Final Exam - Moed A

Total time for the exam: 3 hours!

Important: For **True / False** questions, copy the statement to your notebook and write clearly true or false. You should prove the statement if true, and provide counterexample otherwise.

1) **Multipath Gaussian channel. (24 Points)**
Consider a Gaussian noise channel of power constraint P, where the signal takes two different paths and the received noisy signals, Y_1 and Y_2, are feed into a finite impulse response (FIR) filter which coherently combines the input signals, namely, $Y = a \cdot Y_1 + b \cdot Y_2$, where $a, b \in \mathbb{R}$. The system model is shown in the figure below.

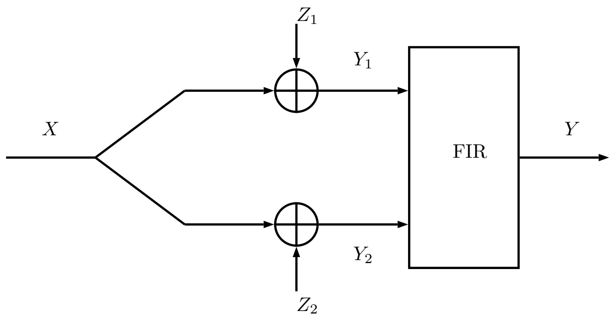


Fig. 1: Channel model

We assume that Z_1 and Z_2 are jointly Gaussian, with zero means, and covariance matrix

$$K = \begin{bmatrix} N & N\rho \\ N\rho & N \end{bmatrix}.$$

a) *Given* $a, b \in \mathbb{R}$, find the capacity $C(a, b, \rho)$ of the channel described above.
b) Evaluate your result in the previous item for $\rho = 0, -1,$ and 1? Explain your results when $a = b$.
c) What is the best filter in the sense of maximizing the capacity, i.e., solve $\max_{a,b} C(a, b, \rho)$, where the maximization is over all $a, b \in \mathbb{R}$. Explain your result. (You may use the inequality $\frac{(a+b)^2}{a^2+b^2+2ab\rho} \leq \frac{2}{1+\rho}$, for any $a, b \in \mathbb{R}$ and $\rho \in [-1, 1]$.)

2) **True or False on entropy identities (24 Points)**:
a) **True**/**False**: For any discrete random variables, $X_1, X_2,$ and X_3,
$$H(X_1, X_2, X_3) \leq \frac{1}{2}\left[H(X_1, X_2) + H(X_2, X_3) + H(X_3, X_1)\right].$$
b) **True**/**False**: For any discrete random variables, $X_1, X_2,$ and X_3,
$$H(X_1, X_2, X_3) \geq \frac{1}{2}\left[H(X_1, X_2|X_3) + H(X_2, X_3|X_1) + H(X_3, X_1|X_2)\right].$$
c) **True**/**False**: For two probability distributions, p_{XY} and q_{XY}, that are defined on $\mathcal{X} \times \mathcal{Y}$, the following holds:
$$D(p_{XY}||q_{XY}) \geq D(p_X||q_X).$$
d) Given are two channels with identical inputs and outputs alphabets ($|\mathcal{X}_i| = |\mathcal{Y}_i|$ for $i = 1, 2$). Their capacities are denoted by C_1 and C_2, respectively, and the capacity of their cascaded version is C_{12} or C_{21} depending on the channel ordering.
   i) **True**/**False**: If $|\mathcal{X}_i| = |\mathcal{Y}_i| = 2$ for all i, then $C_{12} = C_{21} = 0$ if and only if $C_1 = 0$ or $C_2 = 0$.
   ii) **True**/**False**: If $|\mathcal{X}_i| = |\mathcal{Y}_i| = 3$ for all i, then $C_{12} = 0$ if and only if $C_1 = 0$ or $C_2 = 0$.

3) **Shannon code (24 Points)**: Consider the following method for generating a code for a random variable X which takes on m values $\{1, 2, \dots, m\}$ with probabilities $p_1, p_2, \dots, p_m$. Assume that the probabilities are ordered so that $p_1 \geq p_2 \geq \cdots \geq p_m$. Define
$$F_i = \begin{cases} 0 & i = 1 \\ \sum_{k=1}^{i-1} p_k & i = 2, 3, \dots, m+1 \end{cases}, \quad (1)$$
namely, the sum of the probabilities of all symbols less than i. Then the codeword for i is the number $F_i \in [0, 1]$ rounded off to l_i bits, where $l_i = \lceil \log \frac{1}{p_i} \rceil$.
a) **True**/**False**
   i) The code constructed by this process is prefix-free.
   ii) The average length L satisfies $H(X) \leq L < H(X) + 1$.

b) Construct the code for the probability distribution $(0.5, 0.25, 0.125, 0.125)$.

c) **True/False** For dyadic distribution, i.e. $\forall i : p_i = 2^{-l_i}$ for some positive integer $l_i's$, the average length for this code matches $H(X)$.

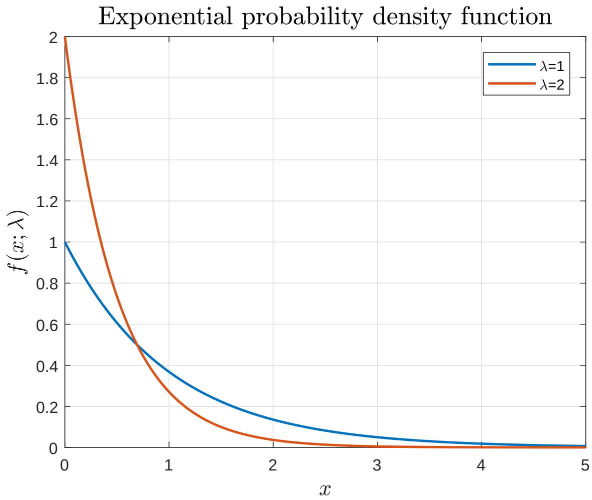
4) **Mixture of exponential distributions and EM (28 Points)**: A mixture of exponential distributions is defined by a vector of parameters $\lambda = [\lambda_1, \ldots, \lambda_k]$ and a latent discrete random variable Z which denotes the number of the exponential distributions hence $Z \in \{1, 2, ..., k\}$. The probability density function (pdf) of the mixture is:

$$P(x; \lambda) = \sum_{j=1}^{K} P(Z = j) f(x|z = j; \lambda_j) \tag{2}$$

where $f(x|z = j; \lambda_j) = \lambda_j \exp(-\lambda_j \cdot x)$, for $x \geq 0$, is the exponential pdf with parameter λ_j.

a) Assume $K = 1$, namely, $f(x; \lambda) = \lambda \exp(-\lambda x)$, for $x \geq 0$. Given a set of observations $\{x_i\}_{i=1}^n$ drawn i.i.d from $f(x; \lambda)$, find the maximum likelihood estimator (MLE) of λ.

b) The exponential distribution is used to model a continuous time interval between two Poisson events. The average rate of the Poisson events is denoted by λ, and the exponential distribution determines how much time will pass until the arrival of a new event.



The figure above has the pdf $f(x|z = j; \lambda_j)$ for $\lambda_j = 1$ and for $\lambda_j = 2$. Draw the mixture of these two distributions, where $P_Z(1) = P_Z(2) = 0.5$.

c) Recall from class that the EM algorithm aims to maximize the log-likelihood function: The E-step at iteration (t):

$$w(i, j) \triangleq Q_i^{(t)}(j) = P_{Z|X;\theta}\left(Z = j|X = x_i; \theta^{(t-1)}\right) \tag{3}$$

where $\theta^{(t-1)}$ are the parameters of the exponential mixture model at iteration $t - 1$.
The M-step at iteration (t):

$$\theta^{(t)} = \arg\max_{\theta} \sum_{i=1}^{n} \mathbb{E}_{Q_{Z_i}^{(t)}}[\log(P_{X,Z}(x_i, z_i; \theta))] \tag{4}$$

$$= \arg\max_{\theta} \sum_{i=1}^{n} \sum_{j=1}^{K} Q_i^{(t)}(j) \log P(x_i, z = j; \theta) \tag{5}$$

You are given training samples $\{x_i\}_{i=1}^n$ from a mixture of exponential distributions. Write explicitly the estimation formulas of the EM algorithm for exponential mixtures, using the training samples.

d) Assume that you have 10 agents in a company that provide services via telephone. Whenever an agent completes a service call, he fills a service form in the company's database and the duration of the call is recorded. After each call, the agent immediately answers a new call. The duration of a service call may depend on the costumer's need, the agent and, the day of the week. Hence, for each agent, these calls are modeled by a mixture of exponential distributions.

You are given a train data $\mathcal{S} = \{(x_i, y_i)\}_{i=1}^n$ where x_i is the duration of the i-th call and $y_i \in \{1, \ldots, 10\}$ is the agent number. You are also given a test data $\mathcal{T} = \{x_j\}_{j=1}^L$.

Write a pseudo code that fits a mixture of K exponential distributions for each agent in the company, using T iterations of expectation maximization. Then, it assigns an agent number for each one of the test samples.

Good Luck!