Testing multiple hypotheses: a Bayesian decision-theoretic approach

Reading:

- Ch. 3 in Kay-II.

Testing Multiple Hypotheses

Choose a parameter-space partitioning:

$$\mathrm{sp}_\Theta(0) \cup \mathrm{sp}_\Theta(1) \cup \cdots \cup \mathrm{sp}_\Theta(M-1) = \mathrm{sp}_\Theta, \quad \mathrm{sp}_{\Theta_i} \cap \mathrm{sp}_{\Theta_j} = \emptyset \ \forall i \neq j.$$

We wish to distinguish among $M > 2$ hypotheses, i.e. identify which hypothesis is true:

$$\begin{aligned} \mathcal{H}_0 \quad &: \quad \Theta \in \mathrm{sp}_\Theta(0) \quad \text{versus} \\ \mathcal{H}_1 \quad &: \quad \Theta \in \mathrm{sp}_\Theta(1) \quad \text{versus} \\ \vdots & \qquad\qquad\qquad\quad \text{versus} \\ \mathcal{H}_{M-1} \quad &: \quad \Theta \in \mathrm{sp}_\Theta(M-1) \end{aligned}$$

and, consequently, our action space consists of M choices. We design a decision rule $\phi(\boldsymbol{x}) : \mathcal{X} \to (0, 1, \ldots, M-1)$:

$$\phi(\boldsymbol{x}) = \begin{cases} 0, & \text{decide } \mathcal{H}_0 \\ 1, & \text{decide } \mathcal{H}_1 \\ \vdots & \\ M-1, & \text{decide } \mathcal{H}_{M-1} \end{cases}$$

where $\phi(\boldsymbol{x})$ partitions the data space $\mathcal{X}$ [i.e. the support of $f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x}\,|\,\theta)$] into M regions:

$$\mathcal{X}_0 = \{\boldsymbol{x} : \phi(\boldsymbol{x}) = 0\}, \ldots, \quad \mathcal{X}_{M-1} = \{\boldsymbol{x} : \phi(\boldsymbol{x}) = M-1\}.$$

We use a piecewise-constant loss function specified via $\mathrm{L}(i \mid j)$, where $\mathrm{L}(i \mid j)$ is the loss of deciding the ith hypothesis when hypothesis j is true. Now, our posterior expected loss takes M values:

$$\underbrace{\rho_m(\boldsymbol{x})}_{\rho(\text{decide } \mathcal{H}_m \mid \boldsymbol{x})} = \int_{\Theta} \mathrm{L}(\theta, \text{decide } \mathcal{H}_m)\, f_{\Theta \mid \boldsymbol{X}}(\theta \mid \boldsymbol{x})\, d\theta$$

$$\stackrel{\substack{\mathrm{L}(\theta, \text{decide } \mathcal{H}_m) \text{ is} \\ \text{piecewise constant in } \theta}}{=} \sum_{i=0}^{M-1} \int_{\mathrm{sp}_{\Theta}(i)} \mathrm{L}(m \mid i)\, f_{\Theta \mid \boldsymbol{X}}(\theta \mid \boldsymbol{x})\, d\theta$$

$$\stackrel{\mathrm{L}(m \mid i) \text{ const. over } \mathrm{sp}_{\Theta}(i)}{=} \sum_{i=0}^{M-1} \mathrm{L}(m \mid i) \int_{\mathrm{sp}_{\Theta}(i)} f_{\Theta \mid \boldsymbol{X}}(\theta \mid \boldsymbol{x})\, d\theta$$

for $m = 0, 1, \ldots, M-1$. Then, the Bayes' decision rule $\phi^{\star}(\boldsymbol{x})$ is defined by the following data-space partitioning:

$$\mathcal{X}_m^{\star} = \{\boldsymbol{x} : \rho_m(\boldsymbol{x}) = \min_{0 \le l \le M-1} \rho_l(\boldsymbol{x})\} \quad m = 0, 1, \ldots, M-1$$

or, equivalently, upon applying the Bayes' rule,

$$\mathcal{X}_m^{\star} = \Big\{\boldsymbol{x} : m = \arg \min_{0 \le m' \le M-1} h_{m'}(\boldsymbol{x})\Big\}$$

where

$$h_m(\boldsymbol{x}) \triangleq \sum_{i=0}^{M-1} \mathrm{L}(m \mid i) \int_{\mathrm{sp}_{\Theta}(i)} f_{\boldsymbol{X} \mid \Theta}(\boldsymbol{x} \mid \theta) \pi(\theta)\, d\theta.$$

Multi-hypothesis Bayes' Decision Rule for 0-1 Loss

0-1 loss: $\mathrm{L}(i \,|\, i) = 0$ and $\mathrm{L}(m \,|\, i) = 1$ for $i \neq m$, i.e.

$$\mathrm{L}(m \,|\, i) = 1 - \delta_{m,i}$$

where $\delta_{m,i}$ is the Kronecker delta symbol. Hence, $\rho_m(\boldsymbol{x})$ can be written as

$$\begin{aligned}
\rho_m(\boldsymbol{x}) &= \sum_{i=0,\, i \neq m}^{M-1} \int_{\mathrm{sp}_\Theta(i)} f_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{x})\, d\theta \\
&= 1 - \int_{\mathrm{sp}_\Theta(m)} f_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{x})\, d\theta
\end{aligned}$$

yielding the following Bayes' decision rule, called the *maximum a posteriori* (MAP) rule:

$$\begin{aligned}
\mathcal{X}_m^\star &= \Big\{ \boldsymbol{x} : m = \arg \max_{0 \leq m' \leq M-1} \int_{\mathrm{sp}_\Theta(m')} f_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{x})\, d\theta \Big\} \\
&= \Big\{ \boldsymbol{x} : m = \arg \max_{0 \leq m' \leq M-1} \Pr_{\Theta \,|\, \boldsymbol{X}}\{\theta \in \mathrm{sp}_\Theta(m') \,|\, \boldsymbol{x}\} \Big\}. \quad (1)
\end{aligned}$$

Preposterior (Bayes) Risk

The preposterior (Bayes) risk for a rule $\phi(\boldsymbol{x})$ is

$$
\begin{aligned}
& \mathrm{E}_{\boldsymbol{X},\Theta}[\underbrace{\mathrm{L}(\Theta,\ \text{decide } \mathcal{H}_{\phi(\boldsymbol{X})})}_{\text{piecewise constant in } \Theta}] \\
&= \sum_{m=0}^{M-1} \sum_{i=0}^{M-1} \int_{\mathcal{X}_m} \mathrm{L}(m \,|\, i) \int_{\mathrm{sp}_\Theta(i)} f_{\boldsymbol{X} \,|\, \Theta}(\boldsymbol{x} \,|\, \theta) \pi(\theta) \, d\theta \, d\boldsymbol{x} \\
&= \sum_{m=0}^{M-1} \int_{\mathcal{X}_m} \underbrace{\sum_{i=0}^{M-1} \mathrm{L}(m \,|\, i) \int_{\mathrm{sp}_\Theta(i)} f_{\boldsymbol{X} \,|\, \Theta}(\boldsymbol{x} \,|\, \theta) \, \pi(\theta) \, d\theta \ d\boldsymbol{x}}_{\triangleq\, h_m(\boldsymbol{x})} \\
&= \sum_{m=0}^{M-1} \int_{\mathcal{X}_m} h_m(\boldsymbol{x}) \, d\boldsymbol{x}
\end{aligned}
$$

and the Bayes' decision rule $\phi^\star(\boldsymbol{x})$ is as before:

$$
\mathcal{X}_m^\star = \Big\{ \boldsymbol{x} : m = \arg \min_{0 \le m' \le M-1} h_{m'}(\boldsymbol{x}) \Big\}.
$$

Then, for an arbitrary rule $\phi(\boldsymbol{x})$,

$$\Big[\sum_{m=0}^{M-1}\int_{\mathcal{X}_m} h_m(x)\,d\boldsymbol{x}\Big] - \Big[\sum_{m=0}^{M-1}\int_{\mathcal{X}_m^\star} h_m(\boldsymbol{x})\,d\boldsymbol{x}\Big] \geq 0$$

which verifies that the Bayes' decision rule $\phi^\star(\boldsymbol{x})$ minimizes the preposterior (Bayes) risk.

Preposterior (Bayes) Risk for 0-1 Loss is the Average Error Probability

For the 0-1 loss, the preposterior (Bayes) risk for a rule $\phi(\boldsymbol{x})$ is

$$
\begin{aligned}
\mathrm{E}_{\boldsymbol{X},\Theta}[\underbrace{\mathrm{L}(\Theta,\ \text{decide } \mathcal{H}_{\phi(\boldsymbol{x})})}_{\text{piecewise constant}}] &= \mathrm{Pr}_{\mathrm{error}} \\
&= \sum_{m=0}^{M-1} \int_{\mathcal{X}_m} \sum_{i=0,\, i\neq m}^{M-1} \int_{\mathrm{sp}_\Theta(i)} f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x}\,|\,\theta)\, \pi(\theta)\, d\theta\, d\boldsymbol{x} \\
&= 1 - \underbrace{\sum_{m=0}^{M-1} \int_{\mathcal{X}_m} \int_{\mathrm{sp}_\Theta(m)} f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x}\,|\,\theta)\, \pi(\theta)\, d\theta\, d\boldsymbol{x}}_{\mathrm{Pr}_{\text{correct decision}}} \qquad (2)
\end{aligned}
$$

which is the *average error probability*, with *averaging* performed over the joint probability density or mass function (pdf or pmf) of the data $\boldsymbol{x}$ and parameter Θ.

Union bound: Suppose we wish to bound from above the minimum average error probability achieved by the Bayes' rule. If we had a binary hypothesis problem, say testing $\mathcal{H}_i$ versus $\mathcal{H}_j$, then the minimum average *pairwise* error probability for this binary problem was obtained in handout # 5:

$$P(i,j) = \int_{\underbrace{\mathcal{X}}_{\text{data space}}} \min\Big\{ \int_{\text{sp}_\Theta(i)} f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x}\,|\,\theta)\pi(\theta)\,d\theta, \int_{\text{sp}_\Theta(j)} f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x}\,|\,\theta)\pi(\theta)\,d\theta \Big\}\,d\boldsymbol{x}.$$

Now,

$$\underbrace{\min \Pr_{\text{error}}}_{\text{min. pr. error achieved by the Bayes' rule}} \le \sum_{j=1}^{M-1}\sum_{i=0}^{j-1} P(i,j)$$

which follows by applying the union-bound inequality:

$$\text{error event} = \cup_{j=1}^{M-1} \cup_{i=0}^{j-1} A(i,j)$$

where $A(i,j)$ is the event of mistakingly deciding $\mathcal{H}_i$ instead of $\mathcal{H}_j$ or vice versa. For union bound, see handout # 2 in EE 420x notes.

If we cannot easily compute $P(i,j)$, we can try to find an upper bound for it using the Chernoff bound, see handout # 5.

Bayesian Decision-theoretic Detection for Multiple Simple Hypotheses and 0-1 Loss

We specialize (1) to simple hypotheses $(\mathrm{sp}_\Theta(m) = \{\theta_m\}, m = 0, 1, \ldots, M-1)$:

$$\mathcal{X}_m^\star = \Big\{ \boldsymbol{x} : m = \arg \max_{0 \le l \le M-1} p_{\Theta \,|\, \boldsymbol{X}}(\theta_l \,|\, \boldsymbol{x}) \Big\}$$

or, equivalently,

$$\mathcal{X}_m^\star = \Big\{ \boldsymbol{x} : m = \arg \max_{0 \le m' \le M-1} [\pi_{m'} f_{\boldsymbol{X} \,|\, \Theta}(\boldsymbol{x} | \theta_{m'})] \Big\}$$

for $m = 0, 1, \ldots, M-1$, where

$$\pi_0 = \pi(\theta_0), \quad \ldots, \pi_{M-1} = \pi(\theta_{M-1})$$

define the prior pmf of the M-ary discrete random variable Θ; recall that Θ takes values from $\{\theta_0, \theta_1, \ldots, \theta_{M-1}\}$. If π_i, $i = 0, 1, \ldots, M-1$ are all equal, i.e.

$$\pi_0 = \pi_1 = \cdots = \pi_{M-1} = \frac{1}{M}$$

the resulting test

$$\begin{aligned}
\mathcal{X}_m^\star &= \Big\{ \boldsymbol{x} : m = \arg \max_{0 \le m' \le M-1} \Big[\frac{1}{M} f_{\boldsymbol{X} \,|\, \Theta}(\boldsymbol{x} \,|\, \theta_{m'}) \Big] \Big\} \\
&= \Big\{ \boldsymbol{x} : m = \arg \max_{0 \le m' \le M-1} \underbrace{f_{\boldsymbol{X} \,|\, \Theta}(\boldsymbol{x} \,|\, \theta_{m'})}_{\text{likelihood}} \Big\} \qquad (3)
\end{aligned}$$

is called the *maximum-likelihood test*; this name is easy to justify after inspecting (3) and noting that the computation of the optimal decision region $\mathcal{X}_m^\star$ requires *maximization* of the likelihood function

$$f_{\boldsymbol{X} \,|\, \Theta}(\boldsymbol{x} \,|\, \theta)$$

with respect to the discrete parameter

$$\theta \in \{\theta_0, \theta_1, \ldots, \theta_{M-1}\}.$$

Hence, $\mathcal{X}_m^\star$ is the region of $\boldsymbol{x}$ that will yield a maximum-likelihood estimate of Θ that is equal to θ_m.

0-1 Loss & Simple Hypotheses

In the familiar 0-1 loss case where $\mathrm{L}(1 \mid 0) = \mathrm{L}(0 \mid 1) = 1$, we know that the preposterior (Bayes) risk of a decision rule $\phi(\boldsymbol{x})$ is equal to the *average error probability*. This average error probability simplifies in simple hypothesis testing [see (2)]

$$\mathrm{Pr}_{\mathrm{error}} = 1 - \mathrm{Pr}_{\mathrm{correct\,decision}} = 1 - \sum_{m=0}^{M-1} \pi_m \int_{\mathcal{X}_m} f_{\boldsymbol{X} \mid \Theta}(\boldsymbol{x} \mid \theta_m)\, d\boldsymbol{x}.$$

Example: Simple Hypotheses, 0-1 Loss, Mean Testing for Multivariate Gaussian Measurements

The measurement vector $\boldsymbol{X} = \boldsymbol{x}$ follows a multivariate Gaussian model:

$$f_{\boldsymbol{X} \mid \boldsymbol{\mu}}(\boldsymbol{x} \mid \boldsymbol{\mu}) = \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu}, C)$$

where C is the known covariance matrix. Hence,

$$\ln f_{\boldsymbol{X} \mid \boldsymbol{\mu}}(\boldsymbol{x} \mid \boldsymbol{\mu}) = -\tfrac{1}{2} \ln |2\,\pi\,C| - \tfrac{1}{2}\,(\boldsymbol{x} - \boldsymbol{\mu})^T\,C^{-1}\,(\boldsymbol{x} - \boldsymbol{\mu}).$$

We wish to test

$$\begin{aligned}
\mathcal{H}_0 &\;:\; \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{versus} \\
\mathcal{H}_1 &\;:\; \boldsymbol{\mu} = \boldsymbol{\mu}_1 \quad \text{versus} \\
&\;\vdots \qquad\qquad\quad \text{versus} \\
\mathcal{H}_{M-1} &\;:\; \boldsymbol{\mu} = \boldsymbol{\mu}_{M-1}
\end{aligned}$$

Define the prior pmf for the hypotheses:

$$\pi_0 = \pi(\boldsymbol{\mu}_0), \quad \ldots, \pi_{M-1} = \pi(\boldsymbol{\mu}_{M-1}).$$

Now, the MAP rule is

$$
\begin{aligned}
\mathcal{X}_m^\star &= \big\{\boldsymbol{x} \,:\, m = \arg \max_{0 \le m' \le M-1} \pi_{m'} \, f_{\boldsymbol{X} \mid \Theta}(\boldsymbol{x} \mid \boldsymbol{\mu}_{m'})\big\} \\
&= \big\{\boldsymbol{x} \,:\, \arg \max_{0 \le m' \le M-1} [\ln \pi_{m'} - \tfrac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu}_{m'})^T \, C^{-1} \, (\boldsymbol{x} - \boldsymbol{\mu}_{m'})]\big\}.
\end{aligned}
$$

If

$$
\pi_0 = \pi_1 = \cdots = \pi_{M-1} = \frac{1}{M}
$$

we obtain the maximum-likelihood rule:

$$
\mathcal{X}_m^\star = \Big\{\boldsymbol{x} \,:\, \arg \min_{0 \le m' \le M-1} (\boldsymbol{x} - \boldsymbol{\mu}_{m'})^T \, C^{-1} \, (\boldsymbol{x} - \boldsymbol{\mu}_{m'})\Big\}.
$$

Finally, if $C = \sigma^2 \, I$ (where the noise variance σ^2 may or may not be known), we obtain the common minimum-distance rule:

$$
\mathcal{X}_m^\star = \big\{\boldsymbol{x} \,:\, \arg \min_{0 \le m' \le M-1} \|\boldsymbol{x} - \boldsymbol{\mu}_{m'}\|_{\ell_2}^2\big\}.
$$

Example: Simple Hypotheses, Multiple DC Levels in WGN

The measurements $X[0], X[1], \ldots, X[N-1]$ are conditionally i.i.d. given $\Theta = \theta$, modeled as

$$X[n] = \Theta\, a + W[n] \quad n = 0, 1, \ldots, N-1$$

where

- a is a known constant,
- Θ takes three values:

$$\theta_0 = -1, \quad \theta_1 = 0, \quad \theta_1 = 1,$$

- $W[n]$ is zero-mean white Gaussian noise with known variance σ^2, i.e.

  $$W[n] \sim \mathcal{N}(0, \sigma^2)$$

  and

- $$\pi_0 = \pi_1 = \pi_2 = \frac{1}{3}.$$

Define $\boldsymbol{x} = [x[0], x[1], \ldots, x[N-1]]^T$. Then,

$$f_{\mathbf{X}\,|\,\Theta}(\boldsymbol{x}\,|\,\theta) = \frac{1}{\sqrt{(2\,\pi\,\sigma^2)^N}} \cdot \exp\Big[-\frac{1}{2\,\sigma^2}\sum_{n=0}^{N-1}(x[n]-\theta\,a)^2\Big]$$

$$= \frac{1}{\sqrt{(2\,\pi\,\sigma^2)^N}} \cdot \exp\Big[-\frac{1}{2\,\sigma^2}\sum_{n=0}^{N-1}(x[n]-\overline{x}+\overline{x}-\theta\,a)^2\Big]$$

$$= \frac{1}{\sqrt{(2\,\pi\,\sigma^2)^N}} \cdot \exp\Big\{-\frac{N}{2\,\sigma^2}\,[s^2(\boldsymbol{x}) + (\overline{x}-\theta\,a)^2]\Big\}$$

where

$$s^2(\boldsymbol{x}) = s^2 = \frac{1}{N}\sum_{n=0}^{N-1}(x[n]-\overline{x})^2, \quad \overline{x} = \frac{1}{N}\sum_{n=0}^{N-1} x[n].$$

Note that $\overline{x}$ is a *suffecient statistic for θ*.

The three hypotheses can be written as:

$$\begin{aligned} \mathcal{H}_0 : &\quad \Theta = \theta_0 = -1 \quad \text{versus} \\ \mathcal{H}_1 : &\quad \Theta = \theta_1 = 0 \quad \text{versus} \\ \mathcal{H}_2 : &\quad \Theta = \theta_2 = 1. \end{aligned}$$

Our Bayes' rule (which is also the maximum-likelihood rule in this case) is

$$
\begin{aligned}
\mathcal{X}_m^\star &= \Big\{ \boldsymbol{x} : m = \arg \max_{m' \in \{0,1,2\}} f_{\boldsymbol{X} \,|\, \Theta}(\boldsymbol{x} \,|\, \theta_{m'}) \Big\} \\
&= \Big\{ \overline{x} : m = \arg \max_{m' \in \{0,1,2\}} f_{\overline{X} \,|\, \Theta}(\overline{x} \,|\, \theta_{m'}) \Big\} \\
&= \Big\{ \overline{x} : m = \arg \min_{m' \in \{0,1,2\}} (\overline{x} - \theta_{m'} \, a)^2 \Big\}
\end{aligned}
$$

i.e.

decide $\mathcal{H}_0$ if $\quad \overline{x} < -\frac{1}{2}\, a$

decide $\mathcal{H}_1$ if $\quad -\frac{1}{2}\, a < \overline{x} < \frac{1}{2}\, a$

decide $\mathcal{H}_0$ if $\quad \overline{x} > \frac{1}{2}\, a.$