Problem 1

For each of the following parts, indicate whether the statement is always true or it can be false by clearly writing "True" or "False." Briefly explain the reasoning behind your answer for partial credit (in case your choice is wrong). Diagrams are welcome.

(a) Let Y be a discrete random variable, with probability mass function generated according to two possible hypotheses H_0 or H_1, as $P_{Y|H_1}(Y|H_1)$ or $P_{Y|H_0}(Y|H_0)$. Assume that the prior probability of H_1, $P(H_1) = 2/3$. Then, the probability of false alarm for the maximum likelihood detector (ML) is higher than the probability of false alarm for the maximum a posteriori detector (MAP).

Solution:

False. The threshold for MAP is lower, so the probability of false alarm is greater for MAP.

(b) Let X and Y be jointly continuous random variables. If $\widehat{x}(Y)$ is an estimator of X, then $\mathbb{E}[(\widehat{x}(Y) - X)^2] \geq \mathbb{E}[(\widehat{x}_{\mathrm{MMSE}}(Y) - X)^2]$.

Solution:

True, the MMSE solution attains the minimum mean-squared error and is unbiased.

(c) If $\rho_{X,Y} = 0$, then $\mathbb{E}[(\widehat{x}_{\mathrm{MMSE}}(Y) - X)^2] = \mathrm{Var}[X]$.

Solution:

False. This only means the LLSE solution is to guess the mean, not necessarily the MMSE solution.

(d) Let $X_1, \ldots, X_n$ be i.i.d. random variables with mean $\mathbb{E}[X_i] = \mu$. Then, the value of sample mean $M_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ always gets closer to the true mean as we include additional samples: for all $k < n$, $|M_n - \mu| < |M_k - \mu|$.

Solution:

False. Only true in expectation.

(e) Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be two independent datasets of different sizes, n_1 and n_2, collected under the same conditions with sample means $M_{n_1}^{(1)}$ and $M_{n_2}^{(2)}$, respectively. Under the null model, we assume the data is i.i.d. Gaussian with mean μ and variance 1. We calculate the Z-statistic for each dataset, $Z^{(1)}$ and $Z^{(2)}$, as well as the p-values. If the p-value for the first dataset is smaller, then its sample mean must be closer to the true mean, $|M_{n_1}^{(1)} - \mu| < |M_{n_2}^{(2)} - \mu|$.

Solution:

False, Z-statistic depends on size as well as sample mean.

(f) $\mathrm{Var}(X)\mathrm{Var}(Y) \geq \big(\mathrm{Cov}(X, Y)\big)^2$.

Solution:

True. The correlation coefficient $\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$ satisfies $\rho_{X,Y} \leq 1$. Therefore, by squaring both sides and using this inequality, we obtain the desired inequality.

(g) Consider a binary classification dataset $(\underline{X}_1, Y_1), \ldots, (\underline{X}_n, Y_n)$ and two classifiers $D_{\mathrm{A}}(\underline{x})$ and $D_{\mathrm{B}}(\underline{x})$. If classifier A has a lower error rate than classifier B on the first half of the dataset $(\underline{X}_1, Y_1), \ldots, (\underline{X}_{n/2}, Y_{n/2})$, then it also will have a lower error rate on the second half of the dataset $(\underline{X}_{n/2+1}, Y_{n/2+1}), \ldots, (\underline{X}_n, Y_n)$.

Solution:

False, these quantities are not related.

(h) In a Markov chain with transition probability matrix $\mathbf{P}$, if $P_{jj} > 0$, then the period of state j is 1.

Solution:

True. A self-cycle will make the period 1.

(i) Suppose we have a Markov chain with two communicating classes. Then, there must be multiple limiting probability distributions for the state probability vector $\underline{p}_t$ as $t \to \infty$.

Solution:

False. One of the communicating classes could be transient, leading to a unique limiting probability state vector.

Problem 2

On the first day of your new job as a data scientist, you are asked to consider the following binary hypothesis testing scenario. The range of the observation Y consists of the integers $R_Y = \{\ldots, -3, -2, -1, 0, +1, +2, +3, \ldots\}$. The likelihoods are

$$P_{Y|H_0}(y) = \frac{1}{2}\left(\frac{1}{3}\right)^{|y|} \qquad P_{Y|H_1}(y) = \frac{1}{5}\left(\frac{2}{3}\right)^{|y|}$$

The hypothesis probabilities are $\mathbb{P}[H_0] = 1/3$ and $\mathbb{P}[H_1] = 2/3$.

(a) Determine the ML rule.

Solution:

We know that, at $y = 0$, $P_{Y|H_0}(0) = 1/2 > P_{Y|H_1}(0) = 1/5$.

The likelihood ratio as a function of y is

$$L(y) = \frac{\frac{1}{5}\left(\frac{2}{3}\right)^{|y|}}{\frac{1}{2}\left(\frac{1}{3}\right)^{|y|}} = \frac{2}{5}(2)^{|y|}$$

This is bigger than the ML threshold 1 for all $|y| \geq 2$. Therefore, the ML rule is

$$D^{\text{ML}}(y) = \begin{cases} 1 & P_{Y|H_1}(y) \geq P_{Y|H_0}(y) \\ 0 & P_{Y|H_1}(y) < P_{Y|H_0}(y) \end{cases} = \begin{cases} 1 & |y| \geq 2 \\ 0 & |y| \leq 1 \end{cases}$$

(b) Determine the MAP rule.

Solution:

The MAP threshold for the likelihood is $\frac{1/3}{2/3} = 0.5$. Using the likelihood ratio from above, we see $\mathcal{L}(1) = 4/5 > 0.5$. Hence, the MAP rule is

$$D^{\text{MAP}}(y) = \begin{cases} 1 & \mathbb{P}[H_1]P_{Y|H_1}(y) \geq \mathbb{P}[H_0]P_{Y|H_0}(y) \\ 0 & \mathbb{P}[H_1]P_{Y|H_1}(y) < \mathbb{P}[H_0]P_{Y|H_0}(y) \end{cases} = \begin{cases} 1 & |y| \geq 1 \\ 0 & y = 0 \end{cases}$$

(c) Determine the probability of error under the ML rule.

Solution:

$$\begin{aligned}
P_{\text{FA}} &= \mathbb{P}[D^{\text{ML}}(Y) = 1|H_0] = 1 - P_{Y|H_0}(0) - P_{Y|H_0}(1) - P_{Y|H_0}(-1) = 1 - 1/2 - 1/3 = 1/6 \\
P_{\text{MD}} &= \mathbb{P}[D^{\text{ML}}(Y) = 0|H_1] = P_{Y|H_1}(0) + P_{Y|H_1}(1) + P_{Y|H_1}(-1) = 1/5 + 4/15 = 7/15 \\
\mathbb{P}[\text{error}_{\text{ML}}] &= P_{\text{FA}}\mathbb{P}[H_0] + P_{\text{MD}}\mathbb{P}[H_1] = \frac{1}{6} \cdot \frac{1}{3} + \frac{7}{15} \cdot \frac{2}{3} = \frac{11}{30}
\end{aligned}$$

(d) Determine the probability of error under the MAP rule.

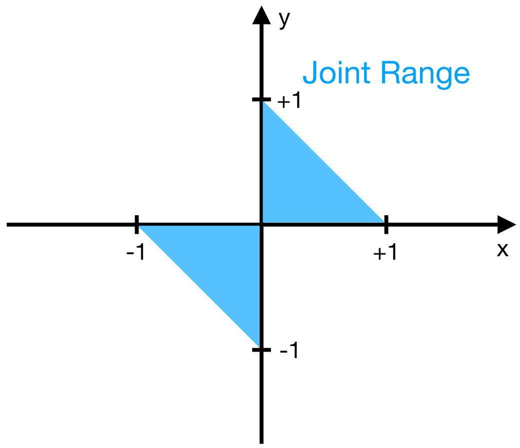
Solution:

$$P_{\text{FA}} = \mathbb{P}[D^{\text{MAP}}(Y) = 1|H_0] = 1 - P_{Y|H_0}(0) = 1/2$$
$$P_{\text{MD}} = \mathbb{P}[D^{\text{MAP}}(Y) = 0|H_1] = P_{Y|H_1}(0) = 1/5$$
$$\mathbb{P}[\text{error}_{\text{MAP}}] = P_{\text{FA}}\mathbb{P}[H_0] + P_{\text{MD}}\mathbb{P}[H_1] = \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{5} \cdot \frac{2}{3} = \frac{3}{10}$$

Problem 3

On your second day at your new data scientist job, you consider the following estimation problem. The joint PDF of jointly continuous random variables X and Y is

$$f_{X,Y}(x,y) = \begin{cases} 1 & x \geq 0,\ y \geq 0,\ x+y \leq 1 \quad \text{OR} \quad x \leq 0,\ y \leq 0,\ x+y \geq -1 \\ 0 & \text{otherwise.} \end{cases}$$



(a) Determine $\mathbb{P}[X \geq Y]$. Your answer can be an integral, but you can also take advantage of the simple structure of the joint PDF.

Solution:

If you draw the line $X = Y$, you see it divides the range so that 1/2 of the area is above the line $Y = X$, and 1/2 of the area is below the line, so $\mathbb{P}[X - Y \geq 0] = 1/2$.

(b) What is the MMSE estimator of X given $Y = y$? Your answer can be an integral, but you can also take advantage of the simple structure of the joint PDF.

Solution:

For any value of y, the density $f_{X|Y}(x|y)$ is uniform on $[0, 1-y]$ if $y > 0$, and uniform on $[-1-y, 0]$ for $y < 0$. For $y = 0$, it can be uniform on [-1,1]. Hence, the estimate is

$$\widehat{x}_{\text{MMSE}}(y) = \begin{cases} \frac{1-y}{2} & y > 0 \\ \frac{-y-1}{2} & y < 0 \\ 0 & y = 0 \end{cases}$$

(c) Determine $\mathbb{E}[X]$. Your answer can be an integral, but you can also take advantage of the simple structure of the joint PDF.

Solution:

By symmetry, $\mathbb{E}[Y] = \mathbb{E}[X] = 0$.

(d) What is the LLSE estimator of X given $Y = y$? Your answer can be an integral.

Solution:

First, let's compute $\mathrm{Var}[X] = \mathrm{Var}[Y]$, by symmetry. Since the means are 0,

$$f_X(x) = \begin{cases} \int_0^{1-x} dy = 1 - x & x > 0 \\ \int_{-1-x}^{0} dy = 1 + x & x \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

Then, $\mathbb{E}[X^2] = \int_{-1}^{1} (x^2)(1 - |x|)\, dx = 2 \int_0^1 x^2(1 - x)\, dx = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$.

$$\mathrm{Cov}[X, Y] = 2 \int_0^1 \int_0^{1-x} xy\, dy\, dx = \int_0^1 x(1 - x)^2\, dx = \frac{1}{2} - \frac{2}{3} + \frac{1}{4} = \frac{1}{12}.$$

The LLSE estimate is now:

$$\widehat{x}_{LLSE}(y) = \frac{\mathrm{Cov}[X, Y]}{\mathrm{Var}[Y]} y = \frac{1}{12} y$$

Problem 4

During your coffee break, you are trying to bound the probability that the total number of successes exceeds a threshold. Let $X_1, \dots, X_{100}$ be i.i.d. Bernoulli $(1/2)$ random variables.

(a) What is the mean of $Y = X_1 + X_2$?

Solution:

$$\mathbb{E}[Y] = 2 \cdot \frac{1}{2} = 1$$

(b) What is the PMF of $Y = X_1 + X_2$?

Solution:

This is a Binomial(2,1/2),

$$P_Y(y) = \begin{cases} 1/4 & y = 0 \\ 1/2 & y = 1 \\ 1/4 & y = 2 \\ 0 & \text{otherwise.} \end{cases}$$

(c) Use the Central Limit Theorem approximation to estimate the probability $\mathbb{P}\Big[\sum_{i=1}^{100} X_i \geq 60\Big]$. You can leave your answer in terms of the standard normal CDF $\Phi(z)$.

Solution:

We treat S_{100} as a Gaussian random variable with known mean 50 and standard deviation $\sqrt{25} = 5$. Then,

$$\mathbb{P}\Big[\sum_{i=1}^{100} X_i \geq 60\Big] \approx 1 - \Phi(\frac{60 - 50}{5}) = 1 - \Phi(2)$$

Problem 5

You measure the sulfate concentration in the local water reservoir over 9 consecutive days and obtain values $X_1, \ldots, X_9$, which are assumed to be i.i.d. Gaussian. (The units are mg/L and omitted below.) The sample mean is $M_9 = 5.1$ and the sample variance is $V_9 = 0.36$.

Let W have a t-distribution with 8 degrees-of-freedom. You may find one or more of the following values useful

- $F_W(-1.4) = \Phi(-1.3) = Q(1.3) = 0.1\,, \quad F_W(1.4) = \Phi(1.3) = 0.9$
- $F_W(-1.9) = \Phi(-1.6) = Q(1.6) = 0.05\,, \quad F_W(1.9) = \Phi(1.6) = 0.95$
- $F_W(-2.3) = \Phi(-2.0) = Q(2.0) = 0.025\,, \quad F_W(2.3) = \Phi(2.0) = 0.975$

(a) Construct a confidence interval for the mean with confidence level 0.95.

Solution:

Given the small number of samples and the Gaussian assumption, if the true mean is μ, we know that $W = \frac{\sqrt{9}(M_9-\mu)}{\sqrt{V_9}}$ follows a Student's t-distribution with 8 degrees of freedom. From the above numbers, that means

$$\mathbb{P}[|W| \leq 2.3] = F_W(2.3) - F_W(-2.3) = 0.95.$$

So,

$$\mathbb{P}[\frac{\sqrt{9}(M_9 - \mu)}{\sqrt{V_9}}| \leq 2.3] = \mathbb{P}[|M_9 - \mu| \leq 2.3\frac{\sqrt{V_9}}{3}] = \mathbb{P}[|M_9 - \mu| \leq \frac{2.3 \cdot 0.6}{3}] = 0.95$$

The interval of interest is $\mu \in [5.1 - 0.46, 5.1 + 0.46] = [4.64, 5.56]$.

(b) Is your sample significantly different from the baseline concentration $\mu = 5.4$ at a significance level of 0.1? Justify your approach and support your answer numerically.

Solution:

Given the small number of samples and the unknown true variance, the proper test is a T-test. Under the given assumptions, if the null hypothesis is $\mu = 5.4$, then $W = \frac{\sqrt{9}(M_9-\mu)}{\sqrt{V_9}}$ follows a Student's t-distribution with 8 degrees of freedom. Given the data, we have $W = \frac{3(5.1-5.4)}{0.6} = -1.5$.

To reject the null hypothesis with level of significance 0.1, we want $F_W(-1.5) < 0.05$. Since $F_W(-1,9) = 0.05$, we fail to reject the null hypothesis at level of significance 0.1.

(c) Say you also go out on the 10^{th} day and collect measurement $X_{10} = 5$. What is the new sample mean M_{10}?

Solution:

$$M_{10} = \frac{1}{10}(9 * M_9 + X_{10}) = \frac{1}{10}(50.9 + 5) = 5.09.$$

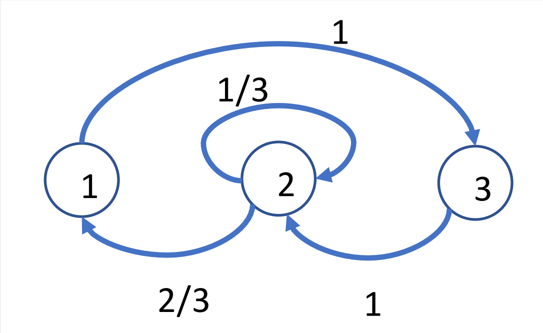
Problem 6

You have managed to model an interesting three-state system via a Markov chain with the following state transition matrix and initial probability state vector:

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 1 \\ 2/3 & 1/3 & 0 \\ 0 & 1 & 0 \end{bmatrix} \qquad \underline{p}_0 = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

(a) Draw the Markov chain, labeling the states as 1, 2, and 3, as well as labeling the arcs with the appropriate transition probabilities.

Solution:



(b) What is the period of state 1?

Solution:

There is a cycle of length 3 (1-3-2-1) and a cycle of length 4 (1-3-2-2-1). The greatest common divisor of those two lengths is 1, so the period of state 1 is 1.

(c) Determine $\mathbb{P}[X_0 = 2, X_1 = 2, X_2 = 3]$.

Solution:

$\mathbb{P}[X_0 = 1, X_1 = 2, X_2 = 3] = \mathbb{P}[X_2 = 3|X_1 = 2]\mathbb{P}[X_1 = 2|X_0 = 1]\mathbb{P}[X_0 = 1] = 0 \cdot \frac{1}{3} \cdot \frac{1}{3} = 0$

(d) 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 $\mathbf{P}^T\underline{\pi} = \underline{\pi}$,

$$\pi_1 = \frac{2}{3}\pi_2;$$
$$\pi_3 = \pi_1$$

From normalization,

$$\pi_1 + \pi_2 + \pi_3 = \pi_1 + \frac{3}{2}\pi_1 + \pi_1 = \frac{7}{2}\pi_1 = 1$$

so $\pi_1 = \pi_3 = 2/7, \pi_2 = 3/7$.