EK381: Exam 3 Review

College of Engineering
Boston University

Binary Hypothesis Testing:

- Two hypotheses H_0 and H_1. Observe a random variable Y. Decide if H_0 or H_1 occurred based only on Y using a decision rule $D(Y)$.

Discrete Case	Continuous Case
$P_{Y\|H_0}(y)$ if H_0 occurs	$f_{Y\|H_0}(y)$ if H_0 occurs
$P_{Y\|H_1}(y)$ if H_1 occurs	$f_{Y\|H_1}(y)$ if H_1 occurs

- Decision Regions:
  $A_0 = \{y \in R_Y : D(y) = 0\}$ $\quad$ $A_1 = \{y \in R_Y : D(y) = 1\}$
- Probability of False Alarm: $P_{\mathsf{FA}} = \mathbb{P}[Y \in A_1 | H_0]$
- Probability of Missed Detection: $P_{\mathsf{MD}} = \mathbb{P}[Y \in A_0 | H_1]$
- Goal is to minimize the probability of error:

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

- Likelihood Ratio (and Log-Likelihood Ratio):

$$L(y) = \frac{P_{Y|H_1}(y)}{P_{Y|H_0}(y)} \qquad \ln\big(L(y)\big) = \ln\left(\frac{P_{Y|H_1}(y)}{P_{Y|H_0}(y)}\right)$$

Maximum Likelihood (ML) Rule:

- Intuition: Choose the hypothesis that best explains the observation.
- In terms of the conditional PMFs for the discrete case,

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

- In terms of the conditional PDFs for the continuous case,

$$D^{\mathsf{ML}}(y) = \begin{cases} 1, & f_{Y|H_1}(y) \geq f_{Y|H_0}(y), \\ 0, & f_{Y|H_1}(y) < f_{Y|H_0}(y). \end{cases}$$

- In terms of the likelihood or log-likelihood ratio,

Likelihood Ratio

$$D^{\mathsf{ML}}(y) = \begin{cases} 1, & L(y) \geq 1, \\ 0, & L(y) < 1. \end{cases}$$

Log-Likelihood Ratio

$$D^{\mathsf{ML}}(y) = \begin{cases} 1, & \ln(L(y)) \geq 0, \\ 0, & \ln(L(y)) < 0. \end{cases}$$

Maximum a Posteriori (MAP) Rule:

- Intuition: Choose the most likely hypothesis given the observation.
- Attains the minimum probability of error.
- In terms of the conditional PMFs for the discrete case,

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

- In terms of the conditional PDFs for the continuous case,

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

- In terms of the likelihood or log-likelihood ratio,

<u>Likelihood Ratio</u>

$$D^{\mathsf{MAP}}(y) = \begin{cases} 1, & L(y) \geq \frac{\mathbb{P}[H_0]}{\mathbb{P}[H_1]}, \\ 0, & L(y) < \frac{\mathbb{P}[H_0]}{\mathbb{P}[H_1]}. \end{cases}$$

<u>Log-Likelihood Ratio</u>

$$D^{\mathsf{MAP}}(y) = \begin{cases} 1, & \ln(L(y)) \geq \ln\left(\frac{\mathbb{P}[H_0]}{\mathbb{P}[H_1]}\right), \\ 0, & \ln(L(y)) < \ln\left(\frac{\mathbb{P}[H_0]}{\mathbb{P}[H_1]}\right). \end{cases}$$

Vector Observations:

- Two hypotheses H_0 and H_1. Observe a random vector $\underline{Y}$. Decide which hypothesis occurred using a decision rule $D(\underline{Y})$.

Discrete Case	Continuous Case
$P_{\underline{Y}\|H_0}(\underline{y})$ if H_0 occurs	$f_{\underline{Y}\|H_0}(\underline{y})$ if H_0 occurs
$P_{\underline{Y}\|H_1}(\underline{y})$ if H_1 occurs	$f_{\underline{Y}\|H_1}(\underline{y})$ if H_1 occurs

- Decision Regions:

$$A_0 = \{\underline{y} \in R_Y : D(\underline{y}) = 0\} \qquad A_1 = \{\underline{y} \in R_Y : D(\underline{y}) = 1\}$$

- Probability of False Alarm: $P_{\mathsf{FA}} = \mathbb{P}[\underline{Y} \in A_1 | H_0]$
- Probability of Missed Detection: $P_{\mathsf{MD}} = \mathbb{P}[\underline{Y} \in A_0 | H_1]$
- Probability of Error: $P_e = P_{\mathsf{FA}}\,\mathbb{P}[H_0] + P_{\mathsf{MD}}\,\mathbb{P}[H_1]$
- The likelihood ratio as well as the ML and MAP rules are the same as on the previous slides, just substitute in $\underline{Y}$ for Y and $\underline{y}$ for y in each equation.

Scalar Estimation:

- We observe a random variable Y and want to estimate an unobserved random variable X using an estimator $\hat{x}(Y)$.
- Goal: Minimize the mean-squared error: $\mathsf{MSE} = \mathbb{E}\Big[\big(X - \hat{x}(Y)\big)^2\Big]$

Minimum Mean-Squared Error (MMSE) Estimator:

- Attains the minimum MSE amongst all possible estimators.
- Given by the conditional expectation: $\hat{x}_{\mathsf{MMSE}}(y) = \mathbb{E}[X|Y = y]$

Linear Least-Squares Error (LLSE) Estimator:

- Attains the minimum MSE amongst all linear estimators.
- $\hat{x}_{\mathsf{LLSE}}(y) = \mathbb{E}[X] + \dfrac{\mathsf{Cov}[X,Y]}{\mathsf{Var}[Y]}\big(y - \mathbb{E}[Y]\big) = \mathbb{E}[X] + \rho_{X,Y}\dfrac{\sigma_X}{\sigma_Y}\big(y - \mathbb{E}[Y]\big)$
- $\mathsf{MSE}_{\mathsf{LLSE}} = \mathsf{Var}[X] - \dfrac{\big(\mathsf{Cov}[X,Y]\big)^2}{\mathsf{Var}[Y]} = \mathsf{Var}[X](1 - \rho^2_{X,Y})$
- For jointly Gaussian X and Y, $\hat{x}_{\mathsf{MMSE}}(y) = \hat{x}_{\mathsf{LLSE}}(y)$.

Vector Estimation:

- We observe a random vector $\underline{Y}$ and want to estimate an unobserved random vector $\underline{X}$ using an estimator $\underline{\hat{x}}(\underline{Y})$.
- Mean-Squared Error: $\text{MSE} = \mathbb{E}\Big[\big(\underline{X} - \underline{\hat{x}}(\underline{Y})\big)^{\mathsf{T}}\big(\underline{X} - \underline{\hat{x}}(\underline{Y})\big)\Big]$

Vector Minimum Mean-Squared Error (MMSE) Estimator:

- Attains the minimum MSE amongst all possible estimators.
- Given by the conditional expectation: $\underline{\hat{x}}_{\text{MMSE}}(\underline{y}) = \mathbb{E}[\underline{X} | \underline{Y} = \underline{y}]$

Vector Linear Least-Squares Error (LLSE) Estimator:

- Attains the minimum MSE amongst all linear estimators.
- $\underline{\hat{x}}_{\text{LLSE}}(\underline{y}) = \mathbb{E}[\underline{X}] + \boldsymbol{\Sigma}_{\underline{X},\underline{Y}}\boldsymbol{\Sigma}_{\underline{Y}}^{-1}\big(\underline{y} - \mathbb{E}[\underline{Y}]\big)$
- Covariance Matrix of $\underline{Y}$: $\boldsymbol{\Sigma}_{\underline{Y}} = \mathbb{E}\big[(\underline{Y} - \mathbb{E}[\underline{Y}])(\underline{Y} - \mathbb{E}[\underline{Y}])^{\mathsf{T}}\big]$
- Cross-Covariance Matrix: $\boldsymbol{\Sigma}_{\underline{X},\underline{Y}} = \mathbb{E}\big[(\underline{X} - \mathbb{E}[\underline{X}])(\underline{Y} - \mathbb{E}[\underline{Y}])^{\mathsf{T}}\big]$
- For jointly Gaussian $\underline{X}$ and $\underline{Y}$, $\underline{\hat{x}}_{\text{MMSE}}(\underline{y}) = \underline{\hat{x}}_{\text{LLSE}}(\underline{y})$.

Sums of Random Variables

- Joint PMF for n discrete random variables $P_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$
- Joint PDF for n continuous random variables $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$
- Expected Value of a Sum: $\mathbb{E}\big[X_1+\cdots+X_n\big]=\sum_{i=1}^{n}\mathbb{E}[X_i]$
- Variance of a Sum: $\text{Var}\big[X_1+\cdots+X_n\big]=\sum_{i=1}^{n}\sum_{j=1}^{n}\text{Cov}[X_i,X_j]$

Independent and Identically Distributed Random Variables:

- Joint PMF $P_{X_1,\ldots,X_n}(x_1,\ldots,x_n)=P_X(x_1)\cdots P_X(x_n)$
- Joint PDF $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)=f_X(x_1)\cdots f_X(x_n)$
- Expected Value of a Sum: $\mathbb{E}\big[X_1+\cdots+X_n\big]=n\mathbb{E}[X]$
- Variance of a Sum: $\text{Var}\big[X_1+\cdots+X_n\big]=n\text{Var}[X]$

- **Weak Law of Large Numbers:** Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with finite mean μ. For any constant $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| > \epsilon \right) = 0$$

- **Strong Law of Large Numbers:** Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with finite mean μ. Then,

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu \right) = 1$$

- **Central Limit Theorem:** Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with finite mean μ and finite variance σ^2. The CDF of $Y_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$ converges to the standard normal CDF,

$$\lim_{n\to\infty} F_{Y_n}(y) = \Phi(y) \ .$$

Parameter Estimation:

- We collect data $X_1, \ldots, X_n$, which we assume is i.i.d. with some unknown true mean μ and true variance σ^2. How can we estimate these parameters?
- Sample Mean: $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
- Sample Variance: $V_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M_n)^2$
- The sample mean and sample variance are unbiased estimators:

$$\mathbb{E}[M_n] = \mu \qquad\qquad \mathbb{E}[V_n] = \sigma^2$$

Confidence Intervals:

- $[M_n \pm \epsilon]$ is a confidence interval for the mean with confidence level $1 - \alpha$ if $\mathbb{P}[\mu - \epsilon \leq M_n \leq \mu + \epsilon] = 1 - \alpha$.

Confidence Interval: Known Variance

- Assumes data is i.i.d. Gaussian(μ, σ^2) with σ^2 known.
- $[M_n \pm \epsilon]$ with $\epsilon = \frac{\sigma}{\sqrt{n}} Q^{-1}(\frac{\alpha}{2})$ is a confidence interval for the mean with confidence level $1 - \alpha$.
- When to use: Variance is known *or* $n > 30$ samples.
- If the variance σ^2 is unknown *and* we have $n > 30$ samples, substitute σ^2 with the sample variance V_n.

Confidence Interval: Unknown Variance

- Assumes data is i.i.d. Gaussian(μ, σ^2) with σ^2 unknown.
- Let $F_{T_{n-1}}(t)$ be the CDF for a Student's t-distribution with $n - 1$ degrees-of-freedom.
- $[M_n \pm \epsilon]$ with $\epsilon = -\frac{\sqrt{V_n}}{\sqrt{n}} F_{T_{n-1}}^{-1}(\frac{\alpha}{2})$ is a confidence interval for the mean with confidence level $1 - \alpha$.
- When to use: Variance is unknown *and* $n \leq 30$ samples.

Significance Testing

- Only have a probability model for the null hypothesis H_0.
- The significance level $0 \leq \alpha \leq 1$ is used to determine when to reject the null hypothesis.
- Given a statistic calculated from the dataset, the p-value is the probability of observing a value at least this extreme under the null hypothesis.
  - If $\text{p-value} < \alpha$, then reject the null hypothesis.
  - If $\text{p-value} \geq \alpha$, then fail to reject the null hypothesis.

One-Sample Z-Test

- Null Hypothesis: $X_1, \ldots, X_n$ is i.i.d. Gaussian(μ, σ^2).
- When to use: Variance σ^2 is known *or* $n > 30$ samples.
- Informally, is the mean not equal to μ?
  1. Calculate the sample mean M_n.
  2. Z-statistic: $Z = \sqrt{n}(M_n - \mu)/\sigma$.
  3. $\text{p-value} = 2\Phi(-|Z|)$.
- If the variance σ^2 is unknown *and* we have $n > 30$ samples, substitute σ^2 with the sample variance V_n.

One-Sample T-Test

- Null Hypothesis: $X_1, \ldots, X_n$ is i.i.d. Gaussian(μ, σ^2).
- When to use: Variance σ^2 is unknown *and* $n \leq 30$ samples.
- Informally, is the mean not equal to μ?
  1. Calculate the sample mean M_n and variance V_n.
  2. Z-statistic: $T = \sqrt{n}(M_n - \mu)/\sqrt{V_n}$.
  3. $\text{p-value} = 2F_{T_{n-1}}(-|T|)$.

Two-Sample Z-Test

- Null Hypothesis: $X_1, \ldots, X_{n_1}$ is i.i.d. Gaussian(μ, σ_1^2) and $Y_1, \ldots, Y_{n_2}$ is i.i.d. Gaussian(μ, σ_2^2).
- When to use: Variances σ_1^2 and σ_2^2 are known *or* $\min(n_1, n_2) > 30$.
- Informally, do the datasets have the same mean?
  1. Calculate the sample means $M_{n_1}^{(1)}$ and $M_{n_2}^{(2)}$.
  2. Z-statistic: $Z = \left(M_{n_1}^{(1)} - M_{n_2}^{(2)}\right) \Big/ \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.
  3. $\text{p-value} = 2\Phi(-|Z|)$.
- If the variances σ_1^2, σ_2^2 are unknown *and* we have $\min(n_1, n_2) > 30$ samples, substitute $\sigma_1^2 = V_{n_1}^{(1)}$ and $\sigma_2^2 = V_{n_2}^{(2)}$.

Two-Sample T-Test

- Null Hypothesis: $X_1, \ldots, X_{n_1}$ is i.i.d. Gaussian(μ, σ^2) and $Y_1, \ldots, Y_{n_2}$ is i.i.d. Gaussian(μ, σ^2). The mean μ is unknown.
- When to use: (Equal) variance σ^2 is unknown *and* $\min(n_1, n_2) \leq 30$.
- Informally, do the datasets have the same mean?

1. Calculate the sample means $M_{n_1}^{(1)}, M_{n_2}^{(2)}$, sample variances $V_{n_1}^{(1)}, V_{n_2}^{(2)}$, and the pooled sample variance

$$\hat{\sigma}^2 = \big((n_1 - 1)V_{n_1}^{(1)} + (n_2 - 1)V_{n_2}^{(2)}\big)/\big(n_1 + n_2 - 2\big)$$

2. T-statistic: $T = \dfrac{\big(M_{n_1}^{(1)} - M_{n_2}^{(2)}\big)}{\sqrt{\hat{\sigma}^2\Big(\dfrac{1}{n_1} + \dfrac{1}{n_2}\Big)}}$
3. $\text{p-value} = 2F_{T_{n_1+n_2-2}}(-|T|)$.

Binary Classification:

- The goal is to decide between two hypotheses, but we do not have access to the underlying probability model.
- Instead, we have a dataset consisting of n samples,

$$\big\{(\underline{X}_1, Y_1), (\underline{X}_2, Y_2), \ldots, (\underline{X}_n, Y_n)\big\}.$$

  The i^{th} sample $(\underline{X}_i, Y_i)$ has an observation vector $\underline{X}_i$ and a label Y_i, which we assume is -1 or $+1$.

- We use this dataset to come up with a classifier $D(\underline{x})$, which is a function that maps any possible observation vector $\underline{x}$ into a guess of its label.
- We measure performance via the error rate, the fraction of misclassified examples (usually reported as a percentage).

Training and Test Error:

- To make sure we are not overfitting, we split our dataset into non-overlapping training and test datasets,

$$\left\{(\underline{X}_{\text{train},1}, Y_{\text{train},1}), (\underline{X}_{\text{train},2}, Y_{\text{train},2}), \dots, (\underline{X}_{\text{train},n_{\text{train}}}, Y_{\text{train},n_{\text{train}}})\right\},$$
$$\left\{(\underline{X}_{\text{test},1}, Y_{\text{test},1}), (\underline{X}_{\text{test},2}, Y_{\text{test},2}), \dots, (\underline{X}_{\text{test},n_{\text{test}}}, Y_{\text{test},n_{\text{test}}})\right\}.$$

- The training set is used to construct our classifier $D(\underline{x})$ and the test set can only be used to evaluate its performance.

$$\text{Training Error} = \text{fraction of misclassified training examples},$$
$$\text{Test Error} = \text{fraction of misclassified test examples}$$

Basic Classifiers:

- The closest average classifier first computes the average vector for each label. Given a new observation vector, it computes the distance to each average and choose the label with the smallest distance.
- The nearest neighbor classifier, when given a new observation vector, computes the distance to every sample in the training set to find the closest point. It then outputs the label of this point as its guess.
- The LDA classifier assumes the observation vectors are Gaussian, with different mean vectors and the same covariance matrix. It estimates these parameters and then applies the resulting ML rule.
- The QDA classifier assumes the observation vectors are Gaussian, with different mean vectors and covariance matrices. It estimates these parameters and then applies the resulting ML rule.

Dimensionality Reduction:

- Principal component analysis allows us to reduce the dimensionality of our observations, by only keeping the (orthogonal) directions corresponding the largest variance.

Markov Chain:

- Sequence of (discrete) random variables $X_0, X_1, X_2, \ldots$ such that, given the history $X_0, \ldots, X_n$, the next state X_{n+1} only depends on the current state X_n,

$$P_{X_{n+1}|X_n,\ldots,X_0}(x_{n+1}|x_n,\ldots,x_0) = P_{X_{n+1}|X_n}(x_{n+1}|x_n)$$

- We assume the range is finite $R_X = \{1, \ldots, K\}$.
- The transition probabilities P_{jk} are the probabilities of moving from state j to state k in one time step. We assume the Markov chain is homogeneous, $P_{X_{n+1}|X_n}(k|j) = P_{jk}$.
- The n-step transition probabilities $P_{jk}(n)$ are the probabilities of moving from state j to state k in exactly n time steps. They can be determined via the Chapman-Kolmogorov equations:

$$P_{jk}(n+m) = \sum_{i=1}^{K} P_{ji}(n) P_{ik}(m)$$

State Transition Matrix:

- The state transition matrix is $\mathbf{P} = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1K} \\ P_{21} & P_{22} & \cdots & P_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ P_{K1} & P_{K2} & \cdots & P_{KK} \end{bmatrix}$
- Row index is for the current state, column index is for the next state.
- All rows must sum to 1.

Probability State Vector:

- The probability state vector is $\underline{p}_t = \begin{bmatrix} P_{X_t}(1) \\ \vdots \\ P_{X_t}(K) \end{bmatrix}$.
- Moving forward one time step: $\underline{p}_{t+1} = \mathbf{P}^{\mathsf{T}} \underline{p}_t$.
- Moving forward n time steps: $\underline{p}_{t+n} = (\mathbf{P}^n)^{\mathsf{T}} \underline{p}_t$.

State Classification:

- State k is accessible from state j if it is possible to reach state k starting from state j in zero or more time steps. (State j is always accessible from itself.) Notation: $j \to k$
- States j and k communicate if $j \to k$ and $k \to j$. Notation: $j \leftrightarrow k$
- A communicating class C is a subset of states such that if $j \in C$, then $k \in C$ if and only if $j \leftrightarrow k$.
- A Markov chain is irreducible if all of its states belong to a single communicating class.
- A state j is transient if there is a state k such that $j \to k$ but $k \nrightarrow j$.
- Any state that is not transient is recurrent.
- The period d of a state j is the greatest common divisor of the length of all cycles from j back to itself.
- If the period is 1, then the state is called aperiodic. The entire Markov chain is aperiodic if all states are aperiodic.

Limiting Probability State Vector:

- For an irreducible, aperiodic Markov chain, there is a unique limiting state probability vector $\underline{\pi} = \lim_{t \to \infty} \underline{p}_t$.

Properties of the Limiting Probability State Vector:

- Normalization: $\sum_{j=1}^{K} \pi_j = 1$
- Any initial probability state probability vector $\underline{p}_0$ will converge to $\underline{\pi}$.
- Steady-State Distribution: $\underline{\pi} = \mathbf{P}^{\mathsf{T}} \underline{\pi}$.

Handling Transient States:

- If there is only one recurrent communicating class and it is aperiodic, then there is still a unique limiting state probability vector. Find by first setting the limiting probabilities of all transient states to 0.

Practice Question #1

Consider the following detection problem.
Under hypothesis H_0, Y is a Geometric$(1/2)$ random variable.
Under hypothesis H_1, Y is a Geometric$(3/4)$ random variable.
The probabilities of the hypotheses are $\mathbb{P}[H_0] = 1/3$ and $\mathbb{P}[H_1] = 2/3$.

(a) Determine the ML rule. The conditional PMFs are

$$P_{Y|H_0}(y) = \begin{cases} \left(\frac{1}{2}\right)^y & y = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \quad P_{Y|H_1}(y) = \begin{cases} \frac{3}{4}\left(\frac{1}{4}\right)^{y-1} & y = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

From these, we can form the likelihood ratio,

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

which is greater than 1 for $y = 1$ and less than 1 for $y \geq 2$.
Therefore, the ML rule is to decide H_1 for $y = 1$ and decide H_0 for $y \geq 2$.

Practice Question #1

Consider the following detection problem.
Under hypothesis H_0, Y is a Geometric$(1/2)$ random variable.
Under hypothesis H_1, Y is a Geometric$(3/4)$ random variable.
The probabilities of the hypotheses are $\mathbb{P}[H_0] = 1/3$ and $\mathbb{P}[H_1] = 2/3$.

(b) Determine the probability of error under the ML rule. For the ML rule, we have $A_0 = \{Y \geq 2\}$ and $A_1 = \{Y = 1\}$. Therefore, the probability of error is

$$\begin{aligned}
P_{\mathsf{FA}} &= \mathbb{P}[Y \in A_1 | H_0] = \mathbb{P}[Y = 1 | H_0] = P_{Y|H_0}(1) = \frac{1}{2} \\
P_{\mathsf{MD}} &= \mathbb{P}[Y \in A_0 | H_1] = \mathbb{P}[Y \geq 2 | H_1] \\
&= 1 - \mathbb{P}[Y = 1 | H_1] = 1 - P_{Y|H_1}(1) = 1 - \frac{3}{4} = \frac{1}{4} \\
\mathbb{P}[\text{error}] &= P_{\mathsf{FA}}\,\mathbb{P}[H_0] + P_{\mathsf{MD}}\,\mathbb{P}[H_1] = \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{2}{3} = \frac{1}{3}\ .
\end{aligned}$$

Practice Question #1

Consider the following detection problem.
Under hypothesis H_0, Y is a Geometric$(1/2)$ random variable.
Under hypothesis H_1, Y is a Geometric$(3/4)$ random variable.
The probabilities of the hypotheses are $\mathbb{P}[H_0] = 1/3$ and $\mathbb{P}[H_1] = 2/3$.

(c) Determine the MAP rule. For the MAP rule, we need to compare the likelihood ratio $L(y)$ to $\frac{\mathbb{P}[H_0]}{\mathbb{P}[H_1]} = \frac{1/3}{2/3} = \frac{1}{2}$. We find that $L(y)$ is greater than $\frac{1}{2}$ for $y = 1, 2$ and $L(y)$ is less than $\frac{1}{2}$ for $y = 3, 4, \ldots$. Therefore, the MAP rule is to decide H_1 for $y = 1, 2$ and decide H_0 for $y \geq 3$.

Consider the following detection problem.
Under hypothesis H_0, Y is a Geometric$(1/2)$ random variable.
Under hypothesis H_1, Y is a Geometric$(3/4)$ random variable.
The probabilities of the hypotheses are $\mathbb{P}[H_0] = 1/3$ and $\mathbb{P}[H_1] = 2/3$.

(d) Determine the probability of error under the MAP rule. For the MAP rule, we have $A_0 = \{Y \geq 3\}$ and $A_1 = \{Y \leq 2\}$. Therefore, the probability of error is

$$\begin{aligned}
P_{\mathsf{FA}} &= \mathbb{P}[Y \in A_1 | H_0] = \mathbb{P}[Y \leq 2 | H_0] \\
&= P_{Y|H_0}(1) + P_{Y|H_0}(2) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4} \\
P_{\mathsf{MD}} &= \mathbb{P}[Y \in A_0 | H_1] = \mathbb{P}[Y \geq 3 | H_1] = 1 - \mathbb{P}[Y \leq 2 | H_1] \\
&= 1 - P_{Y|H_1}(1) - P_{Y|H_1}(2) = 1 - \frac{3}{4} - \frac{3}{16} = \frac{1}{16} \\
\mathbb{P}[\text{error}] &= P_{\mathsf{FA}}\, \mathbb{P}[H_0] + P_{\mathsf{MD}}\, \mathbb{P}[H_1] = \frac{3}{4} \cdot \frac{1}{3} + \frac{1}{16} \cdot \frac{2}{3} = \frac{7}{24}\ .
\end{aligned}$$

Practice Question #2

Consider the following estimation problem. The joint PDF of X and Y is

$$f_{X,Y}(x,y) = \begin{cases} x - y & 1 \le x \le 2,\ 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

(a) What are the marginal PDFs $f_X(x)$ and $f_Y(y)$?

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \begin{cases} \int_0^1 (x-y)\,dy & 1 \le x \le 2 \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \left(xy - \frac{y^2}{2}\right)\Big|_0^1 & 1 \le x \le 2 \\ 0 & \text{otherwise.} \end{cases} = \begin{cases} x - \frac{1}{2} & 1 \le x \le 2 \\ 0 & \text{otherwise.} \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \begin{cases} \int_1^2 (x-y)\,dx & 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \left(\frac{x^2 y}{2} - xy\right)\Big|_1^2 & 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases} = \begin{cases} \frac{3}{2} - y & 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Practice Question #2

Consider the following estimation problem. The joint PDF of X and Y is

$$f_{X,Y}(x,y) = \begin{cases} x - y & 1 \leq x \leq 2,\ 0 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(b) What is the conditional PDF $f_{X|Y}(x|y)$?

$$f_{X|Y}(x|y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{f_Y(y)} & f_Y(y) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \dfrac{x - y}{\frac{3}{2} - y} & 1 \leq x \leq 2,\ 0 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(Remember that the range of conditional PDF will be the same as the range of the joint PDF!)

Practice Question #2

Consider the following estimation problem. The joint PDF of X and Y is

$$f_{X,Y}(x,y) = \begin{cases} x - y & 1 \le x \le 2,\ 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

(c) What is the MMSE estimator of X given $Y = y$?
Remember that the MMSE estimator is just the conditional expectation!

$$\hat{x}_{\mathsf{MMSE}}(y) = \mathbb{E}[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx$$

$$= \int_1^2 x \frac{x-y}{\frac{3}{2}-y}\,dx = \left(\frac{\frac{x^3}{3} - \frac{x^2 y}{2}}{\frac{3}{2}-y} \right)\Bigg|_1^2 = \frac{14-9y}{9-6y}$$

Consider the following estimation problem. The joint PDF of X and Y is

$$f_{X,Y}(x,y) = \begin{cases} x - y & 1 \leq x \leq 2,\ 0 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(d) What is the LLSE estimator of X given $Y = y$?
Remember that the LLSE estimator is a linear function with slope and offset determined through calculating certain means, variances, and covariances. We will use the formula $\hat{x}_{\text{LLSE}}(y) = \mathbb{E}[X] + \dfrac{\text{Cov}[X,Y]}{\text{Var}[Y]}(y - \mathbb{E}[Y])$. Below, we calculate the necessary integrals.

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dy = \int_1^2 x\left(x - \frac{1}{2}\right) dx = \left(\frac{x^3}{3} - \frac{x^2}{4}\right)\Bigg|_2^1 = \frac{19}{12}$$

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y)\, dy = \int_0^1 y\left(\frac{3}{2} - y\right) dy = \left(\frac{3y^2}{4} - \frac{y^3}{3}\right)\Bigg|_0^1 = \frac{5}{12}$$

Practice Question #2

Consider the following estimation problem. The joint PDF of X and Y is

$$f_{X,Y}(x,y) = \begin{cases} x - y & 1 \le x \le 2,\ 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

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

$$\mathbb{E}[Y^2] = \int_{-\infty}^{\infty} y^2 f_Y(y)\,dy = \int_0^1 y^2\left(\frac{3}{2} - y\right) dy = \left(\frac{y^3}{2} - \frac{y^4}{4}\right)\Big|_0^1 = \frac{1}{4}$$

$$\text{Var}[Y] = \mathbb{E}[Y^2] - \big(\mathbb{E}[Y]\big)^2 = \frac{1}{4} - \frac{25}{144} = \frac{11}{144}$$

$$\mathbb{E}[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\,dy$$

$$= \int_1^2 \int_0^1 xy(x-y)\,dy\,dx = \int_1^2 \left(\frac{x^2\,y^2}{2} - \frac{xy^3}{3}\right)_0^1 dx$$

$$= \int_1^2 \left(\frac{x^2}{2} - \frac{x}{3}\right) dx = \left(\frac{x^3}{6} - \frac{x^2}{6}\right)\Big|_1^2 = \frac{2}{3}$$

Practice Question #2

Consider the following estimation problem. The joint PDF of X and Y is

$$f_{X,Y}(x,y) = \begin{cases} x - y & 1 \le x \le 2,\ 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

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

$$\text{Cov}[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \frac{2}{3} - \frac{19}{12} \cdot \frac{5}{12} = \frac{1}{144}$$

$$\hat{x}_{\text{LLSE}}(y) = \frac{19}{12} + \frac{\frac{1}{144}}{\frac{11}{144}}\left(y - \frac{5}{12}\right) = \frac{y}{11} + \frac{17}{11}$$

Practice Question #3

Let Y be a random variable with $\mathbb{E}[Y] = 2$ and $\mathbb{E}[Y^2] = 5$. Let Z be a random variable with $\mathbb{E}[Z] = -1$ and $\mathbb{E}[Z^2] = 3$. Let $\rho_{Y,Z} = -\frac{1}{\sqrt{2}}$ and define $X = 3Y + Z$.

(a) Determine the mean of X.

$$\mathbb{E}[X] = \mathbb{E}[3Y + Z] = 3\mathbb{E}[Y] + \mathbb{E}[Z] = 3 \cdot 2 + (-1) = 5\ .$$

(b) Determine the variance of X. We can use the formula $\mathrm{Var}[aY + bZ] = a^2\,\mathrm{Var}[Y] + b^2\,\mathrm{Var}[Z] + 2ab\,\mathrm{Cov}[Y, Z]$.

$$\begin{aligned}
\mathrm{Var}[Y] &= \mathbb{E}[Y^2] - \big(\mathbb{E}[Y]\big)^2 = 5 - 2^2 = 1 \\
\mathrm{Var}[Z] &= \mathbb{E}[Z^2] - \big(\mathbb{E}[Z]\big)^2 = 3 - (-1)^2 = 2 \\
\mathrm{Cov}[Y, Z] &= \rho_{Y,Z}\sqrt{\mathrm{Var}[Y] \cdot \mathrm{Var}[Z]} = -\frac{1}{\sqrt{2}} \cdot \sqrt{1 \cdot 2} = -1 \\
\mathrm{Var}[X] &= \mathrm{Var}[3Y + Z] = 3^2\,\mathrm{Var}[Y] + 1^2\,\mathrm{Var}[Z] + 2 \cdot 3 \cdot 1\,\mathrm{Cov}[Y, Z] \\
&= 9 \cdot 1 + 1 \cdot 2 + 6 \cdot (-1) = 5\ .
\end{aligned}$$

Practice Question #3

Let Y be a random variable with $\mathbb{E}[Y] = 2$ and $\mathbb{E}[Y^2] = 5$. Let Z be a random variable with $\mathbb{E}[Z] = -1$ and $\mathbb{E}[Z^2] = 3$. Let $\rho_{Y,Z} = -\frac{1}{\sqrt{2}}$ and define $X = 3Y + Z$.

(c) Let $X_1, \ldots, X_{500}$ be i.i.d. random variables with the same distribution as X. Using the Central Limit Theorem approximation, estimate the probability $\mathbb{P}\big[\big|\frac{1}{500}\sum_{i=1}^{500} X_i - \mathbb{E}[X]\big| > \frac{1}{2}\big]$. (You may leave your answer in terms of the Φ function.) Let

$W = \dfrac{1}{500}\displaystyle\sum_{i=1}^{500} X_i - \mathbb{E}[X]$ and note that $\mathbb{E}[W] = 0$ and

$\mathrm{Var}[W] = \dfrac{1}{500}\mathrm{Var}[X] = \dfrac{5}{500} = \dfrac{1}{100}$. Therefore,

$$\mathbb{P}[|W| > \frac{1}{2}] = \mathbb{P}\Big[W > \frac{1}{2}\Big] + \mathbb{P}\Big[W < -\frac{1}{2}\Big] = 1 - F_W\Big(\frac{1}{2}\Big) + F_W\Big(-\frac{1}{2}\Big)$$

$$\approx 1 - \Phi\left(\frac{1/2 - 0}{1/10}\right) + \Phi\left(\frac{-1/2 - 0}{1/10}\right)$$

$$= 1 - \Phi(5) + \Phi(-5) = 2\Phi(-5)$$

Practice Question #4

You are trying out a new blood pressure drug with a control group and an experimental group, each of consisting of 400 samples. The variance is believed to be $\sigma_1^2 = 0.40$ in the control group and $\sigma_2^2 = 0.60$ in the experimental group. For the control group, you obtain sample mean $M_{400}^{(1)} = 2.10$ and for the experimental group you obtain sample mean $M_{400}^{(2)} = 2.02$.

(a) What is the variance of the sample mean Var$[M_{400}^{(1)}]$?

$$\text{Var}[M_{400}^{(1)}] = \frac{1}{400}(0.40) = 0.001$$

(b) Do the groups have different means at a significance level of 0.05? Since the variances are known, a two-sample Z-test is appropriate. The Z-statistic is

$$Z = \frac{(M_n^{(1)} - M_n^{(2)})}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} = \frac{20 \cdot (2.10 - 2.02)}{\sqrt{1}} = 20 \cdot 0.08 = 1.6$$

The p-value is $2\Phi(-|Z|) = 2\Phi(-1.6) = 0.1$ which exceeds the significance level 0.05. Thus, we fail to reject the null hypothesis.

You are trying out a new blood pressure drug with a control group and an experimental group, each of consisting of 400 samples. The variance is believed to be $\sigma_1^2 = 0.40$ in the control group and $\sigma_2^2 = 0.60$ in the experimental group. For the control group, you obtain sample mean $M_{400}^{(1)} = 2.10$ and for the experimental group you obtain sample mean $M_{400}^{(2)} = 2.02$.

(c) Construct a confidence interval for the mean of the control group with confidence level 0.9. First, select γ such that $Q(\gamma) = \alpha/2 = 0.05 \implies \gamma = 1.6$.

Since the variance is known,

$$\left[M_n^{(1)} \pm \frac{\gamma\sigma}{\sqrt{n}}\right] = \left[2.10 \pm \frac{1.6 \cdot \sqrt{0.40}}{20}\right]$$

Practice Question #5

In this problem, you will work through the process of constructing and evaluating an LDA binary classifier by hand. You have been given the following 1-dimensional training and test datasets:

$$\mathbf{X}_{\text{train}} = \begin{bmatrix} +2 \\ 0 \\ -1 \\ -3 \end{bmatrix} \quad \underline{Y}_{\text{train}} = \begin{bmatrix} +1 \\ +1 \\ -1 \\ -1 \end{bmatrix} \quad \mathbf{X}_{\text{test}} = \begin{bmatrix} +4 \\ 0 \end{bmatrix} \quad \underline{Y}_{\text{test}} = \begin{bmatrix} +1 \\ -1 \end{bmatrix}$$

(a) Compute the sample means $\hat{\mu}_+$ and $\hat{\mu}_-$ as well as the sample covariance matrix $\hat{\mathbf{\Sigma}}$, which in this 1-dimensional setting is just a sample variance (and could be denoted by $\hat{\sigma}^2$ instead if you wish).

$$\hat{\mu}_+ = \frac{1}{2}(+2+0) = +1 \qquad \hat{\mu}_- = \frac{1}{2}(-1-3) = -2$$

$$\hat{\mathbf{\Sigma}}_+ = (2-1)^2 + (0-1)^2 = 2$$

$$\hat{\mathbf{\Sigma}}_- = ((-1)-(-2))^2 + (-3-(-2))^2 = 2$$

$$\hat{\mathbf{\Sigma}} = \frac{1}{4-2}((2-1)\hat{\mathbf{\Sigma}}_+ + (2-1)\hat{\mathbf{\Sigma}}_-) = 2$$

Practice Question #5

(b) Work out the LDA classifier. Try to simplify the expression as much as you can. Show your work for full credit.

$$D_{\mathsf{LDA}}(x) = \begin{cases} +1 & 2(\hat{\mu}_+ - \hat{\mu}_-)\hat{\mathbf{\Sigma}}^{-1}x \geq \hat{\mu}_+\hat{\mathbf{\Sigma}}^{-1}\hat{\mu}_+ - \hat{\mu}_-\hat{\mathbf{\Sigma}}^{-1}\hat{\mu}_- \\ -1 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} +1 & 2(+1-(-2))\frac{1}{2}x \geq 1 \cdot \frac{1}{2} \cdot 1 - (-2) \cdot \frac{1}{2} \cdot (-2) \\ -1 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} +1 & x \geq -\frac{1}{2} \\ -1 & \text{otherwise.} \end{cases}$$

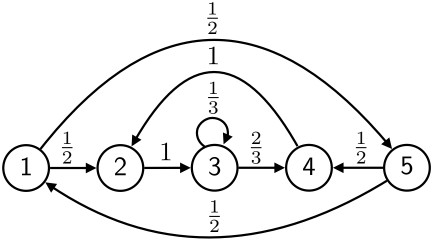
(c) Calculate the LDA training and test error rates.

$$\underline{Y}_{\mathsf{train,guess}} = \begin{bmatrix} +1 \\ +1 \\ -1 \\ -1 \end{bmatrix} \qquad \underline{Y}_{\mathsf{test,guess}} = \begin{bmatrix} +1 \\ +1 \end{bmatrix}$$

Training Error Rate is 0% and Test Error Rate is 50%.

Practice Question #6

Consider the following Markov chain



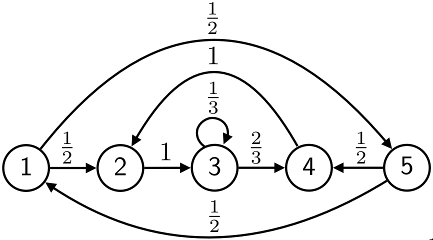
(a) Determine the communicating classes.
$C_1 = \{1, 5\}$ and $C_2 = \{2, 3, 4\}$.

(b) Determine which states are transient and which are recurrent.
States 1 and 5 are transient and states 2, 3, and 4 are recurrent.

(c) Determine the period of each state.
State 3 has a self-cycle and thus has period 1. All states in its communicating class have the same period so states 2 and 4 have period 1 as well. States 1 and 5 have period 2.

Practice Question #6

Consider the following Markov chain



(d) Write down the state transition matrix. $\mathbf{P} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \end{bmatrix}$

(e) Does the Markov chain have a unique limiting probability state vector $\underline{\pi}$?
Yes, even though it is not irreducible, it has only a single recurrent communicating class. This class is aperiodic. Therefore, it has a unique limiting probability state vector where the probabilities of the transient states are set to 0.

(f) Solve for the unique limiting probability state vector $\underline{\pi}$. Since states 1 and 5 are transient, we know that $\pi_1 = \pi_5 = 0$. From the steady-state equation $\mathbf{P}^\mathsf{T}\underline{\pi} = \underline{\pi}$, we get

$$\pi_4 = \pi_2$$
$$\pi_2 + \frac{1}{3}\pi_3 = \pi_3 \implies \pi_3 = \frac{3}{2}\pi_2 \; .$$

Plugging these into the normalization equation, we get

$$\sum_{j=1}^{5} \pi_j = \pi_2 + \frac{3}{2}\pi_2 + \pi_2 = \frac{7}{2}\pi_2 = 1 \implies \pi_2 = \frac{2}{7} \; .$$

Substituting back in, we get $\pi_4 = \dfrac{2}{7}$ and $\pi_3 = \dfrac{3}{7}$ so $\underline{\pi} = \begin{bmatrix} 0 \\ 2/7 \\ 3/7 \\ 2/7 \\ 0 \end{bmatrix}$.