1. ***Binary vector channel***. The output of a binary vector channel with multiplicative noise is is

$$\boldsymbol{X} = \begin{bmatrix} X[0] \\ \dots \\ X[N-1] \end{bmatrix} = \begin{bmatrix} S[0]\,W[0] \\ \dots \\ S[N-1]\,W[N-1] \end{bmatrix}$$

where

$$\boldsymbol{S} = \begin{bmatrix} S[0] \\ \dots \\ S[N-1] \end{bmatrix}$$

and

$$\boldsymbol{W} = \begin{bmatrix} W[0] \\ \dots \\ W[N-1] \end{bmatrix}$$

are the signal and multiplicative noise, respectively. Here,

- $\boldsymbol{S}$ and $\boldsymbol{W}$ are mutually independent,
- $W[n]$, $n = 0, 1, \dots, N-1$ are independent, identically distributed (i.i.d.) Bernoulli random variables with

$$W[n] = \begin{cases} 1, & \text{with probability } \epsilon \\ 0, & \text{with probability } 1-\epsilon \end{cases}$$

  where $0.5 < \epsilon < 1$ is a known constant, and

- Given $\Theta = \theta$, $S[n]$, $n = 0, 1, \dots, N-1$ are i.i.d. Bernoulli random variables with

$$\{S[n] \mid \Theta = \theta\} = \begin{cases} 1, & \text{with probability } \theta \\ 0, & \text{with probability } 1-\theta \end{cases}$$

  where θ is the unknown parameter.

We wish to test

$$\begin{aligned} \mathcal{H}_0 : &\quad \Theta = \theta_0 = 0 \quad \text{versus} \\ \mathcal{H}_1 : &\quad \Theta = \theta_1 = 0.5 \quad \text{versus} \\ \mathcal{H}_2 : &\quad \Theta = \theta_2 = 1. \end{aligned}$$

Assume equiprobable hypotheses:

$$\pi_0 = \pi_1 = \pi_2 = \frac{1}{3}.$$

where

$$\pi_i = p_\Theta(\theta_i), \quad i = 0, 1, 2$$

describes the prior probability mass function (pmf) of Θ. Find the Bayes' decision rule $\phi(\boldsymbol{X})$ for testing the above three hypotheses; express it in terms of ϵ, N, and an integer-valued sufficient statistic $T(\boldsymbol{X})$.

2. ***Binary communication with erasure in an additive white Gaussian noise channel***. Given the state $S = s \in \{0, 1\}$, the measurement $X = x$ follows

$$f_{X \mid S}(x \mid s) = \mathcal{N}(x \mid (-1)^{s+1}, \sigma^2)$$

i.e. $f_{X \mid S}(x \mid 0) = \mathcal{N}(x \mid -1, \sigma^2)$ and $f_{X \mid S}(x \mid 1) = \mathcal{N}(x \mid 1, \sigma^2)$. Assume that the two states are equiprobable:

$$p_S(0) = p_S(1) = \tfrac{1}{2}.$$

We design a decision rule $\phi(x) : \mathcal{X} \to (0, 1, 2)$:

$$\phi(x) = \begin{cases} 0, & \text{decide } S = 0 \\ 1, & \text{decide } S = 1 \\ 2, & \text{erase the signal} \end{cases}$$

and adopt the piecewise-constant cost structure with $\mathrm{L}(0 \mid 1) = \mathrm{L}(1 \mid 0) = 1$, $\mathrm{L}(0 \mid 0) = \mathrm{L}(1 \mid 1) = 0$, and $\mathrm{L}(2 \mid 0) = \mathrm{L}(2 \mid 1) = c$, where $c \geq 0$ is a known constant quantifying the cost of erasure.

(a) Compute the posterior pmf of S, i.e. find $p_{S \mid X}(0 \mid x)$ and $p_{S \mid X}(1 \mid x)$.

(b) Assume that $c < 0.5$. Show that the Bayes' decision rule for this problem has the form:

$$\phi(x) = \begin{cases} 0, & x \leq -\tau \\ 1, & x \geq \tau \\ 2, & -\tau < x < \tau \end{cases}$$

and determine τ.

Hint: Sketch the posterior expected losses $\rho_0(x), \rho_1(x)$, and $\rho_2(x)$.

(c) Find the Bayes' decision rule for the case where $c \geq 0.5$.

3. ***An EM Algorithm for Mixture-density Estimation***. A mixture density is a density of the form

$$f_{X \,|\, \boldsymbol{\theta}, \boldsymbol{\varphi}}(x \,|\, \boldsymbol{\theta}, \boldsymbol{\varphi}) = \sum_{i=1}^{m} \varphi_i \, q_i(x \,|\, \theta_i) \tag{1}$$

which models data from a statistical population that is a mixture of m component densities $q_i(x; \theta_i)$ and mixing proportions φ_i. (For example, think of the distribution of weights of a human population, which is a mixture of the weights of males and weights of females.) In (1), the mixing proportions (probabilities) satisfy $\sum_{i=1}^{m} \varphi_i = 1$ and $\varphi_i \geq 0$. We wish to determine both the individual density parameters $\theta_i, \; i = 1, 2, \ldots, m$ and the mixture parameters $\varphi_i, \; i = 1, 2, \ldots, m$. The vectors of the unknown parameters are

$$\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_m]^T$$

and

$$\boldsymbol{\varphi} = [\varphi_1 . \varphi_2, \ldots, \varphi_m]^T.$$

We collected a set of measurements $X[0] = x[0], X[1] = x[1], \ldots, X[N-1] = x[N-1]$. Assume that $X[0], X[1], \ldots, X[N-1]$ are conditionally i.i.d. given $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ and follow (1). If the data $\boldsymbol{x} = [x[0], x[1], \ldots, x[N-1]]^T$ were labeled according to which distribution generated $x[n]$, then the parameters of the distribution $q_i(x \,|\, \theta_i)$ could be estimated based on the data $x[n]$ associated with it. In many applications, however, the data are not labeled.

Let $\boldsymbol{U} = [U[0], U[1], \ldots, U[N-1]]^T$ be the *unobserved* set of labels (i.e. unobserved data), where

$$U[n] \in \{1, 2, \ldots, m\}, \quad n = 0, 1, \ldots, N-1.$$

We assume that $U[0], U[1], \ldots, U[N-1]$ are conditionally i.i.d. with pmf

$$\varphi_i = \Pr\{U[n] = i\}, \quad i = 1, 2, \ldots, m.$$

Derive an expectation-maximization (EM) algorithm for estimating $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$. Assume that, upon completion of the pth EM step, we obtain the estimates $\boldsymbol{\theta}^{(p)}, \boldsymbol{\varphi}^{(p)}$. Outline of the derivation of the $(p+1)$st EM step:

(a) Show that the marginal log-likelihood function of $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ is

$$L(\boldsymbol{\theta}, \boldsymbol{\varphi}) = \sum_{n=0}^{N-1} \ln f_{X \,|\, \boldsymbol{\theta}, \boldsymbol{\varphi}}(x[n] \,|\, \boldsymbol{\theta}, \boldsymbol{\varphi}).$$

(b) Show that $Q(\boldsymbol{\theta}, \boldsymbol{\varphi} \,|\, \boldsymbol{\theta}^{(p)}, \boldsymbol{\varphi}^{(p)})$ computed in the E step is

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\varphi} \,|\, \boldsymbol{\theta}^{(p)}, \boldsymbol{\varphi}^{(p)}) &= \sum_{i=1}^{m} \sum_{n=0}^{N-1} \ln[\varphi_i \, q_i(x[n] \,|\, \theta_i)] \cdot \frac{\varphi_i^{(p)} \, q_i(x[n] \,|\, \theta_i^{(p)})}{f_{X \,|\, \boldsymbol{\theta}, \boldsymbol{\varphi}}(x[n] \,|\, \boldsymbol{\theta}^{(p)}, \boldsymbol{\varphi}^{(p)})} \\ &= \sum_{i=1}^{m} \left[\sum_{n=0}^{N-1} \frac{\varphi_i^{(p)} \, q_i(x[n] \,|\, \theta_i^{(p)})}{f_{X \,|\, \boldsymbol{\theta}, \boldsymbol{\varphi}}(x[n] \,|\, \boldsymbol{\theta}^{(p)}, \boldsymbol{\varphi}^{(p)})} \right] \cdot \ln \varphi_i \\ &\quad + \sum_{i=1}^{m} \sum_{n=0}^{N-1} \ln[q_i(x[n] \,|\, \theta_i)] \, \frac{\varphi_i^{(p)} \, q_i(x[n] \,|\, \theta_i^{(p)})}{f_{X \,|\, \boldsymbol{\theta}, \boldsymbol{\varphi}}(x[n] \,|\, \boldsymbol{\theta}^{(p)}, \boldsymbol{\varphi}^{(p)})} \end{aligned}$$

(c) Show that the M step for φ_i consists of computing new estimates of φ_is as follows:

$$\varphi_i^{(p+1)} = \frac{1}{N} \sum_{n=0}^{N-1} \frac{\varphi_i^{(p)} \, q_i(x[n] \,|\, \theta_i^{(p)})}{f_{X \,|\, \boldsymbol{\theta}, \boldsymbol{\varphi}}(x[n] \,|\, \boldsymbol{\theta}^{(p)}, \boldsymbol{\varphi}^{(p)})}, \quad i = 1, 2, \ldots, m.$$

Hint: Recall that $\sum_{i=1}^{m} \varphi_i = 1$ and $\varphi_i \geq 0$; these constraints should be incorporated into the estimation of φ_is. Use Lagrange multipliers.

(d) Show that the M step for θ_i consists of computing new estimates of θ_is as follows:

$$\theta_i^{(p+1)} = \arg\max_{\theta_i} \sum_{n=0}^{N-1} \ln[q_i(x[n] \,|\, \theta_i)] \cdot \frac{(\varphi_i)_p \, q_i(x[n] \,|\, \theta_i^{(p)})}{f_{X \,|\, \boldsymbol{\theta}, \boldsymbol{\varphi}}(x[n] \,|\, \boldsymbol{\theta}^{(p)}, \boldsymbol{\varphi}^{(p)})}, \quad i = 1, 2, \ldots, m.$$

4. Suppose that the measurement X is exponentially distributed given the rate parameter $\Theta = \theta$:

$$f_{X \mid \Theta}(x \mid \theta) = \text{Expon}(x \mid \theta).$$

We further assume that the prior distribution of Θ is $\text{Gamma}(\alpha, \beta)$:

$$\pi(\theta) = \text{Gamma}(\theta \mid \alpha, \beta).$$

(a) Suppose that we observe

$$Y = i_{[100,+\infty)}(X)$$

but not observe the exact value of X. What is the posterior pdf of Θ given that $X \geq 100$ (and, consequently, $Y = 1$):

$$f_{\Theta \mid Y}(\theta \mid 1)$$

as a function of α and β? Write down the posterior mean and variance of θ. *Hint:* Find the pmf of Y given $\Theta = \theta$.

(b) Suppose now that we are told that X is exactly 100. What is the posterior pdf

$$f_{\Theta \mid X}(\theta \mid 100)$$

now? What are the posterior mean and variance in this case?

(c) Explain why the posterior variance of Θ is higher in part (b) even though more information has been observed, compared with part (a). Why does this not contradict the iterated-expectations identity

$$\text{var}(X) = \text{E}_Y[\text{var}_{X \mid Y}(X \mid Y)] + \text{var}_Y(\text{E}_{X \mid Y}[X \mid Y])$$

or the corresponding inequality

$$\text{var}_X(X) \geq \text{E}_Y[\text{var}_{X \mid Y}(X \mid Y)]?$$