Hypothesis Testing

Saravanan Vijayakumaran
sarva@ee.iitb.ac.in

Department of Electrical Engineering
Indian Institute of Technology Bombay

Basics of Hypothesis Testing

What is a Hypothesis?

One situation among a set of possible situations

Example (Radar)

EM waves are transmitted and the reflections observed.

Null Hypothesis Plane absent

Alternative Hypothesis Plane present

For a given set of observations, either hypothesis may be true.

What is Hypothesis Testing?

- A statistical framework for deciding which hypothesis is true
- Under each hypothesis the observations are assumed to have a known distribution
- Consider the case of two hypotheses (binary hypothesis testing)

$$\begin{aligned} H_0 &: \mathbf{Y} \sim P_0 \\ H_1 &: \mathbf{Y} \sim P_1 \end{aligned}$$

  $\mathbf{Y}$ is the random observation vector belonging to $\mathbb{R}^n$ for $n \in \mathbb{N}$

- The hypotheses are assumed to occur with given prior probabilities

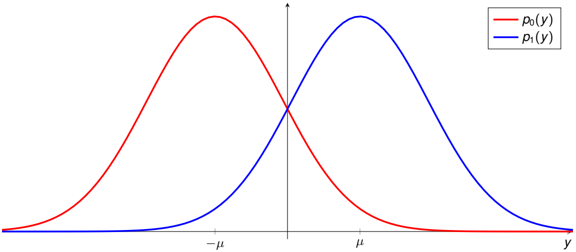
$$\begin{aligned} \Pr(H_0 \text{ is true}) &= \pi_0 \\ \Pr(H_1 \text{ is true}) &= \pi_1 \end{aligned}$$

  where $\pi_0 + \pi_1 = 1$.

Location Testing with Gaussian Error

- Let observation set be $\mathbb{R}$ and $\mu > 0$

$$\begin{aligned} H_0 &: \quad Y \sim \mathcal{N}(-\mu, \sigma^2) \\ H_1 &: \quad Y \sim \mathcal{N}(\mu, \sigma^2) \end{aligned}$$



- Any point in $\mathbb{R}$ can be generated under both H_0 and H_1
- What is a good decision rule for this hypothesis testing problem which takes the prior probabilities into account?

What is a Decision Rule?

- A decision rule for binary hypothesis testing is a partition of $\mathbb{R}^n$ into Γ_0 and Γ_1 such that
$$\delta(\mathbf{y}) = \begin{cases} 0 & \text{if } \mathbf{y} \in \Gamma_0 \\ 1 & \text{if } \mathbf{y} \in \Gamma_1 \end{cases}$$
We decide H_i is true when $\delta(\mathbf{y}) = i$ for $i \in \{0, 1\}$
- For the location testing with Gaussian error problem, one possible decision rule is
$$\begin{aligned} \Gamma_0 &= (-\infty, 0] \\ \Gamma_1 &= (0, \infty) \end{aligned}$$
and another possible decision rule is
$$\begin{aligned} \Gamma_0 &= (-\infty, -100) \cup (-50, 0) \\ \Gamma_1 &= [-100, -50] \cup [0, \infty) \end{aligned}$$
- Given that partitions of the observation set define decision rules, what is the optimal partition?

Which is the Optimal Decision Rule?

- The optimal decision rule minimizes the probability of decision error
- For the binary hypothesis testing problem of H_0 versus H_1, the conditional decision error probability given H_i is true is

$$
\begin{aligned}
P_{e|i} &= \Pr\left[\text{Deciding } H_{1-i} \text{ is true} | H_i \text{ is true}\right] \\
&= \Pr\left[Y \in \Gamma_{1-i} | H_i\right] \\
&= 1 - \Pr\left[Y \in \Gamma_i | H_i\right] \\
&= 1 - P_{c|i}
\end{aligned}
$$

- Probability of decision error is

$$P_e = \pi_0 P_{e|0} + \pi_1 P_{e|1}$$

- Probability of correct decision is

$$P_c = \pi_0 P_{c|0} + \pi_1 P_{c|1} = 1 - P_e$$

Which is the Optimal Decision Rule?

- Maximizing the probability of correct decision will minimize probability of decision error
- Probability of correct decision is

$$\begin{aligned} P_c &= \pi_0 P_{c|0} + \pi_1 P_{c|1} \\ &= \pi_0 \int_{\Gamma_0} p_0(y)\,dy + \pi_1 \int_{\Gamma_1} p_1(y)\,dy \\ &= \pi_0 \int_{\Gamma_0} p_0(y)\,dy + \pi_1 \left[1 - \int_{\Gamma_0} p_1(y)\,dy\right] \\ &= \pi_1 + \int_{\Gamma_0} \left[\pi_0 p_0(y) - \pi_1 p_1(y)\right]\,dy \end{aligned}$$

- To maximize P_c, we choose the partition $\{\Gamma_0, \Gamma_1\}$ as

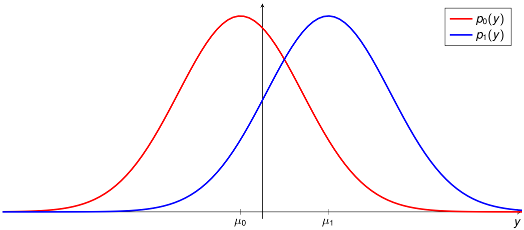
$$\begin{aligned} \Gamma_0 &= \{y \in \mathbb{R} | \pi_0 p_0(y) \geq \pi_1 p_1(y)\} \\ \Gamma_1 &= \{y \in \mathbb{R} | \pi_0 p_0(y) < \pi_1 p_1(y)\} \end{aligned}$$

- The points y for which $\pi_0 p_0(y) = \pi_1 p_1(y)$ can be in either Γ_0 and Γ_1 (the optimal decision rule is not unique)

Location Testing with Gaussian Error

- Let $\mu_1 > \mu_0$ and $\pi_0 = \pi_1 = \frac{1}{2}$

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



$$\begin{aligned} p_0(y) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu_0)^2}{2\sigma^2}} \\ p_1(y) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu_1)^2}{2\sigma^2}} \end{aligned}$$

Location Testing with Gaussian Error

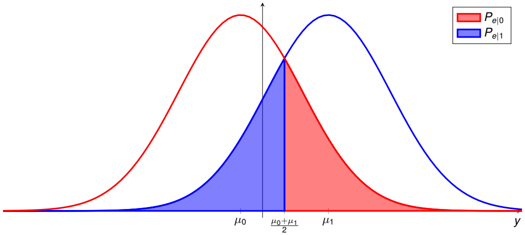
- Optimal decision rule is given by the partition $\{\Gamma_0, \Gamma_1\}$

$$
\begin{aligned}
\Gamma_0 &= \{y \in \mathbb{R} | \pi_0 p_0(y) \geq \pi_1 p_1(y)\} \\
\Gamma_1 &= \{y \in \mathbb{R} | \pi_0 p_0(y) < \pi_1 p_1(y)\}
\end{aligned}
$$

- For $\pi_0 = \pi_1 = \frac{1}{2}$

$$
\begin{aligned}
\Gamma_0 &= \left\{ y \in \mathbb{R} \middle| y \leq \frac{\mu_1 + \mu_0}{2} \right\} \\
\Gamma_1 &= \left\{ y \in \mathbb{R} \middle| y > \frac{\mu_1 + \mu_0}{2} \right\}
\end{aligned}
$$

Location Testing with Gaussian Error



$$P_{e|0} = \Pr\left[Y > \frac{\mu_0+\mu_1}{2} \middle| H_0\right] = Q\left(\frac{\mu_1-\mu_0}{2\sigma}\right)$$

$$P_{e|1} = \Pr\left[Y \leq \frac{\mu_0+\mu_1}{2} \middle| H_1\right] = \Phi\left(\frac{\mu_0-\mu_1}{2\sigma}\right) = Q\left(\frac{\mu_1-\mu_0}{2\sigma}\right)$$

$$P_e = \pi_0 P_{e|0} + \pi_1 P_{e|1} = Q\left(\frac{\mu_1-\mu_0}{2\sigma}\right)$$

This P_e is for $\pi_0 = \pi_1 = \frac{1}{2}$

Location Testing with Gaussian Error

- Suppose $\pi_0 \neq \pi_1$
- Optimal decision rule is still given by the partition $\{\Gamma_0, \Gamma_1\}$

$$
\begin{aligned}
\Gamma_0 &= \{y \in \mathbb{R} | \pi_0 p_0(y) \geq \pi_1 p_1(y)\} \\
\Gamma_1 &= \{y \in \mathbb{R} | \pi_0 p_0(y) < \pi_1 p_1(y)\}
\end{aligned}
$$

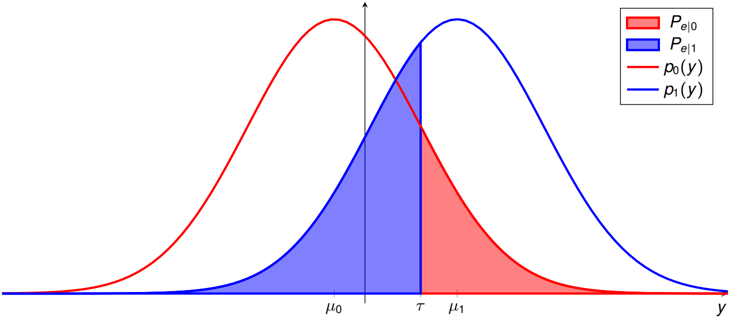
- The partitions specialized to this problem are

$$
\begin{aligned}
\Gamma_0 &= \left\{ y \in \mathbb{R} \,\middle|\, y \leq \frac{\mu_1 + \mu_0}{2} + \frac{\sigma^2}{(\mu_1 - \mu_0)} \log \frac{\pi_0}{\pi_1} \right\} \\
\Gamma_1 &= \left\{ y \in \mathbb{R} \,\middle|\, y > \frac{\mu_1 + \mu_0}{2} + \frac{\sigma^2}{(\mu_1 - \mu_0)} \log \frac{\pi_0}{\pi_1} \right\}
\end{aligned}
$$

Location Testing with Gaussian Error

Suppose $\pi_0 = 0.6$ and $\pi_1 = 0.4$

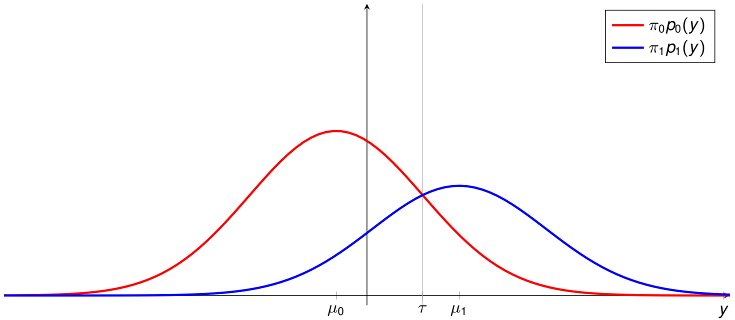
$$\tau = \frac{\mu_1 + \mu_0}{2} + \frac{\sigma^2}{(\mu_1 - \mu_0)} \log \frac{\pi_0}{\pi_1} = \frac{\mu_1 + \mu_0}{2} + \frac{0.4054\sigma^2}{(\mu_1 - \mu_0)}$$



Location Testing with Gaussian Error

Suppose $\pi_0 = 0.6$ and $\pi_1 = 0.4$

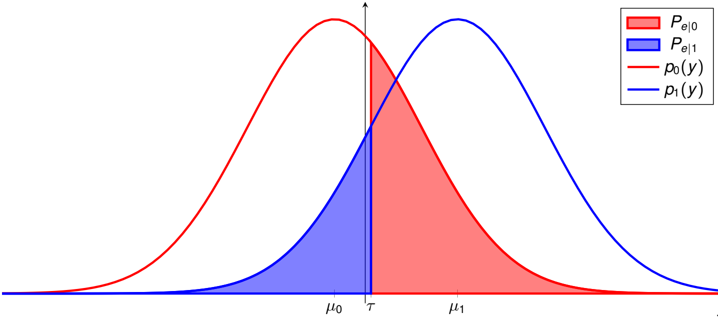
$$\tau = \frac{\mu_1 + \mu_0}{2} + \frac{\sigma^2}{(\mu_1 - \mu_0)} \log \frac{\pi_0}{\pi_1} = \frac{\mu_1 + \mu_0}{2} + \frac{0.4054\sigma^2}{(\mu_1 - \mu_0)}$$



Location Testing with Gaussian Error

Suppose $\pi_0 = 0.4$ and $\pi_1 = 0.6$

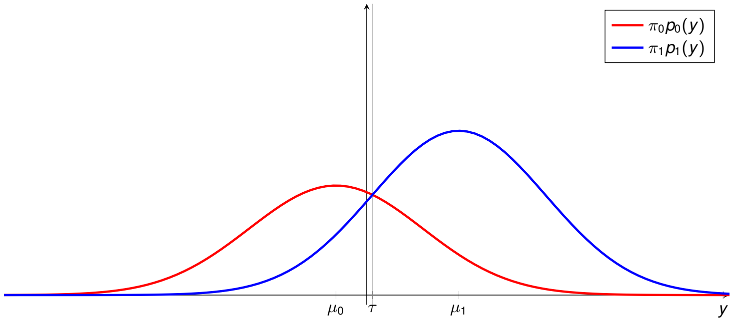
$$\tau = \frac{\mu_1 + \mu_0}{2} + \frac{\sigma^2}{(\mu_1 - \mu_0)} \log \frac{\pi_0}{\pi_1} = \frac{\mu_1 + \mu_0}{2} - \frac{0.4054\sigma^2}{(\mu_1 - \mu_0)}$$



Location Testing with Gaussian Error

Suppose $\pi_0 = 0.4$ and $\pi_1 = 0.6$

$$\tau = \frac{\mu_1 + \mu_0}{2} + \frac{\sigma^2}{(\mu_1 - \mu_0)} \log \frac{\pi_0}{\pi_1} = \frac{\mu_1 + \mu_0}{2} - \frac{0.4054\sigma^2}{(\mu_1 - \mu_0)}$$



M-ary Hypothesis Testing

- M hypotheses with prior probabilities π_i, $i = 1, \ldots, M$

$$\begin{array}{ccc} H_1 & : & \mathbf{Y} \sim P_1 \\ H_2 & : & \mathbf{Y} \sim P_2 \\ \vdots & & \vdots \\ H_M & : & \mathbf{Y} \sim P_M \end{array}$$

- A decision rule for M-ary hypothesis testing is a partition of Γ into M disjoint regions $\{\Gamma_i | i = 1, \ldots, M\}$ such that

$$\delta(\mathbf{y}) = i \text{ if } \mathbf{y} \in \Gamma_i$$

  We decide H_i is true when $\delta(\mathbf{y}) = i$ for $i \in \{1, \ldots, M\}$

- Minimum probability of error rule is

$$\delta_{\text{MPE}}(\mathbf{y}) = \arg \max_{1 \leq i \leq M} \pi_i p_i(\mathbf{y})$$

Maximum A Posteriori Decision Rule

- The a posteriori probability of H_i being true given observation $\mathbf{y}$ is

$$P\left[H_i \text{ is true} \middle| \mathbf{y}\right] = \frac{\pi_i p_i(\mathbf{y})}{p(\mathbf{y})}$$

- The MAP decision rule is given by

$$\delta_{\text{MAP}}(\mathbf{y}) = \arg\max_{1 \le i \le M} P\left[H_i \text{ is true} \middle| \mathbf{y}\right] = \delta_{\text{MPE}}(\mathbf{y})$$

MAP decision rule = MPE decision rule

Maximum Likelihood Decision Rule

- The ML decision rule is given by

$$\delta_{\mathrm{ML}}(\mathbf{y}) = \arg\max_{1 \le i \le M} p_i(\mathbf{y})$$

- If the M hypotheses are equally likely, $\pi_i = \frac{1}{M}$
- The MPE decision rule is then given by

$$\delta_{\mathrm{MPE}}(\mathbf{y}) = \arg\max_{1 \le i \le M} \pi_i p_i(\mathbf{y}) = \delta_{\mathrm{ML}}(\mathbf{y})$$

For equal priors, ML decision rule = MPE decision rule

Irrelevant Statistics

Irrelevant Statistics

- In this context, the term statistic means an observation
- For a given hypothesis testing problem, all the observations may not be useful

Example (Irrelevant Statistic)

$$\mathbf{Y} = \begin{bmatrix} Y_1 & Y_2 \end{bmatrix}^T$$

$$\begin{array}{llll} H_1 : & Y_1 = A + N_1, & Y_2 = N_2 \\ H_0 : & Y_1 = N_1, & Y_2 = N_2 \end{array}$$

where $A > 0$, $N_1 \sim \mathcal{N}(0, \sigma^2)$, $N_2 \sim \mathcal{N}(0, \sigma^2)$.

- If N_1 and N_2 are independent, Y_2 is irrelevant.
- If N_1 and N_2 are correlated, Y_2 is relevant.

- Need a method to recognize irrelevant components of the observations

Characterizing an Irrelevant Statistic

Theorem

For M-ary hypothesis testing using an observation $\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 & \mathbf{Y}_2 \end{bmatrix}$, the statistic $\mathbf{Y}_2$ is irrelevant if the conditional distribution of $\mathbf{Y}_2$, given $\mathbf{Y}_1$ and H_i, is independent of i. In terms of densities, the condition for irrelevance is

$$p(\mathbf{y}_2|\mathbf{y}_1, H_i) = p(\mathbf{y}_2|\mathbf{y}_1) \ \ \forall i.$$

Proof

$$\begin{aligned}
\delta_{\text{MPE}}(\mathbf{y}) &= \arg\max_{1 \le i \le M} \pi_i p_i(\mathbf{y}) = \arg\max_{1 \le i \le M} \pi_i p(\mathbf{y}|H_i) \\
p(\mathbf{y}|H_i) &= p(\mathbf{y}_1, \mathbf{y}_2|H_i) = p(\mathbf{y}_2|\mathbf{y}_1, H_i) p(\mathbf{y}_1|H_i) \\
&= p(\mathbf{y}_2|\mathbf{y}_1) p(\mathbf{y}_1|H_i) \\
\delta_{\text{MPE}}(\mathbf{y}) &= \arg\max_{1 \le i \le M} \pi_i p(\mathbf{y}_2|\mathbf{y}_1) p(\mathbf{y}_1|H_i) = \arg\max_{1 \le i \le M} \pi_i p(\mathbf{y}_1|H_i)
\end{aligned}$$

Example of an Irrelevant Statistic

Example (Independent Noise)

$$\mathbf{Y} = \begin{bmatrix} Y_1 & Y_2 \end{bmatrix}^T$$

$$\begin{array}{llll} H_1 : & Y_1 = A + N_1, & Y_2 = N_2 \\ H_0 : & Y_1 = N_1, & Y_2 = N_2 \end{array}$$

where $A > 0$, $N_1 \sim \mathcal{N}(0, \sigma^2)$, $N_2 \sim \mathcal{N}(0, \sigma^2)$, with N_1, N_2 independent

$$\begin{aligned} p(y_2|y_1, H_0) &= p(y_2) \\ p(y_2|y_1, H_1) &= p(y_2) \end{aligned}$$

Example of a Relevant Statistic

Example (Correlated Noise)

$$\mathbf{Y} = \begin{bmatrix} Y_1 & Y_2 \end{bmatrix}^T$$

$$\begin{array}{lll} H_1: & Y_1 = A + N_1, & Y_2 = N_2 \\ H_0: & Y_1 = N_1, & Y_2 = N_2 \end{array}$$

where $A > 0$, $N_1 \sim \mathcal{N}(0, \sigma^2)$, $N_2 \sim \mathcal{N}(0, \sigma^2)$, $\mathbf{C}_Y = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$

$$\begin{aligned} p(y_2|y_1, H_0) &= \frac{1}{\sqrt{2\pi(1-\rho^2)\sigma^2}} e^{-\frac{(y_2-\rho y_1)^2}{2(1-\rho^2)\sigma^2}}, \\ p(y_2|y_1, H_1) &= \frac{1}{\sqrt{2\pi(1-\rho^2)\sigma^2}} e^{-\frac{[y_2-\rho(y_1-A)]^2}{2(1-\rho^2)\sigma^2}} \end{aligned}$$

References

- Section 3.2, *Fundamentals of Digital Communication*, Upamanyu Madhow, 2008
- Chapter 2, *An Introduction to Signal Detection and Estimation*, H. V. Poor, Second Edition, Springer Verlag, 1994.
- *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*, Steven M. Kay, Prentice Hall, 1998.