Bayesian Multiple Hypothesis Testing

Jessica Fridrich

EECE 566

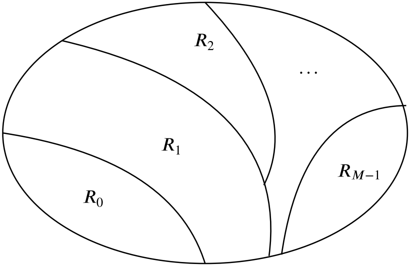
Introduction

- We extend to $M : H_0, H_1, \ldots, H_{M-1}$ hypotheses with known priors $P(H_i)$ and densities $p(\mathbf{x}|H_i)$
  - Pattern recognition or classification with multiple classes
- C_{ij} **cost** of deciding H_i when H_j is true
- Expected Bayes Risk:

$$C = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} C_{ij} P(H_i|H_j) P(H_j)$$

- Find the partition of the space into $R_0, \ldots, R_{M-1}$ so that C is minimal

Bayes risk minimization for multiple HT



- $R_0 \cup R_1 \cup \ldots \cup R_{M-1} = \mathbb{R}^N$
- If $\mathbf{x} \in R_i \Rightarrow$ decide H_i

Bayes risk minimization for multiple HT

$$
\begin{aligned}
C &= \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} C_{ij} P(H_i|H_j) P(H_j) \\
&= \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} C_{ij} \int_{R_i} p(\mathbf{x}|H_j)\mathrm{d}\mathbf{x} P(H_j) \\
&= \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} C_{ij} \int_{R_i} p(\mathbf{x}|H_j) P(H_j)\mathrm{d}\mathbf{x} \\
\text{(Bayes rule)} \quad &= \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} C_{ij} \int_{R_i} p(H_j|\mathbf{x}) p(\mathbf{x})\mathrm{d}\mathbf{x} \\
\text{(BTW)} \quad p(\mathbf{x}) &= \sum_{j=0}^{M-1} p(\mathbf{x}|H_j) P(H_j)
\end{aligned}
$$

Bayes risk minimization for multiple HT

$$
\begin{aligned}
C &= \sum_{i=0}^{M-1} \int_{R_i} \sum_{j=0}^{M-1} C_{ij} p(H_j|\mathbf{x}) p(\mathbf{x}) \mathrm{d}\mathbf{x} \\
&= \sum_{i=0}^{M-1} \int_{R_i} C_i(\mathbf{x}) p(\mathbf{x}) \mathrm{d}\mathbf{x}
\end{aligned}
$$

- Note: $\sum_{j=0}^{M-1} p(H_j|\mathbf{x}) = 1 \implies$
- $C_i(\mathbf{x}) = \sum_{j=0}^{M-1} C_{ij} p(H_j|\mathbf{x})$ is the expected average cost of deciding H_i given $\mathbf{x}$

$\Rightarrow$ Assign $\mathbf{x}$ to R_k where $k = \text{argmin}_i C_i(\mathbf{x})$

Bayes risk minimization for multiple HT, equal costs, zero gains

- Equal costs (without loss of generality): $C_{ij} = 1$ for $i \neq j$, zero gains $C_{ij} = 0$ for $i = j$
- We select

$$
\begin{aligned}
k = \mathrm{argmin}_i C_i(\mathbf{x}) &= \mathrm{argmin}_i \sum_{j=0}^{M-1} C_{ij} p(H_j|\mathbf{x}) \\
\text{(equal costs, zero gains)} &= \mathrm{argmin}_i \sum_{j=0, j\neq i}^{M-1} p(H_j|\mathbf{x}) \\
k &= \mathrm{argmin}_i \sum_{j=0}^{M-1} p(H_j|\mathbf{x}) - p(H_i|\mathbf{x})
\end{aligned}
$$

Bayes risk minimization for multiple HT, equal costs, zero gains

$$k \;=\; \text{argmin}_i\{1 - p(H_i|\mathbf{x})\} = \text{argmax}_i\, p(H_i|\mathbf{x})$$

- $\Rightarrow$ MAP minimizes the total error P_e:

$$P_e = \sum_{i \neq j} P(H_i|H_j)P(H_j)$$

Bayes risk minimization for multiple HT, equal costs, zero gains, equal priors

- If priors are equal: $P(H_i) = 1/M$ for all i. Then MAP turns to ML:

$$
\begin{aligned}
k &= \text{argmax}_i p(H_i|\mathbf{x}) \\
\text{(Bayes rule)} \quad &= \text{argmax}_i \frac{p(\mathbf{x}|H_i) \times 1/M}{p(\mathbf{x})} \\
k &= \text{argmax}_i p(\mathbf{x}|H_i), \text{ ML detector}
\end{aligned}
$$

Example: Multiple DC levels in WGN

$$
\begin{aligned}
H_0: &\quad x_j = -A + \xi_j \\
H_1: &\quad x_j = \xi_j \\
H_2: &\quad x_j = A + \xi_j
\end{aligned}
$$

- $A > 0$, $\xi_j \sim \mathcal{N}(0, \sigma^2)$ iid, $A_i = -A, 0, A$
- Assume: equal priors: $P(H_i) = 1/3$, no gains, equal costs. ML minimizes P_e

$$
p(\mathbf{x}|H_i) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_j - A_i)^2}{2\sigma^2}\right)
$$

Example: Detection of DC level in WGN

$$
\begin{aligned}
\text{ML detector: } k &= \text{argmax}_i p(\mathbf{x}|H_i) \\
&= \text{argmin}_i - \log p(\mathbf{x}|H_i) \\
&= \text{argmin}_i \sum_{j=1}^{n} (x_j - A_i)^2 \\
k &= \text{argmin}_i D_i^2
\end{aligned}
$$

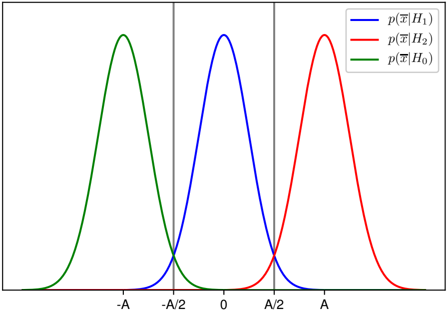
Example: Detection of DC level in WGN

$$
\begin{aligned}
D_i^2 = \sum_{j=1}^{n}(x_j - A_i)^2 \quad &= \quad \sum_{j=1}^{n}(x_j - \overline{\mathbf{x}} + \overline{\mathbf{x}} - A_i)^2 \\
D_i^2 \quad &= \quad \sum_{j=1}^{n}(x_j - \overline{\mathbf{x}})^2 + 2(\overline{\mathbf{x}} - A_i)\underbrace{\cancel{\sum_{j=1}^{n}(x_j - \overline{\mathbf{x}})}}_{=0} \\
&\quad\quad + n(\overline{\mathbf{x}} - A_i)^2, \\
D_i^2 \quad &= \quad \sum_{j=1}^{n}(x_j - \overline{\mathbf{x}})^2 + n(\overline{\mathbf{x}} - A_i)^2
\end{aligned}
$$

- $\sum_{j=1}^{n}(x_j - \overline{\mathbf{x}})^2$ doesn't depend on i $\Rightarrow$ chose $k = \text{argmin}_i(\overline{\mathbf{x}} - A_i)^2$
- Minimum P_e detector is the minimum Euclidean-distance detector between $\overline{\mathbf{x}}$ and $A_i = -A, 0, A$

Example: Detection of DC level in WGN

$$\overline{\mathbf{x}} \sim \begin{cases} \mathcal{N}(-A, \sigma^2/n) & \text{under } H_0 \\ \mathcal{N}(0, \sigma^2/n) & \text{under } H_1 \\ \mathcal{N}(A, \sigma^2/n) & \text{under } H_2 \end{cases}$$

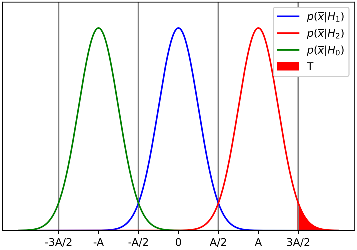


Example: Detection of DC level in WGN

To evaluate P_e, we will compute P_c the probability of making a correct decision $P_e = 1 - P_c$

$$\begin{aligned}
P_c &= \sum_{i=0}^{2} P(H_i|H_i)P(H_i) \\
&= \frac{1}{3}\sum_{i=0}^{2} P(H_i|H_i) \\
&= \frac{1}{3}[\mathsf{Pr}\{\overline{\mathbf{x}} < -A/2|H_0\} + \mathsf{Pr}\{-A/2 < \overline{\mathbf{x}} < A/2|H_1\} \\
&\quad + \mathsf{Pr}\{\overline{\mathbf{x}} > A/2|H_2\}] \\
&= \frac{1}{3}\left[1 - Q\left(\frac{-A/2 + A}{\sqrt{\sigma^2/n}}\right) + Q\left(\frac{-A/2 - 0}{\sqrt{\sigma^2/n}}\right) - Q\left(\frac{A/2 - 0}{\sqrt{\sigma^2/n}}\right)\right. \\
&\quad \left. + Q\left(\frac{A/2 - A}{\sqrt{\sigma^2/n}}\right)\right]
\end{aligned}$$

Example: Detection of DC level in WGN



$$P_c = \frac{1}{3}(1 - T + 1 - 2T + 1 - T) = 1 - \frac{4}{3}T$$

$$P_e = 1 - P_c = \frac{4}{3}T = \frac{4}{3}Q\left(\sqrt{\frac{A^2 n}{4\sigma^2}}\right)$$

$$T = Q\left(\sqrt{\frac{A^2 n}{4\sigma^2}}\right)$$

Example: Detection of DC level in WGN

- Note: For binary HT: $P_e = Q\left(\sqrt{\frac{A^2 n}{4\sigma^2}}\right)$, for ternary HT: $P_e = \frac{4}{3}Q\left(\sqrt{\frac{A^2 n}{4\sigma^2}}\right)$
- Increased probability of error because we need to distinguish between more hypotheses
- M-ary case (Exercise 3.20): $P_e = \frac{2M-2}{M}Q\left(\sqrt{\frac{A^2 n}{4\sigma^2}}\right)$
  - probability of error doubles for $M \to \infty$