Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science
6.437 Inference and Information
Spring 2015

Problem Set 2

Issued: Thursday, February 12, 2015 **Due:** Tuesday, February 24, 2015

Problem 2.1

Let k denote the uptime of a communication link in days. Given that the link is functioning at the beginning of a particular day, there is probability q that it will go down that day, independently of everything else. Thus, the uptime of the link k (in days) obeys a geometric distribution. If the link goes down within 24 hours from the beginning of the first day, we denote the uptime $k = 0$.

It is known that q is one of two values: q_0 or q_1. By observing the actual uptime of the link in an experiment, we would like to determine which of the two values of q applies. The hypotheses are then:

$$\begin{aligned} H_0 &: \ p_{k|H}(k|H_0) = \mathbb{P}(k = k|H = H_0) = q_0(1-q_0)^k \quad k = 0, 1, 2, \cdots \\ H_1 &: \ p_{k|H}(k|H_1) = \mathbb{P}(k = k|H = H_1) = q_1(1-q_1)^k \quad k = 0, 1, 2, \cdots . \end{aligned}$$

Suppose that $q_0 = 1/2$, $q_1 = 1/4$, and we observe $k = k$ in our experiment.

(a) Let $\mathbb{P}(H = H_0) = \mathbb{P}(H = H_1) = 1/2$. Find the decision rule $\hat{H}(k)$ corresponding to the minimum probability of error rule.

(b) Make an approximate labeled sketch of the operating characteristic for the likelihood ratio test (LRT).

(c) Find the decision rule $\hat{H}(k)$ that maximizes $P_\mathrm{D} = \mathbb{P}\left(\hat{H} = H_1 \,\middle|\, H = H_1\right)$, subject to the constraint that $P_\mathrm{F} \le 0.01$, where $P_\mathrm{F} = \mathbb{P}\left(\hat{H} = H_1 \,\middle|\, H = H_0\right)$. What is the corresponding value of P_D?

(d) We are now given $q_0 = 4/7$ and $q_1 = 1/7$. We again assume $\mathbb{P}(H = H_0) = \mathbb{P}(H = H_1) = 1/2$. Find all ML decision rules (including randomized ones), and sketch the corresponding $(P_\mathrm{F}, P_\mathrm{D})$ pairs.

Problem 2.2

Let y be a continuous random variable distributed over the closed interval $[0, 1]$. Under the null hypothesis H_0, y is uniform:

$$p_{y|H}(y|H_0) = \begin{cases} 1, & 0 \le y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Under the alternative hypothesis H_1, the conditional pdf of y is as follows:

$$p_{y|H}(y|H_1) = \begin{cases} 2y, & 0 \leq y \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

The *a priori* probability that y is uniformly distributed is p.

(a) Find the decision rule that minimizes the probability of error.

(b) Find the closed form expression for the operating characteristic of the likelihood ratio test (LRT), i.e., P_{D} as a function of P_{F} for the LRT.

(c) Suppose we require P_{D} to be at least $(1+\varepsilon)P_{\mathrm{F}}$, where $\varepsilon > 0$ is a fixed constant.

  (i) Find $P_{\mathrm{D}}^{\max}(\varepsilon)$, the maximal value of P_{D} that is achievable under this constraint.

  (ii) Find the range of values of ε that lead to non-trivial performance, i.e. $P_{\mathrm{D}}^{\max}(\varepsilon) > 0$.

  (iii) When using the decision rule from part (a), what values of p guarantee that $P_{\mathrm{D}} \geq (1+\varepsilon)\, P_{\mathrm{F}}$?

Problem 2.3

Consider the hypotheses

$$\begin{aligned} H_0 &: y_1 = -d + w_1 \\ H_1 &: y_1 = d + w_1, \end{aligned} \tag{1}$$

where the scalar y_1 is observed, where $d \neq 0$ is some known constant, and where the random variable w_1 represents measurement noise.

(a) Suppose that $d = 1/2$ and that the noise is discrete with distribution

$$p_{w_1}(w_1) = \begin{cases} 1/8 & w_1 = -1 \\ 3/8 & w_1 = 0 \\ 1/2 & w_1 = 1 \end{cases}.$$

Draw a fully-labeled sketch of the operating characteristic of the likelihood ratio test (LRT) for this hypothesis test.

In the remainder of the problem, in addition to y_1 there is a second observation y_2 that behaves under the two hypotheses according to

$$\begin{aligned} H_0 &: y_2 = -d + w_2 \\ H_1 &: y_2 = d + w_2, \end{aligned}$$

where d is the same as in (1), and where w_2 represents the measurement noise in y_2.

(b) For general (arbitrary) $p_{\mathsf{w}_1}(\cdot)$ and d, show that when

$$p_{\mathsf{w}_2}(w_2) = p_{\mathsf{w}_1}(-w_2), \quad \text{for all } w_2, \tag{2}$$

then $(P_{\mathrm{F}}, P_{\mathrm{D}}) = (1-\beta, 1-\alpha)$ is on the operating characteristic of the LRT for y_2 if $(P_{\mathrm{F}}, P_{\mathrm{D}}) = (\alpha, \beta)$ is on the operating characteristic of the LRT for y_1.

Hint: You may find it convenient to first consider the special case when the two LRTs specialize to threshold tests on y_1 and y_2, i.e., $\hat{H}(y_1) = H_1$ when $y_1 > \gamma$ for some γ, and then generalize.

(c) Suppose that d and $p_{\mathsf{w}_1}(\cdot)$ are as given in part (a), and that (2) holds, and further suppose that we can choose (potentially randomly) between measuring either y_1 or y_2, but not both. Plot the efficient frontier of $(P_{\mathrm{F}}, P_{\mathrm{D}})$ operating points for such a scenario.

(d) **(practice)** Does there exist a conditional distribution $p_{\mathsf{w}_2|\mathsf{w}_1}(w_2|w_1)$ such that (2) holds and perfect detection ($P_{\mathrm{F}} = 0, P_{\mathrm{D}} = 1$) is achievable when y_1 and y_2 are simultaneously observed? If so, specify such a conditional distribution; otherwise, explain why not.

Problem 2.4 (practice)

Let y be an *integer* random variable whose values are used to perform a binary hypothesis test using a decision rule of the following form:

$$\hat{H}(y) = \begin{cases} H_1, & y \geq \gamma \\ H_0, & \text{otherwise} \end{cases} \tag{3}$$

where γ is a known constant threshold.

We are told that $\mathbb{P}(\mathsf{H} = H_0) = 2/3$ and $\mathbb{P}(\mathsf{H} = H_1) = 1/3$. We are also given $(P_{\mathrm{F}}, P_{\mathrm{D}})$ value pairs for different values of the threshold γ:

γ	0	1	2	3
P_{F}	1	0.9	0.1	0
P_{D}	1	0.6	0.4	0

Table 1: Performance of decision rule in Eq. (3)

(a) Could this be a likelihood ratio test? Explain.

(b) Calculate the probability of error for each value of the threshold γ.

(c) Determine the probability mass functions $p_{\mathsf{y}|\mathsf{H}}(y|H_0)$ and $p_{\mathsf{y}|\mathsf{H}}(y|H_1)$.

(d) Draw and label the operating characteristic for all achievable points using decision rules of the form given in Eq. (3) and one flip of a fair coin, where the coin allows you to choose between the outcomes listed in Table 1.

(e) Is there a decision rule that achieves smaller probability of error than all decision rules of the form given in Eq. (3)? If yes, specify such a decision rule and find the corresponding probability of error. If not, explain.

Problem 2.5

A 3-dimensional random vector y is observed, and we know that one of the three hypotheses is true:

$$\begin{aligned} H_1 &: \quad \mathsf{y} = \mathbf{m}_1 + \mathsf{w} \\ H_2 &: \quad \mathsf{y} = \mathbf{m}_2 + \mathsf{w} \\ H_3 &: \quad \mathsf{y} = \mathbf{m}_3 + \mathsf{w}, \end{aligned}$$

where

$$\mathsf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \qquad \mathbf{m}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \mathbf{m}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{m}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

and w is a vector whose elements are independently, identically distributed Gaussian random variables with mean zero and variance σ^2.

(a) Let

$$\pi(\mathbf{y}) = \begin{bmatrix} \mathbb{P}(\mathsf{H} = H_1 | \mathsf{y} = \mathbf{y}) \\ \mathbb{P}(\mathsf{H} = H_2 | \mathsf{y} = \mathbf{y}) \\ \mathbb{P}(\mathsf{H} = H_3 | \mathsf{y} = \mathbf{y}) \end{bmatrix} = \begin{bmatrix} \pi_1(\mathbf{y}) \\ \pi_2(\mathbf{y}) \\ \pi_3(\mathbf{y}) \end{bmatrix},$$

and suppose that the Bayes costs are

$$C_{11} = C_{22} = C_{33} = 0, \qquad C_{12} = C_{21} = 1, \qquad C_{13} = C_{31} = C_{23} = C_{32} = 2.$$

(i) Specify the optimum decision rule in terms of $\pi_1(\mathbf{y})$, $\pi_2(\mathbf{y})$, and $\pi_3(\mathbf{y})$.

(ii) Recalling that $\pi_1 + \pi_2 + \pi_3 = 1$, express this rule completely in terms of π_1 and π_2, and sketch the decision regions in the (π_1, π_2) plane.

(b) Suppose that the three hypotheses are equally likely a priori and that the Bayes costs are

$$C_{ij} = 1 - \delta_{ij} = \begin{cases} 1, & i \neq j \\ 0, & i = j \end{cases}.$$

Show that the optimum decision rule can be specified in terms of the pair of sufficient statistics

$$\begin{aligned} \ell_2(\mathbf{y}) &= y_2 - y_1, \\ \ell_3(\mathbf{y}) &= y_3 - y_1. \end{aligned}$$

Hint: To begin, see if you can specify the optimum decision rules in terms of

$$L_i(\mathbf{y}) = \frac{p_{\mathsf{y}|H}(\mathbf{y}|H_i)}{p_{\mathsf{y}|H}(\mathbf{y}|H_1)}, \qquad \text{for } i = 2, 3.$$

Problem 2.6

A binary random variable x with prior $p_{\mathsf{x}}(\cdot)$ takes values in $\{-1, 1\}$. It is observed via n separate sensors; y_i denotes the observation at sensor i. The $\mathsf{y}_1, \ldots, \mathsf{y}_n$ are conditionally independent given x, i.e.,

$$p_{\mathsf{y}_1,\ldots,\mathsf{y}_n|\mathsf{x}}(y_1, \ldots, y_n|x) = \prod_{i=1}^{n} p_{\mathsf{y}_i|\mathsf{x}}(y_i|x).$$

A *local* decision $\hat{x}_i(y_i) \in \{-1, 1\}$ about the value of x is made at each sensor.

(a) In this part of the problem, each sensor sends its local decision to a fusion center. The fusion center combines the local decisions from all sensors to produce a global decision $\hat{x}(\hat{x}_1, \ldots, \hat{x}_n)$. Consider the special case in which: i) $p_{\mathsf{x}}(1) = p_{\mathsf{x}}(-1) = 1/2$; ii) $\mathsf{y}_i = \mathsf{x} + \mathsf{w}_i$, where $\mathsf{w}_1, \ldots, \mathsf{w}_n$ are independent and each uniformly distributed over the interval $[-2, 2]$; and iii) the local decision rule is a simple thresholding of the observation, i.e.,

$$y_i \underset{\hat{x}_i(y_i)=-1}{\overset{\hat{x}_i(y_i)=1}{\gtrless}} 0.$$

Determine the minimum probability of error decision rule, $\hat{x}(\cdot, \ldots, \cdot)$, at the fusion center.

In the remainder of the problem, there is no fusion center. The prior $p_{\mathsf{x}}(\cdot)$, observation model $p_{\mathsf{y}_i|\mathsf{x}}(\cdot|x)$, $i = 1, 2$, and local decision rules $\hat{x}_i(\cdot)$, are no longer restricted as in part (a). However, we restrict our attention to the two-sensor case ($n = 2$).

Consider local decisions $\hat{x}_i(y_i)$, $i = 1, 2$, that minimize the expected cost, where the cost is defined for the two local rules jointly. Specifically, $C(\hat{x}_1, \hat{x}_2, x)$ is the cost of deciding $\hat{x}_1$ at sensor 1 and deciding $\hat{x}_2$ at sensor 2 when the true value of x is x. The cost C strictly increases with the number of errors made by the two sensors, but is not necessarily symmetric.

(b) First, assume the local decision rule $\hat{x}_2(\cdot)$ is given. Show that the choice $\hat{x}_1^*(\cdot)$ for $\hat{x}_1(\cdot)$ that minimizes the expected (joint) cost is a likelihood ratio test of the form

$$\frac{p_{y_1|x}(y_1|1)}{p_{y_1|x}(y_1|-1)} \underset{\hat{x}_1^*(y_1)=-1}{\overset{\hat{x}_1^*(y_1)=1}{\gtrless}} \gamma_1,$$

where γ_1 is a threshold that depends on the rule $\hat{x}_2(\cdot)$. Determine the threshold γ_1.

(c) Assuming, instead, that the local decision rule $\hat{x}_1(\cdot)$ is given, determine the choice $\hat{x}_2^*(\cdot)$ for $\hat{x}_2(\cdot)$ that minimizes the expected joint cost.

(d) Consider a joint cost function $C(\hat{x}_1, \hat{x}_2, x)$ such that the cost is: 0 if both sensors making correct decisions; 1 if exactly one sensor makes an error; and L if both sensors make an error. Determine the value of L such that the optimal local decision rules at the two sensors are decoupled, i.e., the optimal threshold γ_1 does not depend on $\hat{x}_2^*(\cdot)$, and *vice versa.*

Problem 2.7

A random variable y is observed and used to decide between two hypotheses, H_0 and H_1. Under each of these two hypotheses, y is as follows:

$$\begin{aligned} H_0 &: \ \mathsf{y} = s_0 + \mathsf{w} \\ H_1 &: \ \mathsf{y} = s_1 + \mathsf{w} \end{aligned}$$

where s_0 and s_1 are known scalars with $s_0 < s_1$, and w is a zero-mean Gaussian random variable with known variance σ^2. Assume symmetric costs (i.e., $C_{00} = C_{11} = 0$, $C_{10} = C_{01} = 1$).

The *a priori* probabilities $\mathbb{P}(\mathsf{H}_0) = 1 - p$ and $\mathbb{P}(\mathsf{H}_1) = p$ are unknown; hence we wish to conduct a minimax test.

(a) What is the least favorable prior and its corresponding decision rule $\hat{H}_{\mathrm{M}}(y)$?

(b) Assuming now that the priors $1 - p$ and p are known, what is the expected cost of the minimax decision rule $\hat{H}_{\mathrm{M}}(y)$ calculated in part (a), and how does it compare to the expected cost (Bayesian risk) of the optimum Bayes' decision rule $\hat{H}_{\mathrm{B}}(y)$? Determine an expression for the expected costs in terms of the unspecified parameters, and then evaluate both costs when $p = (e^2 + 1)^{-1}$, $s_0 = 0$, $s_1 = 2$, and $\sigma^2 = 1$.