1. **Shannon-type inequalities** This question will show you a definition of a measure space on random variables, which we will use to define *Shannon-type inequalities.*

**This question is a preparation for the next question.**

The idea is simple: we wish to make a software that verifies if an information inequality holds, due to non-negativity of the entropies and mutual information.

We begin with 2 random variables $X_1$ and $X_2$ and treat them as sets $S_1$ and $S_2$, respectively.

![Figure 1: Relations between set - Venn diagram](image)

Let $F_2$ denote the set of all sets that are defined by the usual *set operations* on $S_1$ and $S_2$. Set operations include union, intersection, difference and complement. We define $S_1 \cap S_2^c \triangleq \emptyset$, where $\emptyset$ is the empty set.

(a) Find the minimal number of sets (and the sets themselves), defined by set operations on $S_1$ and $S_2$, such that:

- All the sets are disjoint.
- Every set in $F_2$ can be expressed by union these sets.

For instance, let $A \triangleq (S_1 \cap S_2)$ and $B \triangleq (S_1 \cap S_2^c)$.

It is easy to show that $S_1 = A \cup B$ and $A \cap B = \emptyset$. 

---

1
Use Fig. 2 to show that every information measure on $X_1$ and $X_2$ (or one of them) can be expressed as a linear combination of elements in $\mathcal{U}$, with non-negative coefficients, where

$$\mathcal{U} \triangleq \{H(X_1|X_2), I(X_1;X_2), H(X_2|X_1)\} \quad (1)$$

Conclude that the non-negativity of elements in $\mathcal{U}$ implies non-negativity of all information measures.

The set $\mathcal{U}$ is a set of all elemental information measures and the non-negativity inequalities are elemental inequalities.

We define a measure on $\mathcal{F}_2$, denoted by $\mu^*$, such that $\mu^*$ represent the information measure on the set $S_i$ that corresponds to the random variable $X_i$. For instance, $\mu^*(S_1) \rightarrow H(X_1)$ and $\mu^*(S_1 \cup S_2) \rightarrow H(X_1, X_2)$.

Thus, we interpret the set operation ”union” as ”comma” and $\mu^*$ as $H$ or $I$ (in this case it is $H$).

Find interpretation for the set operations ”intersection”, ”difference” and $\mu^*$.

Shannon-type inequalities are defined to be inequalities that holds due to the non-negativity of information measure.

We’ve seen (for two random variables) how to imply non-negativity of information measures by a small number of constraints, the elemental inequalities. In general (for any number of random variables) the elemental inequalities is a minimal set to imply the non-negativity of information measures.

2. **Inequalities prover in Matlab:** The goal of this question is to write a simple MATLAB program which verifies if an information theory inequality always holds. An information
inequality always holds, if it holds for all joint distributions on the the random variables that are involved. For instance, $H(X) \geq H(X|Y)$ always holds, since it holds for every $P_{X,Y}$.

It is very difficult (numerically) to substitute all joint distributions into an inequality and check if it holds, therefore, we derive a method using linear programming which can be implemented easily in Matlab.

Let $X_1, X_2, X_3$ be discrete random variables, and define

$$h = [H(X_1), H(X_2), H(X_1, X_2), H(X_1, X_3), H(X_2, X_3), H(X_1, X_2, X_3)]^\top$$

as a vector of joint entropies. Note that $h$ is a vector of labels and not values! We also define the vector $h(P_{X_1, X_2, X_3})$ containing the values of entropies in $h$, corresponding to some joint PMF $P_{X_1, X_2, X_3}$.

(a) We show in two steps that every expression including entropies and mutual information terms can be written as a linear combination of $h$:

i. For $\alpha, \beta \subseteq \{1, 2, 3\}$, show that

$$H(X_\alpha|X_\beta)$$

can be written as a linear combination of $h$. (For example, if $\alpha = \{1, 2\}$ then $X_\alpha = (X_1, X_2)$)

ii. Show that

$$I(X_\alpha; X_\beta|X_\gamma),$$

can also be written as a linear combination of $h$.

The representation by $h$ is unique and called canonical form.

(b) Find a vector $f$, such that $f^\top h = I(X_1; X_2, X_3) - I(X_1; X_2|X_3)$.

(c) The following inequalities are called elemental inequalities of 3 discrete random variables:

$$H(X_1|X_2, X_3) \geq 0, I(X_1; X_2) \geq 0, I(X_1; X_2|X_3) \geq 0,$$

$$H(X_2|X_1, X_3) \geq 0, I(X_2; X_3) \geq 0, I(X_2; X_3|X_1) \geq 0,$$

$$H(X_3|X_1, X_2) \geq 0, I(X_1; X_3) \geq 0, I(X_1; X_3|X_2) \geq 0.$$
A linear programming optimization problem has the form:

\[
\begin{align*}
  x^* &= \min_x f^T x \\
  \text{Subject to } A x &\leq b
\end{align*}
\]

and can be solved using linear programming (LP) optimization methods.

i. Formulate the LP minimization problem to check if \( f^T h \geq 0 \) always holds, subject to the elemental information inequalities. Find \( A \) and \( b \) and explain what is \( f \).

ii. Explain why an information inequality expressed by \( f^T h \geq 0 \), always holds if the minimum value of \( f^T h(P_{X_1,X_2,X_3}) \) (the minimum is with respect to \( P_{X_1,X_2,X_3} \)) is zero.

iii. Assume that \( x^* < 0 \), what can we conclude in the inequality?

iv. Write MATLAB code (use the command \texttt{linprog(f,A,b)}) to prove that

\[
\begin{align*}
  I(X_1;X_3|X_2) - I(X_1;X_3) &\geq 0, \\
  H(X_1)+H(X_2)-H(X_1,X_2) &\geq 0, \\
  H(X_1,X_2)-H(X_1|X_2) &\geq 0.
\end{align*}
\]

A Markov chain \( X_1 - X_2 - X_3 \) holds if and only if \( I(X_1;X_3|X_2) = 0 \).

i. Find a vector \( q \) such that

\[
q^T h = I(X_1;X_3|X_2)
\]

ii. Linear optimization methods (as in 8) can also account for equality constraints:

\[
\begin{align*}
  x^* &= \min_x f^T x \\
  \text{Subject to: } A x &\leq b \\
  C x &= d
\end{align*}
\]

where \( C \) is a matrix with coefficients, that represents the equality constraints.

What is the matrix \( C \) and vector \( d \) that represents the Markov chain \( X_1 - X_2 - X_3 \)?

iii. Write a MATLAB code (use the command \texttt{linprog(f,A,b,C,d)}) to proves that

\[
I(X_1;X_2) - I(X_1;X_3) \geq 0.
\]