Lecture-5: Composite Hypothesis Testing

22 Jan 16

Example 0.1. Neyman-Pearson Hypothesis Testing for binary channel (contd. from previous lecture)

Decision rule for Neyman Pearson testing has the form,

$$\tilde{\delta}_{NP}(y) = \begin{cases} 1, & L(y) > \eta_0, \\ \gamma_0, & L(y) = \eta_0, \\ 0, & L(y) < \eta_0, \end{cases} \tag{1}$$

where η_0 is desired threshold for α level Neyman Pearson testing, and $L(y) = \frac{P_1(y)}{P_0(y)}$ is the likelihood ratio. For the case of binary communication channel example (fig. 1), we have,

$$L(y) = \begin{cases} \frac{\lambda_1}{1-\lambda_0}, & y = 0, \\ \frac{1-\lambda_1}{\lambda_0}, & y = 1. \end{cases} \tag{2}$$

Assuming $\lambda_0 + \lambda_1 < 1$, we have, $\frac{\lambda_1}{1-\lambda_0} < 1$, and $\frac{1-\lambda_1}{\lambda_0} > 1$.

Now, we get,

$$P_0(L(y) > \eta) = \begin{cases} 1 & \text{if } \eta < \frac{\lambda_1}{1-\lambda_0}, \\ \lambda_0 & \text{if } \frac{\lambda_1}{1-\lambda_0} \leq \eta \geq \frac{1-\lambda_1}{\lambda_0}, \\ 0 & \text{if } \eta \geq \frac{1-\lambda_1}{\lambda_0}. \end{cases} \tag{3}$$

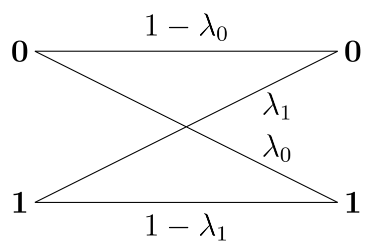

Figure 1: The Binary Channel

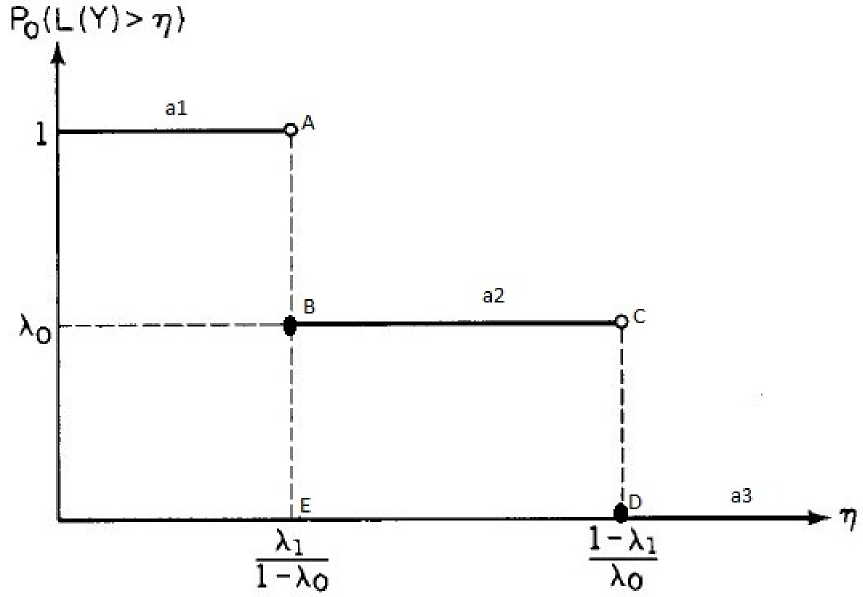

Figure 2: Curve for threshold and randomization selection for a binary channel.

Figure 2 shows a plot of the probability of false alarm. Let η_0 be the smallest number such that,

$$P_0(p_1(Y) > \eta_0 p_0(Y)) \leq \alpha. \tag{4}$$

$$\eta_0 = \begin{cases} \frac{1-\lambda_1}{\lambda_0}, & \alpha \in [0, \lambda_0), \quad \text{(section } a_3) \\ \frac{\lambda_1}{1-\lambda_0}, & \alpha \in [\lambda_0, 1), \quad \text{(section } a_2) \\ \text{arbitrary}, & \alpha = 1. \end{cases} \tag{5}$$

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

$$\gamma_0 = \frac{\alpha - P_0(p_1(Y) > \eta_0 p_0(Y))}{P_0(p_1(Y) = \eta_0 p_0(Y))} \tag{6}$$

If $\alpha \in [0, \lambda_0)$,

$$\gamma_0 = \frac{\alpha - 0}{P_0(p_1(Y) = \eta_0 p_0(Y))} \tag{7}$$

and,

$$P_0(p_1(Y) = \eta_0 p_0(Y)) = \lambda_0 - 0 \tag{8}$$

which is equal to the size of the discontinuity at threshold (CD in fig. 2).

$$\gamma_0 = \frac{\alpha}{\lambda_0} \tag{9}$$

If $\alpha \in [\lambda_0, 1)$,

$$P_0(p_1(Y) = \eta_0 p_0(Y)) = 1 - \lambda_0 \tag{10}$$

which is the size of the discontinuity at threshold (AB in fig. 2). We get, $\gamma_0 = \frac{\alpha - \lambda_0}{1-\lambda_0}$.

$$\gamma_0 = \begin{cases} \frac{\alpha}{\lambda_0}, & \alpha \in [0, \lambda_0), \\ \frac{\alpha-\lambda_0}{1-\lambda_0}, & \alpha \in [\lambda_0, 1), \\ \text{arbitrary}, & \alpha = 1. \end{cases} \tag{11}$$

If $\alpha \in [0, \lambda_0)$,

$$\tilde{\delta}_{NP}(y) = \begin{cases} \frac{\alpha}{\lambda_0}, & \text{if } y = 1, \\ 0, & \text{if } y = 0. \end{cases} \tag{12}$$

If $\alpha \in [\lambda_0, 1]$,

$$\tilde{\delta}_{NP}(y) = \begin{cases} 1, & \text{if } y = 1, \\ \frac{\alpha-\lambda_0}{1-\lambda_0}, & \text{if } y = 0. \end{cases} \tag{13}$$

The detection probability of the Neyman-Pearson test is given by ,

$$P_D(\tilde{\delta}_{NP}) = P_1(L(Y) > \eta_0) + \gamma_0 P_1(L(Y) = \eta_0), \tag{14}$$

$$P_D(\tilde{\delta}_{NP}) = \begin{cases} \frac{\alpha(1-\lambda_1)}{\lambda_0}, & \alpha \in [0, \lambda_0), \\ (1-\lambda_1) + \frac{\lambda_1(\alpha-\lambda_0)}{1-\lambda_0}, & \alpha \in [\lambda_0, 1]. \end{cases}$$

1 Composite Hypothesis Test

Hypothesis testing problems discussed in the previous lectures are sometimes known as 'simple hypothesis testing problems', because, each of the two hypotheses correspond to only a single distribution for the observation. In many hypothesis testing problems, however, there are many possible distributions that can occur under each of the hypotheses. Hypotheses of this type are known as *Composite Hypotheses.*

To model the most general type of composite hypothesis testing problems, we consider a family of probability distributions on Γ indexed by a parameter θ taking values in a parameter set Λ (the set of all possible natures of state), $\{P_\theta;\ \theta \in \Lambda\}$.

Example 1.1. For the simple hypothesis test $\Lambda = \{0,1\}$. More generally,we might have a parameter space that is the union of two disjoint parameter sets Λ_0 and Λ_1 representing the ranges of the parameter under the two hypotheses.

Bayesian Formulation:

In Bayesian formulation of the composite hypothesis testing problem, the parameter is assumed to be a random quantity, Θ, taking on the values in the set Λ, $\Theta \in \Lambda$. The distribution P_θ is interpreted as the conditional distribution of Y given that $\Theta = \theta$.

$$P_\theta\{Y = y\} = P\{Y = y|\Theta = \theta\} \tag{15}$$

we will consider only non-randomized decision rules, i.e., $\theta \in \Lambda_0$ or Λ_1. To choose an optimum decision rule, assign cost to our decisions through a cost function $C_i(\theta)$, where, $C_i(\theta)$ is the cost of choosing decision $i \in \{0, 1\}$ when $Y \sim P_\theta$. Assume that C is nonnegative and bounded.

For a decision rule δ, the conditional risk is defined as,

$$R_\theta(\delta) = \mathbb{E}_\theta[C_{\delta(Y)}(\theta)], \tag{16}$$

where $\mathbb{E}_\theta$ denotes expectation assuming that $Y \sim P_\theta$. Average Bayes risk is defined as

$$r(\delta) = \mathbb{E}[R_\Theta(\delta)]. \tag{17}$$

Bayes rule defined as minimization of $r(\delta)$.

$$\begin{aligned} r(\delta) &= \mathbb{E}[R_\Theta(\delta)], \\ &= \mathbb{E}[\mathbb{E}_\theta[C_{\delta(Y)}(\Theta)|\Theta = \theta]], \\ &= \mathbb{E}[\mathbb{E}[C_{\delta(Y)}(\Theta)|\Theta]], \\ &= \mathbb{E}[C_{\delta(Y)}(\Theta)], \\ &= \mathbb{E}[\mathbb{E}[C_{\delta(Y)}(\Theta)|Y = y]] \quad \forall\{y \in \Gamma\}, \end{aligned} \tag{18}$$

where the last step uses the relation of iterated expectations, $\mathbb{E}\{X\} = \mathbb{E}\{\mathbb{E}\{X|Y\}\}$. Minimizing $r(\delta)$ is same as minimizing $\mathbb{E}[\mathbb{E}[C_{\delta(Y)}(\Theta)|Y = y]$. Since $\delta(y)$ can only be 0 or 1, we see that Bayes rule is given by,

$$\delta(y) = \arg\min_{i \in \{0,1\}} \mathbb{E}[C_i(\Theta)|Y = y]. \tag{19}$$

We choose $\delta(y)$ to be the decision that minimizes the posterior cost.

$$\delta(y) = \mathbb{1}_{\{\mathbb{E}[C_1(\Theta)|Y=y] < \mathbb{E}[C_0(\Theta)|Y=y]\}}, \tag{20}$$

i.e., $\delta(y)$ chooses the hypothesis that is least costly, on the average, given the observations. For example, when $\Lambda = \{0, 1\}$, $\delta(y)$ reduces to the Bayes rule for simple hypothesis test. Assume costs being uniform over the sets Λ_0, Λ_1, i.e., $C_i(\theta) = C_{ij}$, $\forall \theta \in \Lambda_j$, we have,

$$\begin{aligned} &C_{11}P(\Theta \in \Lambda_1|Y = y) + C_{10}P(\Theta \in \Lambda_0|Y = y), \\ &\qquad < C_{01}P(\Theta \in \Lambda_1|Y = y) + C_{00}P(\Theta \in \Lambda_0|Y = y). \end{aligned}$$

Assuming $C_{11} < C_{01}$, we get,

$$\Gamma_1 = \left\{ y \in \Gamma \,\middle|\, \frac{C_{10} - C_{00}}{C_{01} - C_{11}} < \frac{\mathbb{P}(\Theta \in \Lambda_1 | Y = y)}{\mathbb{P}(\Theta \in \Lambda_0 | Y = y)} \right\} \tag{21}$$

where $P(\Theta \in \Lambda_j | Y = y)$ denotes the conditional probability that Θ lies in Λ_j given that $Y = y$. Assume Y has conditional densities $p(y|\Theta \in \Lambda_j)$ for $j \in \{0, 1\}$. Using Bayes rule, we get,

$$P(\Theta \in \Lambda_j | Y = y) = \frac{p(Y = y | \Theta \in \Lambda_j) P(\Theta \in \Lambda_j)}{p(y)}, \; j \in \{0, 1\}. \tag{22}$$

We can write Γ_1 as,

$$\Gamma_1 = \left\{ y \in \Gamma \,\middle|\, \frac{p(Y = y | \Theta \in \Lambda_1) P(\Theta \in \Lambda_1)}{p(Y = y | \Theta \in \Lambda_0) P(\Theta \in \Lambda_0)} > \frac{C_{10} - C_{00}}{C_{01} - C_{11}} \right\}. \tag{23}$$

For general case, let $\Theta \sim W, \; Y \sim \mathbb{P}_\theta, \; \theta \in \Lambda$,

$$dP[Y \leq y | \Theta \in \Lambda_j] = \int_{-\infty}^{y} \int_{\Lambda} P_\theta(y) dW_j(\theta), \tag{24}$$

where,

$$dW_j(\theta) \begin{cases} 0, & \theta \notin \Lambda_j, \\ \frac{dW(\theta)}{\mathbb{P}\{\theta \in \Lambda_j\}}, & \theta \in \Lambda_j. \end{cases} \tag{25}$$