EE5137 Stochastic Processes: Problem Set 11
Assigned: 02/04/21, Due: 09/04/21

There are four (4) non-optional problems in this problem set. This is the last problem set.

1. Exercise 8.15 (Gallager's book) Consider a binary hypothesis testing problem where X is 0 or 1 and a one dimensional observation Y is given by $Y = X + U$ where U is uniformly distributed over $[-1, 1]$ and is independent of X.

   (a) Find $f_{Y|X}(y|0), f_{Y|X}(y|1)$ and the likelihood ratio $\Lambda(y)$.
   Solution: Note that $f_{Y|X}$ is simply the density of U shifted by X, i.e.,

$$f_{Y|X}(y|0) = \begin{cases} 1/2; & -1 \le y \le 1 \\ 0; & \text{elsewhere} \end{cases}, \tag{1}$$

$$f_{Y|X}(y|1) = \begin{cases} 1/2; & 0 \le y \le 2 \\ 0; & \text{elsewhere} \end{cases}. \tag{2}$$

   The likelihood ratio $\Lambda(y)$ is defined only for $-1 \le y \le 2$ since neither conditional density is non-zero outside this range.

$$\Lambda(y) = \frac{f_{Y|X}(y|1)}{f_{Y|X}(y|0)} = \begin{cases} 0; & -1 \le y < 0 \\ 1; & 0 < y \le 1 \\ \infty; & 1 < y \le 2 \end{cases}. \tag{3}$$

   (b) Find the threshold test at η for each $\eta, 0 < \eta < \infty$ and evaluate the conditional error probabilities, $q_0(\eta)$ and $q_1(\eta)$.
   Solution: Since $\Lambda(y)$ has finitely many (three) possible values, all values of η between any adjacent pair lead to the same threshold test. Thus, for $\eta > 1, \Lambda(y) \ge \eta$, if and only if (iff) $\Lambda(y) = \infty$. Thus $\hat{x} = 1$ iff $1 < y \le 2$. For $\eta = 1, \hat{x} = 1$ iff $\Lambda(y) \ge 1$, i.e., iff $\Lambda(y)$ is 1 or ∞. Thus $\hat{x} = 1$ iff $0 \le y \le 2$. For $\eta < 1, \Lambda(y) \ge \eta$ iff $\Lambda(y)$ is 1 or ∞. Thus $\hat{x} = 1$ iff $0 \le y \le 2$. Note that the MAP test is the same for $\eta = 1$ and $\eta < 1$, in both cases choosing $\hat{x} = 1$ for $0 \le y \le 2$.
   Consider $q_1(\eta)$ (the error probabiity using a threshold test at η conditional of $X = 1$). For $\eta > 1$, we have seen that $\hat{x} = 1$ (no error) for $1 < y \le 2$. This occurs with probability 1/2 given $X = 1$. Thus $q_1(\eta) = 1/2$ for $\eta > 1$. Also, for $\eta > 1, \hat{x} = 0$ for $-1 \le y \le 1$. Thus $q_0(\eta) = 0$. Reasoning in the same way for $\eta \le 1$, we have $q_1(\eta) = 0$ and $q_0(\eta) = 1/2$.

   (c) Find the error curve $u(\alpha)$ and explain carefully how $u(0)$ and $u(1/2)$ are found (hint: $u(0) = 1/2$).
   Solution: Each $\eta > 1$ maps into the pair of error probabilities $(q_0(\eta), q_1(\eta)) = (0, 1/2)$. Similarly, each $\eta \le 1$ maps into the pair of error probabilities $(q_0(\eta), q_1(\eta)) = (1/2, 0)$. The error curve contains these points and also contains the supremum of the straight lines of each slope $-\eta$ around $(0, 1/2)$ for $\eta > 1$ and around $(1/2, 0)$ for $\eta \le 1$. The resulting curve is given in Fig. 1.
   Another approach (perhaps more insightful) is to repeat (a) and (b) for the alternative threshold tests that choose $\hat{x} = 0$ in the don't care cases, i.e., the cases for $\eta = 1$ and $0 \le y \le 1$. It can

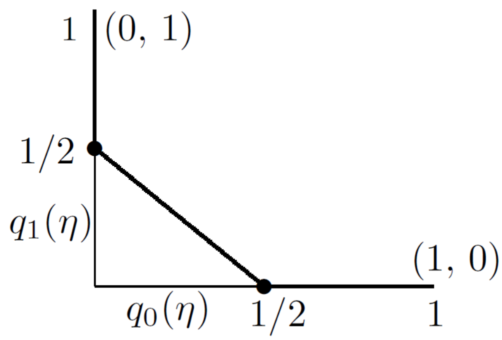


Figure 1: Error curve

be seen that Lemma 8.4.1 and Theorem 8.4.2 apply to these alternative threshold tests also. The points on the straight line between $(0, 1/2)$ and $(1/2, 0)$ can then be achieved by randomizing the choice between the threshold tests and the alternative threshold tests.

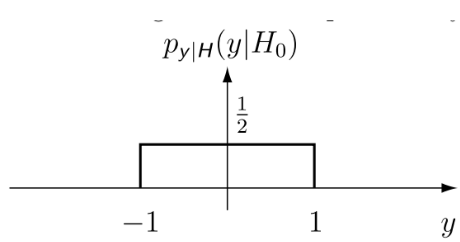
(d) Find a discrete sufficient statistic $v(y)$ for this problem that has 3 sample values.

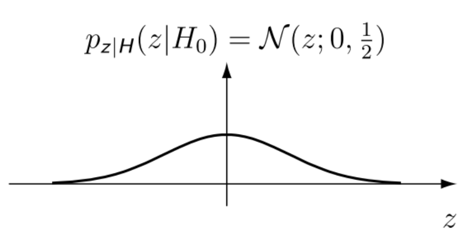
Solution: $v(y) = \Lambda(y)$ is a discrete sufficient statistic with 3 sample values.

(e) Describe a decision rule for which the error probability under each hypothesis is 1/4. You need not use a randomized rule, but you need to handle the don't-care cases under the threshold test carefully.

Solution: The don't care cases arise for $0 \leq y \leq 1$ when $\eta = 1$. With the decision rule of (8.11), these don't care cases result in $\hat{x} = 1$. If half of those don't care cases are decided as $\hat{x} = 0$, then the error probability given $X = 1$ is increased to 1/4 and that for $X = 0$ is decreased to 1/4. This could be done by random choice, or more easily, by mapping $y > 1/2$ into $\hat{x} = 1$ and $y \leq 1/2$ into $\hat{x} = 0$.

2. Consider the problem of deciding between two equally likely hypotheses based on two random variables, Y and Z. Specifically, under H_0, Y and Z are independent and have the following conditional probability densities:



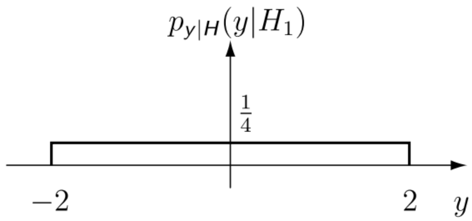


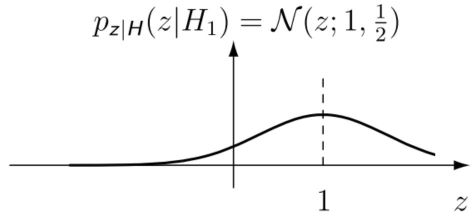
Under H_1, Y and Z are independent and have the following conditional probability densities:

(a) Specify a decision rule for deciding between H_0 and H_1, based on Y and Z, in order to minimize the probability of error.

Solution: Given y and z are conditional independent under either H_0 and H_1, we can express the ML decision rule as:

$$\frac{p_{Y,Z}(y,z|H_1)}{p_{Y,Z}(y,z|H_0)} = \frac{p_Y(y|H_1)p_Z(z|H_1)}{p_Y(y|H_0)p_Z(z|H_0)} \underset{\hat{H}(y)=H_0}{\overset{\hat{H}(y)=H_1}{\gtrless}} 1 \tag{4}$$





- For $|y| > 1, H = H_1$.
- For $|y| \leq 1$,

$$\frac{(1/4)p_Z(z|H_1)}{(1/2)p_Z(z|H_0)} \underset{\hat{H}(y)=H_0}{\overset{\hat{H}(y)=H_1}{\gtrless}} 1 \tag{5}$$

$$z \underset{\hat{H}(y)=H_0}{\overset{\hat{H}(y)=H_1}{\gtrless}} \frac{1+\ln 2}{2}. \tag{6}$$

(b) Compute $P_D = \Pr(\text{decide } H_1|H_1)$ and $P_F = \Pr(\text{decide } H_1|H_0)$ for the decision rule in part (a), expressing your answer in terms of

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, \mathrm{d}t.$$

Solution:

$$P_D = \Pr[\hat{H} = H_1|H = H_1] = \Pr\left[|Y| > 1|H_1\right] + \Pr\left[|Y| \leq 1|H_1\right] \Pr\left[Z \geq \frac{1+\ln 2}{2} \Big| H_1\right] \tag{7}$$

$$= \frac{1}{2} + \frac{1}{2}\int_{\frac{1+\ln 2}{2}}^\infty \frac{1}{\sqrt{2\pi\frac{1}{2}}} e^{-\frac{1}{2\frac{1}{2}}(z-1)^2} = \frac{1}{2} + \frac{1}{2} Q\left(\frac{\ln 2 - 1}{\sqrt{2}}\right) \tag{8}$$

$$P_F = \Pr[\hat{H} = H_1|H = H_0] = \Pr\left[|Y| > 1|H_0\right] + \Pr\left[|Y| \leq 1|H_0\right] \Pr\left[Z \geq \frac{1+\ln 2}{2} \Big| H_0\right] \tag{9}$$

$$= 0 + 1 \cdot \int_{\frac{1+\ln 2}{2}}^\infty \frac{1}{\sqrt{2\pi\frac{1}{2}}} e^{-\frac{1}{2\frac{1}{2}}z^2} = Q\left(\frac{\ln 2 + 1}{\sqrt{2}}\right). \tag{10}$$

3. Let Y_1, Y_2 and Y_3 be three IID Bernoulli random variables with $\Pr(Y_i = 1) = p$ for $i \in \{1, 2, 3\}$. This means that $\Pr(Y_i = y) = p^y(1-p)^{1-y}$ for $y \in \{0, 1\}$. It is known that p can take on two values 1/2 or 2/3. In this problem, we consider the hypothesis test

$$\mathsf{H}_0 : p = 1/2, \qquad \mathsf{H}_1 : p = 2/3$$

based on $(Y_1, Y_2, Y_3) \in \{0, 1\}^3$.

(i) (5 points) Let $T = Y_1 + Y_2 + Y_3$ be the number of ones in the random vector (Y_1, Y_2, Y_3). Let P_0 and P_1 be the distributions of Y_1, Y_2, and Y_3 under hypothesis H_0 and H_1 respectively. Write down the likelihood ratio

$$L(Y_1, Y_2, Y_3) := \frac{P_0(Y_1, Y_2, Y_3)}{P_1(Y_1, Y_2, Y_3)}$$

in terms of T. Hence, argue that T is a sufficient statistic for deciding between H_0 and H_1.

Solution: We have

$$L(Y_1, Y_2, Y_3) = \frac{P_0(Y_1)P_0(Y_2)P_0(Y_3)}{P_1(Y_1)P_1(Y_2)P_1(Y_3)} = \frac{\prod_{i=1}^{3}(\frac{1}{2})^{Y_i}(\frac{1}{2})^{1-Y_i}}{\prod_{i=1}^{3} p^{Y_i}(1-p)^{1-Y_i}} = \frac{1/8}{(2/3)^T(1/3)^{3-T}}$$

Since $L(Y_1, Y_2, Y_3)$ depends only on T, T is a sufficient statistic.

(ii) (4 points) Clearly $T \in \{0, 1, 2, 3\}$. Evaluate the values of the likelihood ratio in terms of T.

Solution: Note that $T \in \{0, 1, 2, 3\}$. Evaluating the likelihood ratio,

$$L(Y_1, Y_2, Y_3) = \begin{cases} 27/8 & T = 0 \\ 27/16 & T = 1 \\ 27/32 & T = 2 \\ 27/64 & T = 3 \end{cases}$$

(iii) (3 points) What is the best probability of missed detection $P_1(\text{declare } \mathsf{H}_0)$ if we allow the probability of false alarm $P_0(\text{declare } \mathsf{H}_1)$ to be $1/8$? What is the corresponding test in terms of T?

Solution: For probability of false alarm to be $1/8$, we need to put the threshold at $(27/64, 27/32)$ and declare that if $T > 2$, then H_1 is declared. This is because $P_0(T > 2) = P_0(T = 3) = 1/8$. Hence, the best probability of detection is $P_1(T > 2) = P_1(T = 3) = (2/3)^3 = 8/27$.

(iv) (7 points) What is the best probability of missed detection $P_1(\text{declare } \mathsf{H}_0)$ if we allow the probability of false alarm $P_0(\text{declare } \mathsf{H}_1)$ to be $1/4$? What is the corresponding test in terms of T?

Hint: You need to consider randomized tests here.

Solution: For probability of false alarm to be $1/4$, we consider that $P_0(T > 1) = 1/2$ and the corresponding probability of detection is $P_1(T > 1) = (2/3)^3 + 3(2/3)^2(1/3) = 20/27$. Hence, we need to randomize between the strategy that places the threshold at $T > 2$ and $T > 1$. Now we find $\alpha \in [0, 1]$ such that

$$\alpha\frac{1}{8} + (1-\alpha)\frac{1}{2} = \frac{1}{4}, \qquad \Longrightarrow \qquad \alpha = \frac{2}{3}$$

Thus, the best probability of detection is

$$\alpha\frac{8}{27} + (1-\alpha)\frac{20}{27} = \frac{12}{27}.$$

The best test in terms of T would be to randomize between $T > 2$ and $T > 1$ where the former has probability $2/3$.

4. A binary random variable X with prior $p_X(\cdot)$ takes values in $\{-1, 1\}$. It is observed via n separate sensors; Y_i denotes the observation at sensor i. The $Y_1, \ldots, Y_n$ are conditionally independent given X, i.e.,

$$p_{Y_1,\ldots,Y_n|X}(y_1, \ldots, y_n|x) = \prod_{i=1}^{n} p_{Y_i|X}(y_i|x).$$

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) $p_X(1) = p_X(-1) = 1/2$; ii) $Y_i = X + W_i$, where $W_1, \ldots, W_n$ 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.

Solution: Since the prior is uniform, the minimum probability of error decision rule is the same as the ML decision rule. Hence we have

$$\frac{p_{\hat{X}_1,\ldots,\hat{X}_n|X}(\hat{x}_1,\ldots,\hat{x}_n \mid 1)}{p_{\hat{X}_1,\ldots,\hat{X}_n|X}(\hat{x}_1,\ldots,\hat{x}_n \mid -1)} \underset{\hat{x}(\hat{x}_1,\ldots,\hat{x}_n)=-1}{\overset{\hat{x}(\hat{x}_1,\ldots,\hat{x}_n)=1}{\gtrless}} 1$$

Now since the observations are conditionally independent and the local decision at each sensor is only a function of the observation at that sensor we have that the local decisions are conditionally independent, i.e.,

$$p_{\hat{X}_1,\ldots,\hat{X}_n|X}(\hat{x}_1,\ldots,\hat{x}_n \mid x) = \prod_{i=1}^{n} p_{\hat{X}_i|X}(\hat{x}_i \mid x).$$

Now since the W_i's are independent and uniform on $[-2,2]$, we have

$$p_{\hat{X}_i|X}(1 \mid 1) = p_{\hat{X}_i|X}(-1 \mid -1) = 3/4, \quad i = 1,\ldots,n$$
$$p_{\hat{X}_i|X}(-1 \mid 1) = p_{\hat{X}_i|X}(1 \mid -1) = 1/4, \quad i = 1,\ldots,n$$

Denoting $n_1 = \sum_i \frac{1}{2}(\hat{x}_i+1)$, i.e., the number of sensors with local decision $\hat{x}_i = 1$, the ML decision rule becomes

$$\frac{(3/4)^{n_1}(1/4)^{n-n_1}}{(1/4)^{n_1}(3/4)^{n-n_1}} \underset{\hat{x}(\hat{x}_1,\ldots,\hat{x}_n)=-1}{\overset{\hat{x}(\hat{x}_1,\ldots,\hat{x}_n)=1}{\gtrless}} 1$$

which after some simplification is equivalent to

$$\sum_{i=1}^{n} \hat{x}_i \underset{\hat{x}(\hat{x}_1,\ldots,\hat{x}_n)=-1}{\overset{\hat{x}(\hat{x}_1,\ldots,\hat{x}_n)=1}{\gtrless}} 0$$

So the minimum probability of error decision rule at the fusion center is a majority rule.

In the remainder of the problem, there is no fusion center. The prior $p_X(\cdot)$, observation model $p_{Y_i|X}(\cdot|x)$, $i = 1,2$, 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. Assuming conditional independence, the expected cost is

$$\begin{aligned}\mathbb{E}\Big[C(\hat{X}_1,\hat{X}_2,X)\Big] &= \mathbb{E}_{Y_1,X}\Big[\mathbb{E}_{Y_2|Y_1,X}\big[C(\hat{X}_1(Y_1),\hat{X}_2(Y_2),X) \bigm| Y_1,X\big]\Big] \\ &= \mathbb{E}_{Y_1,X}\Big[\mathbb{E}_{Y_2|X}\big[C(\hat{X}_1(Y_1),\hat{X}_2(Y_2),X) \bigm| X\big]\Big]\end{aligned}$$

You can define another cost function

$$\tilde{C}(x,\hat{x}_1(y_1)) = \mathbb{E}_{Y_2|X}[C(\hat{x}_1(y_1),\hat{X}_2(Y_2),X)|X = x]$$

(b) First, assume $\hat{x}_2(\cdot)$ 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.

Solution: The optimum decision rule consists in minimizing the expected cost conditioned on a given observation, i.e.,

$$\mathbb{E}_{X|Y_1}\left[\tilde{C}(X,-1) \mid Y_1 = y_1\right] \underset{\hat{x}_1^*(y_1)=-1}{\overset{\hat{x}_1^*(y_1)=1}{\gtrless}} \mathbb{E}_{X|Y_1}\left[\tilde{C}(X,1) \mid Y_1 = y_1\right].$$

The LHS represents the expected cost when $\hat{x}_1$ decides that the hypothesis is $X = -1$ given observation y_1. Obviously if the LHS is larger than the RHS, we decide in favor of $X = 1$ since deciding that $X = 1$ is less costly. Expanding the LHS using the law of total probability, we obtain

$$\mathbb{E}_{X|Y_1}\left[\tilde{C}(X,-1) \mid Y_1 = y_1\right] = \tilde{C}(1,-1)p_{X|Y_1}(1 \mid -1) + \tilde{C}(-1,-1)p_{X|Y_1}(-1 \mid -1)$$

We can write something similar for $\mathbb{E}_{X|Y_1}\left[\tilde{C}(X,1) \mid Y_1 = y_1\right]$. Now, writing $p_{X|Y_1}$ as $p_{Y_1|X}p_X/p_{Y_1}$ (Bayes rule), rearranging and cancelling $p_{Y_1}(y_1)$, we obtain the following likelihood ratio test

$$\frac{p_{Y_1|X}(y_1|1)}{p_{Y_1|X}(y_1|-1)} \underset{\hat{x}_1^*(y_1)=-1}{\overset{\hat{x}_1^*(y_1)=1}{\gtrless}} \frac{p_X(-1)(\tilde{C}(-1,1) - \tilde{C}(-1,-1))}{p_X(1)(\tilde{C}(1,-1) - \tilde{C}(1,1))}$$

Expanding the new cost $\tilde{C}(x, \hat{x}_1(y_1))$ yields

$$\tilde{C}(x, \hat{x}_1(y_1)) = \sum_{\hat{x}_2 \in \{-1,1\}} C(\hat{x}_1(y_1), \hat{x}_2, x)p_{\hat{X}_2|\hat{X}}(\hat{x}_2|x),$$

and the optimal decision rule becomes

$$\frac{p_{Y_1|X}(y_1|1)}{p_{Y_1|X}(y_1|-1)} \underset{\hat{x}_1^*(y_1)=-1}{\overset{\hat{x}_1^*(y_1)=1}{\gtrless}} \frac{p_X(-1)\sum_{\hat{x}_2}[C(1,\hat{x}_2,-1) - C(-1,\hat{x}_2,-1)]p_{\hat{X}_2|X}(\hat{x}_2|-1)}{p_X(1)\sum_{\hat{x}_2}[C(-1,\hat{x}_2,1) - C(1,\hat{x}_2,1)]p_{\hat{X}_2|X}(\hat{x}_2|1)}$$

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

Solution: By symmetry,

$$\frac{p_{Y_2|X}(y_2|1)}{p_{Y_2|X}(y_2|-1)} \underset{\hat{x}_2^*(y_2)=-1}{\overset{\hat{x}_2^*(y_2)=1}{\gtrless}} \frac{p_X(-1)\sum_{\hat{x}_1}[C(\hat{x}_1,1,-1) - C(\hat{x}_1,-1,-1)]p_{\hat{X}_1|X}(\hat{x}_1|-1)}{p_X(1)\sum_{\hat{x}_1}[C(\hat{x}_1,-1,1) - C(\hat{x}_1,1,1)]p_{\hat{X}_1|X}(\hat{x}_1|1)}$$

(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*.

Solution: The answer is $L = 2$. Expanding the threshold γ_1 for the numerator,

$$\begin{aligned}&[C(1,1,-1) - C(-1,1,-1)]p_{\hat{X}_2|X}(1|-1) + [C(1,-1,-1) - C(-1,-1,-1)](1 - p_{\hat{X}_2|X}(1|-1))\\ &= 1 + (L-2)p_{\hat{X}_2|X}(1|-1)\end{aligned}$$

Similarly, for the denominator, we have

$$\begin{aligned}&[C(-1,1,1) - C(1,1,1)]p_{\hat{X}_2|X}(1|1) + [C(-1,-1,1) - C(1,-1,1)](1 - p_{\hat{X}_2|X}(1|1))\\ &= 1 + (2-L)p_{\hat{X}_2|X}(1|1)\end{aligned}$$

Since $p_{\hat{X}_2|X}(\hat{x}_2|x)$ depends on the second sensor's decision rule, if we want the threshold to be independent of this rule for any likelihood model, we have to pick $L = 2$. Using this choice $\gamma_1 = \gamma_2 = 1$.

This was a question I designed for a quiz while I was a Ph.D. student at MIT.

---

5. (Optional) Attempt all the hypothesis testing problems in the past year exam papers.