Lecture 2: Minimax Hypothesis Testing

14 Jan 2016

1 Continued from the Lecture 1

At the onset, we will look into two examples of Bayesian Hypothesis Testing.

Example 1.1 (The Binary Channel). A binary channel is the most common communication channel model used in coding theory and information theory. In this model, a transmitter sends a bit (0 or 1), and the receiver receives it. The bit may be received correctly, or it may be "flipped" with some probability. The probability with which a flipped bit is received is known as "crossover probability".

Consider transmission of a bit over such a Binary Channel. Let the observation at the output of the channel be Y, which can be either 0, or 1. Let the crossover probability be λ_0 when bit 0 is transmitted, i.e., a transmitted 0 is received as 1 with probability λ_0 and as 0 with probability (1 - λ_0), where $0 \leq \lambda_0 \leq 1$. Similarly, let the crossover probability when a 1 is transmitted be λ_1. A Binary Symmetric Channel (BSC) is a special case, where $\lambda_0 = \lambda_1 = \lambda$. Observing Y does not tell us exactly whether the transmitted digit was bit 0 or 1. The goal is to find an optimum decision rule using the Bayesian Hypothesis Testing.

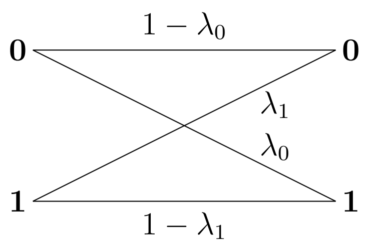

Figure 1: Block diagram for Binary Channel

The two hypothesis H_0, and H_1 depict the transmission of bit 0 and 1 respectively. The observation set is $\Gamma = \{0, 1\}$. The received signal $y \in \Gamma$ is an instance of a Bernoulli random variable Y with probability mass function (pmf) dependent on the transmitted bit,

$$\begin{aligned} Y_0 &\sim \mathcal{B}\left(1-\lambda_0\right) \text{ if } H_0 \text{ is transmitted}, \\ Y_1 &\sim \mathcal{B}\left(1-\lambda_1\right) \text{ if } H_1 \text{ is transmitted}, \end{aligned} \tag{1}$$

where the notation $\mathcal{B}(\lambda)$ denotes the pmf of a Bernoulli random variable p with parameter λ. The pmf of the observation Y can be written compactly as,

$$p_j\left(y\right) = \begin{cases} \lambda_j, & \text{if } y \neq j, \\ \left(1-\lambda_j\right), & \text{if } y = j, \end{cases} \qquad j \in \{0,1\}. \tag{2}$$

The corresponding likelihood ratio is,

$$L(y) = \frac{p_1(y)}{p_0(y)} = \begin{cases} \frac{\lambda_1}{1-\lambda_0} & \text{if } y = 0, \\ \frac{1-\lambda_1}{\lambda_0} & \text{if } y = 1. \end{cases} \tag{3}$$

As discussed in last lecture, a Bayesian decision rule has the form,

$$\delta_B(y) = \mathbb{1}_{\{L(y) \geq \tau\}}, \tag{4}$$

where $\tau = \frac{\pi_0(C_{10}-C_{00})}{\pi_1(C_{01}-C_{11})}$ is a threshold, which depends on the prior probability π_0 and the costs. If λ_0, λ_1, τ are such that $\lambda_1 \geq \tau\left(1-\lambda_0\right)$, then according to the likelihood ratio, a received 0 is decided as transmitted 1. Similarly a received 1 is decided as a transmitted 0 when $(1-\lambda_1) < \tau\lambda_0$.

Consider the simple case of uniform cost, for which, $C_{ij} = 0$ if $i = j$ and $C_{ij} = 1$ if $i \neq j$, and equal priors i.e., $\pi_0 = 1/2$. In this case, $\tau = 1$, and the Bayesian decision rule is given as,

$$\begin{aligned} \delta_B(0) &= \begin{cases} 1 \text{ if } (1-\lambda_1) < \lambda_0, \\ 0 \text{ if } (1-\lambda_1) \geq \lambda_0, \end{cases} \\ \delta_B(1) &= \begin{cases} 1 \text{ if } (1-\lambda_1) \geq \lambda_0, \\ 0 \text{ if } (1-\lambda_1) < \lambda_0. \end{cases} \end{aligned} \tag{5}$$

This can be written in a compact form as,

$$\delta_B(y) = \begin{cases} y, & \text{if } (1-\lambda_1) \geq \lambda_0 \\ \left(1-y\right), & \text{if } (1-\lambda_1) < \lambda_0. \end{cases} \tag{6}$$

For a BSC with $(\lambda_1 = \lambda_0 = \lambda)$,

$$\delta_B(y) = \begin{cases} y, & \text{if } \lambda \leq 0.5 \\ \left(1-y\right), & \text{if } \lambda > 0.5. \end{cases} \tag{7}$$

For the optimal Bayes rule described above, conditional risks have the following expressions,

$$R_0(\delta) = \begin{cases} \lambda_0 & \text{if } (1-\lambda_1) \geq \lambda_0, \\ (1-\lambda_0) & \text{if } (1-\lambda_1) < \lambda_0, \end{cases} \tag{8}$$

and,

$$R_1(\delta) = \begin{cases} \lambda_1 & \text{if } (1-\lambda_1) \geq \lambda_0, \\ (1-\lambda_1) & \text{if } (1-\lambda_1) < \lambda_0. \end{cases} \tag{9}$$

The unconditional risk can be obtained as a weighted sum of the conditional risks. For a BSC, the expression for unconditional risk can be simplified to,

$$r(\delta) = \min(\lambda, 1-\lambda). \tag{10}$$

Example 1.2 (Location Testing with Gaussian Error). Consider the typical communication model denoted by the equation,

$$y = x + n, \tag{11}$$

where n is a white noise signal with mean zero, and variance σ^2. The null hy-

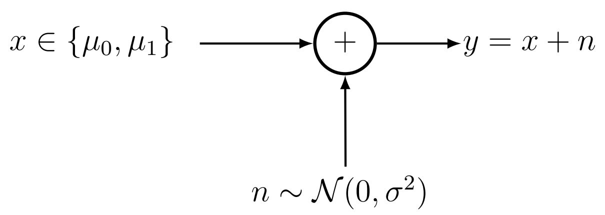

Figure 2: AWGN channel

pothesis (H_0) corresponds to the reception of a signal y with mean μ_0 and under alternative Hypothesis (H_1), y has mean μ_1.

$$H_0 \ : Y \sim \mathcal{N}\left(\mu_0, \sigma^2\right), \tag{12}$$
$$H_1 \ : Y \sim \mathcal{N}\left(\mu_1, \sigma^2\right). \tag{13}$$

The corresponding observation space is $\Gamma = \mathcal{R}$. Assuming $\mu_1 > \mu_0$, we have the following expression for the likelihood ratio $L(y)$,

$$
\begin{aligned}
L(y) = \frac{P_1(y)}{P_0(y)} &= \frac{\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-(y-\mu_1)^2}{2\sigma^2}]}{\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-(y-\mu_0)^2}{2\sigma^2}]} \\
&= \exp\left[\frac{-(y-\mu_1)^2 + (y-\mu_0)^2}{2\sigma^2}\right] \\
&= \exp\left[\frac{(\mu_1-\mu_0)\left(y - \frac{\mu_1+\mu_0}{2}\right)}{\sigma^2}\right]. \qquad (14)
\end{aligned}
$$

For uniform cost and equal priors, $\tau = 1$ and $\Gamma_1 = \{y \in \Gamma | \ L(y) \geq 1\}$. From eqn. (14), we get,

$$
\exp\left[\frac{(\mu_1-\mu_0)\left(y - (\frac{\mu_1+\mu_0}{2})\right)}{\sigma^2}\right] \geq 1, \qquad (15)
$$

$$
\frac{(\mu_1-\mu_0)\left(y - (\frac{\mu_1+\mu_0}{2})\right)}{\sigma^2} \geq 0.
$$

In terms of y, we can write the decision region as,

$$
\Gamma_1 = \left\{ y \in \Gamma : \ y \geq \frac{\mu_1+\mu_0}{2} \right\} \qquad (16)
$$

Thus, the decision rule in this case will be

$$
\delta(y) = \begin{cases} 1 \text{ if } y \geq \frac{\mu_1+\mu_0}{2} \\ 0 \text{ if } y < \frac{\mu_1+\mu_0}{2} \end{cases} \qquad (17)
$$

The corresponding conditional risks are,

$$
\begin{aligned}
R_0(\delta) = P_0(\Gamma_1) = \int_{\tau'}^{\infty} dP_0(x) = 1 - \Phi\left(\frac{\tau'-\mu_1}{\sigma}\right) \\
R_1(\delta) = P_1(\Gamma_0) = \int_{-\infty}^{\tau'} dP_1(x) = \Phi\left(\frac{\tau'-\mu_0}{\sigma}\right) \qquad (18)
\end{aligned}
$$

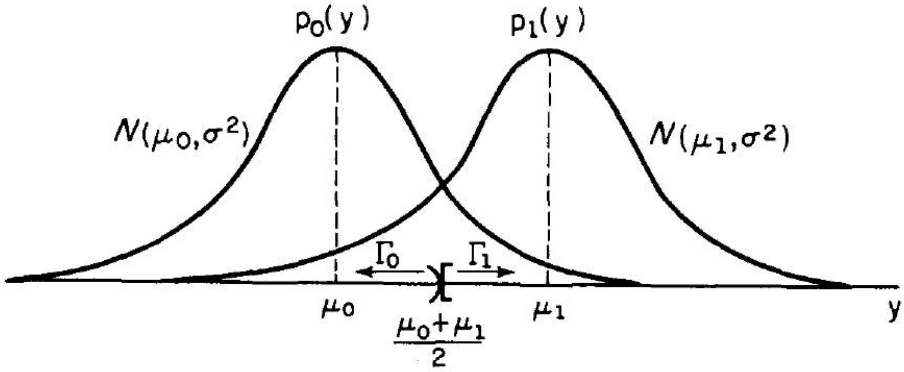

Figure 3: Illustration of decision regions for the uniform cost and equal prior case.

2 Minimax Hypothesis Testing

Bayesian hypothesis testing assumes the knowledge of prior probabilities for the hypotheses. However, in a practical scenario, the priors are not necessarily available at the receiver. Under such circumstances, it is not possible to design a single Bayesian decision criterion that minimizes the average risk or Bayes risk for all possible prior distributions. Hence, it is necessary to develop a separate design criterion. In this section, we look at the "Minimax criterion", which considers the minimization of the maximum of conditional risks $R_0(\delta)$ and $R_1(\delta)$ over all possible decision rules δ.

Definition 2.1. The decision rule (δ) minimizing the max risk given by the expression max $\{R_0(\delta), R_1(\delta)\}$ is known as **Minimax Rule.**

2.1 The Minimax Rule

To derive the Minimax rule, we first consider the unconditional risk for a given decision rule δ and a given prior for H_0, i.e $\pi_0 \in [0,1]$. The average risk for a decision rule δ is,

$$r(\pi_0, \delta) = \pi_0 R_0(\delta) + (1-\pi_0) R_1(\delta), \quad \pi_0 \in [0,1]. \tag{19}$$

The unconditional risk function is shown in Fig. 4. As can be seen, for a fixed δ, the function $r(\pi_0, \delta)$ is a straight line taking values $R_1(\delta)$ at $\pi_0 = 0$ and $R_0(\delta)$ at $\pi_0 = 1$. Thus, it is an affine function, and hence attains maximum value at the extremities,

$$\max_{0 \le \pi_0 \le 1} r(\pi_0, \delta) = \max\{R_0(\delta), R_1(\delta)\}. \tag{20}$$

We can state the minimax criterion as the minimizer of the expression in eqn. (20) over all δ,

$$\min_{\delta} \max_{0 \le \pi_0 \le 1} r(\pi_0, \delta). \tag{21}$$

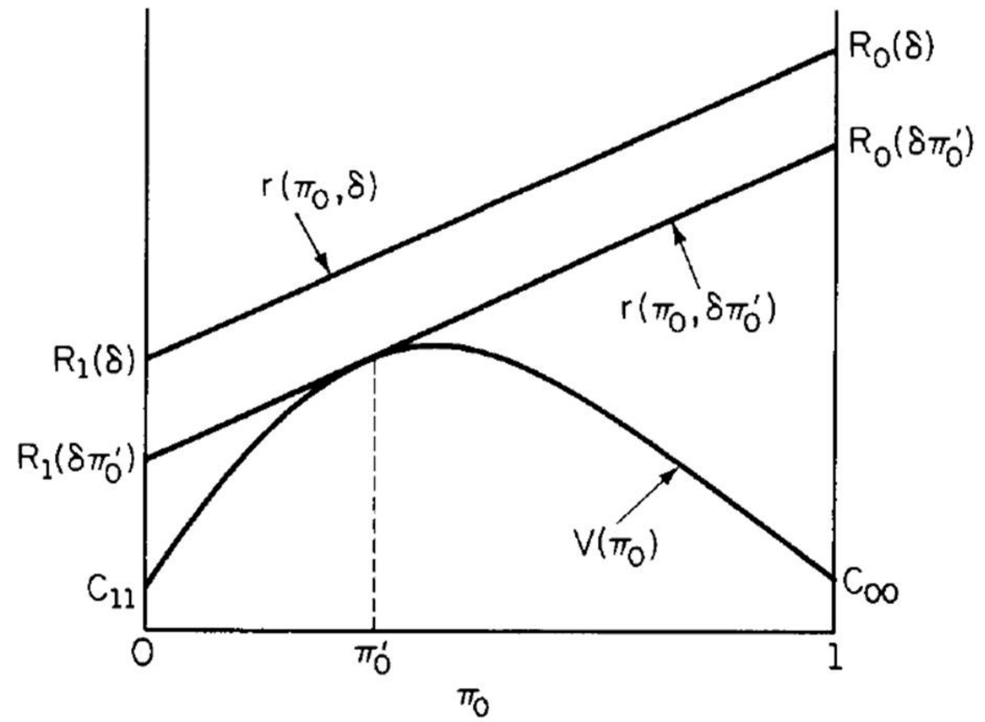

Figure 4: Illustration of the functions $r(\pi_0, \delta)$ *and* $V(\pi_0)$

Now, for each prior $\pi_0 \in \ [0,1]$, let δ_{π_0} denote the optimum Bayes rule corresponding to that prior, and let $V(\pi_0) = r(\pi_0, \delta_{\pi_0})$, be the Bayes risk for the prior π_0. It can be proved that $V(\pi_0)$ is a continuous concave function of π_0 for $\pi_0 \in [0,1]$ with $V(0) = C_{11}$ *and* $V(1) = C_{00}$. The proof is given below.

Lemma 2.2. *The function* $V(\pi) : [0,1] \to \mathcal{R}$ *is concave*

A function is concave, if, for any $\{x, y\}$ in the domain of f and any $\alpha \in [0,1]$,

$$f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha)f(y).$$

Proof. Consider two priors π, π', and a third prior $\pi'' = \alpha\pi + (1-\alpha)\pi'$. We can write,

$$\begin{aligned} V(\pi'') &= r(\pi'', \delta_{\pi''}), \\ &= \alpha r(\pi, \delta_{\pi''}) + (1-\alpha) r(\pi', \delta_{\pi''}). \end{aligned} \tag{22}$$

Since $V(\pi) = r(\pi, \delta_\pi)$ is the minimizer of $r(\pi, \delta)$, we get,

$$V(\pi'') \geq \alpha V(\pi) + (1-\alpha)V(\pi'). \tag{23}$$

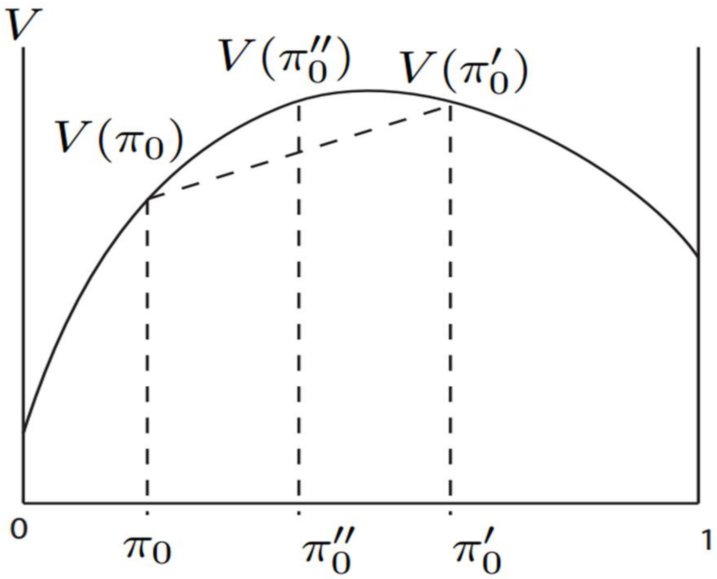

Figure 5: $V(\pi)$ as a concave function

Hence $V(\pi)$ is concave. □

Suppose that $V(\pi_0)$ and $r(\pi_0, \delta)$ are as depicted in Fig. 4. Also shown in Fig. 4 is the line labelled $r(\pi_0, \delta_{\pi_0'})$, that is both parallel to $r(\pi_0, \delta)$ as well as tangent to $V(\pi_0)$. For this case, δ cannot be the minimax rule because the risk line shown as $r(\pi_0, \delta_{\pi_0'})$ lies completely below $r(\pi_0, \delta)$ and thus has a smaller maximum value. Since $r(\pi_0, \delta_{\pi_0'})$ touches $V(\pi_0)$ at $\pi_0 = \pi_0'$, $\delta_{\pi_0'}$ is a Bayes Rule for the prior π_0'. Since a similar tangent line can be drawn for any decision rule δ, it is easily seen that only Bayes Rules can possibly be Minimax rules for Fig. 4.

Moreover, by examination of Fig. 6, we see that the Minimax rule for this case is a Bayes rule corresponding to the prior value π_L that maximizes $V(\pi_0)$ over $\pi_0 \in [0, 1]$. Note that for this prior we have that $r(\pi_0, \delta_{\pi_L})$ is constant over π_0, so,

$$\max\{R_0(\delta_{\pi_L}), R_1(\delta_{\pi_L})\} = R_0(\delta_{\pi_L}) = R_1(\delta_{\pi_L}) \tag{24}$$

The fact that δ_{π_L} is minimax follows from the Fig. 6, since if $\pi_0' < \pi_L$, we have $\max\{R_0(\delta_{\pi_0'}), R_1(\delta_{\pi_0'})\} = R_0(\delta_{\pi_0'}) > R_0(\delta_{\pi_L})$, and if $\pi_0'' > \pi_L$, we have that $\max\{R_0(\delta_{\pi_0''}), R_1(\delta_{\pi_0''})\} = R_1(\delta_{\pi_0''}) > R_1(\delta_{\pi_L})$, as depicted. Because π_L maximizes the minimum Bayes risk, it is also called the **least-favorable prior**. Hence, a minimax decision rule is the Bayes rule for the least-favorable prior.

Proposition 2.3. *Suppose π_L maximizes $V(\pi_0)$ for $\pi_0 \in [0, 1]$. Suppose 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.*

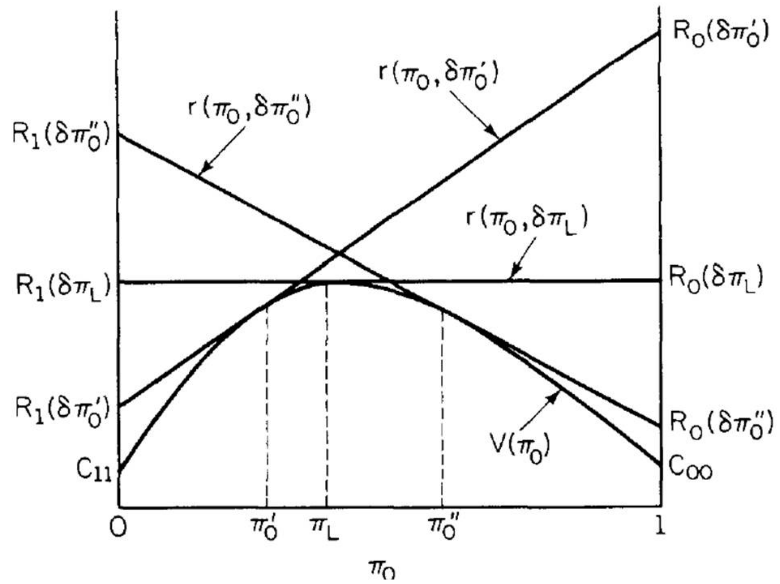

Figure 6: *Illustration of the Minimax Rule when V has an interior maximum*

Proof. Consider the case when $R_1(\delta_{\pi_L}) = R_0(\delta_{\pi_L})$. We know that,

$$V(\pi_L) = \max_{\pi_0 \in [0,1]} \min_{\delta} r(\pi_0, \delta) = r(\pi_L, \delta_{\pi_L}) = r(\pi_0, \delta_{\pi_L}). \tag{25}$$

The second equality follows from the fact that $r(\pi_0, \delta_{\pi_L})$ is a constant in π_0.

$$\begin{aligned} \max_{\pi_0 \in [0,1]} \min_{\delta} r(\pi_0, \delta) &= \max_{\pi_0 \in [0,1]} r(\pi_0, \delta_{\pi_L}), \\ &\geq \min_{\delta} \max_{\pi_0 \in [0,1]} r(\pi_0, \delta). \end{aligned} \tag{26}$$

For every δ, we note that,

$$\max_{\pi_0 \in [0,1]} r(\pi_0, \delta) \geq \max_{\pi_0 \in [0,1]} \min_{\delta} r(\pi_0, \delta). \tag{27}$$

which shows that,

$$\min_{\delta} \max_{\pi_0 \in [0,1]} r(\pi_0, \delta) \geq \max_{\pi_0 \in [0,1]} \min_{\delta} r(\pi_0, \delta). \tag{28}$$

Combining eqns. (26), (28), we see that,

$$\min_{\delta} \max_{\pi_0 \in [0,1]} r(\pi_0, \delta) = \max_{\pi_0 \in [0,1]} \min_{\delta} r(\pi_0, \delta). \tag{29}$$

Indeed we have shown that,

$$r(\pi_L, \delta_{\pi_L}) = \min_{\delta} \max_{\pi_0 \in [0,1]} r(\pi, \delta), \tag{30}$$

which was to be shown.

For $\pi_L = 0$, we note that, $\max_{\pi_0 \in [0,1]} r(\pi_0, \delta_{\pi_L}) = R_1(\delta_{\pi_L}) = r(\pi_L, \delta_{\pi_L})$. Using a similar argument as above, we can note that δ_{π_L} is a minimax rule. Similar argument can be made for $\pi_L = 1$ case. This completes the proof. $\square$

Now, for any $\pi'_0 \in [0,1]$, $r(\pi, \delta_{\pi'_0}) \geq V(\pi)$ since $V(\pi)$ minimizes Bayes risk for all δ. Also $r(\pi, \delta_{\pi'_0})$ is a straight line tangent to V at $\pi = \pi'_0$. Hence, if V is differentiable,

$$\begin{aligned} V'(\pi'_0) &= \frac{d}{d\pi} r(\pi, \delta_{\pi'_0}) \Big|_{\pi=\pi'_0}, \\ &= R_0(\delta_{\pi'_0}) - R_1(\delta_{\pi'_0}). \end{aligned} \tag{31}$$

If V has an interior maximum, i.e., $\pi_L \in (0,1)$, then $V'(\pi_L)$ equals zero, if V is differentiable at π_L.

Thus, under the condition of unknown priors, Minimax rule considers the worst case scenario by taking the least favorable prior π_L into account and minimizes the maximum unconditional risk for that prior.