Problem 1 (Detection) *20 points*

(*This is a two-page problem with two different scenarios.*)

Scenario 1: Under H_0, Y is Gaussian$(1,1)$. Under H_1, Y is Gaussian$(3,1)$. $\mathbb{P}[H_0] = \mathbb{P}[H_1] = 1/2$.

(a) For Scenario 1, work out the decision rule that will result in the smallest possible probability of error. Simplify your expression as much as possible.

$$f_{Y|H_0}(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y-1)^2}{2}\right) \qquad f_{Y|H_1}(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y-3)^2}{2}\right)$$

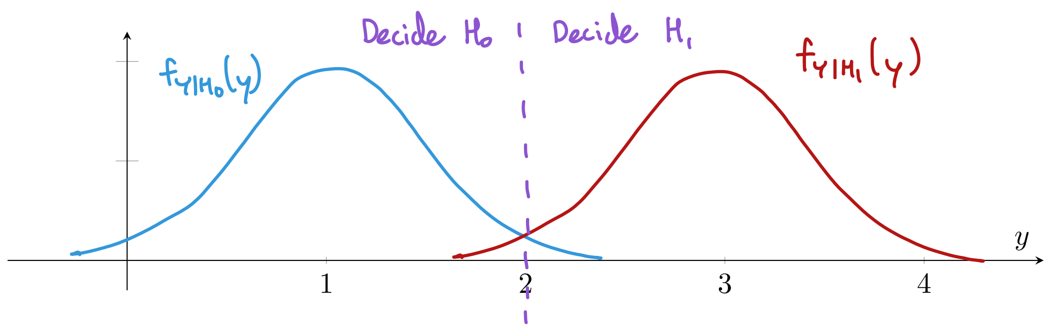
$$\ln(L(y)) = -\frac{(y-3)^2}{2} + \frac{(y-1)^2}{2} = \frac{1}{2}\left(\cancel{-y^2} + 6y - 9 + \cancel{y^2} - 2y + 1\right)$$

$$= 2y - 4$$

Since $\mathbb{P}[H_0] = \mathbb{P}[H_1] = \frac{1}{2}$, ML rule is optimal.

$$D^{ML}(y) = \begin{cases} 1 & 2y-4 \geq 0 \\ 0 & 2y-4 < 0 \end{cases} = \begin{cases} 1 & y \geq 2 \\ 0 & y < 2 \end{cases}$$

(b) Sketch the likelihoods $f_{Y|H_0}(y)$ and $f_{Y|H_1}(y)$ from Scenario 1 on the following plot, along with a vertical line that clearly indicates your decision boundary.



(c) Determine the probability of error for your decision rule for Scenario 1. You may leave your answer in terms of the Φ function.

$$P_{FA} = \mathbb{P}[Y \geq 2 \mid H_0] = 1 - F_{Y|H_0}(2) = 1 - \Phi\left(\frac{2-1}{1}\right) = 1 - \Phi(1) = \Phi(-1)$$

$$P_{MD} = \mathbb{P}[Y < 2 \mid H_1] = F_{Y|H_1}(2) = \Phi\left(\frac{2-3}{1}\right) = \Phi(-1)$$

$$P_e = P_{FA}\,\mathbb{P}[H_0] + P_{MD}\,\mathbb{P}[H_1]$$

$$= \Phi(-1) \cdot \frac{1}{2} + \Phi(-1) \cdot \frac{1}{2} = \Phi(-1)$$

(CONTINUED ON NEXT PAGE)

Scenario 2: Under H_0, $\underline{Y}$ is Gaussian$(\underline{\mu}_0, \mathbf{I})$. Under H_1, Y is Gaussian$(\underline{\mu}_1, \mathbf{I})$ where

$$\underline{\mu}_0 = \begin{bmatrix}1\\1\end{bmatrix},\ \underline{\mu}_1 = \begin{bmatrix}3\\3\end{bmatrix}, \text{ and } \mathbf{I} = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}.\ \mathbb{P}[H_0] = \mathbb{P}[H_1] = 1/2.$$

It will be helpful to know that the PDF for a Gaussian vector $\underline{Y} = \begin{bmatrix}Y_1\\Y_2\end{bmatrix}$ with mean vector $\underline{\mu}$ and covariance matrix $\mathbf{\Sigma}$ is

$$f_{\underline{Y}}(\underline{y}) = \frac{1}{\sqrt{(2\pi)^2 \det(\mathbf{\Sigma})}} \exp\left(-\frac{1}{2}(\underline{y} - \underline{\mu})^\mathsf{T} \mathbf{\Sigma}^{-1} (\underline{y} - \underline{\mu})\right)$$

(d) For Scenario 2, work out the decision rule that will result in the smallest possible probability of error. Simplify your expression as much as possible.

$$f_{\underline{Y}|H_0}(\underline{y}) = \frac{1}{2\pi} \exp\left(-\tfrac{1}{2}\left\| \underline{y} - \begin{bmatrix}1\\1\end{bmatrix}\right\|^2\right) \qquad f_{\underline{Y}|H_1}(\underline{y}) = \frac{1}{2\pi} \exp\left(-\tfrac{1}{2}\left\| \underline{y} - \begin{bmatrix}3\\3\end{bmatrix}\right\|^2\right)$$

$$\ln(L(\underline{y})) = -\tfrac{1}{2}\left(\left\| \underline{y} - \begin{bmatrix}3\\3\end{bmatrix}\right\|^2 - \left\| \underline{y} - \begin{bmatrix}1\\1\end{bmatrix}\right\|^2\right) = 3(y_1+y_2) - 9 - (y_1+y_2) + 1$$
$$= 2(y_1+y_2) - 8$$

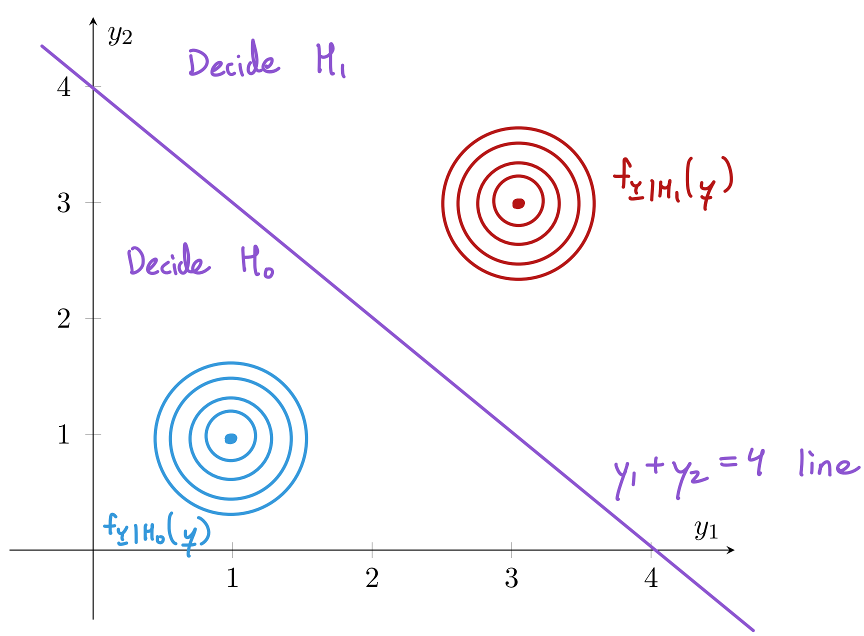
Since $\mathbb{P}[H_0] = \mathbb{P}[H_1] = \frac{1}{2}$, ML rule is optimal.

$$D^{ML}(\underline{y}) = \begin{cases} 1 & \left\|\underline{y} - \begin{bmatrix}3\\3\end{bmatrix}\right\| \leq \left\|\underline{y} - \begin{bmatrix}1\\1\end{bmatrix}\right\| \\ 0 & \left\|\underline{y} - \begin{bmatrix}3\\3\end{bmatrix}\right\| > \left\|\underline{y} - \begin{bmatrix}1\\1\end{bmatrix}\right\| \end{cases}$$

This suffices. Same as the closest average classifier.

$$= \begin{cases} 1 & y_1 + y_2 \geq 4 \\ 0 & y_1 + y_2 < 4 \end{cases}$$

(e) Sketch the likelihoods $f_{\underline{Y}|H_0}(\underline{y})$ and $f_{\underline{Y}|H_1}(\underline{y})$ from Scenario 2 as contour plots on the following plot, along with a line that clearly indicates your decision boundary.



Problem 2 (Estimation) *20 points*

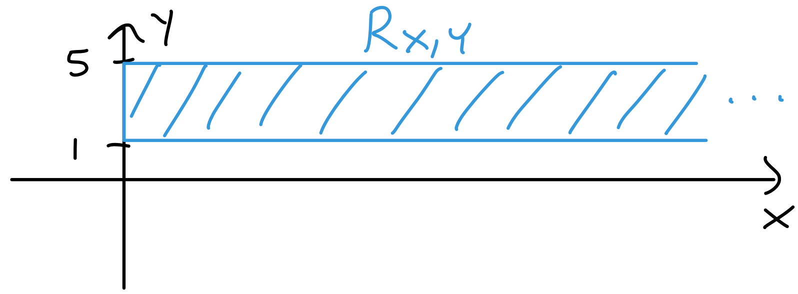
(*This is a two-page problem with a single scenario.*)

Consider the following scenario: X given $Y = y$ is Exponential$(2/y)$ and Y is Uniform$(1,5)$. Note that for parts (b), (c), and (d), it is possible to obtain an answer without integration. However, it is fine if you write your answer in terms of an integral.

(a) What is the range $R_{X,Y}$?

$$R_{X,Y} = \{(x,y) : x \geq 0, \; 1 \leq y \leq 5\}$$

a sketch would also be OK



(b) What is the MMSE estimator of X given $Y = y$?

Recall if Z is Exponential(λ), $\mathbb{E}[Z] = \frac{1}{\lambda}$.

Since X given $Y=y$ is Exponential$\left(\frac{2}{y}\right)$,

$$\mathbb{E}[X \mid Y=y] = \frac{1}{\left(\frac{2}{y}\right)} = \frac{y}{2}$$

an integral is also OK. $f_{X|Y}(x|y) = \begin{cases} \frac{2}{y} e^{-\frac{2}{y}x}, & x \geq 0, \; 1 \leq y \leq 5 \\ 0 & \text{o/w} \end{cases}$

$$\mathbb{E}[X \mid Y=y] = \int_0^\infty \frac{2x}{y} e^{\frac{-2x}{y}} \, dx$$

(c) Calculate $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

$$\mathbb{E}[Y] = \frac{1+5}{2} = 3 \qquad \mathbb{E}[X] = \mathbb{E}[\mathbb{E}(X|Y)] = \mathbb{E}\left[\frac{Y}{2}\right] = \frac{3}{2}$$

Integrals are also OK.

$$\mathbb{E}[Y] = \int_1^5 y \cdot \tfrac{1}{4} \, dy \qquad \mathbb{E}[X] = \int_1^5 \int_0^\infty \frac{x}{2y} e^{\frac{-2x}{y}} \, dx \, dy$$

(CONTINUED ON NEXT PAGE)

(d) What is the LLSE estimator of X given $Y = y$?

Since the MMSE estimator is linear, we know the LLSE estimator is the same, $\hat{x}_{LLSE}(y) = \frac{y}{2}$.

Alternatively, $\hat{x}_{LLSE}(y) = \mathbb{E}[X] + \frac{Cov[X,Y]}{Var[Y]}(y - \mathbb{E}[Y])$

$\mathbb{E}[X]$, $\mathbb{E}[Y]$ in part (c).

$Var[Y] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$ with $\mathbb{E}[Y^2] = \int_1^5 y^2 \frac{1}{4} dy$

or $Var[Y] = \frac{(5-1)^2}{12} = \frac{16}{12} = \frac{4}{3}$.

$Cov[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ with $\mathbb{E}[XY] = \int_1^5 \int_0^\infty \frac{x}{2} e^{\frac{-2x}{y}} dx\, dy$

(e) What is the mean-squared error of the LLSE estimator from part (d)?

$MSE_{LLSE} = Var[X] - \frac{(Cov[X,Y])^2}{Var[Y]}$

$Cov[X,Y]$, $Var[Y]$ in part (d).

$Var[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ with $\mathbb{E}[X^2] = \int_1^5 \int_0^\infty \frac{x^2}{2y} e^{\frac{-2x}{y}} dx\, dy$

Problem 3 (Statistics and Sums of Random Variables) *20 points*

The L & L Candy Company claims that the average bag contains 100 candies. After purchasing 64 bags, you calculate that your sample average is 98 and your sample variance is 16. You decide to use a significance test to determine whether you should believe the company's claim.

(a) What kind of significance test should you use?

> **Solution:**
>
> Since we are comparing the mean of one dataset with more than 30 samples to a known baseline, a one-sample Z-test is appropriate.

(b) Should we reject the null hypothesis at a significance level of 0.01?

> **Solution:**
>
> Since we have more than 30 samples, we will use the sample variance V_n in place of σ^2. The Z-statistic is $Z = \frac{\sqrt{n}(M_n - \mu)}{\sqrt{V_n}} = \frac{\sqrt{64}(98-100)}{\sqrt{16}} = \frac{8 \cdot (-2)}{4} = -4$
> The corresponding p-value is
>
> $$\text{p-value} = 2\Phi(-|Z|) = 2\Phi(-4) < 2\Phi(-2.57) = 0.01$$
>
> so we reject the null hypothesis at a significance level of 0.01.

(c) Using your data, construct a confidence interval for the mean with confidence level 0.99.

> **Solution:**
>
> We use a confidence interval for the mean when the variance is known since we have more than 30 samples, and set $\sigma^2 = V_n$. We know that the confidence interval in this setting is $[M_n \pm \epsilon]$ with $\epsilon = \frac{\sqrt{V_n}}{\sqrt{n}} Q^{-1}(\alpha/2)$. Here, we have $1 - \alpha = 0.99$ so $\alpha = 0.005$. Plugging in, we have $\epsilon = \frac{\sqrt{16}}{\sqrt{64}} Q^{-1}(0.005) = \frac{1}{2} \cdot 2.57 = 1.285$ and confidence interval $\big[98 \pm 1.285\big]$ with confidence level 0.99.

(d) If you buy 32 more bags and find these additional bags have 104 candies on average, what would be your updated sample average over all $64 + 32 = 96$ purchased bags?

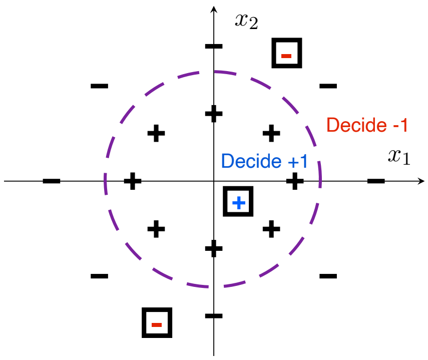
> **Solution:** $M_{96} = \frac{1}{96}(64 \cdot 98 + 32 \cdot 104) = \frac{2}{3} \cdot 98 + \frac{1}{3} \cdot 104 = 98 + \frac{6}{3} = 100$

(e) Let $V_1, \dots, V_9$ be i.i.d. Exponential(1/2) random variables. Using the Central Limit Theorem approximation, estimate the probability that $\sum_{i=1}^{9} V_i$ is less than 30. You may leave your answer in terms of the Φ function.

> **Solution:**
>
> Let $W = \sum_{i=1}^{9} V_i$ and note that $\mathbb{E}[W] = 9 \cdot \mathbb{E}[V] = 9 \cdot 2 = 18$ and $\mathsf{Var}[W] = 9 \cdot \mathsf{Var}[V] = 9 \cdot 4 = 36$.
>
> $$\mathbb{P}[W \leq 30] = F_W(30) \approx \Phi\left(\frac{30 - 18}{\sqrt{36}}\right) = \Phi(2)$$

Problem 4 (Machine Learning) *20 points*

You are given the 16 training data points on the figure, denoted by $+$ and $-$ symbols, along with 3 test data points, denoted by squares (for which you have no labels). The axes intentionally have no tick marks, but you can assume that the scale of the x_1 and x_2 axes are the same.



(a) The *closest average classifier* has a high training error rate. Explain why this is the case.

> **Solution:**
>
> The sample mean vector is the same for both labels, $\hat{\underline{\mu}}_+ = \hat{\underline{\mu}}_- = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Therefore, the closest average classifier will decide that every input should be mapped to $+1$.

(b) Is there a *linear classifier* of the form $D_{\text{linear}}(x_1, x_2) = \begin{cases} +1 & a_1x_1 + a_2x_2 \geq 0 \\ -1 & a_1x_1 + a_2x_2 < 0 \end{cases}$ that can attain training error rate less than $1/4$? Justify your answer.

> **Solution:**
>
> No. Any line that passes through the origin will have approximately half of the $+1$ data points and half of the -1 data points on the same side of the line. The training error rate for a linear classifier will always be about $1/2$.

(c) Consider a *non-linear classifier* of the form $D_{\text{non-linear}}(x_1, x_2) = \begin{cases} +1 & g(x_1, x_2) \leq \beta \\ -1 & g(x_1, x_2) > \beta \end{cases}$ where the function $g(x_1, x_2)$ depends only on x_1, x_2 while the threshold β can depend on the training data. Select a good $g(x_1, x_2)$ for this scenario. Justify your answer.

> **Solution:**
>
> Since the training data is circular, the function $g(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$ is a good choice, since we would then threshold based on the radius. (Also could use $x_1^2 + x_2^2$.)

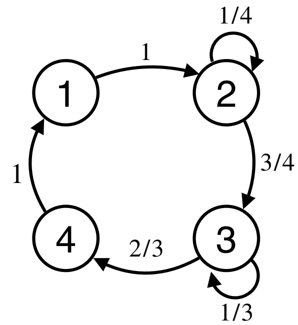
(d) Write a detailed procedure for selecting the threshold β from part (c) using the training data. Your answer should be written explicitly in terms of $(X_{\text{train},i,1}, X_{\text{train},i,2}, Y_{\text{train},i})$ for $i = 1, \ldots, n_{\text{train}}$ (but it does not need to be MATLAB/Python code or pseudo-code).

> **Solution:**
>
> We can set the threshold by first estimating the average radius under each label, and then setting the threshold to the halfway point. Let $n_{\text{train},+}$ be the number of $+1$ training points and let $n_{\text{train},-}$ be the number of -1 training points.
>
> $$L_+ = \big\{i \in \{1, \ldots, n_{\text{train}}\} : Y_{\text{train},i} = +1\big\} \qquad L_- = \big\{i \in \{1, \ldots, n_{\text{train}}\} : Y_{\text{train},i} = -1\big\}$$
> $$\hat{r}_+ = \frac{1}{n_{\text{train},+}} \sum_{i \in L_+} \sqrt{X_{\text{train},i,1}^2 + X_{\text{train},i,2}^2} \qquad \hat{r}_- = \frac{1}{n_{\text{train},-}} \sum_{i \in L_-} \sqrt{X_{\text{train},i,1}^2 + X_{\text{train},i,2}^2}$$
> $$\beta = \frac{\hat{r}_+ + \hat{r}_-}{2}$$

(e) Sketch the decision boundary for your trained non-linear classifier from parts (c) and (d) on the figure above. In each test data square, write the label selected by your classifier.

Problem 5 (Markov Chains) *20 points*

Consider the following discrete-time Markov chain with $\mathbb{P}[X_0 = 3] = \mathbb{P}[X_0 = 4] = 1/2$.



(a) List the communicating classes. For each communicating class, determine the period and whether it is transient or recurrent.

Solution:

$C_1 = \{1, 2, 3, 4\}$ which has period 1 and is recurrent.

(b) Write down the transition matrix.

Solution:

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1/4 & 3/4 & 0 \\ 0 & 0 & 1/3 & 2/3 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

(c) Does a unique limiting state probability vector $\underline{\pi}$ exist? If so, argue why and solve for it. If not, argue why.

Solution:

This is a recurrent, aperiodic Markov chain. Thus, it has a unique limiting state probability vector $\underline{\pi}$. From the steady-state equation $\mathbf{P}^T \underline{\pi} = \underline{\pi}$,

$$\pi_1 = \pi_4; \qquad \pi_1 + \frac{\pi_2}{4} = \pi_2 \implies \pi_1 = \frac{3}{4}\pi_2 \implies \pi_2 = \frac{4}{3}\pi_4; \qquad \frac{2}{3}\pi_3 = \pi_4 \implies \pi_3 = \frac{3}{2}\pi_4$$

From normalization, $\pi_1 + \pi_2 + \pi_3 + \pi_4 = \pi_4 + \frac{4}{3}\pi_4 + \frac{3}{2}\pi_4 + \pi_4 = \frac{29}{6}\pi_4 = 1$.

$$\implies \pi_1 = \frac{6}{29}, \ \pi_2 = \frac{8}{29}, \ \pi_3 = \frac{9}{29}, \ \pi_4 = \frac{6}{29}$$

(d) Determine $\mathbb{P}[X_0 = 3, X_1 = 4, X_2 = 1]$.

Solution: $\mathbb{P}[X_0 = 3, X_1 = 4, X_2 = 1] = \mathbb{P}[X_0 = 3]\, P_{34}\, P_{41} = \frac{1}{2} \cdot \frac{2}{3} \cdot 1 = \frac{1}{3}$

(e) Determine $\mathbb{P}[X_1 = 1 | X_3 = 2]$.

Solution:

$$\mathbb{P}[X_3 = 2 | X_1 = 1] = P_{12}\, P_{23} = 1 \cdot \frac{1}{4} = \frac{1}{4}$$

$$\mathbb{P}[X_1 = 1] = \mathbb{P}[X_0 = 4]\, P_{41} = \frac{1}{2} \cdot 1 = \frac{1}{2}$$

$$\begin{aligned} \mathbb{P}[X_3 = 2] &= \mathbb{P}[X_0 = 3]\, P_{34}\, P_{41}\, P_{12} + \mathbb{P}[X_0 = 4]\, P_{41}\, P_{12}\, P_{22} \\ &= \frac{1}{2} \cdot \frac{2}{3} \cdot 1 \cdot 1 + \frac{1}{2} \cdot 1 \cdot 1 \cdot \frac{1}{4} = \frac{1}{3} + \frac{1}{8} = \frac{11}{24} \end{aligned}$$

$$\mathbb{P}[X_1 = 1 | X_3 = 2] = \frac{\mathbb{P}[X_3 = 2 | X_1 = 1]\, \mathbb{P}[X_1 = 1]}{\mathbb{P}[X_3 = 2]} = \frac{\frac{1}{4} \cdot \frac{1}{2}}{\frac{11}{24}} = \frac{3}{11}$$