the above-mentioned bound, which is known as the level or significance level of the test.

⇒ The Neyman–Pearson design goal is to find the most powerful α-level test of H_0 vs. H_1

⇒ As opposed to the Bayes and minimax criteria, the Neyman–Pearson criterion recognizes a basic asymmetry in importance of the two hypotheses.

⇒ To do so, we need to consider randomized tests.

* A <u>randomized decision rule</u> $\tilde{\delta}$ for H_0 vs. H_1 is defined as a function mapping Γ to the interval $[0, 1]$:

For $y \in \Gamma$, $\tilde{\delta}(y)$ is the conditional probability with which we accept H_1 given the observation $Y = y$.

Example: the randomized minimax rule $\tilde{\delta}_{\pi_L}$ we studied before can be written as follows

$$\tilde{\delta}_{\pi_L}(y) = \begin{cases} 1 & \text{if } L(y) > \tau_L \\ q & \text{if } L(y) = \tau_L \quad \text{(boundary)} \\ 0 & \text{if } L(y) < \tau_L \end{cases}$$

where τ_L is the threshold corresponding to the least-favorable prior π_L.

* A nonrandomized decision rule, δ, is a special case of a randomized decision rule. The difference between the two is that the value of δ is the index of the accepted hypothesis and the value of $\tilde{\delta}$ is the probability with which we accept H_1. They coincide when $\tilde{\delta}$ takes on only the two values 0 and 1.

* For a randomized rule $\tilde{\delta}$, the false-alarm probability is

$$P_F(\tilde{\delta}) = E_0\{\tilde{\delta}(Y)\} = \int_\Gamma \tilde{\delta}(y) p_0(y)\, \mu(dy),$$

where $E_0\{\ \}$ denotes expectation under hypothesis H_0. It is the probability with which it accepts H_1 given that H_0 is true.

The detection probability of $\tilde{\delta}$ is

$$P_D(\tilde{\delta}) = E_1(\tilde{\delta}(Y)) = \int_\Gamma \tilde{\delta}(y) p_1(y)\, \mu(dy).$$

* The general solution to the Neyman–Pearson design problem can be summarized in the following proposition.

Proposition II.D.1: The Neyman – Pearson Lemma

Consider the hypothesis pair $H_0 \sim P_0$ vs $H_1 \sim P_1$ where P_i has density p_j for $j = 0, 1$. Suppose that $\alpha > 0$. Then the following statements are true:

(i) (optimality) Let $\tilde{\delta}$ be any decision rule satisfying $P_F(\tilde{\delta}) \le \alpha$, and let $\tilde{\delta}'$ be any decision rule of the form

$$\tilde{\delta}'(y) = \begin{cases} 1 & \text{if } p_1(y) > \eta\, p_0(y) \\ \gamma(y) & \text{if } p_1(y) = \eta\, p_0(y) \\ 0 & \text{if } p_1(y) < \eta\, p_0(y) \end{cases} \qquad \text{(II.D.5)}$$

where $\eta \ge 0$ and $0 \le \gamma(y) \le 1$ are such that

$$P_F(\tilde{\delta}') = \alpha.$$

Then $P_D(\tilde{\delta}') \ge P_D(\tilde{\delta})$. That is, any size-$\alpha$ decision rule of the form (II.D.5) is a Neyman – Pearson rule.

(ii) (Existence) For any $\alpha \in (0, 1)$ there is a decision rule, $\tilde{\delta}_{NP}$, of the form of (II.D.5) with $\gamma(y) = \gamma_0$ (a constant), for which

$$P_F(\tilde{\delta}_{NP}) = \alpha .$$

(iii) (Uniqueness) Suppose that $\tilde{\delta}''$ is any α-level Neyman–Pearson decision rule for H_0 vs. H_1. Then $\tilde{\delta}''$ must be of the form of (II.D.5) except possibly on a subset of Γ having zero probability (measure) under H_0 and H_1.

Proof: (i) From the definition of $\tilde{\delta}'$, we have

$$(\tilde{\delta}'(y) - \tilde{\delta}(y))(p_1(y) - \eta\, p_0(y)) \geq 0, \ \forall y \in \Gamma.$$

Thus, $\int_\Gamma (\tilde{\delta}'(y) - \tilde{\delta}(y))(p_1(y) - \eta\, p_0(y))\, \mu(dy) \geq 0$

$$\Rightarrow \int_\Gamma \tilde{\delta}' p_1\, d\mu - \int_\Gamma \tilde{\delta} p_1\, d\mu \geq \eta \left[\int_\Gamma \tilde{\delta}' p_0\, d\mu - \int_\Gamma \tilde{\delta} p_0\, d\mu \right].$$

$$\Rightarrow P_D(\tilde{\delta}') - P_D(\tilde{\delta}) \geq \eta [P_F(\tilde{\delta}') - P_F(\tilde{\delta})]$$

$$= \eta(\alpha - P_F(\tilde{\delta}))$$

$$\geq 0 .$$

from the assumption on $\tilde{\delta}$

$$\Rightarrow P_D(\tilde{\delta}') \geq P_D(\tilde{\delta})$$

This proves (i).

(ii) Let η_0 be the smallest number such that

$$P_0(p_1(y) \boxed{>} \eta_0 p_0(y)) \le \alpha.$$

If $P_0(p_1(y) > \eta_0 p_0(y)) < \alpha$, we choose

$$\gamma_0 = \frac{\alpha - P_0(p_1(y) > \eta_0 p_0(y))}{P_0(p_1(y) = \eta_0 p_0(y))},$$

otherwise, choose γ_0 arbitrarily.

We only need to show $0 \le \gamma_0 \le 1$. To do so, we need to show

$$\alpha - P_0(p_1(y) > \eta_0 p_0(y)) \le P_0(p_1(y) = \eta_0 p_0(y)).$$

If $\alpha - P_0(p_1(y) > \eta_0 p_0(y)) > P_0(p_0(y) = \eta_0 p_0(y))$, then $P_0(p_1(y) \ge \eta_0 p_0(y)) < \alpha$

$\ge$ Not $>$ only

Then, η_0 can be reduced to η_0', $\eta_0 > \eta_0'$, by a little such that

$$P_0(p_1(y) > \eta_0' p_0(y)) \le \alpha.$$

This contradicts with the choice of η_0.

Then, define $\tilde{\delta}_{NP}$ to be the decision rule of (II.D.5) with $\eta = \eta_0$ and $\gamma(y) = \gamma_0$.
We have

$$P_F(\tilde{\delta}_{NP}) = E_0\{\tilde{\delta}_{NP}(Y)\}$$

$$= P_0(p_1(Y) > \eta_0\, p_0(Y)) + \gamma_0\, P_0(p_1(Y) = \eta_0\, p_0(Y))$$

$$= \alpha.$$

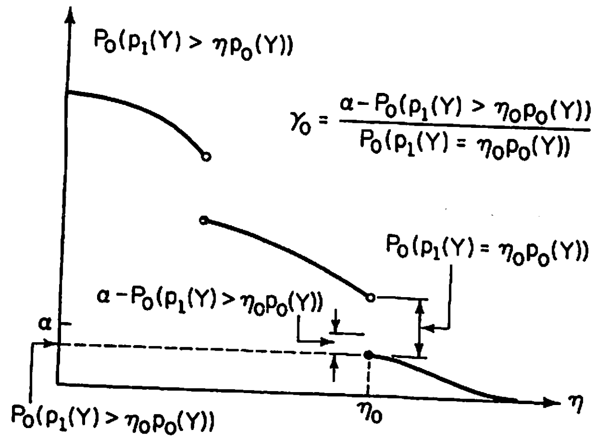


FIGURE II.D.1. Threshold and randomization for an α-level Neyman-Pearson test.

(iii). Suppose that $\tilde{\delta}'$ is an α-level Neyman-Pearson rule of the form (II.D.5) and let $\tilde{\delta}''$ be any other α-level Neyman-Pearson rule. Then

$$P_D(\tilde{\delta}'') = P_D(\tilde{\delta}')$$

(otherwise, only one of them is NP).

We still have

$$\eta(\alpha - P_F(\tilde{\delta}'')) = \eta(P_F(\tilde{\delta}') - P_F(\tilde{\delta}''))$$

$$\leq P_D(\tilde{\delta}') - P_D(\tilde{\delta}'')$$

as before in (i).

$\Rightarrow \alpha - P_F(\tilde{\delta}'') \le 0$

$\Rightarrow P_F(\tilde{\delta}'') = \alpha$, i.e., $\tilde{\delta}''$ must be of size α.

going back $\Rightarrow \int_\Gamma [\tilde{\delta}'(y) - \tilde{\delta}''(y)][p_1(y) - \eta p_0(y)]\, \mu(dy) = 0$

Since $[\tilde{\delta}'(y) - \tilde{\delta}''(y)][p_1(y) - \eta p_0(y)] \ge 0, \forall y \in \Gamma$ as in (i), we have

$(\tilde{\delta}'(y) - \tilde{\delta}''(y))(p_1(y) - \eta p_0(y)) = 0 \quad \forall y \in \Gamma$
except on a zero probability (measure) under H_1 and H_0.

$\Rightarrow \tilde{\delta}''$ and $\tilde{\delta}''$ differ only on the set
$\{y \in \Gamma \mid p_1(y) = \eta p_0(y)\}$

$\Rightarrow \tilde{\delta}''$ is also of the form (II.D.5), possibly differing from $\tilde{\delta}'$ only in the function $\gamma(y)$.

(iii) is, thus, proved.

$\square$

* This result tells us that the optimality of the likelihood ratio test again. The Neyman-Pearson test for a given hypothesis pair differs from the Bayes and minimax tests only in the choice of the threshold and randomization.

Example I: Location Testing with Gaussian Error

$$P_0(p_1(y) > \eta\, p_0(y)) = P_0(L(y) > \eta) = P_0(y > \eta')$$
$$= 1 - \Phi\left(\frac{\eta' - \mu_0}{\sigma}\right)$$

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

Since, for any value α, $0 < \alpha < 1$, there exists η_0', thus η_0, such that

$$P_0(p_1(y) > \eta_0\, p_0(y)) = \alpha$$

with $\eta_0' = \sigma\, \Phi^{-1}(1-\alpha) + \mu_0$, and

$$P(y = \eta_0') = 0.$$

from the proof of Neyman–Pearson Lemma, the randomization γ_0 can be chosen arbitrarily, say $\gamma_0 = 1$.

An α-level Neyman–Pearson test, thus, is

$$\tilde{\delta}_{NP}(y) = \begin{cases} 1 & \text{if } y \geq \eta_0' \\ 0 & \text{if } y < \eta_0' \end{cases}$$

The detection probability of $\tilde{\delta}_{NP}$ is

$$P_D(\tilde{\delta}_{NP}) = E_1(\tilde{\delta}_{NP}(y)) = P_1(y > \eta_0')$$

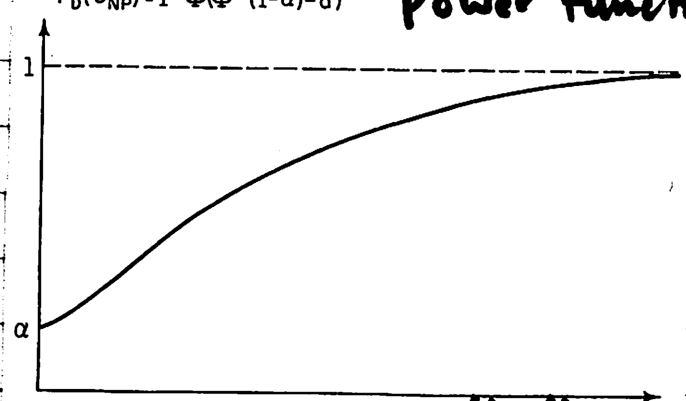
$$= 1 - \Phi\left(\frac{\eta_0' - \mu_1}{\sigma}\right) = 1 - \Phi\left(\Phi^{-1}(1-\alpha) - d\right)$$

where $d = \dfrac{\mu_1 - \mu_0}{\sigma}$ is the SNR.

For a fixed α, the P_D is a function of the SNR d. It is called the power function of the test

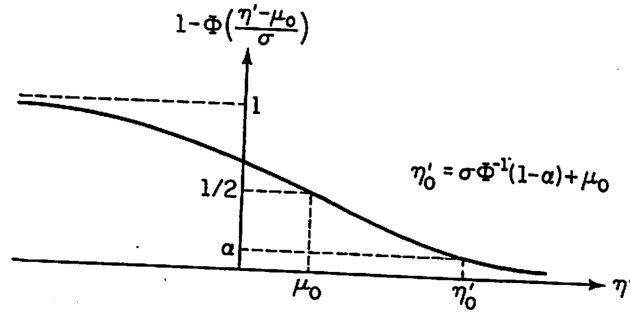
A parametric plot of this relationship is called the receiver operating characteristics (ROCs)

Power funct



$d = \dfrac{\mu_1 - \mu_0}{\sigma}$

II.D.3. Power function for Neyman-Pearson testing of locat



II.D.2. Illustration of threshold (η_0') for Neyman-Pearson test ith Gaussian error.

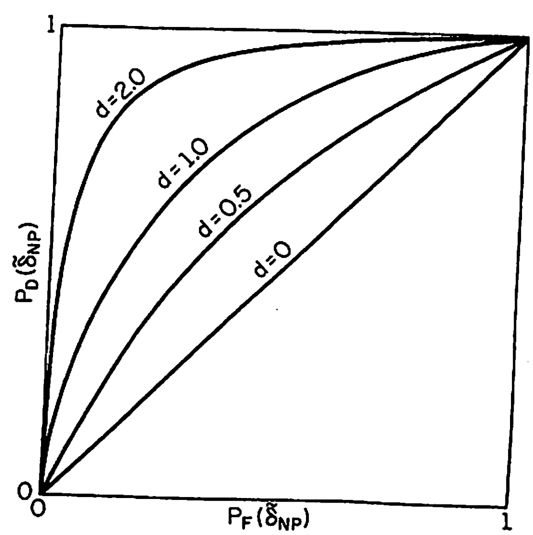


FIGURE II.D.4. Receiver operating characteristics (ROCs) for Neyman-Pearson location testing with Gaussian error [$d = (\mu_1 - \mu_0)/\sigma$].

Example II: The Binary Channel.

In the previous example, the probability density function is continuous and thus the randomization is not necessary:

$$P_0(p_1(Y) > \eta_0 p_0(Y)) \leq \alpha$$

$$\Leftrightarrow \quad P_0(p_1(Y) > \eta_0 p_0(Y)) < \alpha$$

$\Rightarrow$ For the smallest such η_0,

$$P_0(p_1(Y) > \eta_0 p_0(Y)) = \cancel{=}\alpha$$

$\Rightarrow \gamma_0$ is arbitrary

To show the randomization is necessary, we next show the discrete case as before of the binary channel.

Note the threshold η_0 is the smallest number such that $P_0(p_1(Y) > \eta_0 p_0(Y)) \leq \alpha$, i.e.,

$$P_0(L(Y) > \eta_0) \leq \alpha.$$

For simplicity, let us assume $\lambda_0 + \lambda_1 < 1$

$$\Rightarrow \quad \frac{\lambda_1}{1-\lambda_0} < \frac{1-\lambda_1}{\lambda_0}$$

$$\Rightarrow P_0(L(Y) > \eta) = \begin{cases} 1 & \text{if } \eta < \frac{\lambda_1}{1-\lambda_0} \\ \lambda_0 & \text{if } \frac{\lambda_1}{1-\lambda_0} \leq \eta < \frac{1-\lambda_1}{\lambda_0} \\ 0 & \text{if } \eta \geq \frac{1-\lambda_1}{\lambda_0} \end{cases}$$

By inspection, we see that the desired threshold for α-level Neyman-Pearson testing is given by

$$\eta_0 = \begin{cases} \dfrac{1-\lambda_1}{\lambda_0} & \text{if } 0 \le \alpha < \lambda_0 \\ \dfrac{\lambda_1}{1-\lambda_0} & \text{if } \lambda_0 \le \alpha < 1 \\ 0 & \text{if } \alpha = 1 \end{cases}$$

and

$$\gamma_0 = \begin{cases} \dfrac{\alpha}{\lambda_0} & \text{if } 0 \le \alpha < \lambda_0 \\ \dfrac{\alpha-\lambda_0}{1-\lambda_0} & \text{if } \lambda_0 \le \alpha < 1 \\ \text{arbitrary} & \text{if } \alpha = 1. \end{cases}$$

$\Rightarrow$ The resulting Neyman–Pearson test is

$$\tilde{\delta}_{NP}(y) = \begin{cases} \dfrac{\alpha}{\lambda_0} & \text{if } y = 1 \\ 0 & \text{if } y = 0 \end{cases}$$

for $0 \le \alpha < \lambda_0$, and

$$\tilde{\delta}_{NP}(y) = \begin{cases} 1 & \text{if } y = 1 \\ \dfrac{\alpha-\lambda_0}{1-\lambda_0} & \text{if } y = 0 \end{cases}$$

for $\lambda_0 \le \alpha \le 1$.

$$\Rightarrow P_D(\tilde{\delta}_{NP}) = P_1(L(Y) > \eta_0) + \gamma_0 P_1(L(Y) = \eta_0)$$

$$= \begin{cases} \alpha \dfrac{1-\lambda_1}{\lambda_0} & \text{if } 0 \le \alpha < \lambda_0 \\ 1-\lambda_1 + \lambda_1 \dfrac{\alpha - \lambda_0}{1-\lambda_0} & \text{if } \lambda_0 \le \alpha \le 1. \end{cases}$$

When $\lambda_0 = \lambda_1 = \frac{1}{2}$,

$$P_D(\tilde{\delta}_{NP}) = \alpha = P_F(\tilde{\delta}_{NP})$$

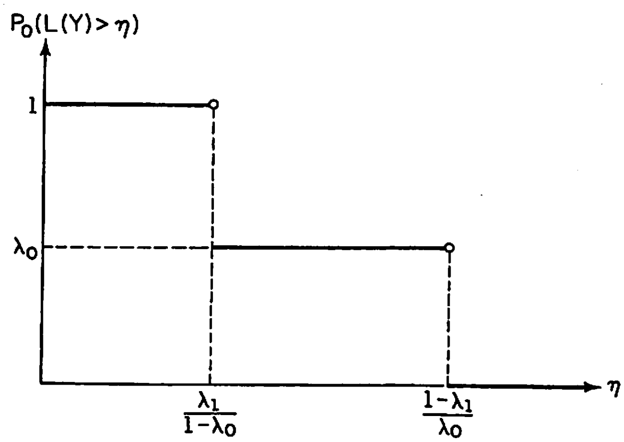
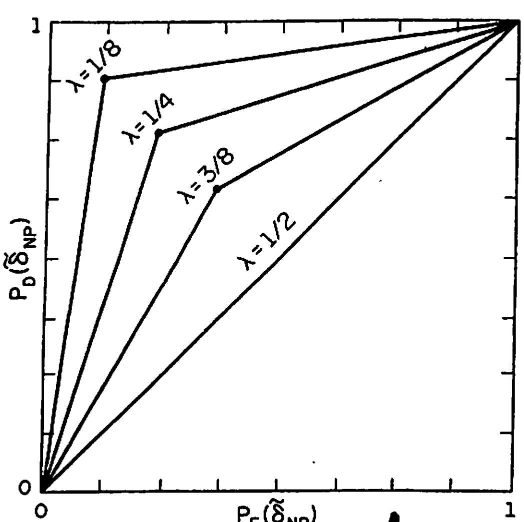


FIGURE II.D.5. Curve for threshold and randomization selection for a binary channel.



α

$\leftarrow \lambda_0 = \lambda_1 = \lambda$

$\alpha = P_F(\tilde{\delta}_{NP})$

FIGURE II.D.6. ROCs for a binary symmetric channel.

Homework

II.F Exercises

1. Find the minimum Bayes risk for the binary channel of Example II.B.1.

2. Suppose Y is a random variable that, under hypothesis H_0, has pdf

$$p_0(y) = \begin{cases} (2/3)(y+1), & 0 \leq y \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

and, under hypothesis H_1, has pdf

$$p_1(y) = \begin{cases} 1, & 0 \leq y \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

   (a) Find the Bayes rule and minimum Bayes risk for testing H_0 versus H_1 with uniform costs and equal priors.

   (b) Find the minimax rule and minimax risk for uniform costs.

   (c) Find the Neyman-Pearson rule and the corresponding detection probability for false-alarm probability $\alpha \in (0,1)$.

3. Repeat Exercise 2 for the situation in which p_j is given instead by

$$p_j(y) = \frac{(j+1)}{2} e^{-(j+1)|y|}, \; y \in \mathbb{R}, j = 0, 1.$$

For parts (a) and (b) assume costs

$$C_{ij} = \begin{cases} 0, & \text{if } i = j \\ 1, & \text{if } i = 1 \text{ and } j = 0 \\ 3/4, & \text{if } i = 0 \text{ and } j = 1, \end{cases}$$

and for part (a) assume priors $\pi_0 = 1/4$ and $\pi_1 = 3/4$.

§4. Multiple Hypothesis Testing

* The binary hypothesis testing can be extended to M-ary hypotheses testing with $M>2$. This can be seen often in communication systems, where M signals, such as M-QAM, are transmitted. It also occurs in pattern recognition with M different patterns. The problem is also called classification or discrimination problem.

* Although an Neyman-Pearson criterion can be formulated for the M-ary hypothesis test, it seems not often used in practice. For more details: Lehmann, E. L., Testing Statistical Hypotheses, New York, John Wiley, 1959

  More commonly used is the Bayes criterion to minimize the average risk.

* Assume that we want to decide among the M possible hypotheses $\{H_0, H_1, \cdots, H_{M-1}\}$. Let $\Gamma_0, \cdots, \Gamma_{M-1}$ be a partition of Γ for a decision rule δ:

$$\Gamma_j = \{y \in \Gamma \mid \delta(y) \text{ chooses } H_j\}.$$

Let $\pi_0, \cdots, \pi_{M-1}$ be the prior probabilities of H_j, i.e., $\pi_j = P(H_j)$.

* The Bayes risk

$$r(\delta, \underline{\Pi}) = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} C_{ij} P_j(\Gamma_i) \pi_j$$

where C_{ij} is the cost

* Let $\pi_j(y) \triangleq P(H_j \text{ true} \mid Y = y) = \dfrac{p_j(y)\pi_j}{p(y)}$

where $p(y) = \sum_{j=0}^{M-1} \pi_j p_j(y)$, Bayes' formula

They are the posterior or a posteriori probabilities of the hypotheses H_j

* $$C_i(y) \triangleq \sum_{j=0}^{M-1} C_{ij} \pi_j(y)$$

is the average cost incurred by choosing hypothesis H_i given $Y = y$

* It can be proved similarly to the Bayes test (binary), that the decision rule that minimizes $r(\delta, \underline{\Pi})$ is to minimizes $C_i(y)$ over $i = 0, 1, \cdots, M-1$.

* Consider the uniform costs:

$$C_{ij} = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases},$$

$$\Rightarrow C_i(y) = \sum_{\substack{j=0 \\ j\neq i}}^{M-1} \pi_j(y)$$

$$= \sum_{j=0}^{M-1} \pi_j(y) - \pi_i(y)$$

$$\Rightarrow \arg\min_{0\le i\le M-1} C_i(y) = \arg\max_{0\le i\le M-1} \pi_i(y)$$

$\Leftrightarrow$ the M-ary maximum a posteriori probability (MAP) decision rule.

* When all the prior probabilities π_j are the same, i.e., $\pi_j = \frac{1}{M}$, $0\le j\le M-1$,

$$\pi_j(y) = \frac{p_j(y)\frac{1}{M}}{p(y)}$$

$$\max_{0\le i\le M-1} \pi_i(y) \Leftrightarrow \max_{0\le i\le M-1} p_i(y)$$

$$= \max_{0\le i\le M-1} p(y|H_i)$$

$\Leftrightarrow$ The M-ary maximum likelihood (ML) decision rule.

$$\boxed{\text{Bayes rule} \xrightarrow[\text{Costs}]{\text{Uniform}} \text{MAP} \xrightarrow[\text{priors}]{\text{Equal}} \text{ML}}$$

equal probable signals

§5. Composite Hypothesis Testing

* What we studied before are called simple hypothesis testing problems because each of the two hypotheses corresponds to only a single distribution for the observation.

In many applications, each of the two hypotheses may correspond to many possible distributions. Hypotheses of this type are called <u>composite hypotheses</u>.

An example of such composite hypotheses is in radar applications, where in the returned signal, if a target present, it has unknown parameters, such as the time of arrivals (related to range) and its Doppler shift (related to velocity).

* To have a general model, we consider a family of probability distributions on Γ indexed by a parameter θ taking values in a parameter set Λ:

$$\{ P_\theta ; \ \theta \in \Lambda \},$$

P_θ is the probability distribution of the observation given that θ is the true parameter value.

The parameter values in Λ represent the set of all possible states of nature. For the simple

hypothesis pair (II.B.1), $\Lambda = \{0, 1\}$. In genera[l]
$\Lambda = \Lambda_0 \cup \Lambda_1$ where $\Lambda_0 \cap \Lambda_1 = \phi$ represent the ranges of the parameter under the two hypotheses.

* For the Bayesian formulation, the parameter Θ is considered a random variable taking values in Λ : $\Theta = \theta$. P_θ is the conditional distribution of Y given $\Theta = \theta$.

We wish to make a binary decision about Θ

* For simplicity, we consider non-randomized decision rules

* Let $C[i, \theta]$ be the cost of choosing decision i when $Y \sim P_\theta$ for $i \in \{0, 1\}$ and $\theta \in \Lambda$.

* For a decision rule δ, define conditional risks

$$R_\theta(\delta) = E_\theta\{C[\delta(y), \theta]\}, \quad \theta \in \Lambda$$

where E_θ denotes expectation assuming that $Y \sim P_\theta$.

* An average or Bayes risk can be defined as

$$r(\delta) = E\{R_\Theta(\delta)\}$$

and a <u>Bayes rule</u> is the one that minimizes $r(\delta)$.

* Since $E_\theta\{C[\delta(Y),\theta]\} = E\{C[\delta(Y),\Theta]\,|\,\Theta=\theta\}$

$$r(\delta) = E[E\{C[\delta(Y),\Theta]\,|\,\Theta\}\} = E\{C[\delta(Y),\Theta]\}$$

$$E(X) = E\{E(X|Y)\}$$

$\Rightarrow$ $r(\delta)$ is simply the cost of using δ averaged over Θ and Y

$\Rightarrow$ Similarly, $r(\delta) = E[E\{C[\delta(Y),\Theta]\,|\,Y\}\}$

$\Rightarrow$ $r(\delta)$ is minimized over δ if for each $y \in \Gamma$, we choose $\delta(y)$ to be the decision that minimizes the posterior cost

$$E\{C[\delta(Y),\Theta]\,|\,Y=y\}$$

* Since $\delta(y)$ is either 0 or 1, a Bayes rule is

$$\delta_B(y) = \begin{cases} 1 & \text{if } E\{C[1,\Theta]\,|\,Y=y\} < E\{C[0,\Theta]\,|\,Y=y\} \\ 0 \text{ or } 1 & \text{if } \quad \cdots \quad = \quad \cdots \\ 0 & \text{if } \quad \cdots \quad > \quad \cdots \end{cases}$$

The interpretation is simple : δ_B chooses the hypothesis that is least costly, on the average, given the observation.

* When $\Lambda = \Lambda_0 \cup \Lambda_1$ and $\Lambda_0 \cap \Lambda_1 = \phi$, Consider the uniform costs:

$$C[i, \theta] = C_{ij}, \quad \theta \in \Lambda_j$$

and assume $C_{11} < C_{01}$. Then,

$$\delta_B(y) = \begin{cases} 1 & \text{if } \dfrac{P(\Theta \in \Lambda_1 | Y = y)}{P(\Theta \in \Lambda_0 | Y = y)} > \dfrac{C_{10} - C_{00}}{C_{01} - C_{11}} \\ 1 \text{ or } 0 & \text{if } \quad .. \quad = \quad .. \\ 0 & \text{if } \quad .. \quad < \quad .. \end{cases}$$

where $P(\Theta \in \Lambda_j | Y = y)$ denotes the conditional probability that Θ lies in Λ_j given $Y = y$.

* Assume Y has conditional densities $p(y|\Theta \in \Lambda_j)$ for $j = 0, 1$.

$$\Rightarrow P(\Theta \in \Lambda_j | Y = y) = \frac{p(y|\Theta \in \Lambda_j) P(\Theta \in \Lambda_j)}{p(y)}, \quad j = 0, 1,$$

with $p(y) = \sum_{j=0}^{1} p(y|\Theta \in \Lambda_j) P(\Theta \in \Lambda_j)$.

$$\Rightarrow \delta_B(y) = \begin{cases} 1 & \text{if } L(y) > \dfrac{\pi_0 (C_{10} - C_{00})}{\pi_1 (C_{01} - C_{11})} \\ 1 \text{ or } 0 & \text{if } \quad .. \quad = \quad .. \\ 0 & \text{if } \quad .. \quad < \quad .. \end{cases}$$

where $\pi_j = P(\Theta \in \Lambda_j)$, priors.

$$L(y) = \frac{p(y \mid \Theta \in \Lambda_1)}{p(y \mid \Theta \in \Lambda_0)}$$ similar to the previous Simple Bayes' rule

* Example 1: Testing on the Radius of a Point in the Plane

Suppose $\Gamma = \mathbb{R}^2$, $Y = (Y_1, Y_2)^T$. The hypotheses are

$$H_0 : \begin{aligned} Y_1 &= \varepsilon_1 \\ Y_2 &= \varepsilon_2 \end{aligned}$$

vs.

$$H_1 : \begin{aligned} Y_1 &= A\cos\Psi + \varepsilon_1 \\ Y_2 &= A\sin\Psi + \varepsilon_2 \end{aligned}$$

where A is a positive constant, Ψ is a random variable distributed uniformly in $[0, 2\pi]$, and ε_1, ε_2 are $\mathcal{N}(0, \sigma^2)$ random variables that are independent of one another and of Ψ.

The observation $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$ can be thought of as a noisy measurement of the coordinates of a point in the plane that is either at the origin or is uniformly distributed on a circle of radius A. (Its applications will be discussed in later Chapters).

The parameter in this case can be taken to be $\Theta = (\theta_1, \theta_2)$ with $\theta_1 \in \{0, A\}$, $\theta_2 \in [0, 2\pi]$.

$$\Lambda = \{0, A\} \times [0, 2\pi]$$

$$\Lambda_0 = \{\theta \in \Lambda \mid \theta_1 = 0\} = \{0\} \times [0, 2\pi]$$

$$\Lambda_1 = \{\theta \in \Lambda \mid \theta_1 = A\} = \{A\} \times [0, 2\pi]$$

The density of Y given $\Theta = \theta$ is the joint density of two independent $N(0, \sigma^2)$ random variables shifted in mean by $\theta_1 \cos\theta_2$ and $\theta_1 \sin\theta_2$, respectively, i.e.,

$$p_\theta(y) = \frac{1}{2\pi\sigma^2} \exp\{-q(y,\theta)/(2\sigma^2)\}, \quad y \in \mathbb{R}^2$$

where $q(y,\theta) \triangleq (y_1 - \theta_1 \cos\theta_2)^2 + (y_2 - \theta_1 \sin\theta_2)^2$

$$\Rightarrow p(y \mid \Theta \in \Lambda_0) = p_\theta(y)\Big|_{\theta=0}$$

$$= \frac{1}{2\pi\sigma^2} \exp\{-(y_1^2 + y_2^2)/(2\sigma^2)\}$$

$$p(y \mid \Theta \in \Lambda_1) = \frac{1}{2\pi} \int_0^{2\pi} p_\theta(y)\Big|_{\theta_1 = A} d\theta_2$$

$$= \frac{1}{4\pi^2\sigma^2} \int_0^{2\pi} \exp\{-q(y,\theta)\big|_{\theta_1 = A} / (2\sigma^2)\} \, d\theta_2$$

$$\Rightarrow L(y) = \frac{p(y|\Theta \in \Lambda_1)}{p(y|\Theta \in \Lambda_0)}$$

$$= \frac{e^{-A^2/(2\sigma^2)}}{2\pi} \int_0^{\pi} \exp\left\{\frac{A}{\sigma^2}(y_1 \cos\theta_2 + y_2 \sin\theta_2)\right\} d\theta_2$$

Let $r = [y_1^2 + y_2^2]^{\frac{1}{2}}$

$$\phi = \tan^{-1}\left(\frac{y_2}{y_1}\right)$$

$$\Rightarrow y_1 = r\cos\phi, \quad y_2 = r\sin\phi$$

$$\Rightarrow L(y) = \frac{e^{-\frac{A^2}{2\sigma^2}}}{2\pi} \int_0^{2\pi} \exp\left\{\frac{Ar}{\sigma^2}\cos(\theta_2 - \phi)\right\} d\theta_2$$

$$= e^{-\frac{A^2}{2\sigma^2}} I_0\left(\frac{Ar}{\sigma^2}\right)$$

where I_0 is the zeroth-order modified Bessel function of the first kind.

$I_0(x)$ is monotone increasing

$$L(y) \underset{<}{\overset{>}{=}} \tau \Leftrightarrow r \underset{<}{\overset{>}{=}} \tau'$$

where $\tau' = \sigma^2 I_0^{-1}\left(\tau e^{A^2/(2\sigma^2)}\right)/A$.

$\Rightarrow$ The Bayes, minimax, and Neyman-Pearson tests are of the form

$$\tilde{\delta}_0(y) = \begin{cases} 1 & \text{if } r > \tau' \\ \gamma & \text{if } r = \tau' \\ 0 & \text{if } r < \tau' \end{cases}$$

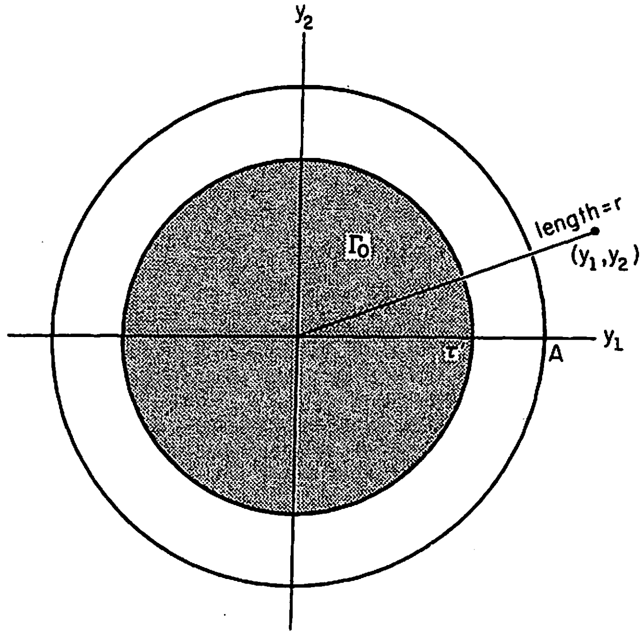


FIGURE II.E.1. Decision regions for Example II.E.1 ($\Gamma_1 = \Gamma_0^c$).

* If there is no available distribution for the parameter, analytical solutions for the other optimal criteria are not easy.

In the case of Neyman-Pearson testing, let $\Lambda = \Lambda_0 \cup \Lambda_1$ for two disjoint sets Λ_0 and Λ_1. For a randomized decision rule $\tilde{\delta}$, we define false-alarm and detection probabilities as follows

$$P_F(\tilde{\delta};\theta) = E_\theta\{\tilde{\delta}(Y)\}, \quad \theta \in \Lambda_0$$

$$P_D(\tilde{\delta};\theta) = E_\theta\{\tilde{\delta}(Y)\}, \quad \theta \in \Lambda_1$$

* Suppose, as in the Neyman-Pearson formulation, that we wish to be assumed that the false-alarm probability does not exceed a given value, α. Then, an ideal test would be the one to maximinize $P_D(\tilde{\delta};\theta)$ for every $\theta \in \Lambda_1$ subject to $P_F(\tilde{\delta};\theta) \le \alpha$, $\theta \in \Lambda_0$.

Such a test is called a uniformly most powerful (UMP) test of level α.

* Although UMP tests are desirable, they usually do not exist. The problem is to maximize $P_D(\tilde{\delta};\theta)$ for all parameters $\theta \in \Lambda_1$ with a single $\tilde{\delta}$, which is not easy.

To see this, let us consider the case when Λ_0 is simple, i.e. $\Lambda_0 = \{\theta_0\}$ for some θ_0. Assume P_θ has density p_θ for $\theta \in \Lambda$. For each $\theta \in \Lambda_1$, the most powerful α-level test for H_0 vs. $Y \sim P_\theta$ has critical region

$$\Gamma_\theta = \{y \in \Gamma \mid p_\theta(y) > \tau p_{\theta_0}(y)\}$$

where τ (possibly a randomization) chosen to give size α.

From the Neyman-Pearson Lemma, we know that this test is essentially unique so that any other α-level test will have smaller power.

Let us choose two elements θ' and θ'' of Λ_1. Then the test with critical region $\Gamma_{\theta'}$ will have smaller power in testing H_0 vs. $Y \sim P_{\theta''}$ than does test with critical region $\Gamma_{\theta''}$ (and vice versa) unless these two critical regions are essentially identical. Thus, the UMP test exists if and only if the critical region Γ_θ is the same for all $\theta \in \Lambda_1$. This is a very strong condition.

* Example 2: UMP Testing of Location

Consider the parametric family of distributions $\{P_\theta ; \theta \in \Lambda\}$ where $P_\theta \sim \mathcal{N}(\theta, \sigma^2)$ and $\Lambda \subseteq \mathbb{R}$.

The hypothesis pair is

$$H_0 : \theta = \mu_0$$

vs.

$$H_1 : \theta > \mu_0$$