Lecture 4: Examples

21 Jan 16

1 Continued from the Lecture 3

The following is the summary of Bayesian, Minimax and Neyman Pearson hypothesis testing:

1.1 Bayesian Hypothesis Testing

Consider the binary hypothesis testing scenario, which has two possible hypotheses H_0 and H_1, corresponding to two possible probability distributions P_0 and P_1, respectively on the observation set (Γ). This problem is written as,

$$\begin{aligned} H_0 &: Y \sim P_0, \\ H_1 &: Y \sim P_1. \end{aligned} \tag{1}$$

The decision rule δ is a function on Γ, given by,

$$\delta(y) = \mathbb{1}_{\{y \in \Gamma_1\}}. \tag{2}$$

We define expected cost incurred by decision rule δ when hypothesis H_j is true as,

$$R_j(\delta) = C_{1j}P_j(\Gamma_1) + C_{0j}P_j(\Gamma_0), \tag{3}$$

where Γ_0 is the rejection region, and Γ_1 is the acceptance region. The Bayes risk or the overall cost incurred by decision rule δ is given by,

$$\begin{aligned} r(\delta) &= \pi_0 R_0(\delta) + \pi_1 R_1(\delta), \\ &= \pi_0 R_0(\delta) + (1-\pi_0) R_1(\delta), \end{aligned} \tag{4}$$

where π_0 and π_1 are known as the priori probabilities of the two hypotheses H_0 and H_1 respectively.

A commonly used cost assignment is the uniform cost given by

$$C_{ij} = \begin{cases} 0, & \text{if} \quad i = j, \\ 1, & \text{if} \quad i \neq j, \end{cases} \tag{5}$$

and the corresponding conditional risks are given by,

$$R_0(\delta) = P_0(\Gamma_1), \quad \text{and} \quad R_1(\delta) = P_1(\Gamma_0).$$

1.2 Minimax Hypothesis Testing

The minimax criterion is given by,

$$\min_{\delta} \max(R_0(\delta), R_1(\delta)). \tag{6}$$

Or equivalently,

$$\begin{aligned} \min_{\delta} \max_{0 \leq \pi_0 \leq 1} r(\pi_0, \delta) &= \max_{0 \leq \pi_0 \leq 1} \min_{\delta} r(\pi_0, \delta), \\ &= \max_{0 \leq \pi_0 \leq 1} V(\pi_0) \end{aligned} \tag{7}$$

where $V(\pi_0) = \min_{\delta} r(\pi_0, \delta)$. The Minimax rule is achieved where π_0 is such that

$$R_0(\delta_{\pi_0}) = R_1(\delta_{\pi_0}) \tag{8}$$

1.3 Neyman-Pearson Hypothesis Testing

The design criterion for Neyman-Pearson hypothesis testing is,

$$\max_{\delta} P_D(\delta) \text{ subject to } P_F(\delta) \leqslant \alpha, \tag{9}$$

where $P_D(\delta)$ is the probability of correct detection and $P_F(\delta)$ which is the probability of false alarm and upper bounded by α. The randomized decision rule is written as,

$$\tilde{\delta}(y) = \begin{cases} 1, & L(y) > \eta, \\ \gamma(y), & L(y) = \eta, \\ 0, & L(y) < \eta, \end{cases} \tag{10}$$

$$\therefore \tilde{\delta}(y) = \mathbb{1}_{\{L(y) > y\}} + \gamma(y)\mathbb{1}_{\{L(y) = y\}}. \tag{11}$$

where $\tilde{\delta}$ is interpreted as the conditional probability with which we accept H_1 for a given observation $Y = y$, $L(y) = \frac{p_1(y)}{p_0(y)}$ is the likelihood function, $\eta \geq 0$ is a certain threshold, and $0 \leq \gamma(y) \leq 1$. with $\eta = \eta_0$ and $\gamma(y) = \gamma_0$, we have,

$$\eta_0 = \inf\{\eta \in \mathbb{R} : P_0\{L(y) > \eta\} \leqslant \alpha\}, \tag{12}$$

$$\gamma_0 = \frac{\alpha - P_0\{L(y) > \eta\}}{P_0\{L(y) = \eta\}}. \tag{13}$$

$P_0(L(y) > \eta)$ as a function of η is shown in figure 1. This can be interpreted as the *complementary distribution function* of the likelihood function and hence right continuous and may have discontinuity. From figure 1, it is clear that $0 \leq \alpha - P_0\{L(y) > \eta\} \leq P_0\{L(y) = \eta\}$ and hence $0 \leq \gamma_0 \leq 1$.

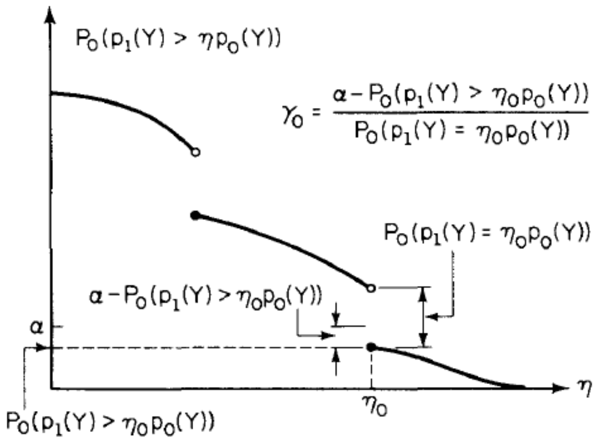

Figure 1: Threshold and randomization for α level Neyman-Pearson test

Example 1.1 (Location testing with Gaussian error). Consider the following problem where we have a real-valued measurement Y, which is corrupted with Gaussian noise (n) having zero mean and standard deviation σ. Here the observation space is real line $\Gamma = \mathcal{R}$.

$$Y = X + n, \tag{14}$$

where $X \in \{\mu_0, \mu_1\}$ is the original signal and $n \sim \mathcal{N}(0, \sigma^2)$. In this example, 'null hypothesis' (H_0) indicates the transmission of signal with mean μ_0 and alternative hypothesis (H_1) indicates transmission of signal with mean μ_1. Without loss of generality, let us assume $\mu_1 > \mu_0$.

$$\begin{aligned} H_0 &: Y \sim \mathcal{N}\left(\mu_0, \sigma^2\right), \\ H_1 &: Y \sim \mathcal{N}\left(\mu_1, \sigma^2\right), \end{aligned} \tag{15}$$

where $\mathcal{N}(\mu_0, \sigma^2)$ is Gaussian distribution with mean μ and variance σ^2. The probability density function has the form, $Pr(X = x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right)$.

Bayesian Hypothesis testing

The likelihood function is given by,

$$L(y) = \frac{p_1(y)}{p_0(y)} = \frac{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_1)^2}{\sigma^2}\right)}{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_0)^2}{\sigma^2}\right)},$$
$$= \exp\left(\frac{\mu_1 - \mu_0}{\sigma^2}\left(y - \frac{\mu_1 + \mu_0}{2}\right)\right). \quad (16)$$

The Bayes rule is given by

$$\delta_B(y) = \mathbb{1}_{\{L(Y) > \tau\}} \quad (17)$$

Where τ is the appropriate threshold expressed in terms of prior probability of Null Hypothesis π_0 as $\tau = \frac{\pi_0}{1-\pi_0}$ (in the case of uniform cost structure). Equivalently eqn. (17) can be written as comparing Y with another threshold $\tau' = L^{-1}(\tau)$. Hence $\delta_B(y) = \mathbb{1}_{\{Y > \tau'\}}$, where,

$$\tau' = \frac{\mu_0 + \mu_1}{2} + \frac{\sigma^2}{\mu_0 - \mu_1} \log(\tau). \quad (18)$$

For example, with uniform costs and equal priors we have $\tau = 1$ and $\tau' = \left(\frac{\mu_0 + \mu_1}{2}\right)$. Thus, in this particular case, the Bayes rule compares the observation to the average of μ_0 and μ_1. If y is greater than or equal to the average, the hypothesis H_1 is chosen, otherwise if y is less than this average, hypothesis H_0 is chosen. This test is illustrated in figure 2. We can write $P_j(\Gamma_1)$ for $j \in \{0, 1\}$ as follows.

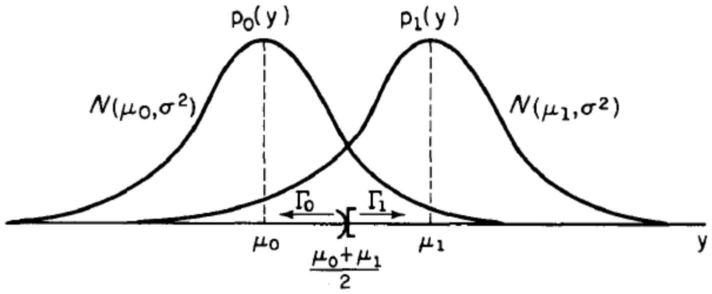

Figure 2: Illustration of location testing with Gaussian error with uniform cost and equal prior

$$
\begin{aligned}
P_j(\Gamma_1) &= \int_{\Gamma_1} dP_j(y) = \int_{\tau'}^{\infty} dP_j(y), \ \text{[since } \Gamma_1 = \{y \in \mathcal{R} | y \geq \tau'\}], \\
&= \int_{\frac{\tau'-\mu_j}{\sigma}}^{\infty} dP(\tau), \\
&= 1 - \Phi\Big(\frac{\tau' - \mu_j}{\sigma}\Big). \qquad (19)
\end{aligned}
$$

Now from eqn. (18), we can write the following

$$
P_j(\Gamma_1) = \begin{cases} 1 - \Phi\Big(\frac{\log(\tau)}{d} + \frac{d}{2}\Big) & \text{if } j = 0, \\ 1 - \Phi\Big(\frac{\log(\tau)}{d} - \frac{d}{2}\Big) & \text{if } j = 1, \end{cases} \qquad (20)
$$

where $d = \frac{\mu_1 - \mu_0}{\sigma}$ is a simple version of *signal-to-noise ratio(SNR)* and Φ denotes the cumulative distribution function of a $\mathcal{N}(0, 1)$. Now the unconditional risk is,

$$
r\big(\pi_0, \delta_{\pi_0}\big) = \pi_0 \left(1 - \Phi\left(\frac{\tau' - \mu_j}{\sigma^2}\right)\right) + (1 - \pi_0)\Phi\Big(\frac{\tau' - \mu_j}{\sigma^2}\Big) \qquad (21)
$$

For equal prior *i.e.* $\pi_0 = \pi_1 = \frac{1}{2}$, we have,

$$
\begin{aligned}
r\Big(\frac{1}{2}, \delta_{\frac{1}{2}}\Big) &= \frac{1}{2}\Big(1 - \Phi\Big(\frac{d}{2}\Big)\Big) + \frac{1}{2}\Phi\Big(-\frac{d}{2}\Big), \\
&= 1 - \Phi\Big(\frac{d}{2}\Big) \ \text{[due to even symmetry of Gaussian].} \qquad (22)
\end{aligned}
$$

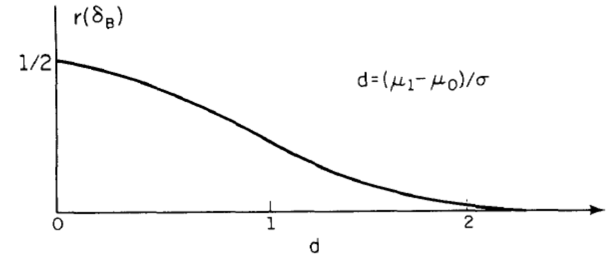

Figure 3: Bayes risk in location testing with Gaussian error

Minimax rule

We know that $V(\pi_0) = r(\pi_0, \delta_{\pi_0})$. Now $V(0) = C_{11}$ and $V(1) = C_{00}$, regardless of the cost structure as it only depends on prior and hence the least favorable prior π_L is in the interior (0,1) in this case. Moreover, since eqn. (21) is a differentiable function of π_0, randomization is unnecessary, and π_L can be found by setting $R_0(\delta_{\pi_L}) = R_1(\delta_{\pi_L})$. [That randomization is unnecessary also follows by noting that $P_0(L(Y) = \tau) = P_1(L(Y) = \tau) = 0$ for any τ since $L(Y)$ is a continuous random variable]. The prior π_0 enters $R_0(\delta_{\pi_0})$ and $R_1(\delta_{\pi_0})$ only through τ', so an equalizer rule is found by solving,

$$1 - \Phi\Big(\frac{\tau' - \mu_0}{\sigma}\Big) = \Phi\Big(\frac{\tau' - \mu_1}{\sigma}\Big). \tag{23}$$

By even symmetry property of Gaussian distribution function, we have,

$$\frac{\tau' - \mu_0}{\sigma} = \frac{\mu_1 - \tau'}{\sigma}. \tag{24}$$

The unique solution is given by the following, which is also clear from the figure 4,

$$\tau' = \frac{\mu_0 + \mu_1}{2}. \tag{25}$$

So the minimax decision rule is $\delta_{\pi_L} = \mathbb{1}_{\{y \geq \frac{\mu_0+\mu_1}{2}\}}$. From (25), it follows that the least favorable prior is $\pi_L = \frac{1}{2}$, and the minimax risk is,

$$V\left(\frac{1}{2}\right) = 1 - \Phi\Big(\frac{\mu_1 - \mu_0}{2\sigma}\Big). \tag{26}$$

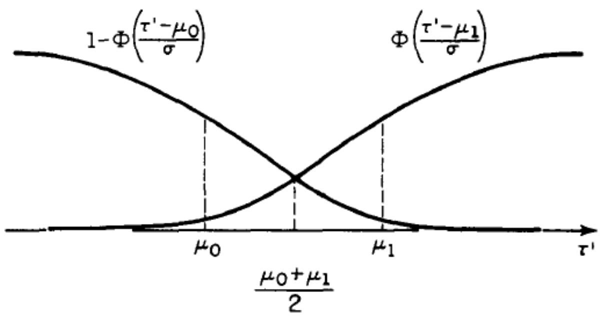

Figure 4: Conditional risk for location testing with Gaussian error and uniform cost

Neyman Pearson rule

Here, we have,

$$\begin{aligned}P_F(\tilde{\delta}_{NP}) &= P_0\{L(Y) > \eta\},\\ &= P_0\{Y > L^{-1}(\eta)\},\\ &= 1 - \Phi\Big(\frac{\eta' - \mu_0}{\sigma}\Big). \end{aligned} \tag{27}$$

where $\eta' = \frac{\mu_0+\mu_1}{2} + \frac{\sigma^2}{\mu_1-\mu_0}\log\eta$, and the curve of eqn. (27) is shown in figure 5. Note that any value of α can be achieved by exactly choosing,

$$\eta_0' = \mu_0 + \sigma\Phi^{-1}(1-\alpha), \tag{28}$$

where Φ^{-1} is the inverse function of Φ. Since $P(Y = \eta_0) = 0$, randomization can be chosen arbitrarily say $\gamma_0 = 1$. An α level Neyman-Pearson test for this case is given by,

$$\begin{aligned}\tilde{\delta}_{NP} &= \begin{cases} 1 - y \geq \eta_0, \\ 0 - y < \eta_0, \end{cases}\\ &= \mathbb{1}_{\{y \geq \eta_0\}}. \end{aligned} \tag{29}$$

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

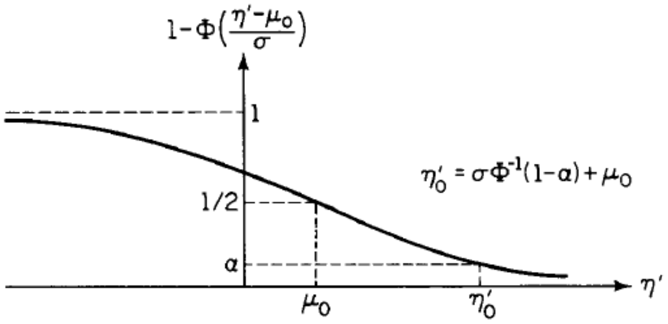

Figure 5: Illustration of threshold η_0' for Neyman-Pearson testing of location with Gaussian error

$$\begin{aligned}P_D(\tilde{\delta}_{NP}) &= P_1\{Y \geq \eta_0\}),\\ &= 1 - \Phi(\frac{\eta' - \mu_1}{\sigma}),\\ &= 1 - \Phi(\Phi^{-1}(1-\alpha) - d), \end{aligned} \tag{30}$$

where $d = \frac{\mu_1 - \mu_0}{\sigma}$ as appeared previously in case of Bayes hypothesis testing. For fixed α, equation (30) gives the detection probability as a function of d. This relationship is sometimes known as the power function for the test of eqn. (30). A plot of this relationship is shown in figure 6. Eqn. (29) also gives the detection probability as a function of the false-alarm probability for fixed d. Again borrowing from radar terminology, a parametric plot of this relationship is called the *receiver operating characteristics*(ROCs). The ROCs for the test of (29) are shown in figure 7.

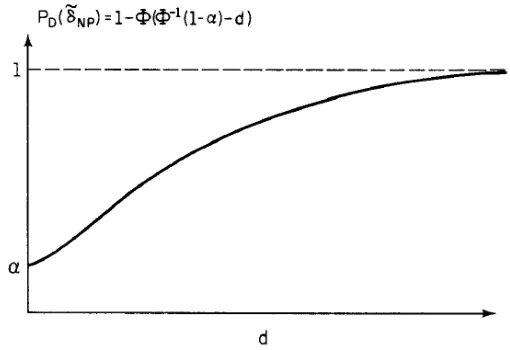

Figure 6: Power function for Neyman-Pearson testing for location testing with Gaussian error

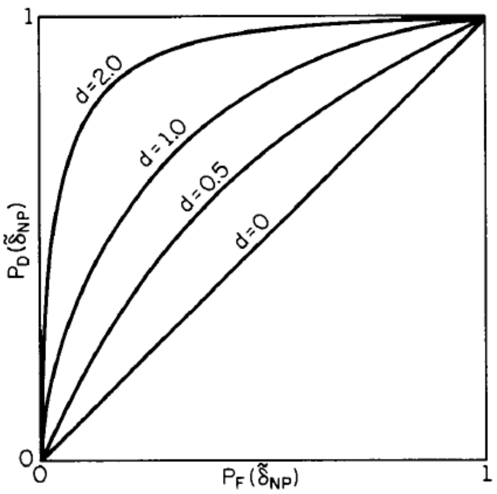

Figure 7: ROC curve for Neyman-Pearson testing for location testing with Gaussian error

Example 1.2 (The Binary Channel). On a Binary Communication Channel a binary digit is to be transmitted. Our observation Y is the output of the channel, which can also be either zero or one. Due to channel noise a transmitted "zero" is received as a "one" with probability λ_0 and as a "zero" with probability (1 - λ_0), where $0 \leq \lambda_0 \leq 1$. Similarly, a transmitted "one" is received as a "zero" with probability λ_1 and as a "one" with probability (1- λ_1). Thus, the observation Y does not always represent which among the "zero" or a "one" transmitted. So we need to develop a technique to optimally detect the transmitted digit.

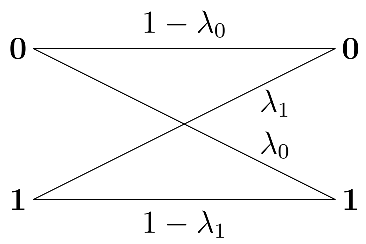

Figure 8: The binary channel

This situation is clearly a Hypothesis Testing problem with the two hypothesis H_0 and H_1 depicted as transmission of a "zero" and transmission of a "one" respectively. The observation set is $\Gamma = \{0, 1\}$. The received signal $Y \in \Gamma$ will have a probability density function as follows:

$$\begin{aligned} Y_0 &\sim (1 - \lambda_0) \text{ if } H_0 \text{ is transmitted,} \\ Y_1 &\sim (1 - \lambda_1) \text{ if } H_1 \text{ is transmitted,} \end{aligned} \tag{31}$$

and the observation Y has densities (i.e., probability mass functions):

$$p_j\left(y\right) = \begin{cases} \lambda_j, & \text{if } y \neq j, \\ \left(1 - \lambda_j\right), & \text{if } y = j, \end{cases} \tag{32}$$

for $j \in \{0, 1\}$.

Bayesian Hypothesis testing

The likelihood ratio is given by,

$$L(y) = \frac{p_1(y)}{p_0(y)} = \begin{cases} \frac{\lambda_1}{1-\lambda_0} & \text{if } y = 0, \\ \frac{1-\lambda_1}{\lambda_0} & \text{if } y = 1, \end{cases} \tag{33}$$

For certain threshold τ, the decision rule is,

$$\delta_B(y) = \begin{cases} \mathbb{1}_{\left\{\frac{\lambda_1}{1-\lambda_0} \geq \tau\right\}} & \text{if } y = 0 \;\; [\text{we write it as } \mathbb{1}_A \text{ (event } A)], \\ \mathbb{1}_{\left\{\frac{1-\lambda_1}{\lambda_0} \geq \tau\right\}} & \text{if } y = 1 \;\; [\text{we write it as } \mathbb{1}_B \text{ (event } B)]. \end{cases} \tag{34}$$

The conditional risks are given by the following equations,

$$\begin{aligned}R_0(\delta_{\pi_0}) &= P_0(\Gamma_1)\\ &= \lambda_0 \mathbb{1}_B + (1-\lambda_0)\mathbb{1}_A \quad (35)\\ R_1(\delta_{\pi_0}) &= P_1(\Gamma_0)\\ &= (1-\lambda_1)\mathbb{1}_{B^c} + \lambda_1 \mathbb{1}_{A^c} \quad (36)\end{aligned}$$

The unconditional risk is given by

$$\begin{aligned}r(\pi_0, \delta_{\pi_0}) = \pi_0\lambda_0\mathbb{1}_B + \pi_0(1-\lambda_0)\mathbb{1}_A + (1-\pi_0)(1-\lambda_1)(1-\mathbb{1}_B) + \\ (1-\pi_0)\lambda_1(1-\mathbb{1}_A),\\ = (1-\pi_0)(1-\lambda_1) - \{(1-\pi_0)(1-\lambda_1) - \pi_0\lambda_0\}\mathbb{1}_B + \\ (1-\pi_0)\lambda_1 - \{(1-\pi_0)\lambda_1 - \pi_0(1-\lambda_0)\}\mathbb{1}_A. \quad (37)\end{aligned}$$

To proceed further, we need the following,

$$\begin{aligned}A &= \left\{\frac{\lambda_1}{1-\lambda_0} \geq \frac{\pi_0}{1-\pi_0}\right\} \text{ means event } A \text{ is true,}\\ B &= \left\{\frac{1-\lambda_1}{\lambda_0} \geq \frac{\pi_0}{1-\pi_0}\right\} \text{ means event } B \text{ is true.}\end{aligned}$$

We know that,

$$\begin{aligned}f(a) &= a\mathbb{1}_{\{a\geq 0\}},\\ &= (a)_+,\\ &= max\{a, 0\}. \quad (38)\end{aligned}$$

So unconditional risk becomes

$$\begin{aligned}r(\pi_0, \delta_{\pi_0}) &= (1-\pi_0)(1-\lambda_1) - \Big\{(1-\pi_0)(1-\lambda_1) - \pi_0\lambda_0\Big\}_+ + \\ &\quad (1-\pi_0)\lambda_1 - \Big\{(1-\pi_0)\lambda_1 - \pi_0(1-\lambda_0)\Big\}_+,\\ &= \min\Big\{(1-\pi_0)(1-\lambda_1), \pi_0\lambda_0\Big\} + \min\Big\{(1-\pi_0)\lambda_1, \pi_0(1-\lambda_0)\Big\}. \quad (39)\end{aligned}$$

Again if $\pi_0 = 1 - \pi_0$, i.e., $\pi_0 = \frac{1}{2}$,

$$r\Big(\frac{1}{2}, \delta_{\frac{1}{2}}\Big) = \min\Big\{(1-\lambda_1), \lambda_0\Big\} + \min\Big\{\lambda_1, (1-\lambda_0)\Big\}. \quad (40)$$

Minimax rule

From equation (39) there are only two possibilities as follows,

$$\begin{aligned} \pi_0(1-\lambda_0) &\le \lambda_1(1-\pi_0), \\ \pi_0\lambda_0 &\le (1-\pi_0)(1-\lambda_1). \end{aligned} \tag{41}$$

Now, we define the quantity $\underline{\pi}$ and $\overline{\pi}$,

$$\underline{\pi} = \min\Big\{\frac{\lambda_1}{1-\lambda_0+\lambda_1}, \frac{1-\lambda_1}{1-\lambda_1+\lambda_0}\Big\},$$
$$\overline{\pi} = \max\Big\{\frac{\lambda_1}{1-\lambda_0+\lambda_1}, \frac{1-\lambda_1}{1-\lambda_1+\lambda_0}\Big\}.$$

The unconditional risk can be written as,

$$r(\pi_0, \delta_{\pi_0}) = \begin{cases} \pi_0, & \text{if } \pi_0 \le \underline{\pi}, \\ 1-\pi_0, & \text{if } \pi_0 \ge \overline{\pi}, \\ \underline{\pi} + \left(\frac{1-\overline{\pi}-\underline{\pi}}{\overline{\pi}-\underline{\pi}}\right)(\pi_0 - \underline{\pi}), & \text{if } \underline{\pi} < \pi_0 < \overline{\pi}. \end{cases} \tag{42}$$

Say $c = \left(\frac{1-\overline{\pi}-\underline{\pi}}{\overline{\pi}-\underline{\pi}}\right)$, Then, if $c > 0$ then $\pi_L = \overline{\pi}$; if $c < 0$, then $\pi_L = \underline{\pi}$; and if $c = 0$ then any q will work, where q is the probability of picking "one" at the threshold. So pick a randomized rule at the threshold.

Now recall that,

$$q = \frac{V'(\pi_L^+)}{V'(\pi_L^+) - V'(\pi_L^-)}, \tag{43}$$

where $V'(\pi_0)$ is the derivative of V with respect to π_0. Now assume $c > 0$, then $q = \frac{-1}{-1-c} = \frac{1}{1+c}$, which is clear from the figure 9. If $\pi_L = \overline{\pi}$, then $V(\overline{\pi}) = 1-\overline{\pi} > \underline{\pi}$;

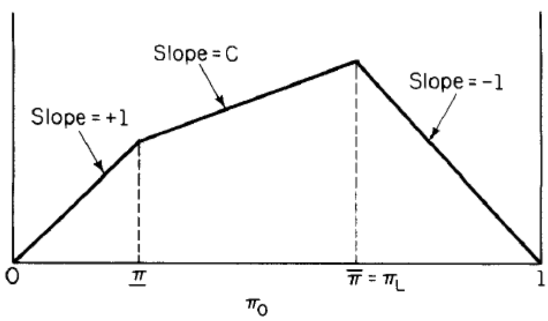

Figure 9: $V(\pi_0)$ for the binary channel

and, if $\pi_L = \underline{\pi}$, then $V(\underline{\pi}) = \underline{\pi} > 1 - \overline{\pi}$.

$$V(\pi_L) = \max\{\underline{\pi}, 1 - \overline{\pi}\}. \tag{44}$$

Now, the decision rule is,

$$\delta_{\pi_0}(y) = \begin{cases} 0, & \forall y \text{ if } \pi_0 \geq \overline{\pi}, \\ 1, & \forall y \text{ if } \pi_0 \leq \underline{\pi}. \end{cases} \tag{45}$$

And if $\pi_0 \in \{\underline{\pi}, \overline{\pi}\}$,

$$\begin{aligned} \delta_{\pi_0}(0) &= \mathbb{1}_{A^c}, \\ \delta_{\pi_0}(1) &= \mathbb{1}_B. \end{aligned} \tag{46}$$

Say $c > 0$, then by inspection, we have $\pi_L = \overline{\pi}$ and $\delta^+_{\pi_L}(y) = 0$, $\Gamma^+_1 = \phi$. The decision rule is,

$$\delta_{\pi_0}(y) = \begin{cases} y, & \text{if } \frac{1-\lambda_1}{1-\lambda_0-\lambda_1} \geq \pi_0 > \frac{\lambda_1}{1-\lambda_0-\lambda_1}, \\ 1 - y, & \text{if } \frac{\lambda_1}{1-\lambda_0-\lambda_1} \geq \pi_0 > \frac{1-\lambda_1}{1-\lambda_0-\lambda_1}. \end{cases} \tag{47}$$