EE5137 Stochastic Processes: Problem Set 11
Assigned: 01/04/21, Due: Never

All the problems here are optional.

1. Exercise 8.1 (Gallager's book) In this exercise, we evaluate $\Pr\{e_\eta|\mathbf{X}=\mathbf{a}\}$ and $\Pr\{e_\eta|\mathbf{X}=\mathbf{b}\}$ for binary detection from vector signals in Gaussian noise directly from (8.40) and(8.41).

   (a) By using (8.40) for each sample value $\mathbf{y}$ of $\mathbf{Y}$, show that

   $$\mathbb{E}[\mathrm{LLR}(\mathbf{Y})|X=\mathbf{a}] = \frac{-(\mathbf{b}-\mathbf{a})^T(\mathbf{b}-\mathbf{a})}{2\sigma^2}. \tag{1}$$

   Hint: Note that, given $\mathbf{X}=\mathbf{a}, \mathbf{Y}=\mathbf{a}+\mathbf{Z}$.
   Solution: Taking the expectation of (8.40) conditional on $\mathbf{X}=\mathbf{a}$,

   $$\begin{aligned}\mathbb{E}[\mathrm{LLR}(\mathbf{Y})|X=\mathbf{a}] &= \frac{(\mathbf{b}-\mathbf{a})^T}{\sigma^2}\mathbb{E}\left[\mathbf{Y}-\frac{\mathbf{b}+\mathbf{a}}{2}\right] &(2)\\ &= \frac{(\mathbf{b}-\mathbf{a})^T}{\sigma^2}\left(\mathbf{a}-\frac{\mathbf{b}+\mathbf{a}}{2}\right), &(3)\end{aligned}$$

   from which the desired result is obvious.

   (b) Defining $\gamma = \|\mathbf{b}-\mathbf{a}\|/(2\sigma)$, show that

   $$\mathbb{E}[\mathrm{LLR}(\mathbf{Y})|X=\mathbf{a}] = -2\gamma^2. \tag{4}$$

   Solution: The result in (a) can be expressed as $\mathbb{E}[\mathrm{LLR}(\mathbf{Y})] = -\|\mathbf{b}-\mathbf{a}\|^2/(2\sigma^2)$, from which the result follows.

   (c) Show that

   $$\mathsf{Var}\left(\mathbb{E}[\mathrm{LLR}(\mathbf{Y})|X=\mathbf{a}]\right) = 4\gamma^2. \tag{5}$$

   Hint: Note that the fluctuation of $\mathrm{LLR}(\mathbf{Y})$ conditional on $\mathbf{X}=\mathbf{a}$ is $(1/\sigma^2)(\mathbf{b}-\mathbf{a})^T\mathbf{Z}$.
   Solution: Using the hint,

   $$\begin{aligned}\mathsf{Var}[\mathrm{LLR}(\mathbf{Y})] &= \frac{1}{\sigma^2}(\mathbf{b}-\mathbf{a})^T\mathbb{E}[\mathbf{Z}\mathbf{Z}^T](\mathbf{b}-\mathbf{a})\frac{1}{\sigma^2} &(6)\\ &= \frac{1}{\sigma^2}(\mathbf{b}-\mathbf{a})^T[\mathbf{I}](\mathbf{b}-\mathbf{a}) &(7)\\ &= \frac{1}{\sigma^2}\|\mathbf{b}-\mathbf{a}\|^2, &(8)\end{aligned}$$

   from which the result follows.

(d) Show that, conditional on $\mathbf{X} = \mathbf{a}$, $\text{LLR}(\mathbf{Y}) \sim \mathcal{N}(-2\gamma^2, 4\gamma^2)$. Show that, conditional on $\mathbf{X} = \mathbf{a}$, $\text{LLR}(\mathbf{Y})/(2\gamma) \sim \mathcal{N}(-\gamma, 1)$.

Solution: Conditional on $\mathbf{X} = \mathbf{a}$, we see that $\mathbf{Y} = \mathbf{a} + \mathbf{Z}$ is Gaussian and thus $\text{LLR}(\mathbf{Y})$ is also Gaussian conditional on $\mathbf{X} = \mathbf{a}$. Using the conditional mean and variance of $\text{LLR}(\mathbf{Y})$ found in (b) and (c), $\text{LLR}(\mathbf{Y}) \sim \mathcal{N}(-2\gamma^2, 4\gamma^2)$.

When the rv $\text{LLR}(\mathbf{Y})$ is divided by 2γ, the conditional mean is also divided by 2γ, and the variance is divided by $(2\gamma)^2$, leading to the desired result.

Note that $\text{LLR}(\mathbf{Y})/2\gamma$, is very different from $\text{LLR}(\mathbf{Y}/2\gamma)$. The first scales the LLR and the second scales the observation $\mathbf{Y}$. If the observation itself is scaled, the result is a sufficient statistic and the LLR is unchanged by the scaling.

(e) Show that the first half of (8.44) is valid, i.e., that

$$\Pr\{e_\eta | \mathbf{X} = \mathbf{a}\} = \Pr\{\text{LLR}(\mathbf{Y}) \geq \ln \eta | \mathbf{X} = \mathbf{a}\} = Q\left(\frac{\ln \eta}{2\gamma} + \gamma\right). \tag{9}$$

Solution: The first equality above is simply the result of a threshold test with the threshold η. The second uses the fact in (d) that $\text{LLR}(\mathbf{Y})/2\gamma$, conditional on $\mathbf{X} = \mathbf{a}$, is $\mathcal{N}(-\gamma, 1)$. This is a unit variance Gaussian rv with mean $-\gamma$. The probability that it exceeds $\ln \eta / 2\gamma$ is then $Q\left(\ln(\eta/2\gamma + \gamma\right)$.

(f) By essentially repeating (a) through (e), show that the second half of (8.44) is valid, i.e., that

$$\Pr\{e_\eta | \mathbf{X} = \mathbf{b}\} = Q\left(-\frac{\ln \eta}{2\gamma} + \gamma\right). \tag{10}$$

Solution: One can simply rewrite each equation above, but care is needed in observing that the likelihood ratio requires a convention for which hypothesis goes on top of the fraction. Thus, here the sign of the LLR is opposite to that in parts (a) to (e). This also means that the error event occurs on the opposite side of the threshold.

2. Exercise 8.4(a)–(b) (Gallager's book)

(a) Show that if $v(\mathbf{y})$ is is a sufficient statistic according to condition 1 of Theorem 8.2.8, then

$$p_{X|YV}(x|\mathbf{y}, v(\mathbf{y})) = p_{X|Y}(x|\mathbf{y}). \tag{11}$$

Solution: Let $V(\mathbf{Y})$ be the rv with sample values $v(\mathbf{y})$. Note that $\{Y = y\}$ is the same event as $\{\mathbf{Y} = \mathbf{y}\} \cap \{V = v(y)\}$. If $\mathbf{Y}$ is discrete and this event has positive probability, then (11) is obvious since the condition on both sides is the same and has positive probability. If Y is a rv with a positive density, then (11) is true if the condition $Y = y$ is replaced with $y - \delta < Y < y$. Then (11) holds if $\lim_{\delta \to 0} (1/\delta) \Pr\{y - \delta < Y \leq y\} > 0$, which is valid since Y has a positive density. This type of argument can be extended to the case where $\mathbf{Y}$ is a random vector and holds whether or not $V(\mathbf{Y})$ is a sufficient statistic. This says that $X \to \mathbf{Y} \to V$ is Markov, which is not surprising since V is simply a function of $\mathbf{Y}$.

(b) Consider the subspace of events conditional on $V(\mathbf{y}) = v$ for a given v. Show that for $\mathbf{y}$ such that $v(\mathbf{y}) = v$,

$$p_{X|\mathbf{Y}V}(x|\mathbf{y}, v(\mathbf{y})) = p_{X|V}(x|v). \tag{12}$$

Solution: We must assume (as a natural extension of (a)), that $V(\mathbf{Y})$ is a sufficient statistic, i.e., that there is a function u such that for each v, $u(v) = \Lambda(\mathbf{y})$ for all $\mathbf{y}$ such that $v(\mathbf{y}) = v$.

We also assume that $\mathbf{Y} = \mathbf{y}$ has positive probability or probability density, since the conditional probabilities don't have much meaning otherwise. Then

$$\frac{p_{X|\mathbf{Y}V}(1|\mathbf{y}, v(\mathbf{y})}{p_{X|\mathbf{Y}V}(0|\mathbf{y}, v(\mathbf{y}))} = \frac{p_{X|\mathbf{Y}}(1|\mathbf{y})}{p_{X|\mathbf{Y}}(0|\mathbf{y})} \tag{13}$$

$$= \frac{p_1}{p_0}\Lambda(\mathbf{y}) \tag{14}$$

$$= \frac{p_1}{p_0}u(v(\mathbf{y})), \tag{15}$$

where we first used (11), then Bayes' law, and then the assumption that $u(v)$ is a sufficient statistic. Since this ratio is the same for all $\mathbf{y}$ for which $v(\mathbf{y})$ has the same value, the ratio is a function of v alone,

$$\frac{p_{X|\mathbf{Y}V}(1|\mathbf{y}, v(\mathbf{y}))}{p_{X|\mathbf{Y}V}(0|\mathbf{y}, v(\mathbf{y}))} = \frac{p_{X|V}(1|v)}{p_{X|V}(0|v)}. \tag{16}$$

Finally, since $p_{X|V}(0|v) = 1 - p_{X|V}(1|v)$ and $p_{X|\mathbf{Y}V}(0|\mathbf{y}, v(\mathbf{y})) = 1 - p_{X|\mathbf{Y}V}(1|\mathbf{y}, v(\mathbf{y}))$, we see that (16) implies (12). Note that this says that $X \to V \to \mathbf{Y}$ is Markov.

3. Exercise 8.5(a)–(b) (Gallager's book)

   (a) Let $\mathbf{Y}$ be a discrete observation random vector and let $v(\mathbf{y})$ be a function of the sample values of $\mathbf{Y}$. Show that

   $$p_{\mathbf{Y}|VX}(\mathbf{y}|v(\mathbf{y}), x) = \frac{p_{\mathbf{Y}|X}(\mathbf{y}|x)}{p_{V|X}(v(\mathbf{y})|x)}. \tag{17}$$

   Solution: We must assume that $p_{VX}(v(\mathbf{y}), x) > 0$ so that the expression on the left is defined. Then, using Bayes' law on $\mathbf{Y}$ and V for fixed x on the left side of (17),

   $$p_{\mathbf{Y}|VX}(\mathbf{y}|v(\mathbf{y}), x) = \frac{p_{V|\mathbf{Y}X}(v(\mathbf{y})|\mathbf{y}x)p_{\mathbf{Y}|X}(\mathbf{y}|x)}{p_{V|X}(v(\mathbf{y})|x)}. \tag{18}$$

   Since V is a deterministic function of $\mathbf{Y}$, the first term in the numerator above is 1, so this is equivalent to (17).

   (b) Using Theorem 8.2.8, show that the above fraction is independent of X if and only if $v(\mathbf{y})$ is a sufficient statistic.

   Solution: Using Bayes' law on the numerator and denominator of (17),

   $$\frac{p_{\mathbf{Y}|X}(\mathbf{y}|x)}{p_{V|X}(v(\mathbf{y})|x)} = \frac{p_{X|\mathbf{Y}}(x|\mathbf{y})p_{\mathbf{Y}}(\mathbf{y})}{p_{X|V}(x|v(\mathbf{y}))p_V(v(\mathbf{y}))}. \tag{19}$$

   If $v(\mathbf{y})$ is a sufficient statistic, then the second equivalent statement of Theorem 8.2.8, i.e.,

   $$p_{X|\mathbf{Y}}(x|\mathbf{y}) = p_{X|V}(x|v(\mathbf{y})), \tag{20}$$

   shows that the first terms on the right side of (19) cancel, showing that the fraction is independent of x. Conversely, if the fraction is independent of x, then the ratio of $p_{X|\mathbf{Y}}(x|\mathbf{y})$ to $p_{X|V}(x|v(\mathbf{y}))$ is a function only of y. Since X is binary, this fraction must be 1, establishing the second statement of Theorem 8.2.8.

4. Exercise 8.6 (Gallager's book)

(a) Consider Example 8.2.5, and let $Z = [A]\mathbf{W}$ where $\mathbf{W} \sim \mathcal{N}(0, I)$ is normalized IID Gaussian and $[A]$ is non-singular. The observation rv $\mathbf{Y}$ is $\mathbf{a}+\mathbf{Z}$ given $\mathbf{X} = \mathbf{a}$ and is $\mathbf{b}+\mathbf{Z}$ given $\mathbf{X} = \mathbf{b}$. Suppose the observed sample value $\mathbf{y}$ is transformed into $\mathbf{v} = [A^{-1}]\mathbf{y}$. Explain why $\mathbf{v}$ is a sufficient statistic for this detection problem (and thus why MAP detection based on $\mathbf{v}$ must yield the same decision as that based on $\mathbf{y}$).

Solution: $\mathbf{v}$ is a sufficient statistic if there is a function u such that $u(\mathbf{v}(\mathbf{y})) = \Lambda(\mathbf{y})$. Since $\mathbf{v}$ is an invertible function of $\mathbf{y}$, we can clearly map $\mathbf{v}$ back into $\mathbf{y}$ and map that into $\Lambda(\mathbf{y})$. In equations, using (8.64),

$$\Lambda(\mathbf{y}) = \frac{f_{\mathbf{V}|\mathbf{X}}(\mathbf{v}(\mathbf{y})|\mathbf{b})}{f_{\mathbf{V}|\mathbf{X}}(\mathbf{v}(\mathbf{y})|\mathbf{a})}. \tag{21}$$

(b) Consider the detection problem where $\mathbf{V} = [A^{-1}]\mathbf{a} + \mathbf{W}$ given $\mathbf{X} = \mathbf{a}$ and $[A^{-1}]\mathbf{b} + \mathbf{W}$ given $\mathbf{X} = \mathbf{b}$. Find the log-likelihood ratio $\mathrm{LLR}(\mathbf{v})$ for a sample value $\mathbf{v}$ of $\mathbf{V}$. Show that this is the same as the log-likelihood ratio for a sample value $\mathbf{y} = [A]\mathbf{v}$ of $\mathbf{Y}$.

Solution: The problem $\mathbf{V} = [A^{-1}]\mathbf{X} + \mathbf{W}$ is the same as that in Example 8.2.3 except that the received signal is $\mathbf{V}$ rather than $\mathbf{Y}$, the transmitted signals are $[A^{-1}]\mathbf{a}$ and $[A^{-1}]\mathbf{b}$, and $\sigma^2 = 1$. The log likelihood ratio, from (8.40) is

$$\mathrm{LLR}(\mathbf{v}) = \left([A^{-1}]\mathbf{b} - [A^{-1}]\mathbf{a}\right)^T \left(\mathbf{v} - \frac{1}{2}\left([A^{-1}]\mathbf{b} + [A^{-1}]\mathbf{a}\right)\right) \tag{22}$$

$$\mathrm{LLR}(\mathbf{v}(\mathbf{y})) = (\mathbf{b}-\mathbf{a})^T[A^{-1}]^T \left([A^{-1}]\mathbf{y} - [A^{-1}]\frac{1}{2}(\mathbf{b}+\mathbf{a})\right) \tag{23}$$

$$= (\mathbf{b}-\mathbf{a})^T[A^{-1}]^T[A^{-1}] \left(\mathbf{y} - \frac{(\mathbf{b}+\mathbf{a})}{2}\right). \tag{24}$$

This is the same as $\mathrm{LLR}(\mathbf{y})$ as given in (8.45) after making the observation that $[K_{\mathbf{Z}}] = [A^T A]$.

(c) Find $\Pr\{e|\mathbf{X} = \mathbf{a}\}$ and $\Pr\{e|\mathbf{X} = \mathbf{b}\}$ for the detection problem in (b) by using the results of Example 8.2.3. Show that your answer agrees with (8.53). Note: the methodology here is to transform the observed sample value to make the noise IID; this approach is often both useful and insightful.

Solution: The error probability using the results of Example 8.2.3 is given by (8.44) in terms of γ, which is defined as

$$\gamma = \frac{1}{2}\|[A^{-1}]\mathbf{b} - [A^{-1}]\mathbf{a}\| \tag{25}$$

$$= \frac{1}{2}\|[A^{-1}](\mathbf{b}-\mathbf{a})\| \tag{26}$$

$$= \frac{1}{2}\sqrt{(\mathbf{b}-\mathbf{a})^T[A^{-1}]^T[A^{-1}](\mathbf{b}-\mathbf{a})}. \tag{27}$$

The result in (8.53) is the same equation, except that γ is defined in (8.50) as

$$\gamma = \sqrt{\frac{(\mathbf{b}-\mathbf{a})^T}{2}[K_{\mathbf{Z}}^{-1}]\frac{(\mathbf{b}-\mathbf{a})}{2}}. \tag{28}$$

Since $[K_{\mathbf{Z}}] = [A^T A]$, these are the same.

5. Exercise 8.7 (Gallager's book) Binary frequency shift keying (FSK) with incoherent reception can be modeled in terms of a 4 dimensional observation vector $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)^T$. $\mathbf{Y} = \mathbf{U} + \mathbf{Z}$ where $\mathbf{Z} \sim \mathcal{N}(0, \sigma^2 I)$ and $\mathbf{Z}$ is independent of X. Under $X = 0, \mathbf{U} = (a\cos\phi, a\sin\phi, 0, 0)^T$, whereas under $X = 1, \mathbf{U} = (0, 0, a\cos\phi, a\sin\phi)^T$. The random variable ϕ is uniformly distributed between 0 and 2π and is independent of X and $\mathbf{Z}$. The a priori probabilities are $p_0 = p_1 = 1/2$.

(a) Convince yourself from the circular symmetry of the situation that the ML receiver calculates the sample values v_0 and v_1 of $V_0 = Y_1^2 + Y_2^2$ and $V_1 = Y_3^2 + Y_4^2$ and chooses $\hat{x} = 0$ if $v_0 \geq v_1$ and chooses $\hat{x} = 1$ otherwise.

Solution: It is easier to visualize the here if we express the 4-rv's as pairs of complex rv's. That is, $U = (ae^{i\phi}, 0)$ under $X = 0$ and $U = (0, ae^{i\phi})$ under $X = 1$. Similarly, $\mathbf{Z} = (R_0 e^{i\theta_0}, R_1 e^{i\theta_1})$ where R_0 and R_1 are Rayleigh rv's and are independent of the other rv's as in the problem description. The received complex rv's are $(\sqrt{V_0}e^{i\psi_0}, \sqrt{V_1}e^{i\psi_1})$. The phases ϕ, θ_0, and θ_1 are each uniformly distributed over $[0, 2\pi)$ and are independent of each other and of the other variables as described. The received phases, ψ_0 and ψ_1 are complicated functions of the other phases, but it is intuitively clear if one draws a picture of $\sqrt{V_0}e^{i\psi_0}$ that ψ_0 is uniformly distributed over $[0, 2\pi)$ and is independent of V_0, V_1, and X. In the same way, ψ_1 is uniform and independent of V_0, V_1, and X. It follows that V_0 and V_1 form a sufficient statistic for X.

(b) Find $\Pr\{V_1 > v_1 | X = 0\}$ as a function of $v_1 > 0$.

Solution: Conditional on $X = 0, V_1 = |Z_1|^2$. Since $|Z_1|^2$ is the sum of the squares of 2 IID real Gaussian rv's each of mean σ^2,

$$\Pr\{V_1 > v_1 | X = 0\} = \exp(-v_1/2\sigma^2). \tag{29}$$

(c) Show that

$$f_{Y_1,Y_2|X,\phi}(y_1, y_2|0, 0) = \frac{1}{2\pi\sigma^2} \exp\left[\frac{-y_1^2 - y_2^2 + 2y_1 a - a^2}{2\sigma^2}\right]. \tag{30}$$

Solution: Conditional on $X = 0$ and $\phi = 0$, Y_1 is $\mathcal{N}(a, \sigma^2)$ and Y_2 is $\mathcal{N}(0, \sigma^2)$ and independent of Y_1. Thus their conditional joint density is as shown.

(d) Show that

$$\Pr\{V_1 > V_0 | X = 0, \phi = 0\} = \int f_{Y_1,Y_2|X,\phi}(y_1, y_2|0, 0) \Pr\{V_1 > y_1^2 + y_2^2 | X = 0\} dy_1 dy_2. \tag{31}$$

Show that this is equal to $(1/2) \exp(-a^2/(4\sigma^2))$.

Solution: The right side of (31) is $\Pr\{V_1 > Y_1^2 + Y_2^2 | X = 0, \phi = 0\}$. Recalling that $V_0 = Y_1^2 + Y_2^2$, we see that (31) is satisfied. Substituting (29) and (30) into (31), we get

$$\Pr\{V_1 > V_0 | X = 0, \phi = 0\} = \int \frac{1}{2\pi\sigma^2} \exp\left[\frac{-y_1^2 - y_2^2 + 2y_1 a - a^2}{2\sigma^2}\right] \left[\frac{\exp(-y_1^2 - y_2^2)}{2\sigma^2}\right] dy_1 dy_2 \tag{32}$$

$$= \frac{1}{2\pi\sigma^2} \exp\left[\frac{-y_1^2 - y_2^2 + 2y_1 a - a^2}{2\sigma^2}\right] dy_1 dy_2 \tag{33}$$

$$= \frac{1}{2} \exp\left(\frac{-a^2}{4\sigma^2}\right). \tag{34}$$

(e) Explain why this is the probability of error (i. e., why the event $V_1 > V_0$ is independent of ϕ), and why $\Pr\{e|X = 0\} = \Pr\{e|X = 1\}$.

Solution: If we use an arbitrary phase ϕ in (29), we can express Z_1, Z_2 in a coordinate system rotated by ϕ to see that (29) is still valid. Thus (34) is valid for all ϕ and thus valid with no conditioning on ϕ. The probability of error conditional on $X = 1$ is obviously the same because of the symmetry between $X = 0$ and $X = 1$.

6. Exercise 8.9 (Gallager's book) A disease has two strains, 0 and 1, which occur with a priori probabilities p_0 and $p_1 = 1 - p_0$ respectively.

(a) Initially, a rather noisy test was developed to find which strain is present for patients with the disease. The output of the test is the sample value y_1 of a random variable Y_1. Given strain 0 $(X = 0), Y_1 = 5 + Z_1$, and given strain 1 $(X = 1), Y_1 = 1 + Z_1$. The measurement noise Z_1 is independent of X and is Gaussian, $Z_1 \sim \mathcal{N}(0, \sigma^2)$. Give the MAP decision rule, i.e., determine the set of observations y_1 for which the decision is $\hat{x} = 1$. Give $\Pr\{e|X = 0\}$ and $\Pr\{e|X = 1\}$ in terms of the function $Q(x)$.

Solution: This is simply a case of binary detection with an additive Gaussian noise rv. To prevent simply copying the answer from Example 8.2.3, the signal a associated with $X = 0$ is 5 and the signal b associated with $X = 1$ is 1. Thus $b < a$, contrary to the assumption in Example 8.2.3. Looking at that example, we see that (8.27), repeated below, is still valid.

$$\mathrm{LLR}(y) = \left[\left(\frac{b-a}{\sigma^2} \right) \left(y - \frac{b+a}{2} \right) \right] \underset{\hat{x}(y)=a}{\overset{\hat{x}(y)=b}{\gtrless}} \ln(\eta). \tag{35}$$

We can get a threshold test on y directly by first taking the negative of this expression and then dividing both sides by the positive term $(a-b)/\sigma^2$ to get

$$y \underset{\hat{x}(y)=a}{\overset{\hat{x}(y)=b}{\gtrless}} \frac{-\sigma^2 \ln(\eta)}{a-b} + \frac{b+a}{2}. \tag{36}$$

We get the same equation by switching the association of $X = 1$ and $X = 0$, which also changes the sign of the log threshold.

(b) A budding medical researcher determines that the test is making too many errors. A new measurement procedure is devised with two observation random variables Y_1 and Y_2. Y_1 is the same as in (a). Y_2, under hypothesis 0, is given by $Y_2 = 5 + Z_1 + Z_2$, and, under hypothesis 1, is given by $Y_2 = 1 + Z_1 + Z_2$. Assume that Z_2 is independent of both Z_1 and X, and that $Z_2 \sim \mathcal{N}(0, \sigma^2)$. Find the MAP decision rule for $\hat{x}$ in terms of the joint observation (y_1, y_2), and find $\Pr\{e|X = 0\}$ and $\Pr\{e|X = 1\}$. Hint: Find $f_{Y_2|Y_1,X}(y_2|y_1, 0)$ and $f_{Y_2|Y_1,X}(y_2|y_1, 1)$.

Solution: Note that Y_2 is simply Y_1 plus the noise term Z_2, and that Z_2 is independent of X and Y_1. Thus, Y_2, conditional on Y_1 and X is simply $\mathcal{N}(Y_1, \sigma^2)$, which is independent of X. Thus Y_1 is a sufficient statistic and Y_2 is irrelevant. Including Y_2 does not change the probability of error.

(c) Explain in laymen's terms why the medical researcher should learn more about probability.

Solution: It should have been clear intuitively that adding an additional observation that is only a noisy version of what has already been observed will not help in the decision, but knowledge of probability sharpens one's intuition so that something like this becomes self evident even without mathematical proof.

(d) Now suppose that Z_2, in (b), is uniformly distributed between 0 and 1 rather than being Gaussian. We are still given that Z_2 is independent of both Z_1 and X. Find the MAP decision rule for $\hat{x}$ in terms of the joint observation (y_1, y_2) and find $\Pr\{e|X = 0\}$ and $\Pr\{e|X = 1\}$.

Solution: The same argument as in (b) shows that Y_2, conditional on Y_1, is independent of X, and thus the decision rule and error probability do not change.

(e) Finally, suppose that Z_1 is also uniformly distributed between 0 and 1. Again find the MAP decision rule and error probabilities.

Solution: By the same argument as before, Y_2, conditional on Y_1 is independent of X, so Y_1 is a sufficient statistic and Y_2 is irrelevant. Since Z_1 is uniformly distributed between 0 and 1, then Y_1 lies between 5 and 6 for $X = 0$ and between 1 and 2 for $X = 1$. There is thus no possibility of error in this case.

7. 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.,

$$f_{Y|X}(y|0) = \begin{cases} 1/2; & -1 \leq y \leq 1 \\ 0; & \text{elsewhere} \end{cases}, \tag{37}$$

$$f_{Y|X}(y|1) = \begin{cases} 1/2; & 0 \leq y \leq 2 \\ 0; & \text{elsewhere} \end{cases}. \tag{38}$$

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

$$\Lambda(y) = \frac{f_{Y|X}(y|1)}{f_{Y|X}(y|0)} = \begin{cases} 0; & -1 \leq y < 0 \\ 1; & 0 < y \leq 1 \\ \infty; & 1 < y \leq 2 \end{cases}. \tag{39}$$

(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 $\eta > 1, \Lambda(y) \geq \eta$, if and only if (iff) $\Lambda(y) = \infty$. Thus $\hat{x} = 1$ iff $1 < y \leq 2$. For $\eta = 1, \hat{x} = 1$ iff $\Lambda(y) \geq 1$, i.e., iff $\Lambda(y)$ is 1 or ∞. Thus $\hat{x} = 1$ iff $0 \leq y \leq 2$. For $\eta < 1, \Lambda(y) \geq \eta$ iff $\Lambda(y)$ is 1 or ∞. Thus $\hat{x} = 1$ iff $0 \leq y \leq 2$. Note that the MAP test is the same for $\eta = 1$ and $\eta < 1$, in both cases choosing $\hat{x} = 1$ for $0 \leq y \leq 2$.

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 $1 < y \leq 2$. 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 $-1 \leq y \leq 1$. Thus $q_0(\eta) = 0$. Reasoning in the same way for $\eta \leq 1$, 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 $\eta \leq 1$ 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 $\eta \leq 1$. The resulting curve is given in Fig. 1.

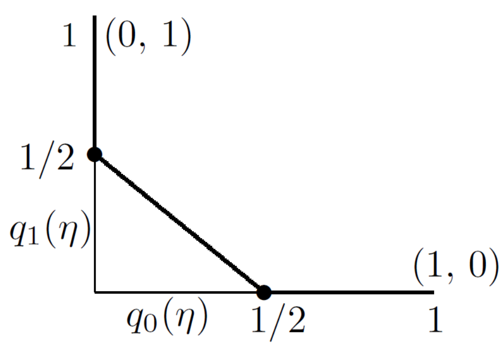


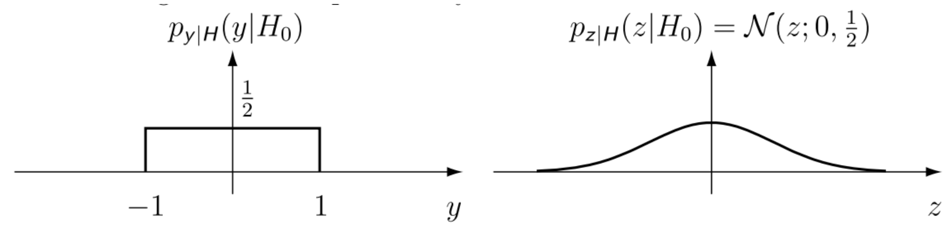
Figure 1: Error curve

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 $0 \leq y \leq 1$. It can 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.

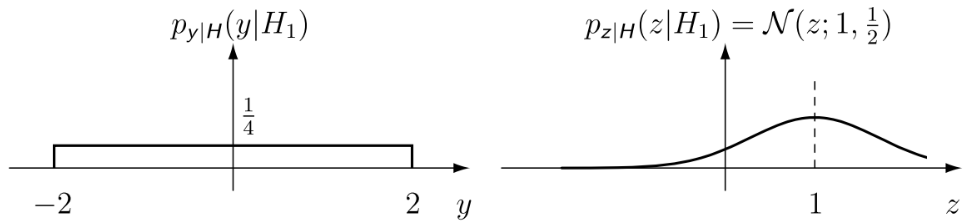
(d) Find a discrete sufficient statistic $v(y)$ for this problem that has 3 sample values.
Solution: $v(y) = \Lambda(y)$ is a discrete sufficient statistic with 3 sample values.

(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 $0 \le y \le 1$ 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 $y \le 1/2$ into $\hat{x} = 0$.

8. 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:

$$\frac{p_{Y,Z}(y, z|H_1)}{p_{Y,Z}(y, z|H_0)} = \frac{p_Y(y|H_1)p_Z(z|H_1)}{p_Y(y|H_0)p_Z(z|H_0)} \underset{\hat{H}(y)=H_0}{\overset{\hat{H}(y)=H_1}{\gtrless}} 1 \tag{40}$$

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

$$\frac{(1/4)p_Z(z|H_1)}{(1/2)p_Z(z|H_0)} \underset{\hat{H}(y)=H_0}{\overset{\hat{H}(y)=H_1}{\gtrless}} 1 \tag{41}$$

$$z \underset{\hat{H}(y)=H_0}{\overset{\hat{H}(y)=H_1}{\gtrless}} \frac{1 + \ln 2}{2}. \tag{42}$$

(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:

$$P_D = \Pr[\hat{H} = H_1|H = H_1] = \Pr\left[|Y| > 1|H_1\right] + \Pr\left[|Y| \le 1|H_1\right]\Pr\left[Z \ge \frac{1+\ln 2}{2}\Big|H_1\right] \tag{43}$$

$$= \frac{1}{2} + \frac{1}{2}\int_{\frac{1+\ln 2}{2}}^\infty \frac{1}{\sqrt{2\pi\frac{1}{2}}} e^{-\frac{1}{2\frac{1}{2}}(z-1)^2}\,\mathrm{d}z = \frac{1}{2} + \frac{1}{2}Q\left(\frac{\ln 2 - 1}{\sqrt{2}}\right) \tag{44}$$

$$P_F = \Pr[\hat{H} = H_1|H = H_0] = \Pr\left[|Y| > 1|H_0\right] + \Pr\left[|Y| \le 1|H_0\right]\Pr\left[Z \ge \frac{1+\ln 2}{2}\Big|H_0\right] \tag{45}$$

$$= 0 + 1 \cdot \int_{\frac{1+\ln 2}{2}}^\infty \frac{1}{\sqrt{2\pi\frac{1}{2}}} e^{-\frac{1}{2\frac{1}{2}}z^2}\,\mathrm{d}z = Q\left(\frac{\ln 2 + 1}{\sqrt{2}}\right). \tag{46}$$

9. 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.

10. 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

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

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 $\mathbb{E}_{X|Y_1}\big[\tilde{C}(X, 1) \mid Y_1 = y_1\big]$. 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.